Human Mutation

Severe neurocognitive and growth disorders due to variation in THOC2, an essential component of nuclear mRNA export machinery

Journal:	*Human Mutation*
Manuscript ID	Draft
Wiley - Manuscript type:	Research Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Gecz, Jozef; University of Adelaide, Adelaide Medical School and the Robinson Research Institute Sharma , Raman; The University of Adelaide, Adelaide Medical School and the Robinson Research Institute Gardner, Alison; The University of Adelaide, Adelaide Medical School and the Robinson Research Institute Homan, Claire; The University of Adelaide; Robinson Research Institute Douglas, Evelyn; Genetics and Molecular Pathology, SA Pathology Mefford, Heather; Division of Genetic Medicine, Department of Pediatrics, University of Washington Wieczorek, Dagmar; Universitatsklinikum Dusseldorf, Institut für Humangenetik; Heinrich-Heine-Universitat Dusseldorf, Institut für Humangenetik Lüdecke, Hermann-Josef; Heinrich-Heine-University, Medical Faculty, Institute of Human Genetics, Düsseldorf, Germany and Institut für Humangenetik, Universitätsklinikum Essen, Universität Duisburg-Essen Stark, Zornitza; Murdoch Children's Research Institute, Victorian Clinical Genetics Services Sadedin, Simon; Murdoch Children's Research Institute and Program in Medical and Population Genetics, Broad Institute of MIT and Harvard Nowak, Catherine Bearce; Boston Children's Hospital douglas, Jessica; Boston Children's Hospital Parsons, Gretchen; Spectrum Health Medical Genetics Mark, Paul; Spectrum Health Medical Genetics Loidi, Lourdes; Fundación Pública Galega de Medicina Xenómica, Hospital Clínico Universitario de Santiago de, Compostela Herman, Gail; Nationwide Children's Hospital and The Ohio State University Mihalic Mosher, Theresa; Nationwide Children's Hospital and The Ohio State University Gillespie, Meredith; Children's Hospital of Eastern Ontario Research Institute, University of Ottawa Brady, Lauren; Department of Pediatrics, McMaster University Medical Centre Tarnopolsky, Mark; McMaster University Medical Center, Dept. Pediatrics and Medicine Madrigal, Irene; Fundación Pública Galega de Medicina Xenómica, Hospital

	Clínico Universitario de Santiago de Compostela Eiris, Jesus; Unidad de Neurología Pediátrica, Departamento de Pediatría, Hospital Clínico Universitario de Santiago de Compostela Domenech Salgado, Laura; Centre de Regulacio Genomica; Universitat Pompeu Fabra Rabionet, Raquel; FUNDACIÓ PRIVADA PER A LA RECERCA I LA DOCÈNCIA SANT JOAN DE DÉU , departament de genètica, microbiologia i estadística de la UB; Universitat de Barcelona Facultat de Biologia, Strom, Tim; Helmholtz Zentrum München, Institute of Human Genetics Ishihara, Naoko; Department of Pediatrics, Fujita Health University School of Medicine Inagaki, Hidehito; Fujita Health University, ICMS Kurahashi, Hiroki; Fujita Health University, Institute for Comprehensive Medical Science, Division of Molecular Genetics Dudding, Tracy; Hunter New England Local Health District, Hunter Genetics ; University of Newcastle, Palmer , Emma; Genetics of Learning Disability Service, Hunter Genetics Field, Michael; Hunter Genetics, The GOLD Service
Key Words:	XLID, THOC2, mRNA export, protein stability, partial loss-of-function variants

RESEARCH ARTICLE

Severe neurocognitive and growth disorders due to variation in *THOC2*, an essential component of nuclear mRNA export machinery

Raman Kumar[1], Alison Gardner[1], Claire Homan[1], Evelyn Douglas[2], Heather Mefford[3], Dagmar Wieczorek[4], Hermann-Josef Lüdecke[4], Zornitza Stark[5], Simon Sadedin[6], The Broad CMG[7], Catherine Bearce Nowak[8], Jessica Douglas[8], Gretchen Parsons[9], Paul Mark[9], Lourdes Loidi[10], Gail Herman[11], Theresa Mihalic Mosher[11], Meredith K Gillespie[12], Lauren Brady[13], Mark Tarnopolsky[13], Irene Madrigal[14], Jesús Eiris[15], Laura Domenech Salgado[16], Raquel Rabionet[16], Tim M Strom[17], Naoko Ishihara[18], Hidehito Inagaki[19], Hiroki Kurahashi[19], Tracy Dudding-Byth[20,21], Elizabeth Palmer[20,22], Michael Field[20], Jozef Gecz[1,23*]

[1]Adelaide Medical School and the Robinson Research Institute, The University of Adelaide, Adelaide, SA5000, Australia; [2]Genetics and Molecular Pathology, SA Pathology, Adelaide, SA5000, Australia; [3]Division of Genetic Medicine, Department of Pediatrics, University of Washington & Seattle Children's Hospital, Seattle, Washington, USA; [4]Heinrich-Heine-University, Medical Faculty, Institute of Human Genetics, Düsseldorf, Germany and Institut für Humangenetik, Universitätsklinikum Essen, Universität Duisburg-Essen, Essen, Germany; [5]Murdoch Children's Research Institute and Department of Pediatrics, University of Melbourne, Melbourne, Australia; [6]Murdoch Children's Research Institute and Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA; [7]Broad's Center for Mendelian Genomics, Cambridge, Massachusetts, USA; [8]Boston Children's Hospital, Longwood Avenue, Boston, MA, USA; [9]Spectrum Health Medical Genetics, Grand Rapids, MI, USA; [10]Fundación Pública Galega de Medicina Xenómica, Hospital Clínico Universitario de Santiago de, Compostela. Santiago de Compostela, Spain; [11]Nationwide Children's Hospital and The Ohio State University, Columbus, OH, USA; [12]Children's Hospital of Eastern Ontario Research Institute, University of Ottawa, Ottawa, Canada; [13]Department of Pediatrics, McMaster University Medical Centre, Hamilton, Canada; [14]Fundación Pública Galega de Medicina Xenómica, Hospital Clínico Universitario de Santiago de Compostela, Santiago de Compostela, Spain; [15]Unidad de Neurología Pediátrica, Departamento de Pediatría, Hospital Clínico Universitario de Santiago de Compostela, Santiago de Compostela, Spain; [16]Centre for Genomic Regulation (CRG), Universitat Pompeu Fabra and CIBERESP, Barcelona Institute for Science and Technology, Barcelona, Spain; [17]Institut für Humangenetik, Universitätsklinikum Essen, Universität Duisburg-Essen, Essen, Germany; [18]Department of Pediatrics, Fujita Health University School of Medicine, Japan; [19]Division of Molecular Genetics, Institute for Comprehensive Medical Science, Fujita Health University, Japan; [20]Genetics of Learning Disability Service, Hunter Genetics, Waratah, NSW 2298, Australia; [21]University of Newcastle, Australia Grow-Up-Well Priority Research Center; [22]School of Women's and Children's Health, University of New South Wales, Randwick, NSW 2031, Australia; [23]Healthy Mothers, Babies and Children, South Australian Health and Medical Research Institute, Adelaide, SA 5000, Australia.

*Correspondence: Jozef Gecz, [1]School of Medicine, The Robinson Research Institute, The University of Adelaide, North Adelaide, SA5005, Australia. E-mail: jozef.gecz@adelaide.edu.au

Abstract

Highly conserved TREX-mediated mRNA export is emerging as a key pathway in neuronal development and differentiation. TREX subunit variants cause perturbed mRNA export from the cell nucleus to the cytoplasm and lead to neurodevelopmental disorders (NDDs). Our previous work implicated the *THOC2* gene on the X-chromosome in intellectual disability (ID). Borderline to severe ID with speech delay, behavioural problems, short stature, elevated BMI and truncal obesity in adulthood were observed. We now report an additional six affected individuals from five unrelated families with pathogenic or likely pathogenic variants in *THOC2*, two *de novo* and three maternally-inherited, extending the genotypic and phenotypic spectrum. These comprise three rare missense *THOC2* variants that affect evolutionarily conserved amino acid residues and reduce protein stability and affected males from two families with canonical splice-site *THOC2* variants that result in C-terminally truncated THOC2 proteins. We present detailed clinical assessment and functional studies on a *de novo* variant in a female with an epileptic encephalopathy and discuss an additional four families with rare variants in *THOC2* with supportive evidence for pathogenicity. Severe neurocognitive features, including movement and seizure disorders were observed in this expanded cohort. Taken together, our data significantly expand existing evidence showing that even subtle alterations to the canonical molecular pathways such as mRNA export, otherwise essential for cellular life, can be compatible with life, but lead to NDDs in humans.

KEY WORDS

XLID; *THOC2*; mRNA export; protein stability; partial loss-of-function variants

INTRODUCTION

Intellectual disability (ID), characterized by substantial limitations in both intellectual functioning and adaptive behaviour, affects 1-3% of the population starting before the age of 18 years and has significant impact on individuals, families and communities (Vissers, et al., 2016). Individuals with ID are more likely than members of the general population to experience poor physical and mental health, have a lower life expectancy, experience inequalities accessing health care and frequently have limited or no specific therapies for their core symptoms (Bittles, et al., 2002; Hosking, et al., 2016). Both genetic and environmental factors contribute to the development of ID (Milani, et al., 2015). Over 130 of the identified >800 ID genes are located on the X-chromosome, and diagnosis of X-linked causes of ID remain critically important for accurate genetic counseling of families (Ropers and Hamel, 2005). Dramatic improvements in high-throughput DNA sequencing technologies and analyses software has led to identification/discovery of new ID genes and additional variants in the known ID genes (Dickinson, et al., 2016; Vissers, et al., 2016). A systematic review of clinical data suggests that ID affected individuals commonly have comorbid neurological, psychiatric and behavioural disorders (Oeseburg, et al., 2011; Vissers, et al., 2016), and disease variants in different parts of a gene can lead to a broad range of complex neurocognitive disorders (Palmer, et al., 2017; Zhu, et al., 2014). This complexity contributes to heterogeneity in clinical symptoms and indistinct boundaries between syndromic and non-syndromic forms of NDD.

In 2015, we reported genetic, molecular and protein structural data on four missense variants in an X-linked essential gene *THOC2* (NM_001081550.1: p.Leu313Phe, p.Leu438Pro, p.Ile800Thr and p.Ser1012Pro) (Kumar, et al., 2015). The affected individuals had a syndromic NDD, characterized by borderline to severe ID, speech delay, short stature and

adult onset truncal obesity (Kumar, et al., 2015). *THOC2* encodes for the THOC2 protein - the largest subunit of the highly conserved TREX (Transcription-Export) mRNA export complex essential for exporting mRNA from the cell nucleus to the cytoplasm (Heath, et al., 2016). The TREX complex is composed of a THO sub-complex (THOC1, THOC2, THOC3, THOC5 and THOC7) and accessory proteins (UAP56, UIF, Aly, CIP29, PDIP3, ZC11A, SRRT, Chtop) (Heath, et al., 2016). The TREX complex, besides its canonical role in mRNA export in the mammalian cells, has been shown to perform critical roles in gene expression, 3' mRNA processing, stress responses, mitotic progression and genome stability as well as developmental processes such as pluripotency maintenance and hematopoiesis (Yamazaki, et al., 2010). We and others have recently demonstrated that subtle perturbations in mRNA export by gene variants or preferential cytoplasmic aggregation can lead to NDDs (Beaulieu, et al., 2013; Coe, et al., 2014; Kumar, et al., 2015), neurodegeneration (Woerner, et al., 2016) or cancer (Chinnam, et al., 2014; Hautbergue, 2017; Liu, et al., 2015; Viphakone, et al., 2015). These alterations can have tissue-specific effects as TREX subunits are shown to have tissue-specific roles; for example, mouse *Thoc5* deficiency interferes with the maintenance of hematopoiesis (Guria, et al., 2011; Mancini, et al., 2010) and *Thoc1* deficiency interferes with testis development (Wang, et al., 2009). Taken together, altered TREX function can have diverse molecular and cellular consequences resulting in a range of diseases. Here we present detailed information on the clinical presentations and functional investigations on an additional eight missense and two splice *THOC2* variants. These data reaffirm and extend our previous findings that *THOC2* variation is responsible for causing complex neurodevelopmental conditions with the core clinical presentation of ID.

MATERIALS AND METHODS

Molecular and cellular studies

RNA extraction, RT-qPCR (primers listed in Supp. Table S1), cycloheximide chase, and THOC2 immunofluorescence staining were performed as reported previously (Kumar, et al., 2015). Molecular studies on the *THOC2* exon35:c.4450-2A>G variant were performed using blood DNA and skin fibroblasts derived from the affected individual and his heterozygous carrier mother. Genomic DNA or cDNA was amplified with KAPA HiFi PCR Kit (Kapa Biosystems) using hTHOC2-4326F/ hTHOC2-4519-R (Supp. Table S1) at 95°C for 3 min, 35 cycles of 98°C-15sec, 59°C-15sec, 72°C-30sec, incubation at 72°C for 10 min, gel purified (MinElute Gel Extraction kit (Qiagen) and Sanger sequenced using the same primers. For the *THOC2* exon28:c.3503+4A>C, blood gDNA from unaffected father, carrier mother and affected son was amplified with TaKaRa ExTaq using THOC2-F/THOC2-R primers (Supp. Table S1) at 94°C for 2 min, 40 cycles of 94°C-30sec, 60°C-30sec, 72°C-30sec, incubation at 72°C for 5 min. The cDNA was generated by reverse transcribing the white blood cell RNAs using Superscript III reverse transcriptase (Life Technologies) and amplified with TaKaRa ExTaq using THOC2-ex27F/THOC2-ex30R (Supp. Table S1) at 94°C for 2 min, 28 cycles of 94°C-30sec, 60°C-30sec, 72°C-30sec, incubation at 72°C for 5 min. The amplified products were analyzed by Sanger sequencing.

Generation of *THOC2* variant expression constructs

Generation of the wild type Myc-tagged human THOC2 expression plasmid was reported earlier (Kumar, et al., 2015). Briefly, the *THOC2* variants were introduced into the existing pCMV-Myc-THOC2 expression construct by overlap PCR method using the primers listed in Supp. Table S1. The variant plasmid sequences were confirmed by Sanger sequencing.

Details relating generation of the *THOC2* variant expression constructs are available on request.

Transient expression and Western blotting

For transient expression experiments, HEK293T and HeLa cells were transfected with expression constructs (400ng pCMV-Myc-THOC2 plasmid and 400ng pEGFP-C1 plasmid/transfection for stability and cycloheximide assays and 4µg/transfection for immunofluorescence staining, IF) using Lipofectamine 3000 reagent according to manufacturer's protocol (Life technologies). Twenty-four hours post-transfection, cells were either fixed with 4% formaldehyde for IF or collected and lysed in buffer containing 50 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.1% Triton-X 100, 1 mM EDTA, 50 mM NaF, 1× Protease inhibitor/no EDTA and 0.1 mM Na_3VO_4 for western blot assay as reported previously (Kumar, et al., 2015).

RESULTS

Identification of *THOC2* variants

We previously implicated four missense *THOC2* variants in 25 individuals with ID and a range of other clinical features (Table 1) (Kumar, et al., 2015). We identified an additional five *THOC2* variants (three missense; *de novo* p.Thr696Ile, *de novo* p.Gly713Asp, maternally-inherited p.His1187Tyr, and two splicing-defective; maternally-inherited chrX:122747409 exon35:c.4450-2A>G and chrX:122757634 exon28:c.3503+4A>C; GenBank: NM_001081550.1) variants in a further six affected individuals, including one pair of monozygotic twins (Table 1, Figures 1-2). Whole exome (WES) or whole genome sequencing (WGS) of probands and parents was used to identify the variants that were confirmed by Sanger sequencing of the PCR amplified variant-carrying region of genomic

DNA of the parents and affected individuals. The previously unreported *THOC2* variants affect amino acids that are highly conserved (Supp. Figure S1), are absent in gnomAD database and are predicted to be pathogenic based on a number of *in silico* prediction criteria (Table 1). We also included in our study a *de novo* missense p.Tyr517Cys variant in a female with moderate-severe ID, speech problems, epileptic encephalopathy, cortical visual impairment, and gait disturbances identified using WES as part of the Epi4K Consortium & Epilepsy Phenome/Genome Project (Epi, et al., 2013) (Table 1). We have also collected further, rare and potentially pathogenic variants through international collaboration (Table1, Figures 1 and 3) and performed functional testing on several of these. The following three variants: p.Arg77Cys, p.Ser1012Pro and p.Asn1261His showed no clear evidence of altered stability of variant THOC2 proteins in our assay.

Clinical presentations

The clinical features of the five previously unreported affected individuals with (likely) pathogenic *THOC2* variants, aged between 3 and 12 years, and the 10 year old female with *de novo* p.Tyr517Cys variant are summarised in Table 2 and photographs, when available, are shown in Figure 2. Detailed clinical information is available in the supplementary data. ID was universal and at least moderate in severity: 2/7 were non-ambulatory and 3/7 non-verbal. Behavioural problems were reported in four individuals, with one meeting diagnostic criteria for autism spectrum disorder. Common additional neurological features include infantile hypotonia (4/7), tremor (2/7) and gait disturbance including balance problems/ataxia and toe walking (4/5 ambulatory individuals). Confirmed seizure disorder was only present in the affected female (individual 7) but suspected in individual 2. Neuro-radiological studies were performed in five individuals and reported as abnormal in three (available data presented in Supp. Figure S2). Growth abnormalities were common including low birth weight (3/7),

microcephaly (2/7) and short stature (2/7). Facial features are shown in Figure 2. Appropriate consent for reporting variants, clinical data and photographs of the affected individuals was obtained from their parents or legal guardians. The research has been approved by the Women's and Children's Health Network Human Research Ethics Committee in Adelaide, Australia.

THOC2 variant protein localisation and stability

Without access to affected individuals' derived cells, we generated Myc-tagged THOC2 missense variant expression constructs to determine protein stability and localisation. The THOC2 protein stability was determined in HEK293T cells and localisation in both the HEK293T and HeLa cells. Total protein lysates of HEK293T cells ectopically-expressing the wild type or variant Myc-THOC2 proteins were western blotted for THOC2, EGFP and β-Tubulin. We used HEK293T cells expressing Myc-p.Ile800Thr THOC2 as a control for protein stability assay as this variant is shown to cause reduced protein stability (Kumar, et al., 2015). The results showed reduced stability of p.Tyr517Cys, p.Thr696Ile, p.Gly713Asp and p.His1187Tyr THOC2 compared to the wild type protein (Figure 3A). Presence of comparable levels of EGFP in the cells transfected with different expression constructs indicated that the reduced levels of THOC2 protein were not due to difference in transfection efficiency (Figure 3). We also determined the turnover rate of Myc-p.Tyr517Cys THOC2 protein by cycloheximide chase. For this assay, the HEK293T cells transfected with pCMV-Myc-WT or pCMV-Myc-p.Tyr517Cys THOC2 and pEGFP-C1 transfection control plasmids were cultured in presence of translation inhibitor cycloheximide for different durations and western blotted for THOC2, EGFP and β-Tubulin. The results showed that p.Tyr517Cys THOC2 turnover rate was 3h compared with 8h for the wild type protein (Figure 3B).

THOC2 variant proteins, similar to the wild type, were mainly localised to the nucleus in both the HEK293T and HeLa cells (Supp. Figure S3).

***THOC2* splice variant: exon35:c.4450-2A>G**

Sanger sequencing of amplified target region from affected son and mother's blood genomic DNA showed that the affected boy inherited chrX:122747409 exon35:c.4450-2A>G variant from his unaffected heterozygous carrier mother (Figure 4C). A -2 A>G change in the intron-exon splicing site boundary (acceptor AG) is predicted to abolish splicing (Ohno, et al., 2017). To validate this possibility, we generated skin fibroblast cultures from the heterozygous carrier mother and the affected son. We PCR amplified their fibroblast cDNAs using primers with binding sequences located within exon 34 and 35. Amplification of a 194 bp DNA fragment from the mother indicated normal splicing but a 537 bp product from the affected son indicated retention of the intron located between these exons (Figure 4A-B). We confirmed this result by Sanger sequencing of the PCR products generated from genomic DNA that showed presence of A/G nucleotides in the carrier mother but only G (A>G) nucleotide in the affected son (Figure 4C). The cDNA sequence showed presence of normally-spliced mRNA in the mother but retention of intronic sequence upstream of the exon 35 in the affected son indicating defective splicing due to presence of -2 G variant at the intron-exon 35 junction sequence (Figure 4C). The presence of normally-spliced mRNA in the unaffected mother is consistent with X-inactivation (94% skewing) of the variant allele in her fibroblasts. We predicted that a retention of intron between exon 34-35 in the affected fibroblasts would result in loss of 110 C-terminal amino acids of the 1593 wild type THOC2 protein (that is, 1483 amino acids); however, overall the variant protein would be 58 amino acid smaller as it would now be a 1535 amino acid protein comprised of 1483 amino acids of wild type THOC2 and 52 translated from intronic sequence in the pre-mRNA (Figure 1A).

Consistent with our prediction, the western blot data showed presence of a slightly smaller THOC2 protein band in the affected son's fibroblasts than his unaffected mother. Many independent western blot runs showed presence of two closely-located THOC2 bands - similar to the fruit fly THO2 (Rehwinkel, et al., 2004) - in the unaffected mother but a single highly intense band in the affected son's fibroblasts (Figure 4D). The observed difference in levels of THOC2 protein was post-translational as we found comparable amounts of *THOC2* mRNA, as assayed by real time RT-qPCR, in the mother and son (Figure. 4E). Finally, we observed no difference in THOC2 localisation in fibroblasts of the affected son and his unaffected mother (Figure 4F).

***THOC2* splice variant: exon28:c.3503+4A>C**

For the second splice variant chrX:122757634 exon28:c.3503+4A>C, molecular studies were performed on the white blood cells of the unaffected father, carrier mother and the male proband. Sanger sequencing of target region amplified from the unaffected father and mother, and affected son's genomic DNA showed that the affected son inherited the A>C change from his unaffected carrier mother who had A/C nucleotides at this position (Figure 5). The intronic nucleotide change A>C at +4 position of the 5' exon-intron donor splicing site sequence is predicted to cause aberrant splicing (https://www.med.nagoya-u.ac.jp/neurogenetics/SD_Score/sd_score.html). To confirm this possibility, we amplified cDNA generated by reverse transcribing blood RNA of the father, mother and the affected son using primers located within exon 27 and exon 30 (Figure 5A; Supp. Table S1). Interestingly, whereas a 491 bp PCR product was observed in highly skewed carrier mother (98:2%) and normal father, 491 bp and 634 bp PCR products were detected in the affected son. A 491 bp amplified product indicated normal splicing in the mother and father, and 491 bp and 634 bp bands suggested partially-defective splicing in the affected son. Amplification

of a 634 bp instead of a 994 bp fragment that would have resulted from a complete retention of intron 28 indicated aberrant splicing event in the affected son (Figure 5). Sanger sequencing of 491 bp and 634 bp PCR products from the mother, father and son confirmed normal splicing in the mother and father and aberrant splicing in the affected son. The sequence showed retention of a 143 bp instead of complete 503 bp fragment due to activation of a cryptic splice site within the intron 28 in the son (Figure 5). Retention of 143 bps from intron 28 in the mRNA is predicted to result in a truncated THOC2 protein containing 1168 wild type amino acids and 7 novel amino acids from the translation of the intronic sequence retained in the defective mRNA, which would be present in addition to the wild type 1593 amino acid protein from the normally-spliced mRNA in the affected son.

Discussion

We present here detailed clinical information, and molecular and functional studies, on five previously unreported *THOC2* variants in six affected males (two *de novo* variants and one maternally inherited variant in monozygotic twins) and on one affected female with a previously reported *de novo* p.Tyr517Cys variant. We present evidence that extends the genotypic spectrum beyond the four *THOC2* missense variants that we reported previously (Kumar, et al., 2015) by including two intronic variants that affect splicing, and four missense variants that affect protein stability in a cell-based assay system. All of the variants meet ACMG criteria for pathogenic or likely pathogenic status (Table 1) (Richards, et al., 2015). These findings, along with the four missense variants reported earlier (Kumar, et al., 2015) add to the existing evidence that alterations in essential mRNA export pathway can cause NDDs (Amos, et al., 2017; Beaulieu, et al., 2013; Kumar, et al., 2015).

We confirm that the core clinical feature of *THOC2*-related disorder in hemizygous males is ID, with several individuals having additional neurocognitive features including infantile hypotonia, behavioural disorders and autism spectrum disorder, and (less commonly) gait disturbance and tremor. Although the range of neurodevelopmental features is similar, our original cohort contained males with borderline to mild ID (Kumar, et al., 2015); whereas, all individuals in this cohort have at least moderate ID. Five of the affected individuals had neuroimaging, with three having evidence of white matter thinning or cortical atrophy, and one individual (individual 2) having extensive cortical gyral abnormalities. Similarly, in the original cohort four of the six individuals with neuroimaging available were reported to have a variety of abnormalities including inferior cerebellar vermis dysplasia, ventriculomegaly, gliosis and mild cervical cord compression. A consistent pattern of extra-neurological features was not evident, although as in the original cohort, several individuals had evidence of restricted growth, including short stature, low birth weight and microcephaly. The truncal obesity, raised BMI and co-morbid anxiety and depression frequently noted in the original cohort was not observed in this new cohort, likely reflecting the relatively lower age range (4-12 years compared to 5-77 years) given that these features generally emerged in late adolescence/adulthood. We specifically evaluated the cohort to assess if a characteristic facial gestalt was evident across individuals with pathogenic or likely pathogenic variants across our original and this expanded clinical cohort (Supp. Figure S4) (Dudding-Byth, et al., 2017). Although a clearly recognizable facial gestalt was not obvious, there are some similarities. Children have a wide forehead with variable ptosis, epicanthic folds, down-slanting palpebral fissures, wide nasal bridge, long philtrum and pointed chin. Adults have a wide forehead with prominent supraorbital fissures with variable horizontal eyebrows, thin nasal bridge and down-slanting palpebral fissures. Age progression of the face shows shortening of the

philtrum with widening of the nasal base. The facial gestalt spectrum associated with *THOC2* pathogenic variants will continue to emerge as more individuals are reported.

The two males with splice variants (individuals 6 and 7) and the monozygotic twins with missense variant p.His1187Tyr (individual 3 and 4) inherited the *THOC2* variant from their mother. As was the case in our original cohort, the heterozygous mothers were unaffected, with normal learning and health. All had highly skewed X cell inactivation (XCI) (Table 1). In comparison, individual 7, who was a female with a *de novo* missense variant (p.Tyr517Cys) had a particularly severe neurocognitive presentation with infantile onset epileptic encephalopathy and profound intellectual disability. At age 10 y she remains non-ambulant and non-verbal. Unfortunately, we did not have access to her genomic DNA for testing the XCI status. We do note that severely affected females with *de novo* variants in X-linked genes have been reported by us and others (Palmer, et al., 2016; Snijders Blok, et al., 2015; Zweier, et al., 2014), and that differences in XCI in blood do not always provide an explanation for the range of phenotypic expression in heterozygous females.

A range of protein-protein interactions are required for mRNA export (Chi, et al., 2013). Proteins with altered stability (Hirayama, et al., 2008), localization (Beaulieu, et al., 2013) (e.g., THOC6 p.Gly46Arg implicated in syndromic ID) or interaction (Chi, et al., 2013) can impact mRNA export and consequently disrupt normal cell function. Likewise, *THOC2* variants can cause neuronal defects due to altered protein interactions, stability, localization and/or structure, which can impact normal THOC2 functions such as mRNA export or its mRNA binding properties. We did not observe mislocalization of the THOC2 variant proteins in cultured cells and did not test alterations in their interaction with the other known or unknown TREX proteins. However, significantly reduced levels of a number of new

(p.Tyr517Cys, p.His1187Tyr, p.Thr696Ile, p.Gly713Asp) and published (p.Leu438Pro, p.Ile800Thr (Kumar, et al., 2015)) missense THOC2 variant proteins are due to impaired protein stability or reduced levels of normal mRNA due to aberrant splicing in exon28:c.3503+4A>C affected individuals (see below). We also noted increased stability of p.Asn1261His THOC2 protein. We and others have shown that THOC2 controls TREX function by maintaining the stability of THOC1, 3, 5 and 7 subunits (Chi, et al., 2013; Kumar, et al., 2015). Reduced levels of THOC2 missense variant proteins are most likely due to enhanced proteasome-mediated degradation as THOC2 is ubiquitinated (Lopitz-Otsoa, et al., 2012). Published evidence shows that molecular chaperons protect proteins from degradation by promoting correct protein folding and assembly. Reduced stability of a set of variant THOC2 proteins may be due to structural changes or impaired protective-association with an unknown chaperon, which can be TREX complex subunit or an unrelated protein. Indeed, reduced levels of THOC2 protein in a girl with a *de novo* X;8 translocation creating a *PTK2-THOC2* gene fusion has also been implicated in non-progressive form of congenital ataxia, cognitive impairment and cerebellar hypoplasia (Di Gregorio, et al., 2013). This is consistent with observations showing that THOC2 depletion has different consequences in different organisms. Some of the examples include reduced levels of Thoc2 leading to significant increase in length of neurites in rat primary hippocampal neurons (Di Gregorio, et al., 2013) and *C. elegans thoc2* knockouts that are completely immobile, slow-growing, sterile, have functional defects in specific sensory neurons and die prematurely (Di Gregorio, et al., 2013). *D. rerio Thoc2* is essential for embryonic development (Amsterdam, et al., 2004) and in *D. melanogaster* S2 cells *Thoc2* knockdown inhibits mRNA export and cell proliferation (Rehwinkel, et al., 2004). THOC2 depletion also results in chromosome alignment, mitotic progression and genomic stability in human HeLa cells (Yamazaki, et al.,

2010). Finally, *Thoc2* and *Thoc5* knock down experiments have shown their role in regulation of embryonic stem cell self-renewal (Wang, et al., 2013).

Compared to his unaffected heterozygous carrier mother, who showed skewed X-inactivation, the exon35:c.4450-2A>G variant caused defective splicing in the male proband resulting with not only a truncated THOC2 protein but also abnormal accumulation in his skin fibroblasts. The truncated 1535 amino acid THOC2 protein lacks C-terminal 110 amino acids (part of a putative RNA binding domain) of the 1593 normal THOC2 protein but added 52 amino acids coded from the un-spliced intron (Figure 5; (Pena, et al., 2012)). Interestingly, whereas cDNA sequencing of exon 27-30 region for the exon28:c.3503+4A>C *THOC2* variant showed normal splicing in mother and father, the affected son showed aberrant splicing resulting in normally spliced (about 2/3rd) and intron 28 retained (nearly 1/3rd) mRNAs. A 143 bp intron 28 retention is predicted to translate a 1175 amino acid THOC2 protein (1168 normal and 7 amino acids coded by the retained intron 28 fragment) completely lacking the C-terminal RNA binding region (Figure 5). Both affected individuals carrying the splice-variants presented with severe neurocognitive features, consistent with the phenotype of the majority of individuals with missense variants with supportive functional evidence. Taken together, exon35:c.4450-2A>G and exon28:c.3503+4A>C *THOC2* splice variants present interesting biological scenarios; the former resulting in 1535 amino acid truncated protein that is present at higher level and the latter with both normal (albeit potentially much reduced) and a 1175 amino acid truncated THOC2 protein. We postulate that the clinical outcomes in the exon35:c.4450-2A>G individual are caused by partial loss of function due to loss of 110 amino acid C-terminal region and accumulation of the truncated THOC2 protein. However, pathogenicity in exon28:c.3503+4A>C affected individual is most likely caused by reduced levels of normal and potential dominant-negative effects of the C-terminally

truncated THOC2 protein. That reduced THOC2 protein levels are associated with ID and other clinical symptoms is emerging as a frequent theme; e.g., due to reduced THOC2 protein stability caused by missense variants (see above) and now an aberrant splicing event. Indeed reduced THOC2 levels are shown to destabilize the TREX complex in humans (Chi, et al., 2013; Kumar, et al., 2015) and removal of any THO subunit causes destabilization of other TREX components in yeast (Pena, et al., 2012). Interestingly, p.Arg1307Trp THOC2 variant was recently reported to effect THOC2-THOC1 binding and thus THOC1 degradation in MDAMB157 breast cancer cells (Lapek, et al., 2017).

Systematic studies on functional effect of Tho2 C-terminal RNA binding region in yeast provide interesting explanation as to how the truncated THOC2 protein can perturb normal mRNA export function in human cells (Pena, et al., 2012) (Figure 6). The yeast data show that whereas ΔTho2 strain does not grow at 37°C (restrictive temperature), $Tho2\Delta_{1408\text{-}1597}$ and $Tho\Delta_{1271\text{-}1597}$ growth is considerably reduced suggesting that C-terminal 1271-1597 amino acids are required for cell survival at restrictive temperature (Pena, et al., 2012). If the exon28:c.3503+A>C variant caused complete splicing defect retaining intron 28 in all mRNAs, the cells would translate only a 1175 amino acids (with 1168 normal) THOC2 protein; essentially lacking the C-terminal region encompassing the RBD domain that when deleted in yeast strain ($Tho\Delta_{1271\text{-}1597}$) totally abolishes their growth at 37°C. However, the affected boy carrying a single allele of the exon28:c.3503+A>C *THOC2* variant, although with severe clinical symptoms, is alive. This could be explained by presence of reduced levels of THOC2 protein produced from translation of about 2/3rd normally-spliced mRNA in the affected white blood cells. Taken together, clinical outcomes in the affected boy may be due to perturbed mRNA export caused by reduced levels of THOC2 protein and perhaps also C-terminally truncated non-functional THOC2 protein translated from about 1/3rd aberrantly

spliced mRNAs that retain intron 28. It is though likely that the severe effects due to truncated and/or reduced THOC2 protein in the affected individuals may be compensated by other TREX proteins and may explain the survival of the girl with congenital ataxia, cognitive impairment and cerebellar hypoplasia due to a *de novo* X;8 translocation creating a *PTK2-THOC2* gene fusion, who expresses reduced levels of the THOC2 protein (Di Gregorio, et al., 2013). For example, REF knockdown increases UIF protein levels (Hautbergue, et al., 2009) and Nxf1 depletion results in increased levels of THOC2 (Viphakone, et al., 2012).

We also identified a set of *THOC2* missense variants that do not reach ACMG criteria for (likely) pathogenicity. This includes four previously unreported variants (p.Arg77Cys; p.Ser1108Leu; p.Arg1121Gly and p.Asn1261His) in five affected individuals (including a brother pair) which are absent from databases of reference individuals (ExAC/ gnomAD) (Lek, et al., 2016), affect highly evolutionarily conserved amino acid residues, and *in silico* analyses are supportive of pathogenicity (Table 1). Consistent with not satisfying the ACMG criteria, they remain variants of uncertain clinical significance in the absence of additional supportive functional evidence. Specifically, the missense variants in individuals 8-10 and 12 did not affect protein stability in our cellular assay system. However, the variants may have a detrimental effect on THOC2 function due to altered protein structure impacting protein-RNA and/or protein-protein interactions with known or unknown TREX subunits. Indeed amyotrophic lateral sclerosis (ALS)-linked Matrin 3 variant proteins exhibit altered interactions with specific TREX proteins and lead to nuclear mRNA export defects of both global mRNA and more specifically, the TDP-43 and FUS mRNAs, linking cellular transport defects to ALS (Boehringer, et al., 2017). In addition, we note that two of the *THOC2* variants (p.Leu313Phe and p.Ser1012Pro) in our original cohort (Kumar, et al., 2015)

segregated with NDDs in multiple individuals in families, but also lacked evidence of a detrimental effect on protein stability. Detailed clinical information on the affected individuals carrying these VOUS is presented in the supplementary data and summarized in Table 2. Similarities in clinical presentation between the majority of affected individuals in this group and with those with (likely) pathogenic variants are apparent, including severe-profound intellectual disability, behavioural difficulties including autistic features, infantile hypotonia, seizure disorders, gait abnormalities, neuroradiological features including white matter abnormalities and ventricular dilatation, microcephaly and short stature. Individual 8 had cerebellar vermal hypoplasia which was also noted in an individual from the original cohort (Kumar, et al., 2015). Individuals 8 and 11 were considered to share overlapping facial features with other members of the cohort (Supp. Figure S4). However, many of these features are relatively common in neurodevelopmental disorders in general, and we do not yet consider the phenotype of *THOC2*-related disorder to be distinctive enough to strongly weight assessment of pathogenicity of novel variants. The challenge of proving causality for previously unreported missense variants in NDD genes is well recognized and speaks to the need for ongoing intertwined clinical and research efforts to clarify causality of VOUS (Wright, et al., 2018).

THOC2 is ubiquitously expressed in all human tissues (Thul, et al., 2017) and is highly expressed in the developing and mature human (Johnson, et al., 2009; Kumar, et al., 2015; Uhlen, et al., 2015) and mouse brain, with higher abundance in frontal cortex and cerebellum (Di Gregorio, et al., 2013; Kumar, et al., 2015). THOC2 is an essential mRNA export factor as its siRNA-mediated depletion results in almost complete retention of mRNAs in the cell nucleus (Chi, et al., 2013), potentially toxic to the cell. These data are consistent with the findings that *THOC2* is a highly-constrained gene (Samocha, et al., 2014) and *THOC* (e.g.

THOC1, *3*, *5*, *6* and *7*) genes are essential for cell survival (Blomen, et al., 2015). Taken together, as *THOC2* knockout cells will not survive due to complete mRNA nuclear retention, we predict that the identified *THOC2* variants represent partial loss-of-function that disrupt normal mRNA export in neuronal and possibly other cell types, potentially causing variable clinical presentations.

TREX complex couples transcription and mRNA biogenesis with nuclear mRNA export and has emerged as an essential pathway in embryogenesis, organogenesis and differentiation (Heath, et al., 2016). For example, Thoc2 and Thoc5 selectively bind and regulate export of mRNAs involved in maintenance of pluripotency (e.g. Nanog, Sox2, Esrrb, and Klf4 mRNAs) (Wang, et al., 2013) and Thoc5 in hematopoiesis and heat shock (Katahira, et al., 2009; Mancini, et al., 2010). Lack of *Thoc1* and *Thoc5* in mice is embryonic lethal (Mancini, et al., 2010; Wang, et al., 2006). Mouse modeling shows that both *Thoc1* and *Thoc5* knockouts are embryonic lethal (Mancini, et al., 2010; Wang, et al., 2006). However, *Thoc1* and *Thoc5* expression in a range of developing and adult tissues may indicate that the two genes have more essential role in early embryonic development compared to less stringent requirement during later stages of embryonic or adult development (Mancini, et al., 2010; Wang, et al., 2006); a functional pattern most likely also true for the *THOC2* gene. Essentiality of *THOC2* gene indicates that *THOC2* knockout will also be lethal. However, reduced levels or perturbed functionality can lead to a range of NDD phenotypes as observed for a cohort of THOC2 variants identified by us. It is now well-established that development of brain depends on tightly regulated and complex sequence of events involving neuronal and glial cell proliferation, migration and maturation (Chiurazzi and Pirozzi, 2016). Therefore, it is not surprising that our *THOC2* variant data and published work (Dickinson, et al., 2016) provides strong evidence that even subtle alterations to the canonical molecular pathways

such as mRNA export, otherwise essential for cellular life, can be tolerated but at a cost of a NDD.

In summary, we present detailed clinical data on seven individuals with *THOC2*-associated ID caused by both missense and splice variants that meet ACMG criteria for (likely) pathogenicity. They have a core phenotype of ID, and common findings of behavioural disorders, infantile hypotonia, gait disturbance and growth impairment, similar to the affected males with THOC2-associated ID we previously reported (Kumar, et al., 2015). Other than the affected female with a *de novo* missense variant, heterozygote carrier females are typically unaffected. We also present clinical data on five individuals with four previously unreported rare missense variants that show clinical overlap with our core group, but where convincing evidence for causality is still required. The significance of these variants may be clarified as further individuals with *THOC2* variants are reported.

Acknowledgements

We thank the individuals and families for their contribution to this study. This work was supported by NHMRC Program Grant APP1091593 and Senior Research Fellowship APP1041920 and Channel 7 Children's Research Foundation to J.G., research grants APP19PI10/01710 from the Fondo de Investigación Sanitaria; the Spanish Ministry of Economy and Competitiveness, 'Centro de Excelencia Severo Ochoa 2017-2021', SEV-2016-0571; a MINECO Severo Ochoa fellowship to LD (SVP-2013-0680066); FORGE Canada and Care4Rare Canada Consortium funded by Genome Canada, the Canadian Institutes of Health Research (CIHR), the Ontario Genomics Institute, Ontario Research Fund, Génome Québec, and Children's Hospital of Eastern Ontario Foundation; the National Human Genome Research Institute, the National Eye

Institute, and the National Heart, Lung and Blood Institute grant UM1 HG008900 to Daniel MacArthur and Heidi Rehm; the Undiagnosed Diseases Program-Victoria (UDP-Vic), Murdoch Children's Research Institute, Melbourne, Australia.

References

Amos JS, Huang L, Thevenon J, Kariminedjad A, Beaulieu CL, Masurel-Paulet A, Najmabadi H, Fattahi Z, Beheshtian M, Tonekaboni SH and others. 2017. Autosomal recessive mutations in THOC6 cause intellectual disability: syndrome delineation requiring forward and reverse phenotyping. Clin Genet 91(1):92-99.

Amsterdam A, Nissen RM, Sun Z, Swindell EC, Farrington S, Hopkins N. 2004. Identification of 315 genes essential for early zebrafish development. Proc Natl Acad Sci U S A 101(35):12792-7.

Beaulieu CL, Huang L, Innes AM, Akimenko MA, Puffenberger EG, Schwartz C, Jerry P, Ober C, Hegele RA, McLeod DR and others. 2013. Intellectual disability associated with a homozygous missense mutation in THOC6. Orphanet J Rare Dis 8(1):62.

Bittles AH, Petterson BA, Sullivan SG, Hussain R, Glasson EJ, Montgomery PD. 2002. The influence of intellectual disability on life expectancy. J Gerontol A Biol Sci Med Sci 57(7):M470-2.

Blomen VA, Majek P, Jae LT, Bigenzahn JW, Nieuwenhuis J, Staring J, Sacco R, van Diemen FR, Olk N, Stukalov A and others. 2015. Gene essentiality and synthetic lethality in haploid human cells. Science 350(6264):1092-6.

Boehringer A, Garcia-Mansfield K, Singh G, Bakkar N, Pirrotte P, Bowser R. 2017. ALS Associated Mutations in Matrin 3 Alter Protein-Protein Interactions and Impede mRNA Nuclear Export. Sci Rep 7(1):14529.

Chi B, Wang Q, Wu G, Tan M, Wang L, Shi M, Chang X, Cheng H. 2013. Aly and THO are required for assembly of the human TREX complex and association of TREX components with the spliced mRNA. Nucleic Acids Res 41(2):1294-306.

Chinnam M, Wang Y, Zhang X, Gold DL, Khoury T, Nikitin AY, Foster BA, Li Y, Bshara W, Morrison CD and others. 2014. The Thoc1 ribonucleoprotein and prostate cancer progression. J Natl Cancer Inst 106(11).

Chiurazzi P, Pirozzi F. 2016. Advances in understanding - genetic basis of intellectual disability. F1000Res 5.

Coe BP, Witherspoon K, Rosenfeld JA, van Bon BW, Vulto-van Silfhout AT, Bosco P, Friend KL, Baker C, Buono S, Vissers LE and others. 2014. Refining analyses of copy number variation identifies specific genes associated with developmental delay. Nat Genet 46(10):1063-71.

Di Gregorio E, Bianchi FT, Schiavi A, Chiotto AM, Rolando M, Verdun di Cantogno L, Grosso E, Cavalieri S, Calcia A, Lacerenza D and others. 2013. A de novo X;8 translocation creates a PTK2-THOC2 gene fusion with THOC2 expression knockdown in a patient with psychomotor retardation and congenital cerebellar hypoplasia. J Med Genet 50(8):543-51.

Dickinson ME, Flenniken AM, Ji X, Teboul L, Wong MD, White JK, Meehan TF, Weninger WJ, Westerberg H, Adissu H and others. 2016. High-throughput discovery of novel developmental phenotypes. Nature 537(7621):508-514.

Dudding-Byth T, Baxter A, Holliday EG, Hackett A, O'Donnell S, White SM, Attia J, Brunner H, de Vries B, Koolen D and others. 2017. Computer face-matching technology using two-dimensional photographs accurately matches the facial gestalt of unrelated individuals with the same syndromic form of intellectual disability. BMC Biotechnol 17(1):90.

Epi KC, Epilepsy Phenome/Genome P, Allen AS, Berkovic SF, Cossette P, Delanty N, Dlugos D, Eichler EE, Epstein MP, Glauser T and others. 2013. De novo mutations in epileptic encephalopathies. Nature 501(7466):217-21.

Guria A, Tran DD, Ramachandran S, Koch A, El Bounkari O, Dutta P, Hauser H, Tamura T. 2011. Identification of mRNAs that are spliced but not exported to the cytoplasm in the absence of THOC5 in mouse embryo fibroblasts. RNA 17(6):1048-56.

Hautbergue GM. 2017. RNA Nuclear Export: From Neurological Disorders to Cancer. Adv Exp Med Biol 1007:89-109.

Hautbergue GM, Hung ML, Walsh MJ, Snijders AP, Chang CT, Jones R, Ponting CP, Dickman MJ, Wilson SA. 2009. UIF, a New mRNA export adaptor that works together with REF/ALY, requires FACT for recruitment to mRNA. Curr Biol 19(22):1918-24.

Heath CG, Viphakone N, Wilson SA. 2016. The role of TREX in gene expression and disease. Biochem J 473(19):2911-35.

Hirayama S, Yamazaki Y, Kitamura A, Oda Y, Morito D, Okawa K, Kimura H, Cyr DM, Kubota H, Nagata K. 2008. MKKS is a centrosome-shuttling protein degraded by disease-causing mutations via CHIP-mediated ubiquitination. Mol Biol Cell 19(3):899-911.

Hosking FJ, Carey IM, Shah SM, Harris T, DeWilde S, Beighton C, Cook DG. 2016. Mortality Among Adults With Intellectual Disability in England: Comparisons With the General Population. Am J Public Health 106(8):1483-90.

Johnson MB, Kawasawa YI, Mason CE, Krsnik Z, Coppola G, Bogdanovic D, Geschwind DH, Mane SM, State MW, Sestan N. 2009. Functional and evolutionary insights into human brain development through global transcriptome analysis. Neuron 62(4):494-509.

Katahira J, Inoue H, Hurt E, Yoneda Y. 2009. Adaptor Aly and co-adaptor Thoc5 function in the Tap-p15-mediated nuclear export of HSP70 mRNA. EMBO J 28(5):556-67.

Kim W, Bennett EJ, Huttlin EL, Guo A, Li J, Possemato A, Sowa ME, Rad R, Rush J, Comb MJ and others. 2011. Systematic and quantitative assessment of the ubiquitin-modified proteome. Mol Cell 44(2):325-40.

Kumar R, Corbett MA, van Bon BW, Woenig JA, Weir L, Douglas E, Friend KL, Gardner A, Shaw M, Jolly LA and others. 2015. THOC2 Mutations Implicate mRNA-Export Pathway in X-Linked Intellectual Disability. Am J Hum Genet 97(2):302-10.

Lapek JD, Jr., Greninger P, Morris R, Amzallag A, Pruteanu-Malinici I, Benes CH, Haas W. 2017. Detection of dysregulated protein-association networks by high-throughput proteomics predicts cancer vulnerabilities. Nat Biotechnol 35(10):983-989.

Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, O'Donnell-Luria AH, Ware JS, Hill AJ, Cummings BB and others. 2016. Analysis of protein-coding genetic variation in 60,706 humans. Nature 536(7616):285-91.

Liu C, Yue B, Yuan C, Zhao S, Fang C, Yu Y, Yan D. 2015. Elevated expression of Thoc1 is associated with aggressive phenotype and poor prognosis in colorectal cancer. Biochem Biophys Res Commun 468(1-2):53-8.

Lopitz-Otsoa F, Rodriguez-Suarez E, Aillet F, Casado-Vela J, Lang V, Matthiesen R, Elortza F, Rodriguez MS. 2012. Integrative analysis of the ubiquitin proteome isolated using Tandem Ubiquitin Binding Entities (TUBEs). J Proteomics 75(10):2998-3014.

Mancini A, Niemann-Seyde SC, Pankow R, El Bounkari O, Klebba-Farber S, Koch A, Jaworska E, Spooncer E, Gruber AD, Whetton AD and others. 2010. THOC5/FMIP, an mRNA export TREX complex protein, is essential for hematopoietic primitive cell survival in vivo. BMC Biol 8:1.

Milani D, Ronzoni L, Esposito S. 2015. Genetic Advances in Intellectual Disability. J Pediatr Genet 4(3):125-7.

Morgan DO. 2013. The D box meets its match. Mol Cell 50(5):609-10.

Oeseburg B, Dijkstra GJ, Groothoff JW, Reijneveld SA, Jansen DE. 2011. Prevalence of chronic health conditions in children with intellectual disability: a systematic literature review. Intellect Dev Disabil 49(2):59-85.

Ohno K, Takeda JI, Masuda A. 2017. Rules and tools to predict the splicing effects of exonic and intronic mutations. Wiley Interdiscip Rev RNA.

Olsen JV, Blagoev B, Gnad F, Macek B, Kumar C, Mortensen P, Mann M. 2006. Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. Cell 127(3):635-48.

Palmer EE, Kumar R, Gordon CT, Shaw M, Hubert L, Carroll R, Rio M, Murray L, Leffler M, Dudding-Byth T and others. 2017. A Recurrent De Novo Nonsense Variant in ZSWIM6 Results in Severe Intellectual Disability without Frontonasal or Limb Malformations. Am J Hum Genet 101(6):995-1005.

Palmer EE, Stuhlmann T, Weinert S, Haan E, Van Esch H, Holvoet M, Boyle J, Leffler M, Raynaud M, Moraine C and others. 2016. De novo and inherited mutations in the X-linked gene CLCN4 are associated with syndromic intellectual disability and behavior and seizure disorders in males and females. Mol Psychiatry.

Pena A, Gewartowski K, Mroczek S, Cuellar J, Szykowska A, Prokop A, Czarnocki-Cieciura M, Piwowarski J, Tous C, Aguilera A and others. 2012. Architecture and nucleic acids recognition mechanism of the THO complex, an mRNP assembly factor. EMBO J 31(6):1605-16.

Rehwinkel J, Herold A, Gari K, Kocher T, Rode M, Ciccarelli FL, Wilm M, Izaurralde E. 2004. Genome-wide analysis of mRNAs regulated by the THO complex in Drosophila melanogaster. Nat Struct Mol Biol 11(6):558-66.

Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E and others. 2015. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med 17(5):405-24.

Ropers HH, Hamel BC. 2005. X-linked mental retardation. Nat Rev Genet 6(1):46-57.

Samocha KE, Robinson EB, Sanders SJ, Stevens C, Sabo A, McGrath LM, Kosmicki JA, Rehnstrom K, Mallick S, Kirby A and others. 2014. A framework for the interpretation of de novo mutation in human disease. Nat Genet 46(9):944-50.

Snijders Blok L, Madsen E, Juusola J, Gilissen C, Baralle D, Reijnders MR, Venselaar H, Helsmoortel C, Cho MT, Hoischen A and others. 2015. Mutations in DDX3X Are a Common Cause of Unexplained Intellectual Disability with Gender-Specific Effects on Wnt Signaling. Am J Hum Genet 97(2):343-52.

Thul PJ, Akesson L, Wiking M, Mahdessian D, Geladaki A, Ait Blal H, Alm T, Asplund A, Bjork L, Breckels LM and others. 2017. A subcellular map of the human proteome. Science 356(6340).

Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, Sivertsson A, Kampf C, Sjostedt E, Asplund A and others. 2015. Proteomics. Tissue-based map of the human proteome. Science 347(6220):1260419.

Viphakone N, Cumberbatch MG, Livingstone MJ, Heath PR, Dickman MJ, Catto JW, Wilson SA. 2015. Luzp4 defines a new mRNA export pathway in cancer cells. Nucleic Acids Res 43(4):2353-66.

Viphakone N, Hautbergue GM, Walsh M, Chang CT, Holland A, Folco EG, Reed R, Wilson SA. 2012. TREX exposes the RNA-binding domain of Nxf1 to enable mRNA export. Nat Commun 3:1006.

Vissers LE, Gilissen C, Veltman JA. 2016. Genetic studies in intellectual disability and related disorders. Nat Rev Genet 17(1):9-18.

Wagner SA, Beli P, Weinert BT, Nielsen ML, Cox J, Mann M, Choudhary C. 2011. A proteome-wide, quantitative survey of in vivo ubiquitylation sites reveals widespread regulatory roles. Mol Cell Proteomics 10(10):M111 013284.

Wang L, Miao YL, Zheng X, Lackford B, Zhou B, Han L, Yao C, Ward JM, Burkholder A, Lipchina I and others. 2013. The THO complex regulates pluripotency gene mRNA export and controls embryonic stem cell self-renewal and somatic cell reprogramming. Cell Stem Cell 13(6):676-90.

Wang X, Chang Y, Li Y, Zhang X, Goodrich DW. 2006. Thoc1/Hpr1/p84 is essential for early embryonic development in the mouse. Mol Cell Biol 26(11):4362-7.

Wang X, Chinnam M, Wang J, Wang Y, Zhang X, Marcon E, Moens P, Goodrich DW. 2009. Thoc1 deficiency compromises gene expression necessary for normal testis development in the mouse. Mol Cell Biol 29(10):2794-803.

Woerner AC, Frottin F, Hornburg D, Feng LR, Meissner F, Patra M, Tatzelt J, Mann M, Winklhofer KF, Hartl FU and others. 2016. Cytoplasmic protein aggregates interfere with nucleocytoplasmic transport of protein and RNA. Science 351(6269):173-6.

Wright CF, McRae JF, Clayton S, Gallone G, Aitken S, FitzGerald TW, Jones P, Prigmore E, Rajan D, Lord J and others. 2018. Making new genetic diagnoses with old data: iterative reanalysis and reporting from genome-wide data in 1,133 families with developmental disorders. Genet Med.

Yamazaki T, Fujiwara N, Yukinaga H, Ebisuya M, Shiki T, Kurihara T, Kioka N, Kambe T, Nagao M, Nishida E and others. 2010. The closely related RNA helicases, UAP56 and URH49, preferentially form distinct mRNA export machineries and coordinately regulate mitotic progression. Mol Biol Cell 21(16):2953-65.

Zhu X, Need AC, Petrovski S, Goldstein DB. 2014. One gene, many neuropsychiatric disorders: lessons from Mendelian diseases. Nat Neurosci 17(6):773-81.

Zweier C, Rittinger O, Bader I, Berland S, Cole T, Degenhardt F, Di Donato N, Graul-Neumann L, Hoyer J, Lynch SA and others. 2014. Females with de novo aberrations in PHF6: clinical overlap of Borjeson-Forssman-Lehmann with Coffin-Siris syndrome. Am J Med Genet C Semin Med Genet 166C(3):290-301.

Figure legends

FIGURE 1 Location of variant amino acids and structural features in THOC2 protein. Ubiquitinated (K139 and K1084 (Kim, et al., 2011) and K345 (Lopitz-Otsoa, et al., 2012; Wagner, et al., 2011)), phosphorylated (S1448 and 1486 (Olsen, et al., 2006)) amino acid residues, potential RNA binding domain (RBD) and destruction box (D-box) and KEN box sequences that interact with the Anaphase Promoting Complex/Cyclosome (APC/C) for protein ubiquitination and subsequent destruction by the proteasome (Morgan, 2013) are shown. Unreported (orange) and published (black: (Kumar, et al., 2015)) missense variants effecting THOC2 protein stability are marked with red lollipops. The positions of two splice variants are also shown (red).

FIGURE 2 Front and side facial views of the affected individuals with *THOC2* variants.

FIGURE 3 Functional testing of *THOC2* missense variants. **A:** THOC2 variant protein stability is reduced in HEK293T cells. pCMV-Myc-THOC2 wild-type or variant expression constructs and pEGFP-C1 plasmid (transfection control) were transfected into HEK293T cells. Total protein lysates of cells 24hr post-transfection were analysed by western blotting with mouse anti-Myc (clone 9E10; Sigma), mouse anti-EGFP (clones 7.1 and 13.1; Roche) and rabbit anti-β-tubulin (loading control; Abcam) antibodies. pCMV-Myc-THOC2 p.Ile800Thr construct expressing the p.Ile800Thr protein shown to have reduced stability was used as a control (Kumar, et al., 2015). Western blot signals were quantified using ImageJ software. Averages of the Myc-THOC2 proteins normalised to the housekeeping β-tubulin signal from two independent runs are shown. **B:** Myc-Tyr517Cys THOC2 protein half-life is significantly reduced in HEK293T cells. pCMV-Myc-THOC2 or pCMV-Myc-THOC2-Tyr517Cys expression constructs and pEGFP-C1 plasmid (transfection control) were

transfected into HEK293T cells. Next day the cells were cultured in the presence of 100 μg/ml translation inhibitor cycloheximide and cells harvested at the time points shown. Total protein lysates were analysed by western blotting with mouse anti-Myc, mouse anti-EGFP and rabbit anti-β-tubulin (loading control) antibodies.

FIGURE 4 Exon35:c.4450-2A>G variant abolishes splicing of intron between exons 34-35. **A:** Part of the *THOC2* gene showing location of the A/G nucleotide in the heterozygous carrier mother and A>G splice variant in the affected son. The C-terminal part of the 1593 amino acid wild type and 1535 amino acid (that contains 1483 normal and 52 amino acids coded by the unspliced intron) THOC2 protein in the affected boy are also shown. **B:** Gel showing a 194 bp RT-PCR product from the normally-spliced mRNA of the heterozygous carrier mother and a 537bp product from defective splicing of mRNA causing retention of an intron between exon 34-35 in the affected son. RT-PCR products from total RNA isolated from passage 3 (lanes 1-2) and 5 (lanes 3-4) fibroblasts. Location of the forward and reverse primers within exons 34 and 35 is shown. **C:** Sanger sequencing chromatograms of PCR products amplified from genomic and cDNA of the affected son and his heterozygous carrier mother using primers located within exon 34 and 35. Genomic DNA around the Exon34-Intron-Exon35 region is shown. **D:** Western blot showing THOC2 protein in the affected son and his carrier mother's skin fibroblasts. TREX subunit UAP56 was used as a loading control. **E:** RT-qPCR showing levels of the THOC2 mRNA in the affected son and his carrier mother's skin fibroblasts. **F:** Immunofluorescence detection of THOC2 in skin fibroblasts of the unaffected mother and affected son.

FIGURE 5 Exon28:c.3503+4A>C variant causes aberrant splicing of intron 28 **A:** Part of the *THOC2* gene showing location of the A/C nucleotide in the heterozygous carrier mother and

A>C splice variant in the affected son. The C-terminal part of the 1593 amino acid wild type and 1175 amino acid (that contains 1168 normal and 7 amino acids coded by the unspliced intron) THOC2 protein in the affected son are also shown. **B:** Gel showing a 491 bp RT- PCR product from the normally-spliced heterozygous carrier mother and unaffected father, and 491 bp and 634 bp (retaining 143 bp of the 503 bp intron 28) RT-PCR products derived from the normally and aberrantly spliced mRNAs, respectively, in the affected son. Location of the forward and reverse primers within exons 27 and 30 is shown. **C:** Sanger sequencing chromatograms of PCR products amplified using primers located within exons 27 and 30 from genomic and cDNA of unaffected father and mother (carrier with 98:2% skewing), and the affected son.

FIGURE 6 Summary of truncated human THOC2 proteins translated from aberrantly-spliced mRNAs and functional outcomes of yeast C-terminal Tho2 deletion strains (Pena, et al., 2012). Blue boxes depict the 52 and 7 amino acids coded by unspliced intron sequences of exon35:c.4550-2T>C and exon28:c.3503+4A>C variants, respectively. WT, +++ = normal, Δ1408-1597aa, ++ and Δ1271-1597aa = reduced growth at restrictive temperature.

TABLE 1 Detailed description of the *THOC2* variants.

TABLE 2 Summary of clinical data.

Supplementary data files

Supplementary file with detailed clinical data

Supplementary Figure S1. THOC2 variant amino acid residues are highly conserved. Amino acid sequence alignments show the variant amino acid residue (red) at the conserved positions (green). The amino acid sequences were aligned using www.uniprot.org/align/.

Supplementary Figure S2 Brain MRI of the individual carrying the exon28: c.3503+4A>C *THOC2* variant. Red arrow shows myelination and yellow arrow shows periventricular white matter lesion at age 2y 11m.

Supplementary Figure S3 A: THOC2 variant proteins are normally localised in HeLa and HEK293T cells. HeLa and HEK293T cells cultured on Poly-L-lysine coated coverslips (BD Biosciences) were transfected with wild type or variant Myc-THOC2 expression plasmids and detected by immunofluorescence using mouse anti-Myc (9E10; Sigma) primary and Donkey anti-rabbit-IgG Alexa 488 secondary antibody (Life Technologies). Only HeLa cell localisations at 630× magnification are shown. **B:** Nuclear and cytoplasmic (HEK293T; average 250 cells each) and nuclear and nuclear plus cytoplasmic (HeLa; average 400 cells each) staining was visually scored and plotted. Note that nuclear plus cytoplasmic instead of only cytoplasmic staining for HeLa was scored because all the cells with cytoplasmic staining also had nuclear staining.

Supplementary Figure S4 Children and adolescents/adults within current and previously reported *THOC2* cohorts. Images connected by vertical grey shading represent the same

individual at different ages.

Supplementary Table S1 Primers list (PCR and overlap PCR mutagenesis).

TABLE 1 Detailed description of the *THOC2* variants.

		Individual	From	Method of identification	Position	NM_001081550	Mode of inheritance	CADD	Provean score	Provean prediction	GERP++	Phylop	gnomAD frequency	Polyphen2*	Variant plasmid tested	Reduced Protein Stability	ACMG pathogenicity classification**
Rare variants with supporting molecular evidence	Missense		Kumar et al 2015 AJHG	X-Chr Exome sequencing	122799566	c.1313T>C:p.Leu438Pro	Maternal inheritance	28.1	-6.08	Deleterious	5.7	1.902	absent	D	YES	YES	LP
		1	Australia	Trio WES	122767853	c.2087C>T:p.Thr696Ile	De novo	27.4	-5.47	Deleterious	5.03	0.963	absent	D	YES	YES	DP
		2	USA	WES	122766890	c.2138G>A:p.Gly713Asp	De novo	31	-4.69	Deleterious	5.73	2.412	absent	D	YES	YES	DP
			Kumar et al 2015 AJHG	X-Chr Exome sequencing	122765621	c.2399T>C:p.Ile800Thr	Maternal inheritance	23.9	-4.01	Deleterious	5.97	2.016	absent	P	YES	YES	LP
		3-4	Canada/Germany/Russia Identical twins	WES	122757079	c.3559C>T:p.His1187Tyr	Maternal inheritance; Mother skewed (99.9:0.1%)	23.1	-5.07	Deleterious	6.07	2.571	absent	P	YES	YES	LP
	Splice	5	Japan	Trio WES	122757634	Exon28:c.3503+4A>C	Maternal inheritance; Mother skewed 98:2%	10.8	N/A	N/A	5.57	1.86	absent	N/A	N/A	ND	LP
		6	Canada	WES	122747561	Exon35:c.4450-2A>G	Maternal inheritance; Mother skewed 94:6%	23.7	N/A	N/A	5.25	1.735	absent	N/A	Fibroblasts of the affected male proband and carrier mother	NO	LP
	Missense	7	USA Epi4K Consortium & Epilepsy Phenome/Genome Project; Nature 501:217-221, 2013	WES	122778639	c.1550A>G:p.Tyr517Cys	De novo	26.6	-7.87	Deleterious	5.84	1.955	absent	D	YES	YES	DP

Rare variants without supporting molecular evidence	Missense	8	USA	WES	122837349	c.229 C>T:p.Arg77Cys	Maternal inheritance	25.8	-3.07	Deleterious	5.09	2.219	absent	P	YES	NO	VOUS
			Kumar et al 2015 AJHG	X-Chr Exome sequencing	122802090	c.937C>T:p.Leu313Phe	Maternal inheritance	23.8	-2.44	Neutral	5.38	2.248	absent	D	YES	NO	VOUS
			Kumar et al 2015 AJHG	X-Chr Exome sequencing	122759786	c.3034T>C:p.Ser1012Pro	Maternal inheritance	23.8	-4.54	Deleterious	5.6	1.87	absent	P	YES	NO	VOUS
		9-10	Spain	WES of two affected brothers	122757818	c.3323C>T:p.Ser1108Leu	Variant not detected in mother's peripheral blood or hair roots so likely germline mosaicism	15.4	0.96	Neutral	5.91	2.493	absent	B	YES	NO	LP or VOUS
		11	Spain	WES parents and affected individual	122757780	c.3361A>G:p.Arg1121Gly	Maternal inheritance	24.2	-6.49	Deleterious	5.91	0.834	absent	D	NO	NO	VOUS
		12	USA	WES	122756613	c.3781A>C:p.Asn1261His	Maternal inheritance; testing of family members shows likely new variant in mother	22.8	-1.81	Neutral	5.34	1.768	absent	D	YES	NO	VOUS

*Probably damaging, D; Possibly damaging, P; Benign, B; N/A, not applicable; ND, not determined **De novo Pathogenic, DP; Likely Pathogenic, LP; Variant of unknown Significance, VOUS

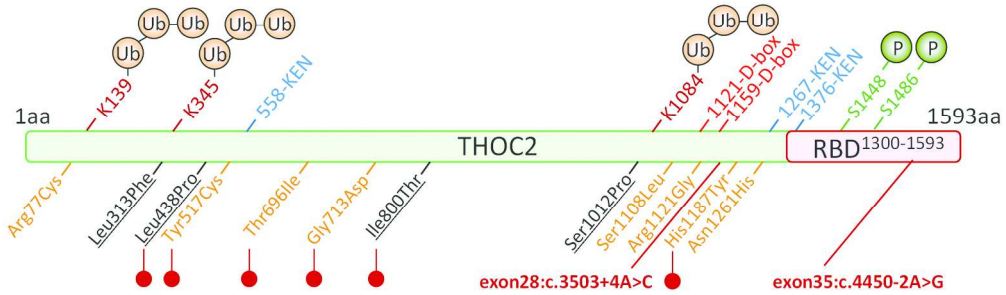
TABLE 2 Summary of clinical data.

	(likely) pathogenic							VOUS				
Individual	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
Variant details	c.2087C>T: p.Thr696Ile	c.2138G>A: p.Gly713Asp	c.3559C>T: p.His1187Tyr Twin 1	c.3559C>T: p.His1187Tyr Twin 2	Exon28: c.3503+4A>C	Exon35:c.4450-2A>G	c.1550A>G: p.Tyr517Cys	c.229C>T: p.Arg77Cys	c.3323C>T: p.Ser1108Leu Sib 1	c.3323C>T: p.Ser1108Leu Sib 2	c.3361A>G: p.Arg1121Gly	c.3781A>C: p.Asn1261His
Gender	Male	Male	Male	Male	Male	Male	Female	Male	Male	Male	Male	Male
Age (years)	12	5	7	7	3	10	10	9	17	12	44	6
Perinatal features												
Gestation (weeks)	36	37	37	37	37	41	NA	41	40	37	Premature	Term
Low birth weight (<2.5kg)	Yes	No	Yes	Yes	No	No	NA	No	No	No	No	No
Birth weight (g)	2000	2650	1990	2420	3018	4365	NA	3770	2880	3480	NA	5200
Neurologic features												
Intellectual disability	Severe	Mod+	Mod+	Mod+	Severe	Severe	Profound	Profound	Severe	Severe	Severe	Mod+
Speech delay	Yes, single words, signs	Yes, non-verbal	Yes	Yes	Yes, non-verbal	Yes, nasal dysarthria	Yes, non-verbal	Yes, non-verbal	Yes	Yes	Yes, severe	Yes, 2-3 word phrases
Hypotonia	No	Yes	NA	NA	Yes	Yes, central hypotonia appendicular spasticity	Yes	Yes, central hypotonia, appendicular spasticity	No	No	Yes	Yes, central hypotonia, appendicular spasticity
Hyperkinesia	No	No	Yes	Yes	Yes	No	No	No	No	Yes	No	No
Tremor	No	Yes, intermittent	No	No	No	Yes	No	No	No	No	No	No
Epilepsy	No	Suspected	No	No	No	No	Yes, epileptic encephalopathy	Yes, infantile spasms	No	No	No	Yes, Lennox-Gestaut
Gait disturbances	No	Yes, gait/balance problems	Yes, toe walking	Yes, toe walking	Non ambulatory	Yes, ataxia/broad based gait, spasticity	Non ambulatory	Non ambulatory	Yes, 'clumsy' and toe walking	Yes, 'clumsy' and toe walking	Non ambulatory	Yes, spastic quadiplegia
Behaviour problems	No	Yes	Yes	Yes	Yes, ASD	No	NR	No	Yes, ASD	Yes, ASD	Yes	Yes, short attention span
Anxiety	No	No	No	No	No	No	NR	No	Yes	Yes	No	No
Depression	No	No	No	No	No	No	NR	No	No	No	No	No

Brain MRI/CT	MRI normal	Thin corpus callosum, dysgyria	ND	ND	Ventricular dilatation, delayed myelination, peri-ventricular white matter abnormality	Slight prominence of cerebrofluid spaces.	MRI normal	Cerebellar vermal hypoplasia, thin corpus callosum, dilated ventricles	ND	ND	ND	Unilateral Periventricular leucomalacia
Growth parameters												
Microcephaly (≤3%)	Yes	Yes,< 1%	No	No	No	Yes,2%	No, 5%	No	No	NA	Yes	No
Short stature (≤3%)	Yes	Yes	No	No	No	No	No	No	Yes (<1%)	No	Yes (<1%)	Yes, (<1%)
Overweight (BMI≥25)	No	No	No	No	No	No	No	No	No	No	Yes (39.9)	No (4%)
Broad high forehead	Yes	Yes	Yes	Yes	No	No	NR	No	Yes	Yes	Yes	No
Other features		Mild joint laxity, subluxed hips, disordered sleep, feeding difficulties (g-tube dependency), laryngomalacia, micrognathia, abnormal palmar creases	Noonan facies, pes planus, hypospadias	Noonan facies, pes planus, hypospadias		Clinodactyly, nystagmus, abnormality soft palate, hypoteloric, nystagmus	Cortical visual impairment	Cryptorchidism, astigmatism, myopia, strabismus. Tracheostomy and gastrostomy dependent: chronic respiratory failure, sialorrhea, dysphagia, esophageal reflux. Type I diabetes.	Pes cavus, nocturnal eneuresis		Osteoarthritis, glaucoma, cataracts, supraventricular tachycardia, cardiomegaly, micropenis	Cryptorchidism, hyperopia, astigmatism, right optic nerve coloboma. Postnatal NICU admission with feeding issues.

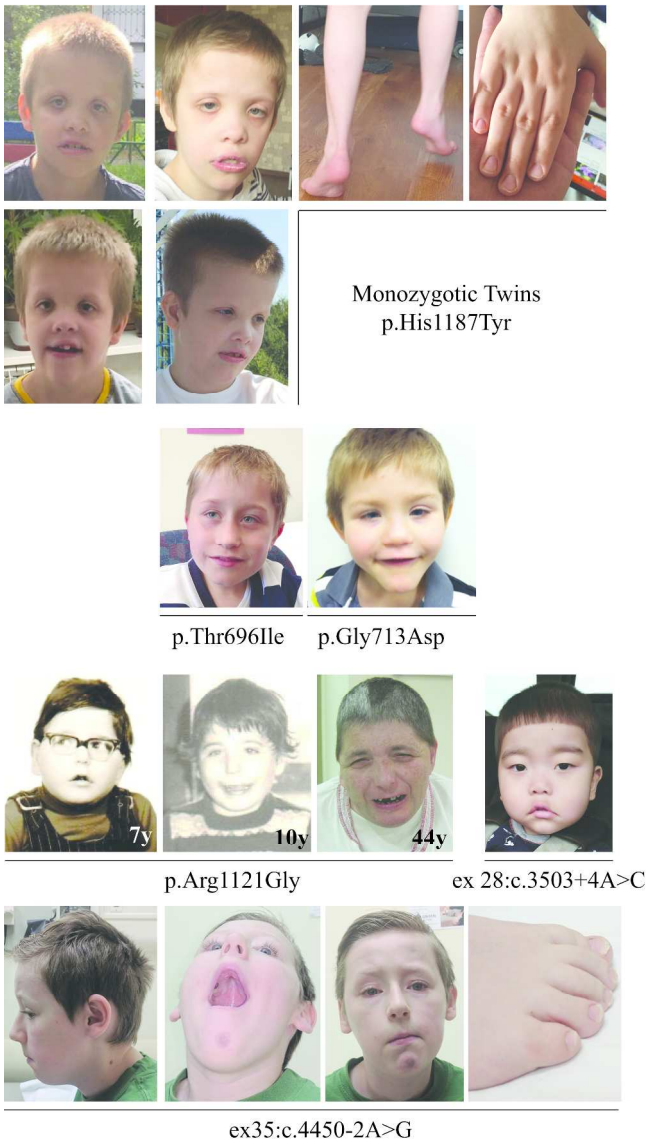
Abbreviations: %: centile; ASD: autism spectrum disorder; CT computerised tomography scan; g-tube: gastrostomy tube; mod+: at least moderate severity; MRI: magnetic resonance imaging; ND: not done; NA: not available; NR: not reported; NICU: neonatal intensive care unit; VOUS: variant of uncertain significance

Figure 1



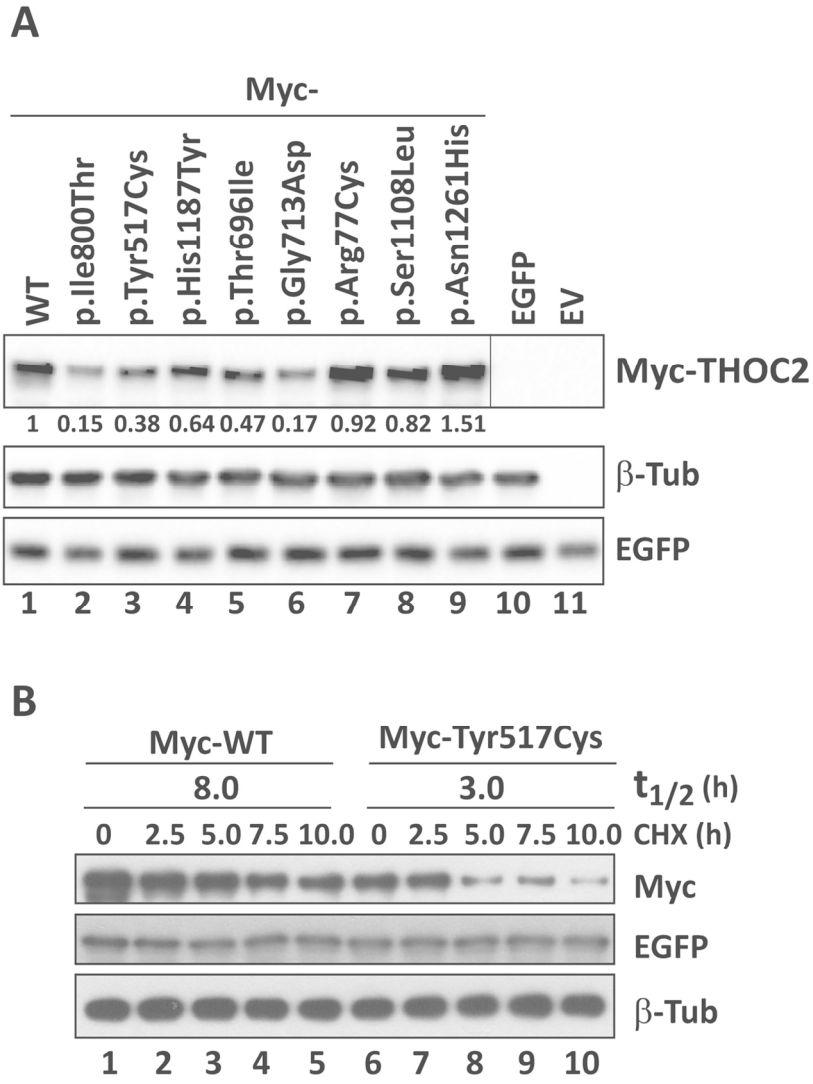
194x70mm (300 x 300 DPI)

Figure 2



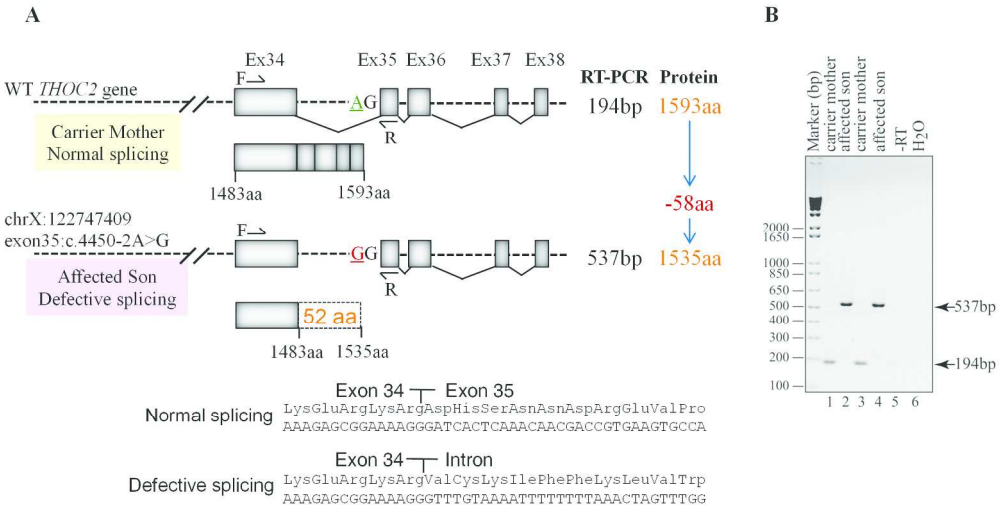
215x402mm (300 x 300 DPI)

Figure 3



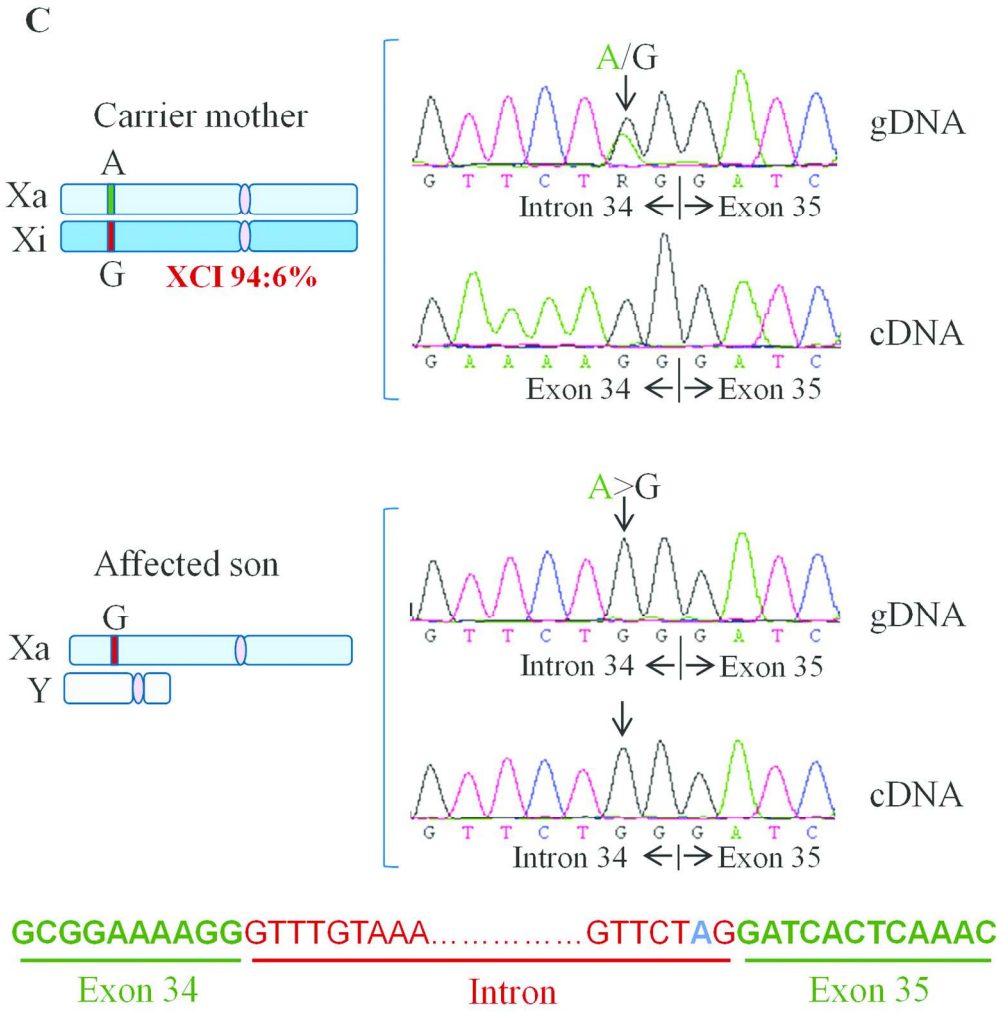
106x156mm (300 x 300 DPI)

Figure 4



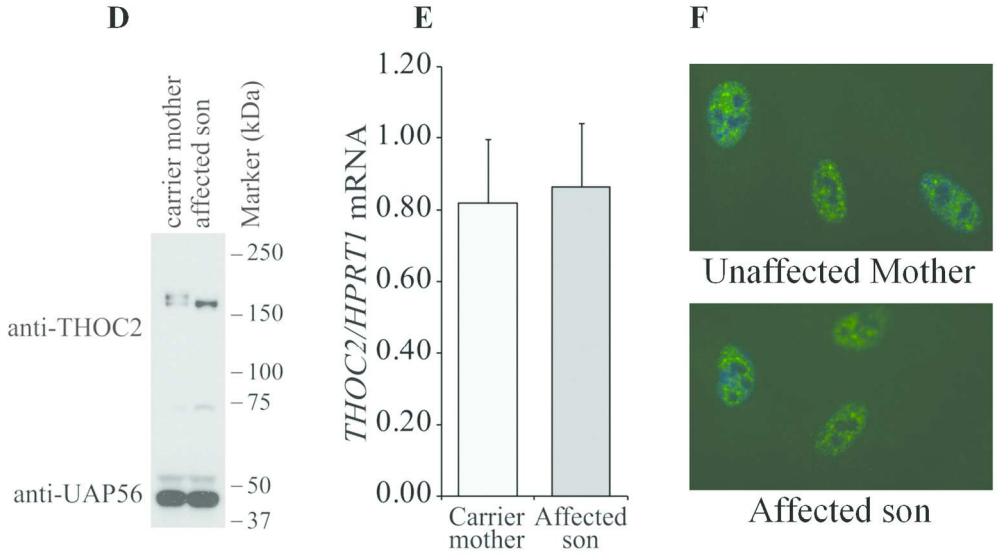
223x124mm (300 x 300 DPI)

Figure 4



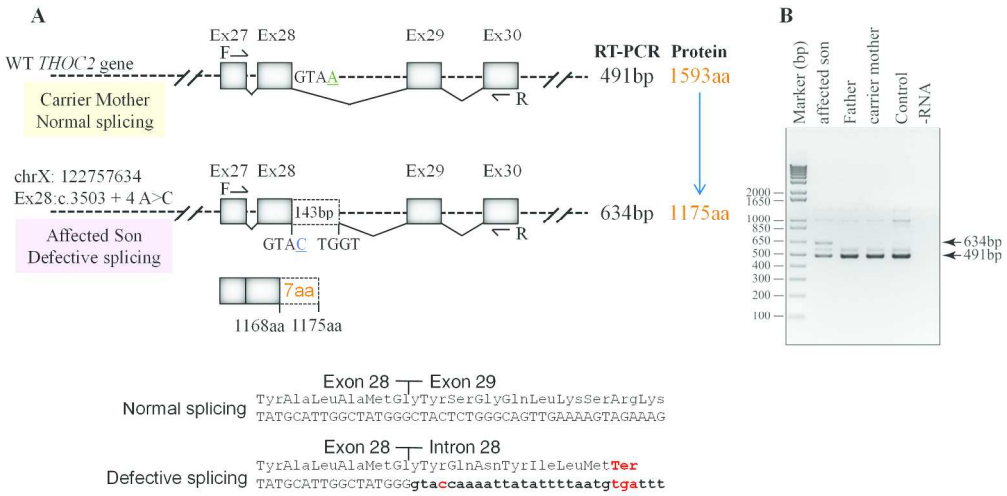
134x148mm (300 x 300 DPI)

Figure 4



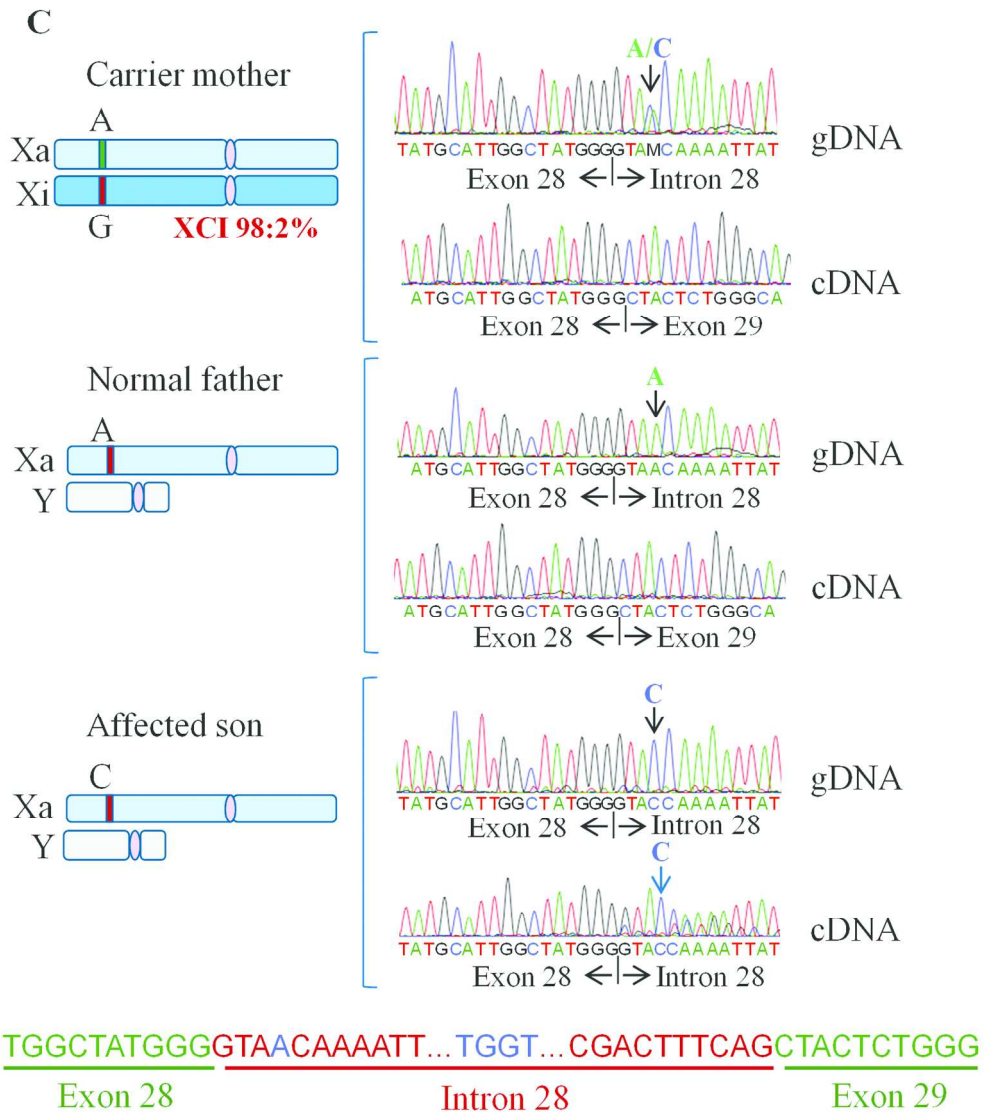
156x99mm (300 x 300 DPI)

Figure 5



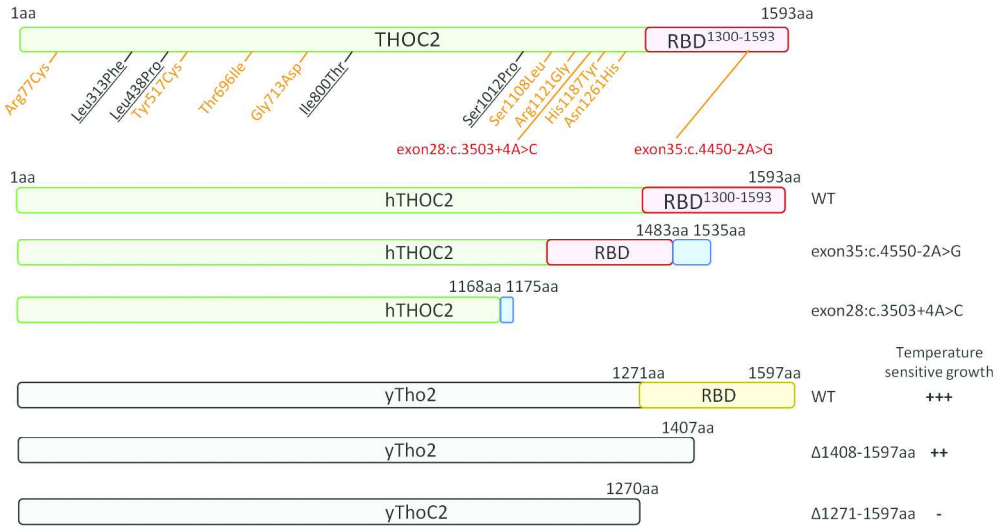
231x127mm (300 x 300 DPI)

Figure 5



139x169mm (300 x 300 DPI)

Figure 6



236x140mm (300 x 300 DPI)

SUPPLEMENTARY DATA FILE

Severe neurocognitive and growth disorders due to variation in *THOC2*, an essential component of nuclear mRNA export machinery

Raman Kumar[1], Alison Gardner[1], Claire Homan[1], Evelyn Douglas[2], Heather Mefford[3], Dagmar Wieczorek[4], Hermann-Josef Lüdecke[4], Zornitza Stark[5], Simon Sadedin[6], The Broad CMG[7], Catherine Bearce Nowak[8], Jessica Douglas[8], Gretchen Parsons[9], Paul Mark[9], Lourdes Loidi[10], Gail Herman[11], Theresa Mihalic Mosher[11], Meredith K Gillespie[12], Lauren Brady[13], Mark Tarnopolsky[13], Irene Madrigal[14], Jesús Eiris[15], Laura Domenech Salgado[16], Raquel Rabionet[16], Tim M Strom[17], Naoko Ishihara[18], Hidehito Inagaki[19], Hiroki Kurahashi[19], Tracy Dudding-Byth[20,21], Elizabeth Palmer[20,22], Michael Field[20], Jozef Gecz[1,23*]

[1]Adelaide Medical School and the Robinson Research Institute, The University of Adelaide, Adelaide, SA5000, Australia; [2]Genetics and Molecular Pathology, SA Pathology, Adelaide, SA5000, Australia; [3]Division of Genetic Medicine, Department of Pediatrics, University of Washington & Seattle Children's Hospital, Seattle, Washington, USA; [4]Heinrich-Heine-University, Medical Faculty, Institute of Human Genetics, Düsseldorf, Germany and Institut für Humangenetik, Universitätsklinikum Essen, Universität Duisburg-Essen, Essen, Germany; [5]Murdoch Children's Research Institute and Department of Pediatrics, University of Melbourne, Melbourne, Australia; [6]Murdoch Children's Research Institute and Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA; [7]Broad's Center for Mendelian Genomics, Cambridge, Massachusetts, USA; [8]Boston Children's Hospital, Longwood Avenue, Boston, MA, USA; [9]Spectrum Health Medical Genetics, Grand Rapids, MI, USA; [10]Fundación Pública Galega de Medicina Xenómica, Hospital Clínico Universitario de Santiago de, Compostela. Santiago de Compostela, Spain; [11]Nationwide Children's Hospital and The Ohio State University, Columbus, OH, USA; [12]Children's Hospital of Eastern Ontario Research Institute, University of Ottawa, Ottawa, Canada; [13]Department of Pediatrics, McMaster University Medical Centre, Hamilton, Canada; [14]Fundación Pública Galega de Medicina Xenómica, Hospital Clínico Universitario de Santiago de Compostela, Santiago de Compostela, Spain; [15]Unidad de Neurología Pediátrica, Departamento de Pediatría, Hospital Clínico Universitario de Santiago de Compostela, Santiago de Compostela, Spain; [16]Centre for Genomic Regulation (CRG), Universitat Pompeu Fabra and CIBERESP, Barcelona Institute for Science and Technology, Barcelona, Spain; [17]Institut für Humangenetik, Universitätsklinikum Essen, Universität Duisburg-Essen, Essen, Germany; [18]Department of Pediatrics, Fujita Health University School of Medicine, Japan; [19]Division of Molecular Genetics, Institute for Comprehensive Medical Science, Fujita Health University, Japan; [20]Genetics of Learning Disability Service, Hunter Genetics, Waratah, NSW 2298, Australia; [21]University of Newcastle, Australia Grow-Up-Well Priority Research Center; [22]School of Women's and Children's Health, University of New South Wales, Randwick, NSW 2031, Australia; [23]Healthy Mothers, Babies and Children, South Australian Health and Medical Research Institute, Adelaide, SA 5000, Australia.

*Correspondence: Jozef Gecz, [1]School of Medicine, The Robinson Research Institute, The University of Adelaide, North Adelaide, SA5005, Australia. E-mail: jozef.gecz@adelaide.edu.au

Detailed clinical data

High confidence in pathogenicity of *THOC2* variants, ACMG classification likely pathogenic or pathogenic.

Individual 1: affected male with *de novo* p.Thr696Ile variant

This is a 12 year old male, the second child of healthy unrelated Caucasian parent. He was delivered at 36 weeks gestation with a birth weight of 2000g (10th centile), length 42cm (<1st centile) and head circumference of 32cm (10-25th centile). He had speech and global developmental delay in infancy, but was not noted to have low tone. Aged 12 he has a severe intellectual disability. He has limited communication with some signs and a few single words. He was able to walk at 4 years of age. He has no history of epileptic seizures.

On examination he has subtle dysmorphic features including a broad high forehead, microcephaly (HC 50cm, 2nd centile), and short stature with a height of 111cm (<<1st centile) recorded at 10 years.

Brain MRI was reported as normal. Previous non diagnostic investigations include SNP chromosomal microarray and fragile X syndrome testing.

Individual 2: affected male with maternally inherited p.Gly713Asp variant

This is a 5 year old male, the first child of healthy parents. He has a maternal half-brother with a diagnosis of attention deficit hyperactivity disorder (ADHD) and a paternal half sister with diagnoses of bipolar disorder and ADHD. Prenatal history was complicated by maternal hyperemesis throughout pregnancy. He was delivered at near term (37+6 weeks), breech by emergency caesarian-section due to prolonged rupture of membranes with a birth weight of 2650gm (10th centile), length 47cm (3-10th percentile) and head circumference 34cm (25-50th percentile).

In the newborn period he was noted to have respiratory distress with biphasic stridor, mild laryngomalacia and micrognathia, as well as feeding difficulties. He has disordered sleep and failure to thrive, requiring gastrostomy tube feeding. He also has lax joints with a history of subluxing hips. He required myringotomy tubes and adenoidectomy. He has difficulty with sleep and requires clonidine to fall asleep and stay asleep.

Additional features include the presence of tremor and gait/balance problems. He has had 3 seizure-like episodes. He had hypotonia, speech and global developmental delay and has likely a moderate intellectual disability per his developmental assessments from neurology, and aged 5 years remains non-verbal. He has some vocalizations and is adept with a communication device. There are no features of autism. There are behavioral problems including occasional aggression thought to be related to frustration due to limited communication abilities. He has a history of hand flapping behaviors that have now resolved.

On examination he is small in all growth parameters with head circumference most significantly affected: he is microcephalic with a head circumference of -3.79 SD. A review of his past head circumference measurements show that he has consistently -3 SDs or more since birth to the present. He has a broad, high forehead and flatness of the occiput. He has prominent ears with rolled helices. He has a flat facial profile. Palpebral fissures are straight and short. He has mild median epicanthal folds. There is a broad nasal root with an upturned tip and a smooth philtrum with a thin upper vermillion border and prominent chin. He has small teeth. He has bilateral upper transverse palmar creases with normal lower palmar creases and an accessory flexion crease on each 5th finger with mild clinodactyly bilaterally. There is a flat hemangioma at the nape of the neck.

On Brain MRI he is noted to have a foreshortened and thin corpus callosum and low brainstem volume. No basal ganglia fusion. There is dysgyria with variability in gyral morphology, changes in orientation and gyral frequency. The gyral pattern appears finely nodular/polymicrogyria-like in the inferior frontal lobes. There are also scattered areas of T2 and FLAIR signal.

Prior non diagnostic investigations include a normal karyotype, normal microarray and a negative Prader-Willi methylation study. EEG was normal.

Individuals 3 and 4: affected monozygotic twins with maternally inherited p.His1187Tyr variants

These are two 7 1/2 year old monozygotic twin brothers, born to healthy unrelated Russian parents after assisted reproduction for unexplained subfertility. The twins were delivered at 37 weeks gestation. Twin one had a birth weight of 1900g (<3rd centile) and length of 46 cm (3rd centile) and twin two a birth weight of 2420g (10th centile) with length of 48 cm (25th centile)

Delays in their motor development were noted in the first year of life: they were only able to sit without support at 12 months of age. Twin 1 walked independently at 1 year 9 months, and twin 1 at 2 years of age. They have a tendency to toe walk. Expressive and receptive language has been significantly delayed and at 7 ½ years of age they have no verbal communication and only limited receptive language. They are described as hyperactive and can have some behavioral difficulties including auto-aggressive behaviors. They have difficulties coordinating the biting and chewing of their food. Both boys had hypospadias.

On examination twin one (aged 5 yr) was 106cm tall (25-50th centile) with a weight of 16.3kg (25-50th centile) and a head circumference of 40.5cm (10-25th centile). Twin two (aged 5 yr) was 107cm tall (25-50th centile) with a head circumference of 51cm (10-25th centile). Both boys were noted to have a flat midface, hypertelorism a short nose, short ears and prominent upper lip. A differential diagnosis of Noonan syndrome had been considered.

Prior non diagnostic investigations included a normal karyotype, normal chromosomal microarray (twin two), normal fragile X PCR. Testing of their mother indicated that she was almost completely skewed for X inactivation (0.013:0.987).

Individual 5: Affected male with maternally inherited c.3503+4A>C variant

This is a three year old male. He was delivered at 37 weeks with a birth weight of 3.018kg (25th centile). He was noted to have infantile hypotonia. Aged 3 he has been diagnosed with a severe intellectual disability. He is non-verbal and has a diagnosis of autism spectrum disorder with some challenging behaviors including a tendency to self-injure. His non-ambulant, but can crawl. He has no history of seizures.

On examination he is noted to have a flat nasal bridge and a down-turned mouth. He has no tremor. He is noted to be hyperkinetic. Aged 3 years 10 months his height was 94.5cm (-1.2SD), weight 13.8kg (-1.0SD) and head circumference was 49cm (-0.8SD)

MRI brain at age 2 years 11 months demonstrates mild dilatation of the lateral ventricles, especially in frontal lobes, mildly delayed myelination and an abnormal white matter lesion in the periventricular area close to the anterior horn.

Previous normal genetic investigations include a 46, XY karyotype. Copy number analysis was performed on the data from exome studies and was non-diagnostic.

Individual 6: Affected male with maternally inherited c.4450-2A>G variant.

This is a 10 year old male born to non-consanguineous parents with a negative family history at 41 weeks gestation after a pregnancy complicated by first trimester bleeding and maternal hypertension. He required oxygen at birth secondary to a small unilateral pneumothorax which resolved spontaneously. Birth weight was 9lb 10 oz (>90th centile).

He had infantile hypotonia and was diagnosed with a severe learning disability. He is in a special class at school and requires an Individualized Education Plan (IEP). He is able to communicate verbally though his speech is affected by an abnormality of the soft palate and nasal dysarthria. Additional neurological features including rest and intention tremor, an ataxic broad based gait, spasticity, and nystagmus. There are no concerns with his behavior.

On examination his head circumference (aged 5 years) was 49 cm (2nd percentile) with a height (aged 10yr) of 142.1 cm (50-85th centile) and BMI of 17.6kg (50-80th percentile). He is measurably hypoteloric with an intercanthal distance of 2.5 cm (-1SD) and an inter-pupillary distance of 4.5 cm (35th percentile). He is noted to have a startle response. He has a thin eyebrows and a slightly flattened nasal bridge. He has large ears (>2 SD). His palate is very high and narrow. He has slight clinodactyly of the fingers. He has flat toenails that are slightly depressed at the cuticle. Brain MRI shows slight prominence of the cerebro-fluid filled spaces.

Previous genetic investigations including Fragile-X, microarray, and mitochondrial DNA sequencing were all negative, though he was found to have borderline complex 1+3 and 4 deficiencies. His mother who carries the c.4450-2A>G splice variant has skewed X inactivation (94:6).

Individual 7: affected female with *de novo* p.Tyr517Cys variant.
This is a 10 year old female, the first child of healthy parents, she has an unaffected younger sister. She has an infantile onset epileptic encephalopathy. Epileptic seizures were diagnosed at 2 ½ months of age: with first seizures consisting of eyelid twitching, followed by clonic activity of the left arm and leg. This was treated by phenobarbitone with reasonable seizure control for one month. Subsequently clusters of epileptic spasms were diagnosed and EEG demonstrated hypsarrhythmia with independent left and right epileptiform activity. Trialled medication included ACTH, vigabatrin and clonazepam. She still has breakthrough seizures.

She was noted to have generalised hypotonia, poor head control and inability to visually fix and follow. She has a severe intellectual disability and attends special education schooling with speech therapy, physiotherapy and occupational therapy. She is non-verbal and non –ambulatory. She uses a gait trainer. She has cortical visual impairment.

At one year of age on examination she was noted to have a head circumference on the 5th centile, with height on the 90th centile and weight on the 25th centile.

Prior non-diagnostic testing included a karyotype, subtelomeric FISH, FISH at the Angelman locus, sequencing of MECP2 and UBE3A and metabolic screening. Brain CT and MRI were reported as normal.

Rare *THOC2* variants without supportive functional data and conflicting *in silico* predictors: variants of uncertain clinical significance on ACMG classification

Individual 8: affected male with maternally inherited p.Arg77Cys variant
This 9 year old male is one of two affected males to a non-consanguineous Caucasian couple. He had a younger brother who died at the age of 4 and had a similar clinical presentation with intellectual disability, seizure disorder and tracheostomy dependency. His mother also has two unaffected daughters by different partners. The mother has a history of early developmental delay but was able to complete high school. She also has bipolar disorder and a speech disorder (stutter). There is an extended maternal family history which could be consistent with X linked inheritance. The proband has a maternal uncle who died before the age of 4 years, who had microcephaly, a severe progressive neurological condition, never being able to walk or talk, and required gastrostomy and tracheostomy. The proband's mother's maternal aunt also had a son who died between the ages of 2 and 4 with, by report, a similar presentation. The mother's sister has learning difficulties as school, bipolar disorder and a speech disorder (stutter). The maternal grandmother completed high school and some tertiary level eductation.

The proband was born at 41 weeks gestation by spontaneous vaginal delivery, with Apgar scores of 8 at 1 and 9 at 5 minutes, and minimal resuscitation (transient oxygen) required at birth. Routine antenatal ultrasounds were reported as normal, his mother was treated with lamotrigine through the pregnancy but there were no other teratogenic exposures. Birth weight was 55.2cm, weight 3.7kg and head circumference 33cm. He was noted to have increased muscle tone and poor suck, and was transferred to the special care nursery requiring nasogastic feeds. A head ultrasound scan was

suggestive of Dandy Walker malformation and he was transferred to a tertiary hospital for ongoing care.

The proband has a profound intellectual disability, hypotonic quadriplegic cerebral palsy (GMFCS level 5) and intractable infantile onset seizure disorder. He has a history of central apnea and chronic respiratory failure and has been dependent on a tracheostomy and ventilator support since the age of 3 years. He has severe dysphagia and esophageal reflux and has been dependent on gastrostomy feeding since the first year of life, initially being able to take some thickened oral feeds. He has sialorrhea and previously had parotid duct ligation and salivary gland excision. He is profoundly delayed and does not make eye contact and smiles occasionally in a non-purposeful manner. He has limited vision, strabismus, myopia, astigmatism. He is non-verbal. He is centrally hypotonic with random non-coordinated movements of the limbs and face, and has developed tongue fasciculations. He previously was able to sit and roll, but has now lost those skills. With age he has increasing peripheral spasticity and bilateral ankle clonus. He has no tremor.

He has an intractable seizure disorder, which first presented as infantile spasms. His current antiepileptic medication is levetiracetam (Keppra) and diazepam (Valium). He also has non epileptic myoclonus and muscular spasms which are treated with a muscular relaxant. EEG recordings have been abnormal including moderate generalized background slowing and occasional bi-central spike and waves.

He has abnormalities on neuroimaging including hypoplasia of the cerebellar vermis, cerebral aqueduct stenosis, stable enlargement of the ventricular system reflecting diffuse white matter loss and thin corpus callosum, as well as hypoplasia of the brainstem, pons and medulla.

Other medical issues include type I diabetes and left cryptorchidism (requiring orchidoplexy). He has a structurally normal heart and kidneys.

On recent examination aged 9 he has a weight of 29.3kg (50^{th} centile), height of 1.132m (25^{th} centile), and is noted to be normocephalic. He has bitemporal narrowing, large ears, a flat nasal bridge, a tented mouth, bilateral ptosis and right sided strabismus. He has tapered fingers with puffy feet and toes.

Prior non diagnostic investigations include a chromosomal microarray, creatine kinase of 408 u/L at birth and 123 u/L aged 2 years, no mutation identified on sanger sequencing of the *PLP1* gene.

Individuals 9 and 10: two brothers with maternally inherited p. Ser1108Leu variants.
These are two brothers born to healthy unrelated Spanish parents, with no additional family history of intellectual disability.

The first brother (individual 9) is currently 17 years of age and has diagnoses on the autistic spectrum and severe intellectual disability. He was born after an uncomplicated pregnancy at 40 weeks gestation. His birth weight was 2880g (10^{th} centile). Apgar scores were 10 at 1 and 5 minutes. He started to walk at 16 months without prior crawling. He had a tendency to toe walk. He is described as 'clumsy'. Development of his pincer grasp was delayed, and he was delayed in being able to chew, and eat and drink independently. His ability to point to objects and his receptive and expressive language development was delayed. He is able to maintain a simple conversation, solve simple puzzles, and identify different parts of the body. He can recognize a few logos. He can get easily frustrated and has some fears. He has no stereotypical movements. He can initiate simple social interactions but does not have symbolic play. He gets anxious prior to novel situations. Eye contact is limited.

He has no tremor or seizures. His sleep pattern is normal. He had nocturnal enuresis until the age of 12.

There have been significant concerns with short stature, and he was referred to an endocrinologist.

On examination he was noted to have a broad forehead and a narrow nasal root with a bulbous tip to his nose, thick lips and a short philtrum. His ears were normally sited. He had a clumsy gait and shortening of the Achilles tendon. His head circumference was 55 cm (50th centile), height 153 cm (<<1st centile) and weight 56kg (3-10th centile).

The second brother (individual 10) is 12 years of age and born at 37 weeks after an uncomplicated pregnancy. He also has diagnoses of autism and severe intellectual disability, and attends a special school. His birth weight was 3480g (50th centile) and Apgar scores were 9 at 1 and 10 and 5 minutes.

There were no significant concerns with his motor development in the first year of life, but he was slow to walk, with this beginning at 18 months. He had a tendency to toe walk. He had significantly delayed expressive and receptive speech delay, and no social pointing until 3 years of age. He has no expressive language, and very delayed receptive language. His social development has been delayed and he tended to use adults to obtain objects that he wanted. His eye contact has been limited. He can engage in simple social interactions with adults but cannot sustain peer relationships. He is described as being restless and has a tendency to 'explosive' reactions, and can have challenging behaviors including mild auto and hetero-aggressive acts when frustrated. He has difficulties adapting to changes in his routine and finds it very difficult to sit still. He has stereotypical movements including raising up his arms, and clutching and shaking his hands when excited. He is described as clumsy. He requires assistance in his activities of daily living.

On examination he is noted to have a broad forehead and widened nasal tip. He has a tendency to toe walk, and has shortening of the Achilles tendon. At the age of 12 years 9 months he was 146cm tall (10th centile), and weighed 36kg (10th centile).

Prior non-diagnostic investigations include a normal chromosomal microarray (qChip Post)

Individual 11: male with maternally inherited p.Arg1121Gly variant

This male individual died aged 44, and was of Spanish heritage. He was not born prematurely and he did not have a low birth weight. He is reported to have been affected by severe intellectual disability, with a history of severe speech delay and infantile hypotonia. He received a diagnosis of cerebral palsy. He was reported to be non-ambulant. He had aggressive behavior with self-harm including hand-biting and face hitting. He lived in an institution for individuals with intellectual disability. He had severe speech and language disorder, with around 10 words. Other health issues included asthma, osteoarthritis, glaucoma, cataracts, supraventricular tachycardia and cardiomegaly.

He was reported to be microcephalic with short stature, height of 134cm (<<1ST centile), and weight (aged 44 years) was 69.8kg (50th centile). He had a raised BMI (39.9) and subtle dysmorphic features including a broad high forehead. He was reported to have micropenis.

Individual 12: male with maternally inherited p.Asn1261His variant

This male individual is the first child to a 26 year old female. He is currently 5 years 9 month of age. The pregnancy was uncomplicated with appropriate prenatal care. There were no known exposures, maternal illnesses or medications used during the pregnancy. Routine prenatal testing was normal, and there was no gestational diabetes. No prenatal ultrasound or genetic testing was performed. Delivery was performed at 42 weeks at home with a midwife and was spontaneous vaginal with vacuum extraction due to large size. Birth weight was 5.2 kg, length 57cm, head circumference 38.5cm. According to the parents, the baby was initially "blue but his color came right back." Formal Apgar scores were not recorded. He had some irregular breathing at 4 hours of age and was admitted to a regional neonatal intensive care unit where he remained for 10 days, mostly for feeding difficulties.

The parents are both healthy and unrelated. They have a daughter, who does not carry the *THOC2* variant, who was admitted to hospital at 8 months of age with a cluster of unresponsive tonic episodes. MRA/MVA showed cortical venous thrombi with superior sagittal sinus thrombosis. The tonic episodes were treated with Keppra, but she was weaned off this after a normal EEG. She has made a good recovery and had normal development.

The THOC2 variant is maternally inherited and not present in the maternal grandmother. The maternal grandfather is deceased, and no DNA was available for testing. There is no extended family history of intellectual disability. The mother has 6 healthy adult full-brothers with normal development.

The patient had delayed early milestones, sitting at 11 months and walking at 2 years 11 months. His first words were at 12 months of age but he did not gain additional words until he was over 2 years if age. He developed overt seizures between 18-24 months, although subtle and or absence seizures may have been present earlier. The seizures were treated with multiple medications but progressed to refractory seizures consistent with Lennox-Gastaut syndrome. Initial EEG was consistent with a diffuse encephalopathy, follow-up EEG was indicative of numerous brief atonic seizures in the setting of a slow background. Seizure frequency and severity have partially responded to oral cannabis which was begun as part of a clinical trial at age 5 years 2 months. Prominent seizure type remains as infrequent atonic seizures. His development has progressed with improved seizure control. He knows more than 100 words and has several two to three word phrases. He can eat with a spoon, help dress himself and ride a tricycle. He attends a regular preschool class and receives occupational, physio and speech therapies. He has a short attention span.

He has central hypotonia with appendicular spasticity and has received a diagnosis of spastic quadriplegia. He is noted to have visual difficulties with hyperopia with astigmatism and a possible right optic nerve coloboma. He has had bilateral orchidoplexy for cryptorchidism.

Aged 6 years 10 months, his weight was 17.9kg (3rd centile), height 108.4 cm (<1st centile) and head circumference at 5 years 10 months was 50 cm (29 centile). He has mild dysmorphic features, including somewhat long facies and prominent ears.

Previous non-diagnostic testing include chromosomal microarray, testing for Fragile X, urinary organic acids, serum amino acids, acylcarnitine profile and transferrin isoforms for carbohydrate deficient glycoprotein disorders. MRI scan performed at 2 was consistent with mild left sided periventricular leukomalacia or delayed myelination. Hearing screening has been normal. Clinical exome sequencing was performed on the proband and his parents at Baylor-Miraca, Houston, TX.

Methods used for identifying the *THOC2* variants

p.Thr696Ile (Individual 1)

Trio whole exome sequencing (WES) and data processing were performed by Genomics Platform at the Broad Institute of Harvard and MIT (Broad Institute, Cambridge, MA, USA). WES was performed using Illumina exome capture (38 Mb target) and the data was processed through a pipeline based on Picard. Single Nucleotide Polymorphism (SNPs) and insertions/deletions (indels) were jointly called across all samples using Genome Analysis Toolkit (GATK) HaplotypeCaller package version 3.4.

p.Gly713Asp (Individual 2)

THOC2 variant was identified using the published method (Tanaka, et al., 2015).

p.His1187Tyr (Individual 3-4)

Genomic DNA was extracted from peripheral blood samples of the affected individuals and their parents using DNA standard extraction kits. Exomes were enriched in solution and indexed with version 5 of the SureSelect XT Human All Exon 50Mb kit (Agilent Technologies). Sequencing was performed as 101 bp paired-end reads on HiSeq2000/2500 systems (Illumina). For one of the twins and his parents, 9.4, 8.7 and 12 Gigabases of sequence were generated, respectively, resulting in an average depth of coverage of 113, 110 and 150, with 96.5%, 96% and 97.5% of the target regions covered at least 20 times. Image analysis and base calling was performed using Illumina Real Time Analysis. Reads were aligned against the human assembly hg19 (GRCh37) using BWA v 0.5.9. Variant calling was performed specifically for the regions targeted by the exome enrichment kit using SAMtools (v 0.1.18), PINDEL (v 0.2.4t), ExomeDepth (v 1.0.0) and custom scripts. Variant quality was determined using the SAMtools varFilter script with default parameters except for the maximum read depth (-D) and the minimum P-value for base quality bias (-2), which were set to 9999 and 1e-400, respectively. In addition, a custom script was applied to mark all variants with adjacent bases of low median base quality. All variants were then annotated using custom Perl scripts. Annotation included known transcripts (UCSC Known Genes and RefSeq genes), known variants (dbSNP v 135), type of mutation and - if applicable - amino acid changes. The annotated variants were integrated into an in-house database. Because the index case has a likewise affected twin brother, and the mother displayed an almost completely skewed X-inactivation, the database was queried to show homozygosity, compound heterozygosity or 'hemizygosity' for X-linked variants. To reduce false positives, variants that were already present in the database, or had a variant quality of less than 30, or did not pass the filter scripts were filtered out. The raw read data of the remaining variants using the Integrative Genomics Viewer (IGV) was manually checked.

Exon35:c.4450-2A>G (Individual 5)

THOC2 variant was identified using the published method (Nishimura, et al., 2014).

Exon28:c.3503+4A>C (Individual 6)

XHMM software on exome sequencing data, was used to analyse for copy number variation. Fifty nanograms of whole blood DNAs from the proband and his parents were used to generate a whole exome library using SureSelect QXT Reagent and SureSelect Clinical Research Exome (Agilent technologies, Santa Clara, CA). Sequencing was carried out using HiSeq 1500 (Illumina, San Diego, CA). After demultiplexing from other sample data, the reads were mapped onto the human reference hg19 using BWA 0.7.15 (Li and Durbin, 2009). Sorting and recalibration of the mapped reads were done using Picard tools 2.7.1 (http://broadinstitute.github.io/picard), and variants were called into a VCF file using GATK 3.6 (Van der Auwera, et al., 2013). Annotations were added using SnpEff and SnpSift (Cingolani, et al., 2012). Low quality data and common SNPs of more than 1% frequency in Japanese or global populations were filtered out by using dbSNP 147, HGVD (Higasa, et al., 2016), iJGVD (Nagasaki, et al., 2015) and ExAC (Lek, et al., 2016) data.

p. Tyr517Cys (Individual 7)

THOC2 variant was identified using the published method (Epi, et al., 2013)

p.Arg77Cys (Individual 8)

THOC2 variant was identified using the published method (Farwell, et al., 2015).

p.Ser1108Leu and p.Arg1121Gly (Individual 9-10 and 11)

Both patients had prior investigation with the chromosomal microarray (qChip Post). Carriers of variants p.Ser1108Leu and p.Arg1121Gly were identified by WES. In short, DNA from

the proband and family members (parents for p.Arg1121Gly and affected sibling for Ser1108Leu) were fragmented with covaris, captured using the Agilent SureSelect 50Mb kit and TruSeq adapters, pooled in groups of three samples and sequenced on an Illumina HiSeq2000 lane with 2x75bp reads. Fastq files were processed using the in-house *ediva* pipeline: they were aligned to the hg19 reference genome, duplicate-marked, locally realigned and recalibrated, and variants were called using GATK HaplotypeCaller (Van der Auwera, et al., 2013) followed by quality assessment with VQSR and retaining tranche 99.9. Variants were annotated with ediva and filtered according to familial segregation (*de novo*, recessive and X-linked variants in the trio, shared recessive and X-linked variants in the sibpair), MAF < 0.05% in any population/database, and scored based on functionality predictors. The trio case p.Arg1121Gly was also a carrier of a *de novo* variant in *SALL3*, while the *THOC2* variant p.Ser1108Leu was the only likely candidate passing the filters in the sib-pair. Identified variants were further confirmed by Sanger sequencing in the carriers.

p.Asn1261His (Individual 12)

THOC2 variant was identified using the published method (Yang, et al., 2014).

References

Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin) 6(2):80-92.

Epi KC, Epilepsy Phenome/Genome P, Allen AS, Berkovic SF, Cossette P, Delanty N, Dlugos D, Eichler EE, Epstein MP, Glauser T and others. 2013. De novo mutations in epileptic encephalopathies. Nature 501(7466):217-21.

Farwell KD, Shahmirzadi L, El-Khechen D, Powis Z, Chao EC, Tippin Davis B, Baxter RM, Zeng W, Mroske C, Parra MC and others. 2015. Enhanced utility of family-centered diagnostic exome sequencing with inheritance model-based analysis: results from 500 unselected families with undiagnosed genetic conditions. Genet Med 17(7):578-86.

Higasa K, Miyake N, Yoshimura J, Okamura K, Niihori T, Saitsu H, Doi K, Shimizu M, Nakabayashi K, Aoki Y and others. 2016. Human genetic variation database, a reference database of genetic variations in the Japanese population. J Hum Genet 61(6):547-53.

Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, O'Donnell-Luria AH, Ware JS, Hill AJ, Cummings BB and others. 2016. Analysis of protein-coding genetic variation in 60,706 humans. Nature 536(7616):285-91.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25(14):1754-60.

Nagasaki M, Yasuda J, Katsuoka F, Nariai N, Kojima K, Kawai Y, Yamaguchi-Kabata Y, Yokozawa J, Danjoh I, Saito S and others. 2015. Rare variant discovery by deep whole-genome sequencing of 1,070 Japanese individuals. Nat Commun 6:8018.

Nishimura N, Gotoh T, Oike Y, Yano M. 2014. TMEM65 is a mitochondrial inner-membrane protein. PeerJ 2:e349.

Tanaka AJ, Cho MT, Millan F, Juusola J, Retterer K, Joshi C, Niyazov D, Garnica A, Gratz E, Deardorff M and others. 2015. Mutations in SPATA5 Are Associated with Microcephaly, Intellectual Disability, Seizures, and Hearing Loss. Am J Hum Genet 97(3):457-64.

Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J and others. 2013. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr Protoc Bioinformatics 43:11 10 1-33.

Yang Y, Muzny DM, Xia F, Niu Z, Person R, Ding Y, Ward P, Braxton A, Wang M, Buhay C and others. 2014. Molecular findings among patients referred for clinical whole-exome sequencing. JAMA 312(18):1870-9.

Supplementary Figure S1. THOC2 variant amino acid residues are highly conserved. Amino acid sequence alignments show the variant amino acid residue (red) at the conserved positions (green). The amino acid sequences were aligned using www.uniprot.org/align/.

p.Arg77Cys

Mutant	DISEFCEDMPS	82
Hom	DISEFREDMPS	82
Mus	DISEFREDMPS	82
Rattus	DISEFREDMPS	82
Pan	DISEFREDMPS	82
Macaca	DISEFREDMPS	82
Canis	DISEFREDMPS	82
Equus	DISEFREDMPS	82
Sus	DISEFREDMPS	82
Bos	DISEFREDMPS	82
Loxo	DVSEFREDMPS	82
Gallus	DIIEFREDMPS	79
Xenopus	EIIAFRDDMPS	79

p.Tyr517Cys

Mutant	FKTFPCQHRYR	522
Hom	FKTFPYQHRYR	522
Mus	FKTFPYQHRYR	522
Rattus	FKTFPYQHRYR	522
Pan	FKTFPYQHRYR	522
Macaca	FKTFPYQHRYR	522
Canis	FKTFPYQHRYR	522
Equus	FKTFPYQHRYR	522
Sus	FKTFPYQHRYR	522
Bos	FKTFPYQHRYR	522
Loxo	FKTFPYQHRYR	522
Gallus	FKTFPYQYRYR	519
Xenopus	FKCFPYQYRYR	518

p.Thr696Ile

Mutant	ITEEMIMEQLE	701
Hom	ITEEMTMEQLE	701
Mus	ITEEMTMEQLE	701
Rattus	ITEEMTMEQLE	701
Pan	ITEEMTMEQLE	701
Macaca	ITEEMTMEQLE	701
Canis	ITEEMTMEQLE	701
Equus	ITEEMTMEQLE	701
Sus	ITEEMTMEQLE	701
Bos	ITEEMTMEQLE	701
Loxo	ITEEMTMEQLE	701
Gallus	ITEEMTMEQLE	698
Xenopus	ITEEMTMEQLE	697

p.Gly713Asp

Mutant	QLKAEDGYFGQ	718
Hom	QLKAEGGYFGQ	718
Mus	QLKAEGGYFGQ	718
Rattus	QLKAEGGYFGQ	718
Pan	QLKAEGGYFGQ	718
Macaca	QLKAEGGYFGQ	718
Canis	QLKAEGGYFGQ	718
Equus	QLKAEGGYFGQ	718
Sus	QLKAEGGYFGQ	718
Bos	QLKAEGGYFGQ	718
Loxo	QLKAEGGYFGQ	718
Gallus	QLKAEGGYFGQ	715
Xenopus	QLKAEGGYFGQ	714

p.Ser1108Leu

Mutant	KLTKAPVHCLE	1113
Hom	KLTKASVHCLE	1113
Mus	KLTKASVHCLE	1113
Rattus	KLTKASVHCLE	1113
Pan	KLTKASVHCLE	1113
Macaca	KLTKASVHCLE	1113
Canis	KLTKASVHCLE	1113
Equus	KLTKASVHCLE	1113
Sus	KLTKASVHCLE	1113
Bos	KLTKASVHCLE	1113
Loxo	KLTKASVHCLE	1113
Gallus	KLTKASVHCLE	1110
Xenopus	KLTKASVHCLE	1109

p.Arg1121Gly

Mutant	EYTHIGNILIV	1126
Hom	EYTHIRNILIV	1126
Mus	EYTHIRNILIV	1126
Rattus	EYTHIRNILIV	1126
Pan	EYTHIRNILIV	1126
Macaca	EYTHIRNILIV	1126
Canis	EYTHIRNILIV	1126
Equus	EYTHIRNILIV	1126
Sus	EYTHIRNILIV	1126
Bos	EYTHIRNILIV	1126
Loxo	EYTHIRNILIV	1126
Gallus	EYTHIRNILIV	1123
Xenopus	EYTHIRNILIV	1122

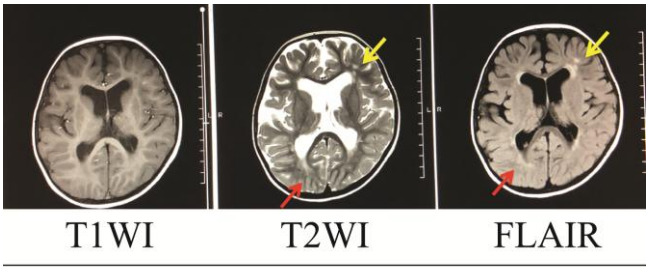
p.His1187Tyr

Mutant	PENEFYHKDPP	1192
Hom	PENEFHHKDPP	1192
Mus	PENEFHHKDPP	1192
Rattus	PENEFHHKDPP	1192
Pan	PENEFHHKDPP	1192
Macaca	PENEFHHKDPP	1192
Canis	PENEFHHKDPP	1192
Equus	PENEFHHKDPP	1192
Sus	PENEFHHKDPP	1192
Bos	PENEFHHKDPP	1192
Loxo	PENEFHHKDPP	1192
Gallus	PENEFHHKDPP	1189
Xenopus	PENDFHHKDPP	1188

p.Asn1261His

Mutant	GNSGSHSNKAV	1266
Hom	GNSGSNSNKAV	1266
Mus	GNSGSNSNKAV	1266
Rattus	GNSGSNSNKAV	1266
Pan	GNSGSNSNKAV	1266
Macaca	GNSGSNSNKAV	1266
Canis	GNSSSNSSKAV	1266
Equus	GNSGSNSSKAV	1266
Sus	GNSGSNSSKTV	1266
Bos	GNSGSNSSKTV	1266
Loxo	GNSGSTSNKTV	1266
Gallus	GNSASNSKIIK	1265
Xenopus	GNSASSSKVLK	1264

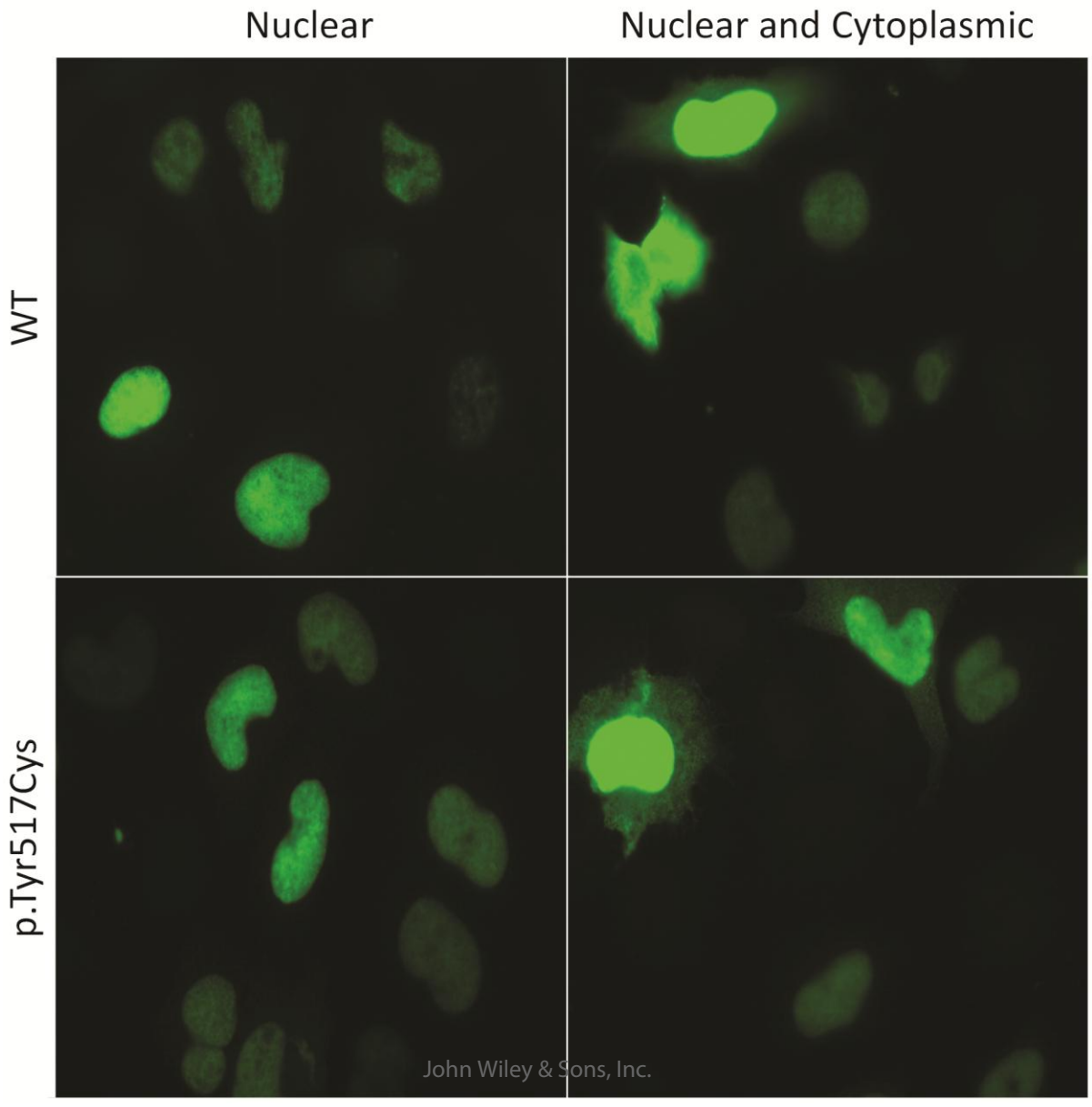
Supplementary Figure S2 Brain MRI of the individual carrying the exon28: c.3503+4A>C *THOC2* variant. Red arrow shows myelination and yellow arrow shows periventricular white matter lesion at age 2y 11m.

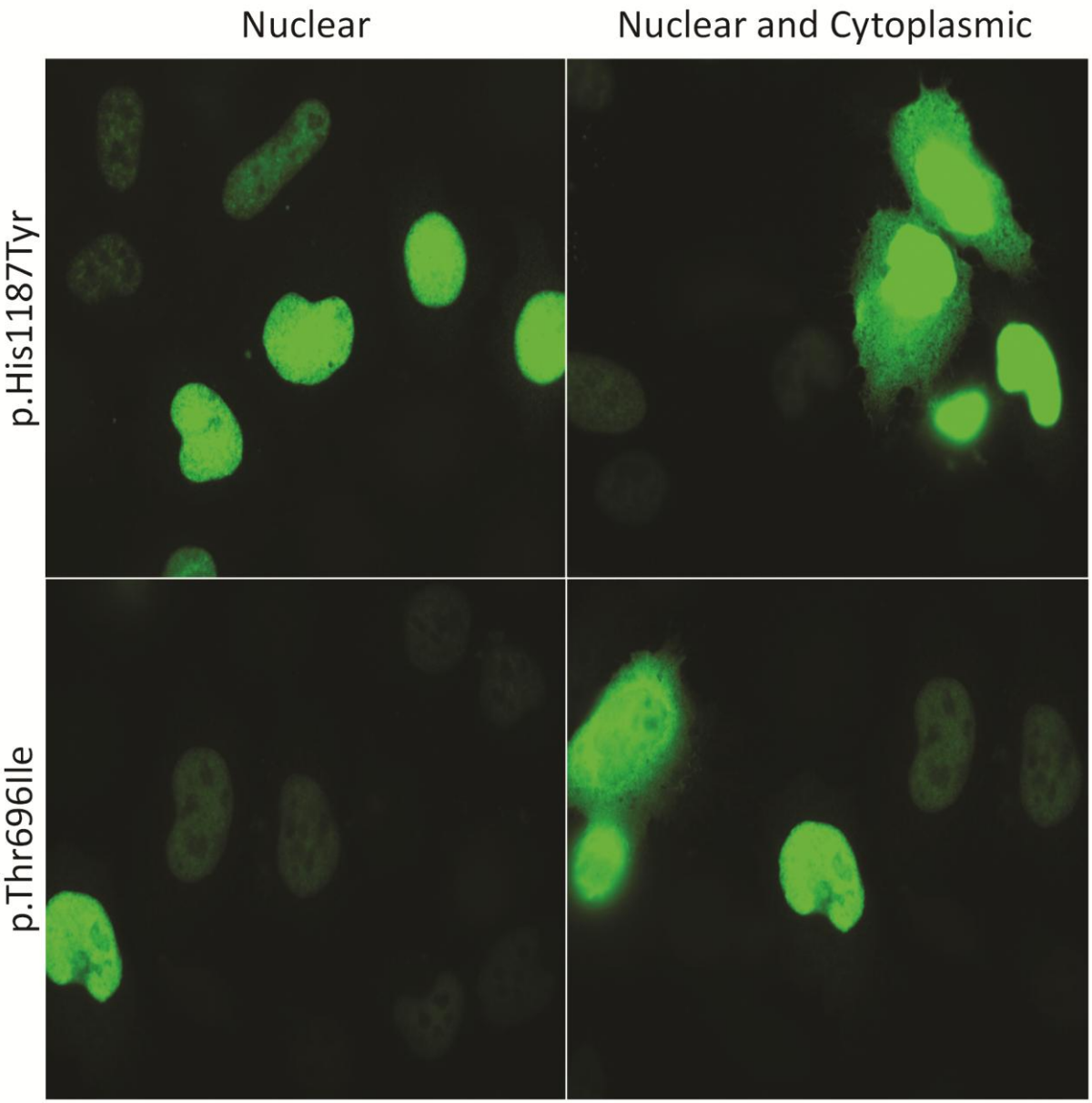


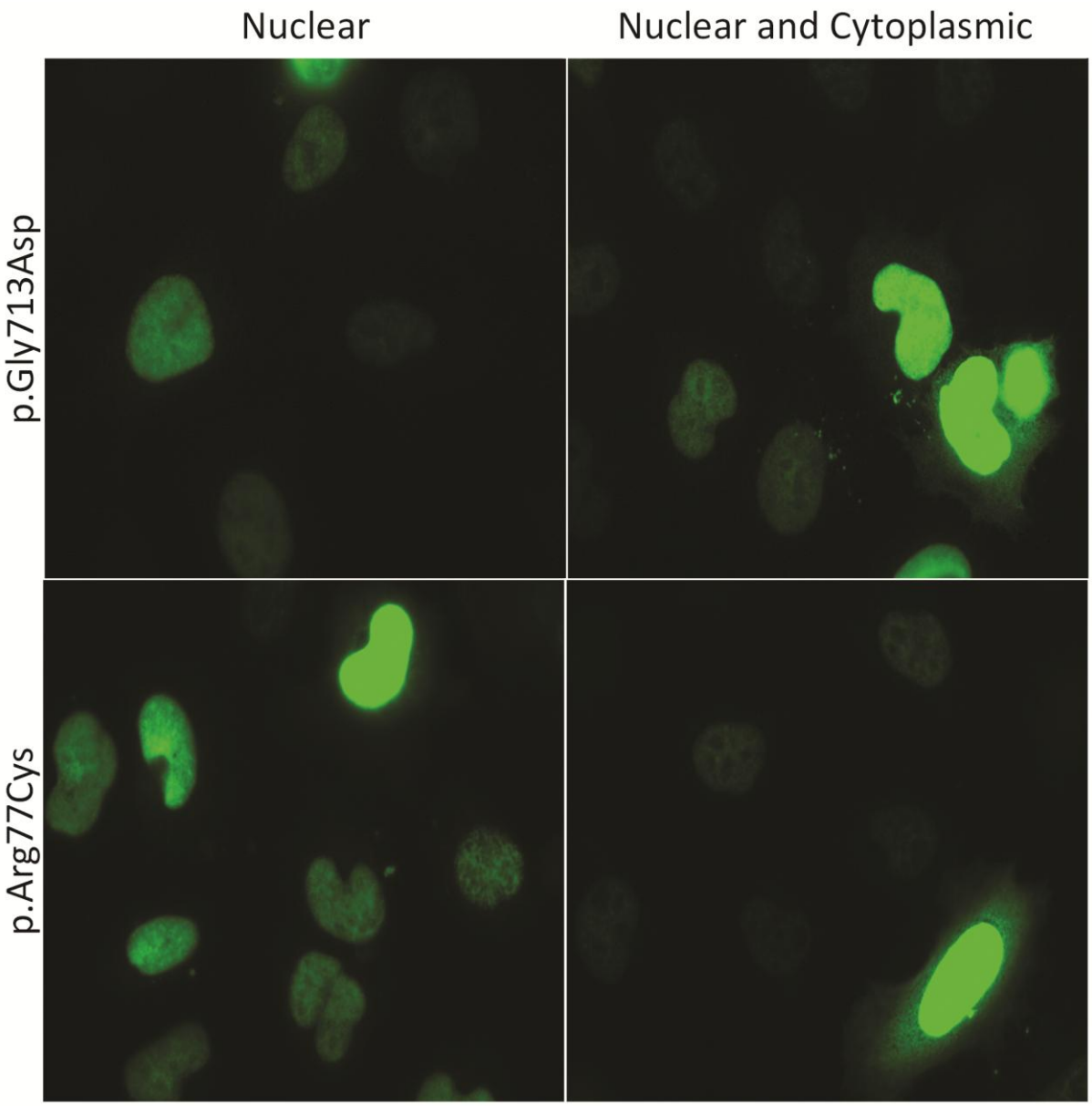
ex 28:c.3503+4A>C

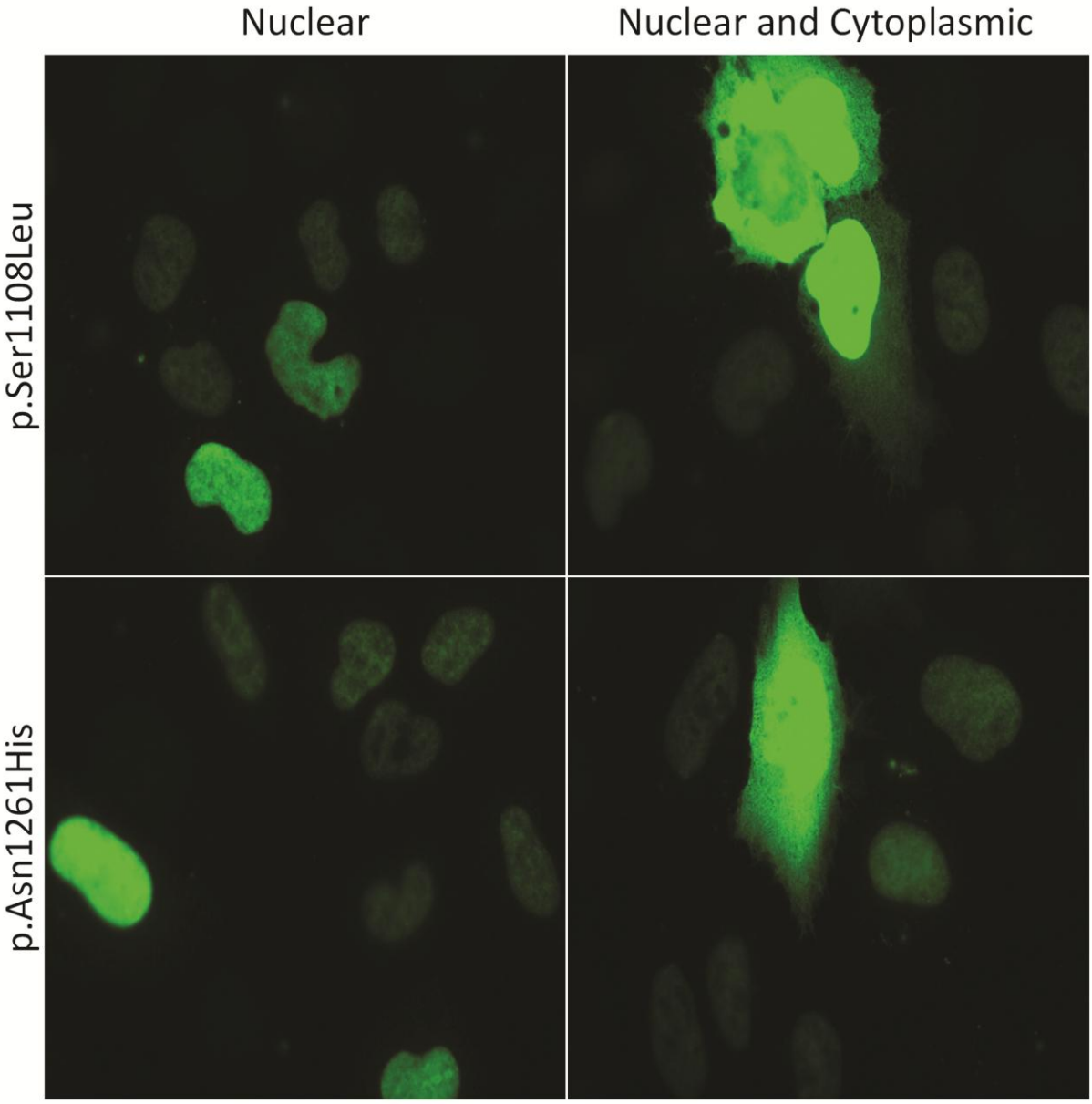
Supplementary Figure S3 A: THOC2 variant proteins are normally localised in HeLa and HEK293T cells. HeLa and HEK293T cells cultured on Poly-L-lysine coated coverslips (BD Biosciences) were transfected with wild type or variant Myc-THOC2 expression plasmids and detected by immunofluorescence using mouse anti-Myc (9E10; Sigma) primary and Donkey anti-rabbit-IgG Alexa 488 secondary antibody (Life Technologies). Only HeLa cell localisations at 630× magnification are shown. **B:** Nuclear and cytoplasmic (HEK293T; average 250 cells each) and **C:** nuclear and nuclear plus cytoplasmic (HeLa; average 400 cells each) staining was visually scored and plotted. Note that nuclear plus cytoplasmic instead of only cytoplasmic staining for HeLa was scored because all the cells with cytoplasmic staining also had nuclear staining.

A

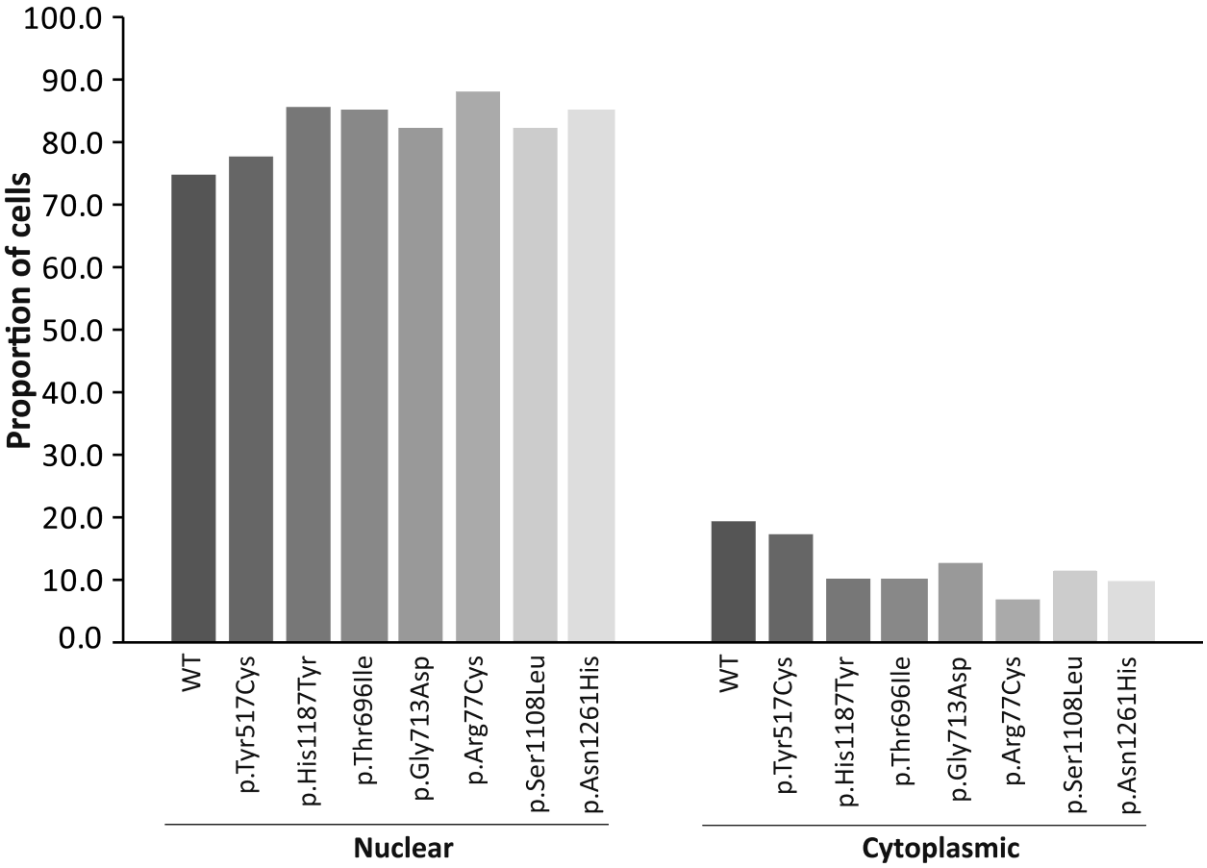



Nuclear
Nuclear and Cytoplasmic
p.His1187Tyr
p.Thr696Ile

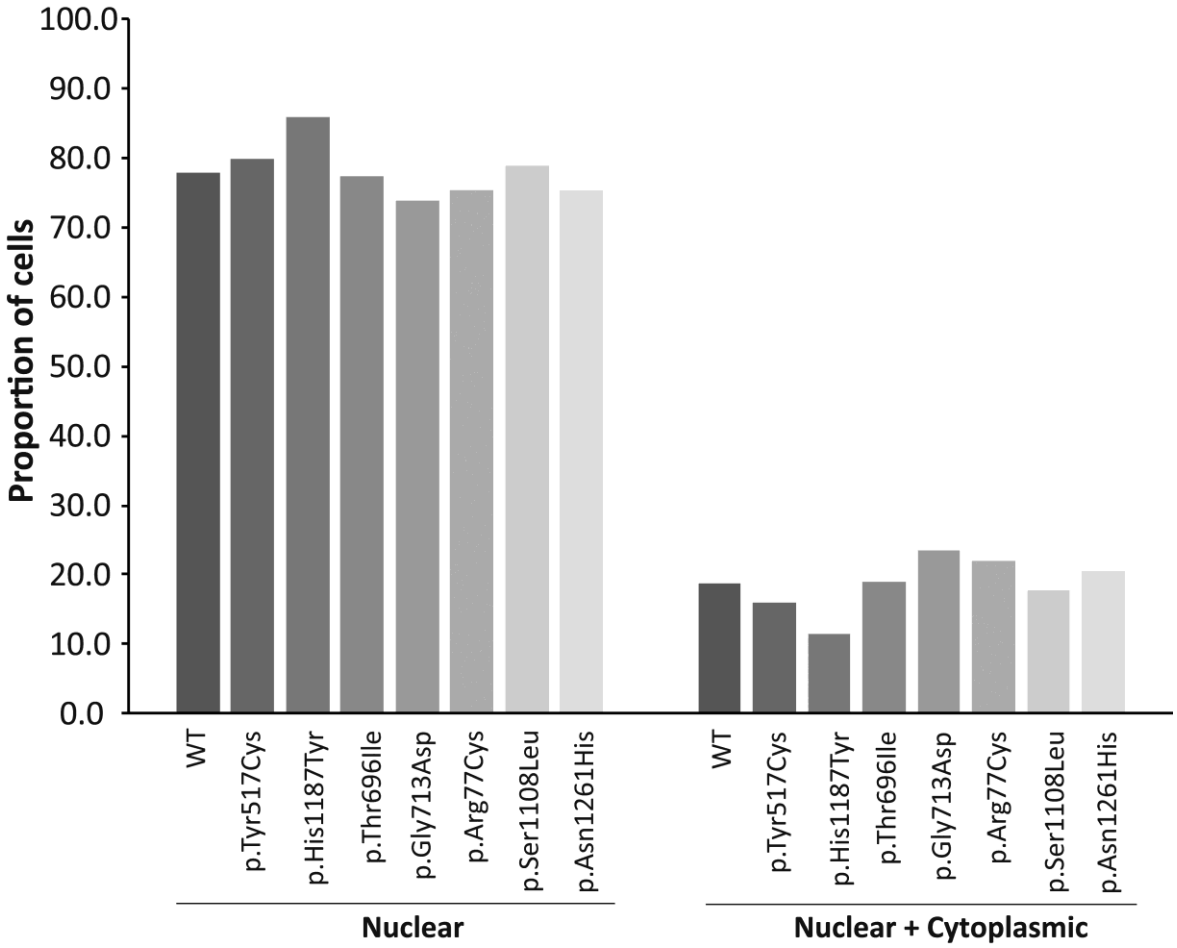

Nuclear
Nuclear and Cytoplasmic
p.Gly713Asp
p.Arg77Cys


Nuclear
Nuclear and Cytoplasmic
p.Ser1108Leu
p.Asn1261His

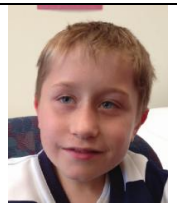
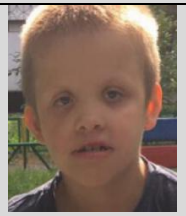
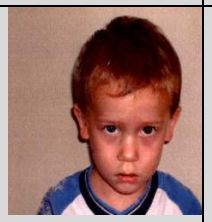
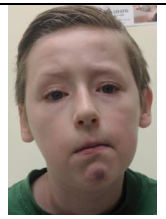
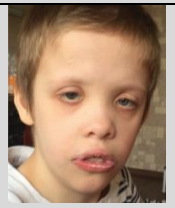
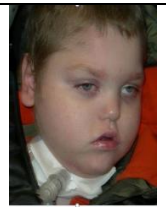
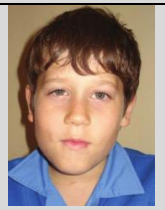
B

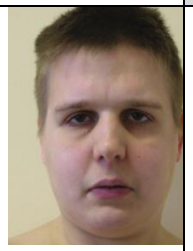
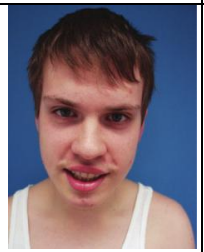
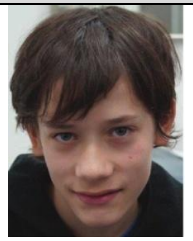
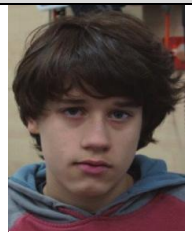
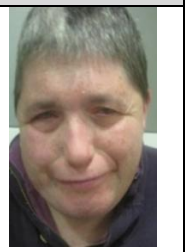
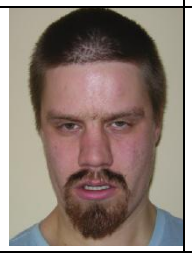
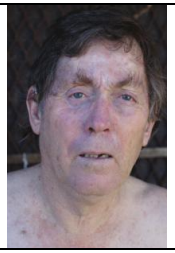
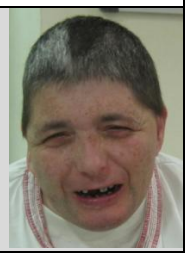

100.0
90.0
80.0
70.0
60.0
50.0
40.0
30.0
20.0
10.0
0.0
Proportion of cells
WT
p.Tyr517Cys
p.His1187Tyr
p.Thr696Ile
p.Gly713Asp
p.Arg77Cys
p.Ser1108Leu
p.Asn1261His
Nuclear
WT
p.Tyr517Cys
p.His1187Tyr
p.Thr696Ile
p.Gly713Asp
p.Arg77Cys
p.Ser1108Leu
p.Asn1261His
Cytoplasmic

C


100.0
90.0
80.0
70.0
60.0
50.0
40.0
30.0
20.0
10.0
0.0
Proportion of cells
WT
p.Tyr517Cys
p.His1187Tyr
p.Thr696Ile
p.Gly713Asp
p.Arg77Cys
p.Ser1108Leu
p.Asn1261His
Nuclear
WT
p.Tyr517Cys
p.His1187Tyr
p.Thr696Ile
p.Gly713Asp
p.Arg77Cys
p.Ser1108Leu
p.Asn1261His
Nuclear + Cytoplasmic

Supplementary Figure 4 Children and adolescent/adults within current and previously reported* *THOC2* cohorts. Images connected by vertical grey shading represent the same individual at different ages.

Children	**Individual**	**1**	**2**	**3 (twin 1)**	**4 (twin 2)**	**5**	**MRX12/V-7***	**11**
	Variant	p.Thr696lle	p.Gly713Asp	p.His1187Tyr	p.His1187Tyr	ex:28c.3503+4A>C	p.Leu438Pro	Arg1121Gy
	ACMG Pathogenicity Classification	*De novo* pathogenic	*De novo* pathogenic	Likely pathogenic	Likely pathogenic	Likely pathogenic	Likely Pathogenic	VOUS
	Age	12	5	4	4	3	-	-
								
	Individual	**6**	**L:2_II-5***	**3_twin 1**	**8**	**MRX12_V-8***	**MRX12_V-7***	**11**
	Variant	Ex35:c.4450-2A>G	p.lle800Thr	p.His1187Tyr	p.Arg77Cys	p.Leu438Pro	p.Leu438Pro	Arg1121Gy
	Pathogenicity	Likely pathogenic	Likely pathogenic	Likely pathogenic	VOUS	Likely pathogenic	Likely pathogenic	VOUS
	Age	10	-	-	9	9	-	-
								
Adolescent	**Individual**	**L22_III-7***	**L:2_II-5***	**D45_III-7***	**MRX12_V-**	**MRX12_V-**	**MRX12_V-**	**11**

and adults					**2***	**8***	**7***	
	Variant	p.lle800Thr	p.lle800Thr	p.Leu313phe	p.Leu438Pro	p.Leu438Pro	p.Leu438Pro	Arg1121Gy
	Pathogenicity	Likely pathogenic	Likley Pathogenic	VOUS	Likely pathogenic	Likely pathogenic	Likely pathogenic	VOUS
	Age	28	44	30	30	-	14	
								
	Individual	**L33_III-3***	**L22_III-2***	**L22_III-1***	**MRC12_III-17***	**MRX12_IV-2***		**11**
	Variant	p.lle800Thr	p.lle800Thr	p.lle800Thr	p.Leu438Pro	p.Leu438Pro		p.Arg1121 Gly
	Pathogenicity	Likely pathogenic	Likley pathogenic	Likley pathogenic	Likely pathogenic	Likely pathogenic		VOUS
	Age	34	42	43				44
								

*Previously published families (Kumar, et al., 2015).

Supplementary Table S1 Primers list (PCR and overlap PCR mutagenesis).

Name	5' → 3'
p.Tyr517Cys	
hTHOC2-1592-A1550G-R	GTTTCATTCTTCCACTGGCCATACAGACGATATCTATGTTGACATGGAAATGT
p.Ser1108Leu Mutagenesis	
hTHOC2_3302_C3323T_F	ATTACAAACTAACCAAGGCATTGGTACATTGCCTTGAAACAG
hTHOC2_3343_G3323A_R	CTGTTTCAAGGCAATGTACCAATGCCTTGGTTAGTTTGTAAT
p.His1187Tyr Mutagenesis	
hTHOC2_3537_C3559T_F	CATGATACCTGAAAATGAGTTTTATCACAAGGATCCCCCTCCGAGGAAT
hTHOC2_3585_G3559A_F	ATTCCTCGGAGGGGGATCCTTGTGATAAAACTCATTTTCAGGTATCATG
p.Asn1261His	
hTHOC2_3758_A3781C_F	CAAGCAATGGAAATAGTGGATCCCACAGCAACAAAGCTGT
hTHOC2_3797_T3781G_F	ACAGCTTTGTTGCTGTGGGATCCACTATTTCCATTGCTTG
p.Thr696Ile	
hTHOC2-2149-2088T-R	CAAAATAACCACCCTCAGCTTTTAGCTGCTCTCCACCAGTCATAGCCTCTAGTTGCTCCATTaTCA
p.Gly713Asp	
hTHOC2_2112_2138G>A_F	TGGTGGAGAGCAGCTAAAAGCTGAGGATGGTTATTTTGGTCAGA
hTHOC2-2216-R	GCAAGATCATGGTCCAATAGAGCATC
p.Arg77Cys	
hTHOC2_208_C229T_F	CTTAGTGACATTAGTGAATTTtGTGAGGATATGCCCTCCATT
hTHOC2_249_G229A_R	AATGGAGGGCATATCCTCACaAAATTCACTAATGTCACTAAG
RT-qPCR THOC2 mRNA expression	
THOC2-RT-PCR-F	GAGAATGAAGGTTATGCCAAGCTGA
THOC2-RT-PCR-R	CGTGTTCTGGCCTGCATTCAAACAC
RT-qPCR Housekeeping mRNA expression	
HPRT1/F	TGACACTGGCAAAACAATGCA
HPRT1/R	GGTCCTTTTCACCAGCAAGCT

PCR amplification and Sanger sequencing of amplified cDNA and gDNA of carrier mother and affected son (c.4450-2A>G)	
hTHOC2-4326F	TACTCCTCCACCACTGTCCAAGAG
hTHOC2-4519-R	TTGTTCCATTCTCCTCTTTACGTCTC
PCR amplification and Sanger sequencing of amplified gDNA of normal father, carrier mother and affected son (c.3503+4A>C)	
hTHOC2-F	GAATCTGGGTCAAGCTTTGG
hTHOC2-R	CTCCCATAATCTCCTGAAGG
PCR amplification and Sanger sequencing of amplified cDNA of normal father, carrier mother and affected son (c.3503+4A>C)	
hTHOC2-ex27F	GGATTTGATGGTGGAAATAAGG
THOC2-ex30R	ACAGCTTTCACACCACACTG
hTHOC2-4570-R	TTGGATAAGGAGATTCACAAGGAC