

Whole Exome

REFERRING HEALTHCARE PROFESSIONAL

NAME	HOSPITAL

PATIENT

NAME	DOB	AGE	GENDER	ORDER ID
		1	Male	

PRIMARY SAMPLE TYPE	SAMPLE COLLECTION DATE	CUSTOMER SAMPLE ID

SUMMARY OF RESULTS

PRIMARY FINDINGS

Analysis of whole exome sequence variants in previously established disease genes

The patient is heterozygous for *MPZ* c.532del, p.(Val178Trpfs*74), which is likely pathogenic.

Del/Dup (CNV) analysis

Negative.

SECONDARY FINDINGS

The patient is heterozygous for *PMS2* c.1297A>T, p.(Lys433*), which is likely pathogenic.

Please see [APPENDIX 3: Secondary Findings](#) for further details

PRIMARY FINDINGS: SEQUENCE ALTERATIONS IN ESTABLISHED DISEASE GENES

GENE	TRANSCRIPT	NOMENCLATURE	GENOTYPE	CONSEQUENCE	INHERITANCE	CLASSIFICATION
MPZ	NM_000530.7	c.532del, p.(Val178Trpfs*74)	HET	frameshift_variant	AD	Likely pathogenic
	ID	ASSEMBLY	POS	REF/ALT		
		GRCh37/hg19	1:161276170	AC/A		
	gnomAD AC/AN	POLYPHEN	SIFT	MUTTASTER	PHENOTYPE	
	0/0	N/A	N/A	N/A	Charcot-Marie-Tooth disease, Dejerine-Sottas disease	

SEQUENCING PERFORMANCE METRICS

SAMPLE	MEDIAN COVERAGE	PERCENT >= 20X
Index	162	99.25

TEST INFORMATION

Blueprint Genetics Whole Exome Test consists of sequence analysis of all protein coding genes in the genome for the proband, coupled with Whole Exome Deletion/Duplication (CNV) Analysis. The test targets all protein coding exons, exon-intron boundaries (\pm 20 bps) and selected noncoding, deep intronic variants (listed in Appendix). This test should be used to detect single nucleotide variants and small insertions and deletions (INDELS) up to 220 bps and copy number variations defined as single exon or larger deletions and duplications. This test should not be used for the detection of repeat expansion disorders or diseases caused by mitochondrial DNA (mtDNA) mutations. The test does not recognize balanced translocations or complex inversions, and it may not detect low-level mosaicism.

Analysis of Whole Exome Test is primarily focused on established disease genes that have been previously associated with genetic disorders. The genes with known clinical association include those curated by Blueprint Genetics (BpG) and included in BpG diagnostic panels (>2400 genes). These genes are supplemented with genes included in The [Clinical Genomics Database](#) (>3350 genes) and the Developmental Disorders Genotype-Phenotype Database ([DD2GP](#)) (>1640 genes). Total number of genes that are considered as clinically associated in the Whole Exome analysis is >3750 (and the number is constantly updated).

If analysis of exome variants in previously established disease genes is inconclusive, exome variant data are also analyzed for variants that are not located within known clinically associated genes but have properties that make them candidates for potentially disease-causing variants (please see Appendix: Summary of the Test). If over time other patients with similar phenotype and variants in the same gene are identified, the variant may be reclassified as a likely cause of the disorder.

STATEMENT

CLINICAL HISTORY

The patient is a 1-year-old boy with severe hypotonia, areflexia and sensorimotor demyelinating neuropathy. Consent has been received to report secondary findings for this individual.

CLINICAL REPORT

Whole-exome sequence analysis of variants in previously established disease genes

Given that there is no reported family history of the disease, the exome data of the patient were analysed for rare heterozygous variants (potential de novo variants) and variants following a recessive inheritance pattern.

Sequence analysis using the Blueprint Genetics (BpG) Whole Exome identified a heterozygous frameshift variant *MPZ* c.532del, p.(Val178Trpfs*74).

***MPZ* c.532del, p.(Val178Trpfs*74)**

This variant is absent in [gnomAD](#), a large reference population database (n>120,000 exomes and >15,000 genomes) which aims to exclude individuals with severe pediatric disease. The variant deletes 1 base pair in exon 4 (of 6) of *MPZ* and generates a frameshift, leading to a stretch of 74 abnormal amino acids and an eventual new termination codon downstream of the last exon. This variant is predicted to elongate the protein by 3 amino acids. To the best of our knowledge, this variant has not been described in the medical literature or reported in disease-related variation databases such as [ClinVar](#) or [HGMD](#). However, a number of similar frameshift variants in *MPZ*, all leading to a long stretch of aberrant amino acids and a short elongation of the protein, have been described in multiple individuals with *MPZ*-related disease (PMIDs: [8816708](#), [9055797](#), [9187667](#), [15261887](#), [19259128](#), [22018721](#), [25614874](#), [27353517](#)). Some of these patients have been reported to have childhood onset neuropathy with an often severe presentation (PMID: [8816708](#), [9055797](#), [15261887](#), [19259128](#), [22018721](#)), while a milder adult-onset, phenotype has been reported for others (PMID: [31211173](#)). One patient with a heterozygous frameshift variant p.(Leu175Serfs*76) was noted to have gross motor delay in early childhood and presented at age 8 with a severe, early onset peripheral neuropathy. Based on his early onset, severity of clinical presentation, and NCV findings, this patient was diagnosed with Dejerine-Sottas syndrome (PMID: [8816708](#)). Another patient with a de novo variant p.(Arg185Alafs*66) showed moderate upper and lower limb weakness at the age of 1.5 years, as well as diminished tendon reflexes and macrocephaly (PMID: [19259128](#)). Functional studies of some of these variants have shown that they lead to retention of the mutant *MPZ* protein in the endoplasmic reticulum and induce apoptosis (PMIDs: [16252242](#), [23344956](#)).

MPZ

The *MPZ* gene (MIM #[159440](#)) encodes myelin protein zero, which is the major structural protein of peripheral myelin and accounts for approximately 50% of the total protein present in the sheath of peripheral nerves (PMID: [16856148](#)). *MPZ* is required for proper myelination and plays a major structural role in connecting adjacent myelin lamellae in the myelin sheath. Heterozygous pathogenic variants in *MPZ* cause a spectrum of autosomal dominant Charcot-Marie-Tooth (CMT) phenotypes, including congenital forms of hypomyelinating neuropathy (MIM #[145900](#), #[618184](#), also referred to as Dejerine-Sottas syndrome), the demyelinating form of CMT (MIM #[118200](#), #[180800](#)), intermediate phenotypes (MIM #[607791](#)), and the axonal type of CMT disease (MIM #[607677](#), [607736](#)). In general, CMT disease refers to a group of disorders characterized by a chronic motor and sensory neuropathy. The affected individuals typically have distal wasting, weakness and sensory loss that starts in the lower limbs and progresses slowly in a length-dependent manner. CMT is one of the most common inherited neuromuscular diseases, with a population prevalence of 1 in 2,500 (PMID: [24018473](#)). CMT is divided into demyelinating (CMT1) and axonal (CMT2) forms based on median nerve motor conduction velocity, with CMT2 usually associated with less severe symptoms.

Both adult-onset and severe, early onset forms of *MPZ*-related disease have been described. Severe early-onset disease is

typically characterized by early-onset hypotonia, severely delayed motor development, muscle weakness with areflexia, and severely decreased nerve conduction velocities (NCV) resulting from improper myelination of axons (PMID: 8816708). Two early-onset MPZ-related phenotypes are annotated on OMIM: congenital hypomyelinating neuropathy type 2 (CHN; MIM #618184) and Dejerine-Sottas syndrome (DSS; MIM #145900). Clinically, there is significant phenotypic overlap between these two disorders. However, the disorders have sometimes been classified differently, noting that CHN is characterized by hypo- or amyelination resulting from a congenital defect in myelin formation, whereas DSS has features of continuous myelin breakdown, with demyelination and remyelination (PMID: 17825553).

There are currently over 250 variants in *MPZ* annotated as disease-causing (DM) in the HGMD Professional variant database (version 2020.2), which includes mainly missense and truncating variants (nonsense, frameshift, splicing). It has been suggested that *MPZ* variants can be separated in two groups: partial loss-of-function variants causing late onset disease, and abnormal gain-of-function variants causing more severe early onset disease (PMID: 18337304).

Mutation nomenclature is based on GenBank accession NM_000530.7 (*MPZ*) with nucleotide one being the first nucleotide of the translation initiation codon ATG.

Upon request, filtered variant files and raw data files from the whole exome analysis can also be provided.

CONCLUSION

MPZ c.532del, p.(Val178Trpfs*74) is classified as likely pathogenic, based on the established association between the gene and the patient's phenotype, the variant's absence in reference populations, and its consequence (frameshift). Disease caused by *MPZ* variants is inherited in an autosomal dominant manner. *MPZ*-related disease may be caused by a de novo variant. Genetic counseling and family member testing are recommended.

STEP	DATE
Order date	
Sample received	
Sample in analysis	
Reported	

On Sep 03, 2020 the statement has been prepared by our geneticists and physicians, who have together evaluated the sequencing results:



Kirsty Wells, PhD, HCPC registered clinical scientist
Senior Geneticist



Juha Koskenvuo, MD, PhD
Lab Director, Chief Medical Officer

APPENDIX 3: SECONDARY FINDINGS

The patient was opted-in for an analysis of secondary findings, which are sequence variants unrelated to the indication for ordering the sequencing but of medical value for patient care. Whole Exome data of the patient were analyzed for secondary findings in 59 genes according to recommendations of American College of Medical Genetics and Genomics (ACMG; PMID [27854360](#)).

SECONDARY FINDINGS: SEQUENCE ALTERATIONS

GENE	TRANSCRIPT	NOMENCLATURE	GENOTYPE	CONSEQUENCE	INHERITANCE	CLASSIFICATION
PMS2	NM_000535.6	c.1297A>T, p.(Lys433*)	HET	stop_gained	AD,AR	Likely pathogenic
ID		ASSEMBLY	POS	REF/ALT		
		GRCh37/hg19	7:6027099	T/A		
gnomAD AC/AN		POLYPHEN	SIFT	MUTTASTER	PHENOTYPE	
0/0		N/A	N/A	disease causing	Colorectal cancer; hereditary nonpolyposis	

NOTES REGARDING SECONDARY FINDINGS

The secondary findings analysis in this individual identified a heterozygous nonsense variant c.1297A>T, p.(Lys433*) in *PMS2*. This variant is classified as likely pathogenic*.

PMS2 c.1297A>T, p.(Lys433*)

This variant is absent in [gnomAD](#). The variant generates a premature stop codon in exon 11 (of 15) and is predicted to lead to loss of normal protein function, either through protein truncation or nonsense-mediated mRNA decay. This variant has been reported in the literature in an individual undergoing panel testing for hereditary cancer predisposition syndrome (PMID: [31159747](#)), and has also been detected by other laboratories in the context of clinical testing (ClinVar variation ID: [216073](#)). Loss of function of *PMS2* in an established disease mechanism in hereditary cancer predisposition syndrome.

PMS2

PMS2 encodes a key component of the mismatch repair system, which functions to correct DNA mismatches and small insertions and deletions occurring during DNA replication and homologous recombination. The protein forms heterodimers with *MLH1* to form the MutL-alpha heterodimer. (GeneCards - [PMS2](#)) Germline pathogenic variants in *PMS2* (MIM #[600259](#)) are associated with hereditary nonpolyposis colorectal cancer (HNPCC), also known as Lynch syndrome (MIM # [614337](#), GeneReviews [NBK1211](#)) and autosomal recessive constitutional mismatch repair deficiency (CMMRD; MIM #[276300](#)).

Lynch syndrome is inherited in an autosomal dominant manner and is characterized by an increased risk for colon cancer and cancers of the endometrium, ovary, stomach, small intestine, hepatobiliary tract, urinary tract, brain, and skin. Lynch syndrome can be caused by germline pathogenic variants in four mismatch-repair (MMR) genes: *MLH1*, *MSH2*, *PMS2* and *MSH6*, or rarely, by germline deletions in the *EPCAM* gene (PMID: [21309036](#)). *MLH1* and *MSH2* germline pathogenic variants are identified in approximately 90% of families with Lynch syndrome; *MSH6* pathogenic variants in about 7%-10%; and *PMS2* pathogenic variants in fewer than 5%. The loss of MMR function leads to somatic DNA replication errors in repetitive sequences (microsatellites). Thus, microsatellite instability (MSI) is a hallmark in Lynch syndrome associated tumors. Traditionally, MSI testing and immunohistochemistry on tumor tissue have been used to identify the probability of Lynch syndrome and to identify which gene is most likely to have a causative germline variant.

In a large cohort study by Broeke et al the cumulative risk (CR) of colorectal cancer for male pathogenic variant carriers by age 70 years was 19%. The CR among female carriers was 11% for colorectal cancer and 12% for endometrial cancer. The mean age of CRC development was 52 years. (PMID: [25512458](#)) Recommendations for surveillance of mutation positive individuals include colonoscopy with removal of precancerous polyps every one to two years beginning between ages 20 and 25 years or two to five years before the earliest age of diagnosis in the family, whichever is earlier (GeneReviews [NBK1211](#)). The efficacy of surveillance for cancer of the endometrium, ovary, stomach, duodenum, and urinary tract is unknown.

The association of breast cancer as a part of Lynch syndrome is still debated (PMID: [23510156](#), [25737380](#), [25673086](#)). A recent study by Roberts et al showed that pathogenic variants in *MSH6* and *PMS2* implicated in Lynch syndrome associate with an increased risk for breast cancer (PMID: [29345684](#)). They calculated that approximately 37 percent of women with pathogenic *PMS2* variant will develop breast cancer by the age of 60 years, compared to approximately 15 percent of women in the general population. In addition, several studies have also shown that breast tumors in women with Lynch syndrome show microsatellite instability and loss of one or more mismatch repair proteins more frequently compared with sporadic breast tumors (PMID: [8646682](#), [19123071](#), [19575290](#), [22034109](#), [22691310](#)).

Constitutional mismatch repair deficiency (CMMRD; MIM #[276300](#)) syndrome is a severe autosomal recessive childhood cancer predisposition syndrome that results from homozygous or compound heterozygous pathogenic germline variants in one of the four MMR genes: *MLH1*, *MSH2*, *MSH6* or *PMS2*. Due to the constitutional defect in MMR capacity, individuals with biallelic pathogenic MMR gene variants have a high risk of developing a diverse spectrum of malignancies in childhood and adolescence. The spectrum includes mainly hematological malignancies, brain/central nervous system (CNS) tumors and colorectal cancer and other cancers that are typically seen in Lynch Syndrome patients at a later age. Also a variety of other malignancies have been observed in CMMRD patients. Many of the CMMRD patients, but not all, show features reminiscent of NF1, particularly multiple café au lait maculae (PMID: [24737826](#)).

Currently there are approximately 300 *PMS2* variants annotated as disease causing in the HGMD Professional variant database (version 2020.2). These include all kinds of variants: 8% of them are missense variants; 13% nonsense variants; 13% splicing variants; 26% small deletions, insertions, or indels, 35% gross deletions, 2% gross insertions, and 2% complex rearrangements. The majority (61%) of the variants are associated with colorectal cancer.

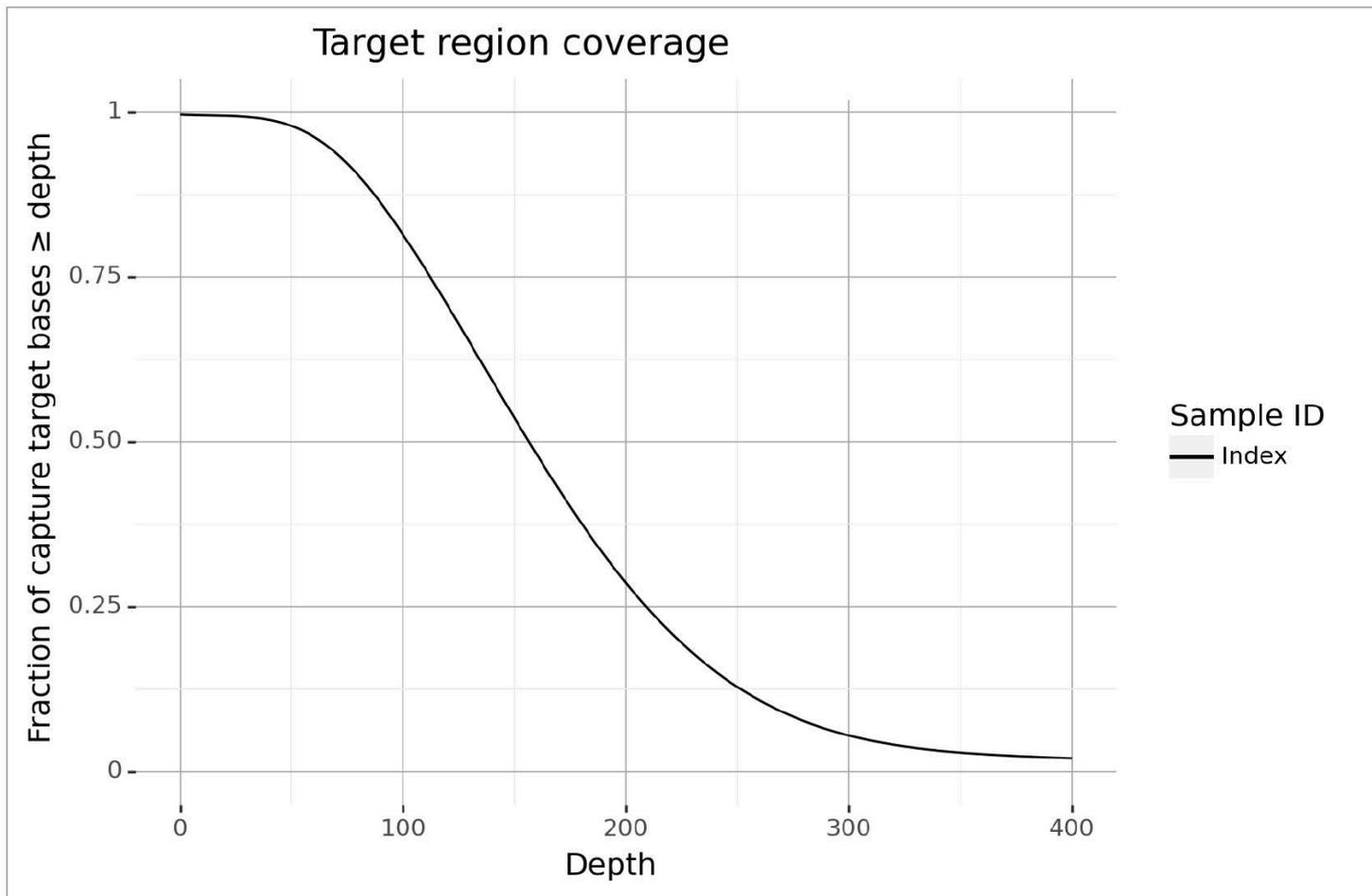
Mutation nomenclature is based on GenBank accession NM_000535.6 (*PMS2*) with nucleotide one being the first nucleotide of the translation initiation codon ATG.

Secondary findings: Conclusion

PMS2 c.1297A>T, p.(Lys433*) is classified as likely pathogenic, based on the variant's absence in reference populations and its consequence (nonsense). Heterozygous disease-causing variants in *PMS2* cause Lynch syndrome, which is inherited in an autosomal dominant manner and characterized by an increased risk for colorectal and other cancers (GeneReviews [NBK1211](#)). Genetic counseling is recommended.

*According to ACMG guidelines, 'Known Pathogenic' and 'Expected Pathogenic' variants are reported as secondary findings for *PMS2*. Known Pathogenic variants are defined as previously established pathogenic variants. For a subset of genes, also 'Expected Pathogenic' variants are analyzed and reported, and these represent any variants that are predicted to cause definite loss of function in genes where loss of function is the disease mechanism. These sequence variants include nonsense, frameshift and essential splice site variants.

Readability of the coverage plot may be hindered by faxing. A high quality coverage plot can be found with the full report on nucleus.blueprintgenetics.com.



APPENDIX 5: SUMMARY OF THE TEST

WHOLE EXOME

Laboratory process: When required, the total genomic DNA was extracted from the biological sample using bead-based method. DNA quality and quantity were assessed using electrophoretic methods. After assessment of DNA quality, qualified genomic DNA sample was randomly fragmented using non-contact, isothermal sonochemistry processing. Sequencing library was prepared by ligating sequencing adapters to both ends of DNA fragments. Sequencing libraries were size-selected with bead-based method to ensure optimal template size and amplified by polymerase chain reaction (PCR). Regions of interest (exons and intronic targets) were targeted using hybridization-based target capture method. The quality of the completed sequencing library was controlled by ensuring the correct template size and quantity and to eliminate the presence of leftover primers and adapter-adapter dimers. Ready sequencing libraries that passed the quality control were sequenced using the Illumina's sequencing-by-synthesis method using paired-end sequencing (150 by 150 bases). Primary data analysis converting images into base calls and associated quality scores was carried out by the sequencing instrument using Illumina's proprietary software, generating CBCL files as the final output.

Bioinformatics and quality control: Base called raw sequencing data were transformed into FASTQ format using Illumina's software (bcl2fastq). Sequence reads of each sample were mapped to the human reference genome (GRCh37/hg19). Burrows-Wheeler Aligner (BWA-MEM) software was used for read alignment. Duplicate read marking, local realignment around indels, base quality score recalibration and variant calling were performed using GATK algorithms (Sentieon) for nDNA. Variant data for were annotated using a collection of tools (VcfAnno and VEP) with a variety of public variant databases including but not limited to gnomAD, ClinVar and HGMD. The median sequencing depth and coverage across the target regions for the tested sample were calculated based on MQ0 aligned reads. The sequencing run included in-process reference sample(s) for quality control, which passed our thresholds for sensitivity and specificity. The patient's sample was subjected to thorough quality control measures including assessments for contamination and sample mix-up. Copy number variations (CNVs), defined as single exon or larger deletions or duplications (Del/Dups), were detected from the sequence analysis data using a proprietary bioinformatics pipeline. The difference between observed and expected sequencing depth at the targeted genomic regions was calculated and regions were divided into segments with variable DNA copy number. The expected sequencing depth was obtained by using other samples processed in the same sequence analysis as a guiding reference. The sequence data were adjusted to account for the effects of varying guanine and cytosine content.

Interpretation: Our variant classification follows the [Blueprint Genetics Variant Classification Schemes](#) modified from the [ACMG guideline 2015](#). Minor modifications were made to increase the reproducibility of the variant classification and to improve the clinical validity of the report. Likely benign and benign variants were not reported. The pathogenicity potential of the identified variants were assessed by considering the predicted consequence, the biochemical properties of the codon change, the degree of evolutionary conservation as well as a number of reference population databases and mutation databases such as, but not limited, to the 1000 Genomes Project, [gnomAD](#), [ClinVar](#) and HGMD. For missense variants, *in silico* variant prediction tools such as SIFT, PolyPhen, MutationTaster were used to assist with variant classification. In addition, the clinical relevance of any identified CNVs was evaluated by reviewing the relevant literature and databases such as 1000 Genomes Project, [Database of Genomic Variants](#), ExAC, [DECIPHER](#). The clinical evaluation team assessed the pathogenicity of the identified variants by evaluating the information in the patient referral, reviewing the relevant literature and manually inspecting the sequencing data if needed. Reporting was carried out using HGNC-approved gene nomenclature and mutation nomenclature following the HGVS guidelines.

In addition to analysis of variants in previously established disease genes, variants in genes where disease association has not yet been established were considered as potentially disease-causing using the following scheme:

- For probands who were whole-exome sequenced with parents, all coding region de novo variants were considered as candidate variants.
- Novel (absent in gnomAD) heterozygous, truncating variants (nonsense, frameshift, canonical splice site variants) in genes predicted to be intolerant for loss-of-function variation based on ExAC variant data. Genes were determined as intolerant if probability of loss-of-function intolerance score $pLI \geq 0.9$. The closer pLI is to one, the more LoF intolerant the gene appears to be. Genes with $pLI \geq 0.9$ are defined as an extremely LoF intolerant set of genes.

- Rare (<1% MAF in gnomAD), truncating homozygous or (predicted) compound heterozygous variants, or a combination of rare truncating and rare missense variant that is predicted deleterious by multiple *in silico* tools.

In addition, only variants in genes whose known expression pattern and function are considered relevant for the phenotype are included (e.g., variants in genes exclusively expressed in a muscular tissue are not considered as a candidate for a central nervous system disease). Candidate variants are not validated by Sanger sequencing, but their quality is inspected by visualization of sequence reads and evaluation of quality metrics, and only likely true variants are reported.

For proband and family members who were opted-in for analysis of secondary findings from the WES data, 59 clinically actionable genes were analyzed and reported for secondary findings according to recommendations by ACMG (PMID 27854360) with minor modifications aiming to increase the clarity of the classifications of the reportable variants (please see our website/clinical interpretation). Secondary findings are not analyzed or reported for deceased individuals or fetal samples.

Confirmation of sequence alterations: Sequence variants classified as pathogenic, likely pathogenic and variants of uncertain significance (VUS) were confirmed using bi-directional Sanger sequencing when they did not meet our stringent NGS quality metrics for a true positive call. In addition, prenatal case with diagnostic findings were confirmed.

Confirmation of copy number variants: CNVs (Deletions/Duplications) were confirmed using a digital PCR assay if they covered less than 10 exons (heterozygous), less than 3 exons (homo/hemizygous) or were not confirmed at least three times previously at our laboratory. Furthermore, CNVs of any size were not confirmed when the breakpoints of the call could be determined.

Analytic validation: This laboratory-developed test has been independently validated by Blueprint Genetics. The sensitivity of this panel is expected to be in the same range as the validated whole exome sequencing laboratory assay used to generate the panel data (sensitivity for SNVs 99.65%, and indels 1-50 bps 99.07%, one-exon deletion 92.3% and two exons CNV 100%, and specificity >99.9% for most variant types). It does not detect very low level mosaicism as a variant with minor allele fraction of 14.6% can be detected in 90% of the cases.

Test restrictions: A normal result does not rule out the diagnosis of a genetic disorder since some DNA abnormalities may be undetectable by the applied technology. Test results should always be interpreted in the context of clinical findings, family history, and other relevant data. Inaccurate or incomplete information may lead to misinterpretation of the results. **Technical**

limitations: This test does not detect the following: complex inversions, gene conversions, balanced translocations, repeat expansion disorders unless specifically mentioned, noncoding variants deeper than ± 20 base pairs from exon-intron boundary unless otherwise indicated (please see the list of noncoding variants covered by the test). Additionally, this test may not reliably detect the following: low level mosaicism, stretches of mononucleotide repeats, indels larger than 50bp, single exon deletions or duplications, and variants within pseudogene regions/duplicated segments. The sensitivity of this test may be reduced if DNA is extracted by a laboratory other than Blueprint Genetics. Laboratory error is also possible. Please see the Analytic validation above.

Regulation and accreditations: This test was developed and its performance characteristics determined by Blueprint Genetics (see Analytic validation). It has not been cleared or approved by the US Food and Drug Administration. This analysis has been performed in a CLIA-certified laboratory (#99D2092375), accredited by the College of American Pathologists (CAP #9257331) and by FINAS Finnish Accreditation Service, (laboratory no. T292), accreditation requirement SFS-EN ISO 15189:2013. All the tests are under the scope of the ISO 15189 accreditation (excluding mtDNA testing and digital PCR confirmation).

Please refer to Appendix 8 of the report in Nucleus ordering and reporting portal for full list of noncoding variants included in the Whole Exome analysis.

Please refer to Appendix 8 of the report in Nucleus ordering and reporting portal for full list of noncoding variants included in the Whole Exome analysis.

GLOSSARY OF USED ABBREVIATIONS:

AD = autosomal dominant

AR = autosomal recessive

gnomAD = genome aggregation database (reference population database; >138,600 individuals)

gnomAD AC/AN = allele count/allele number in the genome aggregation database (gnomAD)

HEM = hemizygous

HET = heterozygous

HOM = homozygous

ID = rsID in dbSNP

MutationTaster = *in silico* prediction tools used to evaluate the significance of identified amino acid changes.

Nomenclature = HGVS nomenclature for a variant in the nucleotide and the predicted effect of a variant in the protein level

OMIM = Online Mendelian Inheritance in Man®

PolyPhen = *in silico* prediction tool used to evaluate the significance of amino acid changes.

POS = genomic position of the variant in the format of chromosome:position

SIFT = *in silico* prediction tool used to evaluate the significance of amino acid changes.

Transcript = GenBank accession for reference sequence used for variant nomenclature
