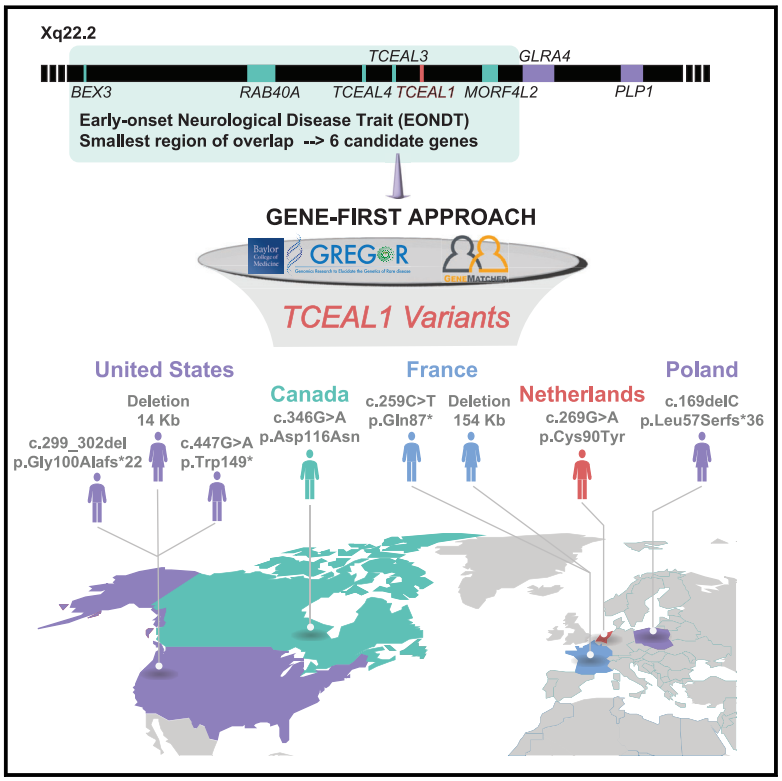
TCEAL1 loss-of-function results in an X-linked dominant neurodevelopmental syndrome and drives the neurological disease trait in Xq22.2 deletions

Graphical abstract



Authors

Hadia Hijazi, Linda M. Reis, Davut Pehlivan, ..., Jennifer E. Posey, Elena V. Semina, James R. Lupski

Correspondence

esemina@mcw.edu (E.V.S.), jlupski@bcm.edu (J.R.L.)

Applying a gene-first approach and worldwide gene-matching, we identify eight individuals with variants in *TCEAL1*, a candidate gene for the early-onset neurological disease trait (EONDT) in females with Xq22.2 deletion. The neurodevelopmental disorder observed overlaps that described in Xq22.2 deletion females, implicating *TCEAL1* as the driver gene.

Hijazi et al., 2022, The American Journal of Human Genetics *109*, 2270–2282
December 1, 2022
https://doi.org/10.1016/j.ajhg.2022.10.007

CellPress

TCEAL1 loss-of-function results in an X-linked dominant neurodevelopmental syndrome and drives the neurological disease trait in Xq22.2 deletions

Hadia Hijazi,[1,25,30] Linda M. Reis,[2,30] Davut Pehlivan,[1,3,4,5] Jonathan A. Bernstein,[6] Michael Muriello,[2] Erin Syverson,[2] Devon Bonner,[6] Mehrdad A. Estiar,[7,8] Ziv Gan-Or,[7,8,9] Guy A. Rouleau,[7,8,9] Ekaterina Lyulcheva,[10] Lynn Greenhalgh,[10] Marine Tessarech,[11,12] Estelle Colin,[11,12] Agnès Guichet,[11,12] Dominique Bonneau,[11,12] R.H. van Jaarsveld,[13] A.M.A. Lachmeijer,[13] Lyse Ruaud,[14,15] Jonathan Levy,[15] Anne-Claude Tabet,[15] Rafal Ploski,[16] Małgorzata Rydzanicz,[16] Łukasz Kępczyński,[17] Katarzyna Połatyńska,[18] Yidan Li,[1] Jawid M. Fatih,[1] Dana Marafi,[1,26] Jill A. Rosenfeld,[1,19] Zeynep Coban-Akdemir,[1,27] Weimin Bi,[1,19] Richard A. Gibbs,[1,20] Grace M. Hobson,[21] Jill V. Hunter,[22] Claudia M.B. Carvalho,[1,28] Jennifer E. Posey,[1,29] Elena V. Semina,[2,23,29,*] and James R. Lupski[1,4,20,24,29,*]

Summary

An Xq22.2 region upstream of *PLP1* has been proposed to underly a neurological disease trait when deleted in 46,XX females. Deletion mapping revealed that heterozygous deletions encompassing the smallest region of overlap (SRO) spanning six Xq22.2 genes (*BEX3*, *RAB40A*, *TCEAL4*, *TCEAL3*, *TCEAL1*, and *MORF4L2*) associate with an early-onset neurological disease trait (EONDT) consisting of hypotonia, intellectual disability, neurobehavioral abnormalities, and dysmorphic facial features. None of the genes within the SRO have been associated with monogenic disease in OMIM. Through local and international collaborations facilitated by GeneMatcher and Matchmaker Exchange, we have identified and herein report seven *de novo* variants involving *TCEAL1* in seven unrelated families: three hemizygous truncating alleles; one hemizygous missense allele; one heterozygous *TCEAL1* full gene deletion; one heterozygous contiguous deletion of *TCEAL1*, *TCEAL3*, and *TCEAL4*; and one heterozygous frameshift variant allele. Variants were identified through exome or genome sequencing with trio analysis or through chromosomal microarray. Comparison with previously reported Xq22 deletions encompassing *TCEAL1* identified a more-defined syndrome consisting of hypotonia, abnormal gait, developmental delay/intellectual disability especially affecting expressive language, autistic-like behavior, and mildly dysmorphic facial features. Additional features include strabismus, refractive errors, variable nystagmus, gastroesophageal reflux, constipation, dysmotility, recurrent infections, seizures, and structural brain anomalies. An additional maternally inherited hemizygous missense allele of uncertain significance was identified in a male with hypertonia and spasticity without syndromic features. These data provide evidence that *TCEAL1* loss of function causes a neurological rare disease trait involving significant neurological impairment with features overlapping the EONDT phenotype in females with the Xq22 deletion.

The Xq22.2 sub-band is ~1.2 Mb in size and contains 19 annotated genes of which one, proteolipid protein 1 (*PLP1* [MIM: 300401]), is associated with two allelic Mendelian neurological disease traits, Pelizaeus-Merzbacher disease (PMD [MIM: 312080]) and spastic paraplegia type 2 (SPG2 [MIM: 312920]).[1] No other monogenic

[1]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, USA; [2]Department of Pediatrics and Children's Research Institute, Medical College of Wisconsin and Children's Wisconsin, Milwaukee, WI, USA; [3]Section of Pediatric Neurology and Developmental Neuroscience, Department of Pediatrics, Baylor College of Medicine, Houston, TX, USA; [4]Department of Pediatrics, Baylor College of Medicine, Houston, TX, USA; [5]Jan and Dan Duncan Neurological Research Institute at Texas Children's Hospital, Houston, TX, USA; [6]Department of Pediatrics, Division of Medical Genetics, Stanford School of Medicine, Stanford, CA, USA; [7]Department of Human Genetics, McGill University, Montreal, QC, Canada; [8]The Neuro (Montreal Neurological Institute-Hospital), McGill University, Montreal, QC, Canada; [9]Department of Neurology & Neurosurgery, McGill University, Montreal, QC, Canada; [10]Liverpool Centre for Genomic Medicine, Liverpool Women's Hospital, Liverpool, UK; [11]Department of Medical Genetics, Angers University Hospital, Angers, France; [12]Mitovasc Unit, UMR CNRS 6015-INSERM 1083, University of Angers, Angers, France; [13]Department of Genetics, University Medical Center Utrecht, Utrecht, the Netherlands; [14]INSERM UMR1141, Neurodiderot, University of Paris, 75019 Paris, France; [15]APHP.Nord, Robert Debré University Hospital, Department of Genetics, 75019 Paris, France; [16]Department of Medical Genetics, Medical University of Warsaw, Warsaw, Poland; [17]Department of Genetics, Polish Mother's Memorial Hospital – Research Institute, Łódź, Poland; [18]Department of Developmental Neurology an Epileptology, Polish Mother's Memorial Hospital – Research Institute, Łódź, Poland; [19]Baylor Genetics, Houston, TX, USA; [20]Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX, USA; [21]Department of Research, Nemours Children's Health, Wilmington, DE, USA; [22]E.B. Singleton Department of Pediatric Radiology, Texas Children's Hospital, Houston, TX, USA; [23]Departments of Ophthalmology and Visual Sciences and Cell Biology, Neurobiology and Anatomy, Medical College of Wisconsin, Milwaukee, WI, USA; [24]Texas Children's Hospital, Houston, TX, USA
[25]Present address: Division of Genomic Diagnostics, Department of Pathology and Laboratory Medicine, Children's Hospital of Philadelphia, Philadelphia, PA, USA
[26]Present address: Department of Pediatrics, Faculty of Medicine, Kuwait University, P.O. Box 24923, 13110 Safat, Kuwait
[27]Present address: Human Genetics Center, Department of Epidemiology, Human Genetics, and Environmental Sciences, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX, USA
[28]Present address: Pacific Northwest Research Institute, Seattle, WA, USA
[29]Senior authors
[30]These authors contributed equally
*Correspondence: esemina@mcw.edu (E.V.S.), jlupski@bcm.edu (J.R.L.)
https://doi.org/10.1016/j.ajhg.2022.10.007.

Table 1. Clinical phenotypes of individuals with *TCEAL1* variants

	Individual 1	Individual 2	Individual 3	Individual 4	Individual 5	Individual 6	Individual 7	Individual 8
Sex	**male**	**male**	**male**	**male**	**female**	**female**	**female**	**male**
Ancestry	White (US)	White (European)	White (European)	White (European)	White/Chinese	Algerian	White (European)	White (European)
Consanguinity	–	–	–	two AOH regions totaling 27 Mb by SNP array	–	–	–	–
Age	5 y/o	10 y/o	17 y/o	7 y/o	17 y/o	9 y/o	6 y/o	12 y/o
Variant type	nonsense	frameshift	nonsense	missense	CNV	CNV	frameshift	missense
Variant	c.447G>A (p.Trp149*)	c.299_302del (p.Gly100Alafs*22)	c.259C>T (p.Gln87*)	c.269G>A (p.Cys90Tyr)	∼14 kb DEL of *TCEAL1*	∼154 kb DEL of *TCEAL1*, *TCEAL3*, *TCEAL4*	c.169delC (p.Leu57Serfs*36)	c.346G>A (p.Asp116Asn)
In silico prediction	damaging	damaging	damaging	mixed	damaging	damaging	damaging	mixed
Zygosity	hemi	hemi	hemi	hemi	het	het	het	hemi
Coordinates (hg19)	ChrX: 102, 885,291	ChrX: 102,885,138	ChrX: 102,885,103[a]	ChrX: 102,885,113	ChrX: 102,879,326–102,893,312	ChrX: 102,774, 750–102,929,222	ChrX: 102, 885,012	ChrX: 102, 885,190
Inheritance	*de novo*	*de novo*	*de novo*	*de novo*	*de novo*	*de novo*	*de novo*	maternal
Birth history								
Prenatal complications	maternal HELLP	maternal PUPPP	none	none	–	–	–	–
Gestational age, delivery	37 weeks, C-section	41 weeks, C-section	term	40 weeks, NSVD	37 weeks, NSVD	40 weeks	38 weeks, C-section	term
Postnatal complications	–	+, hyperbilirubinemia	–	difficulty feeding (short frenulum)	+, hyperbilirubinemia	–	–	–
Neurological anomalies								
DD/ID (HP: 0012758/ HP: 0001249)	++, moderate	+++, severe	+++, severe	++, moderate to severe	+, mild to moderate	+++, severe	++, mild	–
Neurodevelopmental regression (HP: 0002376)	–	+, lost single words	–	–	+	–	–	–
Seizures (HP: 0001250)	–	+ (developed in early childhood)	–	+ (one absence with apnea)	–	–	+	–
Verbal skills	single words only, communication device and signs	currently non-verbal	non-verbal	single words only, max 30 words	single words, short sentences	single words only, ∼10 words	single words, short sentences	full sentences and normal vocabulary for age

(*Continued on next page*)

Table 1. ***Continued***

	Individual 1	Individual 2	Individual 3	Individual 4	Individual 5	Individual 6	Individual 7	Individual 8
Sex	**male**	**male**	**male**	**male**	**female**	**female**	**female**	**male**
Abnormal muscle tone (HP: 0003808)	hypotonia (HP: 0001319)	hypotonia (HP: 0001319)	hypotonia (HP: 0001319)	hypotonia (HP: 0001319)	N/A	hypotonia (HP: 0001319)	hypotonia (HP: 0001319), lower leg spasticity	hypertonia (HP: 0001276), spasticity (HP: 0001257), ankle clonus
Gait disturbance (HP: 0001288)	+	+, non-ambulatory	+, ataxic with support	+	+	−	+	+, toe walking
Behavioral abnormalities (HP: 0000708)	+	+	+	+	+	+	+	−
Autism/autistic-like behavior (HP: 0000717)	+	+	+	−	+	−	−	−
Abnormal myelination (HP: 0012447)	−	−	N/A	+	+	−	+	−
Structural brain anomalies	possible mild foreshortening of corpus callosum	−	N/A	delayed myelinization of terminal zones of lateral ventricles, slight diminishing of white matter parietooccipital	abnormal myelination for age (HP: 0012447)	−	−	−
Ocular anomalies								
Astigmatism (HP: 0000483)	+	+	+	−	+	+	+	−
Nystagmus (HP: 0000639)	+	−	−	−	−	−	−	−
Strabismus (HP: 0000486)	+	+	+	−	−	+	−	−
Myopia/hyperopia (HP: 0000545/HP: 0000540)	+, mild myopia	+, hyperopia	−	−	−	−	−	−
Iris coloboma (HP: 0000612)	−	−	−	+	−	−	−	−
Dysmorphic features								
Broad forehead (HP: 0000337)	+	+	+	+	−	+	−	−
Other dysmorphic features	+	+	deep-set eyes, very bright blue eyes	−	+	+	+	−

(*Continued on next page*)

Table 1. *Continued*

Sex	Individual 1 male	Individual 2 male	Individual 3 male	Individual 4 male	Individual 5 female	Individual 6 female	Individual 7 female	Individual 8 male
Other								
Gastrointestinal abnormality (HP: 0011024)	–	+, GERD, constipation, G-tube	+, chewing difficulty, constipation	+, constipation	N/A	+, regurgitation	–	–
Abnormality of the immune system (HP: 0002715)	+, recurrent ear infections	+, recurrent chest infections (aspiration)	+, recurrent ear infections	+, recurrent respiratory and ear infections	+, oral allergy syndrome, recurrent infection	–	–	–
Other findings	–	–	growth retardation (onset 5 y/o), no puberty onset, hyperlaxity	not toilet trained at age 7, hypermobile fingers	hypertriglyceridemia, microcytic anemia	–	premature puberty	urinary incontinence

Abbreviations (alphabetical order) and symbols: AOH, absence of heterozygosity; CNV, copy-number variant; DD, developmental delay; DEL, deletion; HELLP, hemolysis, elevated liver enzymes, and low platelets; ID, intellectual disability; N/A, not available; NSVD, normal spontaneous vaginal delivery; PUPPP, pruritic urticarial papules and plaques of pregnancy; SNV, single-nucleotide variant; y/o, years old. GenBank: NM_001006639.2 is used for variant nomenclature; *in silico*: SIFT/PolyPhen (damaging, tolerated, or mixed predictions).
[a]Converted from hg38 coordinates (chrX: 103,630,175).

associations with Mendelian diseases have been reported in OMIM for genes mapped to the Xq22.2 interval. However, we and others have described the association of heterozygous intergenic Xq22.2 deletions that encompass the smallest region of overlap (SRO) containing six contiguous genes (*BEX3*, *RAB40A*, *TCEAL4*, *TCEAL3*, *TCEAL1*, and *MORF4L2*) with the emerging early-onset neurological disease trait (EONDT) in 46,XX females, consisting of hypotonia at birth, severe intellectual disability, neurobehavioral abnormalities, and mildly dysmorphic facial features.[2,3] Two of the six genes, *TCEAL1* and *MORF4L2*, have been specifically prioritized in previous studies as potential candidates because of their inclusion in a smaller Xq22.2 deletion that was reported in a female with an EONDT-like phenotype.[4] To date, no individuals with damaging variants impacting *MORF4L2* alone have been reported.

TCEAL1 (transcription elongation factor A-like 1 [MIM: 300237]) is a single coding-exon gene that encodes a 21-kDa nuclear phosphoprotein, referred to as TCEAL1 (or p21/SIIR). The encoded protein is related to the S-II class of transcription elongation factors as it consists of three main functional domains along its length of 157 amino acids (aa), an arginine/serine (RS) domain, a zinc-finger-like (ZnF-L) domain, and a helix-turn-helix (HTH) domain, and has a predicted RNA polymerase II binding site.[5–7] Previous knockout functional studies of the different domains of TCEAL1 in the context of the Rous sarcoma virus were conducted in transfected COS-1 cells where it was shown that loss of function (LoF) of the C-terminal domain of TCEAL1, RS, and the middle domain, ZnF-L, lead to down-regulation of the promoter activity of the virus; whereas LoF of the N-terminal domain, HTH, had minimal effects on the promoter activity of the virus.[6] These data suggest that TCEAL1, particularly the RS and ZnF-L domains, may play a role in transcriptional regulation.[6]

A total of four males and three females were identified with *de novo* disruption of *TCEAL1* (Table 1 and supplemental note, individuals 1–7) along with an inherited variant of uncertain significance in an additional male (individual 8). Written informed consent for research studies and/or reporting of clinical features was obtained for all individuals, including photo publication if applicable; all study procedures were approved by a local research review board and adhered to the Declaration of Helsinki. The seven individuals with *de novo* variants presented with neurological anomalies, which overlap those observed in EONDT females with contiguous gene deletions that include *TCEAL1* and *PLP1*. In particular, both males and females with *de novo TCEAL1* variants, as well as EONDT females, presented with DD/ID, behavioral abnormalities including autism or autistic-like features, abnormal muscle tone, and gait disturbance (Table 1).[2–4] Strabismus, a thin corpus callosum, and delayed or hypomyelination were less common in the present cohort compared to EONDT females who have deletions that include both *TCEAL1* and *PLP1*.[2–4]

Figure 1. Clinical photographs
(A and B) Individual 1 at 5 years old showing mildly dysmorphic facial features including broad forehead, deep-set eyes, telecanthus, prominent bow-shaped upper lip, mildly coarse facial features, and mildly low-set ears.
(C and D) Individual 2 at 3 years old with mildly dysmorphic features including long palpebral fissures, deep-set eyes, prominent bow-shaped upper lip, and brachycephaly.
(E–G) Individual 3 at 7 (E) and 17 (F and G) years old with similarly mild dysmorphic features including broad forehead, deep-set eyes, and bow-shaped upper lip.
(H) Individual 4 demonstrating a broad forehead, deep-set eyes, and bow-shaped upper lip.
(I and J) Individual 6 showing frontal bossing, bilateral epicanthus, hypertelorism, deep-set eyes, horizontal eyebrows, and fleshy earlobes.
(K and L) Individual 7 demonstrating a broad forehead, telecanthus, low-set ears, and widely spaced teeth.

In males, DD/ID ranged from moderate to severe with particular weakness in expressive language; the female individuals in this study showed similar but perhaps somewhat milder impairment. Individuals were identified independently through trio-exome or genome sequencing or review of the Baylor Hopkins Center for Mendelian Genomics (BHCMG) research exome sequencing (rES) database[8] (~15,000 exomes), and the Baylor Genetics (BG) diagnostic laboratory (clinical ES, cES, ~15,000 exomes) and CMA databases (>75,000 personal genomes) and connected through GeneMatcher[9,10] and the Matchmaker Exchange.[11,12] Careful review identified an overlapping phenotype of developmental delay/intellectual disability (DD/ID) including affected expressive language (7/7, 100%), neurobehavioral abnormalities (7/7, 100%) including autism or autistic-like behavior (4/7, 57.1%), and dysmorphic craniofacial features (7/7, 100%) that include a broad forehead, deep-set eyes, telecanthus, a prominent bow-shaped upper lip, slightly low-set ears, mild coarsening of facies, and brachycephaly (Figure 1). Individuals also demonstrated hypotonia (6/6, 100%), motor stereotypies (5/5, 100%), and abnormal gait or non-ambulatory status (6/7, 85.7%). Abnormal myelination or structural brain anomalies were observed in 3 of 6 individuals (50.0%, Figures 2A and 2B). Three individuals (3/7, 42.9%) reported seizures (Figure 2C). Additional affected

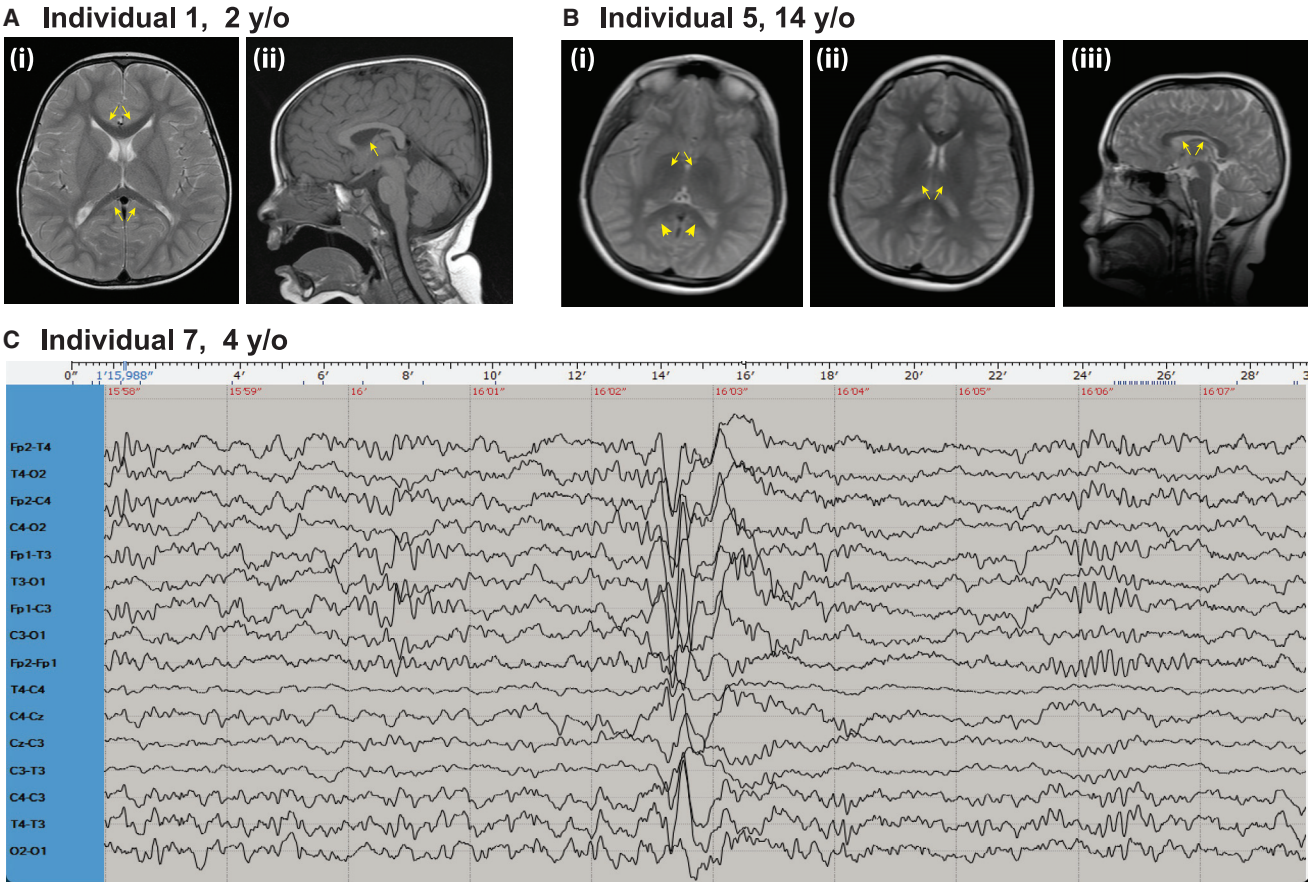


Figure 2. Brain MRIs and EEG

(A) Individual 1 brain MRI (i) T1 sequence axial view showing mildly reduced corpus callosum (CC) length (arrows) measuring 50.7 mm (between −1.0 and −2.0 SD). The CC thickness appears appropriate for age. (ii) T2 sequence midsagittal view showing normal myelination of the internal capsule (arrow).

(B) Individual 5 brain MRI (i) rapid sequence axial image showing lack of myelination of the internal capsule (arrows) and a bilateral T2 hyperintensity at the posterior limb of the internal capsule suggestive of gliosis (arrowheads). (ii) Rapid sequence axial image showing lack of myelination of the internal capsule (arrows). (iii) Rapid sequence sagittal image showing borderline to low-normal corpus callosum thickness (arrows).

(C) Individual 7 electroencephalogram (EEG) performed at 4 years of age, showing focal paroxysmal activity in the left frontocentrotemporal area with sharp waves and slow waves. y/o, years old.

organ systems and features (Table S1) present in more than half of individuals included ocular anomalies (astigmatism, nystagmus, strabismus, myopia or hyperopia, iris coloboma; 6/7, 85.7%), gastrointestinal abnormalities (gastroesophageal reflux disease [GERD], constipation, or regurgitation; 4/7, 57.1%), and recurrent infections (ear, respiratory; 5/7, 71.4%).

In contrast, individual 8 (Table 1, Table S1, supplemental note), a male with a maternally inherited missense variant (c.346G>A [p.Asp116Asn] [GenBank: NM_001006639.2]), demonstrated a distinct phenotype characterized by spasticity with hypertonia, hyperreflexia, bilateral ankle clonus, and a toe-walking gait in the absence of any developmental delay, intellectual disability, or dysmorphic craniofacial features. Notably, his mother, who is heterozygous for the same *TCEAL1* missense variant, was unaffected—a finding that is distinct from the observation of heterozygous LoF variants leading to disease trait expression observed in female individuals 5, 6, and 7 in the present cohort. The observation of a distinct neurological phenotype in individual 8, in combination with the unaffected status of this individual's mother who is heterozygous for the variant, raises the possibility that an alternative molecular mechanism (i.e., LoF hypomorphic or null allele; or antimorphic in carrier females versus gain of function [GoF hypermorphic] or novel function [neomorphic allele]) may be responsible for his unique phenotype compared to individuals 1–7. It is also possible that this variant is a rare benign allele with a CADD score = 24 or the personal genome of individual 8 has another unrecognized variant mapping at this or another locus that contributes to the phenotype. Detailed clinical descriptions of all eight individuals are provided as supplemental text.

TCEAL1 variants were identified by exome sequencing (ES), genome sequencing (GS), array comparative genomic hybridization (aCGH), and/or single-nucleotide polymorphism (SNP)-array (Table S2, see supplemental information for detailed methods). Six *de novo* putative LoF variants

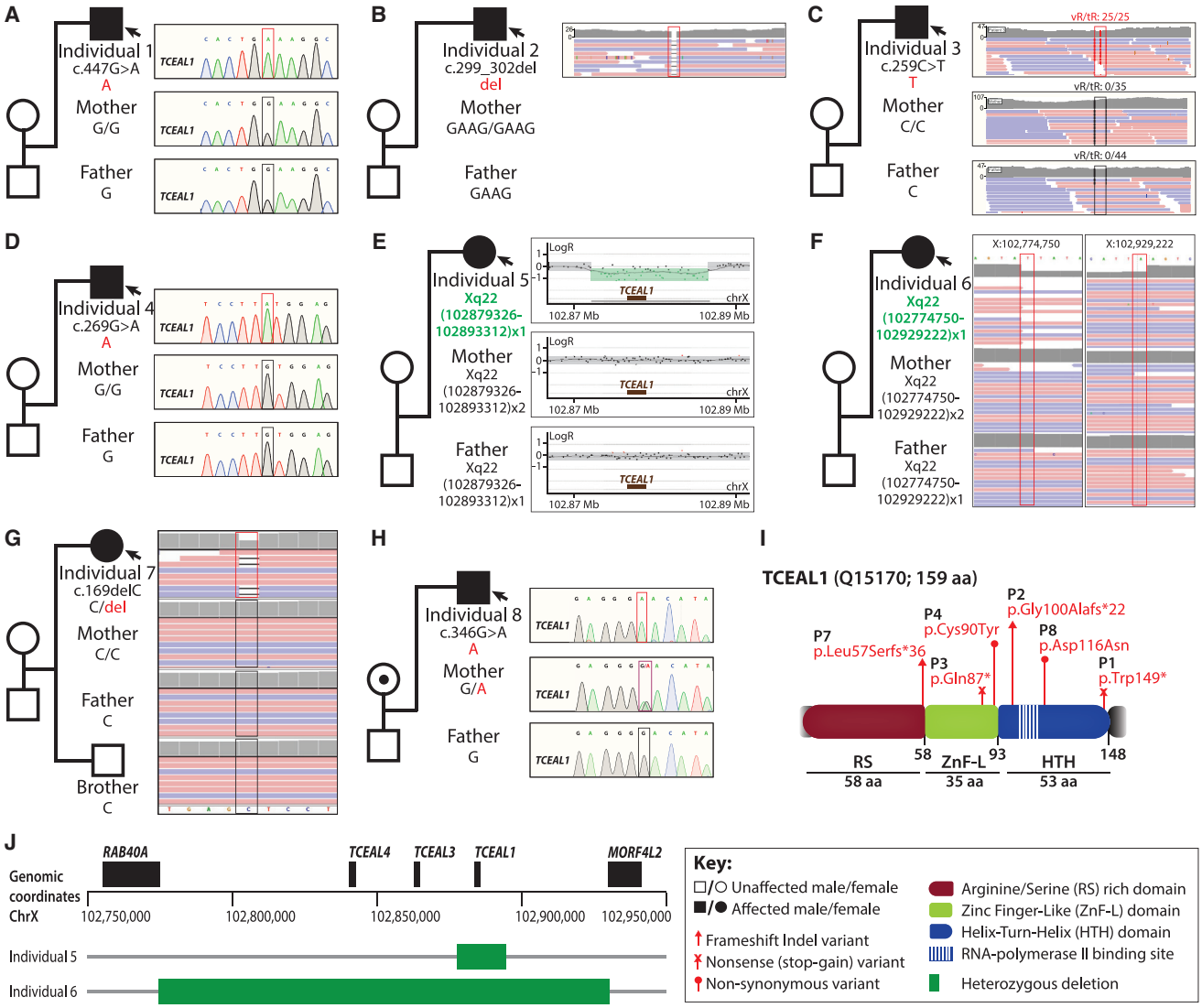


Figure 3. Description of *TCEAL1* variants and segregation within families

(A) Individual 1 with a *de novo* hemizygous nonsense variant that maps to the HTH domain.
(B) Individual 2 with a *de novo* hemizygous frameshift variant that maps to the HTH domain.
(C) Individual 3 with a *de novo* hemizygous nonsense variant that maps to the ZnF-L domain.
(D) Individual 4 with a *de novo* hemizygous missense variant that maps to the ZnF-L domain.
(E) HD-aCGH on individual 5 and parents revealing an ~14 kb *de novo* heterozygous deletion that encompasses the entirety of *TCEAL1*.
(F) Individual 6 with an ~154 kb *de novo* heterozygous deletion that encompasses *TCEAL1*, *TCEAL3*, and *TCEAL4*.
(G) Individual 7 with a *de novo* heterozygous frameshift variant that maps to the RS domain.
(H) Individual 8 with a maternally inherited hemizygous missense variant that maps to the HTH domain.
(I) Three domains are indicated: the arginine/serine (RS) rich domain (red), the zinc finger-like (ZnF-L) domain (green), and the helix-turn-helix (HTH) domain (blue). The RNA polymerase II binding site is indicated by vertical stripes within the HTH domain. Described missense, nonsense, and frameshift variants are mapped to the linear protein structure of TCEAL1. Q15170 is the UniProt accession ID used.
(J) Location of deletions identified in individuals 5 and 6 within the Xq22.2 region.
Abbreviations: tR, total read depth; vR, variant read depth. GenBank: NM_001006639.2 is used for variant nomenclature.

(four truncating and two deletions) were identified in *TCEAL1* (GenBank: NM_001006639.2); the gene contains three exons, of which only the third exon is coding (Figure 3). The truncating variants were two hemizygous nonsense variants (c.447G>A [p.Trp149*] [GenBank: NM_001006639.2] and c.259C>T [p.Gln87*] [GenBank: NM_001006639.2]) and two frameshift variants (hemizygous c.299_302del [p.Gly100Alafs*22] [GenBank: NM_001006639.2] and heterozygous c.169delC [p.Leu57Serfs*36] [GenBank: NM_001006639.2]) in individuals 1, 3, 2, and 7, respectively. These truncating variants occurred at the C-terminal end of the RS domain, within the ZnF-L and HTH domains, or just outside the HTH domain (Figure 3I). These variants may lead to complete LoF of TCEAL1 in males (individuals 1–3). However, because *TCEAL1* is a single-coding-exon gene, it may escape nonsense-mediated decay (NMD).[13] It is possible

that these truncating variants with PTCs (premature termination codons) might lead to proteins with partial function or even a GoF involving loss of the HTH domain of TCEAL1. Little is known about the function of this domain in TCEAL1, although HTH domains generally act as DNA-binding sites that regulate transcription.[14]

While the numbers of individuals are too small to draw definitive conclusions, there did appear to be a correlation between the location of the truncating intragenic variant and the severity of the neurological phenotype in males, i.e., a "polarity effect" (see Figure 1 of Inoue et al.[15]) as has been observed at the *SOX10* locus with peripheral demyelinating neuropathy, central dysmyelination, Waardenburg syndrome, and Hirschsprung disease (PCWH [MIM: 609136]). Individuals 2 and 3 with truncating variants within the ZnF-L or early in the HTH domain both had severe/profound cognitive impairment, no independent ambulation or meaningful words, and gastrointestinal abnormalities, while individual 1, with truncation just after the HTH domain, had moderate delay, abnormal but independent gait, single words, and normal gastrointestinal functioning. This observation is consistent with the contention that the mRNA will most likely escape nonsense-mediated decay (because the variants occur in the third and final exon, which is the only coding exon), so the longer proteins would be expected to be more functional. It also suggests that the C terminus of the protein is important in the function of TCEAL1, despite the lack of identified domains within this region. Further identified variants will be needed to determine whether this deduction of association remains true and suggests a focus for possible future functional experiments on the HTH domain of TCEAL1.

In contrast to the male subjects, two of the three females reported herein had *de novo* heterozygous *TCEAL1* deletions and one had a frameshift variant allele located N-terminal to the termination codons identified in male individuals. Overall, these three females had a somewhat milder clinical presentation than the male individuals despite their heterozygous LoF variants, supporting a role for *TCEAL1* gene dosage[16,17] in the observed variable severity of disease.

Individual 5 was found to have a heterozygous 14 kb *TCEAL1* deletion. Clinically performed X chromosome inactivation (XCI) studies were indeterminate as the studied markers were not informative. No CNVs with boundaries comparable to the deletion identified in individual 5 were found in public structural variant databases, including gnomAD, DECIPHER, and the Database of Genomic Variants (DGV). While partial gene deletions have not been reported to date, this may be due to methodological limitations of CNV detection in the studied cohorts. Given its small single-coding-exon size (coding exon is ~900 bp), partial (i.e., intragenic) CNVs in *TCEAL1* are likely to be missed. The average probe coverage of clinical-grade CMA does not meet the validation threshold within the coding interval of *TCEAL1* (~900 bp); therefore, it is possible that affected individuals with intragenic *TCEAL1* deletions are missed due to technical limitations, a challenge often faced in studies of short single-coding-exon genes.[18] Given the frequency of gene deletion and inherent genomic instability of this region, future screening of *TCEAL1* in affected individuals that are optimized for both DNA sequencing and copy-number variant assessment are warranted.

Mapping of the 14 kb deletion in individual 5 via high density aCGH (HD-aCGH, Figure 3E) and subsequent junction-PCR and Sanger sequencing (Figure 4) identified the precise (nucleotide-level) coordinates as chrX: 102,879,326–102,893,312 (GRCh37/hg19) and a 77 bp homology defined by the presence of 77 bp of sequence shared by the distal and proximal breakpoints. The deletion was *de novo*, confirmed via HD-aCGH and trio junction PCR of the parent-proband trio (Figures 3E and 4). Repeat sequences, specifically highly similar intrachromosomal repeats (HSIRs), have been implicated in predisposing a 90 kb hotspot on Xq22.2 to genomic instability and the formation of potentially pathogenic deletions.[2] HSIRs are defined as intrachromosomal repeat sequences that are typically >700 bp in length with 95%–100% identity between pairs.[2] An HSIR that is ~140 kb in length, RepX-i1010, is of particular note as ~50% of proximal breakpoints of pathogenic Xq22.2 deletions are clustered within it.[2] Notably, *TCEAL1* is entirely embedded within RepX-i1010.

We therefore sought to examine the breakpoints of the *TCEAL1* deletion in individual 5 for the possibility of overlapping repeats. The breakpoints of the deletion were aligned to the haploid reference genome and examined for overlapping repeats in four different repeat datasets, (1) the (HSIR) dataset,[2] (2) the Segmental Dup dataset,[19,20] (3) the Repbase (Repeat Masker) dataset,[21] and (4) the SelfChain dataset.[22,23] Intriguingly, we found both breakpoints overlap a pair of SelfChain repeats (101408 and 101409), i.e., the junction forms a fusion SelfChain, and the deletion along with the SelfChain pair are entirely embedded within RepX-i1010. The fusion SelfChain resulting from the deletion suggests that homology of the two SelfChains may have stimulated genomic instability and deletion formation.[24] These findings emphasize the unique enrichment of RepX-i1010 for smaller repeat constituents and the role for repeats generally in predisposing Xq22.2 to the formation of pathogenic structural variants, specifically genomic deletions.

Individual 6 was found to have a heterozygous 154 kb deletion of *TCEAL1*, *TCEAL3*, and *TCEAL4*. No comparable deletions were identified in public structural variant databases, including gnomAD, DECIPHER, and the DGV. Mapping of the 154 kb deletion in individual 6, initially identified by SNP-array and genomic sequencing, confirmed the *de novo* status of the deletion and confirmed deletion coordinates of chrX: 102,774,750–102,929,222 marked by a 1 bp microhomology (GRCh37/hg19, Figure 5). XCI studies performed in a

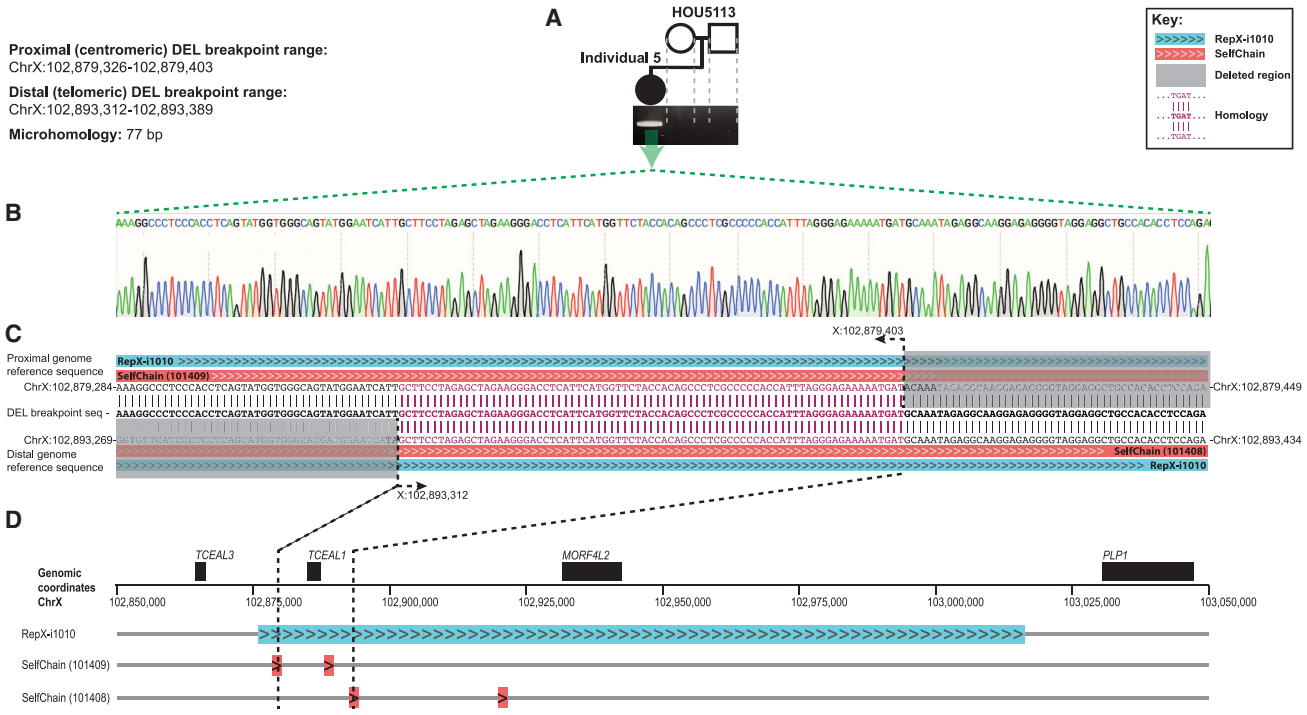


Figure 4. Demonstration of *de novo* status and breakpoint sequence of *TCEAL1* deletion and local genomic architecture at the *TCEAL1* locus in individual 5

(A) Junction PCR confirms the *de novo* nature of the *TCEAL1* deletion in individual 5.

(B) Nucleotide resolution of the deletion breakpoint-junction demonstrates a 77 bp homology.

(C) The human genome reference sequence (GRCh37/hg19) surrounding the proximal and distal breakpoint sites is provided (top and bottom), mapped to the sequence of the deletion breakpoint-junction (middle). Chromosome coordinates are provided for the start and end points of the displayed portions of the proximal and distal genome reference sequence, as well as for the nucleotide positions at the start and end of the 77 bp homology. Breakpoints map to SelfChain repeat pairs that are in direct orientation (indicated by ">>>" highlighted in red). Furthermore, the entire deletion and both SelfChain repeats are embedded within the larger repeat, RepX-i1010 (indicated by ">>>" highlighted in blue).

(D) The positions of RepX-i1010 and both SelfChain repeats within Xq22.2 are demonstrated. Black dotted lines illustrate the positions of the breakpoints in individual 5.

clinical diagnostic laboratory on individual 6 further indicated a skewed XCI of 85%/15% at the *AR* locus (data not shown), although the identity of the preferentially inactivated allele (reference allele or variant allele) was not determined through this testing.

Two additional hemizygous missense variants were identified including one *de novo* (individual 4, c.269G>A [p.Cys90Tyr] [GenBank: NM_001006639.2]) and one maternally inherited (individual 8, c.346G>A [p.Asp116Asn] [GenBank: NM_001006639.2]). Notably, the maternally inherited missense variant was identified in individual 8, who also demonstrated a phenotype distinctive from the remainder of the cohort. This variant was found to have a CADD phred score of 24.1 and was present in one heterozygous individual (one female out of 183,451 individuals) in gnomAD (MAF = 5.451×10^{-6}), suggesting the potential that this variant could be damaging in hemizygous individuals.

In mice, the knockout of *Tceal1* has been studied in female and male mice where affected offspring appeared to exhibit a neurological phenotype, involving behavioral abnormalities (in homozygous females), abnormal response to stimulus (in hemizygous males), and ophthalmologic anomalies (in homozygous females) (data sourced from the International Mouse Phenotyping Consortium; IMPC[25,26]). Overall, the reported phenotypes (neurological anomalies, autistic-like behavioral abnormalities, strabismus, and refractive vision abnormalities) in the present cohort show overlap with those reported in homozygous female and hemizygous male mice with *Tceal1* knockout.[26] Although phenotype data for heterozygous *Tceal1* knockout mice have not been reported, the neurological phenotypes described in hemizygous males and homozygous females support a potential role of *TCEAL1* in the human nervous system.

TCEAL1 was previously implicated as a possible contributor to the contiguous gene deletion syndromes caused by Xq22.2 intergenic microdeletions, but without clear evidence of its monogenic involvement.[2–4] Indeed, in DECIPHER, there are a total of 70 individuals with deletions or duplications encompassing the *TCEAL1* locus. Despite this large number of individuals with CNVs in this region, each of these CNVs either extends proximally beyond the *TCEAL* genes in this region to include *RAB40A* and/or distally to include *PLP1*, an established disease gene. Thus, these data alone are not informative to

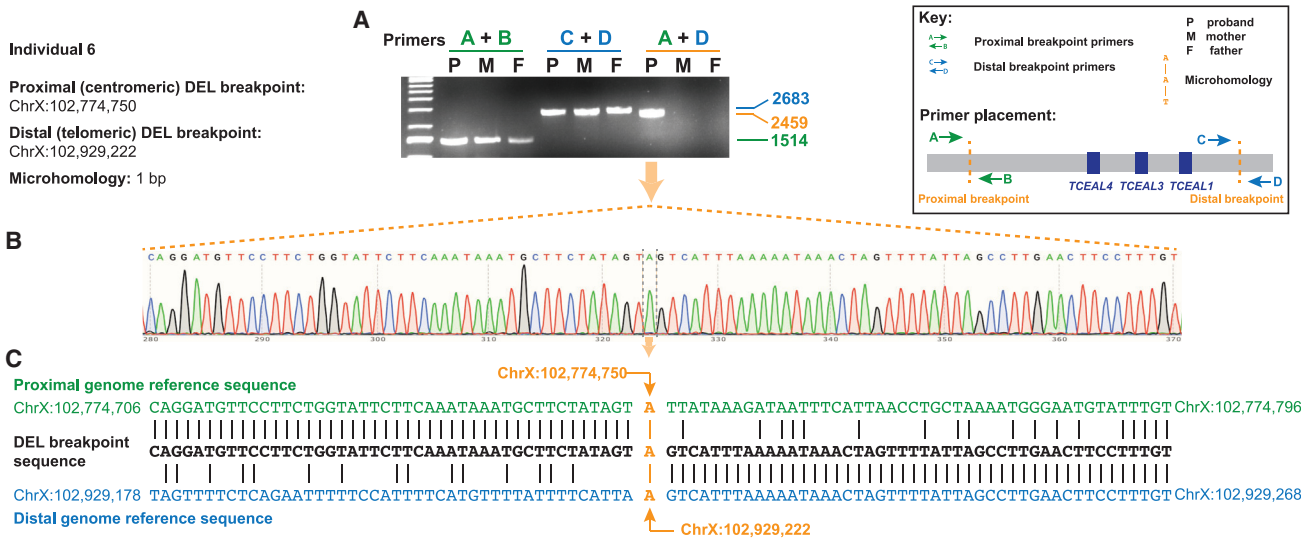


Figure 5. Demonstration of *de novo* deletion CNV and breakpoint sequencing in individual 6
(A) Junction PCR confirms the *de novo* nature of the deletion CNV in individual 6.
(B) Nucleotide resolution of the deletion breakpoint-junction demonstrates a 1 bp microhomology.
(C) The human genome reference sequence (GRCh37/hg19) surrounding the proximal and distal breakpoint sites is provided (top and bottom), mapped to the sequence of the deletion breakpoint-junction (middle). Chromosome coordinates are provided for the start and end points of the displayed portions of the proximal and distal genome reference sequence, as well as for the nucleotide position of the 1 bp microhomology.

conclude that *TCEAL1* is solely responsible for the phenotypes reported in these individuals in DECIPHER.

The present study identifies the role of *TCEAL1* in Mendelian disease through identification of *de novo TCEAL1* variants in seven unrelated individuals with neurological disease as well as one variant of uncertain significance (VUS) in a male individual harboring a maternally inherited variant allele. These data provide evidence that *TCEAL1* loss of function causes a rare disease trait involving significant neurological impairment with features overlapping the EONDT phenotype in females with the Xq22 deletion. Notably, the *TCEAL* gene family shares a common ancestral sequence with the *BEX* gene family, and their protein structures share a common C-terminal domain that can support homodimerization and heterodimerization of TCEAL and BEX proteins.[27,28] This shared C-terminal domain suggests the possibility that other members of the *TCEAL* and *BEX* gene families may be associated with rare human disease traits in the future. TCEAL1 is one of nine members of the TCEAL protein family (TCEAL1 through TCEAL9), and the genes for these proteins all map to the Xq22.1-Xq22.2 region. Despite the encoding genes' proximity, TCEAL1 has limited overall protein sequence identity with a majority of the TCEAL/BEX proteins, whereas TCEAL8 and TCEAL9 share the highest amino acid similarity (52.83%, 53.33% respectively; Table S3). Compared to other TCEAL protein family members encoded within the SRO, TCEAL1 shares 31.25% homology with TCEAL4 and no significant similarity with TCEAL3. This limited sequence similarity supports the possibility that TCEAL1 may not share redundant function with other TCEAL protein family members, supporting a role for haploinsufficiency and LoF in human neurological disease traits.

Importantly, *TCEAL1* is expected to be subject to the effect of XCI in females by virtue of its X-linked location. The process of XCI can greatly influence ChrX gene expression and X-linked disease penetrance in 46,XX females because of the random chance of silencing or expressing heterozygous mutant alleles.[29,30] Thus, the expression of pathogenic heterozygous *TCEAL1* variants in females is influenced not only by gene dosage but also expression of that gene dosage: the direction and influence of XCI (i.e., whether the mutated or the wild-type X chromosome is inactivated), specifically in disease-relevant organs and tissues. It is possible that unaffected females heterozygous for putatively damaging variants in *TCEAL1* may benefit from skewed XCI favoring expression of the wild-type allele. In contrast, affected females may demonstrate skewed XCI favoring expression of the damaging *TCEAL1* allele. XCI studies performed in a clinical diagnostic laboratory on individual 6 further indicated a skewed XCI of 85%/15% at the *AR* locus (data not shown), although the identity of the preferentially inactivated allele (reference allele or variant allele) was not determined through this testing. The detection of XCI skewing (85%/15%) at the *AR* locus in female individual 6 supports the possibility that both *TCEAL1* gene dosage and expression of that gene dosage, modulated by XCI, may influence the severity and expression of disease associated with *TCEAL1*.

The X chromosome, like its heterologous/heterogametic ChrY sex chromosome, is enriched with repeat sequences, specifically long interspersed nuclear elements (LINEs) that

facilitate spreading of XCI to all parts of the chromosome.[31,32] However, the enrichment for LINEs is a double-edged sword, as these repetitive sequences also render an increased susceptibility to the formation of copy-number variants (CNVs).[33–36] We and others have previously pointed to the prominence of repeat sequences in predisposing Xq22 to the formation of pathogenic CNVs,[2,37–39] of which perhaps the most significant to this study is the ~140 kb repeat, RepX-i1010, that was implicated in a 90 kb genomic instability hotspot within Xq22.2 where nearly half of all breakpoints of pathogenic Xq22 deletions cluster.[2] Here, we identified an additional individual with an Xq22.2 deletion with breakpoints that also map to RepX-i1010.

These data provide compelling evidence that loss of gene function, specifically of *TCEAL1*, contributes to the EONDT-like neurological disease traits in females with Xq22 deletions, potentially acting as a "driver gene" in the microdeletion syndrome. Comparison of clinical features of affected individuals suggests a more defined X-linked dominant rare disease trait consisting of neurological impairment, ocular and gastrointestinal anomalies, and mildly dysmorphic facial features and warrants further consideration of this gene in the molecular diagnosis of similarly affected individuals. Furthermore, the *TCEAL1* locus appears to be highly susceptible to the formation of genomic deletions by virtue of the overlapping and surrounding genomic architecture, specifically the RepX-i1010 HSIR, and thus may be a *de novo* structural variant mutagenesis hotspot.

Data and code availability

All *TCEAL1* variants reported herein have been deposited to ClinVar, accession IDs ClinVar: SCV002558856; ClinVar: SCV002558857; ClinVar: SCV002558858; ClinVar: SCV002558859; ClinVar: SCV 002558860; ClinVar: SCV002558861; ClinVar: SCV002558862; ClinVar: SCV002558863. For subjects who have provided written informed consent for sharing of their genomic data in controlled access databases, these data will be deposited to AnVIL and/or dbGaP under accession dbGAP: phs000711.v5.p1.

Supplemental information

Supplemental information can be found online at https://doi.org/10.1016/j.ajhg.2022.10.007.

Acknowledgments

We thank all individuals, their families, and the referring physicians who submitted samples for testing. No additional compensation was received for these contributions. The authors thank the contributors to MyGene2, GeneMatcher, and other Matchmaker Exchange databases and the Genome Aggregation Database (gnomAD) and the groups that provided exome and genome variant data to these resources. A full list of contributing groups to gnomAD can be found at https://gnomad.broadinstitute.org/about. This work was supported in part by the US National Human Genome Research Institute (NHGRI)/National Heart, Lung, and Blood Institute (NHLBI) grant UM1 HG006542 to the Baylor-Hopkins Center for Mendelian Genomics (BHCMG); US National Institute of Neurological Disorders and Stroke (NINDS) grants R01 NS058529 and R35 NS105078 and National Institute of General Medical Sciences (NIGMS) grant R01 GM106373 to J.R.L.; NHGRI U01 HG011758 to the Baylor College of Medicine Genomics Research Elucidate Genetics of Rare disease (BCM-GREGoR) consortium and grant K08 HG008986 to J.E.P.; by NIGMS T32 GM007526-42 to D.M.; and by the National Eye Institute grants R01EY025718 and EY015518 to E.V.S. and 1UL1RR031973 from the Clinical and Translational Science Award (CTSA) program. D.P. is supported by International Rett Syndrome Foundation (IRSF grant #3701-1). This study makes use of data generated by the DECIPHER[40] community and the Deciphering Developmental Disorders (DDD) study. Please refer to the supplemental acknowledgments for full acknowledgment and details.

Declaration of interests

J.R.L. serves on the Scientific Advisory Board of Baylor Genetics (BG); J.A.R. and W.B. report affiliation with BG. Baylor College of Medicine (BCM) and Miraca Holdings have formed a joint venture with shared ownership and governance of Baylor Genetics (BG), which performs clinical microarray analysis (CMA) and clinical exome sequencing (cES) and molecular diagnostic whole-genome sequencing (WGS). J.R.L. has stock ownership in 23andMe, is a paid consultant for the Regeneron Genetics Center, and is a co-inventor on multiple United States and European patents related to molecular diagnostics for inherited neuropathies, eye diseases, genomic disorders, and bacterial genomic fingerprinting. H.H., D.P., Y.L., J.M.F., D.M., J.A.R., Z.H.C.A., W.B., R.A.G., C.M.B.C., J.E.P., and J.R.L. report affiliation with the Department of Molecular and Human Genetics at Baylor College of Medicine. The Department of Molecular and Human Genetics at Baylor College of Medicine derives revenue from molecular genetic and personal genome (CMA, cES, WGS) genomic testing offered in BG. Z.G.O. serves on the scientific advisory boards and receives consultancy fees from Bial Biotech Inc. and Handl Therapeutics. He has received consultancy fees from Neuron23, Ono Therapeutics, and UCB.

Received: February 22, 2022
Accepted: October 13, 2022
Published: November 10, 2022

Web resources

CADD – Combined Annotation Dependent Depletion, https://cadd.gs.washington.edu/snv
DECIPHER, https://decipher.sanger.ac.uk
DGV – Database of Genomic Variants, http://dgv.tcag.ca/
GeneMatcher, https://genematcher.org/
gnomAD, https://gnomad.broadinstitute.org/
IGV, http://software.broadinstitute.org/software/igv/
Matchmaker Exchange, https://matchmakerexchange.org/
MyGene2, https://mygene2.org
Online Mendelian Inheritance in Man, https://www.omim.org/
UCSC Genome Browser, https://genome.ucsc.edu/

References

1. Lupski, J.R. (2022). Biology in balance: human diploid genome integrity, gene dosage, and genomic medicine. Trends Genet. *38*, 554–571.

2. Hijazi, H., Coelho, F.S., Gonzaga-Jauregui, C., Bernardini, L., Mar, S.S., Manning, M.A., Hanson-Kahn, A., Naidu, S., Srivastava, S., Lee, J.A., et al. (2020). Xq22 deletions and correlation with distinct neurological disease traits in females: further evidence for a contiguous gene syndrome. Hum. Mutat. *41*, 150–168.
3. Yamamoto, T., Wilsdon, A., Joss, S., Isidor, B., Erlandsson, A., Suri, M., Sangu, N., Shimada, S., Shimojima, K., Le Caignec, C., et al. (2014). An emerging phenotype of Xq22 microdeletions in females with severe intellectual disability, hypotonia and behavioral abnormalities. J. Hum. Genet. *59*, 300–306.
4. Labonne, J.D.J., Graves, T.D., Shen, Y., Jones, J.R., Kong, I.K., Layman, L.C., and Kim, H.G. (2016). A microdeletion at Xq22.2 implicates a glycine receptor *GLRA4* involved in intellectual disability, behavioral problems and craniofacial anomalies. BMC Neurol. *16*, 132.
5. Pillutla, R.C., Shimamoto, A., Furuichi, Y., and Shatkin, A.J. (1999). Genomic structure and chromosomal localization of *TCEAL1*, a human gene encoding the nuclear phosphoprotein p21. Genomics *56*, 217–220.
6. Yeh, C.H., and Shatkin, A.J. (1994). Down-regulation of Rous sarcoma virus long terminal repeat promoter activity by a HeLa cell basic protein. Proc. Natl. Acad. Sci. USA *91*, 11002–11006.
7. Yeh, C.H., and Shatkin, A.J. (1994). A HeLa-cell-encoded p21 is homologous to transcription elongation factor SII. Gene *143*, 285–287.
8. Posey, J.E., O'Donnell-Luria, A.H., Chong, J.X., Harel, T., Jhangiani, S.N., Coban Akdemir, Z.H., Buyske, S., Pehlivan, D., Carvalho, C.M.B., Baxter, S., et al. (2019). Insights into genetics, human biology and disease gleaned from family based genomic studies. Genet. Med. *21*, 798–812.
9. Sobreira, N., Schiettecatte, F., Valle, D., and Hamosh, A. (2015). GeneMatcher: a matching tool for connecting investigators with an interest in the same gene. Hum. Mutat. *36*, 928–930.
10. Wohler, E., Martin, R., Griffith, S., Rodrigues, E.D.S., Antonescu, C., Posey, J.E., Coban-Akdemir, Z., Jhangiani, S.N., Doheny, K.F., Lupski, J.R., et al. (2021). PhenoDB, GeneMatcher and VariantMatcher, tools for analysis and sharing of sequence data. Orphanet J. Rare Dis. *16*, 365.
11. Sobreira, N.L.M., Arachchi, H., Buske, O.J., Chong, J.X., Hutton, B., Foreman, J., Schiettecatte, F., Groza, T., Jacobsen, J.O., Haendel, M.A., et al. (2017). Matchmaker Exchange. Curr. Protoc. Hum. Genet. *95*.
12. Philippakis, A.A., Azzariti, D.R., Beltran, S., Brookes, A.J., Brownstein, C.A., Brudno, M., Brunner, H.G., Buske, O.J., Carey, K., Doll, C., et al. (2015). The matchmaker exchange: a platform for rare disease gene discovery. Hum. Mutat. *36*, 915–921.
13. Cusack, B.P., Arndt, P.F., Duret, L., and Roest Crollius, H. (2011). Preventing dangerous nonsense: selection for robustness to transcriptional error in human genes. PLoS Genet. *7*, e1002276.
14. Aravind, L., Anantharaman, V., Balaji, S., Babu, M.M., and Iyer, L.M. (2005). The many faces of the helix-turn-helix domain: transcription regulation and beyond. FEMS Microbiol. Rev. *29*, 231–262.
15. Inoue, K., Khajavi, M., Ohyama, T., Hirabayashi, S.i., Wilson, J., Reggin, J.D., Mancias, P., Butler, I.J., Wilkinson, M.F., Wegner, M., and Lupski, J.R. (2004). Molecular mechanism for distinct neurological phenotypes conveyed by allelic truncating mutations. Nat. Genet. *36*, 361–369.
16. Lupski, J.R. (2015). Structural variation mutagenesis of the human genome: Impact on disease and evolution. Environ. Mol. Mutagen. *56*, 419–436.
17. Ricard, G., Molina, J., Chrast, J., Gu, W., Gheldof, N., Pradervand, S., Schütz, F., Young, J.I., Lupski, J.R., Reymond, A., and Walz, K. (2010). Phenotypic consequences of copy number variation: insights from Smith-Magenis and Potocki-Lupski syndrome mouse models. PLoS Biol. *8*, e1000543.
18. Boone, P.M., Bacino, C.A., Shaw, C.A., Eng, P.A., Hixson, P.M., Pursley, A.N., Kang, S.H.L., Yang, Y., Wiszniewska, J., Nowakowska, B.A., et al. (2010). Detection of clinically relevant exonic copy-number changes by array CGH. Hum. Mutat. *31*, 1326–1342.
19. Bailey, J.A., Gu, Z., Clark, R.A., Reinert, K., Samonte, R.V., Schwartz, S., Adams, M.D., Myers, E.W., Li, P.W., and Eichler, E.E. (2002). Recent segmental duplications in the human genome. Science *297*, 1003–1007.
20. Bailey, J.A., Yavor, A.M., Viggiano, L., Misceo, D., Horvath, J.E., Archidiacono, N., Schwartz, S., Rocchi, M., and Eichler, E.E. (2002). Human-specific duplication and mosaic transcripts: the recent paralogous structure of chromosome 22. Am. J. Hum. Genet. *70*, 83–100.
21. Jurka, J. (2000). Repbase update: a database and an electronic journal of repetitive elements. Trends Genet. *16*, 418–420.
22. Chiaromonte, F., Yap, V.B., and Miller, W. (2002). Scoring pairwise genomic sequence alignments. Pac. Symp. Biocomput., 115–126.
23. Schwartz, S., Kent, W.J., Smit, A., Zhang, Z., Baertsch, R., Hardison, R.C., Haussler, D., and Miller, W. (2003). Human-mouse alignments with BLASTZ. Genome Res. *13*, 103–107.
24. Zhou, W., Zhang, F., Chen, X., Shen, Y., Lupski, J.R., and Jin, L. (2013). Increased genome instability in human DNA segments with self-chains: homology-induced structural variations via replicative mechanisms. Hum. Mol. Genet. *22*, 2642–2651.
25. IMPC (2020). International mouse phenotyping consortium, *Tceal1* knockout mice. https://www.mousephenotype.org/data/genes/MGI:2385317.
26. Dickinson, M.E., Flenniken, A.M., Ji, X., Teboul, L., Wong, M.D., White, J.K., Meehan, T.F., Weninger, W.J., Westerberg, H., Adissu, H., et al. (2016). High-throughput discovery of novel developmental phenotypes. Nature *537*, 508–514.
27. Manford, A.G., Mena, E.L., Shih, K.Y., Gee, C.L., McMinimy, R., Martínez-González, B., Sherriff, R., Lew, B., Zoltek, M., Rodríguez-Pérez, F., et al. (2021). Structural basis and regulation of the reductive stress response. Cell *184*, 5375–5390.e16.
28. Navas-Pérez, E., Vicente-García, C., Mirra, S., Burguera, D., Fernàndez-Castillo, N., Ferrán, J.L., López-Mayorga, M., Alaiz-Noya, M., Suárez-Pereira, I., Antón-Galindo, E., et al. (2020). Characterization of an eutherian gene cluster generated after transposon domestication identifies *Bex3* as relevant for advanced neurological functions. Genome Biol. *21*, 267.
29. Carrel, L., and Willard, H.F. (2005). X-inactivation profile reveals extensive variability in X-linked gene expression in females. Nature *434*, 400–404.
30. Lupski, J.R., Garcia, C.A., Zoghbi, H.Y., Hoffman, E.P., and Fenwick, R.G. (1991). Discordance of muscular dystrophy in monozygotic female twins: evidence supporting

asymmetric splitting of the inner cell mass in a manifesting carrier of Duchenne dystrophy. Am. J. Med. Genet. *40*, 354–364.

31. Chow, J.C., Ciaudo, C., Fazzari, M.J., Mise, N., Servant, N., Glass, J.L., Attreed, M., Avner, P., Wutz, A., Barillot, E., et al. (2010). LINE-1 activity in facultative heterochromatin formation during X chromosome inactivation. Cell *141*, 956–969.
32. Lyon, M.F. (2003). The Lyon and the LINE hypothesis. Semin. Cell Dev. Biol. *14*, 313–318.
33. Stankiewicz, P., and Lupski, J.R. (2006). The genomic basis of disease, mechanisms and assays for genomic disorders. Genome Dyn. *1*, 1–16.
34. Startek, M., Szafranski, P., Gambin, T., Campbell, I.M., Hixson, P., Shaw, C.A., Stankiewicz, P., and Gambin, A. (2015). Genome-wide analyses of LINE-LINE-mediated nonallelic homologous recombination. Nucleic Acids Res. *43*, 2188–2198.
35. Szafranski, P., Kośmider, E., Liu, Q., Karolak, J.A., Currie, L., Parkash, S., Kahler, S.G., Roeder, E., Littlejohn, R.O., DeNapoli, T.S., et al. (2018). LINE- and *Alu*-containing genomic instability hotspot at 16q24.1 associated with recurrent and nonrecurrent CNV deletions causative for ACDMPV. Hum. Mutat. *39*, 1916–1925.
36. P. Stankiewicz and J.R. Lupski, eds. (2020). The Genomic Basis of Medicine (Oxford University Press). https://doi.org/10.1093/med/9780198746690.003.0030.
37. Beck, C.R., Carvalho, C.M.B., Banser, L., Gambin, T., Stubbolo, D., Yuan, B., Sperle, K., McCahan, S.M., Henneke, M., Seeman, P., et al. (2015). Complex genomic rearrangements at the *PLP1* locus include triplication and quadruplication. PLoS Genet. *11*, e1005050.
38. Lee, J.A., Inoue, K., Cheung, S.W., Shaw, C.A., Stankiewicz, P., and Lupski, J.R. (2006). Role of genomic architecture in *PLP1* duplication causing Pelizaeus-Merzbacher disease. Hum. Mol. Genet. *15*, 2250–2265.
39. Bahrambeigi, V., Song, X., Sperle, K., Beck, C.R., Hijazi, H., Grochowski, C.M., Gu, S., Seeman, P., Woodward, K.J., Carvalho, C.M.B., et al. (2019). Distinct patterns of complex rearrangements and a mutational signature of microhomeology are frequently observed in *PLP1* copy number gain structural variants. Genome Med. *11*, 80.
40. Firth, H.V., Richards, S.M., Bevan, A.P., Clayton, S., Corpas, M., Rajan, D., Van Vooren, S., Moreau, Y., Pettett, R.M., and Carter, N.P. (2009). DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. Am. J. Hum. Genet. *84*, 524–533.

The American Journal of Human Genetics, Volume *109*

Supplemental information

TCEAL1 loss-of-function results in an X-linked dominant neurodevelopmental syndrome and drives the neurological disease trait in Xq22.2 deletions

Hadia Hijazi, Linda M. Reis, Davut Pehlivan, Jonathan A. Bernstein, Michael Muriello, Erin Syverson, Devon Bonner, Mehrdad A. Estiar, Ziv Gan-Or, Guy A. Rouleau, Ekaterina Lyulcheva, Lynn Greenhalgh, Marine Tessarech, Estelle Colin, Agnès Guichet, Dominique Bonneau, R.H. van Jaarsveld, A.M.A. Lachmeijer, Lyse Ruaud, Jonathan Levy, Anne-Claude Tabet, Rafal Ploski, Małgorzata Rydzanicz, Łukasz Kępczyński, Katarzyna Połatyńska, Yidan Li, Jawid M. Fatih, Dana Marafi, Jill A. Rosenfeld, Zeynep Coban-Akdemir, Weimin Bi, Richard A. Gibbs, Grace M. Hobson, Jill V. Hunter, Claudia M.B. Carvalho, Jennifer E. Posey, Elena V. Semina, and James R. Lupski

SUPPLEMENTAL NOTE

SUPPLEMENTAL NOTE: CASE REPORTS

Individual 1 was a 5-year-old male with hypotonia, especially of the lower limbs, unsteady gait, and global developmental delay with particular difficulties in expressive language. He was born at 37 weeks as the first child to healthy parents via C-section due to maternal HELLP (Hemolysis, Elevated Liver enzymes, and Low Platelets) syndrome and breech presentation. There were no prenatal exposures to medications or drugs reported. He walked unassisted at 18 months of age; his gait was broad based with small steps and toe-walking and asymmetric running characterized by dragging of the right foot. At 4 years chronological age, fine motor, gross motor, and receptive language skills were in the 20-24 month level while expressive language was at a 15 month level. He had a handful of single words but his speech was primarily babbling and he used signs and a communication tablet. Some autistic behavior including repetitive/self-stimulating behavior and episodes of frustration were noted, but he was very social with good eye contact. Ophthalmology exam identified intermittent high frequency small amplitude nystagmus with a large latent component and large increase in amplitude and frequency in lateral gazes, intermittent exotropia, and mild myopia with astigmatism and the fundus exam identified a crowded disc and tortuous vessels. He had recurrent otitis media but subsequently had normal hearing. Height was normal at 25-50th centile and weight was 90-97th centile. Physical exam identified mildly dysmorphic facial features with a widow's peak, broad forehead, deep set eyes, mildly coarse facial features, and mildly low-set ears (Figure 1A, B), and a single café-au-lait macule on the upper right leg. Brain MRI at 34 months of age was essentially normal with possible mild foreshortening of the corpus callosum (CC). The CC thickness, myelination of the internal capsule, optic nerves, and vestibulo-cochlear nerves were appropriate for age (Figure 2A). Spine MRI at the same age (34 months) showed mild scoliosis, which may have been positional or secondary to hypotonia, and a prominent central canal. Metabolic workup including urine organic acids, plasma amino

acids, acylcarnitine, very long chain fatty acids, and serum transferrin isoelectric focusing, as well as chromosomal microarray and Fragile X testing were normal.

Trio analysis of ES data identified a *de novo* hemizygous nonsense variant in *TCEAL1*, c.447G>A p.(Trp149*), which was not present in gnomAD. Sanger sequencing confirmed the hemizygous variant in the proband and its absence in both parents. Based on its location within the Xq22 critical region, robust expression in brain, and absence of loss-of-function variants within gnomAD, this was considered a good candidate to potentially explain the observed clinical phenotype in individual 1.

Individual 2 was a 10-year-old male with generalized hypotonia and severe global developmental delay. He was born at 41 weeks via C-section after a failed induction; pregnancy was complicated by maternal PUPPP (Pruritic Urticarial Papules and Plaques of Pregnancy). He sat unaided at the age of 2 years and 3 months, but never crawled or walked. He demonstrated some developmental regression, having had a handful of single words as a young child, but was later non-verbal. He displayed an array of unusual, repetitive behaviors from an early age. These included abnormal hand movements, self-stimulating behaviors (some suggestive of gratification behavior), breath-holding episodes, hyperventilation episodes, and severe aerophagia (swallowing air). He also had a disturbed sleep cycle with sleep deprivation and more recently developed self-harming behaviors. He developed a seizure disorder in middle childhood. Ophthalmological examination identified hyperopia and astigmatism. He had longstanding problems with gut dysmotility and constipation, ultimately resulting in an ileostomy and colostomy, and a small epigastric hernia. He also had recurrent lung/airway infections that were found to be secondary to poor swallow and micro-aspirations necessitating gastric tube insertion, recurrent thumb dislocation, and episodic hyponatremia. Physical exam at 3 years of age found height at 25-50th centile, weight >99th centile, and head circumference at the 91-98th centile as well as mildly dysmorphic features including long palpebral fissures, brachycephaly, prominent bow-shaped upper lip, tapered fingers and shortened thumb (Figure 1C-D). Brain MRI at 10 years of age was normal.

Trio analysis identified a *de novo* hemizygous indel variant, c.299_302del p.(Gly100Alafs*22) (Figure 3), which was not present in gnomAD. Sanger sequencing was not able to be performed but the variant was independently identified as a high quality hemizygous *de novo* variant by two separate studies (DDD and 100KGP projects) in both exome and genome sequencing.

Individual 3 was a 17-year-old male presenting with global developmental delay, hypotonia, unsteady gait and growth retardation. He is the second child of a healthy unrelated couple. The pregnancy was uneventful and he was born at term, with a weight at 3380g (53rd centile), a length at 51cm (72nd centile) and a head circumference at 35.5cm (79th centile) He showed a global developmental delay in the first year, with a sitting at 15 months and global hypotonia. At evaluation, he could stand and walk some steps with assistance, but with substantial ataxia. He had no language but could communicate by eye contact and by gestures of intention. He had never had seizures. He had a history of recurrent otitis media but with normal hearing. He had some autistic traits with repetitive and stereotyped behaviors and was treated by Risperidone, but he was social and didn't have a major behavioral disorder. Ophthalmology exam identified strabismus and astigmatism. He had difficulty chewing and constipation. His height was between -1 and -2SD until 5 y/o when he showed a statural shift. The endocrine investigations showed normal growth hormone secretion but low levels of IGF1. At 17 years-old, he has short stature (145cm, -4SD) with normal weight and head circumference (55.5cm). He also had small testicles, with no pubertal onset. Physical examination identified dysmorphic facial features with deep set eyes, mildly coarse facial features, microdontia and sparse teeth. He also had very bright blue eyes, which contrasted with the dark eyes of both of his parents. A complete metabolic workup, including urine organic acids, and plasma and urine amino acids, was normal. Chromosomal microarray and Angelman, Prader-Willi and Fragile X testing were also normal.

Trio analysis of whole exome sequencing identified a *de novo* hemizygous nonsense variant in *TCEAL1*, c.259C>T, p.(Gln87*), which was absent from gnomAD and predicted pathogenic. Sanger

sequencing was not able to be performed but the variant was present in 25 of 25 reads by NGS in the proband and was not present in 35 and 44 reads in the parents.

Individual 4 was a 7-year-old male with hypotonia, hypermobility and moderate to severe developmental delay. He was born after an uncomplicated pregnancy as a second child to healthy parents. After an initial normal start, he presented with joint hypermobility, feeding difficulties and constipation within the first half a year. He was known to have a history of respiratory tract and ear infections. Examination at one year of age revealed hypotonia and motor developmental delay. Later examinations noted global developmental delay including speech and language delays. At 4 years old, he started walking, communicated mainly through gestures and some single words, and had developed some stereotypic movements and tics. He has an iris coloboma. He had a broad forehead but no other dysmorphic features and growth parameters were within normal range. Brain MRI showed delayed myelinization of the terminal zones of the lateral ventricles and slight diminishing of white matter in the parietooccipital zones.

Individual 5 (BAB13799; Family HOU5113) was a 17-year-old female with mild to moderate intellectual disability. She was born at 37 weeks via NSVD to a 30-year-old G2P0 mother and a 34-year-old father after an uncomplicated pregnancy. There were no reported prenatal exposures. Birth weight was 6 lbs. 10 oz. (SD -0.52), birth length was ~19 inches (SD -0.48), and birth head circumference is not recalled. Apgar scores were not recalled. The neonatal period was complicated by hyperbilirubinemia, requiring five days of inpatient phototherapy. Medical record review indicated a peak bilirubin level of 17 mg/dL.

There were concerns for developmental delay from infancy. The individual crawled at 13 months and walked at 22 months. The timing of her first word was not recalled, but at age two she had approximately 10 words. She had been in special education classes since age 2 years and received speech therapy for moderate/severe expressive and receptive language disorder, and occupational therapy to address hand and truncal weakness. She was diagnosed with autism at age 7 years, previously having been diagnosed with pervasive developmental disorder-not otherwise specified (PDD-NOS). She was noted to have vocal and motor tics in childhood.

The individual had a significant change in her neurobehavioral features at age 11 years. Approximately two weeks after a febrile illness characterized by myalgia and sore throat, the individual demonstrated acute changes in her behavior and mood as well as fatigue, facial and shoulder tics, intrusive thoughts and staring spells. An evaluation was conducted for infectious causes of her symptoms including streptococcal infection and Lyme disease. No evidence of recent infection was identified, however the neurobehavioral features subsided after approximately 6 weeks of treatment with clarithromycin, but later recurred. After their recurrence, the individual was treated with antibiotics, and later intravenous immune globulin and steroids and the immunomodulators mycophenolate or methotrexate. These episodes were responsive to steroids, but not prevented by immunomodulation.

In recent years, symptoms continued to wax and wane, at times dramatically, requiring approximately 3-4 oral steroid bursts per year. At 17 years old, she had an episode that did not resolve with oral steroids and has since been treated with low dose guanfacine and naltrexone as well as penicillin shots every 3-4 weeks. Additional infectious work up was performed and was unrevealing. Over the course of 4+ years of symptoms she was treated with multiple medications, including clozapine, desvenlafaxine, ziprasidone, sertraline, lithium, lorazepam, clonazepam and lurasidone with limited response if any. Non-prescription cannabidiol was helpful for relief of OCD symptoms (e.g., compulsive scribbling with pen on paper). Over time many of the individual's symptoms of fatigue and cognitive difficulties improved. However, there were persistent auditory hallucinations.

Clinical genetics evaluation included karyotype, fragile X testing, DNA microarray and trio-ES and GS. Additional evaluations included EEG, MRI of brain, and EKG. She had a history of microcytic anemia and is an alpha-thalassemia carrier. Eye exam showed slight astigmatism. Audiology screenings were normal. The individual was diagnosed with oral allergy syndrome triggered by most fresh fruits and vegetables. This individual had a history of hypertriglyceridemia.

Physical exam at age 14y and 16y revealed epicanthal folds, slight overbite, narrow hands with slender fingers, clinodactyly of the 4^{th} and 5^{th} toes, hallux valgus, mild shortening of the lateral toes, a

single café au lait spot on the back and an unsteady gait. Anthropometric measurements at age 16 years: OFC 57.5 cm, >98th centile (+2.32 SD); weight 62.4kg, 76th centile; and height 159.4 cm, 30th centile. The individual was of European and Chinese ancestry. She had a younger sister who was reportedly in good health. Family history was negative for close relatives with psychiatric disease or developmental disorders.

Review of the chromosomal microarray data identified a *de novo* heterozygous ~14 Kb deletion of ChrX:102,879,326-102,893,312 (hg19) with 77 bp of microhomology at the breakpoint site; *TCEAL1* was the only gene within the deleted region. This deletion was not present in the Database of Genomic Variants (http://dgv.tcag.ca/dgv/app/home) and was confirmed through further characterization of the breakpoints as described below.

Individual 6 was a 9-year-old female with severe intellectual disability and obesity. She was the fourth child to healthy Algerian non-consanguineous parents. She was born at 40 weeks of gestation. Birth weight was 3950 g (92th centile), birth length 51.5 cm (84th centile) and occipito-frontal circumference (OFC) 36 cm (87th centile). Neonatal period was marked by axial hypotonia. Psychomotor development was delayed: she sat at 12 months of age, walked unassisted at 19 months of age, and spoke her first words at 4 years of age. At 18 months of age, she had an accelerated weight and length gain with increase of the weight curve up to +5SD. She had surgeries at 7 years of age for bilateral genu valgum due to obesity. At 8 years of age fine motor skills were in the 20-month level, gross motor skills were in the 14-month level and expressive and receptive language were below a 12-month level. She had autistic spectrum disorder features (stereotypies, disturbances in social interactions) without a diagnosis of autism due to her severe intellectual disability. At last examination at 9 years of age, she spoke about ten words with some associations of words. Height was 136 cm (+2.5SD), weight 74 kg (>+5SD), OFC at 55.5 cm (+2.5SD) and body mass index at 40 (morbid obesity). She walked with small steps and open feet. Physical examination identified dysmorphic facial features including frontal bossing, bilateral epicanthus, hypertelorism, deep set eyes, horizontal eyebrows, fleshy earlobes and diastema. Ophthalmology examination including

electroretinogram showed astigmatism and strabismus. X-rays showed congenital brachymetatarsia involving the fourth metatarsal bone in the right foot. Head CT at 36 months of age and brain MRI at 4 years of age were normal. Metabolic workup including purines and pyrimidines, creatinine, urinary organic acids, mucopolysaccharides and oligosaccharides in the urine, as well as Fragile X testing, *GNAS* gene screening and a next-generation sequencing panel of cognitive deficiency genes were normal.

SNP array showed a 154 Kb *de novo* Xp22.2 deletion encompassing 3 genes: *TCEAL4, TCEAL3* and only one OMIM gene *TCEAL1* (*300237). No duplication encompassing *TCEAL1* was reported in DGV or gnomAD Structural Variant databases. Genome sequencing confirmed the Xq22.2 *de novo* deletion seq[GRCh37] Xq22.2(102,774,750_102,929,222)x1 *dn*. No additional pathogenic variants were identified, including all known genes involved in neurodevelopmental disorder.

Individual 7 was a 6-year-old female with global developmental delay, mild intellectual disability, profound speech delay and epilepsy. She was born at 38 weeks as a second child (third pregnancy, one spontaneous abortion) to unrelated and excluding refractory error in the mother, otherwise healthy parents via C-section due to the mother's myopia. Mother reported decreased fetal movements in comparison to the pregnancy with the previous child. Postnatal hypotonia was also noted. Both motor and intellectual development were delayed. The individual walked independently from the age of 18 months. Until the age of 13 months, she only crawled. Speech development was delayed: until the age of 3 years, she did not speak any words, but rather used non-verbal communication, a few onomatopoeic sounds, and proto-declarative behavior. Until the age of 5 years, her vocabulary consisted of less than 20 words. After the oxcarbazepine treatment initiation, her vocabulary range increased and she began to combine words into sentences. She reacted to others' emotions, liked to attract others' attentions, enjoyed physical contact (hugging), and reacted aggressively when ignored. She attended municipal kindergarten as well as rehabilitation classes. In the Stanford-Binet test, she was assessed as having mild intellectual disability. She also displayed self-stimulation, episodic temper tantrums and severely increased appetite. EEG revealed focal paroxysmal activity in the left frontocentrotemporal area with sharp waves and slow waves. After the

first atypical absence seizures at the age of 4 years, she was diagnosed with epilepsy and treated with oxcarbazepine with dose escalation to 150 mg – 0 – 450 mg. Good seizure control was achieved on this regimen. Head MRI at 6 years of age revealed a few small foci of increased signal in F2/FLAIR images, located in subcortical and periventricular areas of the white matter of both hemispheres and right dentate nucleus – with no evidence of contrast enhancement. Signs of upper respiratory tract inflammation was also noted: thickening of the maxillary sinus mucosa and enlargement of the pharyngeal tonsils. Slight facial dysmorphic features, including telecanthus, broad forehead, low-set ears and widely spaced teeth were also observed, and a single lentigo on the scalp. Other health problems included obesity, astigmatism and premature puberty (breast enlargement).

Methylation status of the Prader-Willi/Angelman critical region was normal (commercially available MS-MLPA Probemix ME028), previous *MAGEL2* sequencing for Schaaf-Yang syndrome revealed no pathogenic or potentially pathogenic variants. Singleton ES performed in the proband revealed a heterozygous ultra-rare (absent in gnomAD database) missense variant in *TMEM240* and frameshift variants in *ST6GALNAC6* and *TCEAL1*. With segregation studies, the *TMEM240* and *ST6GALNAC6* variants were found to be inherited from the proband's healthy father (the *TMEM240* variant was also observed in the proband's healthy brother, data not shown) and were classified as benign. However, we noted that, being extremely rare, the presence of these variants in the father confirmed parenthood. The frameshift variant in *TCEAL1*, c.169delC p.(Leu57Serfs*36), was absent in both parents and the proband's brother, thus it was considered as a *de novo* event.

Individual 8 was a 12-year-old male of European ancestry who came to medical attention for features of spastic paraparesis. He was born to a non-consanguineous couple who were 35 (father) and 36 (mother) years of age at term, and there were no immediate concerns during the neonatal period. Birth weight was 3480 g (44.1%, -0.15 SD). He met his early developmental milestones during the first year of life. At the age of 13 months, he began walking independently, and his gait was described as idiopathic toe-walking of childhood. He would walk only short distances, and began to crawl increasingly. Sometime after he began

walking, lower extremity spasticity was noted, with additional difficulties in fine motor skills involving the upper extremities. By the age of 8 years, he had mild dysarthria when speaking English (his second language), no dysphagia, full extra-ocular movements, normal visual acuity and auditory functioning, and presumed normal intellect based on his performance in the third grade. He did not demonstrate any bowel incontinence, but did develop urinary incontinence. His motor exam was notable for 4+/5 weakness involving the shoulders and lower extremity flexors with evidence for spasticity of the lower extremities more severely than the upper extremities. He was hyperreflexic with an upgoing Babinski, and a normal sensory exam. Rapid alternating movements were slow with his upper extremities, but there was no dysmetria. His gait was very spastic. He was also noted to have mild scoliosis. Clinical investigation included a normal brain MRI, normal CT scan of the head, and normal EMG.

His family history was notable for two parents in good health, and two older sisters, ages 12y and 10y, who were also reported to be in good health. Neurodegenerative conditions were reported on both maternal and paternal sides of the family: progressive supranuclear palsy (PSP) in the maternal grandmother (deceased beyond 60 years of age from the same), Parkinson disease (PD) in the paternal grandfather (deceased at 42 years of age from the same), and one maternal cousin who is presently living with multiple sclerosis (MS).

Review of exome data identified a hemizygous missense variant in *TCEAL1*, c.346G>A p.(Asp116Asn) (Figure 3), which was present in gnomAD in one female European individual with a population specific allele frequency of 0.000012 and an overall allele frequency of 0.00000545. The variant was predicted to be deleterious by SIFT but benign by PolyPhen-2 and had a CADD phred score of 24.1 and a REVEL score of 0.139 (Table 1, Table S2). The variant was found to be inherited from the unaffected mother. Further analysis of the ES data by ExomeDepth did not reveal any pathogenic copy number variants.

SUPPLEMENTAL TABLES

Table S2 Detailed Molecular Information

	Individual 1	**Individual 2**	**Individual 3**	**Individual 4**	**Individual 5**	**Individual 6**	**Individual 7**	**Individual 8**
Sex	M	M	M	M	F	F	F	M
Ancestry	White (U.S.)	White (European)	White (European)	White (European)	White + Chinese	Algerian	White (European)	White (European)
Consanguinity	No	No	No	No	No	No	No	No
Age at most recent clinical evaluation	5 y/o	10 y/o	17 y/o	7 y/o	17 y/o	8 y/o	6 y/o	12 y/o
Variant type	SNV Nonsense	Indel Frameshift	SNV Nonsense	SNV (Missense)	CNV (single gene) Deletion	Xq22 deletion	SNV Frameshift	SNV Missense
Variant	NM_001006639: c.447G>A p.(Trp149*)	NM_001006639: c.299_302del p.(Gly100Alafs*22)	NM_001006639: c.259C>T p.(Gln87*)	NM_001006639: c.269G>A p.(Cys90Tyr)	Single-gene deletion of *TCEAL1*	Deletion of *TCEAL1, TCEAL3, TCEAL4*	NM_001006639: c.169delC p.(Leu57Serfs*36)	NM_001006639: c.346G>A p.(Asp116Asn)
Zygosity	Hemizygous 39 of 39 reads Sanger confirmed	Hemizygous 12 of 12 reads	Hemizygous 25 of 25 reads	Hemizygous Sanger confirmed	Heterozygous	Heterozygous	Heterozygous 42 out of 79 reads	Hemizygous Sanger confirmed
Coordinates (hg19)	ChrX: 102,885,291	ChrX: 102,885,138	Chr X: 102,885,103†	ChrX: 102,885,113	ChrX: 102,879,326- 102,893,312	ChrX: 102,774,750- 102,929,222	ChrX: 102,885,012	ChrX: 102,885,190
Inheritance	*De novo* Sanger confirmed	*De novo* by NGS	*De novo* by NGS	*De novo* Sanger confirmed	*De novo* HD-aCGH confirmed	*De novo* SNP-array, WGS	*De novo* NGS-based amplicon deep sequencing	Maternally inherited Sanger confirmed
Population frequency	NP (gnomAD)	NP (gnomAD)	NP (gnomAD)	NP (gnomAD)	NP (DGV, DECIPHER, gnomAD SV)	NP (DGV, gnomAD)	NP (gnomAD)	1/81,897 European alleles (gnomAD)
In silico missense predictions	Truncation	Truncation	Truncation	SIFT: Tolerated (0.09) Poly-Phen-2: Probably damaging (0.997)	Full gene deletion	Full gene deletion	Truncation	SIFT: Deleterious (0.1) PolyPhen-2: benign (0.429)

CADD and REVEL scores	37; NA	NA	36; NA	23.8; 0.146	NA	NA	NA	24.1; 0.139
Genetic testing performed	Trio ES	Trio ES	Trio ES	Trio ES	Trio-CMA (BG) Trio HD-aCGH	Trio GS Trio SNP array	Singleton ES	Singleton ES
Other genetic/ environmental factors contributing to phenotype	None	None	None	None	Carrier for alpha thalassemia trait	None	None	None
Other variants identified but interpreted as probably non-pathogenic	None known/reported	None known/reported	None known/reported	Two regions of homozygosity (7.77 Mb and 19.38 Mb) with no known clinical significance	Het SNV in *HAAO* (NM_012205.3: c.685T>C p.(W229R)) - AR disease (MIM# 617660) mainly skeletal/non-neurological	None known/reported	*TMEM240* NM_001114748.2: c.481A>C p.(Lys161Gln) inherited from a healthy father	None known/reported

BG: Baylor Genetics, CADD: Combined Annotation Dependent Depletion, CMA: chromosomal microarray analysis, HD-aCGH: high definition array comparative genomic hybridization, CNV: copy number variant, DGV: database of genomic variants, ES: exome sequencing, F: female, GS: genome sequencing, M: male, NA: not applicable, NP: not present; REVEL: rare exome variant ensemble learner, SNV: single nucleotide variant, WGS: whole genome sequencing. †converted from hg38 coordinates (chrX:103,630,175).

NM_001006639.2 is used for variant nomenclature.

Table S3 Amino Acid identity of TCEAL proteins with TCEAL1

The TCEAL protein family includes 9 proteins (TCEAL1, TCEAL2, TCEAL3, TCEAL4, TCEAL5, TCEAL6, TCEAL7, TCEAL8, and TCEAL9) located in Xq22.1-Xq22.2. The percent amino acid identity of each protein with TCEAL1 is shown.

Gene	**Genomic Region**	**% amino acid identity with TCEAL1**
TCEAL2	Xq22.1	42.11%
TCEAL3	Xq22.2	No significant similarity
TCEAL4	Xq22.2	31.25%
TCEAL5	Xq22.1	No significant similarity
TCEAL6	Xq22.1	No significant similarity
TCEAL7	Xq22.1	35.87%
TCEAL8	Xq22.1	52.83%
TCEAL9	Xq22.2	53.33%

SUPPLEMENTAL METHODS

GENOMIC METHODS

Genomic methods for each subject are summarized in Supplemental Table 2. **Individuals 1-4 and 7-8** underwent exome sequencing (ES) or genome sequencing (GS). ES in **individual 1** and his parents was completed through Psomagen (Rockville, MD), with trio analysis completed using SNP & Variation Suite (Golden Helix, Bozeman, MT) including gnomAD tracts for general population frequency and dbNSFP for *in silico* predictions as described.[1] Query of the MatchMaker exchange identified seven additional individuals hosted by MyGene2[2], GeneMatcher[3,4] and DECIPHER.[5,6] For **individual 2**, ES and trio analysis were completed through the Deciphering Developmental Disorders project, and genome sequencing (GS) and trio analysis were completed through the 100,000 genomes project. For **individual 3**, clinical ES and trio analysis was performed in the Genetics department of Angers University Hospital using a Twist Biosicence (San Francisco, CA) in-solution enrichment methodology (Twist Human Core Exome EF Multiplex + Twist Human RefSeq Panel), followed by 101 bases of paired-end reads via a massively parallel sequencing approach on Illumina NextSeq550. Bcl were converted as FASTQ using GenerateFASTQ (v.1.0.0). Quality control was done using Fastqc (v.0.11.9). Sequence reads were mapped to the human genome build (hg38 / GRCh38) and analyzed with a dedicated in-house pipeline integrating various modules for coverage analysis, variant calling, annotation. Variant prioritization was performed with VarAFT 2.16.[7] For **individual 4**, clinical trio ES was performed using SureSelect V5 (Agilent Technologies) capture reagent, and sequenced on an Illumina HiSeq 2500. Coding and splice site exome variants were prioritized based on predicted *de novo* or autosomal recessive disease transmission and following the diagnostic standards of the University Medical Center Utrecht. **individual 7** underwent singleton clinical ES performed using Twist Human Core Exome EF Multiplex + Twist Human RefSeq Panel (Twist Bioscience) and paired-end (2x100 bp) sequenced on Illumina NovaSeq 6000. Family study was performed by next generation sequencing-based deep amplicon sequencing. **individual 8** underwent singleton ES with CNV analysis using ExomeDepth[8].

For all SNVs, i*n silico* analysis included SIFT and Polyphen-2 predictions of the effect of missense variants as well as CADD[9] and REVEL[10] scores. Variant segregation within families by PCR and subsequent Sanger sequencing was performed for two of the four ES-identified *TCEAL1* SNVs. Primer pairs were used for the segregation PCRs as follows:

Individual	**Forward Primer (5'-3')**	**Reverse Primer (5'-3')**
Individual 1	AGCAGCCTCCTTGTGGAGTA	ATCTTTCATGCAAATGTGTAGGGC
Individual 4	TGAGGAGCTCTTGCCTGA	GTATTGCCATTGCCACCT
Individual 8	CCTCGGAGGAGGAGTTCTTT	CAGAAAGTCCAGGTGGCAAT

PCR and Sanger sequencing was not possible for individual 2 (independent exome sequencing and genome sequencing trio analyses identified the *de novo* variant), individual 3 (variant was present in 25/25 NGS reads and not present in either parent), or individual 7 (variant was present in 3716 reads out of 8061 NGS-base amplicon deep sequencing reads and not present in either parent or healthy brother, with depth of coverage at 13,897x, 7514x, and 13084x, respectively).

Individual 5 was identified through a screen for single-gene CNVs in the clinical microarray database of the Baylor Genetics (BG) diagnostic laboratory (consisting of > 75,000 clinical genomes). The deletion in individual 5 was initially detected by clinical-grade Chromosomal Microarray Analysis (CMA; version 11.2) at BG. This CMA version has a total of 400k probes spread genome-wide at alternating average densities of '> 4 probes per exon' in >4,200 clinically-relevant genes and '1 probe per 30 Kb' genome-wide in non-coding regions.[11] Due to the lower probe resolution of CMA array in non-coding regions, a higher resolution custom array was utilized for the precise mapping of deletion breakpoints; this high-density (HD) array was designed on the Agilent eArray website (https://earray.chem.agilent.com/suredesign/) and interrogates copy-number variation (CNV) in Xq22 via a total of 40,208 60-mer probes spanning chrX:97,915,511–113,400,000 (NCBI build 36) at an average distribution of one probe per 386 bp (format 4 × 44 K).[12] It was utilized to assay the genomes of this subject (individual 5, BAB13799), and her parents

(BAB13800 and BAB13801). The experimental procedures were performed according to the manufacturer's protocol (Agilent Oligonucleotide Array-Based CGH for Genomic DNA Analysis, Version 7.2) with some modifications.[13] Sex-matched controls (GM15510 as female control and GM10851 as male control, both from the Coriell Institute) were used with the trio. The scanned array images were processed by the Agilent Feature Extraction software (version10) and the extracted files were analyzed by the Genomic Workbench (version 7.0.4.0). Maximum proximal and distal boundaries of the heterozygous X-linked deletion were defined as follows, proximally as the last probe with a log2 ratio of 1 prior to the deleted probes (avg. log2 ratio of −1) and distally as the first probe with a log2 ratio of 1 after the deleted probes.

Precise mapping of the deletion breakpoints to the nucleotide level was accomplished by PCR and Sanger sequencing using inward facing primers that completely span the deletion in individual 5 (i.e. primers that bind outside of the maximum boundary coordinates deduced from HD-aCGH results). Parental samples were studied alongside the proband's to confirm specificity of the PCR product in the proband. The Jct-PCR primer-pair sequences are as follows, F (5'->3'): GAGGCCTTGCATTGTCTTTT and R (5'->3'): AAAGCTGCTCCCAGGAAAGT. The obtained sequence (i.e. the deletion breakpoint-junction sequence or Jct-sequence) was aligned to the haploid reference genome (GRCh37/hg19) and breakpoint mapping coordinates were evaluated for overlapping repeat sequences using four different repeat datasets, 1) the highly similar intrachromosomal repeats (HSIR) dataset,[14] 2) the Segmental Dup dataset,[15,16] 3) the Repbase (Repeat Masker) dataset,[17] and 4) the SelfChain dataset.[18,19]

Individual 6 underwent a parent-proband clinical trio-SNP array (Illumina OmniExpress 700K) as well as trio-genome sequencing for fine-mapping of breakpoints. X chromosome inactivation (XCI) studies were performed clinically by analysis of methylation status at the following loci: *PCSK1N* (Xp11.23), *AR* (Xq11.2-2q12), *ZDHHC15* (Xq13.3), and *SLITRK4* (Xq27.3).

INFORMED CONSENT

Written informed consent for research studies and/or reporting of clinical features was obtained for all Individuals, including photo publication if applicable; additional clinical details of these individuals are provided herein. Human studies were approved by the Institutional Review Boards of Children's Wisconsin (protocol 124172) (individual 1), UK Research Ethics Committee (REC) (protocol 10/H0305/83 granted by the Cambridge South REC and GEN/284/12 granted by the Republic of Ireland REC) and Health Research Authority REC (individual 2), Baylor College of Medicine (protocol H-29697) (individual 5), and McGill University Health Centre (protocol 31-01-008703) (individual 8); the remaining variants were identified through clinical testing with informed consent to publish. Genomic DNA for all experiments was isolated from blood according to standard procedures. All available neuroimages were reviewed by an expert board-certified neuro-radiologist (JVH).

SUPPLEMENTAL ACKNOWLEDGEMENTS

This study makes use of data generated by the DECIPHER community and the DDD study.

The DDD study presents independent research commissioned by the Health Innovation Challenge Fund [grant number HICF-1009-003], a parallel funding partnership between Wellcome and the Department of Health, and the Wellcome Sanger Institute [grant number WT098051]. The views expressed in this publication are those of the author(s) and not necessarily those of Wellcome or the Department of Health. The study has UK Research Ethics Committee approval (10/H0305/83, granted by the Cambridge South REC, and GEN/284/12 granted by the Republic of Ireland REC). The research team acknowledges the support of the National Institute for Health Research, through the Comprehensive Clinical Research Network.

This study makes use of data generated by the DECIPHER community. A full list of centres who contributed to the generation of the data is available from https://deciphergenomics.org/about/stats and via email from contact@deciphergenomics.org. Funding for the DECIPHER project was provided by Wellcome. Those who carried out the original analysis and collection of the Data bear no responsibility for the further analysis or interpretation of the Data.

SUPPLEMENTAL REFERENCES

1. Reis, L.M., Sorokina, E.A., Thompson, S., Muheisen, S., Velinov, M., Zamora, C., Aylsworth, A.S., and Semina, E.V. (2019). *De Novo* Missense Variants in *WDR37* Cause a Severe Multisystemic Syndrome. Am J Hum Genet *105*, 425-433.
2. MyGene2. NHGRI/NHLBI University of Washington-Center for Mendelian Genomics, Seattle, WA (https://mygene2.org/MyGene2/). http://www.mygene2.org.
3. Azzariti, D.R., and Hamosh, A. (2020). Genomic Data Sharing for Novel Mendelian Disease Gene Discovery: The Matchmaker Exchange. Annu Rev Genomics Hum Genet *21*, 305-326.
4. Sobreira, N., Schiettecatte, F., Valle, D., and Hamosh, A. (2015). GeneMatcher: a matching tool for connecting investigators with an interest in the same gene. Hum Mutat *36*, 928-930.
5. Firth, H.V., Richards, S.M., Bevan, A.P., Clayton, S., Corpas, M., Rajan, D., Van Vooren, S., Moreau, Y., Pettett, R.M., and Carter, N.P. (2009). DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. Am J Hum Genet *84*, 524-533.
6. Swaminathan, G.J., Bragin, E., Chatzimichali, E.A., Corpas, M., Bevan, A.P., Wright, C.F., Carter, N.P., Hurles, M.E., and Firth, H.V. (2012). DECIPHER: web-based, community resource for clinical interpretation of rare variants in developmental disorders. Hum Mol Genet *21*, R37-44.
7. Desvignes, J.P., Bartoli, M., Delague, V., Krahn, M., Miltgen, M., Beroud, C., and Salgado, D. (2018). VarAFT: a variant annotation and filtration system for human next generation sequencing data. Nucleic Acids Res *46*, W545-W553.
8. Plagnol, V., Curtis, J., Epstein, M., Mok, K.Y., Stebbings, E., Grigoriadou, S., Wood, N.W., Hambleton, S., Burns, S.O., Thrasher, A.J., et al. (2012). A robust model for read count data in exome sequencing experiments and implications for copy number variant calling. Bioinformatics *28*, 2747-2754.
9. Rentzsch, P., Witten, D., Cooper, G.M., Shendure, J., and Kircher, M. (2019). CADD: predicting the deleteriousness of variants throughout the human genome. Nucleic Acids Res *47*, D886-D894.
10. Ioannidis, N.M., Rothstein, J.H., Pejaver, V., Middha, S., McDonnell, S.K., Baheti, S., Musolf, A., Li, Q., Holzinger, E., Karyadi, D., et al. (2016). REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. Am J Hum Genet *99*, 877-885.
11. Gambin, T., Yuan, B., Bi, W., Liu, P., Rosenfeld, J.A., Coban-Akdemir, Z., Pursley, A.N., Nagamani, S.C.S., Marom, R., Golla, S., et al. (2017). Identification of novel candidate disease genes from *de novo* exonic copy number variants. Genome Med *9*, 83.

12. Carvalho, C.M., Bartnik, M., Pehlivan, D., Fang, P., Shen, J., and Lupski, J.R. (2012). Evidence for disease penetrance relating to CNV size: Pelizaeus-Merzbacher disease and manifesting carriers with a familial 11 Mb duplication at Xq22. Clin Genet *81*, 532-541.
13. Carvalho, C.M., Zhang, F., Liu, P., Patel, A., Sahoo, T., Bacino, C.A., Shaw, C., Peacock, S., Pursley, A., Tavyev, Y.J., et al. (2009). Complex rearrangements in patients with duplications of *MECP2* can occur by fork stalling and template switching. Hum Mol Genet *18*, 2188-2203.
14. Hijazi, H., Coelho, F.S., Gonzaga-Jauregui, C., Bernardini, L., Mar, S.S., Manning, M.A., Hanson-Kahn, A., Naidu, S., Srivastava, S., Lee, J.A., et al. (2020). Xq22 deletions and correlation with distinct neurological disease traits in females: Further evidence for a contiguous gene syndrome. Hum Mutat *41*, 150-168.
15. Bailey, J.A., Gu, Z., Clark, R.A., Reinert, K., Samonte, R.V., Schwartz, S., Adams, M.D., Myers, E.W., Li, P.W., and Eichler, E.E. (2002). Recent segmental duplications in the human genome. Science *297*, 1003-1007.
16. Bailey, J.A., Yavor, A.M., Viggiano, L., Misceo, D., Horvath, J.E., Archidiacono, N., Schwartz, S., Rocchi, M., and Eichler, E.E. (2002). Human-specific duplication and mosaic transcripts: the recent paralogous structure of chromosome 22. Am J Hum Genet *70*, 83-100.
17. Jurka, J. (2000). Repbase update: a database and an electronic journal of repetitive elements. Trends Genet *16*, 418-420.
18. Chiaromonte, F., Yap, V.B., and Miller, W. (2002). Scoring pairwise genomic sequence alignments. Pac Symp Biocomput, 115-126.
19. Schwartz, S., Kent, W.J., Smit, A., Zhang, Z., Baertsch, R., Hardison, R.C., Haussler, D., and Miller, W. (2003). Human-mouse alignments with BLASTZ. Genome Res *13*, 103-107.