

1 **Title**

2 Full-length isoform sequencing for resolving the molecular basis of Charcot-Marie-Tooth 2A

3

4

5 **Authors**

6 Andrew B. Stergachis, MD PhD<sup>1,2,3,\*</sup>, Elizabeth E. Blue, PhD<sup>1,3,4\*</sup>, Madelyn A Gillentine, PhD<sup>5</sup>, Lee-kai  
7 Wang, BS<sup>6</sup>, Ulrike Schwarze, MD<sup>7</sup>, Adriana Sedeño Cortés, MS<sup>1</sup>, Jane Ranchalis, BS<sup>1</sup>, Aimee Allworth,  
8 MS<sup>1</sup>, Austin E. Bland, MS<sup>1</sup>, Sirisak Chanprasert, MD<sup>1</sup>, Jingheng Chen, MPH BS<sup>4</sup>, Daniel Doherty, MD  
9 PhD<sup>3,8</sup>, Andrew B. Folta, BA<sup>1</sup>, Ian Glass, MD<sup>3,8</sup>, Martha Horike-Pyne, MPH<sup>1</sup>, Alden Y. Huang, PhD<sup>6</sup>, Alyna T.  
10 Khan, BA<sup>4,9</sup>, Kathleen A. Leppig, MD<sup>10</sup>, Danny E. Miller, MD PhD<sup>3,7,8</sup>, Ghayda Mirzaa, MD<sup>3,8,11</sup>, Azma  
11 Parhin, MD<sup>1</sup>, Wendy Raskind, MD PhD<sup>1</sup>, Elisabeth A. Rosenthal, PhD<sup>1</sup>, Sam Sheppeard, BS<sup>1</sup>, Samuel  
12 Strohbehn, BS<sup>1</sup>, Virginia P. Sybert, MD<sup>1</sup>, Thao T. Tran, BS<sup>7</sup>, Mark Wener, MD<sup>7</sup>, University of Washington  
13 Center for Mendelian Genomics (UW-CMG), Undiagnosed Diseases Network (UDN), Peter H. Byers,  
14 MD<sup>1,7</sup>, Stanley F. Nelson, MD<sup>6</sup>, Michael J. Bamshad, MD<sup>3,8</sup>, Katrina M. Dipple, MD PhD<sup>3,8</sup>, Gail P. Jarvik,  
15 MD PhD<sup>1,2,3</sup>, Suzanne Hoppins, PhD<sup>12</sup>, & Fuki M. Hisama, MD<sup>1,3,\*</sup>,  
16  
17

16

17

18 **Affiliations**

- 19 1. University of Washington School of Medicine, Department of Medicine, Seattle, WA, USA  
20 2. University of Washington School of Medicine, Genome Sciences, Seattle, WA, USA  
21 3. Brotman Baty Institute for Precision Medicine, Seattle, WA, USA  
22 4. University of Washington, Institute of Public Health Genetics, Seattle, WA, USA  
23 5. Seattle Children's Hospital, Department of Laboratories, Seattle, WA, USA  
24 6. Institute for Precision Health, David Geffen School of Medicine, University of California Los  
25 Angeles, Los Angeles, CA  
26 7. University of Washington School of Medicine, Department of Laboratory Medicine and Pathology,  
27 Seattle, WA, USA  
28 8. University of Washington, Department of Pediatrics, Seattle, WA, USA  
29 9. University of Washington, Department of Biostatistics, Seattle, WA, USA  
30 10. Group Health Cooperative, Kaiser Permanente Washington, Seattle, WA, USA  
31 11. Seattle Children's Research Institute, Center for Integrative Brain Research, Seattle, WA, USA  
32 12. University of Washington School of Medicine, Department of Biochemistry, Seattle, WA, USA

33 ¥ Corresponding Authors

34 \* Equal contributions

35 **Abstract**

36 *Objectives:* Transcript sequencing of patient derived samples has been shown to improve the diagnostic  
37 yield for solving cases of likely Mendelian disorders, yet the added benefit of full-length long-read  
38 transcript sequencing is largely unexplored.

39  
40 *Methods:* We applied short-read and full-length isoform cDNA sequencing and mitochondrial functional  
41 studies to a patient-derived fibroblast cell line from an individual with neuropathy that previously lacked  
42 a molecular diagnosis.

43  
44 *Results:* We identified an intronic homozygous *MFN2* c.600-31T>G variant that disrupts a branch point  
45 critical for intron 6 splicing. Full-length long-read isoform cDNA sequencing after treatment with a  
46 nonsense-mediated mRNA decay (NMD) inhibitor revealed that this variant creates five distinct altered  
47 splicing transcripts. All five altered splicing transcripts have disrupted open reading frames and are  
48 subject to NMD. Furthermore, a patient-derived fibroblast line demonstrated abnormal lipid droplet  
49 formation, consistent with *MFN2* dysfunction. Although correctly spliced full-length *MFN2* transcripts  
50 are still produced, this branch point variant results in deficient *MFN2* protein levels and autosomal  
51 recessive Charcot-Marie-Tooth disease, axonal, type 2A (CMT2A).

52  
53 *Discussion:* This case highlights the utility of full-length isoform sequencing for characterizing the  
54 molecular mechanism of undiagnosed rare diseases and expands our understanding of the genetic basis  
55 for CMT2A.

56  
57 **Introduction**

58 Transcript sequencing is emerging as a powerful clinical tool, and recent studies report that transcript  
59 sequencing can increase diagnostic yield by 2-24% versus DNA sequencing alone when evaluating  
60 suspected Mendelian disorder cases<sup>1-5</sup>. These studies uniformly use short-read sequencing approaches  
61 to identify transcripts with aberrant expression levels or spliced products<sup>6</sup>. However, the full-length  
62 transcripts produced by these aberrantly spliced products are not typically evident using short-read  
63 sequencing, as exon skipping and alternative splice site usage within multi-intronic genes can create  
64 transcripts with potential dominant-negative or loss-of-function impacts. Distinguishing among these  
65 full-length transcript outcomes is important for appropriately evaluating conditions whereby distinct  
66 phenotypes and inheritance patterns are associated with dominant-negative or loss-of-function variants  
67 in the same gene.

68  
69 Charcot-Marie-Tooth 2A (CMT2A) is the most common subtype of CMT2 and is an axonal peripheral  
70 nerve disorder characterized by motor, sensory, or autonomic neuropathy. ~90% of CMT2A cases follow  
71 an autosomal dominant inheritance pattern associated with dominant-negative variants in mitofusin 2  
72 (*MFN2*)<sup>7</sup>. In contrast, autosomal recessive inheritance is associated with biallelic loss-of-function (LOF)  
73 *MFN2* variants, which typically do not result in a clinical phenotype in the heterozygous state. Splicing  
74 variants in *MFN2* can cause both dominant and recessive forms of CMT2A<sup>8-10</sup>, indicating the need to  
75 accurately identify the full-length transcript effect of novel splicing variants.

76  
77 Recent advances in highly accurate long-read, full-length transcript sequencing has the potential to aid  
78 in the evaluation of splice-site variants. We report a patient who was found by a combination of short-  
79 read and full-length transcript sequencing to have a homozygous branch point variant in *MFN2* that  
80 results in deficient *MFN2* protein levels via the creation of five distinct altered transcripts that are all  
81 subject in nonsense-mediated decay (NMD).

82

## 83 **Methods**

### 84 *Exome sequencing and analysis:*

85 Quad exome sequencing (proband, unaffected mother, unaffected brother, unaffected paternal-half-  
86 brother) was performed on DNA (Baylor College of Medicine) through the Undiagnosed Diseases  
87 Network (UDN). In addition to clinical exome analysis, researchers reprocessed the exome data and  
88 performed queries focused on UPD(1) and genes previously implicated in CMT and similar disorders.

89

### 90 *Short-read transcript/RNA sequencing and analysis:*

91 RNA extraction, library preparation, and short-read sequencing were performed on cultured skin  
92 fibroblasts from the proband as previously described<sup>5</sup>. A control dataset of short-read transcript/RNA  
93 sequencing from 236 skin fibroblast samples, generated at UCLA, was used to identify RNA expression  
94 outliers and aberrant splicing products using OUTRIDER<sup>11</sup> and IRFinder<sup>12</sup> respectively.

95

### 96 *Full length isoform long-read transcript sequencing and analysis:*

97 Cultured skin fibroblasts were grown in DMEM with 10% FBS and then incubated in the presence of  
98 cycloheximide (100µg/ml, Sigma-Aldrich) for 6 hours before RNA extraction using the RNeasy Mini kit  
99 (Qiagen). Complementary DNA (cDNA) synthesis was performed following the ISO-Seq protocol, which  
100 preserves 3' and 5' end information (PacBio, Menlo Park, CA). A PacBio SMRTbell library was  
101 constructed using these PCR-amplified full-length cDNA transcripts and sequenced using a Sequel II. Iso-  
102 Seq data was processed using the Iso-Seq3 pipeline, mapped to GRCh38, and visualized using IGV.

103

### 104 *Sanger validation:*

105 RNA was isolated as above, and cDNA was synthesized with random hexamers and SuperScript™ III  
106 reverse transcriptase (Invitrogen). The *MFN2* region of interest was amplified by PCR with a sense  
107 primer in exon 5 (5'-GCCATGAGGCCTTCTCCTT) and an antisense primer in exon 8 (5'-  
108 AGACGCTCACTCACCTTGTG). PCR products were separated on 7% polyacrylamide gel. Normal and all  
109 abnormal products were excised from the gel and DNA was retrieved by submersion of gel slices in  
110 100µl water at room temperature overnight. Eluted PCR products were amplified using the same  
111 primers in exons 5 and 8 and amplicons were subjected to Sanger sequencing.

112

### 113 *Lipid droplet analysis:*

114 Skin fibroblast cultures from the proband and an unrelated patient who does not harbor variants in  
115 *MFN2* were separately plated on glass bottom dishes (MatTek). After 48 hours of culture, cells were  
116 incubated with 0.1µg/ml Mitotracker Red CMX Ros (Molecular Probes), 5mM BODIPY 493/503  
117 (Invitrogen) and 3 drops of NucBlue (Invitrogen) for 15 min at 37°C with 5% CO<sub>2</sub>. Cells were  
118 subsequently moved into complete media for ≥45 minutes, then imaged using a Z-series step size of  
119 0.3µm on a Nikon Ti-E widefield microscope with a 63X NA 1.4 oil objective (Nikon), solid-state light  
120 source (Spectra X, Lumencor), and an sCMOS camera (Zyla 5.5 megapixel). Each line was imaged on  
121 three separate occasions by a blinded experimenter (n>100 cells per experiment). Images were  
122 deconvolved using 7 iterations of 3D Landweber deconvolution. The number and fluorescence intensity  
123 of lipid droplets on deconvolved images was quantified using Spot Detection Analysis (Nikon Elements).  
124 Maximum intensity projections were generated using ImageJ software (NIH). All quantification was  
125 performed by an experimenter blinded to sample identification.

126

### 127 *Standard Protocol Approvals, Registrations, and Patient Consents:*

128 This study was approved by the National Institutes of Health (NIH) Institutional Review Board (IRB) (IRB  
129 # 15HG0130), and written informed consent was obtained from all participants in the study.

130

131 **Results**

132 *Clinical phenotype associated with MFN2 deep intronic variant:*

133 We evaluated a 42-year-old woman who initially presented with abnormal “foot-slapping” gait at one  
134 year of age that progressed into distal leg weakness requiring a wheelchair by age 8. She underwent  
135 spinal fusion and Harrington rod placement for scoliosis in her teens and developed respiratory  
136 involvement in her thirties. She had normal cognitive development and no family history of  
137 neuromuscular disease. Electromyography and nerve conduction velocity studies at age 2 years revealed  
138 distal motor and sensory polyneuropathy, with positive waves and fibrillation. Nerve and muscle biopsy  
139 revealed marked denervation atrophy. Neurologic exam at age 42 showed a quadriparetic woman with  
140 normal facial strength, hypophonia, severe muscle wasting of arms and legs, and 1-2/5 proximal motor  
141 strength and 0/5 distal strength. Sensation was present but reduced to all modalities distally and  
142 reflexes were absent throughout.

143

144 *Identification of a deep intronic MFN2 variant using short-read transcript sequencing:*

145 Initial genetic evaluation revealed paternal uniparental isodisomy of chromosome 1 (UPD[1]), while  
146 panel testing for genes associated with neuromuscular disorders was non-diagnostic. She was enrolled  
147 into the UDN, where initial exome analysis did not identify a strong candidate variant. Short-read  
148 transcript sequencing of RNA isolated from a patient-derived fibroblast line identified *MFN2*, located on  
149 chromosome 1, as an expression outlier in comparison to sequencing data from control fibroblast lines  
150 (Z-score -6.9) (**Figure 1A**) with about 2-fold lower expression relative to control fibroblasts. In addition,  
151 *MFN2* exhibited increased retention of intron 6 (Z-score 8.6) (**Figure 1B**). Re-analysis of the exome data  
152 identified a homozygous *MFN2* c.600-31T>G variant within intron 6 that is absent from population  
153 databases and is predicted to disrupt the U nucleotide in the  $\gamma$ UnAy consensus branch point sequence<sup>13</sup>  
154 (**Figure 1B**).

155

156 *Identification of the splicing impact of an MFN2 branch point variant:*

157 Heterozygous LOF *MFN2* variants are not typically associated with disease, and this patient’s *MFN2*  
158 transcript level was only reduced to 51% of normal (**Figure 1A**). As branch-point variants can induce  
159 complex splicing alterations<sup>14</sup>, we performed full-length isoform sequencing (ISO-Seq) to determine the  
160 identity of all full-length *MFN2* spliced transcripts. ISO-Seq data from patient-derived cells treated with  
161 the NMD inhibitor cycloheximide revealed five distinct altered *MFN2* transcripts that each use a distinct  
162 splice acceptor site in lieu of the canonical exon 7 splice acceptor site (**Figure 1C**). Notably, all five  
163 altered splicing transcripts have disrupted open reading frames that make them subject to NMD, and  
164 none of them are present within control fibroblast cell lines (**Figure 1D**). Overall, these data  
165 demonstrate that this branch point variant does not create a significant amount of a stable abnormal  
166 protein but does substantially reduce the amount of normal protein.

167

168 *An MFN2 branch point variant causes insufficient MFN2 levels:*

169 To determine whether this branch point variant results in insufficient *MFN2* protein levels, we analyzed  
170 patient-derived fibroblast cells for hallmarks of *MFN2* dysfunction. *MFN2* is essential for mitochondrial  
171 dynamics, and pathogenic *MFN2* variants are associated with diverse mitochondrial phenotypes,  
172 including impaired mitochondrial respiration and movement, as well as increased lipid droplet  
173 formation<sup>15</sup>. We found that patient-derived fibroblast cells had both increased number and intensity of  
174 lipid droplets compared to control cells (**Figure 2**), which is consistent with the idea that this branch  
175 point variant results in insufficient functional *MFN2* protein.

176

177

178 **Discussion**

179 This study describes a pathogenic intronic branch point variant that alters splicing of intron 6 in both  
180 copies of *MFN2* sufficiently to alter the total functional capacity of the encoded protein – expanding our  
181 understanding of the molecular basis of CMT2A. Full-length transcript sequencing allowed us to identify  
182 and quantitate the outcomes of abnormal intron 6 splicing, exposing that this is a leaky variant that  
183 results in some normally spliced transcripts. Our data suggests that the overall lower transcript  
184 abundance is sufficient to produce this phenotype. This is consistent with a lack of symptoms in her  
185 father, who is heterozygous for this variant, and indicates that this variant results in an autosomal  
186 recessive LOF mechanism.

187

188 Recent advances in long-read sequencing technology and full-length transcript sequencing have the  
189 potential to transform clinical workflows for evaluating patients with unsolved likely Mendelian  
190 conditions. This study provides a proof-of-concept for the utility of full-length transcriptome data to  
191 identify disease-associated variants and to characterize the mechanism by which these variants cause  
192 disease. Further studies are needed to fully evaluate the utility of full-length transcript data in clinical  
193 practice.

194

195 **Study Funding**

196 A.B.S holds a Career Award for Medical Scientists from the Burroughs Wellcome Fund and is a Pew  
197 Biomedical Scholar. This study was supported by NIH grants 1U01HG010233, 1DP5OD029630,  
198 R01GM118509 and U01HG007703, in addition to funds from the Collagen Diagnostic Laboratory,  
199 University of Washington. Fibroblast data and analysis were partially supported by the California Center  
200 for Rare Diseases within the UCLA Institute of Precision Health. Sequence data analysis was supported by  
201 the University of Washington Center for Mendelian Genomics (UW-CMG), which was funded by NHGRI  
202 grant UM1 HG006493. The content is solely the responsibility of the authors and does not necessarily  
203 represent the official views of the National Institutes of Health.

204

205 **Disclosure**

206 The authors report no competing interests.

207

208 **References**

- 209 1. Cummings BB, Marshall JL, Tukiainen T, et al. Improving genetic diagnosis in Mendelian disease  
210 with transcriptome sequencing. *Sci Transl Med*. *Sci Transl Med*; 2017;9.
- 211 2. Gonorazky HD, Naumenko S, Ramani AK, et al. Expanding the Boundaries of RNA Sequencing as a  
212 Diagnostic Tool for Rare Mendelian Disease. *Am J Hum Genet*. *Am J Hum Genet*; 2019;104:466–  
213 483.
- 214 3. Frésard L, Smail C, Ferraro NM, et al. Identification of rare-disease genes using blood  
215 transcriptome sequencing and large control cohorts. *Nat Med*. *Nat Med*; 2019;25:911–919.
- 216 4. Kremer LS, Bader DM, Mertes C, et al. Genetic diagnosis of Mendelian disorders via RNA  
217 sequencing. *Nat Commun*. *Nat Commun*; 2017;8.
- 218 5. Lee H, Huang AY, Wang L kai, et al. Diagnostic utility of transcriptome sequencing for rare  
219 Mendelian diseases. *Genet Med*. *Genet Med*; 2020;22:490–499.
- 220 6. Lee H, Nelson SF. The frontiers of sequencing in undiagnosed neurodevelopmental diseases. *Curr*  
221 *Opin Genet Dev*. *Curr Opin Genet Dev*; 2020;65:76–83.
- 222 7. Züchner S, Mersiyanova I v., Muglia M, et al. Mutations in the mitochondrial GTPase mitofusin 2  
223 cause Charcot-Marie-Tooth neuropathy type 2A. *Nat Genet*. *Nat Genet*; 2004;36:449–451.

- 224 8. Park SY, Kim SY, Hong YH, Cho SI, Seong MW, Park SS. A novel double mutation in cis in MFN2  
225 causes Charcot-Marie-Tooth neuropathy type 2A. *Neurogenetics*. *Neurogenetics*; 2012;13:275–  
226 280.
- 227 9. Kotruchow K, Kabzińska D, Hausmanowa-Petrusewicz I, Kochański A. A late-onset and mild form  
228 of Charcot-Marie-Tooth disease type 2 caused by a novel splice-site mutation within the  
229 Mitofusin-2 gene. *Acta Myologica*. Pacini Editore; 2013;32:166.
- 230 10. Boaretto F, Vettori A, Casarin A, et al. Severe CMT type 2 with fatal encephalopathy associated  
231 with a novel MFN2 splicing mutation. *Neurology*. *Neurology*; 2010;74:1919–1921.
- 232 11. Brechtmann F, Mertes C, Matusevičiūtė A, et al. OUTRIDER: A Statistical Method for Detecting  
233 Aberrantly Expressed Genes in RNA Sequencing Data. *Am J Hum Genet*. *Am J Hum Genet*;  
234 2018;103:907–917.
- 235 12. Middleton R, Gao D, Thomas A, et al. IRFinder: assessing the impact of intron retention on  
236 mammalian gene expression. *Genome Biol*. *Genome Biol*; 2017;18.
- 237 13. Corvelo A, Hallegger M, Smith CWJ, Eyras E. Genome-wide association between branch point  
238 properties and alternative splicing. *PLoS Comput Biol*. *PLoS Comput Biol*; 2010;6.
- 239 14. Zhang P, Philippot Q, Ren W, et al. Genome-wide detection of human variants that disrupt  
240 intronic branchpoints. *Proc Natl Acad Sci U S A*. *Proc Natl Acad Sci U S A*; 2022;119.
- 241 15. Zaman M, Shutt TE. The Role of Impaired Mitochondrial Dynamics in MFN2-Mediated Pathology.  
242 *Front Cell Dev Biol*. *Front Cell Dev Biol*; 2022;10.  
243  
244

245 **Figure Legend**

246

247 **Figure 1 | Identification of a homozygous *MFN2* branch point variant that disrupts *MFN2* splicing.**

248 (A) Short-read RNA sequencing identified *MFN2* as an expression outlier in this patient's sample,  
249 exhibiting 51% of the RNA expression level seen in control fibroblast samples.

250 (B) Genetic testing of *MFN2* identified a deep intronic homozygous variant in intron 6 of *MFN2*. Short-  
251 read RNA sequencing identified that the intron retention ratio of intron 6 of *MFN2* was significantly  
252 abnormal compared to controls.

253 (C) Long-read full-length transcript sequencing (ISO-Seq) of this patient's sample after treatment with  
254 the non-sense mediated decay (NMD) inhibitor cycloheximide (CHX) identified 6 major *MFN2*  
255 transcripts. The predicted protein impact and transcript count of each are indicated to the right. Inset  
256 below shows the alternative splice acceptor sites used for each transcript, as well as the sequence  
257 context of the patient's variant relative to the canonical branch point sequence.

258 (D) DNA agarose gel showing altered spliced products affecting exon 7 of *MFN2* before and after CHX  
259 treatment, as well as in a control sample.

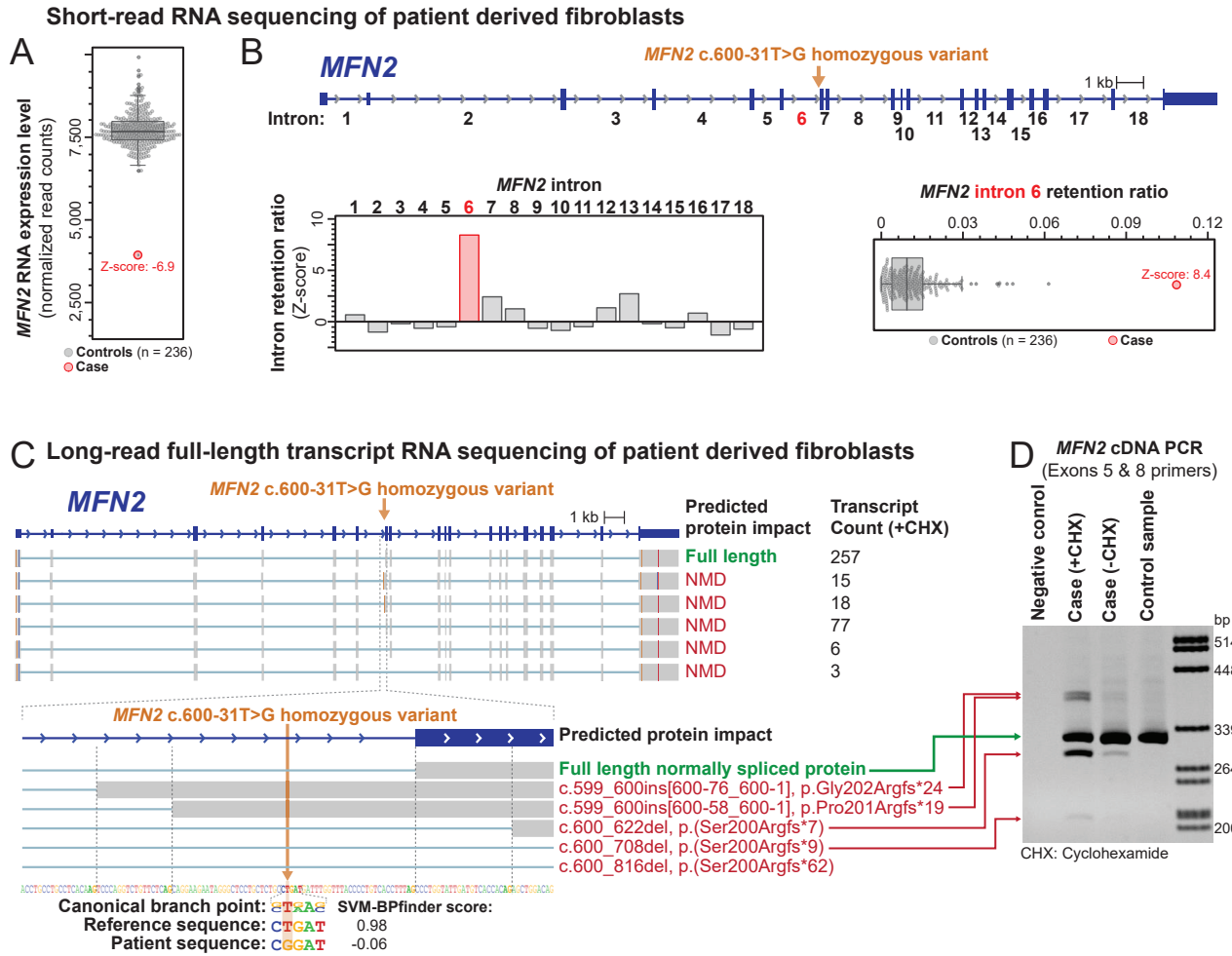
260

261 **Figure 2 | *MFN2* branch point variant results in abnormal lipid droplet formation.**

262 (A) Representative images of control and proband fibroblast cells. Mitochondria were labeled with  
263 Mitotracker CMXRos, lipid droplets with Bodipy 493/503, and nuclei with NucBlue. Images represent  
264 maximum intensity projections. Scale bar = 5  $\mu$ m.

265 (B) Fold increase of lipid droplet fluorescence intensity and number in proband compared to control.

# Figure 1





## Figure 2

