

BRIEF COMMUNICATION

RNaseq analysis for the diagnosis of muscular dystrophy

Hernan Gonorazky^{1,2,3,a}, Minggao Liang^{2,4,a}, Beryl Cummings^{5,6,a}, Monkol Lek^{5,6}, Johann Micallef², Cynthia Hawkins⁷, Raveen Basran⁷, Ronald Cohn^{2,3,4}, Michael D. Wilson^{2,4}, Daniel MacArthur^{5,6}, Christian R. Marshall⁷, Peter N. Ray^{2,4,7} & James J. Dowling^{1,2,3,4}

¹Division of Neurology, Hospital for Sick Children, Toronto, Ontario, Canada M5G A04

²Program of Genetics and Genome Biology, Hospital for Sick Children, Toronto, Ontario, Canada M5G A04

³Department of Paediatrics, University of Toronto, Toronto, Ontario, Canada M5G A04

⁴Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada M5G A04

⁵Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts 02114

⁶Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, Massachusetts

⁷Pediatric Laboratory Medicine, Hospital for Sick Children, Toronto, Ontario, Canada M5G A04

Correspondence

James J. Dowling, 15-09703 PGCRL, 686 Bay Street, Toronto, Ontario, Canada M5G0A4.
Tel: +1 (416) 813-8980; Fax: +1 (416) 813-6334; E-mail: james.dowling@sickkids.ca

Funding Information

This work was supported by the SickKids Centre for Brain and Mental Health Chase-an-Idea grant (Dowling).

Received: 23 October 2015; Accepted: 25 October 2015

Annals of Clinical and Translational Neurology 2016; 3(1): 55–60

doi: 10.1002/acn3.267

^aContributed equally.

Introduction

Neurogenetic diseases are a common cause of severe disability, associated with multiple morbidities, early mortality, and significant economic and societal costs.^{1,2} Muscular dystrophies (MDs) are prototypical neurogenetic disorders in that they are clinically and genetically heterogeneous but united by a common set of clinical and diagnostic observations (muscle weakness, elevated Creatine Phosphokinase (CPK), and dystrophic changes on muscle biopsy).³ A critical existing issue in the MD field is that a significant number of children (~50%) with suspected disease do not yet have identified gene mutation(s) despite extensive genetic analysis.⁴ This fact presents a major barrier to clinical care by prolonging the diagnostic odyssey for patients and families, by limiting care recommendations and prognostic information, and

Abstract

The precise genetic cause remains elusive in nearly 50% of patients with presumed neurogenetic disease, representing a significant barrier for clinical care. This is despite significant advances in clinical genetic diagnostics, including the application of whole-exome sequencing and next-generation sequencing-based gene panels. In this study, we identify a deep intronic mutation in the *DMD* gene in a patient with muscular dystrophy using both conventional and RNA-seq-based transcriptome analyses. The implications of our data are that non-coding mutations likely comprise an important source of unresolved genetic disease and that RNAseq is a powerful platform for detecting such mutations.

by exposing patients to potentially unnecessary testing.⁵ It also hinders research into disease pathogenesis and therapy development.

Whole-exome sequencing (WES), along with next-generation sequencing-based gene panels, represent a significant technical advance that has greatly improved genetic diagnostics.⁶ However, there remains a large group of patients whose genetic cause has not been uncovered despite thorough investigation with these modalities.^{4,7} One potential and underexplored source of mutation is sequence variation that alters RNA expression and/or processing. Such mutations may occur at exon/intron boundaries, and thus be captured by WES, or occur outside of the standard exome (such as in deep intronic and intergenic regions). To date, while 15% of mutations in the Human Gene Mutation Database are predicted to alter splicing and/or gene expression, nearly all of these are

described at splice-site junctions. Only a small number of cases exist where mutations in intronic or intergenic sequences have been identified as the cause of disease.^{8–10} These aspects of the genome are seldom examined in the context of neurogenetic disease; furthermore, next-generation sequencing based on RNA transcript analysis (RNA-seq) has yet to be studied in this context. Here, we present a case of a deep intronic mutation as the cause of Duchenne MD, and show that RNAseq is an effective modality for detecting this mutation.

Methods and Results

We encountered a 6-year-old boy with the insidious onset of gait abnormalities. He had been normally developing until age 2.5 years when he began experiencing increased trips and falls. After assessment by a family practitioner, a serum CPK study was sent, and was found to be elevated at >16,000 μL (normal 55–300). His physical examination was notable for shoulder and calf muscle hypertrophy, limb-girdle muscle weakness, mild bilateral ankle contractures, a positive Gowers sign, and a waddling gait. On the basis of these features, we initiated clinical genetic testing for Duchenne MD,¹¹ including multiplex ligation-dependent probe amplification-based deletion/duplication analysis of the *DMD* gene and direct Sanger sequencing of *DMD* coding sequence and intron/exon boundaries. These investigations did not reveal a causative mutation.

Because his clinical picture was consistent with the diagnosis of MD, we next sent a next-generation sequencing-based gene panel that examined all coding exons of the known limb-girdle MD genes (35 genes, Emory Genetics Laboratory).⁴ No causative mutation was detected using this strategy. We then performed a diagnostic muscle biopsy, which revealed histologic changes consistent with a MD (Fig. 1A). Immunostaining showed absent dystrophin expression in the majority of fibers (Fig. 1B and C), suggestive of a mutation at the *DMD* locus. We therefore used the remaining muscle biopsy material to analyze the *DMD* transcript, using both a standard approach and RNA-seq.

For conventional analysis, we performed overlapping RT-PCR reactions covering the entire *DMD* gene (Fig. 1D). We detected altered cDNA products (size greater than predicted) with all primer sets that included both exons 37 and 38. We sequenced one of the altered products and uncovered a novel 51 base pair insertion between exons 37 and 38 corresponding to a discrete sequence fragment within intron 37 (Fig. 1E). We next analyzed genomic DNA corresponding to the area surrounding the inserted sequence, and identified a single sequence variant in intron 37 (g.chrX:32,366,860 A>C [c.5326-215 T>G]). This variant creates a novel splice

acceptor site, which then pairs with a cryptic splice donor site in the intron to create an aberrant 51 base pair exon (Fig. 1D). This aberrant exon creates a frameshift and premature stop codon in the *DMD* coding sequence (Fig. 1F), and is consistent with the absent protein expression seen by immunostaining. On the basis of this, we concluded that this sequence variant is causative for disease in our patient.

RNAseq analysis proved to be equally as informative. We isolated RNA from ~15 mg of biopsy from our patient and from three controls (RNA integrity value (RIN) values ranging from 7.2 to 9.5). We generated libraries with an Illumina, San Diego, CA, USA. TruSeq V2 mRNA library kit and performed RNAseq on an Illumina HiSeq 2000 to average depth of 50–100 million paired end reads of transcript sequence per sample.¹² Reads were aligned to the hg19 reference assembly using RNA-STAR¹³ and Gencode Release 19 transcriptome annotations.¹⁴ To detect unannotated transcripts and isoforms, we performed unguided transcriptome assembly for each sample using Cufflinks¹⁵; these were then annotated against the reference transcriptome. Unsupervised cluster analysis based on gene expression levels definitively distinguished the patient sample from controls (Fig. 2A). Examination of the patient's transcriptome identified altered transcript levels (>threefold) in 1197 genes relative to controls (FDR-adjusted $P > 0.05$), with *DMD* the third most significantly reduced transcript (Fig. 2B). Examination of the *DMD* transcript showed relatively uniform reduction in all exons (Fig. 2C), whereas dedicated interrogation of *DMD* transcript identified a single novel change corresponding to an insertion of transcribed sequence from intron 37 (Fig. 2C and D). We then determined the specific location and length of this sequence, and it corresponded exactly to the altered fragment we identified by the conventional analysis. In all, RNA-seq rapidly and correctly identified the causative abnormality in the *DMD* transcript.

To test the potential utility of RNAseq as a clinical diagnostic technique, we then analyzed the data in a semiblinded fashion. RNAseq source files from our patient and controls were given random numbers and sent blinded (without clinical or genetic information) for analysis to the McArthur laboratory. We (the McArthur group) then independently and unambiguously identified the *DMD* transcript abnormality, including detection of the unique sequence variant in the intron that was found by genomic DNA sequencing to be the causative gene mutation (Fig. 3A and B). As we were interested in RNA-seq as a potential clinical diagnostic tool, we additionally documented comprehensive coverage of several major muscle specific transcripts, including *DMD* and other large, complex and likely low abundance transcripts (ex:

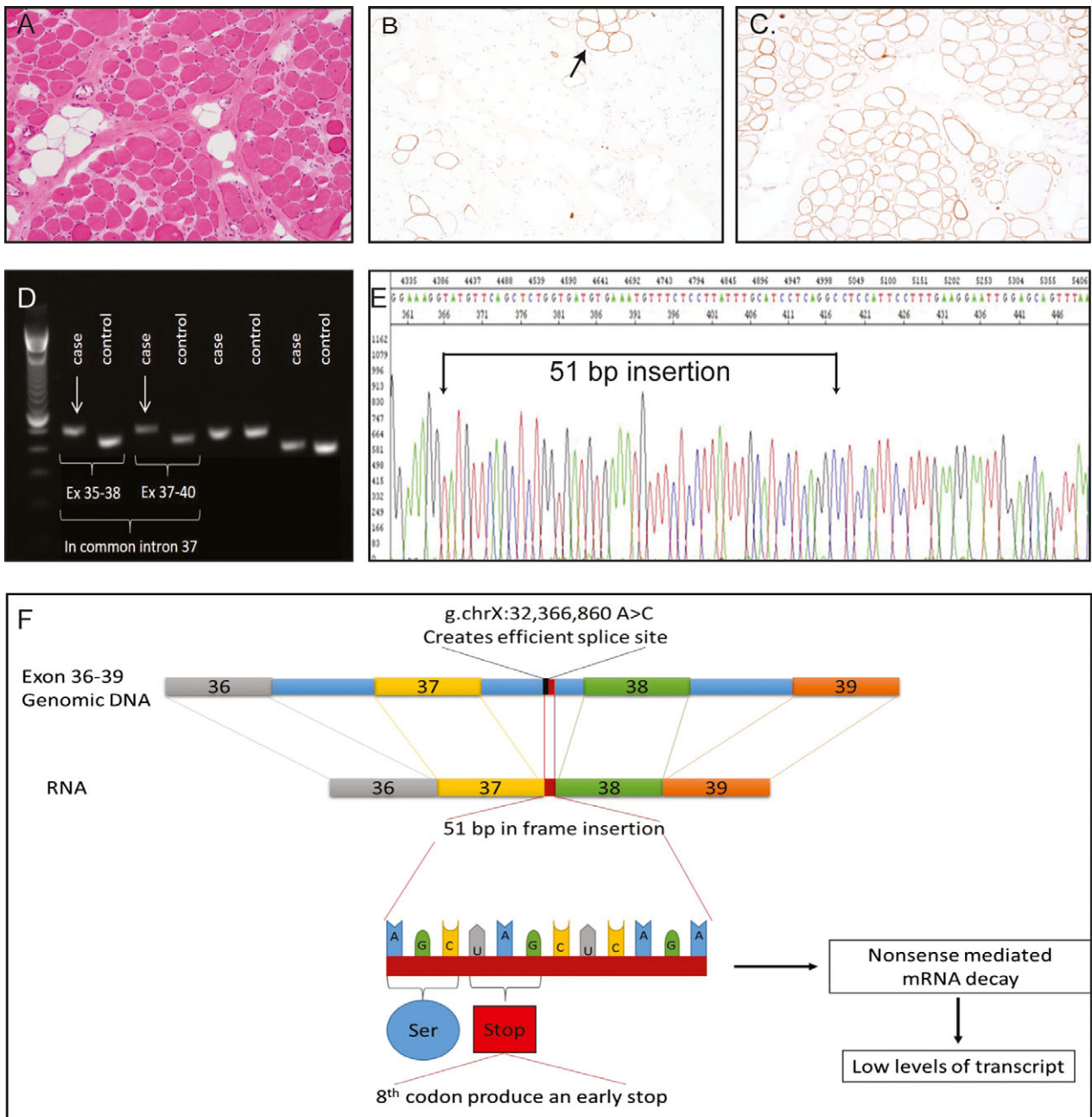


Figure 1. Deep intronic mutation in the *DMD* gene as a cause of muscular dystrophy. (A–C) Diagnostic muscle biopsy results. (A) Hematoxylin and Eosin staining revealed a typical dystrophic pattern (areas of fibrosis, fatty infiltrate, and degenerating and regenerating fibers). (B) IHC for dystrophin showing absent expression, with the exception of some revertant fibers (arrow). (C) Immunohistochemistry (IHC) for a-sarcoglycan showing a normal staining pattern. (D) RT-PCR analysis using cDNA from patient muscle and overlapping primer sets that span multiple *DMD* exons. The primer sets including both exons 37 and 38 yielded larger than expected bands (arrows). (E) Sanger sequencing of an RT-PCR fragment with exons 37/38 revealed a 51 base pair insertion of intron 37 sequence. (F) Schematic of the mutation and its consequences.

LAMA2, *RYR1* and *NEB*, Fig. 3C). We also determined that at our read depth of 100 million reads we found in our patient sample ~20 reads supporting the intron 37 inclusion *DMD* transcript (as opposed to zero in the con-

trols or 150 samples from GTEx). These additional data thus confirm that RNA-seq can identify a disease associated transcript variant in a nonbiased fashion, and provide baseline measurements that suggest RNAseq can

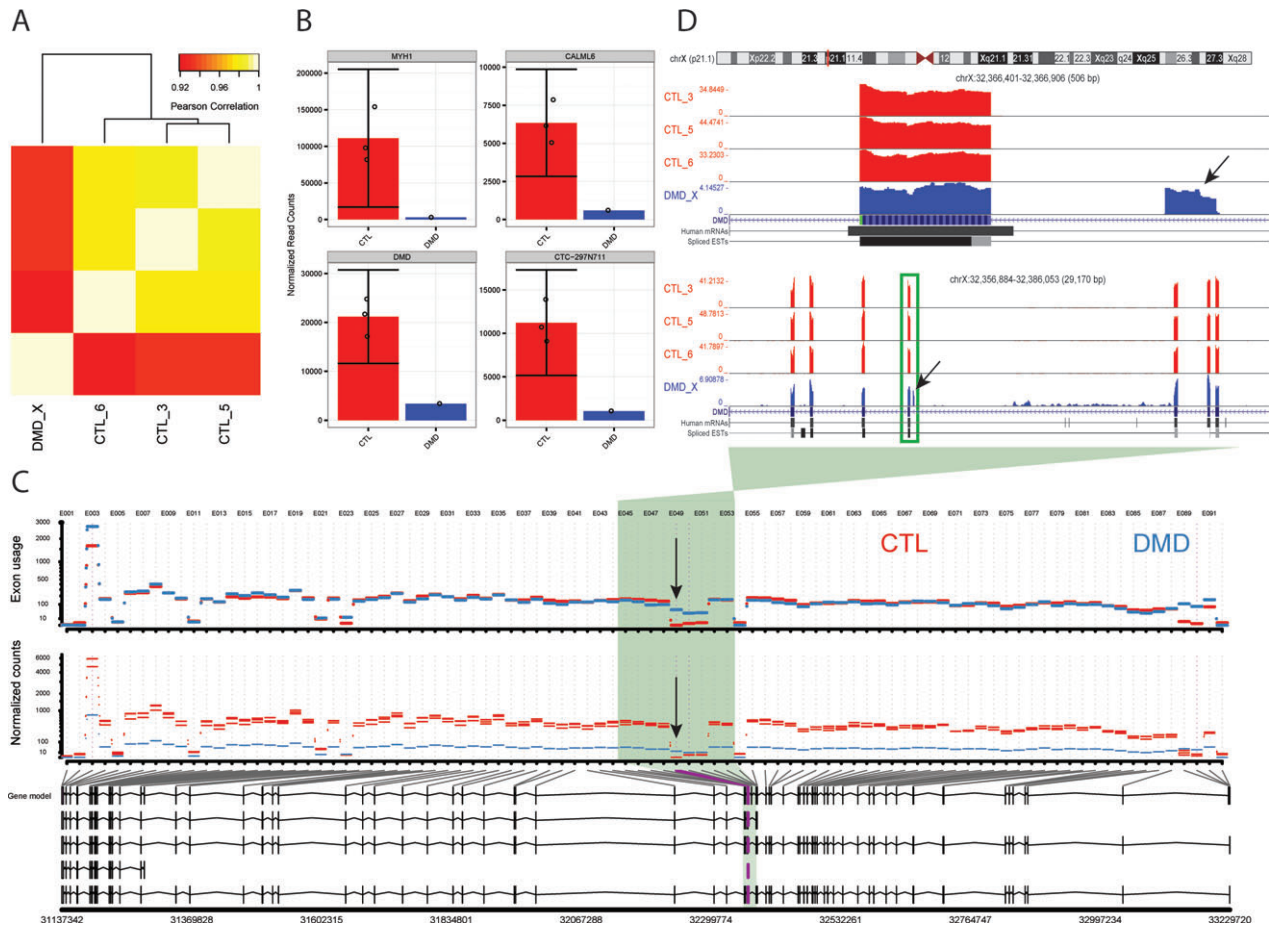


Figure 2. RNA-seq analysis of muscle from a patient with an intronic *DMD* mutation. (A) Pairwise correlation heatmap of RNA-seq samples based on the Pearson correlation of log gene expression values for all genes. Gene expression (counts) was determined and normalized by effective library size using DESeq2. (B) Counts for the top four downregulated genes in *DMD* patient versus controls as ranked by significance. Bar plots show counts for the patient (blue) and sample means for controls (red). Error bars represent 95% confidence intervals. Circles represent counts for individual samples. (C) Per-exon read counts and differential exon usage for *DMD* in patient versus control samples. The novel *DMD* exon was detected as the most differentially expressed exon between the patient and controls (E049 at chrX:32366809-32366856; marked by arrow). Additional novel exons were detected (E050, E090, E091) but likely represent transcriptional noise. (D) UCSC genome browser screenshot of raw RNA-seq signal normalized to library size at the *DMD* locus corresponding to the shaded region in (C) (upper panel). A zoom-in of the boxed region including the novel *DMD* exon is shown in the lower panel (novel exon marked by arrow).

serve as a clinical diagnostic tool for a range of muscle disorders.

Discussion

This case study illustrates two important concepts. First, it adds to a small but growing group of cases where non-coding mutations lying outside the intron/exon boundary are identified as the cause of neurogenetic disease.^{8–10,16} This provides support to our prediction that a significant fraction of the currently “unsolved” cohort of patients with MD and other similar genetic diseases are due to mutations that alter RNA processing and/or expression.

Identification of such mutations will likely require analysis of nonexomic sequence sources, including RNA (as was done in this case) and genomic noncoding DNA.

Second, our study demonstrates for the first time the potential utility of RNAseq for the identification of mutations that alter RNA transcript processing and/or expression.¹⁷ Using RNAseq, we were able to quickly and accurately detect a noncoding mutation in *DMD*. Given that essentially all-known muscle disease genes are adequately captured and represented by RNAseq, the technology is likely applicable across the broad spectrum of muscle disease. This is particularly true given that muscle biopsy material is available for analysis in many cases.

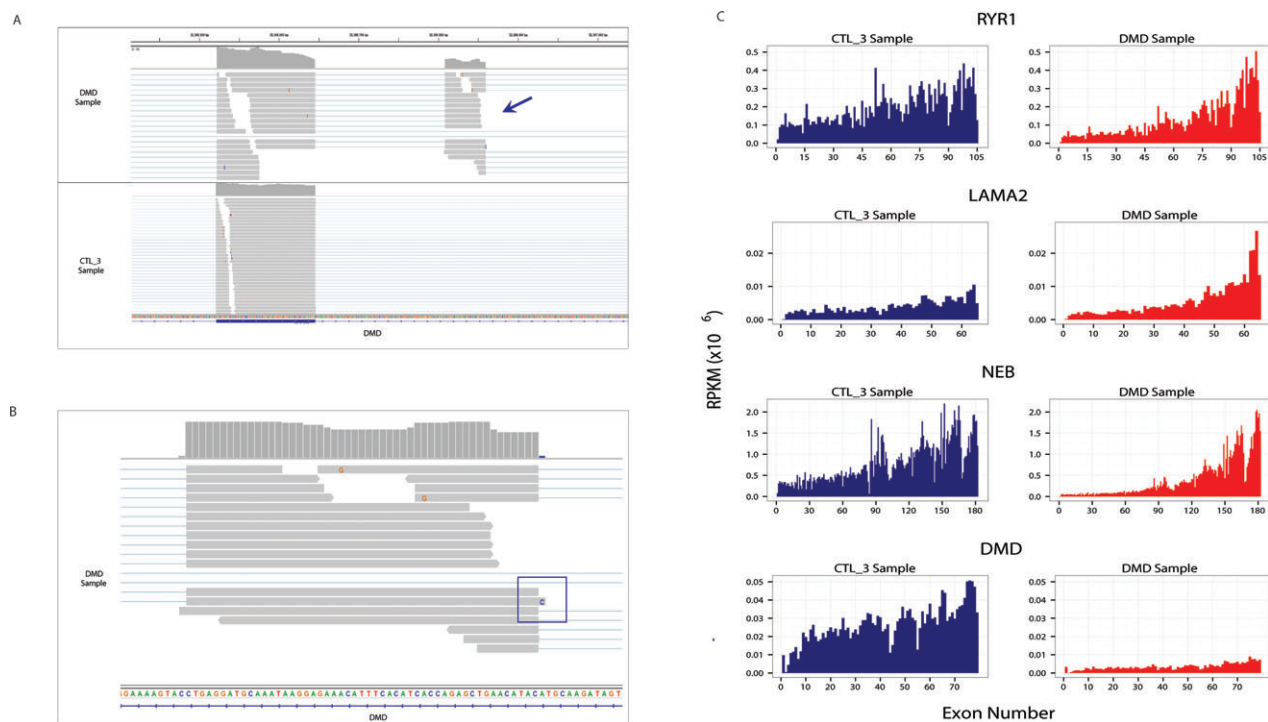


Figure 3. RNA-seq as a tool for mutation detection in muscular dystrophy. (A) RNA-seq reads for the *DMD* transcript from the patient and a representative control. Note the existence of an aberrant transcript fragment in all reads from the DMD patient (arrow). (B) Examination at the resolution of the individual base pair of the aberrant transcript fragment. Direct evaluation reveals the causative sequence variant in the patient (box). (C) exon usage data plot (normalized as reads per million) for four large, muscle-specific transcripts. Note that there is adequate coverage of these large transcripts throughout the gene. The exception is *DMD*, which shows uniformly low expression in the patient sample (control in blue, patient in red).

Also, it is possible to perform RNA analysis on myotubes derived from myoD-driven transdifferentiation of fibroblasts,¹⁸ therefore obviating the requirement for muscle biopsy material for generating usable RNAseq data. This opens the broader possibility of application of RNAseq to a range of neurogenetic disorders, using for example, neurons derived in vitro either through direct reprogramming¹⁹ or from induce Pluripotent stem (iPS) cells.²⁰

Lastly, of note, the specific mutation identified in our case is unusual as it occurs deep in an intron and it creates a novel exon with splice acceptor and donor independent from the surrounding exons. Importantly, investigation of this genomic variant by combined annotation-dependent depletion (CADD) analysis places it only in the 10th percentile in terms of pathogenicity, meaning that it would likely be ranked only in the top 300,000 of disease-relevant variants in our patient's genome. Based on this, it is possible that this mutation would not have been considered pathogenic in the absence of transcript data, and instead would have been coded as a VOUS or even a benign variant. Given the potentially high rate of VOUS in the non-coding genome, this raises the question as to whether whole-genome sequencing on its own will be adequate to

uncover such mutations, or if RNA analysis will be a critical required element for establishing pathogenicity of non-coding mutations. With the speed and potential accuracy of RNAseq (as well as the price, which is equivalent to WES), it is possible that it may represent a linked or even preferred modality.

Acknowledgments

We thank Etsuko Tsuchiya, Marianne Eliou, Jennifer Orr, and Dax Tori (Donnelly Sequencing Centre) for technical support. This work was supported by the SickKids Centre for Brain and Mental Health Chase-an-Idea grant (Dowling). Patient information and material was collected in compliance with an REB-approved protocol. ML and MDW were supported by a Canada Research Chair and NSERC grant 436194-2013.

Author Contributions

H. G. and J. M. presented the clinical data, helped with RNA-seq analysis, and assisted in manuscript writing. M. L., M. W., B. C., M. L., and D. McA. all worked on

RNA-seq analysis and interpretation. C. R. M., R. B., and P. D. R. performed conventional RNA analysis on the patient biopsy. C. H. interpreted the muscle biopsy. R. C. helped with data interpretation and clinical evaluation. J. J. D. conceived the study, helped with all data interpretation, and wrote the manuscript. All authors assisted in manuscript editing.

Conflict of Interest

None declared.

References

- Landfeldt E, Lindgren P, Bell CF, et al. The burden of Duchenne muscular dystrophy: an international, cross-sectional study. *Neurology* 2014;83:529–536.
- Lopez-Bastida J, Oliva-Moreno J. Cost of illness and economic evaluation in rare diseases. *Adv Exp Med Biol* 2010;686:273–282.
- Chelly J, Desguerre I. Progressive muscular dystrophies. *Handb Clin Neurol* 2013;113:1343–1366.
- Ankala A, da Silva C, Gualandi F, et al. A comprehensive genomic approach for neuromuscular diseases gives a high diagnostic yield. *Ann Neurol* 2015;77:206–214.
- Wong SH, McClaren BJ, Archibald AD, et al. A mixed methods study of age at diagnosis and diagnostic odyssey for Duchenne muscular dystrophy. *Eur J Hum Genet* 2015;1294–300.
- Jiang T, Tan MS, Tan L, Yu JT. Application of next-generation sequencing technologies in Neurology. *Ann Transl Med* 2014;2:125.
- Yang Y, Muzny DM, Reid JG, et al. Clinical whole-exome sequencing for the diagnosis of Mendelian disorders. *N Engl J Med* 2013;369:1502–1511.
- Baskin B, Banwell B, Khater RA, et al. Becker muscular dystrophy caused by an intronic mutation reducing the efficiency of the splice donor site of intron 26 of the dystrophin gene. *Neuromuscul Disord* 2009;19:189–192.
- Ruggieri A, Ramachandran N, Wang P, et al. Non-coding VMA21 deletions cause X-linked myopathy with excessive autophagy. *Neuromuscul Disord* 2015;25:207–211.
- Trabelsi M, Beugnet C, Deburgrave N, et al. When a mid-intronic variation of DMD gene creates an ESE site. *Neuromuscul Disord* 2014;24:1111–1117.
- Bushby K, Finkel R, Birnkrant DJ, et al. Diagnosis and management of Duchenne muscular dystrophy, part 1: diagnosis, and pharmacological and psychosocial management. *Lancet Neurol* 2010;9:77–93.
- Irimia M, Weatheritt RJ, Ellis JD, et al. A highly conserved program of neuronal microexons is misregulated in autistic brains. *Cell* 2014;159:1511–1523.
- Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 2013;29:15–21.
- Harrow J, Frankish A, Gonzalez JM, et al. GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res* 2012;22:1760–1774.
- Trapnell C, Hendrickson DG, Sauvageau M, et al. Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat Biotechnol* 2013;31:46–53.
- Kevelam SH, Taube JR, van Spaendonck RM, et al. Altered PLP1 splicing causes hypomyelination of early myelinating structures. *Ann Clin Transl Neurol* 2015;2:648–661.
- Ku CS, Wu M, Cooper DN, et al. Exome versus transcriptome sequencing in identifying coding region variants. *Expert Rev Mol Diagn* 2012;12:241–251.
- Waddell LB, Monnier N, Cooper ST, et al. Using complementary DNA from MyoD-transduced fibroblasts to sequence large muscle genes. *Muscle Nerve* 2011;44:280–282.
- Tsunemoto RK, Eade KT, Blanchard JW, Baldwin KK. Forward engineering neuronal diversity using direct reprogramming. *EMBO J* 2015;34:1445–1455.
- Bellin M, Marchetto MC, Gage FH, Mummery CL. Induced pluripotent stem cells: the new patient? *Nat Rev Mol Cell Biol* 2012;13:713–726.