Genetics in Medicine (2023) 25, 100003

ELSEVIER

Genetics in Medicine
An Official Journal of the ACMG

www.journals.elsevier.com/genetics-in-medicine

ARTICLE

Clustered variants in the 5′ coding region of *TRA2B* cause a distinctive neurodevelopmental syndrome

ARTICLE INFO

Article history:
Received 21 September 2022
Received in revised form
14 December 2022
Accepted 15 December 2022
Available online 20 December 2022

Keywords:
Epilepsy
Infantile spasms
Intellectual disability
Molecular genetics
TRA2B

ABSTRACT

Purpose: Transformer2 proteins (Tra2α and Tra2β) control splicing patterns in human cells, and no human phenotypes have been associated with germline variants in these genes. The aim of this work was to associate germline variants in the *TRA2B* gene to a novel neurodevelopmental disorder.

Methods: A total of 12 individuals from 11 unrelated families who harbored predicted loss-of-function monoallelic variants, mostly de novo, were recruited. RNA sequencing and western blot analyses of Tra2β-1 and Tra2β-3 isoforms from patient-derived cells were performed. Tra2β1-GFP, Tra2β3-GFP and *CHEK1* exon 3 plasmids were transfected into HEK-293 cells.

Results: All variants clustered in the 5′ part of *TRA2B*, upstream of an alternative translation start site responsible for the expression of the noncanonical Tra2β-3 isoform. All affected individuals presented intellectual disability and/or developmental delay, frequently associated with infantile spasms, microcephaly, brain anomalies, autism spectrum disorder, feeding difficulties, and short stature. Experimental studies showed that these variants decreased the expression of the canonical Tra2β-1 isoform, whereas they increased the expression of the Tra2β-3 isoform, which is shorter and lacks the N-terminal RS1 domain. Increased expression of Tra2β-3-GFP were shown to interfere with the incorporation of *CHEK1* exon 3 into its mature transcript, normally incorporated by Tra2β-1.

Conclusion: Predicted loss-of-function variants clustered in the 5′ portion of *TRA2B* cause a new neurodevelopmental syndrome through an apparently dominant negative disease mechanism involving the use of an alternative translation start site and the overexpression of a shorter, repressive Tra2β protein.

Introduction

Most eukaryotic genes comprise exon and intron sequences. Productive gene expression relies on a process called splicing that joins exons together into messenger RNAs (mRNAs) and discards intron sequences. Splicing is catalyzed by an RNA-protein complex called the spliceosome which also generates multiple different mRNAs with different exon contents through alternative splicing. Pathogenic variants in core spliceosome components, as well as in splicing regulatory proteins, have been associated with several human genetic diseases, including retinitis pigmentosa and syndromic conditions.[1-3] Alternative splicing plays a key role in development of the nervous system, and defects in splicing cause multiple neurologic disorders.[4]

The transformer2 (Tra2) proteins are an important group of splicing regulators conserved across the animal

*Correspondence and requests for materials should be addressed to Francis Ramond, Service de Génétique, Hôpital Nord, CHU Saint-Etienne, Saint-Etienne CEDEX 2, France. *E-mail address:* Francis.Ramond@chu-st-etienne.fr OR David Elliott, Newcastle University Biosciences Institute, Newcastle University, Central Parkway, Newcastle NE1 3BZ, England. *E-mail address:* David.Elliott@ncl.ac.uk

A full list of authors and affiliations appears at the end of the paper.

doi: https://doi.org/10.1016/j.gim.2022.100003

kingdom.[5] Vertebrates have 2 distinct transformer2 proteins encoded by separate genes. These are Tra2β protein (transformer 2 β homolog, encoded by the *TRA2B* gene) and Tra2α (encoded by the *TRA2A* gene). Genetic knockout of *TRA2B* prevents embryonic development in mice although heterozygote knockouts are apparently normal.[6] Mice with cortex-specific knockout of *TRA2B* have brain abnormalities but survive to adulthood,[7] whereas more extensive neuronal knockout of *TRA2B* causes death shortly after birth with severe brain abnormalities.[8,9] Tra2β protein also regulates splicing of exons controlling smooth muscle physiology.[10,11] No germline pathogenic variants have thus far been identified for Tra2 proteins. However, Tra2 proteins are related to proteins in the SR family of splicing regulators including SRRM4, which is highly expressed in the mammalian nervous system and associated with neurodevelopmental defects in mutant mice.[4,12]

Tra2 proteins frequently activate splicing inclusion by directly binding to exons via a single RNA-recognition motif (RRM).[5,13,14] Physiological RNA binding sites for Tra2β protein have been globally mapped across the human transcriptome at individual nucleotide resolution, showing that AGAA is the most frequent Tra2β-RNA binding site in humans (although CAA can also be recognized when present within RNA stem-loops).[5,14] Tra2β protein also contains 2 RS domains (regions rich in arginine/serine dipeptides that are important in protein−protein interactions). These are RS1 at the N-terminus and RS2 at the C-terminus (Figure 1). After binding to exons within pre-mRNAs, Tra2β protein stabilizes binding of other spliceosome components via protein−protein interactions involving its RS1 domain, leading to splicing inclusion of exons. The RS2 domain of Tra2β is shorter and not required in vitro for activation of mouse Tra2β target exons.[8] Taken together, these data indicate a key functional requirement for the RRM domain and the RS1 region in splicing control by Tra2β.

The human *TRA2B* gene is located on human chromosome 3 and contains 9 protein-coding exons (Figure 1). Full-length Tra2β protein (called Tra2β-1, NP_004584.1) is translated from an open reading frame initiating within exon 1 and terminating within exon 9 (transcript variant 1, NM_004593.3). The *TRA2B* gene also produces a number of still poorly characterized splice isoforms. In particular, *TRA2B* transcript variant 2 (NM_001243879.2) skips exon 2 and uses a second shorter open reading frame initiating from a downstream ATG located within exon 3.[17,18] This shorter open reading frame encodes a truncated version of the Tra2β protein (called Tra2β-3, NP_001230808.1) that still contains the RRM and RS2 domains but lacks RS1.[15] Tra2β-1 is expressed fairly ubiquitously throughout the body, including the developing human brain, whereas the Tra2β-3 mRNA isoform is more tissue specific and developmentally regulated.[8,15] Importantly, while Tra2β-1 protein frequently activates exon splicing, the functional role and importance of Tra2β-3 protein is not well understood.

The objective of this study was to identify the genetic basis of unexplained neurodevelopmental disorders. We identified individuals in multiple unrelated families with a neurodevelopmental syndrome that have genetic variants mapping within the *TRA2B* gene. This is the first association of genetic germline variants in *TRA2B* causing any genetic disease. Molecular analysis of patient cell lines and transfected cells suggest that clustered variants in the *TRA2B* gene initiate a dominant negative disease mechanism caused by increased expression of the Tra2β-3 protein isoform.

Abbreviation

cDNA – complementary DNA
CDS – coding sequence
DD – developmental delay
DMEM – Dulbecco's Modified Eagle Medium
GFP – green fluorescent protein
gnomAD – genome aggregation database
iCLIP – individual-nucleotide resolution Cross-Linking and Immuno Precipitation
ID – intellectual disability
mRNA – messenger RNA
PCR – polymerase chain reaction
PSI – percentage splicing inclusion
RNAseq – RNA sequencing
RRM – RNA-recognition motif
RS domain – arginine serine rich domain
SR protein – serine arginine rich protein
SRRM4 – serine arginine repetitive matrix 4
Tra2 – transformer2
Tra2α – transformer2 alpha homolog protein
Tra2β – transformer2 β homolog protein
Tra2β-1 – tra2β protein isoform 1
Tra2β-3 – tra2β protein isoform 3

Materials and Methods

Selection of patient cohort

A total of 12 individuals from 11 distinct families with variants in the *TRA2B* gene were enrolled through GeneMatcher.[19] These affected individuals were identified through patient care related to indications including developmental delay (DD) and intellectual disability (ID) and evaluated for genetic causes using exome sequencing in a clinical context. No causative pathogenic or likely pathogenic variants were found in genes already known to be implicated in human diseases. Variants in *TRA2B* were deemed as a potential disease-causing candidates because of their molecular characteristics (absent from Genome Aggregation Database [gnomAD],[20] de novo in most cases, and predicted loss of function in a gene predicted to be intolerant to loss of function). Clinical and molecular data from these individuals are summarized in Figure 2 and Supplemental Table 1. Clinical data were retrospectively retrieved from medical records. Magnetic resonance imaging and electroencephalograms were performed and

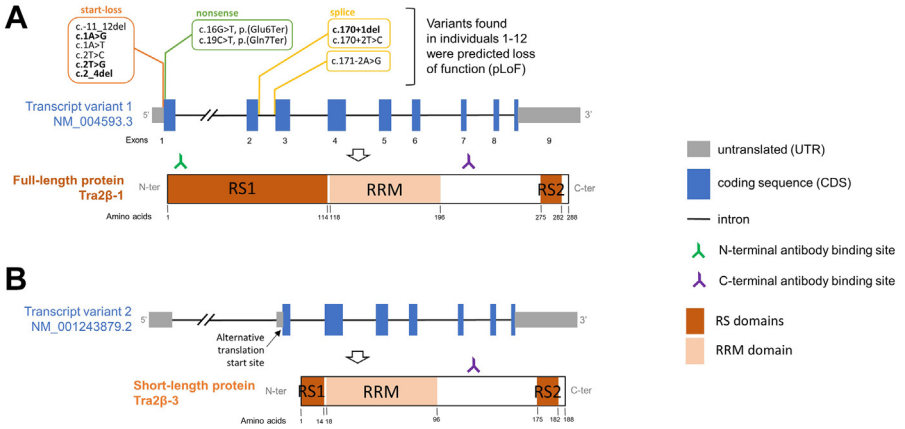

Figure 1 ***TRA2B* gene/protein structure.** Exon/intron structure of 2 alternative transcripts of *TRA2B* are represented in panel A and B. Positions of the variants reported in this study are represented above the transcripts (color code corresponds to predicted effect: orange indicates start-loss, green indicates nonsense, yellow indicates splice-affecting. Further details are given in Supplemental Table 2). Protein isoform encoded by each alternative transcript is represented underneath the transcript. The structures of these proteins are annotated according to the information in Uniprot.[15,16] A. Transcript variant 1 (NM_004593.3, Matched Annotation from NCBI and EMBL-EBI [MANE] Select) coding for the full-length protein (Tra2β-1). B. Transcript variant 2 (NM_001243879.2) coding for the short length protein (Tra2β-3). All variants reported in this study were predicted to be loss of function variants (panel A), ie, predicted to produce an absent protein, either in the absence of the translation start site (start-loss variants) or through nonsense-mediated RNA decay (nonsense and splice-affecting variants). However, an N-truncated protein is produced in the presence of all 4 variants as assessed from western bloting (variants in bold letters in panel A), probably corresponding to Tra2β-3 through the use of an alternative translation start site within *TRA2B* exon 3. RRM, RNA-recognition motif; RS, arginine serine rich domain.

the data were interpreted by trained clinicians through standard care.

Exome sequencing

Genomic DNA was extracted from peripheral white blood cells or buccal swab from affected individuals and parents and used for exome sequencing following standard procedures.[21-24] Clinical variant prioritization was as previously described.[22-24] The GenBank identifier for transcript variant 1 (NM_004593.3) was used to identify variants in the *TRA2B* gene. Variants were confirmed through Sanger sequencing. In most cases (11/12), the *TRA2B* sequence variants detected in affected individuals were established to be de novo through either Sanger sequencing or exome sequencing of parents' DNAs (Supplemental Table 1).

Culture of cells from patient and control individuals

Lymphoblastoid cells derived from patient cells were grown in suspension using Dulbecco's Modified Eagle Medium (DMEM) and 10% fetal calf serum. Patient-derived fibroblasts were grown as adherent cells in DMEM and 10% fetal calf serum.

RNA sequencing of lymphoblastoid cells from affected and control individuals

RNA from lymphoblastoid cells from individual 5 and control lymphoblastoid cells were extracted using a Monarch Total RNA Miniprep Kit (T2010S, NEB) according to manufacturer's instructions and was resuspended in nuclease-free water. RNA quality was checked using an Agilent 2100 Bioanalyzer, and mRNA libraries were prepared using a TruSeq mRNA Library Kit (Illumina). Paired-end sequencing was performed on an Illumina HiSeq 2000 and aligned to the human genome GRCh38 using Spliced Transcripts Alignment to a Reference.[25,26]

Western blotting

Endogenous Tra2β proteins were detected through western blot analysis using the following primary antibodies and dilutions: N-terminal antibody, Tra2β, 1:2000 dilution (ab31353, Abcam); C-terminal antibody, Tra2β, 1:1000 dilution (VPA00344, Bio-Rad). β-actin was detected as a loading control (1:2000 dilution, A5441, Sigma-Aldrich). GFP was detected using the Living Colors A.v. Monoclonal Antibody (JL-8) from Clontech (catalog number 632381) (1:2000 dilution).

Splicing analysis in mammalian cells

The open reading frame of Tra2β-1 cloned into the expression vector pGFP3 was used as previously described.[8] The open reading frame of Tra2β-3 was amplified from a geneblock purchased from IDT, using the primers 5′-AAAAAAAAGAATTCATGTCTACTCGCAG GCGT-3′ and 5′-AAAAAAAAGTCGACTTAATAGCGA CGAGGTGA-3′ and then cloned into pGFP3.[27] Analysis of *CHEK1* minigene splicing patterns were carried out in

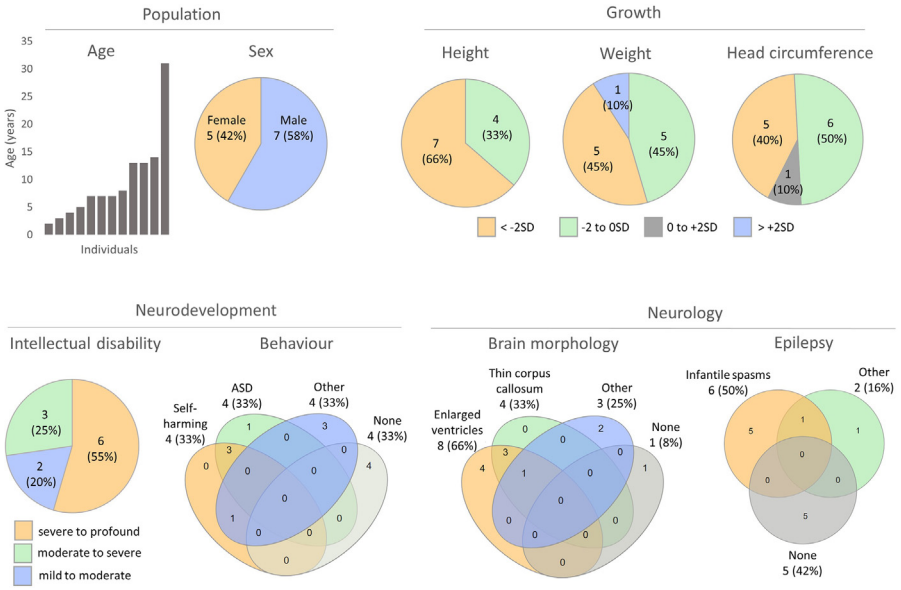

Figure 2 Main clinical characteristics of the 12 individuals reported in this study. Number of individuals affected by a phenotypic trait are indicated, followed in parenthesis by the proportion of these individuals in the cohort. Individuals with no data are excluded. Current age is represented as a histogram. Sex, height, weight, head circumference, and intellectual skills are represented as circle plots. The color legend is indicated underneath the plots when necessary. Behavior, brain morphology, and type of epilepsy are represented as Venn diagrams because some individuals have multiple phenotypes. Because of multiple features in some individuals, the total number can be more than 12 and the sum of proportions can be more than 100%. A more detailed phenotypic and genetic description is available in Supplemental Table 1. ASD, autism spectrum disorder.

HEK-293 cells grown in culture with DMEM and 10% fetal calf serum that were cotransfected as previously reported[5] with plasmids encoding GFP, Tra2β-1-GFP, and Tra2β-3-GFP. Protein expression in transfected cells was monitored using an α-GFP antibody (1:2000 dilution, Living Colors). Replicating HEK293 cells were transfected with the *CHEK1* exon 3 minigene[5] (250 ng) alongside either 100 ng pGFP3 (lane 1 in Figure 3), 250 ng pTra2β1-GFP3 (lane 2 in Figure 3), 250 ng Tra2β1-GFP + 250 ng Tra2 β3-GFP (lane 3 in Figure 3), 250 ng Tra2 β1-GFP + 500 ng Tra2 β3-GFP (lane 4 in Figure 3), 250 ng Tra2β1-GFP + 1 µg Tra2 β3-GFP (lane 5 in Figure 3), or 250 ng Tra2β3-GFP (lane 6 in Figure 3). RNA was prepared using a standard Trizol RNA extraction (catalog number 15596026, Thermo Fisher Scientific). Polymerase chain reaction (PCR) was carried out using a OneStep RT-PCR Kit (210212, Qiagen). RNA was extracted from patient cell lines using Monarch Total RNA Miniprep Kit (T2010S, NEB). Complementary DNA was synthesized in a 10 µl reaction using 500 ng total RNA, using a Superscript VILO cDNA Synthesis Kit (11754050, Invitrogen) and following manufacturer's instructions. Endogenous splicing profiles were monitored using PCR with primers for *CHEK1* exons 2 and 4 and capillary gel electrophoresis on the Qiaxcel Advanced capillary electrophoresis system (catalog number 9002123, Qiagen), and percentage splicing inclusion (PSI) values were calculated as previously described.[5,8] Statistical significance was assessed using *t* tests and GraphPad prism.

Results

Identification of a novel neurodevelopmental syndrome caused by localized variants in *TRA2B*

In this study, 12 individuals from 11 distinct families with variants in the *TRA2B* gene were enrolled through GeneMatcher.[19] All 12 affected individuals carried heterozygous variant located upstream of exon 3 in *TRA2B* (NM_004593.3), which were identified through exome sequencing (Figure 1). These variants would meet the criteria to be classified as likely pathogenic or pathogenic according to American College of Medical Genetics and Genomics guidelines if *TRA2B* were an established disease gene (Supplemental Tables 1 and 2). All of these variants were absent from the general population databases (gnomAD),[20] were predicted to be loss of function variants, and had not been reported previously. Of these 11 distinct variants, 6 were expected to disrupt the translation start site in exon 1 (start-loss variants), 3 were expected to disrupt the splicing process at the exon 2/3 boundary (splice-affecting variants), and the remaining 2 were expected to produce a premature stop codon (truncating variants). Remarkably, none of the 11 variants were located downstream of the alternative start codon responsible for the expression of Tra2β-3. In 10 out of 11 families, the *TRA2B* variant arose de novo. In 1 family, the variant was transmitted from the affected father (individual 8) to his son (individual 9). The grandparents were not available for segregation analysis at

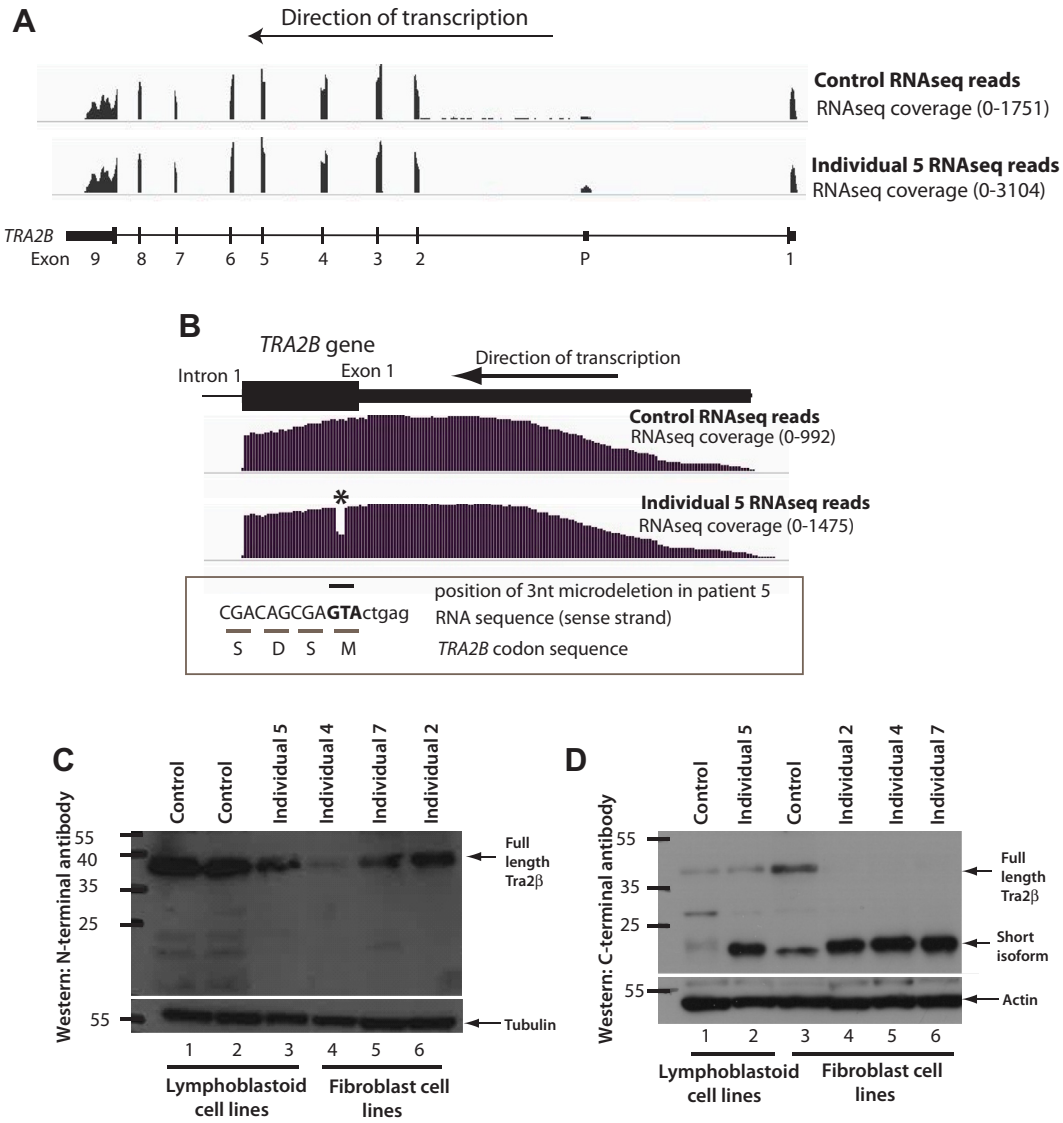

Figure 3 Expression analysis of disease-associated variants. A. Screenshot from the Integrative Genomics Viewer genome browser[28] showing RNAseq reads mapping to *TRA2B* exon 1 from lymphoblastoid cell lines from individual 5 and a control individual. B. The dip in RNAseq reads corresponding to the expression of the allele containing the c.2_4del variant is indicated as an asterisk. C. Western blot showing Tra2β-1 protein expression detected in control and individual cell lines using an antibody specific to the N-terminus of the Tra2β-1 protein isoform. D. Western blot showing Tra2β-1 and Tra2β-3 protein expression detected in control and individual cell lines using an antibody specific to the C-terminal peptide sequence shared between the 2 isoforms. RNASeq, RNA sequencing.

the time of publication. The 12 individuals included 5 females and 7 males, aged between 2 and 14 years with the exception of an adult aged 31 years (father of individual 8).

Clinical and molecular data from these individuals are summarized in Figure 2 and detailed in Supplemental Tables 1 and 2. All individuals (12/12, 100%) had a history of DD or ID. The degree of DD/ID was severe to profound for 6 individuals, moderate to severe for 2, mild to moderate for 3, and unknown for 1. Axial or global hypotonia of neonatal to infancy onset was reported in 10 individuals (10/12, 83%). Motor milestones were delayed in all individuals, with independent walking achieved between ages 20 months and 6 years in 6 individuals (6/11, 55%), not yet achieved in 5 (5/11, 45%), and unknown in 1. Of those 5 individuals who were unable to walk independently, 4 had the potential to achieve walking in the future, either because they could already walk with assistance or stand or because they were too young to walk at the time of publication. Language skills were substantially impaired for most, with 7 individuals (7/12, 58%) being nonverbal at age 3 to 14 years, including 5 individuals older than 7 years. Four individuals acquired some degree of spoken language (4/12, 33%). Behavioral problems were present in 8 individuals (8/12, 66%). Some individuals had

multiple behavioral problems (Figure 2, behavior). In total, autism spectrum disorders were present in 4 individuals (4/12, 33%), self-harming in 4 (4/12, 33%), and frustration intolerance in 2.

Epilepsy was present in 7 individuals (7/12, 58%) with the age at onset between 3 and 9 months. The initial presentation was with infantile spasms in 6 individuals (6/12, 50%) and an unclassified epileptic encephalopathy in 1. Electroencephalograms showed hypsarrhythmia in 3 individuals. In 3 individuals infantile spasms resolved during late infancy. Two of them remained seizure-free and drug-free in adolescence, whereas the third one was seizure-free at the age of 6 after steroids and sodium valproate treatment. For 3 individuals, epilepsy evolved unfavorably: individual 3 had an unspecified epileptic encephalopathy, individual 2 became resistant to multiple antiepileptic drugs and to vagus nerve stimulation at age 6 years after an initial period free of seizures under ketogenic diet from age 1 to 5, and individual 11 transitioned from infantile spasms to a Lennox-Gastaut syndrome. For individual 1 with infantile spasms, the course of epilepsy is unknown. Head circumference was below average in 10 individuals, including microcephaly from −2 to −4 SD in 5 of them (5/12, 42%). All individuals underwent neuroimaging with the exception of individual 9 (Supplemental Table 1). Brain anomalies were noted in 10 individuals (10/11, 90%), including 6 exhibiting mild diffuse brain atrophy or enlargement of the ventricles (6/11, 54%), 4 with thin or hypoplastic corpus callosum (4/11, 36%), 2 with delayed myelination, and 1 with cerebellar hypoplasia. Individual 11, who underwent surgery for intractable seizures before genetic diagnosis, had histopathologically confirmed focal cortical dysplasia.

Growth was below average for almost all individuals. Height and weight were below −2 SD for 4 individuals (4/12, 33%). Feeding difficulties were reported in 7 individuals (7/12, 58%), including 1 infant requiring percutaneous gastrostomy. Chronic constipation was present in 5 individuals (5/12, 42%). Sleep disturbances in the form of night-time awakenings were reported in 3 individuals (3/12, 25%). One of them responded favorably to melatonin. Minor cardiac malformations were noted in 2 individuals (2/12, 17%). Vision was frequently affected: 3 individuals had a cortical visual impairment, 2 had an intermittent exotropia, and 2 had astigmatism. Hearing was unaffected in all individuals. Skeletal abnormalities were common and mild, including scoliosis in 4 individuals (4/12, 33%), hand and/or feet brachydactyly in 3 (3/12, 25%), cutaneous syndactyly in 2 (2/12, 17%), and pectus excavatum in 1.

Facial similarities were noted between some individuals (Figure 4), sharing some resemblance with Angelman syndrome or chromatinopathies. The following genetic abnormalities were specifically investigated and excluded in some individuals because of their consistent phenotype: Angelman syndrome through methyl-specific PCR (individuals 3, 4, 6, and 7), mitochondrial DNA sequencing (individuals 3, 5, and 11), *ARX* (individual 5), *MECP2* (individual 6), and Coffin-Siris (individual 7).

Detection of an N-terminal truncated Tra2β protein isoform within affected individuals

To investigate the molecular mechanism responsible for the phenotype in our cohort, we first assessed the effects of the variants on the *TRA2B* transcript through RNA sequencing (RNAseq) analysis in lymphoblastoid cells derived from individual 5. In this individual, a heterozygous deletion of 3 nucleotides (c.2_4del) was expected to remove the start codon of the Tra2β-1 protein isoform (Supplemental Table 2). Alignment of RNAseq reads to the human genome did not highlight any reduction in the overall expression level or change in the splicing pattern of *TRA2B* mRNA between individual 5 and control individuals (Figure 3A). RNAseq analysis revealed an approximately 50% dip in read coverage over the deletion, which was not observed in the control cell line (Figure 3B). Hence, the transcript corresponding to the c.2_4del variant is expressed at similar levels to the wild-type allele and not targeted for mRNA degradation.

Because all the variants in our cohort were predicted to disrupt the expression of Tra2β-1 isoform, we next assessed the effect of the variants on the Tra2β protein using western blot analysis. Protein was extracted from fibroblasts or lymphoblastoid cells derived from individual 2 (c.170+1del splice variant), individual 4 (c.2T>G start-loss variant), individual 5 (c.2_4del start-loss variant), and individual 7 (c.1A>G start-loss variant) (Supplemental Table 2). Consistent with the predicted effect of the 5′ *TRA2B* variants, we observed decreased levels of full-length Tra2β-1 protein (~39 kDa) in these affected individuals compared with controls using an antibody specific to the N-terminus of the protein (Figure 3C). To monitor whether the patient-associated variants also affected the expression of Tra2β-3 we used an antibody raised against the C-terminus of the protein, which binds to both Tra2β-1 and Tra2β-3 isoforms. This C-terminus antibody confirmed the decrease in Tra2β-1 levels within affected individuals' cells and also showed increased expression of a shorter protein (~20 kDa) of the size predicted for the Tra2β-3 isoform. This shorter protein isoform was barely detected in control lymphoblastoid cells and also more weakly observed in control fibroblast cells than in patient-derived fibroblasts (Figure 3D). These data suggest that while reducing the expression of Tra2β-1, as expected, the variants found in our cohort also dramatically increased the expression of the otherwise poorly expressed Tra2β-3 protein isoform.

We next investigated whether this Tra2β protein isoform exchange had an effect on Tra2β function. One of the best characterized splicing targets of human Tra2β is within the checkpoint kinase 1 gene (*CHEK1*). Tra2β-1 activates splicing inclusion of *CHEK1* exon 3, and small interfering RNA−mediated depletion of Tra2β induces a 55% reduction

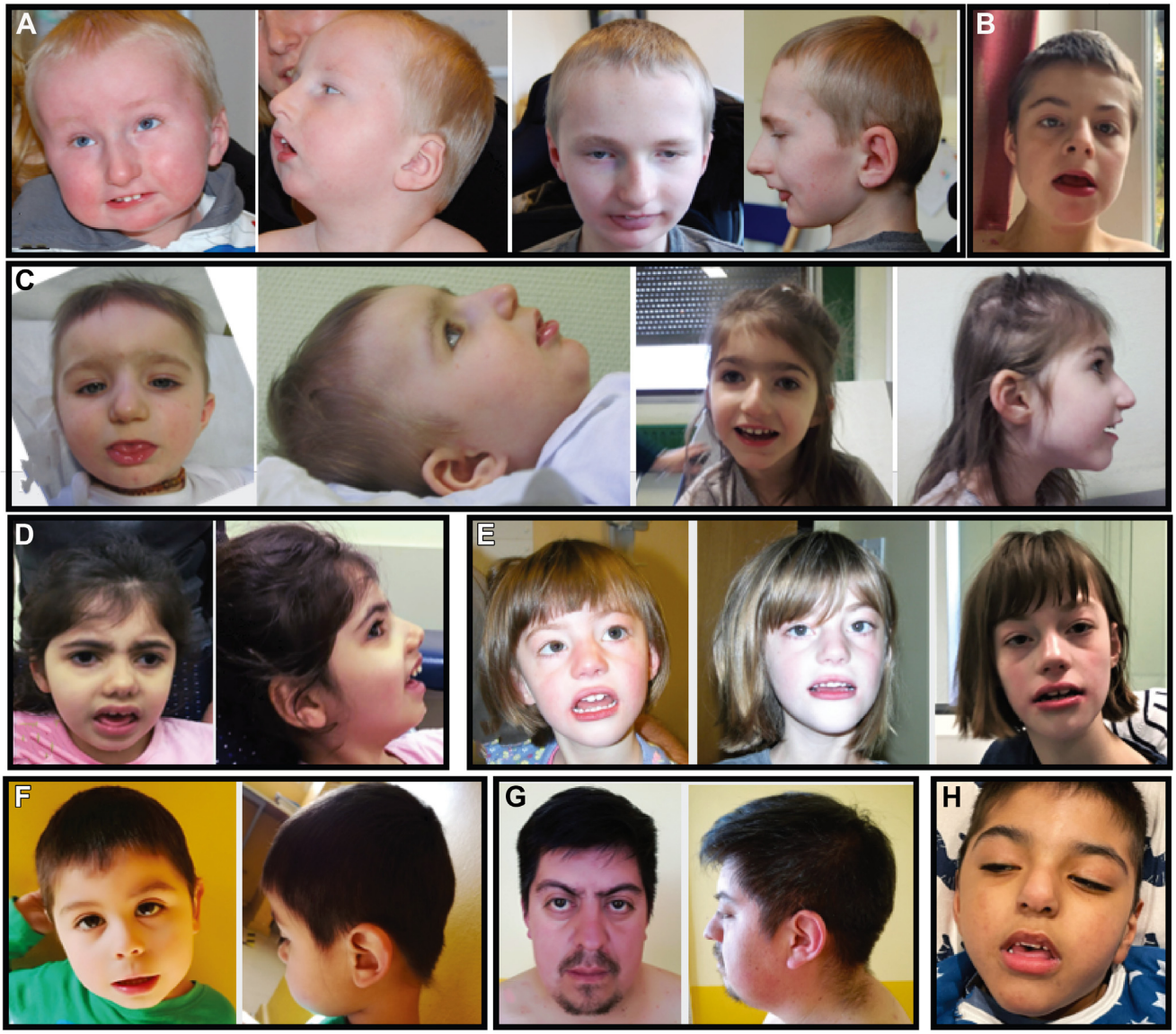

Figure 4 Facial pictures of *TRA2B* individuals. A. Individual 2 at age 1 and 13 years. B. Individual 5 (age unspecified). C. Individual 4 at age 1 and 5 years. D. Individual 3 at age 6 years. E. Individual 7 at age 3, 10, and 13 years. F. Individual 8 at age 4 years. G. Individual 9 at age 31 years. H. Individual 11 at age 3 years. The following features were considered common but not constant in the cohort by 4 independent clinical geneticists: long and slightly upslanting palpebral fissures, deep set eyes, epicanthi, high nose bridge, bulbous nose tip, anteverted nares, smooth and short philtrum, thin upper lip, fleshy lower lip, open mouth, widely spaced teeth, low set fleshy ears, and rectangular end of the chin.

in the splicing inclusion of this exon in breast cancer cells.[5] Both analysis of exon junction reads within our RNAseq data (Figure 5A) and experimental isoform-specific reverse transcription PCR (Figure 5B) detected a reduced PSI of *CHEK1* exon 3 in individual 5–derived lymphoblastoid cells compared with control lymphoblastoid cells. However, the size of reduction in *CHEK1* exon 3 splicing inclusion within lymphoblastoid cells from individual 5 was less than that previously detected after small interfering RNA–mediated depletion of both Tra2α and Tra2β in MDA-MB-231 cells.[5]

The N-terminal truncated Tra2β-3 protein isoform operates as a splicing repressor

Because of (1) the striking discrepancy between the lack of phenotype in *Tra2b* heterozygous knockout mice[6] compared with the severe defects detected in affected individuals described in this study, (2) overexpression of Tra2β-3 protein in affected individuals, and (3) the unchanged expression of *TRA2B* canonical transcript (NM_004593.3) in individual 5 when compared with controls, we hypothesized that the main cause of the disease was not a Tra2β-1

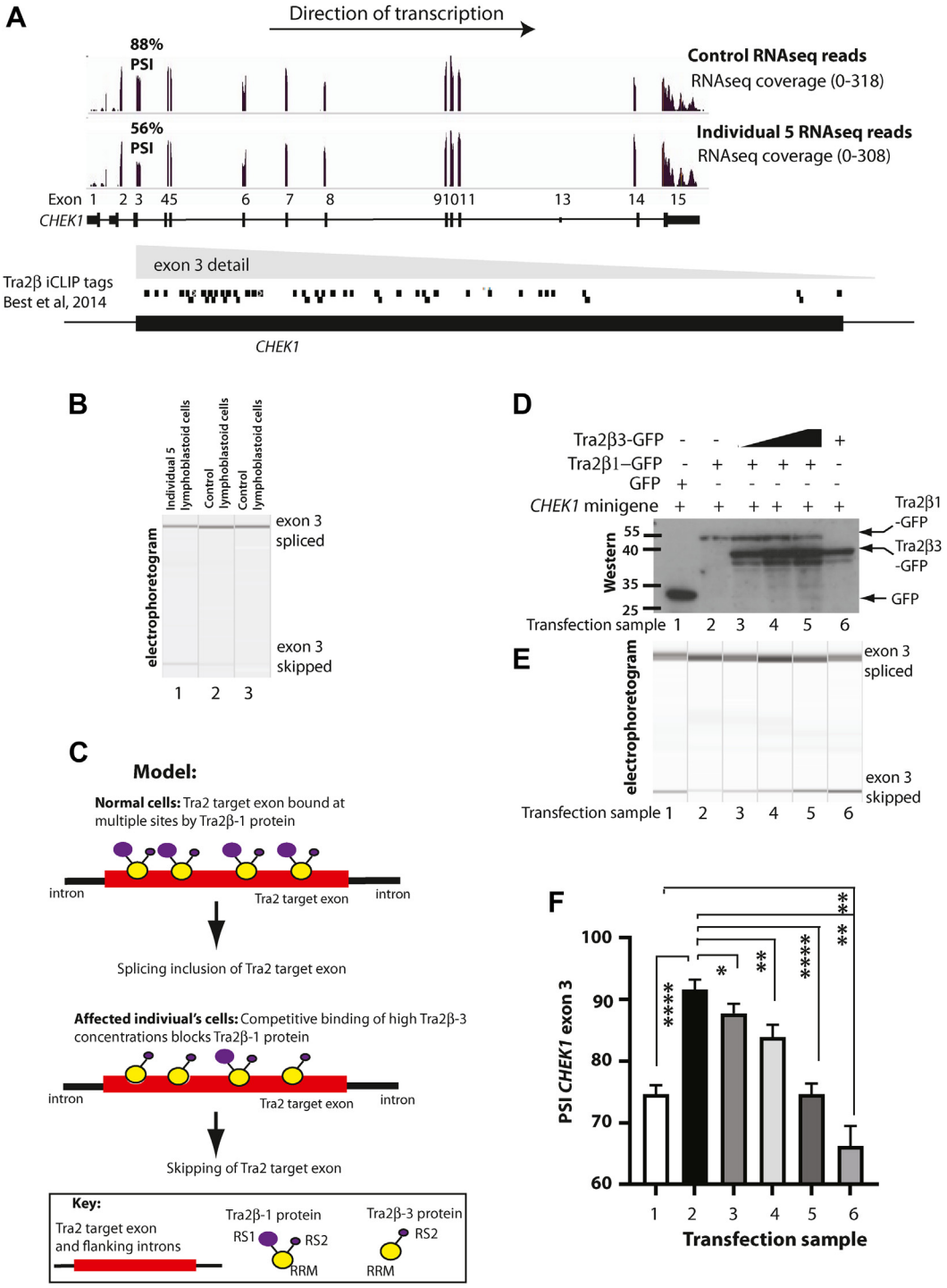

Figure 5 Molecular analysis of disease mechanism. A. Screenshot of the Integrative Genomics Viewer (IGV) genome browser[28] showing RNAseq reads from individual 5 and a control lymphoblastoid control cell line mapping to the *CHEK1* gene. The corresponding *CHEK1* gene exon/intron structure and position of Tra2β-1 protein binding sites identified by individual-nucleotide resolution Cross-Linking and ImmunoPrecipitation (iCLIP)[5] are indicated. The percentage splicing inclusion (PSI) levels for *CHEK1* exon 3 are shown for each of the RNAseq tracks, after calculation from exon junction reads displayed in IGV.[28] A more detailed picture of *CHEK1* exon 3 is shown underneath, better resolving the multiple Tra2β-1 protein binding sites (revealed by iCLIP tags[5]) that map to this exon. B. Experimental detection of splicing inclusion of *CHEK1* exon 3 in lymphoblastoid cells from individual 5 and a control lymphoblastoid cell line using reverse transcription polymerase chain reaction (RT-PCR) primers for *CHEK1* exons 2 and 4 for analysis as previously described.[5] C. Model showing the proposed dominant negative disease mechanism in our cohort. Tra2β-1 protein isoform normally binds to multiple target sites within Tra2-responsive exons to activate their splicing inclusion. In our cohort the Tra2β-3 protein variant lacking the RS1 domain would competitively bind the same target sites as Tra2β-1, reducing the levels of splicing activation of associated exons and increasing exon

haploinsufficiency but rather a consequence of increased levels of Tra2β-3. Supporting this idea, previous in vitro experiments have shown that an engineered version of the Tra2β protein that is missing RS1 (similar to Tra2β-3) can operate as a splicing repressor on some Tra2β-1 target exons identified within mouse testis and brain.[8] At the molecular level, Tra2β-1 protein activates exon splicing via binding to recognition sites inside target exons.[5] iCLIP (individual-nucleotide resolution cross-linking and immunoprecipitation) and mutagenesis experiments have shown that most Tra2β-dependent exons have multiple nonredundant binding sites for Tra2β protein.[5,8] This includes *CHEK1* exon 3 (Tra2β binding sites within this exon that have been identified through iCLIP experiments[5] are shown in Figure 5B). Mutational inactivation of individual Tra2β binding sites leads to reduced splicing inclusion of Tra2β-responsive exons, even though other potential binding sites remain, consistent with requirement of binding of multiple Tra2β proteins for splicing activation.[5,8] The Tra2β-3 protein isoform lacks the RS1 domain that is required to recruit the spliceosome and enable splicing activation, but it still contains the RRM and therefore can bind to Tra2β target exons. In our cohort, the excess of the Tra2β-3 isoform could thus occupy Tra2β binding sites, competing with Tra2β-1 for binding to target exons, and thereby reducing splicing inclusion of Tra2β target exons in key genes (Figure 5C).

To evaluate this hypothesis of a dominant negative mechanism, we precisely cloned the human Tra2β-3 reading frame into an expression vector in frame with the reading frame for GFP and transfected this expression construct into HEK293 cells. As expected, western blotting of transfected cell extracts detected a shorter protein in cells expressing Tra2β-3 GFP fusion protein than the protein in cells expressing Tra2β-1-GFP fusion protein (Figure 5D).[5] To monitor relative splicing activities of these fusion proteins, we also cotransfected a Tra2β-1-responsive minigene containing *CHEK1* exon 3 plus flanking intron sequences[5] alongside the expression constructs (Figure 5D and 5E). Exactly consistent with the previous data, cotransfection of the Tra2β-1-GFP expression vector activated splicing inclusion of *CHEK1* exon 3[5] (Figure 5E and 5F, compare lanes 1 and 2). Supporting our proposed dominant negative model, cotransfection of the Tra2β-3-GFP expression vector with the minigene resulted in significantly reduced *CHEK1* exon 3 splicing inclusion (Figure 5E and 5F, compare lanes 1 and 6). Moreover, transfection of increasing amounts of Tra2β-3-GFP led to progressively decreased splicing inclusion of *CHEK1* exon 3 even with cotransfection with Tra2β-1-GFP (Figure 5E and 5F, compare lanes 2-5). These experiments are consistent with a molecular mechanism in which the Tra2β-3 isoform is not able to activate splicing inclusion of *CHEK1* exon 3, and increased levels of the Tra2β-3 protein isoform detected in patients' cells operate to counteract the splicing activity of the reduced endogenous Tra2β-1 protein isoform expressed from the other allele.

Discussion

In this study, we have identified variants in the *TRA2B* gene that are very likely responsible for a new neurodevelopmental syndrome. This syndrome involves ID/DD of mostly severe degree, frequently associated with epilepsy (particularly infantile spasms), behavioral abnormalities (including autism spectrum disorder and self-injuries), brain abnormalities (including microcephaly, brain atrophy or enlargement of the ventricles, and corpus callosum hypoplasia), feeding disorders, growth retardation, nonspecific visual abnormalities, and some less frequently observed features. Among the 11 independent families reported in this article, the variants occurred de novo in 10 and was inherited from a similarly affected parent in 1. Some individuals share a degree of facial resemblance although not very specific. The mode of transmission is autosomal dominant, with possibly full penetrance.

The genetic variants associated with this neurodevelopmental syndrome are strictly localized within the 5′ coding region of the *TRA2B* gene. Our data suggest a dominant negative disease mechanism in which localized heterozygous variants in *TRA2B* cause the overexpression of an otherwise poorly expressed Tra2β-3 protein isoform. In this model, increased expression of Tra2β-3 protein changes the splicing inclusion of key Tra2β-dependent exons during development by competing with the canonical Tra2β-1 protein. Such splicing changes then cause the neurodevelopmental defects observed in our cohort. The precise genes and exons that might be responsible for the neurodevelopmental syndrome remain to be identified, although Tra2α and Tra2β are known to regulate exon inclusion of genes, such as *NIPBL*, *ATRX*, and *SON*, important in syndromes with intellectual deficiency among others.[5] Haploinsufficiency or loss of function of these genes are responsible for Cornelia de Lange type 1 (OMIM 122470), alpha-thalassemia/mental retardation (OMIM 301040), and ZTTK (OMIM 617140) syndromes, respectively, whose manifestations overlap to those seen in our individuals. However, a limitation of this study is that we cannot rule out a gain of function mechanism via increased expression of the Tra2β-3 protein isoform causing changes independent of Tra2β-1. Another limitation is the absence of demonstration of causality through animal models. However, we believe that the specific molecular pattern of the variants, their rarity and

skipping. D. Western blots of protein extracted from HEK293 cells after transfection of expression constructs and minigenes (full details given in Materials and Methods). This western blot was probed for GFP. E. Capillary gel electrophoretogram after RT-PCR analysis of minigene splicing patterns. F. Bar chart showing mean PSI levels of *CHEK1* exon 3 from each group of transfected HEK293 cells. All data represented by bar charts were generated from 3 biological replicates and error bars represent SEM. All probability *P* values were calculated using an independent two-sample *t* test between the PSI levels as indicated on the bar chart (statistical significance: $*P < .05$, $**P < .01$, $****P < .0001$). RNASeq, RNA sequencing; RRM, RNA-recognition motif.

mode of inheritance, their effect on Tra2β proteins, and the phenotypic resemblance among individuals, are convincing enough to support their involvement in a new syndrome.

Considering possible disease mechanisms, we expect that haploinsufficiency of the *TRA2B* gene would not result in the same phenotype as seen in the individuals described in this study. Indeed, heterozygous deletions of *TRA2B* are reported in the Database of Genomic Variants (http://dgv.tcag.ca/) in 21 supposedly healthy individuals (esv2660289, nsv1142591) within a cohort of 1153.[29] Further supporting our hypothesis, a 455-kilobase deletion encompassing *TRA2B* and several nearby genes was inherited from an asymptomatic parent in an individual with an apparently nonsyndromic neurodevelopmental disorder entered in the Database of genomic variation and phenotype in humans using Ensembl resources (DECIPHER) (https://www.deciphergenomics.org; patient 276055).[30] Unfortunately, we were unable to obtain more information on the individual and his family, who were lost to follow-up. Finally, no copy number variants of modest size spanning the *TRA2B* gene are present in affected individuals in the Clinical Genome Resource (https://clinicalgenome.org)[31] and dbVar databases (https://www.ncbi.nlm.nih.gov/dbvar/).[32] Thus, heterozygous deletions of the entire *TRA2B* gene leading to its haploinsufficiency might not be pathogenic. This is not inconsistent with the predicted intolerance of *TRA2B* to loss-of-function variants based on the 0.99 probability of being loss-of-function intolerant score reported for this gene in the gnomAD (https://gnomad.broadinstitute.org/).[20] Indeed, variants such as those identified in our cohort would be classified as pathogenic in gnomAD in case they are located upstream of exon 3 of *TRA2B*.

In conclusion, to our knowledge, this study is the first description of a genetic disease associated with alterations in a *TRA2*-family gene and further confirms the essential role of alternative splicing control mechanisms in human neurodevelopment.[4] We describe an original dominant negative function mechanism involving the activation of an alternate translation start site in the presence of predicted loss-of-function heterozygous variants clustered in the 5′ part of the gene.

Data Availability

The data that support the findings of this study are available from the corresponding author upon request.

Acknowledgments

We thank all of the individuals and their families for their participation in this study. V.F. and L.S.M. acknowledge Paulina Morales and Angela Peña for their technical support.

Funding

This work was supported by the Biotechnology and Biological Sciences Research Council grants BB/S008039/1 and BB/W002019/1. R.G. is supported by the Tuscany Region Call for Health 2018 (grant DECODE-EE) and the Brain project by Fondazione CA.RI.FI. R.G. and A.V. are members of the European Reference Network for rare and complex epilepsies. V.F. was funded by the Vice-chancellorship for Research and Development of the University of Chile (Project code #0313/2022).

Author Information

Conceptualization: F.R., C.E.P., D.J.E.; Funding Acquisition: D.J.E.; Investigation: F.R., C.D., M.G., O.W., A.V., R.G., D.F., R.L.P., M.L., A.B., U.G., M.B., D.W., P.D.T., V.F., L.S.M., C.M.F., P.M., S.A.H., S.V.M., E.T., B.O.-J., L.B.S., N.O., N.R.S., O.H., L.B., S.H., C.E.P., T.H., D.J.E.; Supervision: F.R., D.J.E.; Writing - original draft: F.R, V.F., D.J.E.; Writing - review & editing: C.D., M.G., O.W., A.V., R.G., D.F., R.L.P., A.B., P.D.T., S.H., T.H.

Ethics Declaration

This study was approved by the Medical Ethics Committee of the Saint-Etienne University Hospital, acting as a central Institutional Review Board. Written informed consent for DNA analysis and the use of medical data for this publication was obtained from parents or legal representatives of the children. All participants and families consented to take part in the study and gave permission to publish photographs when corresponded.

Conflict of Interest

S.V.M. and E.T. are employees of GeneDx, LLC. All other authors declare no conflicts of interest.

Additional Information

The online version of this article (https://doi.org/10.1016/j.gim.2022.100003) contains supplementary material, which is available to authorized users.

Authors

Francis Ramond[1,*], Caroline Dalgliesh[2], Mona Grimmel[3], Oded Wechsberg[4,5], Annalisa Vetro[6], Renzo Guerrini[6], David FitzPatrick[7], Rebecca L. Poole[8], Marine Lebrun[1], Allan Bayat[9,10], Ute Grasshoff[3], Miriam Bertrand[3],

Dennis Witt[3], Peter D. Turnpenny[11], Víctor Faundes[12], Lorena Santa María[12], Carolina Mendoza Fuentes[13], Paulina Mabe[14], Shaun A. Hussain[15], Sureni V. Mullegama[16], Erin Torti[16], Barbara Oehl-Jaschkowitz[17], Lina Basel Salmon[4,18,19,20], Naama Orenstein[4,18], Noa Ruhrman Shahar[19], Ofir Hagari[19], Lily Bazak[19], Sabine Hoffjan[21], Carlos E. Prada[22,23], Tobias Haack[3,24], David J. Elliott[2,*]

Affiliations

[1]Service de Génétique, Hôpital Nord, CHU Saint-Etienne, Saint-Etienne, France; [2]Newcastle University Biosciences Institute, Newcastle University, Newcastle upon Tyne, United Kingdom; [3]Institute of Medical Genetics and Applied Genomics, University of Tuebingen, Tübingen, Germany; [4]Pediatric Genetics Unit, Schneider Children's Medical Center of Israel, Petach Tikva, Israel; [5]Maccabi Healthcare Services, Tel Aviv, Israel; [6]Neuroscience Department, Meyer Children's Hospital and University of Florence, Florence, Italy; [7]MRC Human Genetics Unit, The University of Edinburgh, Edinburgh, Scotland, United Kingdom; [8]NHS Education for Scotland South East Region, South East of Scotland Clinical Genetics Service, Edinburgh, United Kingdom; [9]Institute for Regional Health Services, University of Southern Denmark, Odense, Denmark; [10]Department of Epilepsy Genetics and Personalized Medicine, The Danish Epilepsy Center, Dianalund, Denmark; [11]Clinical Genetics, Royal Devon University Healthcare NHS Foundation Trust, Exeter, United Kingdom; [12]Laboratorio de Genética y Enfermedades Metabólicas, Instituto de Nutrición y Tecnología de los Alimentos, Universidad de Chile, Santiago, Chile; [13]Unidad de Endocrinología, División de Pediatría, Escuela de Medicina, Pontificia Universidad Católica de Chile, Santiago, Chile; [14]Unidad de Neurología, Hospital de Niños Dr. Exequiel González Cortés, Santiago, Chile; [15]Division of Pediatric Neurology, University of California, Los Angeles, Los Angeles, CA; [16]Clinical Genomics Program, GeneDx, MD; [17]Practice of Human Genetics, Homburg (Saar), Germany; [18]Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel; [19]The Raphael Recanati Genetic Institute, Rabin Medical Center, Beilinson Hospital, Petach Tikva, Israel; [20]Pediatric Immunogenetics, Felsenstein Medical Research Center, Petach Tikva, Israel; [21]Abteilung für Humangenetik, Ruhr-Universitat Bochum, Bochum, Germany; [22]Division of Genetics, Birth Defects and Metabolism, Ann & Robert H. Lurie Children's Hospital of Chicago, Chicago, IL; [23]Department of Pediatrics, Feinberg School of Medicine of Northwestern University, Chicago, IL; [24]Centre for Rare Diseases, University of Tuebingen, Tuebingen, Germany

References

1. Yang C, Georgiou M, Atkinson R, et al. Pre-mRNA processing factors and retinitis pigmentosa: RNA splicing and beyond. *Front Cell Dev Biol*. 2021;9:700276. http://doi.org/10.3389/fcell.2021.700276
2. Wood KA, Eadsforth MA, Newman WG, O'Keefe RT. The role of the U5 snRNP in genetic disorders and cancer. *Front Genet*. 2021;12:636620. http://doi.org/10.3389/fgene.2021.636620
3. Homsy J, Zaidi S, Shen Y, et al. De novo mutations in congenital heart disease with neurodevelopmental and other congenital anomalies. *Science*. 2015;350(6265):1262-1266. http://doi.org/10.1126/science.aac9396
4. Vuong CK, Black DL, Zheng S. The neurogenetics of alternative splicing. *Nat Rev Neurosci*. 2016;17(5):265-281. http://doi.org/10.1038/nrn.2016.27
5. Best A, James K, Dalgliesh C, et al. Human Tra2 proteins jointly control a CHEK1 splicing switch among alternative and constitutive target exons. *Nat Commun*. 2014;5:4760. http://doi.org/10.1038/ncomms5760
6. Mende Y, Jakubik M, Riessland M, et al. Deficiency of the splicing factor Sfrs10 results in early embryonic lethality in mice and has no impact on full-length SMN/Smn splicing. *Hum Mol Genet*. 2010;19(11):2154-2167. http://doi.org/10.1093/hmg/ddq094
7. Roberts JM, Ennajdaoui H, Edmondson C, Wirth B, Sanford JR, Chen B. Splicing factor TRA2B is required for neural progenitor survival. *J Comp Neurol*. 2014;522(2):372-392. http://doi.org/10.1002/cne.23405
8. Grellscheid S, Dalgliesh C, Storbeck M, et al. Identification of evolutionarily conserved exons as regulated targets for the splicing activator Tra2β in development. *PLoS Genet*. 2011;7(12):e1002390. http://doi.org/10.1371/journal.pgen.1002390
9. Storbeck M, Hupperich K, Gaspar JA, et al. Neuronal-specific deficiency of the splicing factor Tra2b causes apoptosis in neurogenic areas of the developing mouse brain. *PLoS One*. 2014;9(2):e89020. http://doi.org/10.1371/journal.pone.0089020
10. Shukla S, Fisher SA. Tra2beta as a novel mediator of vascular smooth muscle diversification. *Circ Res*. 2008;103(5):485-492. http://doi.org/10.1161/CIRCRESAHA.108.178384
11. Fu K, Mende Y, Bhetwal BP, et al. Tra2β protein is required for tissue-specific splicing of a smooth muscle myosin phosphatase targeting subunit alternative exon. *J Biol Chem*. 2012;287(20):16575-16585. http://doi.org/10.1074/jbc.M111.325761
12. Quesnel-Vallières M, Irimia M, Cordes SP, Blencowe BJ. Essential roles for the splicing regulator nSR100/SRRM4 during nervous system development. *Genes Dev*. 2015;29(7):746-759. http://doi.org/10.1101/gad.256115.114
13. Cléry A, Jayne S, Benderska N, Dominguez C, Stamm S, Allain FHT. Molecular basis of purine-rich RNA recognition by the human SR-like protein Tra2-β1. *Nat Struct Mol Biol*. 2011;18(4):443-450. http://doi.org/10.1038/nsmb.2001
14. Tsuda K, Someya T, Kuwasako K, et al. Structural basis for the dual RNA-recognition modes of human Tra2-β RRM. *Nucleic Acids Res*. 2011;39(4):1538-1553. http://doi.org/10.1093/nar/gkq854
15. Nayler O, Cap C, Stamm S. Human transformer-2-beta gene (SFRS10): complete nucleotide sequence, chromosomal localization, and generation of a tissue-specific isoform. *Genomics*. 1998;53(2):191-202. http://doi.org/10.1006/geno.1998.5471
16. UniProt Consortium. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res*. 2021;49(D1):D480-D489. http://doi.org/10.1093/nar/gkaa1100
17. Harrow J, Frankish A, Gonzalez JM, et al. GENCODE: the reference human genome annotation for the ENCODE project. *Genome Res*. 2012;22(9):1760-1774. http://doi.org/10.1101/gr.135350.111
18. Pruitt KD, Brown GR, Hiatt SM, et al. RefSeq: an update on mammalian reference sequences. *Nucleic Acids Res*. 2014;42(Database issue):D756-D763. http://doi.org/10.1093/nar/gkt1114

19. Sobreira N, Schiettecatte F, Valle D, Hamosh A. GeneMatcher: a matching tool for connecting investigators with an interest in the same gene. *Hum Mutat*. 2015;36(10):928-930. http://doi.org/10.1002/humu.22844
20. Karczewski KJ, Francioli LC, Tiao G, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*. 2020;581(7809):434-443. Published correction appears in *Nature*. 2021;590(7846):E53. Published correction appears in *Nature*. 2021;597(7874):E3-E4. https://doi.org/10.1038/s41586-020-2308-7
21. Guillen Sacoto MJ, Tchasovnikarova IA, Torti E, et al. De novo variants in the ATPase module of MORC2 cause a neurodevelopmental disorder with growth retardation and variable craniofacial dysmorphism. *Am J Hum Genet*. 2020;107(2):352-363. http://doi.org/10.1016/j.ajhg.2020.06.013
22. Falb RJ, Müller AJ, Klein W, et al. Bi-allelic loss-of-function variants in KIF21A cause severe fetal akinesia with arthrogryposis multiplex. *J Med Genet*. Published online November 5, 2021. http://doi:10.1136/jmedgenet-2021-108064
23. Faundes V, Newman WG, Bernardini L, et al. Histone lysine methylases and demethylases in the landscape of human developmental disorders. *Am J Hum Genet*. 2018;102(1):175-187. http://doi.org/10.1016/j.ajhg.2017.11.013
24. Faundes V, Jennings MD, Crilly S, et al. Impaired eIF5A function causes a Mendelian disorder that is partially rescued in model systems by spermidine. *Nat Commun*. 2021;12(1):833. http://doi.org/10.1038/s41467-021-21053-2
25. Karolchik D, Barber GP, Casper J, et al. The UCSC Genome Browser database: 2014 update. *Nucleic Acids Res*. 2014;42(Database issue):D764-D770. http://doi.org/10.1093/nar/gkt1168
26. Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29(1):15-21. http://doi.org/10.1093/bioinformatics/bts635
27. Venables JP, Dalgliesh C, Paronetto MP, et al. SIAH1 targets the alternative splicing factor T-STAR for degradation by the proteasome. *Hum Mol Genet*. 2004;13(14):1525-1534. http://doi.org/10.1093/hmg/ddh165
28. Robinson JT, Thorvaldsdóttir H, Winckler W, et al. Integrative genomics viewer. *Nat Biotechnol*. 2011;29(1):24-26. http://doi.org/10.1038/nbt.1754
29. MacDonald JR, Ziman R, Yuen RKC, Feuk L, Scherer SW. The Database of Genomic Variants: a curated collection of structural variation in the human genome. *Nucleic Acids Res*. 2014;42(Database issue):D986-D992. http://doi.org/10.1093/nar/gkt958
30. Firth HV, Richards SM, Bevan AP, et al. DECIPHER: database of chromosomal imbalance and phenotype in humans using Ensembl resources. *Am J Hum Genet*. 2009;84(4):524-533. http://doi.org/10.1016/j.ajhg.2009.03.010
31. Rehm HL, Berg JS, Brooks LD, et al. ClinGen — the Clinical Genome Resource. *N Engl J Med*. 2015;372(23):2235-2242. http://doi.org/10.1056/NEJMsr1406261
32. Lappalainen I, Lopez J, Skipper L, et al. DbVar and DGVa: public archives for genomic structural variation. *Nucleic Acids Res*. 2013;41(Database issue):D936-D941. http://doi.org/10.1093/nar/gks1213