Paralog Studies Augment Gene Discovery: *DDX* and *DHX* Genes

Ingrid Paine,[1,39] Jennifer E. Posey,[1,39] Christopher M. Grochowski,[1] Shalini N. Jhangiani,[2] Sarah Rosenheck,[3] Robert Kleyner,[3] Taylor Marmorale,[3] Margaret Yoon,[3] Kai Wang,[4] Reid Robison,[5] Gerarda Cappuccio,[6,7] Michele Pinelli,[6,7] Adriano Magli,[8] Zeynep Coban Akdemir,[1] Joannie Hui,[9] Wai Lan Yeung,[10] Bibiana K.Y. Wong,[11,12] Lucia Ortega,[13] Mir Reza Bekheirnia,[1,14,15] Tatjana Bierhals,[16] Maja Hempel,[16] Jessika Johannsen,[17] René Santer,[17] Dilek Aktas,[18] Mehmet Alikasifoglu,[18] Sevcan Bozdogan,[19] Hatip Aydin,[20] Ender Karaca,[21] Yavuz Bayram,[22] Hadas Ityel,[23] Michael Dorschner,[24] Janson J. White,[25] Ekkehard Wilichowski,[26] Saskia B. Wortmann,[27,28,29] Erasmo B. Casella,[30] Joao Paulo Kitajima,[31] Fernando Kok,[31,32] Fabiola Monteiro,[31]

(Author list continued on next page)

Members of a paralogous gene family in which variation in one gene is known to cause disease are eight times more likely to also be associated with human disease. Recent studies have elucidated *DHX30* and *DDX3X* as genes for which pathogenic variant alleles are involved in neurodevelopmental disorders. We hypothesized that variants in paralogous genes encoding members of the DExD/H-box RNA helicase superfamily might also underlie developmental delay and/or intellectual disability (DD and/or ID) disease phenotypes. Here we describe 15 unrelated individuals who have DD and/or ID, central nervous system (CNS) dysfunction, vertebral anomalies, and dysmorphic features and were found to have probably damaging variants in DExD/H-box RNA helicase genes. In addition, these individuals exhibit a variety of other tissue and organ system involvement including ocular, outer ear, hearing, cardiac, and kidney tissues. Five individuals with homozygous (one), compound-heterozygous (two), or *de novo* (two) missense variants in *DHX37* were identified by exome sequencing. We identified ten total individuals with missense variants in three other DDX/DHX paralogs: *DHX16* (four individuals), *DDX54* (three individuals), and *DHX34* (three individuals). Most identified variants are rare, predicted to be damaging, and occur at conserved amino acid residues. Taken together, these 15 individuals implicate the DExD/H-box helicases in both dominantly and recessively inherited neurodevelopmental phenotypes and highlight the potential for more than one disease mechanism underlying these disorders.

Introduction

Paralogs can often functionally compensate for each other, and this compensation provides genetic robustness but also allows for tissue-specific effects of ubiquitously expressed genes.[1] There are many examples of paralogous gene contributions for well-characterized syndromes, including connective tissue diseases like Ehlers-Danlos (EDSCL1 [MIM: 130000], EDSVASC [MIM: 130050], EDSARTH2 [MIM: 617821], EDSARTH1 [MIM: 130060], PMGEDSV [MIM: 618343], and EDSCV [MIM: 225320]) and Steel syndromes (STLS [MIM: 615155]) (collagen gene family),[2–4] and channelopathies such as Dravet syndrome (EIEE6 [MIM: 607208]) (sodium channel gene family). Mosaic screens in *Drosophila* have indicated that homozygous-lethal *Drosophila* genes for which there are multiple human orthologs are enriched eight-fold for association with human disease. Furthermore, evolutionarily conserved genes with more than one human paralog are three times more likely to be associated with OMIM diseases.[5] One candidate gene family that is both orthologously conserved among species and has multiple human paralogs is the DExD/H-box domain superfamily 2 of the RNA helicases; the family has 58 members, all containing 8–9 conserved amino acid motifs constituting the helicase core region.

The DExD/H-box RNA helicase family is a group of conserved proteins (Figure 1) named for the consensus amino acid sequence DExD or DExH in their Walker B motif (motif 2) within the helicase core domain,[7] which

[1]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA; [2]Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX 77030, USA; [3]Stanley Institute for Cognitive Genomics, Cold Spring Harbor Laboratory, NY 11724, USA; [4]Raymond G. Perelman Center for Cellular and Molecular Therapeutics, Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA; [5]Utah Foundation for Biomedical Research, Salt Lake City, UT 84107, USA; [6]Department of Translational Medicine, University of Naples "Federico II," 80131 Napoli, Italy; [7]Telethon Institute of Genetics and Medicine, 80078 Pozzuoli, Italy; [8]Department of Pediatric Ophthalmology, University of Salerno, 84081 Baronissi SA, Italy; [9]Department of Pediatrics, Prince of Wales Hospital, the Chinese University of Hong Kong, Hong Kong SAR, China; [10]Department of Pediatrics and Adolescent Medicine, Alice Ho Miu Ling Nethersole Hospital, Hong Kong SAR, China; [11]Department of Obstetrics and Gynecology, Baylor College of Medicine, Houston, TX 77030, USA; [12]The Jan and Dan Duncan Neurological Research Institute, Texas Children's Hospital, Houston, TX 77030, USA; [13]Medical Genetics Department, Cook Children's Hospital, Fort Worth, TX 76104, USA; [14]Department of Pediatrics, Section of Pediatric Renal, Baylor College of Medicine, Houston, TX 77030, USA; [15]Department of Genetics, Texas Children's Hospital, Houston, TX 76104, USA; [16]Institute of Human Genetics, University Medical Center Hamburg-Eppendorf, 20246 Hamburg, Germany; [17]Department of Pediatrics, University Medical Center Hamburg-Eppendorf, 20246 Hamburg, Germany;

(Affiliations continued on next page)

Donna M. Muzny,[1,2] Michael Bamshad,[25,33,34] Richard A. Gibbs,[1,2] V. Reid Sutton,[1] University of Washington Center for Mendelian Genomics, Baylor-Hopkins Center for Mendelian Genomics, Telethon Undiagnosed Diseases Program, Hilde Van Esch,[35] Nicola Brunetti-Pierri,[6,7] Friedhelm Hildebrandt,[23] Ariel Brautbar,[13] Ignatia B. Van den Veyver,[1,11,12] Ian Glass,[34] Davor Lessel,[16] Gholson J. Lyon,[3,5,36] and James R. Lupski[1,2,37,38,*]

facilitates the ATP-dependent unwinding of RNA secondary structures. The DExD/H-box RNA helicase family is conserved from bacteria to humans, and the core domains are conserved among family members (Figure 1). This gene family is thought to have expanded evolutionarily by both whole-genome duplication and tandem amplification, resulting in different degrees of variation among gene family members. The two core domains are involved in RNA binding, NTP-binding and hydrolysis, helicase activity, and substrate recognition. Additionally, both the N- and C-terminal domains of these proteins are involved in specific protein-protein interactions that often define the specific biological role of the protein,[7] although the functions of each individual member are not well understood. DExD/H-box RNA helicases are ubiquitously expressed; however, some members share patterns of elevated or specific expression in certain tissues (e.g., cerebellum or testis). The diversity among these helicases could potentially allow for the regulation of many different pathways in development and disease.

Two members of the DExD/H-box RNA helicase family, *DHX30* (MIM: 616423) and *DDX3X* (MIM: 300160), have had multiple different pathogenic variant alleles associated with neurodevelopmental disorders including developmental delay and intellectual disability (DD and/or ID), speech impairment, brain abnormalities, and gait perturbations.[8–12] Twelve unrelated individuals with sporadic neurodevelopmental disorder with severe motor impairment and absent language (NEDMIAL [MIM: 617804]) were each found to have *de novo* missense *DHX30* variants, of which four variants were found to be recurrent and located within nucleotide-interacting motifs of the helicase core region.[8] To date, at least 75 individuals have been reported to have neurodevelopmental delay (MRX102 [MIM: 300958]) in association with *DDX3X* variants.[9–13] Curiously, a majority of affected individuals for this X-linked condition are female, although a handful of males who inherited a *DDX3X* variant from an unaffected mother have been described, suggesting gender-specific variant pathogenicity. Identified variants have included missense, stop-gain, and frameshifting variants, suggesting a potential loss-of-function mechanism. Moreover, mutations of *DDX3X* are one of the most common causes of unexplained intellectual disability (ID), accounting for about 1%–3% of instances in females.[10,11] Furthermore, a third member of this paralogous gene family, *DHX38* (MIM: 605584), was reported recently as a candidate gene in three unrelated families with early-onset autosomal-recessive retinitis pigmentosa (RP84 [MIM: 618220]) in association with homozygous missense *DHX38* variants.[14,15] In light of these findings and the correlation between gene paralogs and human disease phenotypes, we hypothesized that variant alleles in other members of the DExD/H-box RNA helicase gene family might also underlie neurodevelopmental disease traits, and we investigated their potential role in trait manifestations of human disease biology. Through the Baylor Hopkins Center for Mendelian Genomics (BHCMG) and collaborations facilitated by GeneMatcher,[16,17] we provide evidence supporting that *DHX37* (MIM: 617362), *DHX16* (MIM: 603405), *DDX54* (MIM: 611665), and *DHX34* (MIM: 615475) are neurodevelopmental genes (Figure 1). In addition, three single individuals with potentially pathogenic variants in *DHX8*, *DHX58*, and *DDX47* are presented in Table 1 and the Supplemental Individual Reports in order to bring these genes to the attention of the field.

Material and Methods

Participants

All individuals and available parents and relatives participating in this study provided informed consent, including consent to publish photographs, for this study. The BHCMG exome sequencing

[18]DAMAGEN Genetic Diagnostic Center, 06690 Ankara, Turkey; [19]Department of Medical Genetics, Cukurova University Faculty of Medicine, 01330 Adana, Turkey; [20]Department of Medical Genetics, Medical Faculty of Namik Kemal University, Tekirdag 59100, Turkey; [21]Department of Genetics, University of Alabama at Birmingham, Birmingham, AL 35294, USA; [22]Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York City, NY 10029, USA; [23]Division of Nephrology, Department of Pediatrics, Harvard Medical School, Boston, MA 02115, USA; [24]Center for Precision Diagnostics, University of Washington, Seattle, WA 98195, USA; [25]Department of Pediatrics, University of Washington, Seattle, WA 98195, USA; [26]Department of Pediatrics and Pediatric Neurology, Georg-August-Universität Göttingen, 37075 Göttingen, Germany; [27]Institute of Human Genetics, Technical University München, 81675 Munich, Germany; [28]Institute of Human Genetics, Helmholtz Zentrum Munchen, 85764 Neuherberg, Germany; [29]University Children's Hospital, Paracelsus Medical University, 5020 Salsburg, Austria; [30]Children's Institute, Hospital das Clinicas, University of Sao Paulo, 05405-000 Sao Paulo, Brazil; [31]Mendelics Genomic Analysis, 04013-000 Sao Paulo, Brazil; [32]Department of Neurology, University of Sao Paulo School of Medicine, 01246-903 Sao Paulo, Brazil; [33]Department of Genome Sciences, University of Washington, Seattle, WA 98195, USA; [34]Division of Genetic Medicine, University of Washington, Seattle, WA 98195, USA; [35]Center for Human Genetics, University Hospitals Leuven, 3000 Leuven, Belgium; [36]Institute for Basic Research in Developmental Disabilities, Staten Island, NY 10314, USA; [37]Department of Pediatrics, Baylor College of Medicine, Houston, TX 77030, USA; [38]Texas Children's Hospital, Houston, TX 77030, USA
[39]These authors contributed equally to this work
*Correspondence: jlupski@bcm.edu
https://doi.org/10.1016/j.ajhg.2019.06.001.

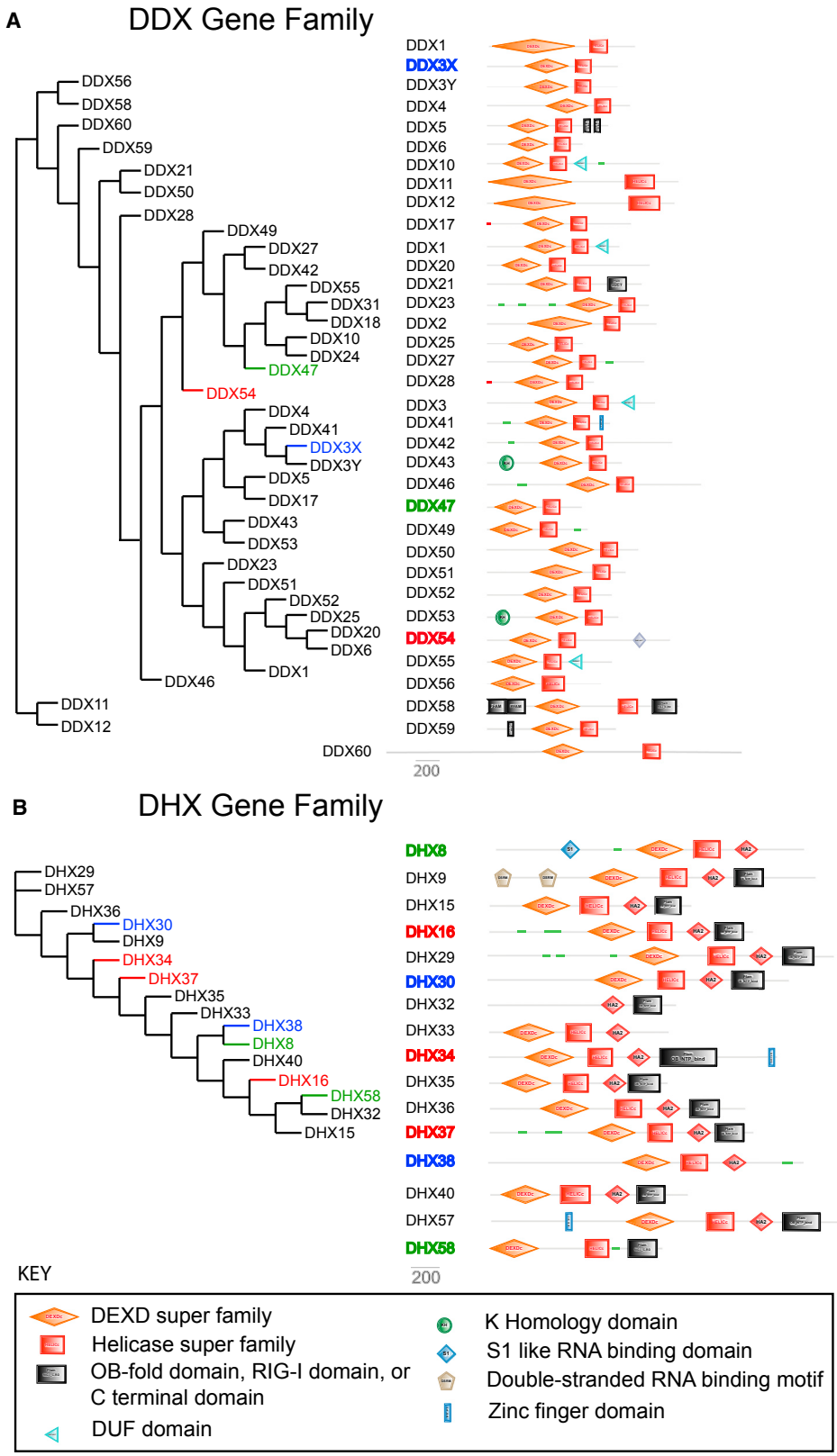


Figure 1. DExD/H-Box RNA Helicase Protein Family

(A and B) Phylogeny trees for DEAD-box (A, left) and DEAH-box (B, left) genes were generated with Phylogeny.fr.[6] The trees are presented with branch length ignored to facilitate clarity. Protein schematics for each DEAD-box (A, right) and DEAH-box (B, right) gene family are shown with known functional domains indicated. Genes that have been previously associated with human disease are indicated in blue text. Gene family members for which evidence is presented in this manuscript are indicated in red text. Gene family members for which only one individual is presented in this manuscript are indicated in green text.

Table 1. Observed Clinical Features by Gene and Variant

Individual	Gene	Variant	DD and/or ID (12/15)	CNS (13/15)	Seizures (6/15)	Ocular (5/15)	Ears (4/15)	Cardiac (3/15)	Kidney (4/15)	Dysmorphic Features (8/15)	Vertebra (3/15)	Other (11/15)	Sex
#1	*DHX37*	Hom p.Asn419Lys	+	+	–	–	–	–	–	–	–	None	m
#2		Comp Het p.Val731Met and p.Leu467Val	+	+	+	+	–	–	–	+	+	Asymmetric cerebellar hypoplasia, ophthalmoplegia	f
#3		Comp Het p.Arg93Gln and p.Glu167Ser	+	+	–	–	–	+	–	+	+	IVC thrombus	m
#4		*De novo* p.Asp382Gly	+	+	+	+	+	–	–	+	+	Macrodontia, aspiration	f
#5		*De novo* p.Thr1094Met	+	+	–	–	–	+	–	+	–	Tooth agenesis	m
#6	*DHX16*	*De novo* p.Phe582Ile	NA	+	+	+	–	–	–	–	–	None	f
#7		*De novo* p.Gln697His	NA	–	–	–	+	–	+	+	–	Short limbs	f
#8		*De novo* p.Gly427Glu	+	+	–	+	Deafness	–	–	–	–	Contractures, growth delay, aspiration	m
#9		*De novo* p.Thr674Met	+/–	+	+	+	Sensorineural hearing loss	–	–	–	–	Myopathy	m
#10	*DDX54*	p.Leu298Phe, *de novo* p.Asn216Ser in *cis*	+	+	+	–	–	–	–	–	–	Hand tremor	f
#11		Comp Hetp.Trp20Arg and p.Arg611Gln	+	+	–	–	–	–	–	+	–	None	f
#12		Hom p.Val286Met	–	–	–	–	–	–	+	–	–	Failure to thrive	f
#13	*DHX34*	Hom p.Asn441Ser	+	+	–	–	–	–	+	+	–	Short stature, hyperextensibility, hip dislocation	f
#14		Hom p.Gln156*	+	+	–	–	–	+	+	–	–	Polydactyly, respiratory failure, failure to thrive	m
#15		*De novo* p.Arg776Pro	+	+	+	–	–	–	–	+	–	None	f
[a]	*DDX47*	Comp Het p.Asp8Phe and p.Gln107Glu	NA	+	+	–	–	–	–	–	–	None	f
	DHX58	Hom p.Trp8Cys	+	+	+	–	–	–	–	+	–	None	f
	DHX8	Hom p.Arg168Trp	–	–	–	–	–	–	+	–	–	None	m

Genes identified through paralog study design are listed in Individuals 1–15. Abbreviations are as follows: Hom = homozygous, Comp Het = compound heterozygous, DD = developmental delay, ID = intellectual disability, CNS = central nervous system, and IVC = inferior vena cava.
[a]Three additional potential candidate genes for which a potentially etiologic variant was identified in only a single family are listed at the bottom of the table.

(ES) database was used as the cohort for primary analysis and included 9,338 individuals with ES data at the time of the study. ES was performed as previously described[18] in the Baylor College of Medicine (BCM) Human Genome Sequencing Center,[19–21] and identified variants were confirmed by orthogonal Sanger dideoxy sequencing. Individuals were consented under BCM IRB-approved protocol H-29697. All exomes for which informed consent for deposition into controlled-access databases has been provided will be deposited into dbGap under the BHCMG dbGaP Study Accession phs000711.v5.p1.

Primary Analysis

The 9,338 BHCMG ES individuals were analyzed for rare, probably damaging variants in the set of 58 DExD/H-box encoding genes. Variants were prioritized on the basis of their predicted effect on protein function, as well as minor allele frequencies across several databases including the internal BHCMG database, the National Heart, Lung, and Blood Institute (NHLBI) Grand Opportunity Exome Sequencing Project (ESP), the 1000 Genomes Project (TGP),[22] the Atherosclerosis Risk in Communities (ARIC) database,[23] the Exome Aggregation Consortium (ExAC) database, and the genome aggregation database (gnomAD), conservation (Phylop)[24] and functional prediction (PolyPhen-2, MutationTaster, Sorting Intolerant From Tolerant [SIFT], and likelihood ratio test [LRT]) algorithms,[25–28] as well as the Combined Annotation Dependent Depletion (CADD)[29] phred score. We used DNMFinder[9] and BafCalculator[30] to identify *de novo* variants from trio-ES (parents + proband) and to identify and quantify regions of absence of heterozygosity (AOH). We used HMZDelFinder[31] and XHMM (eXome-Hidden Markov Model)[32] to identify copy number variants (CNVs) from exome variant data.

Identification of Additional Families

Candidate genes were then entered into GeneMatcher[16,17] to identify additional people with similar neurodevelopmental phenotypes and probably damaging variants. This approach led to the identification of rare, predicted deleterious variants in four gene family members, *DHX37* (five people), *DHX16* (four people), *DDX54* (three people), and *DHX34* (three people) (Figure 1 and Table 1).

Additional methods details are available in the Supplemental Text.

Exome Sequencing

Exome sequencing (ES) of genomic DNA obtained from peripheral blood or saliva was performed at the Human Genome Sequencing Center (HGSC) at BCM through the BHCMG initiative, GeneDx, and the University of Washington Center for Mendelian Genomics (UWCMG). Sequencing performed at BCM used 0.5ug of DNA, and an Illumina paired-end pre-capture library was constructed according to the manufacturer's protocol (version *1005361_D*), including modifications in the BCM-HGSC protocol. Pre-captured libraries were pooled and hybridized to the HGSC VCRome 2.1 design (42Mb NimbleGen, Cat. No. 06266380001) according to the manufacturer's protocol (NimbleGen SeqCap EZ Exome Library SR User's Guide). Paired-end sequencing was performed with the Illumina HiSeq 2000 platform. The samples achieved 97% of the targeted exome bases covered to a depth of 20× or greater and had a sequencing yield of 7.7 Gb. Trio whole-exome sequencing (trio-WES) experiments and data annotation and interpretation in individuals 10 and 13 were performed with a SureSelect Human All Exon (Agilent) and sequenced on a HiSeq2500 platform (Illumina), as described before.[33]

For individual 15, exome sequencing was performed with a SureSelect Human All Exon V4 Kit 51M (Agilent) for enrichment and a HiSeq2500 (Illumina) at Macrogen, South Korea. A total of 125,956,332 101-bp reads were generated. These reads were aligned to the UCSC human reference assembly (hg19). 98% of the exome was covered at least 10×, and the average coverage was 150×. Single-nucleotide variants (SNVs) and small insertions and deletions were detected with GATK HaplotypeCaller. The detected variants were annotated by Mendelics in-house annotation software. CNVs were recomputed with the ExomeDepth R module.

Absence of Heterozygosity Calculation

To calculate segments and total areas of AOH in each known consanguineous individual, we used BafCalculator.[31] First, from all SNVs that passed quality filters in a single variant call file, we extracted the B-allele frequency (i.e., ratio of variant reads/total reads) and transformed this ratio by subtracting 0.5 and taking the absolute value for each data point. After such a transformation, values >0.47 were considered indicative of homozygous variants (expected value = 0.5) corresponding to either alternative or reference alleles, whereas lower values probably indicated heterozygous alleles. Transformed B-allele frequency data were then processed by circular binary segmentation (CBS) implemented in the DNAcopy R Bioconductor package.[34]

Phasing Variants

Primers flanking approximately 75 bp upstream and downstream of the identified variants were implemented for PCR amplification with the QIAGEN HotStart Polymerase. A standard TOPO TA cloning vector (Life Technologies) was applied to the PCR product to allow for cloning of both alleles independently. DNA was extracted from each individual clone after a 37°C overnight incubation and Sanger sequenced to detect the presence or absence of both variants in a single clone.

Droplet digital PCR (ddPCR) was performed on a Bio-Rad QX200 system via the "drop-phase" method.[35] Two independent TaqMan probes that used a distinct fluorescent tag were designed to selectively amplify only in the presence of either the first variant (FAM tag) or the second variant (HEX tag). The assay was run with a QX200 AutoDG ddPCR system from Bio-Rad according to normal protocols for a TaqMan reaction. A second reaction was generated from the same protocols but with the inclusion of a restriction enzyme (*Sph*I) that would selectively cut between the two variants. Droplet generation was performed on a QX200 AutoDG and run on a standard thermocycler under TaqMan cycling conditions. Individual positive droplet populations were quantified with the QuantaSoft software suite from Bio-Rad.

Results

In an analysis of 9,338 individuals with ES from the BHCMG database for rare, probably damaging variants in each of the 58 human DExD/H-box encoding genes, we identified seven candidate neurodevelopmental genes: *DHX37, DHX16, DDX54, DHX34, DDX47, DHX58,* and *DHX8*. Additional families identified through GeneMatcher[16,17] provided support for four of these genes

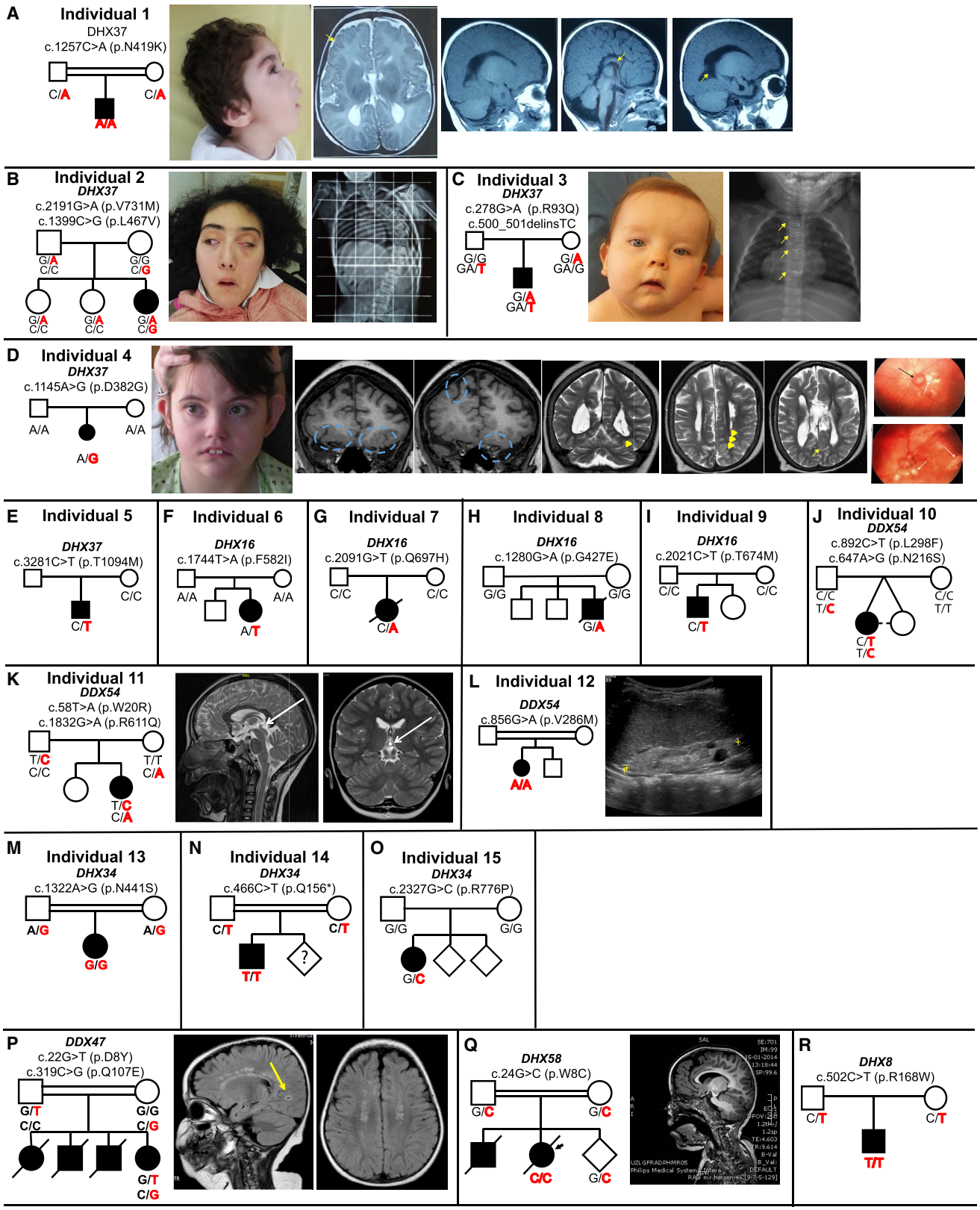


Figure 2. Individual Pedigrees and Clinical Findings

(A–O) Pedigrees, genotypes, and available consented clinical images are shown for individuals 1–15.

(A) Individual 1: arrows indicate polymicrogyria, dysgenesis of the corpus callosum, and cerebellar volume loss.

(B) Individual 2: an x-ray image shows scoliosis.

(C) Individual 3: arrows indicate segmentation anomalies of the vertebra.

(legend continued on next page)

as genes in which probably damaging variants are associated with neurodevelopmental disorders; these genes were *DHX37* (five families), *DHX16* (four families), *DDX54* (three families), and *DHX34* (three families). Three additional genes were identified in a single family each and remain potential candidate genes for which additional studies of variants are needed to further clarify their role in the observed genotype-phenotype relationships. Variants and observed clinical features are detailed in Tables 1 and S1. Our findings suggest that gene-paralog-based approaches can implicate human candidate genes in which variation might cause disease presentation.

Variants in *DHX37* Are Identified in Five Families

We identified three individuals who have bi-allelic variants in *DHX37* and who share a phenotype of DD and/or ID and central nervous system (CNS) anomalies. Two out of three individuals also exhibit vertebral anomalies. Individual 1 is homozygous, and the variant (GenBank: NM_032656.3; c.1257C>A [p.Asn419Lys]) (Tables 1 and S1) was confirmed to be inherited from unaffected carrier parents (Figure 2A). The second individual (individual 2) is an adult female who presented with DD and/or ID, congenital and non-progressive impairment of eye, eyelid, and facial movements, asymmetric cerebellar hypoplasia, seizures, and scoliosis. Individual 2 was found by ES to be compound heterozygous for *DHX37* variants (GenBank: NM_032656.3; c.2191G>A [p.Val731Met] and GenBank: NM_032656.3; c.1399C>G [p.Leu467Val]), which were confirmed to each be inherited from different unaffected parents (Figure 2B). The third individual (individual 3) presented with DD and/or ID, hypotonia, vertebral anomalies, and dysmorphic features and was found to be compound heterozygous for *DHX37* variants (GenBank: NM_032656.3; c.278G>A [p.Arg93Gln] and GenBank: NM_032656.4; c.500_501inv [p.Glu167Ala]), which were confirmed to be each inherited from different unaffected parents (Figure 2C). All of these variants are rare if present in TGP, ARIC, ExAC, or gnomAD, are present in heterozygous states in the unaffected siblings, and all but one (*DHX37* GenBank: NM_032656.3; c.2191G>A [p.Val731Met]) were deemed probably damaging by four algorithms (PolyPhen-2,[25] MutationTaster,[27] SIFT,[26] and LRT[28]) and the CADD score[29] (Table S1).

Two individuals were identified who had heterozygous variants in *DHX37*. The first individual (individual 4) had been given a clinical diagnosis of Aicardi syndrome (AIC [MIM: 304050]) and presented with DD and/or ID, brain anomalies (Supplemental Individual Report), chorioretinal lacunae, seizures, scoliosis, and dysmorphic features (Figure 2D). Trio ES identified a mosaic variant in *DHX37* (GenBank: NM_032656.3; c.1145A>G [p.Asp382Gly]). Sanger sequencing of the parents and proband confirmed that this variant occurred *de novo* (Figure 2D), and further deep sequencing confirmed a ratio of 20:80 variant:total reads (vR:tR). The second individual with a variant in *DHX37* (individual 5) presented with DD and/or ID, hypotonia, dysmorphic features, and Wolff-Parkinson White (WPW [MIM: 194200]) syndrome. Analysis of the ES data identified a *DHX37* variant (GenBank: NM_032656.3; c.3281C>T [p.Thr1094Met]), which was confirmed to not be inherited from the mother (Figure 2E); a paternal sample was unavailable for Sanger confirmation. Copy number variant analysis performed on exome variant data from individuals 4 and 5 with XHMM[36] and HMZDelFinder[31] did not identify CNVs in *DHX37*. In total, we identified five individuals who had *DHX37* variants predicted to be damaging (Figure 3) and who all share a phenotype of DD and/or ID and CNS dysfunction. Three out of five individuals also have scoliosis, and two have cardiac phenotypes (Table 1).

Variants in *DHX16* Are Identified in Four Families

We identified four individuals with *de novo* variants in *DHX16* (Figure 3). Three individuals share features of CNS anomalies and seizures (Table 1). The first individual (individual 6) is a female who has a clinical diagnosis of Aicardi syndrome and who presented with agenesis of the corpus callosum, seizures, and chorioretinal lacunae. Trio ES identified a variant in *DHX16* (GenBank: NM_003587.4; c.1744T>A [p.Phe582Ile]) (Figure 2F). The second individual (individual 7) is a female who presented with dysmorphic features and cystic kidneys; she was in the 5th percentile for length at birth and died on day of life (DOL) 16. Trio ES identified a variant in *DHX16* (GenBank: NM_003587.4; c.2091G>T [p.Gln697His]) (Figure 2G). Sanger sequencing of both trios confirmed the variants in individuals 6 and 7 to be *de novo*. The third individual (individual 8) presented with severe hypotonia and joint contractures, sensorineuronal deafness, a mixed axonal sensory and demyelinating motor neuropathy with feeding difficulties, and growth delay. He died at 4 months of age of respiratory failure. Trio-ES identified a *de novo* variant in *DHX16* (GenBank: NM_003587.4; c.1280G>A [p.Gly427Glu]) (Figure 2H). Additional samples were

(D) Individual 4: circles indicate polymicrogyria, arrowheads indicate polymicrogyria, the arrow indicates a cyst, white arrows indicate chorioretinal lacunae, and black arrows indicate optic nerve colobomas.
(K) Individual 11: white arrows indicate pineal gland cysts.
(L) Individual 12: ultrasound images show polycystic kidneys.
(P–R) Additional candidate variants were identified in single families for three genes, *DDX47* (MIM: 615428), *DHX58* (MIM: 608588), and *DHX8* (MIM: 600396), and these are indicated at the bottom of the figure.
(P) The MRI images from the individual with a candidate *DDX47* variant show occipital infarct (yellow arrow) and white matter abnormalities.
(Q)The MRI image from the individual with a candidate *DHX58* variant shows a thin corpus callosum.

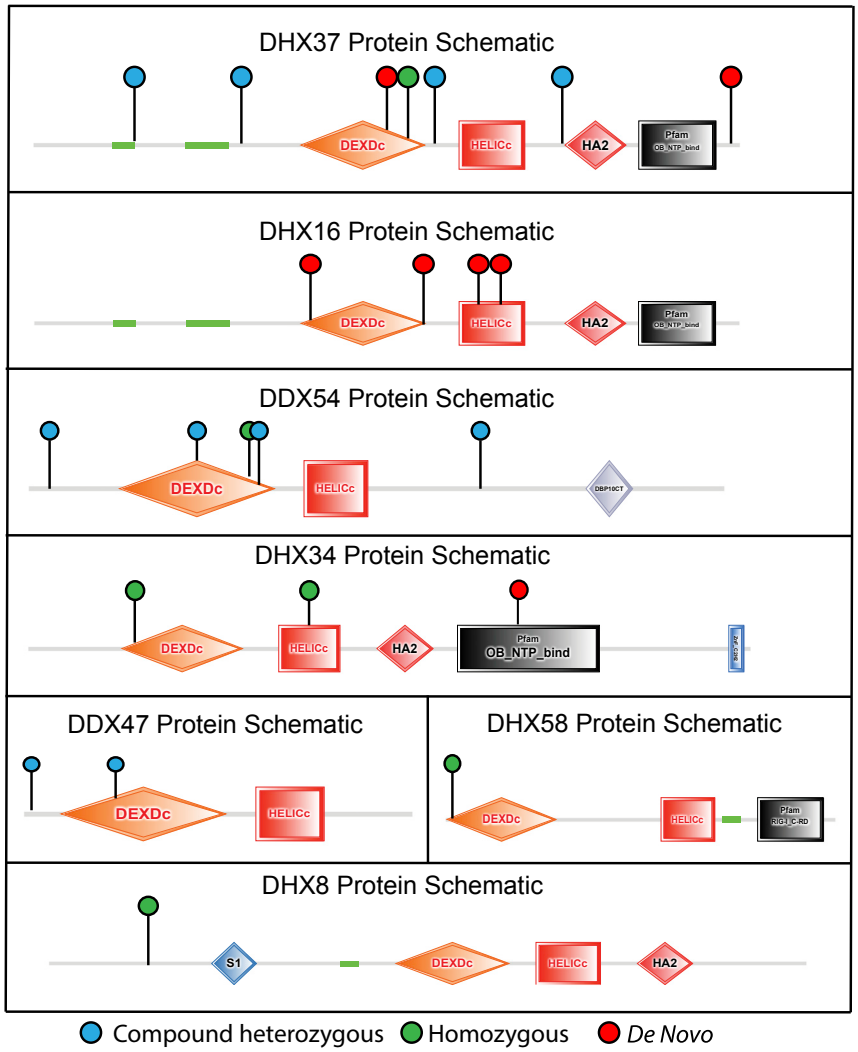


Figure 3. Variant Location Schematic
Protein schematics showing the functional domains of the seven candidate genes with the location of the variants indicated on each. A red lollipop indicates a *de novo* variant, a blue lollipop indicates compound heterozygosity for variants, and a green lollipop indicates homozygosity for the variant.

unavailable for Sanger sequencing. The fourth individual (individual 9) is a male who presented with tonic-clonic seizures, neuropathy, myopathy, developmental delay without intellectual disability, and retinopathy. Trio ES found a *de novo* variant in *DHX16* (GenBank: NM_003587.4; c.2021C>T [p.Thr674Met]) (Figure 2I). These variants are not present in TGP, ARIC, ESP, ExAC or gnomAD and are predicted to be deleterious (PolyPhen2, MutationTaster, SIFT, and LRT) (Table S1).

Variants in *DDX54* Are Identified in Three Families

We identified three individuals with *DDX54* variants: one mono-allelic, and two bi-allelic (Figure 3). Two out of three of the individuals share a phenotype of DD and/or ID and CNS anomalies, and all three are female (Table 1). The first individual (individual 10) is a twin (sister is unaffected, zygosity unknown) who is 23 years old and has DD and/or ID, dystonia, and seizures. Trio ES identified two variants in *DDX54* at position GenBank: NM_001111322.1; c.647A>G (p.Asn216Ser), confirmed to be inherited from the father (Figure 2J), and at position GenBank: NM_001111322.1; c.892C>T (p.Leu298Phe), confirmed to be *de novo*. Cloning of long-range PCR amplicons encompassing these variants and subsequent Sanger dideoxy sequencing of three independent clones demonstrated that the *de novo* variant is also present on the paternal allele in *cis* (Figure 4A). Droplet digital PCR was then used as an orthogonal confirmation of this finding, and it demonstrated that the number of droplets positive for both variants was consistent with linkage of the two variants on the same allele, resulting in co-segregation of the variants within droplets. *Sph*I genomic DNA digestion targeting a restriction site between the variant sites demonstrated a pattern consistent with unlinked variants independently segregating within droplets (Figures 4B–4E). The second individual (individual 11) presented with DD and/or ID and pineal cysts. ES analysis identified compound heterozygosity for variants in *DDX54* inherited from the unaffected father (GenBank: NM_001111322.1; c.58T>A [p.Trp20Arg]) and the unaffected mother (GenBank: NM_001111322.1; c.1832G>A [p.Arg611Gln]) (Figure 2K). The third individual (individual 12) presented with bilateral multicystic dysplastic kidneys and failure to thrive. ES analysis identified homozygosity for a *DDX54* variant (GenBank: NM_001111322.1; c.856G>A [p.Val286Met]) (Figure 2L), but parental samples were unavailable for segregation analysis.

Variants in *DHX34* Are Identified in Three Families

We identified three individuals with variants in *DHX34* (Figure 3). Individual 13 is a dysmorphic female who has DD and/or ID, CNS involvement, and a solitary kidney and a possible dual diagnosis[30] including homozygosity for an inherited *DHX34* variant (GenBank: NM_014681.5; c.1322A>G [p.Asn441Ser]) (Figure 2M) from heterozygous parents and homozygosity for a *CEP97* variant ([MIM: 615864], GenBank: NM_024548.2; c.1148A>G [p.His383Arg], Chr3: 101476598A>G);[18] both of the variants occur in blocks of AOH (of 2.7 Mb

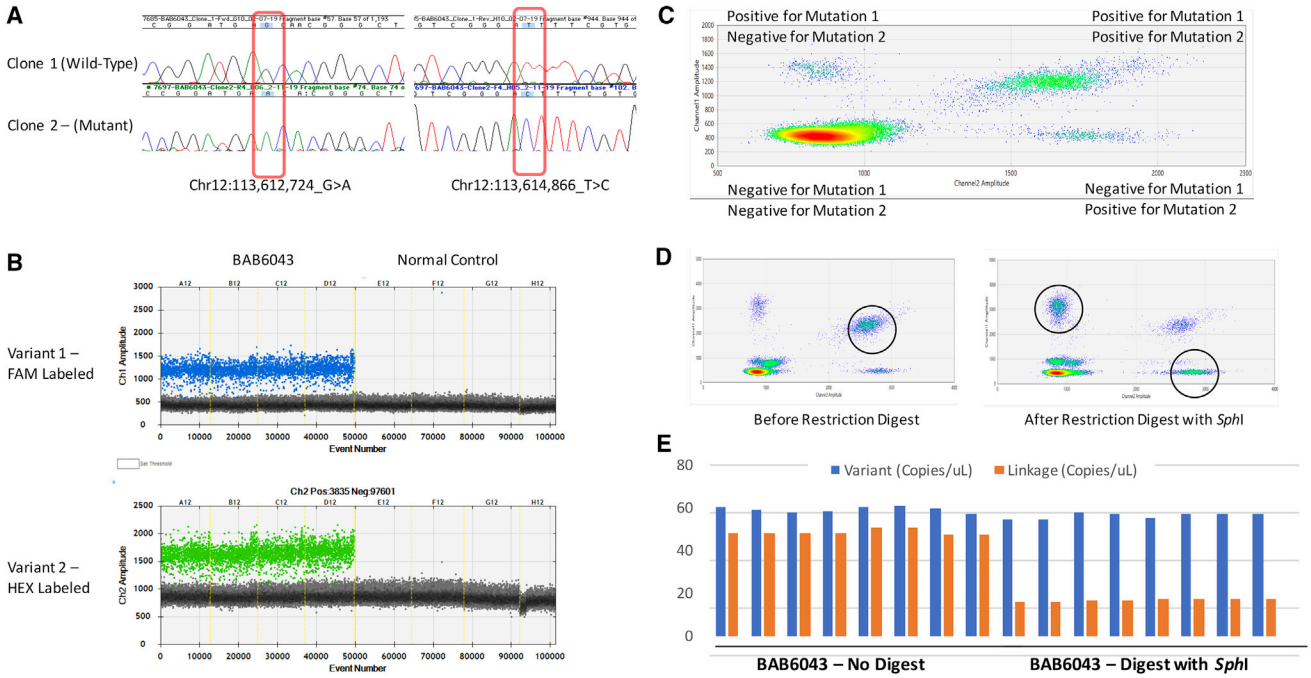


Figure 4. Variant Inheritance Confirmation

Phasing of *DDX54* variant c.892C>T (p.Leu298Phe) (Chr12: 113,612,724_G>A, paternally inherited) and c.647A>G (p.Asn216Ser) (Chr12: 113,614,866_T>C, *de novo*).

(A) Dideoxy Sanger sequencing of two clones generated from long-range PCR amplification of both variants demonstrates one clone containing neither variant and one clone containing both variants, consistent with an in *cis* variant allele configuration.

(B) Droplet digital PCR probes were designed with specificity to each variant, and optimization demonstrates that both variants are detected in individual 10, but not in a wild-type control sample.

(C) Two-dimensional droplet distribution demonstrates dense clustering in the +/+ cluster, indicating co-segregation of both variants within droplets.

(D and E) This clustering is not observed when genomic DNA is treated with *Sph*I digestion prior to ddPCR. An *Sph*I restriction site is located between the variants, and digestion leads to independent segregation of variants within droplets.

and 31.7 Mb, respectively). Individual 14 has congenital bilateral polycystic kidneys, polydactyly, hypothyroidism, failure to thrive, and severe respiratory problems, again with a probable dual diagnosis including homozygous nonsense variants in *TPO* ([MIM: 600044], GenBank: NM_000547.5; c.1618C>T [p.Arg540*]) and in *DHX34* (GenBank: NM_014681.5; c.466C>T [p.Gln156*]) (Figure 2N). Additionally, we identified a heterozygous *de novo* *DHX34* variant (GenBank: NM_014681.5; c.2327G>C [p.Arg776Pro]) in a seven-year-old female with DD and/or ID, hypotonia, and mild dysmorphism (individual 15) (Figure 2O). It is notable that all three individuals (individuals 13, 14, and 15) with rare variant alleles in *DHX34* share a core phenotype of DD and/or ID, but each display somewhat distinct additional clinical features. Individuals 13 and 14, found to have homozygosity for variants in *DHX34*, also share renal involvement, whereas only individual 15 has seizures. The extent to which these additional features can be attributed to *DHX34* alone, with distinct phenotypes attributed to mono-allelic and bi-allelic variation, is not yet fully clear; however, the renal involvement observed in association with *DHX16* (in individual 7) and *DDX54* (in individual 12) supports a potential role for DExD/H-box RNA helicase genes in renal development.

Expression and Potential Functional Impact of Variants in RNA-Helicase Genes

We queried the GTEx database for each of our candidate genes to determine their respective expression patterns (Figure S1). *DHX37*, *DHX16*, *DDX54*, and *DHX34* all share a similar ubiquitous expression pattern with some minor differences. Whereas *DHX37*, *DDX54*, and *DHX34* have the lowest expression in brain tissue compared to other tissues (Figure S1, yellow boxes), *DHX16* has increased expression specifically within the cerebellum. Additionally, data from the Allen Brain Atlas indicate that all of these genes show the highest expression in the brain during embryonic and fetal development (Figure S2).

Given that identified pathogenic variants in *DHX30* and *DDX3X* differ in their position with regard to the conserved motifs,[8] we evaluated the location of the variants described here. Although many variants occur in the established functional domains (Figure 3), the majority of variants do not occur in the conserved helicase motifs (Figure S3), a pattern similar to those seen in *DDX3X*.[10,11] There are two notable exceptions: the *de novo* *DHX16* GenBank: NM_003587.4; c.1280G>A (p.Gly427Glu) variant (in individual 8) and the *DHX34* GenBank: NM_014681.5; c.1322A>G (p.Asn441Ser) variant found in homozygosity (in individual 13). Both

of these variants lead to the substitution of an amino acid residue that is both highly conserved across paralogs and that is used to define the helicase motif of the protein in this gene family. These observations strongly suggest that they disrupt gene function. That many DExD/H-box RNA helicase family genes might play multiple cellular roles in processes such as synthesis, nuclear processing and export, translation, and the storage, splicing, and decay of RNAs, as well as ribonucleoprotein (RNP) assembly, allows for significant variation in pathogenicity and disease presentation depending on which function is affected.

Discussion

Members of paralogous gene families often share molecular functions and can have overlapping tissue expression patterns, leading us to hypothesize that paralog studies might provide an approach to gene discovery. We tested this hypothesis using the DExD/H-box RNA helicase family genes (Figure 1), of which two (*DDX3X* and *DHX30*) have been previously associated with neurodevelopmental disorders and one (*DHX38*) has recently been discovered to underlie an eye disease phenotype. Analysis of the BHCMG exome variant database and then the identification of further individuals through GeneMatcher[16,17] led to the elucidation of four neurodevelopmental genes (*DHX37*, *DHX16*, *DDX54*, *DHX34*) each found in three or more families; this also led to finding three additional individuals with potential disease-causing variants in three other genes. These findings support the utility of a paralog-based approach applied to implicate variation genes associated with human disease.

DHX37, DHX16, DDX54, and *DHX34* as Neurodevelopmental Genes

The age range of our 15 affected individuals, none of whom have received a definitive molecular diagnosis, was 1 day to 34 years. The individuals included nine females and six males. The majority (12/15) exhibit DD and/or ID phenotypes, but three were too young to assess (Table 1). Furthermore, 13/15 had involvement of the CNS, including brain anomalies and hypotonia. These two predominant neurological features, DD and/or ID and CNS dysfunction are common disease traits to all four genes (*DHX37*, *DHX16*, *DDX54*, and *DHX34*) in which presumably pathogenic variants were identified in three or more unrelated affected subjects. Interestingly, vertebral anomalies seem to be specific to individuals with variants in *DHX37* (individuals 2–4, Table 1). Two individuals with *DHX34* variants share a solitary kidney as a feature along with the DD and/or ID and CNS involvement. Of note, two other individuals (individual 7 [*DHX16* variant] and individual 12 [*DDX54* variant]) also have kidney involvement (including solitary kidney and cystic kidney[22]). *DHX16*, *DDX54,* and *DHX34* also share additional growth-associated features such as short limbs (individual 7), growth delay (individual 9), short stature (individual 13), and failure to thrive (individuals 11 and 13) (Table 1 and the Supplemental Individual Reports).

Of particular note, *DHX16* and *DHX37* were both identified in individuals (individuals 4 and 6) with a clinical diagnosis of Aicardi syndrome. Both individuals demonstrated the classic diagnostic triad of chorioretinal lacunae, infantile spasms, and agenesis of the corpus callosum, as well as other commonly-observed features of Aicardi syndrome, including nonspecific DD and/or ID, which many individuals reported here share.[37] Despite seemingly well-defined clinical features, the molecular etiology of Aicardi syndrome has not yet been definitively identified, an observation which might reflect underlying locus heterogeneity for this condition or the limited specificity of this clinical syndromic assignment. In the present study, we report *de novo* variants in *DHX37* and *DHX16*. Other individuals with rare variation in these genes demonstrate neurodevelopmental phenotypes, but do not have hallmark features of Aicardi syndrome, and it is not yet clear whether this results simply from allelic heterogeneity at the locus or other molecular mechanisms impacting genotype-phenotype relationships. Nevertheless, our observations raise the intriguing possibility that the set of disease traits that previously defined Aicardi syndrome might not include a pathognomonic feature or, potentially, that this clinical diagnosis might not have a single, molecular etiology.

Mono-allelic and Bi-allelic Variation in Three RNA Helicase Genes

The occurrence of both mono- and bi-allelic variation at three of the described loci (*DHX37, DDX54,* and *DHX34*) within this study might reveal a broader characteristic of neurodevelopmental conditions associated with DExD/H-box RNA helicase family genes. This phenomenon has been previously described for many loci,[38,39] and resulting phenotypes might be similar, such as observed in myotonia congenita (MIM: 160800, 255700) due to mono- or bi-allelic variation in *CLCN1* (MIM: 118425), or distinct, such as observed in Charcot-Marie-Tooth disease type 2B1 (CMT2B1 [MIM: 605588]) due to bi-allelic variation in *LMNA* (MIM: 150330) or Emery-Dreifuss muscular dystrophy 2 (EDMD2 [MIM: 181350]) due to mono-allelic variation in *LMNA*. In some individuals, the phenotype associated with bi-allelic variation is more severe, as has been observed in Harel-Yoon syndrome[40] (HAYOS [MIM: 617183]) due to *ATAD3A* (MIM: 612316) bi-allelic variants, and segregating in a pattern consistent with Mendelian expectations for a recessively inherited disease trait from unaffected carrier parents, whereas in others, the more severe phenotype is associated with dominantly inherited conditions, such as microphthalmia/coloboma and skeletal dysplasia syndrome (MCSKS [MIM: 615877]) associated with variation in *MAB21L2* (MIM: 604357).[41] Distinct functional outcomes for individual variant alleles (i.e., amorphs, hypomorphs, hypermorphs, antimorphs, and neomorphs),[42] and distinct locations of individual variants within protein functional domains and/or transcripts

underlie many of these observations and underscore the biological importance of understanding gene function in the context of human allelic series.

RNA Helicase Family Members and Potential Mechanisms for Molecular Pathogenesis

Variants in *DHX37* have been shown to alter normal behavior in zebrafish through a specific effect on the splicing of glycine receptors.[24] Mutant zebrafish had decreased levels of glycine receptor subunit mRNA,[43] and deletions of the glycine receptor 4 locus, *GlyRA4*, have also been linked to ID, behavior problems, and craniofacial anomalies.[44] Disruption of GlyR function by antibodies has been linked to epileptic encephalopathies and Stiff-Person syndrome (SPS [MIM: 18450]), and the majority of individuals with GlyR-abs suffer from epilepsy.[45] All of the individuals with variants in *DHX37* have CNS anomalies, and 2/5 have seizures and dysmorphic features, consistent with perturbations of GlyR function. The ocular, ear, cardiac, and vertebral findings point to the extra-CNS role of *DHX37* in development.

DHX16 is a member of the spliceosome complex B and is required for pre-mRNA splicing. A dominant-negative mutation in *DHX16* results in intron retention in transcripts that then remain in the nucleus and are not subject to nonsense-mediated decay (NMD).[46] Furthermore, unspliced genes identified in this screen are involved in neurodevelopment; these include GABA(A) receptor-associated protein (*GABARAP* [MIM: 605125]), which has been linked to hypoplastic brain development in zebrafish, and survival of motor neuron-related-splicing factor 30 (*SMNDC1* [MIM: 603519]), which is associated with spinal muscular atrophy[46] (SMA [MIM: 253300, 253400]), for which a newly developed drug, nusinersen, also targets splicing.[47] *DHX16* shows increased expression levels in the cerebellum compared to other brain regions (Figure S1). In addition to the CNS dysfunction in our cohort of individuals with *DHX16* mutations, 2/4 also have defects, including growth delay and short limbs at birth, that could be related to successful cell cycle progression (Table 1). This phenomenon has already been noted in microcephaly-affected individuals for whom the first seven loci identified (*MCPH1* [MIM: 607117], *ASPM* [MIM: 605481], *CDK5RAP2* [MIM: 608201], *CENPJ* [MIM: 609279], *STIL* [MIM: 181590], *WDR62* [MIM: 613583], and *CEP152* [MIM: 613529]) are all involved in centrosome activity.[48]

DDX54 has been shown to be essential for proper myelination in the CNS.[49] Knockdown of *DDX54* in the murine brain resulted in aberrant distribution of myelin basic protein (*MBP* [MIM: 159430]) and myelin-associated glycoprotein (*MAG* [MIM: 159460]) and the formation of filamentous MAG in the corpus callosum. It is probable that DDX54 plays an essential role in the transcription and mRNA processing of these genes.[49] DDX54 is also part of the spliceosome complex B and plays a specific role in splicing during DNA damage repair (DDR), wherein it increases cell survival by promoting transcription of DNA damage-response genes.[50] Interestingly, DDX54 has also been shown to negatively regulate nuclear-receptor-mediated transcription in a hormone-dependent manner, specifically binding to ERalpha and beta to repress transcriptional activation of estrogen-regulated genes.[51] This is particularly noteworthy given that all three of our individuals with variants in *DDX54* are female. Two of these three individuals have CNS involvement, including dystonia, seizures, and severe brain malformations, which can be directly related to DDX54's role in proper myelination. It is less clear how DDX54 might affect kidney development; however, given the similar clinical findings in one of the individuals with a variant in *DHX16* and two of the individuals with variants in *DHX34* and the positive expression of both of these genes in kidney tissue (Figure S1), it is possible these genes play an important role in kidney development and cellular differentiation.

DHX34 acts as a scaffold required for successful NMD of aberrant mRNA transcripts.[52–54] Interestingly, an NMD partner of DHX34, NBAS, has been previously associated with a human syndrome, short stature, optic nerve atrophy, and Pelger-Huet anomaly (SOPH [MIM: 614800]), that includes features of short stature, dysmorphism, optic atrophy, and leukocyte anomalies.[55] The phenotypes of individuals 12 and 13, who have homozygosity for variants in *DHX34*, include short stature, facial dysmorphism, and microcephaly, although both individuals 13 and 14 have additional variants that could be contributing to their disease (Supplemental Individual Reports).

Notably, of these four genes, only *DHX37* has a high probability of loss-of-function (LOF) intolerance ([pLI] score of 0.99 in gnomAD). *DHX16, DDX54*, and *DHX34* each have pLI scores of 0, indicating that these genes are not significantly intolerant to LOF, although it is possible that a fraction of the observed LOF variants in these genes escape NMD, leading to a gain-of-function (GOF) or dominant-negative phenotype rather than LOF.[56] In the present study, we report *de novo* missense variants in *DHX37, DHX16, DDX54*, and *DHX34* in association with neurodevelopmental disorders. These variant characteristics (missense and *de novo*), in conjunction with the dominant inheritance of their associated phenotypes, is strongly supportive of a GOF rather than LOF effect for the identified variants. Taken together with the observation of both mono- and bi-allelic variation at three of these four loci (*DHX37, DDX54,* and *DHX34*), these observations are consistent with the predicted tolerance to LOF observed in the gnomAD database.

Conclusion

Together, DExD/H-box helicase genes play common roles, namely splicing, in RNA metabolism, but they also might be involved in transcriptional regulation. They also each have unique roles in development, cell fate, and survival, and this finding provides a potential explanation for the diversity of phenotypes exhibited by this cohort. Functional redundancy of paralogs is well established as a hypothesis

for enabling deleterious mutations to persist in a population. More recently, it has been shown that functional compensation coupled with tissue-specific expression patterns can result in tissue-specific disease in ubiquitously expressed genes, such as DExD/H-box helicases. Indeed, this was demonstrated in this gene family in human cell lines for *DDX3X* and *DDX3Y* (MIM: 400010).[57] In fact, 69% of tissue-specific, inherited disease could be explained by differential expression of paralogs in the affected tissue.[1] This might explain the occurrence of different disease phenotypes, both among individuals with variants in the same DExD/H-box helicase genes and also between individuals with the same disease phenotypes and variants in different DExD/H-box helicase genes, both of which we describe here. The increasing evidence associating RNA binding proteins (RBPs) of different types to neuropsychiatric disorders might point to a global mechanism for the pathogenesis of ID in humans and support an analogous 'gene paralogy' approach for ID-associated gene discovery.

Given the potential role for paralogous genes in human disease, we utilized a 'gene first' approach[58,59] and searched the BHCMG database for *de novo* and bi-allelic variants in DExD/H-box RNA helicases. We further amplified our resulting cohort through the use of GeneMatcher.[16,17] This approach has enabled a connection of similar phenotypes across genes; for example, we identified two individuals with a clinical diagnosis of Aicardi syndrome, one with a variant in *DHX16* and one with a variant in *DHX37*, and four individuals with congenital anomalies of the kidneys and urinary tract (CAKUT). Although a 'phenotype first' approach is valuable in large cohorts or when mining electronic health records,[60] in the case of gene discovery in paralog gene families, it could have easily missed the connection between these individuals. The genes for the remaining DExD/H-box helicase family members, as well as other RNA-processing genes, might harbor pathogenic variation associated with Mendelian conditions, particularly neurodevelopmental disorder phenotypes. Additionally, a paralog approach to gene discovery can widely be applied for other disease phenotypes already associated with known genes.

Accession Numbers

The dbGaP accession number for all exome sequences reported in this paper and for which informed consent for data sharing in controlled-access databases has been provided is dbGaP: phs000711.v5.p1.

Supplemental Data

Supplemental Data can be found online at https://doi.org/10.1016/j.ajhg.2019.06.001.

Acknowledgments

This work was supported in part by grants UM1 HG006542 (J.R.L) and UM1 HG006493 (M.B.) from the National Human Genome Research Institute (NHGRI) and the National Heart, Lung, and Blood Institute (NHLBI) to the Baylor Hopkins Center for Mendelian Genomics and the University of Washington Center for Mendelian Genomics, R01 NS058529 and R35 NS105078(J.R.L.) from the National Institute of Neurological Disorders and Stroke (NINDS), U54-HG003273 (R.A.G.) from NHGRI, and Telethon Undiagnosed Diseases Program (TUDP) GSP15001 (N.B.-P.) from the Telethon Foundation, and also by the Aicardi Syndrome Foundation. I.P. was supported by 2T32NS043124-16 through the National Institutes of Health. J.E.P. was supported by NHGRI K08 HG008986. F.H. was supported by the National Institutes of Health (DK088767). M.R.B. was supported by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) K12 DK083014. D.L was supported by the Werner Otto Stiftung and the German Research Foundation (DFG; LE 4223/1). The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by the National Cancer Institute, NHGRI, NHLBI, the National Institute on Drug Abuse, the National Institute of Mental Health, and NINDS. The data used for the analyses described in this manuscript were obtained from the GTEx Portal on 10/29/18. The authors would like to thank Hans-Jurgen Kreienkamp for the help in identifying helicase core motifs and the Genome Aggregation Database (gnomAD) and the groups that provided exome and genome variant data to this resource. A full list of contributing groups can be found at https://gnomad.broadinstitute.org/about.

Declaration of Interests

J.R.L. serves on the scientific advisory board for Baylor Genetics. J.R.L. has stock ownership in 23andMe, is a paid consultant for Regeneron Pharmaceuticals, and is a coinventor on multiple US and European patents related to molecular diagnostics for inherited neuropathies, eye diseases, and bacterial genomic fingerprinting.

Received: December 10, 2018
Accepted: May 31, 2019
Published: June 27, 2019

Web Resources

Allen Brain Atlas, http://portal.brain-map.org/
Baylor College of Medicine Human Genome Sequencing Center, https://www.hgsc.bcm.edu
Baylor College of Medicine Lupski Lab, https://github.com/BCM-Lupskilab
ExAC Browser, http://exac.broadinstitute.org/
Exome Variant Server, https://evs.gs.washington.edu/EVS/
gnomAD Browser, https://gnomad.broadinstitute.org/
Online Mendelian Inheritance in Man, https://OMIM.org
Phylogeny.fr, http://phylogeny.lirmm.fr/phylo_cgi/index.cgi
SMART, http://SMart.embl-heidelberg.de

References

1. Barshir, R., Hekselman, I., Shemesh, N., Sharon, M., Novack, L., and Yeger-Lotem, E. (2018). Role of duplicate genes in determining the tissue-selectivity of hereditary diseases. PLoS Genet. *14*, e1007327.
2. Beighton, P., De Paepe, A., Steinmann, B., Tsipouras, P., Wenstrup, R.J.; and Ehlers-Danlos National Foundation

(USA) and Ehlers-Danlos Support Group (UK) (1998). Ehlers-Danlos syndromes: Revised nosology, Villefranche, 1997. Am. J. Med. Genet. *77*, 31–37.
3. Malfait, F., Francomano, C., Byers, P., Belmont, J., Berglund, B., Black, J., Bloom, L., Bowen, J.M., Brady, A.F., Burrows, N.P., et al. (2017). The 2017 international classification of the Ehlers-Danlos syndromes. Am. J. Med. Genet. C. Semin. Med. Genet. *175*, 8–26.
4. Gonzaga-Jauregui, C., Gamble, C.N., Yuan, B., Penney, S., Jhangiani, S., Muzny, D.M., Gibbs, R.A., Lupski, J.R., and Hecht, J.T. (2015). Mutations in *COL27A1* cause Steel syndrome and suggest a founder mutation effect in the Puerto Rican population. Eur. J. Hum. Genet. *23*, 342–346.
5. Yamamoto, S., Jaiswal, M., Charng, W.L., Gambin, T., Karaca, E., Mirzaa, G., Wiszniewski, W., Sandoval, H., Haelterman, N.A., Xiong, B., et al. (2014). A *drosophila* genetic resource of mutants to study mechanisms underlying human genetic diseases. Cell *159*, 200–214.
6. Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., Dufayard, J.F., Guindon, S., Lefort, V., Lescot, M., et al. (2008). Phylogeny.fr: Robust phylogenetic analysis for the non-specialist. Nucleic Acids Res. *36*, W465-9.
7. Chen, J., Zhang, Y., Liu, J., Xia, M., Wang, W., and Shen, F. (2014). Genome-wide analysis of the RNA helicase gene family in *Gossypium raimondii*. Int. J. Mol. Sci. *15*, 4635–4656.
8. Lessel, D., Schob, C., Küry, S., Reijnders, M.R.F., Harel, T., Eldomery, M.K., Coban-Akdemir, Z., Denecke, J., Edvardson, S., Colin, E., et al.; DDD study; and C4RCD Research Group (2017). *De novo* missense mutations in *DHX30* impair global translation and cause a neurodevelopmental disorder. Am. J. Hum. Genet. *101*, 716–724.
9. Eldomery, M.K., Coban-Akdemir, Z., Harel, T., Rosenfeld, J.A., Gambin, T., Stray-Pedersen, A., Küry, S., Mercier, S., Lessel, D., Denecke, J., et al. (2017). Lessons learned from additional research analyses of unsolved clinical exome individuals. Genome Med. *9*, 1–15.
10. Wang, X., Posey, J.E., Rosenfeld, J.A., Bacino, C.A., Scaglia, F., Immken, L., Harris, J.M., Hickey, S.E., Mosher, T.M., Slavotinek, A., et al.; Undiagnosed Diseases Network (2018). Phenotypic expansion in *DDX3X* - a common cause of intellectual disability in females. Ann. Clin. Transl. Neurol. *5*, 1277–1285.
11. Snijders Blok, L., Madsen, E., Juusola, J., Gilissen, C., Baralle, D., Reijnders, M.R.F., Venselaar, H., Helsmoortel, C., Cho, M.T., Hoischen, A., et al.; DDD Study (2015). Mutations in *DDX3X* are a common cause of unexplained intellectual disability with gender-specific effects on Wnt signaling. Am. J. Hum. Genet. *97*, 343–352.
12. Posey, J.E., Harel, T., Liu, P., Rosenfeld, J.A., James, R.A., Coban Akdemir, Z.H., Walkiewicz, M., Bi, W., Xiao, R., Ding, Y., et al. (2017). Resolution of disease phenotypes resulting from multilocus genomic variation. N. Engl. J. Med. *376*, 21–31.
13. Dikow, N., Granzow, M., Graul-Neumann, L.M., Karch, S., Hinderhofer, K., Paramasivam, N., Behl, L.-J., Kaufmann, L., Fischer, C., Evers, C., et al. (2017). *DDX3X* mutations in two girls with a phenotype overlapping Toriello-Carey syndrome. Am. J. Med. Genet. A. *173*, 1369–1373.
14. Ajmal, M., Khan, M.I., Neveling, K., Khan, Y.M., Azam, M., Waheed, N.K., Hamel, C.P., Ben-Yosef, T., De Baere, E., Koenekoop, R.K., et al. (2014). A missense mutation in the splicing factor gene DHX38 is associated with early-onset retinitis pigmentosa with macular coloboma. J. Med. Genet. *51*, 444–448.
15. Latif, Z., Chakchouk, I., Schrauwen, I., Lee, K., Santos-Cortez, R.L.P., Abbe, I., Acharya, A., Jarral, A., Ali, I., Ullah, E., et al. (2018). Confirmation of the role of DHX38 in the etiology of early-onset retinitis pigmentosa. IOVS *59*, 4552–4557.
16. Sobreira, N., Schiettecatte, F., Boehm, C., Valle, D., and Hamosh, A. (2015). New tools for Mendelian disease gene identification: PhenoDB variant analysis module; and GeneMatcher, a web-based tool for linking investigators with an interest in the same gene. Hum. Mutat. *36*, 425–431.
17. 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.
18. Karaca, E., Harel, T., Pehlivan, D., Jhangiani, S.N., Gambin, T., Coban Akdemir, Z., Gonzaga-Jauregui, C., Erdin, S., Bayram, Y., Campbell, I.M., et al. (2015). Genes that affect brain structure and function identified by rare variant analyses of Mendelian neurologic disease. Neuron *88*, 499–513.
19. Reid, J.G., Carroll, A., Veeraraghavan, N., Dahdouli, M., Sundquist, A., English, A., Bainbridge, M., White, S., Salerno, W., Buhay, C., et al. (2014). Launching genomics into the cloud: Deployment of Mercury, a next generation sequence analysis pipeline. BMC Bioinformatics *15*, 30.
20. Bainbridge, M.N., Wang, M., Wu, Y., Newsham, I., Muzny, D.M., Jefferies, J.L., Albert, T.J., Burgess, D.L., and Gibbs, R.A. (2011). Targeted enrichment beyond the consensus coding DNA sequence exome reveals exons with higher variant densities. Genome Biol. *12*, R68.
21. Challis, D., Yu, J., Evani, U.S., Jackson, A.R., Paithankar, S., Coarfa, C., Milosavljevic, A., Gibbs, R.A., and Yu, F. (2012). An integrative variant analysis suite for whole exome next-generation sequencing data. BMC Bioinformatics *13*, 8.
22. Abecasis, G.R., Altshuler, D., Auton, A., Brooks, L.D., Durbin, R.M., Gibbs, R.A., Hurles, M.E., McVean, G.A., De La Vega, F.M., Donnelly, P., et al.; 1000 Genomes Project Consortium (2010). A map of human genome variation from population-scale sequencing. Nature *467*, 1061–1073.
23. (1989). The Atherosclerosis Risk in Communities (ARIC) Study: Design and objectives. The ARIC investigators. Am. J. Epidemiol. *129*, 687–702.
24. Pollard, K.S., Hubisz, M.J., Rosenbloom, K.R., and Siepel, A. (2010). Detection of nonneutral substitution rates on mammalian phylogenies. Genome Res. *20*, 110–121.
25. Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S., and Sunyaev, S.R. (2010). A method and server for predicting damaging missense mutations. Nat. Methods *7*, 248–249.
26. Ng, P.C., and Henikoff, S. (2003). SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Res. *31*, 3812–3814.
27. Schwarz, J.M., Cooper, D.N., Schuelke, M., and Seelow, D. (2014). MutationTaster2: Mutation prediction for the deep-sequencing age. Nat. Methods *11*, 361–362.
28. Chun, S., and Fay, J.C. (2009). Identification of deleterious mutations within three human genomes. Genome Res. *19*, 1553–1561.
29. Kircher, M., Witten, D.M., Jain, P., O'Roak, B.J., Cooper, G.M., and Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. Nat. Genet. *46*, 310–315.
30. Karaca, E., Posey, J.E., Coban Akdemir, Z., Pehlivan, D., Harel, T., Jhangiani, S.N., Bayram, Y., Song, X., Bahrambeigi, V.,

Yuregir, O.O., et al. (2018). Phenotypic expansion illuminates multilocus pathogenic variation. Genet. Med. *20*, 1528–1537.
31. Gambin, T., Akdemir, Z.C., Yuan, B., Gu, S., Chiang, T., Carvalho, C.M.B., Shaw, C., Jhangiani, S., Boone, P.M., Eldomery, M.K., et al. (2017). Homozygous and hemizygous CNV detection from exome sequencing data in a Mendelian disease cohort. Nucleic Acids Res. *45*, 1633–1648.
32. Fromer, M., Moran, J.L., Chambert, K., Banks, E., Bergen, S.E., Ruderfer, D.M., Handsaker, R.E., McCarroll, S.A., O'Donovan, M.C., Owen, M.J., et al. (2012). Discovery and statistical genotyping of copy-number variation from whole-exome sequencing depth. Am. J. Hum. Genet. *91*, 597–607.
33. Hempel, M., Cremer, K., Ockeloen, C.W., Lichtenbelt, K.D., Herkert, J.C., Denecke, J., Haack, T.B., Zink, A.M., Becker, J., Wohlleber, E., et al. (2015). De novo mutations in CHAMP1 cause intellectual disability with severe speech impairment. Am. J. Hum. Genet. *97*, 493–500.
34. Olshen, A.B., Venkatraman, E.S., Lucito, R., and Wigler, M. (2004). Circular binary segmentation for the analysis of array-based DNA copy number data. Biostatistics *5*, 557–572.
35. Regan, J.F., Kamitaki, N., Legler, T., Cooper, S., Klitgord, N., Karlin-Neumann, G., Wong, C., Hodges, S., Koehler, R., Tzonev, S., and McCarroll, S.A. (2015). A rapid molecular approach for chromosomal phasing. PLoS ONE *10*, e0118270.
36. Miyatake, S., Koshimizu, E., Fujita, A., Fukai, R., Imagawa, E., Ohba, C., Kuki, I., Nukui, M., Araki, A., Makita, Y., et al. (2015). Detecting copy-number variations in whole-exome sequencing data using the eXome Hidden Markov Model: An 'exome-first' approach. J. Hum. Genet. *60*, 175–182.
37. Sutton, V.R., Hopkins, B.J., Eble, T.N., Gambhir, N., Lewis, R.A., and Van Den Veyver, I.B. (2005). Facial and physical features of Aicardi syndrome: Infants to teenagers. Am. J. Med. Genet. *138 A*, 254–258.
38. Harel, T., Yesil, G., Bayram, Y., Coban-Akdemir, Z., Charng, W.-L., Karaca, E., Al Asmari, A., Eldomery, M.K., Hunter, J.V., Jhangiani, S.N., et al.; Baylor-Hopkins Center for Mendelian Genomics (2016). Monoallelic and biallelic variants in *EMC1* identified in individuals with global developmental delay, hypotonia, scoliosis, and cerebellar atrophy. Am. J. Hum. Genet. *98*, 562–570.
39. 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.; Centers for Mendelian Genomics (2019). Insights into genetics, human biology and disease gleaned from family based genomic studies. Genet. Med. *21*, 798–812.
40. Harel, T., Yoon, W.H., Garone, C., Gu, S., Coban-Akdemir, Z., Eldomery, M.K., Posey, J.E., Jhangiani, S.N., Rosenfeld, J.A., Cho, M.T., et al.; Baylor-Hopkins Center for Mendelian Genomics; and University of Washington Center for Mendelian Genomics (2016). Recurrent *de novo* and biallelic variation of *ATAD3A,* encoding a mitochondrial membrane protein, results in distinct neurological syndromes. Am. J. Hum. Genet. *99*, 831–845.
41. Rainger, J., Pehlivan, D., Johansson, S., Bengani, H., Sanchez-Pulido, L., Williamson, K.A., Ture, M., Barker, H., Rosendahl, K., Spranger, J., et al.; UK10K; and Baylor-Hopkins Center for Mendelian Genomics (2014). Monoallelic and biallelic mutations in *MAB21L2* cause a spectrum of major eye malformations. Am. J. Hum. Genet. *94*, 915–923.
42. Muller, H.J. (1932). Further studies on the nature and causes of gene mutations. Proc. 6th Int. Congr. Genet., 213–255.
43. Hirata, H., Ogino, K., Yamada, K., Leacock, S., and Harvey, R.J. (2013). Defective escape behavior in DEAH-box RNA helicase mutants improved by restoring glycine receptor expression. J. Neurosci. *33*, 14638–14644.
44. 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.
45. Swayne, A., Tjoa, L., Broadley, S., Dionisio, S., Gillis, D., Jacobson, L., Woodhall, M.R., McNabb, A., Schweitzer, D., Tsang, B., et al. (2018). Antiglycine receptor antibody related disease: A case series and literature review. Eur. J. Neurol. *25*, 1290–1298.
46. Gencheva, M., Lin, T.Y., Wu, X., Yang, L., Richard, C., Jones, M., Lin, S.B., and Lin, R.J. (2010). Nuclear retention of unspliced pre-mRNAs by mutant DHX16/hPRP2, a spliceosomal DEAH-box protein. J. Biol. Chem. *285*, 35624–35632.
47. Chiriboga, C.A., Swoboda, K.J., Darras, B.T., Iannaccone, S.T., Montes, J., De Vivo, D.C., Norris, D.A., Bennett, C.F., and Bishop, K.M. (2016). Results from a phase 1 study of nusinersen (ISIS-SMN(Rx)) in children with spinal muscular atrophy. Neurology *86*, 890–897.
48. Gilmore, E.C., and Walsh, C.A. (2013). Genetic causes of microcephaly and lessons for neuronal development. Wiley Interdiscip. Rev. Dev. Biol. *2*, 461–478.
49. Zhan, R., Yamamoto, M., Ueki, T., Yoshioka, N., Tanaka, K., Morisaki, H., Seiwa, C., Yamamoto, Y., Kawano, H., Tsuruo, Y., et al. (2013). A DEAD-box RNA helicase Ddx54 protein in oligodendrocytes is indispensable for myelination in the central nervous system. J. Neurosci. Res. *91*, 335–348.
50. Milek, M., Imami, K., Mukherjee, N., Bortoli, F., Zinnall, U., Hazapis, O., Trahan, C., Oeffinger, M., Heyd, F., Ohler, U., et al. (2017). DDX54 regulates transcriptome dynamics during DNA damage response. Genome Res. *27*, 1344–1359.
51. Rajendran, R.R., Nye, A.C., Frasor, J., Balsara, R.D., Martini, P.G.V., and Katzenellenbogen, B.S. (2003). Regulation of nuclear receptor transcriptional activity by a novel DEAD box RNA helicase (DP97). J. Biol. Chem. *278*, 4628–4638.
52. Melero, R., Hug, N., López-Perrote, A., Yamashita, A., Cáceres, J.F., and Llorca, O. (2016). The RNA helicase DHX34 functions as a scaffold for SMG1-mediated UPF1 phosphorylation. Nat. Commun. *7*, 10585.
53. Hug, N., and Cáceres, J.F. (2014). The RNA helicase DHX34 activates NMD by promoting a transition from the surveillance to the decay-inducing complex. Cell Rep. *8*, 1845–1856.
54. Longman, D., Hug, N., Keith, M., Anastasaki, C., Patton, E.E., Grimes, G., and Cáceres, J.F. (2013). DHX34 and NBAS form part of an autoregulatory NMD circuit that regulates endogenous RNA targets in human cells, zebrafish and *Caenorhabditis elegans*. Nucleic Acids Res. *41*, 8319–8331.
55. Maksimova, N., Hara, K., Nikolaeva, I., Chun-Feng, T., Usui, T., Takagi, M., Nishihira, Y., Miyashita, A., Fujiwara, H., Oyama, T., et al. (2010). Neuroblastoma amplified sequence gene is associated with a novel short stature syndrome characterised by optic nerve atrophy and Pelger-Huët anomaly. J. Med. Genet. *47*, 538–548.
56. Coban-Akdemir, Z., White, J.J., Song, X., Jhangiani, S.N., Fatih, J.M., Gambin, T., Bayram, Y., Chinn, I.K., Karaca, E., Punetha, J., et al.; Baylor-Hopkins Center for Mendelian Genomics (2018). Identifying genes whose mutant transcripts cause dominant disease traits by potential gain-of-function alleles. Am. J. Hum. Genet. *103*, 171–187.

57. Wang, T., Birsoy, K., Hughes, N.W., Krupczak, K.M., Post, Y., Wei, J.J., Lander, E.S., and Sabatini, D.M. (2015). Identification and characterization of essential genes in the human genome. Science *350*, 1096–1101.
58. White, J., Beck, C.R., Harel, T., Posey, J.E., Jhangiani, S.N., Tang, S., Farwell, K.D., Powis, Z., Mendelsohn, N.J., Baker, J.A., et al. (2016). *POGZ* truncating alleles cause syndromic intellectual disability. Genome Med. *8*, 3.
59. Bostwick, B.L., McLean, S., Posey, J.E., Streff, H.E., Gripp, K.W., Blesson, A., Powell-Hamilton, N., Tusi, J., Stevenson, D.A., Farrelly, E., et al.; Members of the Undiagnosed Diseases Network (2017). Phenotypic and molecular characterisation of *CDK13*-related congenital heart defects, dysmorphic facial features and intellectual developmental disorders. Genome Med. *9*, 73.
60. Blair, D.R., Lyttle, C.S., Mortensen, J.M., Bearden, C.F., Jensen, A.B., Khiabanian, H., Melamed, R., Rabadan, R., Bernstam, E.V., Brunak, S., et al. (2013). A nondegenerate code of deleterious variants in Mendelian loci contributes to complex disease risk. Cell *155*, 70–80.

The American Journal of Human Genetics, Volume *105*

Supplemental Data

Paralog Studies Augment Gene Discovery:

DDX and *DHX* Genes

Ingrid Paine, Jennifer E. Posey, Christopher M. Grochowski, Shalini N. Jhangiani, Sarah Rosenheck, Robert Kleyner, Taylor Marmorale, Margaret Yoon, Kai Wang, Reid Robison, Gerarda Cappuccio, Michele Pinelli, Adriano Magli, Zeynep Coban Akdemir, Joannie Hui, Wai Lan Yeung, Bibiana K.Y. Wong, Lucia Ortega, Mir Reza Bekheirnia, Tatjana Bierhals, Maja Hempel, Jessika Johannsen, René Santer, Dilek Aktas, Mehmet Alikasifoglu, Sevcan Bozdogan, Hatip Aydin, Ender Karaca, Yavuz Bayram, Hadas Ityel, Michael Dorschner, Janson J. White, Ekkehard Wilichowski, Saskia B. Wortmann, Erasmo B. Casella, Joao Paulo Kitajima, Fernando Kok, Fabiola Monteiro, Donna M. Muzny, Michael Bamshad, Richard A. Gibbs, V. Reid Sutton, University of Washington Center for Mendelian Genomics, Baylor-Hopkins Center for Mendelian Genomics, Telethon Undiagnosed Diseases Program, Hilde Van Esch, Nicola Brunetti-Pierri, Friedhelm Hildebrandt, Ariel Brautbar, Ignatia B. Van den Veyver, Ian Glass, Davor Lessel, Gholson J. Lyon, and James R. Lupski

Supplemental Note: Individual Reports

Individual 1 was previously published in Karaca *et al.,* Neuron 2015,[6] and presented with severe DD/ID, microcephaly, dysgenesis of the corpus callosum and hypotonia. A homozygous variant in *DHX37* (NM_032656.3:c.1257C>A: p.Asn419Lys:Chr12:125453449 C>A) was proposed as the candidate causal variant at that time. This patient was born to unaffected consanguineous parents. Brain MRI revealed cortical volume loss, severe dysgenesis of the corpus callosum, colpocephaly, delayed myelination, thickened sulci, polymicrogyria, and cerebellar volume loss (Figure 2A).

Individual 2 was obtained through GeneMatcher and was seen at Federico II University Hospital in Naples and enrolled in the Telethon Undiagnosed Diseases Program (TUDP) for ES. This patient is a 19-year-old female who presented with severe DD/ID, asymmetric cerebellar hypoplasia, generalized tonic-clonic seizures, ophthalmoplegia, and severe scoliosis (Figure 2B). She was born after an uncomplicated pregnancy and she was first evaluated at 5 moths of life for developmental delay and hypotonia. At 2 years of age she was found to have strabismus, bilateral ptosis, nystagmus but no changes at the eye fundus were observed. At 6 years, she developed generalized seizures and muscle spasms that were partially controlled by anti-epileptic drugs. She never walked and she was wheelchair-bound, she was not able to manipulate objects and she did not develop any language. She had severe scoliosis and dislocation of the right hip. She had bilateral ophthalmoplegia and blindness. She underwent repeated brain MRI that showed hypoplasia of the right cerebellar hemisphere and asymmetric posterior cranial fossa. The MRI of the orbits revealed hypoplasia of extraocular muscles including superior, medial and inferior rectus. Visual evoked potentials were markedly altered.

Brainstem auditory potentials and nerve conduction study were normal. The EEG showed signs of epileptic activities in the left occipital region. Cardiac and abdominal ultrasounds were normal. Before ES, she underwent multiple metabolic and genetic testing including levels of lactate, amino acids and acylcarnitine in blood, urinary organic acids, array-CGH, and Sanger sequencing of *KIF21A*, *TUBB3*, *ROBO3* and *SLC35A3* which were all negative. Trio ES identified compound heterozygous variants in *DHX37* (NM_032656.3:c.2191G>A;p.Val731Met and NM_032656.3:c.1399C>G;p.Leu467Val) which were inherited from each unaffected parent. The two unaffected sisters were found to carry only the c.2191G>A variant.

Individual 3 was obtained through GeneMatcher and was seen at Cook Children's Hospital in Fort Worth, TX. This patient first presented at six weeks gestation when no heart beat was detected on fetal ultrasound. Ultimately, the birth was complicated by bicornuate uterus and breach presentation. The patient was delivered at 39 weeks gestation by C-section with no complications. Perimembranous VSD and secundum ASD were identified after birth and were corrected surgically at 6 months of age. Skeletal survey revealed segmentation anomalies at T3, and L5, and 13 pairs of ribs. C3 and C4 vertebral bodies were found to be flattened and dysplastic. Abdominal ultrasound showed a chronic mural thrombus. This individual also presented with plagiocephaly which was treated with a helmet. He is also mildly dysmorphic with facial asymmetry, telecanthus, and epicanthal folds, and pectus excavatum (Figure 2C). At one year old he was also found to be short for his age (30.51 in, 3rd percentile) with a head circumference in the 34th percentile (48 cm). While noted to have hypotonia, he was able to roll over.

Individual 4 is a female diagnosed with Aicardi syndrome. The pregnancy was reported to be normal; however enlarged brain ventricles were documented in the fetus in a 3D ultrasound during the second trimester. After an unremarkable vaginal birth, infantile spasms developed but were not diagnosed until 2 months of age. An MRI at 2 days of life revealed complete agenesis of the corpus callosum, and an ophthalmologic exam at 2 months revealed bilateral chorioretinal lacunae and optic nerve coloboma in the right eye. This patient also presented with ventriculomegaly, colpocephaly, cerebellar dysplasia, intracranial cysts, polymicrogyria, heterotopia, choreoathetosis, abnormal neural migration, silent aspiration requiring G-tube placement, and severe DD/ID. At baseline, she presented with flat affect, consisting of no communication methods (both verbal and physical), with poor executive functioning. She was generally flaccid in tone and her only means of movement was rolling, since she cannot sit up independently or stand. She also presented with scoliosis at 4 years and 10 months of age involving T8 to L4 with a 62 degree angle, impacting respiratory function. She was treated with a VEPTR rod placement until age 10 at which point spinal fusion was performed from T10-L5 along the thoracic and lumbar nerves. She also presented with dysmorphic features including a short philtrum, large eyes, anteverted nares, macrodontia, downturned corners of the mouth, protruding ears with abnormal pinna and a broad neck (Figure 2D). Interestingly, this patient had precocious puberty, growth attenuation and advanced bone age. At age 9 years and 5 months, her bone age was measured to be 15 years. Her height has not changed from 52 inches at 9 years old. In addition to the complete agenesis of the corpus callosum, the proband also has other abnormalities present on the most recent brain MRI which was taken at age 10.

vetriculomegaly is still present on the MRI along with colpocephaly demonstrated by disproportionate enlargement of the occipital horns of the lateral ventricles. There is also dysgenesis of the cerebellum creating an enlarged 3rd and 4th ventricle. There also appears to be some asymmetry between the right and left hemispheres, subependymal gray matter, polymicrogyria, and a small interhemispheric cyst.

Trio ES identified a mosaic variant in *DXH37* at position NM_032656.3:c.1145A>G:p.Asp382Gly (Chr12:125455894T>C). Deep sequencing using MiSeq yielded 1 million reads at this locus (247388 called a C and 745504 called a T), giving an approximate C to T ratio of confirmed 20:80 vR:tR.

Individual 5 was originally seen at Cook Children's Hospital in Fort Worth, Texas at 15 months old. He presented with hypotonia and developmental delay. At that time, he babbled but had no words, and did not yet pull up to standing. A physical exam found him to be small for his age in both height (78.1 cm, 10th centile) and weight (4.6 kg, 9th centile), with an FOC of 45.9 cm (14.5th centile) but proportional. The head showed mild plagiocephaly and flattening of the right occipital region. He had undergone testing for Prader-Willi and Angelman syndromes, fragile X, and myotonic dystrophy which were all normal. The initial ES report indicated a *de novo* heterozygous variant of unknown significance in *SETBP1* (c.2621A>G; p.Asn874Gly). The patient was eventually enrolled in the BHCMG program at Baylor College of Medicine due to the non-diagnostic ES result. Examination at the age of 10 years revealed developmental delay, speech delay, dysmorphic features including hypertelorism, a pointed chin and midface hypoplasia, hypotonia, tooth agenesis, hypoplastic nails and Wolff-Parkinson White syndrome.

ES data identified a variant in *DHX37* NM_032656.3:c.3281C>T:p.Thr1094Met (Chr12:125434541G>A).

Individual 6 is a female who presented with infantile spasms at 10 weeks of age following an uneventful birth. A brain MRI performed at 2 months of age revealed agenesis of the corpus callosum and subependymal heterotopia. At five months of age the head circumference was 15 ¾ cm (10-25th percentile). At this time, she showed poor head control, hypotonia, poor visual tracking, and slow social smile. Ophthalmologic exam showed chorioretinal lacunae, and areas of depigmentation around the optic nerve. Given the presence of the cardinal features, she was given a clinical diagnosis of Aicardi syndrome. She was enrolled in the BHCMG and received Trio ES on a research basis. Trio ES identified a variant in *DHX16* NM_003587.4:c.1744T>A:p.Phe582Ile (Chr6:30627820 A>T).

Individual 7 was a female infant born by emergency C-section at 32 weeks gestation due to non-reassuring fetal heart tones. The birth measurements showed the proband to be below the fifth percentile for length (37 cm) and limbs appeared short in comparison to the body. The proband also presented with bilateral epicanthal folds and posteriorly rotated, simple auricles. A prenatal SNP array indicated no AOH, and postnatal brain ultrasound was normal. An abdominal ultrasound showed a normal liver but enlarged kidneys with poor corticomedullary differentiation and multiple cysts. This patient passed away at DOL 16 and no autopsy was performed. Trio ES identified a variant in *DHX16* at Chr 6:30624786 C>A position NM_003587.4:c.2091G>T:p.Gln697His.

Individual 8 is a now deceased 4 month old male infant. He presented with a severe congenital hypotonia, bilateral talipes equinovarus, and flexion contractures of the hands and knees with diminished limb movements, along with poor postnatal growth and feeding, requiring nasogastric gavage. Recurrent respiratory distress persisted including choking on his oro-pharyngeal secretions and aspiration. Electrophysiological studies (diminished nerve conduction velocities, and absent sensory nerve action potentials) and audiology, identified both an axonal sensory and denervating motor neuropathy, and sensorineural deafness. High-frequency, low-amplitude horizontal nystagmus was also noted on ophthalmologic review. An MRI of the brain in life was normal. At four months of age, he was admitted with acute chronic respiratory failure in the setting of a coronavirus pneumonitis and despite intensive care support passed away. Post mortem examination showed normal central cerebral myelination, and in the sural nerve, a paucity of large myelinated axons rimmed only by thin myelin sheaths or completely lacking myelin with atrophic myofibers consistent with denervation in the quadriceps musculature. ES performed at the Northwest Clinical Genomics Laboratory at the University of Washington identified a *de novo* variant in *DHX16* NM_003587.4:c.1280G>A:p.Gly427Glu (Chr6:30632615).

Individual 9 is a male proband, the first child of German parents. Both parents and the younger sister have no relevant medical complaints. The pregnancy and delivery, postnatal adaptation and development were normal until the 10th month of life. At that time, he presented with a febrile seizure, which re-occurred at the 15th and 17th month of life and led to initiation of anti-epileptic treatment with desoxyphenobarbital. At the age of four years he had a generalized

tonic-clonic seizure, at that time the EEG showed multifocal hypersynchronic activity. He could stand without support at 11 months and started to walk with support at 13 months. By the age of 20 months an unsteady gait and hypertrophic calf muscles were noted, and he stopped walking. The CK was unremarkable (177 U/l, Ref. < 270). Based on a muscle biopsy showing myopathy with atrophy of type I fibers and mild signs of inflammation, the diagnosis of a myositis was made and treated with cortisol. This led to prompt improvement with regaining of muscle strength and unsupported walking. At the age of five years the gait was noted to be broad based and atactic, he could not jump, had bilateral pointed feet and no deep tendon reflexes. Follow-up muscle biopsy one and three years later showed a degenerative myopathy with isolated necrotic fibers without signs of inflammation.

The CK reached a maximum at age seven years (689 U/l, Ref <270). Electron microscopy showed glycogen storage and fatty vacuoles. At the age of seven years ophthalmological investigation revealed pale pupils, shrunken vessel and pigmentary alterations. Neurophysiological investigations at the same time showed extensive myopathic changes in the EMG, pathological visual evoked potentials (VEP), auditory evoked potentials (AEP), and a pathological EEG with severe signs of generalized and multifocally diminished seizure threshold. The SSEP of the N. tibialis was pathological indicating a conductance deficit. At the age of nearly nine years sensorineural deafness was diagnosed. IQ-testing at that time was unremarkable. The vision decreased over the years with corresponding changes at the fundus starting at age 4 years showing “salt-and-pepper-fundus”, at age six years “Fundus flavimaculatus”, and the electroretinogram showed no activity on either side. Later a tapetoretinal degeneration was described. He suffered an absence epileptic state at age 14 years. Since then progredient

unsteadiness, distal muscle atrophy, multiple contractures and tunnel vision were seen. In summary, this patient has a multisystem involvement with developmental delays and normal intellectual capacity; epilepsy, retinopathy, sensorineural deafness, myopathy and neuropathy. He is currently 34 years old, wheelchair-bound, blind and seizure-free. Trio ES revealed the following *de novo* variant in *DHX16* (NM_003587.4:c.2021C>T:p.Thr674Met).

Individual 10 is one of a set of female twins born full term with a vaginal delivery. The sister is unaffected but it is unknown whether the twins are monozygotic. The proband was noted to have developmental delay in the first year of life. She first sat unsupported at eight months old, stood at two years old, and walked unsupported at three years old. At three years old, she was diagnosed with epilepsy which resolved with age. She was noted to have right sided dystonia and hand tremors at five years old. Her metabolic workup was normal including prolactin, AFP levels, plasma amino acids, carnitine, and hexosaminidase enzyme levels, as was her karyotype and fragile X screen, echocardiogram and eye exam. An initial brain MRI was reportedly normal, however a repeat MRI six years later showed abnormal signal changes in bilateral cerebral peduncles, thalami and in the peritrigonal preiventricular white matter. This signal remained unchanged in a followup MRI eight years later. Trio ES identified two variants in *DDX54* NM_001111322.1:c.647A>G:p.Asn216Ser (Chr 12:113614866 T>C), which was confirmed by Sanger to be inherited from the father (Figure 2), and NM_001111322.1:c.892C>T:p.Leu298Phe (Chr12:113612724 G>A), which was confirmed to be *de novo*.

Individual 11 is a second child of healthy parents of Caucasian and African descent. The girl was born after uncomplicated pregnancy at term with normal measurements. The first month of life were complicated by unexplained recurrent vomiting but normal weight gain. The gross motor development was mildly delayed; she started walking with 17 months. At the age of two years a delay in speech development was diagnosed, followed by the diagnosis of mild global development delay at the age of three years. In addition, she showed a deficit in fine motor skills and behavioral problems. Hearing tests were normal as well as an eye examination. At the age of 6 years she was affected by severe speech impairment, moderate intellectual disability, poor fine motor skills, and repetitive movements and behavior. Mild facial dysmorphisms were noted including broad forehead, broad eyebrows, up-slanting palpebral fissures, high nasal bridge and broad nasal tip, and small teeth. Further, she had a 3x3 cm hypopigmentation at back and right thigh. All measurements were within the normal range. An extended metabolic work up gave normal results except for a slight elevation of glutamine in plasma. Eye examination and ultrasound of heart, abdominal organs and kidney were normal. MRI of the brain revealed multiple pineal cysts without other malformations. Spectroscopy of brain was normal.

Karyotyping and SNP-array were both normal. Trio-ES identified compound heterozygous variants in *DDX54*, a paternally inherited NM_001111322.1:c.58T>A;p.Trp20Arg and a maternally inherited NM_001111322.1:c.1832G>A;p.Arg611Gln. The latter was confirmed by Sanger sequencing.

Individual 12 was born in a consanguineous family of Arabic decent. Renal ultrasound showed poor corticomedullary differentiation, multiple cysts and both kidneys were small for age. ES analysis identified a homozygous variant in *DDX54* NM_001111322.1:c.856G>A:p.Val286Met (chr12:113612859) (Figure 2L). Parental samples were unavailable for segregation analysis.

Individual 13 is a 12 year old female born to consanguineous Turkish parents after an uneventful birth. She presented with short stature (120cm, <3%ile), microcephaly (OFC 46.5cm, <3%ile), hyperextensibility, hip dislocations and a solitary kidney. She also was found to have dysmorphic features including a prominent forehead, high frontal hairline and a narrow palate. ES identified a homozygous variant in *DHX34* at position NM_014681.5:c.1322A>G:p.Asn441Ser (19:47863274A>G). Sanger sequencing confirmed her parents are heterozygous carriers of this variant allele. Additionally, analysis of her exome for evidence of absence of heterozygosity (AOH) showed these variants occur in an AOH stretch of 2656788 bp (342673086 bp total). This patient was previously reported [6] to have another homozygous variant contained within an AOH block in *CEP97* NM_024548.2:c.1148A>G:p.His383Arg (Chr3:101476598A>G), which was proposed as a novel candidate gene for the neurodevelopmental phenotype.

Individual 14 is the first child of multiple to consanguineous parents. At the 23rd week of pregnancy bilateral polycystic kidneys were discovered by ultrasound, followed by the diagnosis of oligohydramnion in the week 26. The boy was born at 37 weeks of pregnancy with a birth weight of 3030 g (-0.3 z), birth length of 48 cm (-1.1 z) and OFC of 32 cm (-1.8 z). At birth postaxial polydactyly of his right hand and both feet were observed. The postnatal period was

complicated by respiratory failure requiring assisted ventilation, pneumothorax, pulmonary hypertension, and renal failure requiring peritoneal dialysis. Further analysis revealed hypothyroidism and a suspicion of deafness. Over the course of disease the renal function recovered to a compensated renal failure. Nevertheless, the pulmonary function did not show a substantial improvement and lead to recurrent hypercapnia and ventilation dependence. Lung biopsy showed emphysema without interstitial fibrosis. The psychomotor development was poor and he displayed a failure to thrive. At age of three months he had a weight of 4350 g (-1.9 z), a length of 54 cm (-2.0z z) and an OFC of 35 cm (-4.5 z). Kidney ultrasound revealed hyperechoic kidneys without cortex / medulla differentiation, with mild hydronerphosis of left kidney. Ultrasound of the liver showed diffuse hyperechoic structure, probably due to an arterial perfusions deficiency. Echocardiography showed a small persistent foramen ovale. Electroencephalography was normal. His karyotype was normal. Trio-ES revealed a homozygous nonsense mutation NM_000547.5:c.1618C>T:p.Arg540* in *TPO* as the probable cause of the hypothyroidism. In addition, we identified a homozygous nonsense mutation NM_014681.5:c.466C>T, p.Gln156* in *DXH34* within an AOH block of 12931821 bp.

Individual 15 is an eight year-old female, the oldest of three siblings, born from non-consanguineous parents, with Ashkenazi and Sephardim ancestry. Family history is unremarkable.

She was born at term by caesarean section due to breech presentation, after an uneventful pregnancy. Her anthropometric measurements at birth were a weight of 2720g, length of 48 cm and head circumference of 34.5 cm. There were no complications during the neonatal period

and she was released from the hospital 72 hours after birth. Hypotonia was noticed within a few months of life and her neuropsychomotor development was subsequently delayed. She walked without support at 19 months and spoke only a few words by the third year of age. She presently has intellectual deficiency, with significant compromise in expressive language, and displays good social interaction and a friendly disposition. She had no history of epilepsy but recently presented an isolated seizure episode and Valproic acid was initiated after an electroencephalogram (EEG) revealing frequently occurring epileptic activity multifocal with generalized projection.

Laboratory studies previously performed included a normal peripheral blood G-banded female karyotype (46,XX), a normal CGH-array and a brain MRI disclosing no abnormalities.

Physical examination at seven years and 11 months revealed height of 119 cm (-1.5 S.D), weight of 19kg (-2 S.D.), head circumference of 51 cm (-0.49 S.D.), and no significant dysmorphisms.

Potential candidate gene *DDX47* identified in a single family. The proband is a 5.5 year old old female born to consanguineous Turkish parents after three sequential infant losses before the age of one year. All children shared a phenotype of hypotonia and epilepsy. At 5.5 years of age, she weighed 21kg, has seizures about once a week, and is severely hypotonic (unable to lift her head). A cranial MRI revealed a T2 signal abnormality in the bilateral periventricular deep white matter and a lacunar infarct with gliosis measuring 4mm on the right occipital lobe (Figure 2). Previous *GLUT1* sequencing and array-CGH (60k, Agilent Inc.) were normal. ES analysis identified compound heterozygous variants in *DDX47* at position

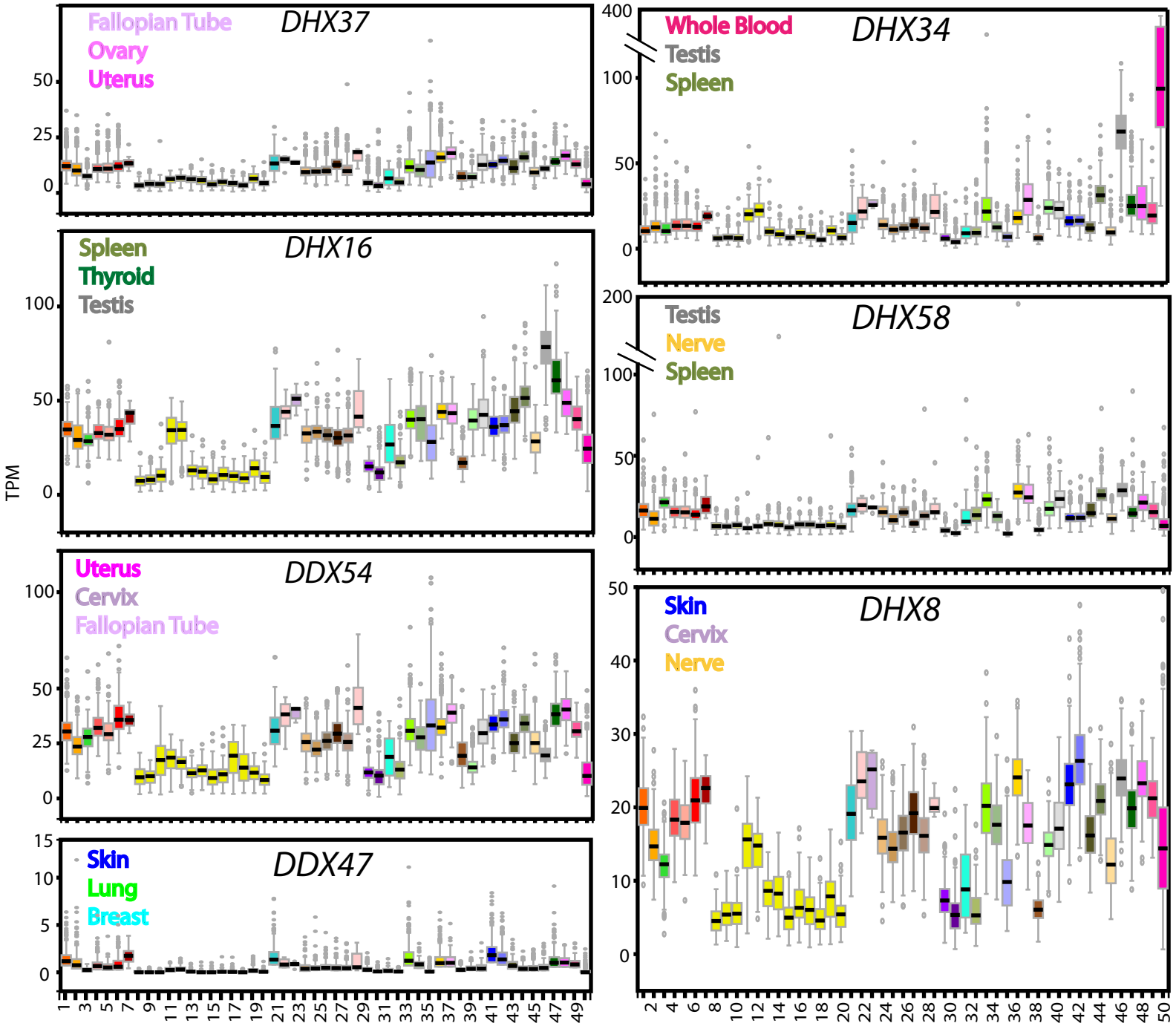
NM_016355.3:c.22G>T;p.Asp8Tyr (12:12966323G>T) which was confirmed by Sanger sequencing to be inherited from the father, and NM_016355.3:c.319C>G;p.Gln107Glu (12:12974279C>G) which was confirmed to be inherited from the mother. DNA from deceased siblings was not available for sequencing. We also identified a homozygous variant in *SLC13A5* XM_005256612.1:c.15331A>G;p.Thr511Ala (17:6589651T>C) which occurs in an AOH block measuring 504308 bp (491 Mb genome-wide total AOH).

Potential candidate gene *DHX58* identified in a single family. The proband was a seven year old female born to consanguineous parents. She was found prenatally to have a thin corpus callosum and midmuscular VSD. At birth she weighed 2.8kg, was 45 cm long and her OFC was 32.5cm. She had recurrent infections since birth including adenovirus, rhinovirus, Staphylococcus and CMV infections. She was reported to have intermittent neutropenia but sequencing of ELA-2, GCP-3, GCSF-R and HAX-1 were all normal. She presented with severe developmental delay with limited motor development, no independent sitting or standing and no speech. She has physically normal eyes with delayed visual maturation. She presented with febrile seizure at 5 months of age which progressed into tonic-clonic seizures. At three years of age her weight, height, and OFC were all below the 3rd percentile (10.3 kg, 83.5cm, and 45.5 cm). She passed away from status epilepticus at seven years old. An older brother had previously died from a similar disease with an additional kidney involvement.

Potential candidate gene *DHX8* identified in a single family. The proband is a nearly eight year old male of Indian descent. He originally presented with a febrile urinary tract infection at two

months of age, at which time a renal ultrasound found a 6.1 cm kidney, moderate hydronephrosis, and distal ureteral dilation. A repeat ultrasound at five years old showed a 9 cm single kidney.

Figure S1: Expression Profiles of each Candidate DExD/H-Box Helicase Genes. Expression profiles of each candidate were queried in the GTEx database and are presented here.



1. Adipose-subcutaneous
2. Adipose- Visceral
3. Adrenal Gland
4. Artery - Aorta
5. Artery - Coronary
6. Artery - Tibial
7. Bladder
8. Brain - Amygdala
9. Brain - Anterior cingulate cortex
10. Brain - Caudate
11. Brain - Cerebellar Hemisphere
12. Brain - Cerebellum
13. Brain - Cortex
14. Brain - Frontal Cortex
15. Brain - Hippocampus
16. Brain - Hypothalamus
17. Brain - Nucleus accumbens
18. Brain - Putamen
19. Brain - Spinal cord
20. Brain - Substantia nigra
21. Breast
22. Cervix - Ectocervix
23. Cervix - Endocervix
24. Colon - Sigmoid
25. Colon - -Transverse
26. Esophagus - Gastroesophageal Junction
27. Esophagus - Musculans
28. Fallopian Tube
29. Heart - Atrial Appendage
30. Heart - Left Ventricle
31. Kidney - Cortex
32. Liver
33. Lung
34. Minor Salivary Gland
35. Muscle - Skeletal
36. Nerve - Tibial
37. Ovary
38. Pancreas
39. Pituitary
40. Prostate
41. Skin - Not Sun Exposed
42. Skin - Sun Exposed
43. Small Intestine - Terminal Ileum
44. Spleen
45. Stomach
46. Testis
47. Thyroid
48. Uterus
49. Vagina
50. Whole Blood

Figure S2: Unsupervised Clustering of DDX and DHX Genes. Unsupervised hierarchical clustering of *DHX8*, *DHX16*, *DHX34*, *DHX37*, *DDX47*, *DDX54*, and *DHX58* based on mRNA levels in the brain tissue.

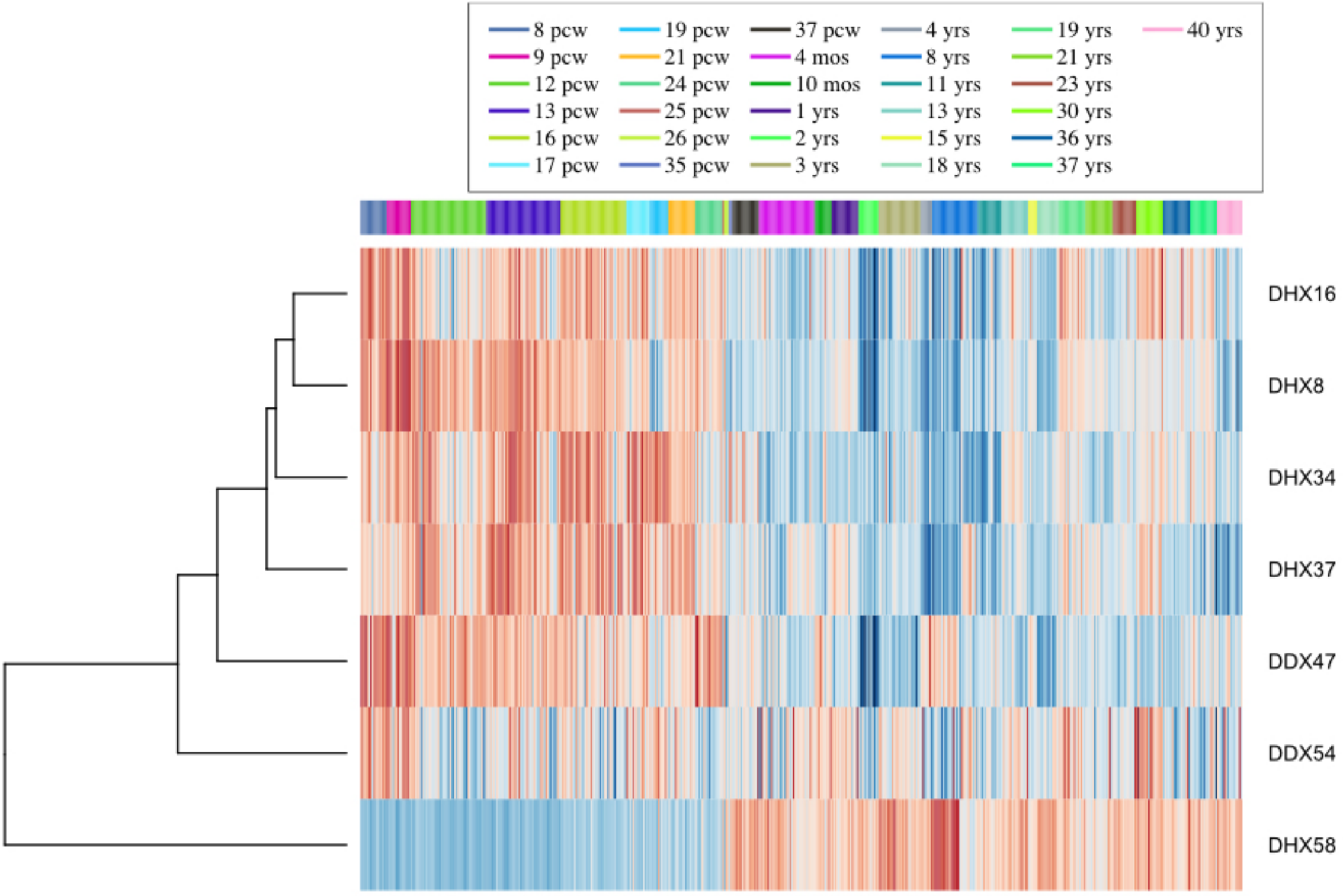
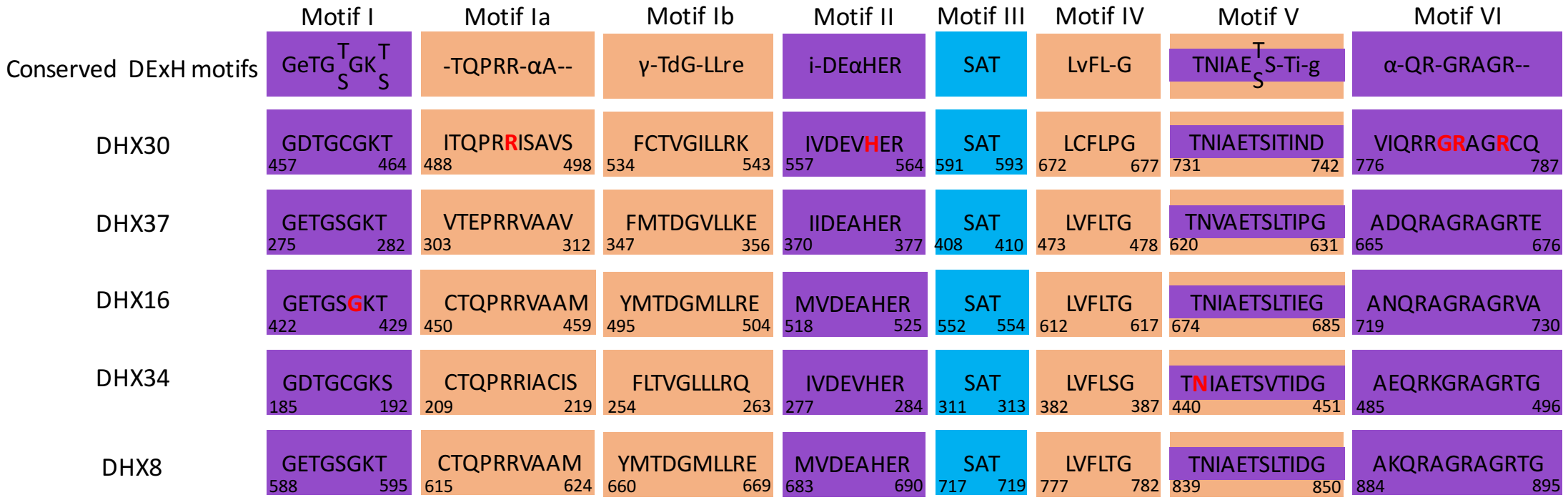

8 pcw
9 pcw
12 pcw
13 pcw
16 pcw
17 pcw
19 pcw
21 pcw
24 pcw
25 pcw
26 pcw
35 pcw
37 pcw
4 mos
10 mos
1 yrs
2 yrs
3 yrs
4 yrs
8 yrs
11 yrs
13 yrs
15 yrs
18 yrs
19 yrs
21 yrs
23 yrs
30 yrs
36 yrs
37 yrs
40 yrs
DHX16
DHX8
DHX34
DHX37
DDX47
DDX54
DHX58

Figure S3: Conserved Motif Schematic. Schematic protein structure of DExD/H-box RNA helicases showing conserved motifs of the helicase core region. Nucleotide-interacting motifs (I, II and VI) are shown in purple, nucleic acid-binding motifs (Ia, Ib and IV) in orange, motif V, which binds nucleic acid and interacts with nucleotides, in purple and orange, and motif III, which couples ATP hydrolysis to RNA unwinding, in blue. The position of the first and last amino acid within each motif is denoted below left and right, respectively. The patient-mutated residues are marked in red.


Motif I
Motif Ia
Motif Ib
Motif II
Motif III
Motif IV
Motif V
Motif VI
Conserved DExH motifs
GeTG T/S GK T/S
-TQPRR-αA--
γ-TdG-LLre
i-DEαHER
SAT
LvFL-G
TNIAE T/S S-Ti-g
α-QR-GRAGR--
DHX30
GDTGCGKT 457 464
ITQPRRISAVS 488 498
FCTVGILLRK 534 543
IVDEVHER 557 564
SAT 591 593
LCFLPG 672 677
TNIAETSITIND 731 742
VIQRRGRAGRCQ 776 787
DHX37
GETGSGKT 275 282
VTEPRRVAAV 303 312
FMTDGVLLKE 347 356
IIDEAHER 370 377
SAT 408 410
LVFLTG 473 478
TNVAETSLTIPG 620 631
ADQRAGRAGRTE 665 676
DHX16
GETGSGKT 422 429
CTQPRRVAAM 450 459
YMTDGMLLRE 495 504
MVDEAHER 518 525
SAT 552 554
LVFLTG 612 617
TNIAETSLTIEG 674 685
ANQRAGRAGRVA 719 730
DHX34
GDTGCGKS 185 192
CTQPRRIACIS 209 219
FLTVGLLLRQ 254 263
IVDEVHER 277 284
SAT 311 313
LVFLSG 382 387
TNIAETSVTIDG 440 451
AEQRKGRAGRTG 485 496
DHX8
GETGSGKT 588 595
CTQPRRVAAM 615 624
YMTDGMLLRE 660 669
MVDEAHER 683 690
SAT 717 719
LVFLTG 777 782
TNIAETSLTIDG 839 850
AKQRAGRAGRTG 884 895

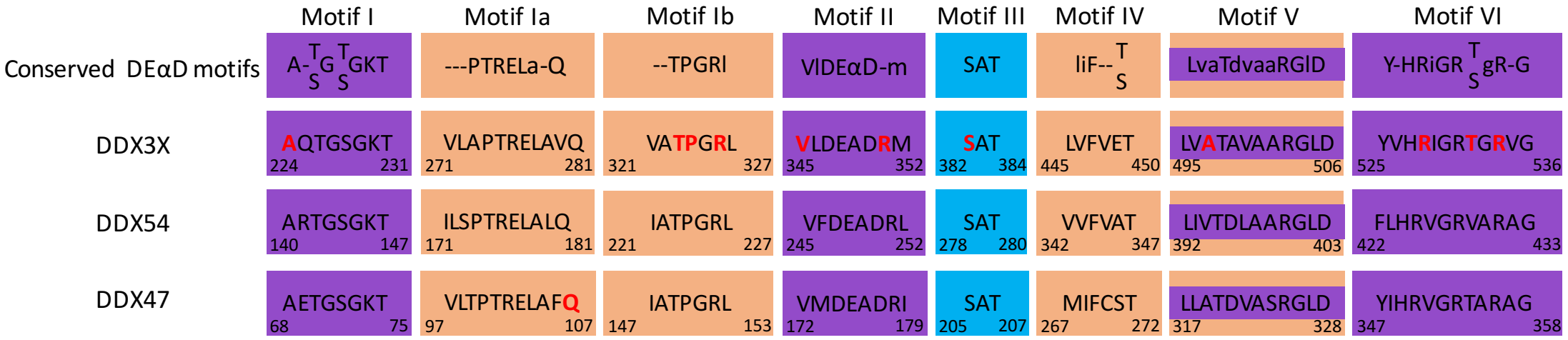

Motif I
Motif Ia
Motif Ib
Motif II
Motif III
Motif IV
Motif V
Motif VI
Conserved DEαD motifs
A-T/S G T/S GKT
---PTRELa-Q
--TPGRl
VIDEαD-m
SAT
liF-- T/S
LvaTdvaaRGlD
Y-HRiGR T/S gR-G
DDX3X
AQTGSGKT 224 231
VLAPTRELAVQ 271 281
VATPGRL 321 327
VLDEADRM 345 352
SAT 382 384
LVFVET 445 450
LVATAVAARGLD 495 506
YVHRIGRTGRVG 525 536
DDX54
ARTGSGKT 140 147
ILSPTRELALQ 171 181
IATPGRL 221 227
VFDEADRL 245 252
SAT 278 280
VVFVAT 342 347
LIVTDLAARGLD 392 403
FLHRVGRVARAG 422 433
DDX47
AETGSGKT 68 75
VLTPTRELAFQ 97 107
IATPGRL 147 153
VMDEADRI 172 179
SAT 205 207
MIFCST 267 272
LLATDVASRGLD 317 328
YIHRVGRTARAG 347 358

Table S1. Genotype Table. All variants for which a minor allele frequency > 0 was identified in ExAC or gnomAD had no stances of homozygous inheritance. *DHX37* was observed to have a probability of loss-of-function intolerance (pLI) of 0.99. While *DHX16*, *DDX54*, and *DHX34* each have a pLI score of 0 in gnomAD, they each display somewhat lower observed:expected ratios of LOF variants (*DHX16* 31:63.7; *DDX54* 22:46.3; *DHX34* 28:48.8). ND denotes no data; NA denotes not applicable.

Supplemental References

1. Hempel, M., Cremer, K., Ockeloen, C.W., Lichtenbelt, K.D., Herkert, J.C., Denecke, J., Haack, T.B., Zink, A.M., Becker, J., Wohlleber, E., et al. (2015). De Novo Mutations in CHAMP1 Cause Intellectual Disability with Severe Speech Impairment. Am. J. Hum. Genet. *97*, 493–500.

2. Hamosh, A., Sobreira, N., Hoover-Fong, J., Sutton, V.R., Boehm, C., Schiettecatte, F., and Valle, D. (2013). PhenoDB: A New Web-Based Tool for the Collection, Storage, and Analysis of Phenotypic Features. Hum. Mutat. *34*, 566–571.

3. Sobreira, N., Schiettecatte, F., Boehm, C., Valle, D., and Hamosh, A. (2015). New Tools for Mendelian Disease Gene Identification: PhenoDB Variant Analysis Module; and GeneMatcher, a Web-Based Tool for Linking Investigators with an Interest in the Same Gene. Hum. Mutat. *36*, 425–431.

4. Gambin, T., Akdemir, Z.C., Yuan, B., Gu, S., Chiang, T., Carvalho, C.M.B., Shaw, C., Jhangiani, S., Boone, P.M., Eldomery, M.K., et al. (2017). Homozygous and hemizygous CNV detection from exome sequencing data in a Mendelian disease cohort. Nucleic Acids Res. *45*, 1633–1648.

5. Olshen, A.B., Venkatraman, E.S., Lucito, R., and Wigler, M. (2004). Circular binary segmentation for the analysis of array-based DNA copy number data. Biostatistics *5*, 557–572.

6. Karaca, E., Harel, T., Pehlivan, D., Jhangiani, S.N., Gambin, T., Coban Akdemir, Z., Gonzaga-Jauregui, C., Erdin, S., Bayram, Y., Campbell, I.M., et al. (2015). Genes that Affect Brain Structure and Function Identified by Rare Variant Analyses of Mendelian Neurologic Disease. Neuron *88*, 499–513.