Deleterious, protein-altering variants in the transcriptional coregulator *ZMYM3* in 27 individuals with a neurodevelopmental delay phenotype

Authors

Susan M. Hiatt, Slavica Trajkova,
Matteo Rossi Sebastiano, ..., Lance H. Rodan,
Richard M. Myers, Gregory M. Cooper

Correspondence

shiatt@hudsonalpha.org (S.M.H.),
gcooper@hudsonalpha.org (G.M.C.)

We identified 27 individuals with neurodevelopmental disorders (NDD) with rare variants in *ZMYM3*, an X chromosome gene encoding a transcriptional regulator. Some variants recurrently affect the same codons, and computational and experimental analyses suggest the variants impair ZMYM3 function. Our data strongly support *ZMYM3* as an NDD gene.

Deleterious, protein-altering variants in the transcriptional coregulator *ZMYM3* in 27 individuals with a neurodevelopmental delay phenotype

Susan M. Hiatt,[1,*] Slavica Trajkova,[2] Matteo Rossi Sebastiano,[3] E. Christopher Partridge,[1] Fatima E. Abidi,[4] Ashlyn Anderson,[1] Muhammad Ansar,[5,49] Stylianos E. Antonarakis,[6] Azadeh Azadi,[7] Ruxandra Bachmann-Gagescu,[8] Andrea Bartuli,[9] Caroline Benech,[10] Jennifer L. Berkowitz,[11] Michael J. Betti,[12] Alfredo Brusco,[2] Ashley Cannon,[13] Giulia Caron,[3] Yanmin Chen,[11] Meagan E. Cochran,[1] Tanner F. Coleman,[1] Molly M. Crenshaw,[14] Laurence Cuisset,[15] Cynthia J. Curry,[16] Hossein Darvish,[17,18] Serwet Demirdas,[19] Maria Descartes,[13] Jessica Douglas,[20] David A. Dyment,[21] Houda Zghal Elloumi,[11] Giuseppe Ermondi,[3]

(Author list continued on next page)

Summary

Neurodevelopmental disorders (NDDs) result from highly penetrant variation in hundreds of different genes, some of which have not yet been identified. Using the MatchMaker Exchange, we assembled a cohort of 27 individuals with rare, protein-altering variation in the transcriptional coregulator *ZMYM3*, located on the X chromosome. Most (n = 24) individuals were males, 17 of which have a maternally inherited variant; six individuals (4 male, 2 female) harbor *de novo* variants. Overlapping features included developmental delay, intellectual disability, behavioral abnormalities, and a specific facial gestalt in a subset of males. Variants in almost all individuals (n = 26) are missense, including six that recurrently affect two residues. Four unrelated probands were identified with inherited variation affecting Arg441, a site at which variation has been previously seen in NDD-affected siblings, and two individuals have *de novo* variation resulting in p.Arg1294Cys (c.3880C>T). All variants affect evolutionarily conserved sites, and most are predicted to damage protein structure or function. *ZMYM3* is relatively intolerant to variation in the general population, is widely expressed across human tissues, and encodes a component of the KDM1A-RCOR1 chromatin-modifying complex. ChIP-seq experiments on one variant, p.Arg1274Trp, indicate dramatically reduced genomic occupancy, supporting a hypomorphic effect. While we are unable to perform statistical evaluations to definitively support a causative role for variation in *ZMYM3*, the totality of the evidence, including 27 affected individuals, recurrent variation at two codons, overlapping phenotypic features, protein-modeling data, evolutionary constraint, and experimentally confirmed functional effects strongly support *ZMYM3* as an NDD-associated gene.

Introduction

Neurodevelopmental disorders (NDD) as a group affect 1%–3% of children, but individual NDD syndromes are typically rare and often result from highly penetrant genetic variation affecting one of many NDD-associated loci.[1,2] While exome- and genome-sequencing tests have provided molecular diagnoses for many individuals with NDDs, the diagnostic yield from sequencing remains below 50%.[3] Various hypotheses exist to explain this diagnostic limitation, one of which is that some NDD-associated genes have yet to be identified. The wide availability of sequencing tests, coupled with data sharing, has allowed identification of many new NDD genes over the last few years.[4]

ZMYM3 (MIM: 300061) lies on the X chromosome and encodes a member of a transcriptional corepressor complex that includes HDAC1, RCOR1, and KDM1A.[5,6] ZMYM3 has been hypothesized to function as a scaffolding protein, coordinating interactions between deacetylases and demethylases, in addition to RNASEH2A.[6] Knockout of *Zmym3* in male mice results in infertility due to a defect in the metaphase-to-anaphase transition during

[1]HudsonAlpha Institute for Biotechnology, Huntsville, AL 35806, USA; [2]Department of Medical Sciences, University of Torino, 10126 Torino, Italy; [3]Molecular Biotechnology and Health Sciences Department, Università degli Studi di Torino, via Quarello 15, 10135 Torino, Italy; [4]Greenwood Genetic Center, Greenwood, SC 29646, USA; [5]Department of Ophthalmology, University of Lausanne, Jules Gonin Eye Hospital, Fondation Asile des Aveugles, Lausanne, Switzerland; [6]Department of Genetic Medicine and Development, University of Geneva, Geneva, Switzerland; [7]Obstetrics and Gynecology Department, Golestan University of Medical Sciences, Gorgan, Iran; [8]Institute of Medical Genetics, University of Zurich, Schlieren 8952, Switzerland; [9]Genetics and Rare Diseases Research Division, Ospedale Pediatrico Bambino Gesù, IRCCS, 00146 Rome, Italy; [10]Univ Brest, Inserm, EFS, UMR 1078, GGB, 29200 Brest, France; [11]GeneDx, LLC, Gaithersburg, MD 20877, USA; [12]Vanderbilt University Medical Center, Nashville, TN 37232, USA; [13]Department of Genetics, University of Alabama at Birmingham, Birmingham, AL, USA; [14]Pediatrics and Medical Genetics, University of Colorado, Aurora CO, USA; [15]Service de Médecine Génomique des Maladies de Système et d'Organe, Département Médico-Universitaire BioPhyGen, Hôpital Cochin, APHP, Université Paris Cité, Paris, France; [16]Genetic Medicine, UCSF/Fresno, Fresno, CA 93701, USA; [17]Neuroscience Research Center, Faculty of Medicine, Golestan University of Medical Sciences, Gorgan, Iran; [18]Nikagene Genetic Diagnostic Laboratory, Gorgan, Golestan, Iran; [19]Department of Clinical Genetics, Erasmus MC University Medical

(Affiliations continued on next page)

Marie Faoucher,[22,23] Emily G. Farrow,[24] Stephanie A. Felker,[1] Heather Fisher,[25] Anna C.E. Hurst,[13]
Pascal Joset,[26] Melissa A. Kelly,[27] Stanislav Kmoch,[28] Benjamin R. Leadem,[11] Michael J. Lyons,[4]
Marina Macchiaiolo,[9] Martin Magner,[29] Giorgia Mandrile,[30] Francesca Mattioli,[31] Megan McEown,[1]
Sarah K. Meadows,[1] Livija Medne,[32] Naomi J.L. Meeks,[33] Sarah Montgomery,[34] Melanie P. Napier,[11]
Marvin Natowicz,[35] Kimberly M. Newberry,[1] Marcello Niceta,[9] Lenka Noskova,[28] Catherine B. Nowak,[20]
Amanda G. Noyes,[11] Matthew Osmond,[21] Eloise J. Prijoles,[4] Jada Pugh,[1] Verdiana Pullano,[2] Chloé Quélin,[36]
Simin Rahimi-Aliabadi,[37] Anita Rauch,[8,38] Sylvia Redon,[10,39,40] Alexandre Reymond,[31] Caitlin R. Schwager,[41]
Elizabeth A. Sellars,[42] Angela E. Scheuerle,[43] Elena Shukarova-Angelovska,[44] Cara Skraban,[32]
Elliot Stolerman,[4] Bonnie R. Sullivan,[41] Marco Tartaglia,[9] Isabelle Thiffault,[24] Kevin Uguen,[10,39,40]
Luis A. Umaña,[43] Yolande van Bever,[19] Saskia N. van der Crabben,[45] Marjon A. van Slegtenhorst,[19]
Quinten Waisfisz,[46,47] Camerun Washington,[4] Lance H. Rodan,[20,48] Richard M. Myers,[1]
and Gregory M. Cooper[1,*]

spermatogenesis.[7] ZMYM3 was found to be necessary for the regulation of various meiotic genes in this process. ZMYM3 has also been found to promote DNA repair, as it regulates the localization of BRCA1 at damaged chromatin.[8] ZMYM3 was originally identified as an NDD candidate gene in a female with a balanced X;13 translocation affecting the 5′ UTR of one isoform of ZMYM3.[9] The proband presented with ID, scoliosis, spotty abdominal hypopigmentation, slight facial asymmetry, clinodactyly, and history of a possible febrile seizure at age one year. Additionally, Philips et al. reported a family with three NDD-affected brothers carrying a missense variant in ZMYM3 (GenBank: NM_005096.3; c.1321C>T [p.Arg441Trp]).[10] The brothers displayed developmental delay, a sleeping disorder, microcephaly, genitourinary anomalies, and facial dysmorphism.

Given the extremely low prevalence for any given Mendelian NDD, data sharing to facilitate cohort building is essential and has had a large impact on rare disease gene discovery over the last decade.[11] Here we describe a cohort of individuals with rare variants in ZMYM3, assembled from submissions to GeneMatcher[12] and PhenomeCentral.[13] We provide strong evidence for an X-linked, ZMYM3-associated NDD based on phenotypic, computational, and experimental analysis of variants observed in 27 individuals.

Subjects and methods

ZMYM3 was submitted to GeneMatcher (https://genematcher.org/) by HudsonAlpha in 2018, and follow-up discussion of cases from either research studies or clinical sequencing was performed via e-mail over the course of four years. Some matches originated from GeneMatcher,[12] while others originated from PhenomeCentral.[13] Over the course of the collaboration, some affected individuals were excluded from the cohort due to segregation of the variant of interest in unaffected male family members, including two individuals harboring GenBank: NM_005096.3; c.2063G>A (p.Arg688His), a variant that was initially identified as a VUS but later reclassified to likely benign after observation in an unaffected male relative. Additionally, one of the individuals with p.Arg688His variation presented with developmental regression and facial dysmorphism that was dissimilar to the phenotypes of other probands described here.

Approval for human subject research was obtained from all local ethics review boards, and informed consent for publication (including photos, where applicable) was obtained at individual sites. Exome sequencing (ES), genome sequencing (GS), or panel testing was performed on DNA extracted from blood, buccal cells, or muscle tissue using typical clinical or research protocols, as described in supplemental material and methods.

For protein modeling, the wild-type 3D protein structure was downloaded from AlphaFoldDB (https://alphafold.ebi.ac.uk/),[14] which was included with the reference from UniProt (accession

Center, Rotterdam, the Netherlands; [20]Boston Children's Hospital, Boston, MA, USA; [21]Children's Hospital of Eastern Ontario Research Institute, Ottawa, ON, Canada; [22]Service de Génétique Moléculaire et Génomique, CHU, Rennes 35033, France; [23]Univ Rennes, CNRS, IGDR, UMR 6290, Rennes 35000, France; [24]Children's Mercy Kansas City, Center for Pediatric Genomic Medicine, Kansas City, KS, USA; [25]Children's Medical Center, Dallas, TX, USA; [26]Medical Genetics, Institute of Medical Genetics and Pathology, University Hospital Basel, Basel, Switzerland; [27]HudsonAlpha Clinical Services Lab, LLC, Huntsville, AL 35806, USA; [28]Research Unit for Rare Diseases, Department of Pediatrics and Inherited Metabolic Disorders, 1st Faculty of Medicine, Charles University in Prague, Prague, Czech Republic; [29]Department of Pediatrics and Inherited Metabolic Disorders, General University Hospital and First faculty of Medicine, Charles University, Prague, Czech Republic; [30]Medical Genetics Unit and Thalassemia Center, San Luigi University Hospital, University of Torino, Orbassano, Italy; [31]Center for Integrative Genomics, University of Lausanne, Lausanne, Switzerland; [32]Childrens Hospital of Philadelphia, Philadelphia, PA, USA; [33]Section of Genetics & Metabolism, Department of Pediatrics, University of Colorado Anschutz Medical Campus, Aurora, CO 80045, USA; [34]Division of Genetics and Metabolism, Children's Health, Dallas, TX, USA; [35]Pathology & Laboratory Medicine, Genomic Medicine, Neurological and Pediatrics Institutes, Cleveland Clinic, Cleveland, OH, USA; [36]Service de Génétique Clinique, Centre de Référence Maladies Rares CLAD-Ouest, CHU Hôpital Sud, Rennes, France; [37]Department of Pharmacology and Toxicology, College of Pharmacy, University of Utah, Salt Lake City, UT 84112, USA; [38]University Children's Hospital Zurich, University of Zurich, Zurich 8032, Switzerland; [39]Service de Génétique Médicale et Biologie de la Reproduction, CHU de Brest, Brest, France; [40]Centre de Référence Déficiences Intellectuelles de causes rares, Brest, France; [41]Division of Genetics, Children's Mercy Kansas City, Kansas City, MO, USA; [42]Genetics and Metabolism, Arkansas Children's Hospital, Little Rock, AR 72202, USA; [43]Department of Pediatrics, Division of Genetics and Metabolism, University of Texas Southwestern Medical Center, Dallas, TX, USA; [44]Department of Endocrinology and Genetics, University Clinic for Children's Diseases, Medical Faculty, University Sv. Kiril i Metodij, Skopje, Republic of Macedonia; [45]Amsterdam University Medical Centers, Department of Clinical Genetics, Amsterdam, the Netherlands; [46]Department of Human Genetics, Amsterdam University Medical Centers, VU University Amsterdam, Amsterdam, The Netherlands; [47]Amsterdam Neuroscience, Amsterdam, The Netherlands; [48]Harvard Medical School, Boston, MA 02115, USA; [49]Advanced Molecular Genetics and Genomics Disease Research and Treatment Centre, Dow University of Health Sciences, Karachi, Pakistan

*Correspondence: shiatt@hudsonalpha.org (S.M.H.), gcooper@hudsonalpha.org (G.M.C.)
https://doi.org/10.1016/j.ajhg.2022.12.007.

Figure 1. Observed variation along the length of ZMYM3

The 1,370 aa ZMYM3 (Q14202, GenBank: NP_005087.1) is annotated with MYM-type zinc fingers (1–9, orange) and Cre-like domain (blue) as described by UniProt and InterPro.

(A) Hemizygous variants observed in males in this study are noted above the protein model, with *de novo* variants in red. [a]Note that p.Arg441Gln was observed in three unrelated males. Hemizygous variants that were previously reported in males are shown below the protein.[10,25]

(B) Maternally inherited (black) or *de novo* (red) heterozygous variants observed in females in this study are noted above the protein model.

number: Q14202). When visualization and coloring were not possible with the online tool, structures were visualized and colored and the sequence was mutated with Chimera v.1.15, rotamer builder tool.[15] Structure superposition was obtained in Chimera with the tool Matchmaker. Structure refinement was performed with the Chimera tool Dock Prep with standard settings, as previously described.[16] Depiction of molecular surfaces was defined as VdW surface and colored according to the electrostatic potential. For additional analyses, see supplemental material and methods. For eukaryotic linear motif analysis (ELM), the UniProt accession (Q14202) was submitted to the online ELM server (http://elm.eu.org/) with standard settings (100 as probability cut-off, species *Homo sapiens*).

For ChIP-seq experiments, we edited the genomic DNA at the *ZMYM3* endogenous locus in HepG2 cells to introduce the variant (the "variant" experiment) or to reintroduce the reference sequence (the "control" experiment), simultaneously with a 3X FLAG tag, 2A self-cleaving peptide, and neomycin resistance gene, using a modified version of the previously published CRISPR epitope tagging ChIP-seq (CETCh-seq) protocol.[17] We nucleofected cells and selected for correctly edited cells using neomycin, confirmed edits by PCR and Sanger sequencing of genomic DNA, and performed ChIP-seq as previously described[18] with duplicate experiments for each condition (see supplemental material and methods). We performed peak calling using SPP[19] and Irreproducible Discovery Rate (IDR),[20] using ENCODE-standardized pipelines for analysis and quality-control.[21] We performed additional differential binding analyses using the R package csaw v.1.28.0.[22]

As an additional control, we used the standard CETCh-seq ZMYM3 experiment in HepG2 available on the ENCODE portal (ENCSR505DVB), with these data processed to match (i.e., down-sampled to 20M reads) the other CETCh-seq experiments described here. See supplemental material and methods for additional details.

Results

ZMYM3 variants

Through a collaboration facilitated by the MatchMaker Exchange,[11] we identified 22 unique variants in *ZMYM3* in 27 affected individuals from 25 unrelated families (Figure 1 and Table 1). All observed variants had high CADD scores (average 24.3, range 19–32, Table S1), indicating that they rank among the 1.25% most highly deleterious SNVs in the human reference assembly, similar to most known highly penetrant NDD-associated variants.[23] All SNVs also had high conservation scores (average GERP score of 4.97, range 3.47–5.22), suggesting they affect positions under selective constraint throughout mammalian evolution.[24]

Twenty-four of these 27 individuals are males that harbor hemizygous missense variants, including two sets of affected brothers. For most males (n = 17), variants were inherited from heterozygous carrier mothers. In four males, the *ZMYM3* variant arose *de novo*, while inheritance could not be defined for three. All variants are rare, with three or fewer total alleles and no hemizygous males or homozygous females in gnomAD[26] or TopMed/Bravo (https://bravo.sph.umich.edu/freeze8/hg38/) (Table S1).

In addition, we identified three heterozygous *ZMYM3* variants in three unrelated, affected females (Figure 1 and

Table 1. Individual variant and phenotypic data for the entire cohort

Individual	Sex	Age (years)	Zygosity	Inheritance	Mother's NDD-related phenotype	Variant (NM_005096.3; NP_005087.1)	Speech delay	Motor delay	ID	ASD traits	Behavioral problems	Facial dys-morphism	GU anomalies	Other
1	male	3	hemizygous	maternal	none	c.205G>A (p.Asp69Asn)	yes	yes	N/A	N/A	no	no	urinary tract dilatation of left kidney on ultrasound	congenital heart defects
21	male	18.2	hemizygous	unknown	none	c.507A>T (p.Arg169Ser)	yes	no	yes	yes	yes	yes	hypospadias	–
2	male	14	hemizygous	maternal	Hx of LD	c.721G>A (p.Glu241Lys)	yes	yes	no	no	no	yes	no	history of growth hormone resistance and IGF1 deficiency (basis unknown), fasting and heat intolerance, excessive fatigue
3	male	8	hemizygous	maternal	none	c.905G>A (p.Arg302His)	yes	yes	N/A	yes	yes	yes	pielonephritis, vesicoureteral reflux	GERD
4a	male	21	hemizygous	maternal	none	c.1183C>A (p.Arg395Ser)	yes	yes	yes	yes	yes	yes	hypospadias	–
4b	male	16	hemizygous	maternal	none	c.1183C>A (p.Arg395Ser)	yes	no	yes	yes	yes	yes	no	–
5	male	7	hemizygous	maternal	none	c.1192C>T (p.Pro398Ser)	yes	yes	yes	yes	yes	no	no	weight <1%ile
22	male	4	hemizygous	unknown	unknown	c.1321C>T (p.Arg441Trp)	yes	no	yes	yes	no	yes	no	mild short stature
6	male	7.42	hemizygous	maternal	ADHD	c.1322G>A (p.Arg441Gln)	yes	yes	yes	yes	yes	yes	single renal cyst	constipation
7	male	13	hemizygous	maternal	none	c.1322G>A (p.Arg441Gln)	yes	yes	yes	yes	yes	yes	cryptorchidism, enuresis	short stature
8	male	15	hemizygous	maternal	Hx of LD	c.1322G>A (p.Arg441Gln)	yes	yes	yes	yes	yes	yes	hypospadias, ambiguous genitalia	short stature
23	male	16	hemizygous	de novo	N/A	c.1360T>C (p.Cys454Arg)	yes	yes	no	no	yes	yes	vesicoureteral reflux	short stature, microcephaly, myopia, retinopathy, GI dysmotility
9a	male	6	hemizygous	maternal	none	c.2193G>C (p.Glu731Asp)	yes	no	yes	yes	yes	yes	no	–
9b	male	4	hemizygous	maternal	none	c.2193G>C (p.Glu731Asp)	yes	no	yes	yes	yes	yes	no	–
10	male	2.5	hemizygous	maternal	dyslexia	c.2794A>G (p.Ile932Val)	yes	no	N/A	N/A	yes	yes	no	GERD, constipation
11	male	19	hemizygous	maternal	none	c.3371G>A (p.Arg1124Gln)	yes	no	yes	no	yes	yes	ectopic kidney	short stature, kyphoscoliosis
12	male	5	hemizygous	maternal	none	c.3409T>A (p.Tyr1137Asn)	yes	yes	yes	no	yes	yes	no	microcephaly
24	male	8	hemizygous	maternal	ADHD, Hx of delays	c.3518G>A (p.Ser1173Asn)	yes	yes	yes	yes	no	yes	no	–

(Continued on next page)

Table 1. *Continued*

Individual	Sex	Age (years)	Zygosity	Inheritance	Mother's NDD-related phenotype	Variant (NM_005096.3; NP_005087.1)	Speech delay	Motor delay	ID	ASD traits	Behavioral problems	Facial dys-morphism	GU anomalies	Other
13	male	3.42	hemizygous	*de novo*	none	c.3605T>A (p.Val1202Asp)	yes	yes	yes	no	yes	yes	cryptorchidism	microcephaly, short stature, weight <3%ile, kyphosis, long bone defects, Madelung deformity
14	male	62	hemizygous	unknown	none	c.3638T>C (p.Met1213Thr)	yes	yes	yes	no	no	no	enuresis	microcephaly, scoliosis, reflux
15	male	16.25	hemizygous	*de novo*	none	c.3820C>T (p.Arg1274Trp)	yes	yes	yes	yes	yes	yes	no	microcephaly, scoliosis
16	male	0	hemizygous	*de novo*	none	c.3880C>T (p.Arg1294Cys)	N/A	N/A	N/A	N/A	N/A	N/A	N/A	deceased
25	male	8.5	hemizygous	maternal	none	c.3970C>T (p.Arg1324Trp)	yes	no	yes	yes	yes	no	no	–
17	male	8	hemizygous	maternal	none	c.4029G>A (p.Met1343Ile)	yes	yes	no	yes	yes	no	no	GI dysmotility, joint laxity, pain & swelling, dysautonomic symptoms
						Total	23/23	15/23	17/20	15/21	18/23	18/23	11/23	–
18	female	1.5	heterozygous, skewed XCI	maternal	none	c.671_674dup (p.Leu226TrpfsTer8)	yes	yes	N/A	N/A	no	yes	no	GERD
19	female	1.42	heterozygous	*de novo*	none	c.2255A>G (p.Tyr752Cys)	yes	yes	N/A	N/A	no	yes	no	–
20	female	3	heterozygous, skewed XCI	*de novo*	unknown	c.3880C>T (p.Arg1294Cys)	yes	yes	N/A	N/A	N/A	yes	pyelectasis	volvulus of midgut, pancreatic cysts
						Total	3/3	3/3			0/2	3/3	1/3	

Individuals 4a and 4b are full siblings; individuals 9a and 9b are full siblings. ID, intellectual disability; ASD, autism spectrum disorder; GU, genitourinary; N/A, not assessed; GERD, gastroesophageal reflux disease; Hx, history; LD, learning disability; ADHD, attention deficit-hyperactivity disorder.

Table 1). All three of these variants are absent from population databases. Two of these variants arose *de novo*, while one was inherited from an apparently unaffected mother. As variation observed in males was often inherited from unaffected heterozygous mothers (with presumed random X-inactivation), we hypothesized that the three affected female individuals might have skewed X-inactivation that could result in expression of primarily the variant *ZMYM3* allele. In two of the three females, X-inactivation testing targeting either the *AR* locus[27] or the *RP2* locus[28] was performed, and in both, skewed X-inactivation was observed. In individual 20, a female carrying a *de novo* p.Arg1294Cys variant (GenBank: NM_005096.3; c.3880C>T), 97% skewing at the *AR* locus was observed. In the case of the maternally inherited p.Leu226TrpfsTer8 variant (individual 18, GenBank: NM_005096.3;c.671_674dup), >94% skewing was observed in both the proband and her unaffected, heterozygous mother at the *RP2* locus. Both mother and daughter were heterozygous for two *RP2* alleles (366/362), and in both, the 366 allele was inactivated (see supplemental note: case reports for additional details). Due to the presence of skewing in both the proband and her unaffected mother, it is possible that this predicted loss-of-function allele is benign. However, skewing of the precise *ZMYM3* alleles was not tested in these individuals.

Phenotypic characterization

Of the 24 identified males, one was a fetus terminated at 26 weeks gestational age with a *de novo* variant in *ZMYM3* (GenBank: NM_005096.3; c.3880C>T [p.Arg1294Cys]) and a very severe phenotype (supplemental note: case reports). For this reason, we did not include this male in further phenotypic comparisons. Of the remaining 23 affected males, all were reported to have developmental delay (23/23), with speech delay (23/23) being more prominent than motor delay (15/23) (Table 1 and supplemental note: case reports). Of those who could be assessed, 17/20 showed intellectual disability, and most were diagnosed with autism or were reported to have autistic traits (15/21). Most males had behavioral concerns at some point in development (18/23). Most affected males were also reported to have at least mild facial dysmorphism (18/23), some of which were highly similar to the individuals reported in Philips et al.[10] (Figure 2). Similarities include thick eyebrows, deeply set eyes, long palpebral fissures, protruding ears, and a high anterior hairline. Other variable features include genitourinary anomalies (n = 11 individuals), short stature (n = 6), microcephaly (n = 5), scoliosis/kyphosis (n = 4), and functional gastrointestinal problems (n = 6) (Table 1). See supplemental note: case reports for additional clinical features for each case.

Among the affected females, all three displayed developmental delay and some facial dysmorphism, but many of their additional features were variable and do not lead to a clear syndromic picture (Table 1, Figure 2, and supplemental note: case reports).

Additionally, while most variants in affected male probands were inherited from apparently unaffected heterozygous carrier mothers (10/15 mothers), five heterozygous mothers were reported to have a history of learning disabilities, attention deficit-hyperactivity disorder (ADHD), or dyslexia (Table 1, supplemental note: case reports, and Figure S1).

Protein modeling

ZMYM3 encodes a DNA-binding transcriptional coregulator with multiple protein isoforms, the longest of which is 1,370 amino acids (Q14202, GenBank: NP_005087.1). This isoform has nine MYM-type zinc fingers and a C-terminal Cre-Like domain (Figure 1). As most of the observed variants are missense (21/22 unique variants), we performed computational modeling to assess the potential effects of these changes. Homology-based protein modeling using AlphaFold[14] indicates that 17 of the 21 missense variants lie in ordered regions, and the majority have intermediate to high predicted local distance difference test (pLDDT) scores,[29] indicating that there is a moderate to high degree of confidence in further computational predictions (Figures 3 and S2 and Table S2).

We assessed flexibility, stability, solvent exposure, and deformation energy of the variant protein models (Figures S3–S6). A general trend toward protein destabilization (negative folding energy differential) was observed for several variants, while p.Arg1274Trp was predicted to be stabilizing (Figure S3). We observed patterns somewhat consistent with solvent exposure across the 21 unique missense variants (Table S2). Six of the seven variants leading to the highest destabilization (p.Arg441Gln, p.Glu731Asp, p.Tyr752Cys, p.Arg1124Gln, p.Tyr1137Asn, p.Met1213Thr) (Figure S3) are buried residues in high confidence regions of the protein. While disruption of each of these rigid residues is predicted to be destabilizing, some are due to likely increased flexibility (p.Glu731Asp, p.Tyr752Cys, p.Tyr1137Asn) while others are predicted to be more rigid (p.Arg441Gln, p.Arg1124Gln, p.Met1213Thr). This result is consistent with the observation that substitutions of amino acids within the protein core are often associated with folding destabilization.

Conversely, the remaining 14 variants affect exposed residues; seven of these lie in low confidence regions or have very low pLDDT values (p.Asp69Asn, p.Arg169Ser, p.Glu241Lys, p.Arg302His, p.Arg395Ser, p.Pro398Ser, p.Arg1274Trp). These wild-type residues are predicted to be flexible, and in most cases the observed mutation is predicted to lead to a more rigid structure. The remaining seven are rigid residues, with the observed mutations associated with varying predicted effects. More detailed surface analyses indicated that several variants result in significant changes of polarity, charge, and hydrophobicity (Table S2 and Figure S6). In particular, p.Arg1274Trp is predicted to have major effects, resulting in stabilization of an exposed residue through the substitution of a polar, charged, and flexible arginine with a neutral, aromatic, and hydrophobic tryptophan moiety (Figure S6).

In addition to structural analysis, we submitted the sequences to the eukaryotic linear motif (ELM).[30] This resource
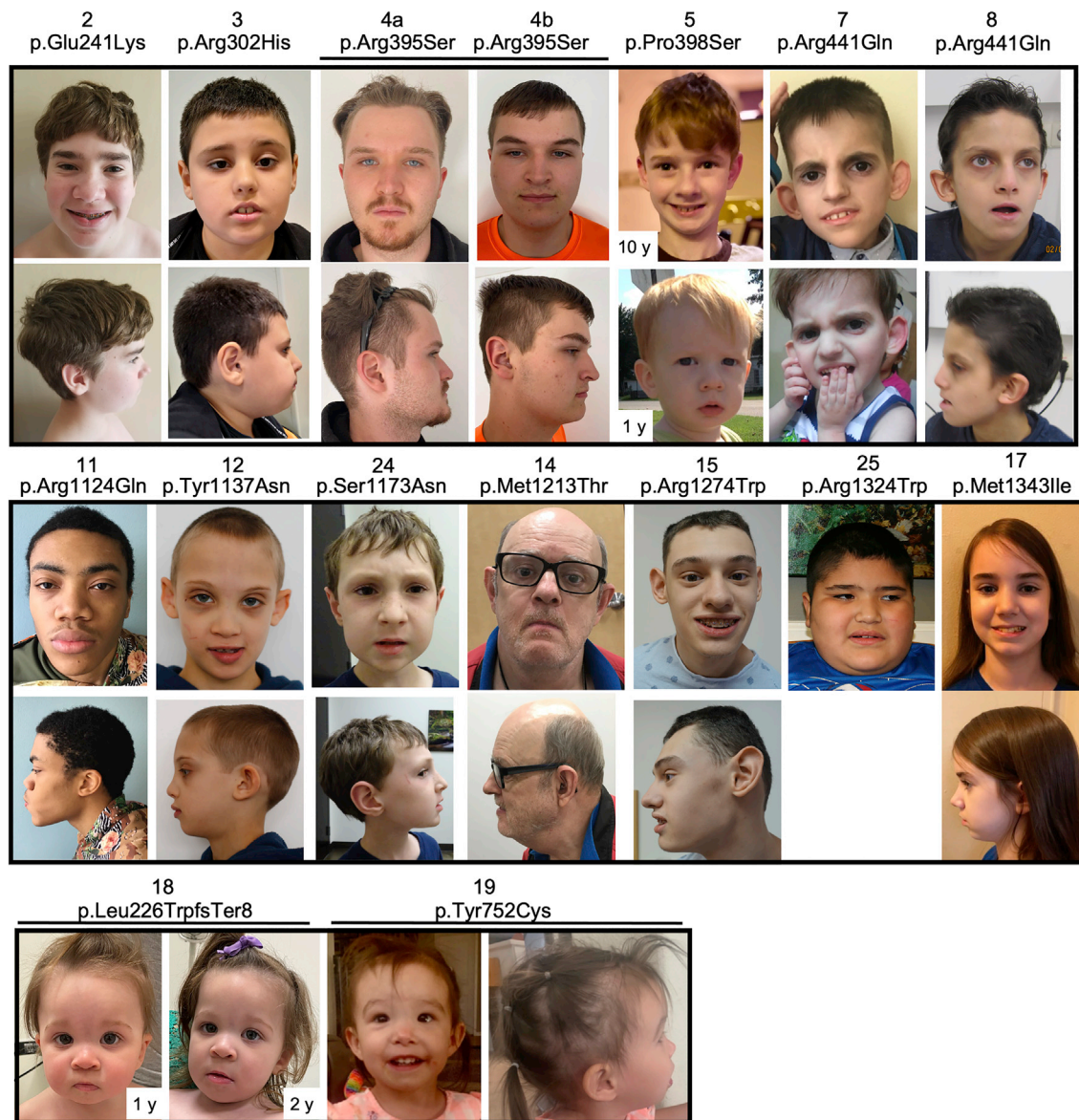
Figure 2. Facial features of a subset of individuals with *ZMYM3* variation
Individual ID and protein effect are noted for each. Note deep-set eyes, long palpebral fissures, large/prominent/cupped ears, and tall forehead.

annotates short amino acid motifs predicted to mediate binding to other proteins or to be affected by post-translational modifications (phosphorylation, cleavage sites, ubiquitination, etc.). Intersecting this information with the position of our mutations suggests that several of the variants alter motifs (Tables S2 and S3) and that modifications of residues Arg302, Ser1173, Val1202, Met1213, Arg1274, and Met1343 are predicted to possibly disrupt multiple interactions.

Genome-wide occupancy of selected ZMYM3 variant transcription factors

A key role of ZMYM3 is to function as a component of the KDM1A/RCOR1 chromatin-modifying complex that regulates gene expression by binding to specific loci throughout the genome.[5] Therefore, we sought to measure the impact of variation on ZMYM3 genome-wide DNA association, hypothesizing that proband-observed variants may alter ZMYM3 genome-wide occupancy patterns. Given the time and expense of these experiments, we chose three variants for testing: p.Arg441Trp, a previously reported variant[10] that affects a residue where we have seen recurrent variation (p.Arg441Trp and p.Arg441Gln); p.Arg1274Trp, a *de novo* variant within the Cre-like domain that was found in an individual with notable facial similarities to those individuals with Arg441 variation; and p.Arg688His, which early in our collaboration was seen in two affected individuals. Subsequently, segregation studies in one family indicated that the p.Arg688His variant was present in an unaffected maternal uncle, suggesting that it is likely benign.
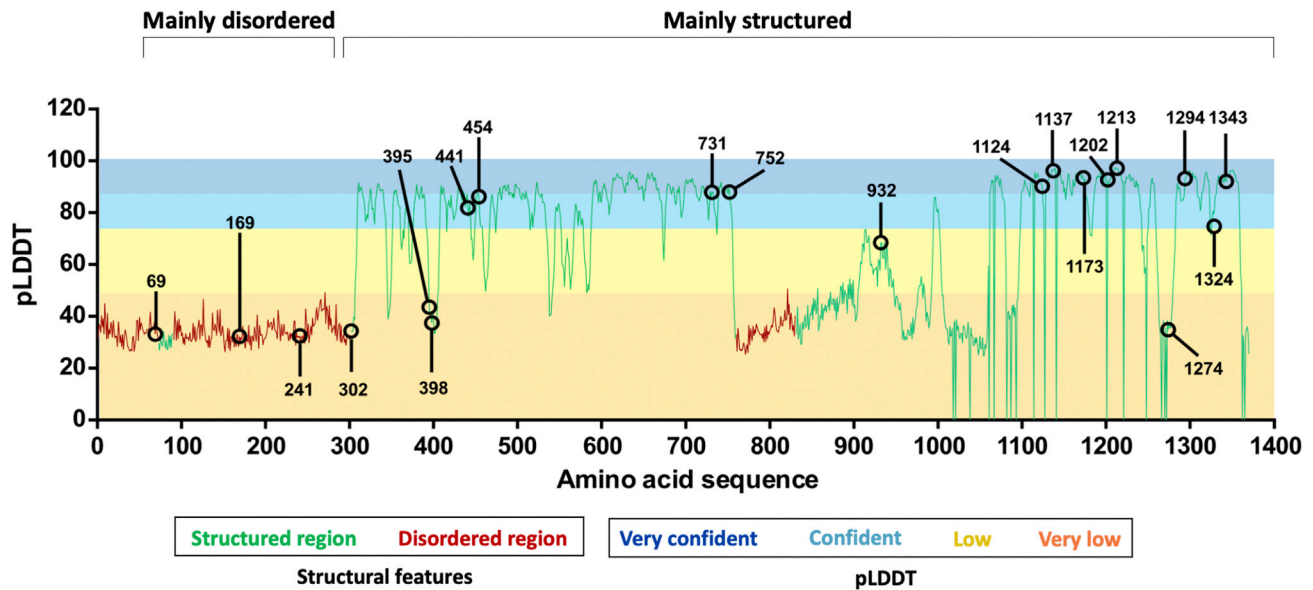
Figure 3. Missense variants in ZMYM3 mainly lie in ordered regions
Three disordered regions (red line) were identified (aa 1–72, 90–301, and 759–830), while the remainder of the protein is predicted to be structured (green line). AlphaFold produces a per-residue confidence score (predicted local distance difference test, pLDDT) between 0 and 100, which is plotted along the length of the ZMYM3 protein. Horizontal bars and shading indicate confidence ranges for pLDDT scores. Missense variants observed here are noted on the graph, and while residues 69, 169, 241, and 302 lie in disordered regions, the remainder of residues lie in structured regions.

For each of these, we introduced the variant into the *ZMYM3* gene in the genomic DNA of cultured HepG2 cells using a modified version of the CRISPR epitope tagging ChIP-seq (CETCh-seq) protocol.[17] We simultaneously introduced a "super-exon" consisting of all exons of *ZMYM3* downstream (relative to coding direction) of the exon in which the variant resides, along with an FLAG epitope tag and selectable resistance gene. These modifications result in cells that express the ZMYM3 protein with the variant residue and a carboxyl terminus FLAG tag for immunoprecipitation, as well as a neomycin resistance gene product for selection of correctly edited cells. As a control for each super-exon edit, we performed the same protocol but reintroduced the reference sequence instead of the missense variant. Genomic DNA modifications were confirmed by PCR and Sanger sequencing. The key advantage of this approach is that the control and variant ZMYM3 proteins are produced from the endogenous genomic loci, each modified by the same super-exon, and that the antibody used (along with other experimental and analytical steps) is the same; the only difference between the variant and control experiments is the presence of the missense variant of interest. For both p.Arg688His and p.Arg1274Trp, we successfully obtained correctly edited cells and performed chromatin immunoprecipitation followed by high-throughput sequencing (ChIP-seq) and peak calling as previously described[18,31]; however, for p.Arg441Trp, we were unable to obtain edited cells. As an additional control, we also analyzed data from a standard CETCh-seq experiment on ZMYM3 (ZMYM3[CETCh]) in HepG2 cells (ENCODE dataset ENCSR505DVB).

When comparing ZMYM3[p.Arg1274Trp]-variant to ZMYM3[p.Arg1274Trp]-control, we observed a large difference in the number of peaks called between the experiments. The control experiment yielded 16,214 peaks and the variant only 3,699 peaks (Table S4); most (68%) of the peaks in the variant experiment were also called in control, suggesting the variant protein occupies a subset of the sites occupied by the control protein. We know from extensive previous ChIP-seq analyses that many loci exhibit read-depth levels near (above or below) peak-calling thresholds, resulting in situations where experiments are more similar than they appear when only considering peak-call overlaps. Thus, we performed additional, more quantitative comparisons, such as global read-depth correlations, which also support a global, variant-specific reduction of occupancy (see supplemental material and methods). Further, we performed a differential occupancy analysis using the R package csaw.[22] Rather than relying on peak calls, csaw performs a sliding window analysis to detect regions with significantly different read-depths between experiments; csaw identified 25,845 genomic regions with sufficient reads for analysis in the p.Arg1274Trp experiments. Among these regions, 13,225 showed differential read-depth between control and variant experiments at FDR < 0.05. All but 19 of these sites (99.9%) had higher read counts in the control than in the variant. We also intersected csaw regions with the union of peak calls between control and variant experiments, resulting in 11,259 genomic regions; of these, 6,631 show significantly more reads in control than in variant, and only three
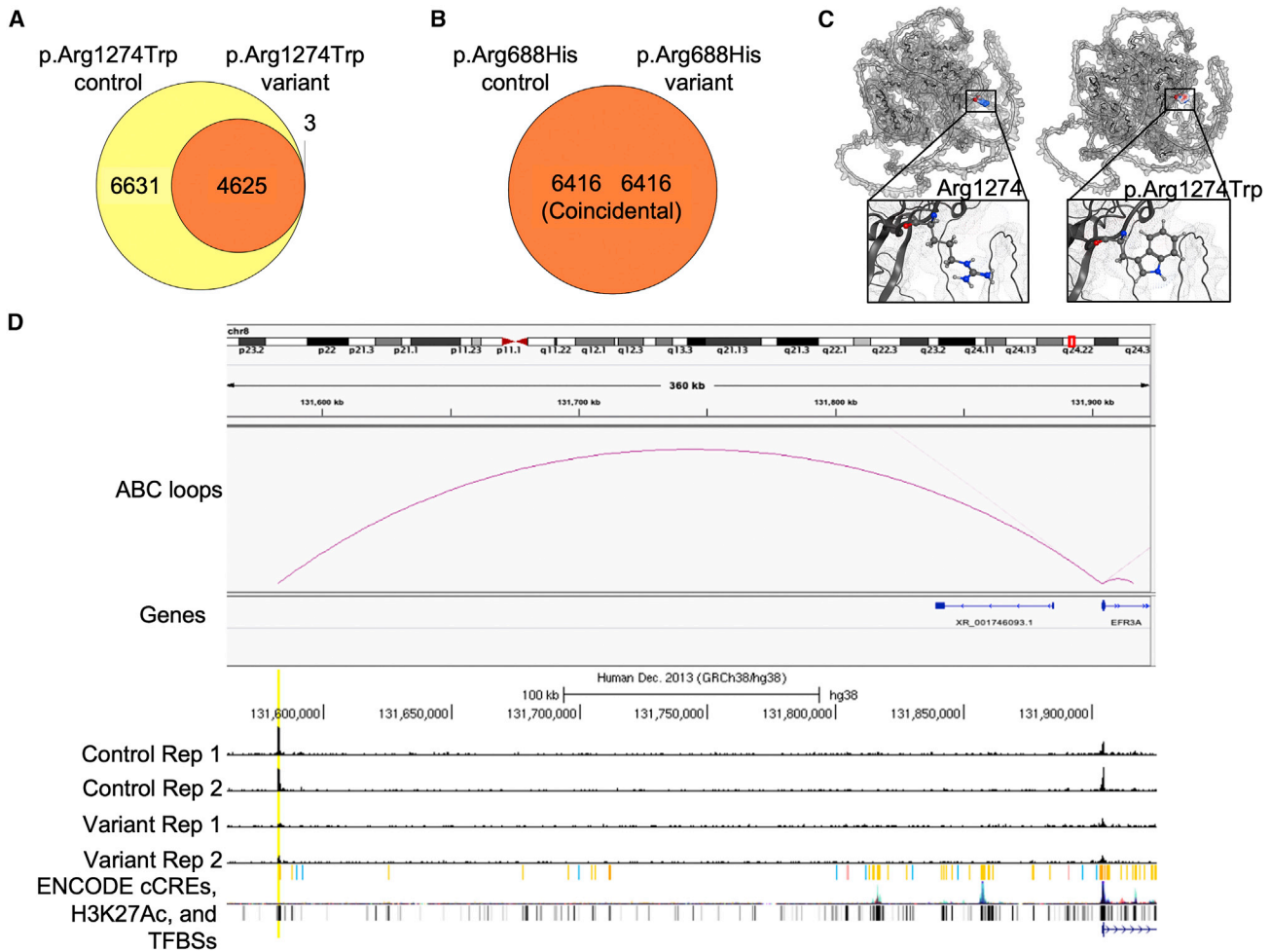
Figure 4. p.Arg1274Trp is a hypomorphic variant, while p.Arg688His has similar genome occupancy to that of wild type
(A and B) Genomic regions called by csaw between experiments, then overlapped with IDR 0.05 peaks called in either experiment. Yellow color indicates regions determined by csaw to have significantly higher differential binding (at FDR < 0.05) in control, orange indicates regions with no differential binding, and red indicates regions with higher differential binding in variant. For p.Arg1274Trp (A), there are 6,631 regions with higher binding in control, 4,625 regions with no differential binding, and 3 regions with higher binding in variant. For p.Arg688His (B), all 6,416 regions had no differential binding.
(C) Protein modeling of Arg1274 (left) and p.Arg1274Trp (right). The van der Waals protein surface is depicted in light gray, and residues in position 1,274 are colored by partial charge (blue, positive; red, negative; white, neutral). Magnified squares show zoomed-in view of the side chains.
(D). Genome browser track for ZMYM3-p.Arg1274Trp-variant ChIP-seq experiments. Human genome (hg38) chr8:131,561,953–131,925,404 is displayed. Top track is activity-by-contact ("ABC loops") showing predicted interaction between enhancer element on left and TSS for the gene *EFR3A* on right. "Genes" track is RefSeq gene model. "Control Rep 1" and "Control Rep 2" are aligned bam reads from ZMYM3-p.Arg1274Trp-control experiments, "Variant Rep 1" and "Variant Rep 2" are aligned bam reads from ZMYM3-p.Arg1274Trp-variant experiments. All bam files are downsampled to an equal number of reads in each replicate, and all four replicate tracks are scaled from 0 to 60 vertically. "ENCODE" tracks are shown below the ChIP-seq tracks: "cCREs" represent candidate *cis*-regulatory elements colored by ENCODE standards, "H3K27Ac" is layered H3K27Ac signal from seven ENCODE cell lines, and "TFBSs" are ENCODE TF clusters (340 factors, 129 cell types). The putative enhancer element identified as the most significant loss of binding in the variant experiment is highlighted in yellow, showing the ENCODE distal enhancer cCRE call, the ABC loop to the TSS of *EFR3A*, and the difference in binding with the two control replicates showing strong signal and the two variant replicates showing substantially less binding.

were significantly higher in variant (Figure 4A). Finally, we performed immunocytochemistry on control and variant p.Arg1274Trp-edited cells to assess ZMYM3 localization. While ZMYM3 is predominantly nuclear in p.Arg1274Trp-control cells, as expected for a DNA-binding transcriptional regulator, ZMYM3 is predominantly cytoplasmic in p.Arg1274Trp-variant cells (Figure S7). Thus, the reduction of ZMYM3^p.Arg1274Trp genomic occu-

pancy appears mediated, at least in part, by reduced nuclear localization.

We similarly analyzed the ZMYM3^p.Arg688His-control and ZMYM3^p.Arg688His-variant experiments. Both control and variant p.Arg688His experiments yielded fewer peaks than ZMYM3^p.Arg1274Trp-control and ZMYM3^CETCh experiments (Figure S8), suggesting that the super-exon insertion at this location may by itself impact activity. However, there

appears to be little to no functional impact from the missense variant. Pearson correlation coefficients of read counts of each of the two replicates of control and variant ranged from 0.71 to 0.84, indicating a high degree of overall similarity among these experiments (Figure S9). Similarly, analysis of csaw regions intersected with peak calls gave 6,416 genomic sites with sufficient reads for analysis, none of which were significantly different (FDR < 0.05) between control and variant (Figure 4B). As such, p.Arg688His does not appear to alter ZMYM3 genomic occupancy, a result consistent with its presence in an unaffected male.

Discussion

Here we describe 27 NDD-affected individuals with protein-altering variation in *ZMYM3*, mostly (n = 24) hemizygous males. Six of these variants arose *de novo*, but most were inherited from unaffected or mildly affected heterozygous mothers. All variants presented here are rare in the general population and predicted to be deleterious. Many of the variants are predicted to interfere with protein structure or function. *ZMYM3* is relatively intolerant to both missense variation (gnomAD missense Z = 4.31) and loss-of-function variation (RVIS = 8.46,[32] pLOEUF = 0.11[26]), further supporting the potential for the variants observed here to have phenotypic effects. Using ChIP-seq, we have also provided functional analyses showing that one variant, p.Arg1274Trp, acts as a hypomorphic variant with greatly reduced genome occupancy compared to its control, and that one likely benign variant, p.Arg688His, has genome occupancy similar to its control experiment.

Among the variants in our cohort, there are two sets of alleles affecting the same codon. At Arg441, a residue that lies within a zinc finger domain that functions in DNA binding, we found substitutions (p.Arg441Trp or p.Arg441Gln) in four unrelated males. Three additional affected males with p.Arg441Trp in one family have been previously reported.[10] Overlapping phenotypic features of these seven individuals include developmental delay (mainly speech), nocturnal enuresis, and microcephaly. In addition, the facial features in these individuals are quite similar. The other recurrent variant that we observed here is p.Arg1294Cys, observed as *de novo* in an aborted male fetus and *de novo* in a female with 97% skewed X inactivation. p.Arg1294Cys has also been submitted to ClinVar[33] as a VUS (SCV000297052.2) by a different group than those that identified p.Arg1294Cys variation for this study. We thus believe the ClinVar submission represents a third, independent report of p.Arg1294Cys pathogenicity, although we are unable to confirm this (see supplemental material and methods).

The biological context of *ZMYM3* is supportive of disease relevance. ZMYM3 is part of a transcriptional corepressor complex that includes HDAC1, RCOR1, and KDM1A.[5,6] Additional interactors in this complex can include ZMYM2 and REST. Variation in two of these five genes

(*KDM1A* and *ZMYM2*) has been robustly associated with neurodevelopmental disorders.[25,34] Additionally, ZMYM3 has since been shown to physically interact with RNASEH2A; variation in *RNASEH2A* (MIM: 606034) has been associated with Aicairdi-Gouteres syndrome 4 (AGS4 [MIM: 610333]). Specifically, a cluster of pathogenic variants found in individuals with AGS4 have been shown to disrupt binding of RNASEH2A to ZMYM3.[6] Residues within the PV-rich domain of ZMYM3 (codons 862–943) have been shown to be necessary for this interaction. p.Ile932Val, observed in our cohort, lies in this region and may disrupt this interaction.

Recently, Connaughton et al. demonstrated a connection between loss-of-function variation in *ZMYM2* (MIM: 602221), a paralog of *ZMYM3* with 44% protein identity, to congenital anomalies of the kidney and urinary tract, with extra-renal features or NDD findings (MIM: 619522).[25] This same publication also reported two male probands who had hemizygous variants of uncertain significance in *ZMYM3*, resulting in p.Gly673Asp and p.Val866Met, although the latter does appear in Bravo/TopMed in a homozygous state (Figure 1). Phenotypic overlap of individuals with variation in *ZMYM2* and *ZMYM3* presented here include developmental delay, microcephaly, and ID. Some similarity of facial features is also shared with the *ZMYM2* cohort, including one ZMYM2 proband with protuberant ears. In addition to ZMYM2 and ZMYM3, the ZMYM-family of proteins includes two additional members, ZMYM4 and QRICH1. Variation in *QRICH1* (MIM: 617387) has been associated with Ververi-Brady syndrome (MIM: 617982), which has features including developmental delay, intellectual disability, non-specific facial dysmorphism, and hypotonia.[35]

Variants observed in this cohort lie across the length of the protein, and modeling data suggest that while several may affect protein structure, several also likely affect protein interactions, which are key in the biological function of ZMYM3. ChIP-seq data for ZMYM3[p.Arg1274Trp] indicate a large reduction in genome-wide occupancy specific to the variant protein, even though the variant is not within any direct DNA-binding domains. Leung et al. have previously shown that this specific residue is necessary for interaction with RAP80, a ubiquitin-binding protein that plays a role in the DNA damage response.[8] The authors also showed that ZMYM3[p.Arg1274Gln] had increased cytoplasmic localization compared to wild-type protein, consistent with our results showing that ZMYM3[p.Arg1274Trp] is predominantly cytoplasmic (Figure S7). While the observed widespread reduction in genomic occupancy indicates a global hypomorphic effect, individual binding event differences may be of particular interest. For example, one of the most significant differential binding events, as determined by csaw, occurs at a regulatory element on chromosome 8 (Figure 4D); this region is annotated as a distal enhancer by the ENCODE Consortium,[36] and, according to activity-by-contact (ABC) analysis,[37] this region connects to and is likely a regulatory element for the gene

EFR3A (MIM: 611798). Pathogenic variants in *EFR3A* have been associated with autism spectrum disorders,[38] with phenotypes that overlap those described here.

A key limitation of this study is the location of *ZMYM3* on chromosome X and the fact that most of the probands observed here inherited their *ZMYM3* variant from an unaffected or mildly affected parent, which makes the statistical evaluation of pathogenicity difficult. We cannot, for example, use *de novo* variant enrichment testing, a powerful means of inferring pathogenicity for dominant NDDs.[39] Traditional association or burden testing also cannot be done given the absence of systematically ascertained and matched cases and controls. Additionally, none of the families described here are large enough to support linkage studies. Testing in other family members may nevertheless be informative for each individual variant's interpretation (Figure S1); this additional information may be useful for flagging potential benign variants within these families, particularly those present in a hemizygous state in unaffected male relatives as was observed for p.Arg688His. X chromosome inactivation studies in additional females, both affected and unaffected, may also be informative.

Despite the above limitations, the totality of the evidence presented here is strong. This includes 27 affected probands that exhibit overlapping phenotypic features, some of which are shared with four previously reported individuals, bringing the total number of NDD-affected individuals known to harbor rare protein-altering variation in *ZMYM3* to at least 31. Six probands described here have variants that arose *de novo*, two of which result in the same missense effect (p.Arg1294Cys). Also, both p.Arg441Trp and p.Arg441Gln were seen in this study; thus, like at Arg1294, there have necessarily been at least two independently arising variants at Arg441 in affected individuals. We further describe protein-modeling data, evolutionary constraint analyses, and experimentally confirmed functional effects, all of which support the phenotypic relevance of the observed variation. While additional analyses are necessary to ultimately confirm these findings and adjudicate the pathogenicity of each individual variant, we provide substantial evidence that *ZMYM3* is an NDD-associated gene.

Data and code availability

The published article includes all variant information pertinent to this study. ChIP-seq data are available via the NCBI Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/geo/, accession GSE216752).

Supplemental information

Supplemental information can be found online at https://doi.org/10.1016/j.ajhg.2022.12.007.

Acknowledgments

Declaration of interests

References

1. Ropers, H.H. (2008). Genetics of intellectual disability. Curr. Opin. Genet. Dev. *18*, 241–250. https://doi.org/10.1016/j.gde.2008.07.008.
2. Cooper, G.M., Coe, B.P., Girirajan, S., Rosenfeld, J.A., Vu, T.H., Baker, C., Williams, C., Stalker, H., Hamid, R., Hannig, V., et al. (2011). A copy number variation morbidity map of developmental delay. Nat. Genet. *43*, 838–846. https://doi.org/10.1038/ng.909.
3. Srivastava, S., Love-Nichols, J.A., Dies, K.A., Ledbetter, D.H., Martin, C.L., Chung, W.K., Firth, H.V., Frazier, T., Hansen, R.L., Prock, L., et al. (2019). Meta-analysis and multidisciplinary consensus statement: exome sequencing is a first-tier clinical diagnostic test for individuals with neurodevelopmental disorders. Genet. Med. *21*, 2413–2421. https://doi.org/10.1038/S41436-019-0554-6.
4. Bamshad, M.J., Nickerson, D.A., and Chong, J.X. (2019). Mendelian gene discovery: fast and furious with no end in sight. Am. J. Hum. Genet. *105*, 448–455. https://doi.org/10.1016/J.AJHG.2019.07.011.

5. Hakimi, M.-A., Dong, Y., Lane, W.S., Speicher, D.W., and Shiekhattar, R. (2003). A candidate X-linked mental retardation gene is a component of a new family of histone deacetylase-containing complexes. J. Biol. Chem. *278*, 7234–7239. https://doi.org/10.1074/jbc.M208992200.

6. Shapson-Coe, A., Valeiras, B., Wall, C., and Rada, C. (2019). Aicardi-Goutières Syndrome associated mutations of RNase H2B impair its interaction with ZMYM3 and the CoREST histone-modifying complex. PLoS One *14*, e0213553. https://doi.org/10.1371/JOURNAL.PONE.0213553.

7. Hu, X., Shen, B., Liao, S., Ning, Y., Ma, L., Chen, J., Lin, X., Zhang, D., Li, Z., Zheng, C., et al. (2017). Gene knockout of Zmym3 in mice arrests spermatogenesis at meiotic metaphase with defects in spindle assembly checkpoint. Cell Death Dis. *8*, e2910. https://doi.org/10.1038/cddis.2017.228.

8. Leung, J.W.C., Makharashvili, N., Agarwal, P., Chiu, L.-Y., Pourpre, R., Cammarata, M.B., Cannon, J.R., Sherker, A., Durocher, D., Brodbelt, J.S., et al. (2017). ZMYM3 regulates BRCA1 localization at damaged chromatin to promote DNA repair. Genes Dev. *31*, 260–274. https://doi.org/10.1101/gad.292516.116.

9. van der Maarel, S.M., Scholten, I.H., Huber, I., Philippe, C., Suijkerbuijk, R.F., Gilgenkrantz, S., Kere, J., Cremers, F.P., and Ropers, H.H. (1996). Cloning and characterization of DXS6673E, a candidate gene for X-linked mental retardation in Xq13.1. Hum. Mol. Genet. *5*, 887–897.

10. Philips, A.K., Sirén, A., Avela, K., Somer, M., Peippo, M., Ahvenainen, M., Doagu, F., Arvio, M., Kääriäinen, H., Van Esch, H., et al. (2014). X-exome sequencing in Finnish families with Intellectual Disability - Four novel mutations and two novel syndromic phenotypes. Orphanet J. Rare Dis. *9*, 49. https://doi.org/10.1186/1750-1172-9-49.

11. Boycott, K.M., Azzariti, D.R., Hamosh, A., and Rehm, H.L. (2022). Seven years since the launch of the matchmaker exchange: the evolution of genomic matchmaking. Hum. Mutat. *43*, 659–667. https://doi.org/10.1002/HUMU.24373.

12. Hamosh, A., Wohler, E., Martin, R., Griffith, S., Rodrigues, E.d.S., Antonescu, C., Doheny, K.F., Valle, D., and Sobreira, N. (2022). The impact of GeneMatcher on international data sharing and collaboration. Hum. Mutat. *43*, 668–673. https://doi.org/10.1002/HUMU.24350.

13. Osmond, M., Hartley, T., Johnstone, B., Andjic, S., Girdea, M., Gillespie, M., Buske, O., Dumitriu, S., Koltunova, V., Ramani, A., et al. (2022). PhenomeCentral: 7 years of rare disease matchmaking. Hum. Mutat. *43*, 674–681. https://doi.org/10.1002/HUMU.24348.

14. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature *596*, 583–589. https://doi.org/10.1038/S41586-021-03819-2.

15. Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera–a visualization system for exploratory research and analysis. J. Comput. Chem. *25*, 1605–1612. https://doi.org/10.1002/JCC.20084.

16. Rossi Sebastiano, M., Ermondi, G., Hadano, S., and Caron, G. (2022). AI-based protein structure databases have the potential to accelerate rare diseases research: AlphaFoldDB and the case of IAHSP/Alsin. Drug Discov. Today *27*, 1652–1660. https://doi.org/10.1016/J.DRUDIS.2021.12.018.

17. Savic, D., Partridge, E.C., Newberry, K.M., Smith, S.B., Meadows, S.K., Roberts, B.S., Mackiewicz, M., Mendenhall, E.M., and Myers, R.M. (2015). CETCh-seq: CRISPR epitope tagging ChIP-seq of DNA-binding proteins. Genome Res. *25*, 1581–1589. https://doi.org/10.1101/GR.193540.115.

18. Meadows, S.K., Brandsmeier, L.A., Newberry, K.M., Betti, M.J., Nesmith, A.S., Mackiewicz, M., Partridge, E.C., Mendenhall, E.M., and Myers, R.M. (2020). Epitope tagging ChIP-seq of DNA binding proteins using CETCh-seq. Methods Mol. Biol. *2117*, 3–34. https://doi.org/10.1007/978-1-0716-0301-7_1/COVER.

19. Kharchenko, P.V., Tolstorukov, M.Y., and Park, P.J. (2008). Design and analysis of ChIP-seq experiments for DNA-binding proteins. Nat. Biotechnol. *26*, 1351–1359. https://doi.org/10.1038/nbt.1508.

20. Li, Q., Brown, J.B., Huang, H., and Bickel, P.J. (2011). Measuring Reproducibility of High-Throughput Experiments. Ann. Appl. Stat. *5*, 1752–1779. https://doi.org/10.1214/11-AOAS466.

21. Landt, S.G., Marinov, G.K., Kundaje, A., Kheradpour, P., Pauli, F., Batzoglou, S., Bernstein, B.E., Bickel, P., Brown, J.B., Cayting, P., et al. (2012). ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. Genome Res. *22*, 1813–1831. https://doi.org/10.1101/GR.136184.111.

22. Lun, A.T.L., and Smyth, G.K. (2016). csaw: a Bioconductor package for differential binding analysis of ChIP-seq data using sliding windows. Nucleic Acids Res. *44*, e45. https://doi.org/10.1093/NAR/GKV1191.

23. 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. https://doi.org/10.1038/ng.2892.

24. Cooper, G.M., Stone, E.A., Asimenos, G., NISC Comparative Sequencing Program, Green, E.D., Batzoglou, S., and Sidow, A. (2005). Distribution and intensity of constraint in mammalian genomic sequence. Genome Res. *15*, 901–913. https://doi.org/10.1101/GR.3577405.

25. Connaughton, D.M., Dai, R., Owen, D.J., Marquez, J., Mann, N., Graham-Paquin, A.L., Nakayama, M., Coyaud, E., Laurent, E.M.N., St-Germain, J.R., et al. (2020). Mutations of the transcriptional corepressor ZMYM2 cause syndromic urinary tract malformations. Am. J. Hum. Genet. *107*, 727–742. https://doi.org/10.1016/J.AJHG.2020.08.013.

26. Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., et al. (2016). Analysis of protein-coding genetic variation in 60, 706 humans. Nature *536*, 285–291. https://doi.org/10.1038/nature19057.

27. Bertelsen, B., Tümer, Z., and Ravn, K. (2011). Three new loci for determining x chromosome inactivation patterns. J. Mol. Diagn. *13*, 537–540. https://doi.org/10.1016/J.JMOLDX.2011.05.003.

28. Machado, F.B., Machado, F.B., Faria, M.A., Lovatel, V.L., Alves Da Silva, A.F., Radic, C.P., De Brasi, C.D., Rios, Á.F.L., de Sousa Lopes, S.M.C., da Silveira, L.S., et al. (2014). 5meCpG epigenetic marks neighboring a primate-conserved core promoter short tandem repeat indicate X-chromosome inactivation. PLoS One *9*, e103714. https://doi.org/10.1371/JOURNAL.PONE.0103714.

29. Mariani, V., Biasini, M., Barbato, A., and Schwede, T. (2013). lDDT: a local superposition-free score for comparing protein structures and models using distance difference tests.

Bioinformatics *29*, 2722–2728. https://doi.org/10.1093/BIO-INFORMATICS/BTT473.

30. Kumar, M., Michael, S., Alvarado-Valverde, J., Mészáros, B., Sámano-Sánchez, H., Zeke, A., Dobson, L., Lazar, T., Örd, M., Nagpal, A., et al. (2022). The Eukaryotic Linear Motif resource: 2022 release. Nucleic Acids Res. *50*, D497–D508. https://doi.org/10.1093/NAR/GKAB975.

31. Johnson, D.S., Mortazavi, A., Myers, R.M., and Wold, B. (2007). Genome-wide mapping of in vivo protein-DNA interactions. Science *316*, 1497–1502. https://doi.org/10.1126/SCIENCE.1141319/SUPPL_FILE/JOHNSON.SOM-5-30.PDF.

32. Petrovski, S., Wang, Q., Heinzen, E.L., Allen, A.S., and Goldstein, D.B. (2013). Genic intolerance to functional variation and the interpretation of personal genomes. PLoS Genet. *9*, e1003709. https://doi.org/10.1371/journal.pgen.1003709.

33. Landrum, M.J., Lee, J.M., Benson, M., Brown, G., Chao, C., Chitipiralla, S., Gu, B., Hart, J., Hoffman, D., Hoover, J., et al. (2016). ClinVar: public archive of interpretations of clinically relevant variants. Nucleic Acids Res. *44*, D862–D868. https://doi.org/10.1093/nar/gkv1222.

34. Chong, J.X., Yu, J.H., Lorentzen, P., Park, K.M., Jamal, S.M., Tabor, H.K., Rauch, A., Saenz, M.S., Boltshauser, E., Patterson, K.E., et al. (2016). Gene discovery for Mendelian conditions via social networking: de novo variants in KDM1A cause developmental delay and distinctive facial features. Genet. Med. *18*, 788–795. https://doi.org/10.1038/GIM.2015.161.

35. Kumble, S., Levy, A.M., Punetha, J., Gao, H., Ah Mew, N., Anyane-Yeboa, K., Benke, P.J., Berger, S.M., Bjerglund, L., Campos-Xavier, B., et al. (2022). The clinical and molecular spectrum of QRICH1 associated neurodevelopmental disorder. Hum. Mutat. *43*, 266–282. https://doi.org/10.1002/HUMU.24308.

36. ENCODE Project Consortium, Kundaje, A., Aldred, S.F., Collins, P.J., Davis, C.A., Doyle, F., Epstein, C.B., Frietze, S., Harrow, J., Kaul, R., et al. (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57–74. https://doi.org/10.1038/NATURE11247.

37. Nasser, J., Bergman, D.T., Fulco, C.P., Guckelberger, P., Doughty, B.R., Patwardhan, T.A., Jones, T.R., Nguyen, T.H., Ulirsch, J.C., Lekschas, F., et al. (2021). Genome-wide enhancer maps link risk variants to disease genes. Nature *593*, 238–243. https://doi.org/10.1038/S41586-021-03446-X.

38. Gupta, A.R., Pirruccello, M., Cheng, F., Kang, H.J., Fernandez, T.V., Baskin, J.M., Choi, M., Liu, L., Ercan-Sencicek, A.G., Murdoch, J.D., et al. (2014). Rare deleterious mutations of the gene EFR3A in autism spectrum disorders. Mol. Autism. *5*, 31. https://doi.org/10.1186/2040-2392-5-31.

39. Samocha, K.E., Robinson, E.B., Sanders, S.J., Stevens, C., Sabo, A., McGrath, L.M., Kosmicki, J.A., Rehnström, K., Mallick, S., Kirby, A., et al. (2014). A framework for the interpretation of de novo mutation in human disease. Nat. Genet. *46*, 944–950. https://doi.org/10.1038/ng.3050.