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.

Hiatt et al., 2023, The American Journal of Human Genetics *110*, 215–227
February 2, 2023
https://doi.org/10.1016/j.ajhg.2022.12.007

CellPress

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]Obestetrics 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 cutoff, 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., downsampled 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 dysmorphism	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 dysmorphism	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

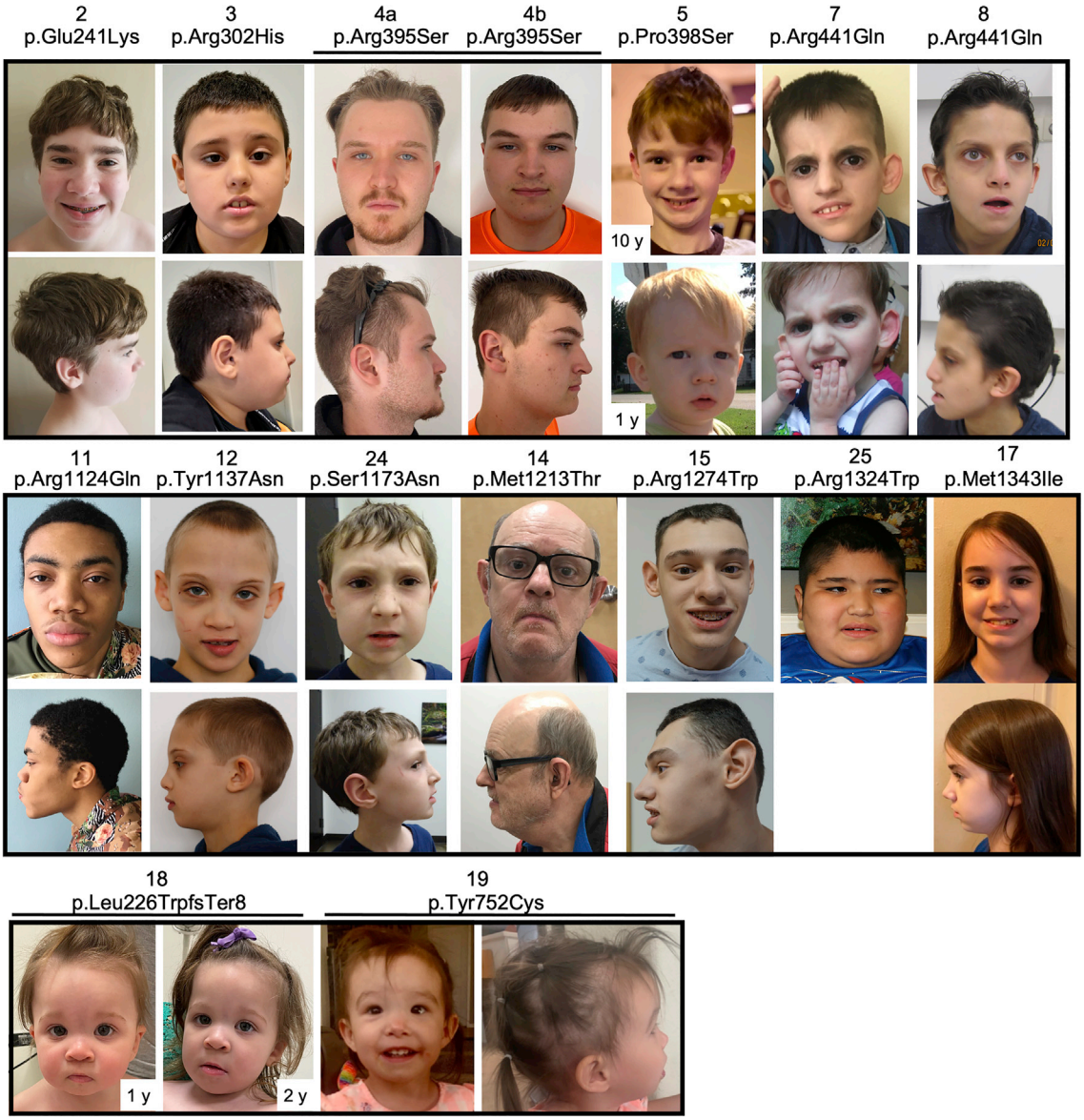


Figure 2. Facial features of a subset of individuals with *ZMYM3* variation
Individual ID and protein effect are noted for each. Note deep-set eyes, long palpebral fissures, large/prominent/cupped ears, and tall forehead.

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.

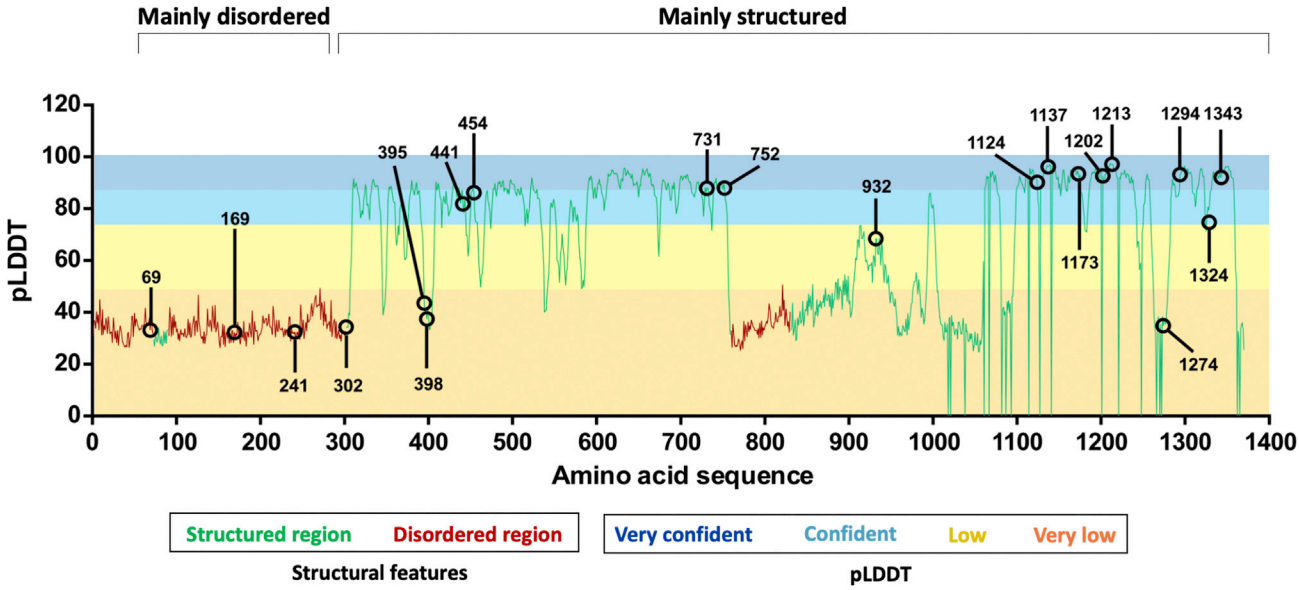


Figure 3. Missense variants in ZMYM3 mainly lie in ordered regions
Three disordered regions (red line) were identified (aa 1–72, 90–301, and 759–830), while the remainder of the protein is predicted to be structured (green line). AlphaFold produces a per-residue confidence score (predicted local distance 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

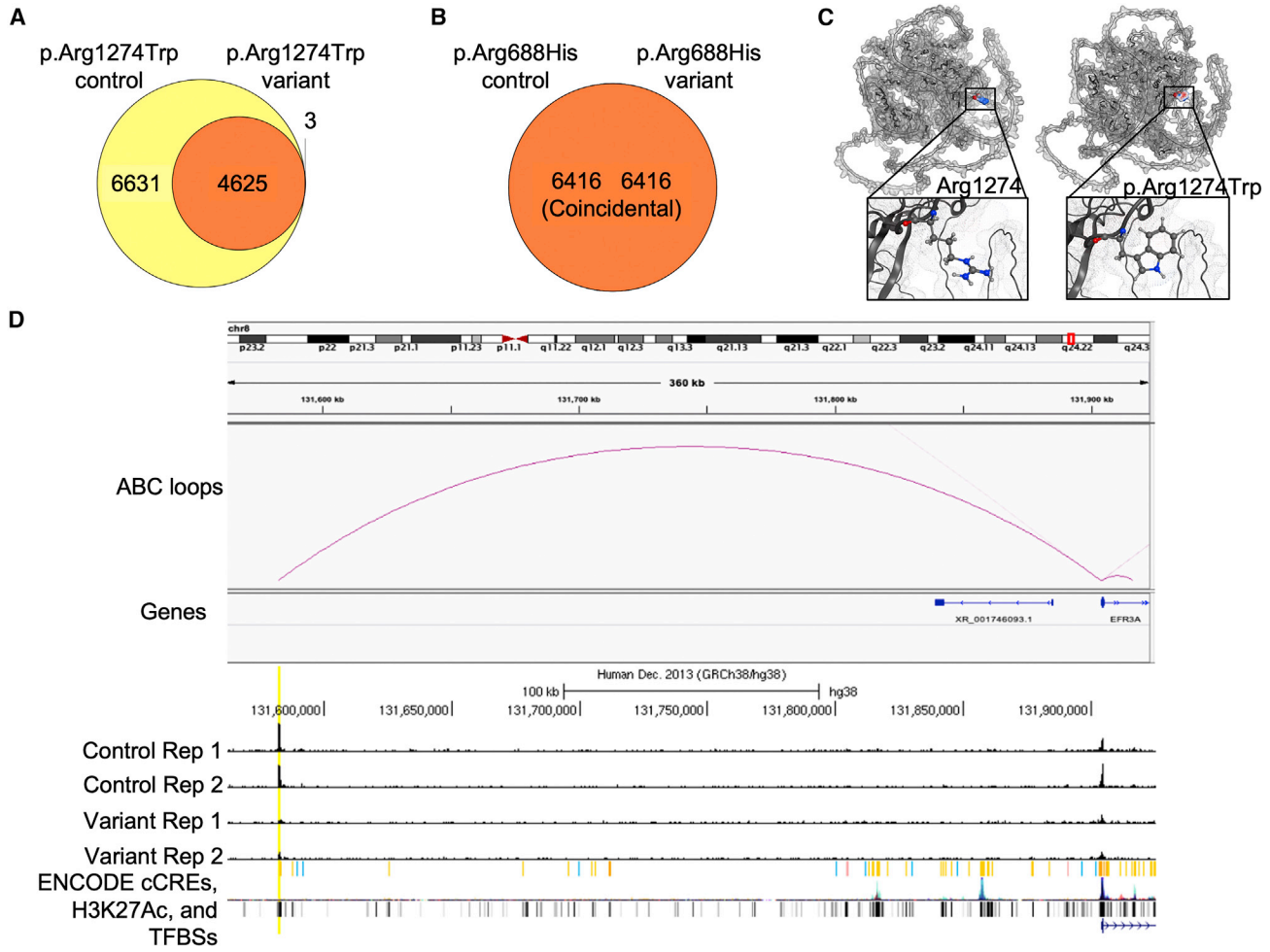


Figure 4. p.Arg1274Trp is a hypomorphic variant, while p.Arg688His has similar genome occupancy to that of wild type

(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 occupancy 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 Aicardi-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

We thank all the families who participated in this study. This work was supported by many funding sources, including the following: Alabama Genomic Health Initiative, an Alabama-State earmarked project (F170303004) through the University of Alabama in Birmingham (S.M.H., A.C., A.C.E.H., M.D., and G.M.C.); National Human Genome Research Institute (NHGRI) UM1HG007301 (S.M.H., G.M.C.); National Institute of Mental Health (NIMH) F31MH126628 (S.A.F.); National Institute of Nursing Research (Program EXCELES, ID Project No. LX22NPO5107), funded by the European Union, Next Generation EU (L.N.); UNCE/MED/007 of Charles University in Prague (L.N.); Ministry of Health of the Czech Republic, NV19-07-00136 (S.K.); Italian Ministry of Health (Ricerca 5x1000, RCR-2020-23670068_001, and RCR-2021-23671215) (M.T.); Italian Ministry of Research (FOE 2019) (M.T.), and PRIN2020 (code 20203P8C3X) (A. Brusco); Swiss National Science Foundation (31003A_182632) (A. Reymond); Blackswan Foundation (A. Reymond); ChildCare Foundation (S.E.A.); CRT Foundation (Program "Erogazioni Ordinarie" 2019) (G.C. and G.E.); Italian Ministry of University and Research (Assegni, Tornata 2022, Bando: BMSS.2022.06/XXIV) (M.R.S.); RVO VFN 64165, Czech Ministry of Health (M. Magner); Swiss National Science Foundation grant 320030_179547 (A. Rauch); and The Genesis Foundation for Children (C.B.N., J.D.). L.N. and S.K. thank the National Center for Medical Genomics (LM2018132) for exome-sequencing analyses. Sequencing and analysis of one individual in this study was made possible by the generous gifts to Children's Mercy Research Institute and Genomic Answers for Kids program at Children's Mercy Kansas City. Reanalysis of exome sequencing for individual 14 was performed on a research basis by the Care4Rare Canada Consortium.

Declaration of interests

J.L.B., Y.C., B.R.L., M.P.N., A.G.N., and H.Z.E. are employees of GeneDx, LLC. S.E.A. is a cofounder and CEO of MediGenome, the Swiss Institute of Genomic Medicine. All other authors declare no competing interests.

Received: September 1, 2022
Accepted: December 8, 2022
Published: December 30, 2022

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/BIOINFORMATICS/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.

The American Journal of Human Genetics, Volume *110*

Supplemental information

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

Susan M. Hiatt, Slavica Trajkova, Matteo Rossi Sebastiano, E. Christopher Partridge, Fatima E. Abidi, Ashlyn Anderson, Muhammad Ansar, Stylianos E. Antonarakis, Azadeh Azadi, Ruxandra Bachmann-Gagescu, Andrea Bartuli, Caroline Benech, Jennifer L. Berkowitz, Michael J. Betti, Alfredo Brusco, Ashley Cannon, Giulia Caron, Yanmin Chen, Meagan E. Cochran, Tanner F. Coleman, Molly M. Crenshaw, Laurence Cuisset, Cynthia J. Curry, Hossein Darvish, Serwet Demirdas, Maria Descartes, Jessica Douglas, David A. Dyment, Houda Zghal Elloumi, Giuseppe Ermondi, Marie Faoucher, Emily G. Farrow, Stephanie A. Felker, Heather Fisher, Anna C.E. Hurst, Pascal Joset, Melissa A. Kelly, Stanislav Kmoch, Benjamin R. Leadem, Michael J. Lyons, Marina Macchiaiolo, Martin Magner, Giorgia Mandrile, Francesca Mattioli, Megan McEown, Sarah K. Meadows, Livija Medne, Naomi J.L. Meeks, Sarah Montgomery, Melanie P. Napier, Marvin Natowicz, Kimberly M. Newberry, Marcello Niceta, Lenka Noskova, Catherine B. Nowak, Amanda G. Noyes, Matthew Osmond, Eloise J. Prijoles, Jada Pugh, Verdiana Pullano, Chloé Quélin, Simin Rahimi-Aliabadi, Anita Rauch, Sylvia Redon, Alexandre Reymond, Caitlin R. Schwager, Elizabeth A. Sellars, Angela E. Scheuerle, Elena Shukarova-Angelovska, Cara Skraban, Elliot Stolerman, Bonnie R. Sullivan, Marco Tartaglia, Isabelle Thiffault, Kevin Uguen, Luis A. Umaña, Yolande van Bever, Saskia N. van der Crabben, Marjon A. van Slegtenhorst, Quinten Waisfisz, Camerun Washington, Lance H. Rodan, Richard M. Myers, and Gregory M. Cooper

Supplemental Note: Case Reports

Individual 1, p.Asp69Asn

Patient was initially referred to the Genetics clinic at 11 months of age for evaluation due to a history of hypoplastic right ventricle following a bidirectional Glenn procedure. He also had a history of hypocalcemia, which had resolved, but consideration for 22q11.2 deletion syndrome was entertained, and a chromosome microarray was obtained which was normal. He was noted at that time to have developmental delays, specifically of gross motor skills as he was unable to crawl, and axial hypotonia. He had no other major findings, and his exam was otherwise mostly normal. Family history was notable for the maternal grandmother and maternal uncle with significant delays in speech acquisition in childhood, although they had no other delays and no current intellectual disabilities. He had a healthy younger full sister and three paternal half-sisters. At the follow-up visit at 19 months of age, he had made little to no developmental progress. Interim brain MRI had revealed abnormalities of the corpus callosum. There was no history of any seizures or loss of previously acquired skills. His developmental delays and findings on MRI were felt to be out of proportion with his history of the complex heart defect and excellent postsurgical progress. Thus, further expanded gene testing was ordered that revealed a maternally inherited *ZMYM3* variant. No additional variants were identified. At three years old, he continues to have gross motors delays and can pull to stand. He is also now noted to have significant speech delay, using babbling and vocalizations but no purposeful words.

Individual 2, p.Glu241Lys

Individual 2 is a 14 year old male with learning difficulties, imbalance, excessive fatigability, heat intolerance, oculomotor dysfunction, and history of growth retardation secondary to insulin-like growth factor 1 (IGF1) deficiency.

This male was born at 35 5/7 weeks gestation by spontaneous vertex vaginal delivery with birth weight 2.61 kg and length 45.7 cm to a G2P1 29 y/o mother and 38 y/o father. The prenatal history was notable for short femurs and humeri and a pericardial effusion. His mother had type I diabetes and hypertension during the pregnancy; the diabetes was under good metabolic control and the hypertension treated with several anti-hypertensive agents. At 20 days of age he was noted to have asymptomatic bilateral pulmonary artery stenoses and a small PFO. Torticollis was noted at 2 months of age. Since early childhood, his clinical phenotype has included dysmorphisms, developmental and neurological issues, and endocrine/growth issues.

A dysmorphology evaluation at 22 months showed head circumference 49 cm (68%; Z = +0.46), weight 8.7 kg (<1%; Z = -3.06) and height 80.3 cm (7%; Z = -1.49). Other notable measurements then included inner canthal distance 2.4 cm (3%), interpupillary distance 4.6 cm (~50%), right hand 9 cm (2%) and right foot 12.2 cm (8%). The anterior fontanelle was fingertip open and there was midfacial hypoplasia, upslanting palpebral fissures, broad nasal root, unilateral single palmar crease, 5th finger clinodactyly, hypermobility of the hips and ankles and mild hypotonia. A dysmorphology evaluation at 9.25 y/o, while receiving treatment with Increlex (recombinant human insulin-like growth factor 1), showed head circumference 55.1 cm (97%; Z = +1.88), weight 27.6 kg (39%; -0.27 SD), height 128.6 cm (21%; -0.81 SD),

ICD 3.0 cm (50%), IPD 5.7 cm (75-97%), palpebral fissure 3.2 cm (>2SD), right hand 14.5 cm (25-50%) and right foot 19.5 cm (3-25%). There was relative macrocephaly and brachycephaly, a prominent forehead with subtle bossing, deep-set eyes, upslanting palpebral fissures, a depressed and broad nasal root, bulbous nasal tip and small chin. Since then, there has been continued height, weight and head growth and the facial features have changed, possibly related to age and/or his medical treatment. Exam at 14 y/o showed head circumference 60.2 cm (>99%; Z = +3.69), weight 71.7 kg (94.5%; Z = 1.60), height 162.5 cm (42.7%; Z = -0.18) and distinctive craniofacial features including marked macrocephaly with prominent forehead, deep-set eyes with upslanting and long palpebral fissures, synophros, increased interpupillary distance and bulbous nasal tip.

Underweight and short stature were noted during infancy; weight and length at 15 months were 7.79 kg (<1%; -3.03 SD) and 70.4 cm (<1%; -2.84 SD). A skeletal survey at 21 months was normal apart from a widely patent anterior fontanelle. Serum IGF1 at 23 months was <25 (NL: 63-279 ng/mL) and repeated on a separate occasion; routine serum chemistries, CBC, serum CRP and WSR and TSH were normal and IGFBP3 low normal. His early childhood clinical history was notable for heat and fasting intolerance and excessive fatigability. He underwent treatment with growth hormone for 7 months beginning at 29 months, without any increase in growth rate or in serum IGF1; GH was discontinued at 3 y/o. Shortly thereafter he was started on Increlex, with good responses in terms of increased serum IGF1 and in growth velocity. He has since been treated with Increlex, with continued good response for linear growth. His height at 13 y/o was 158 cm (57%; Z = +0.18). He had a tendency for hypoglycemia that predated treatment with Increlex and has a longstanding history, still current, of heat intolerance that is associated with irritability, poor attention, and reduced quality of thinking. He also lacks endurance for both gross and fine motor tasks of extended duration compared to age-matched peers. Recent growth parameters are noted above. Recent endocrine data suggest recent endogenous production of IGF1. The basis for the previous short stature is still unexplained.

Developmentally, the proband did not roll over or crawl until one year and walked independently at about 19 months of age; early social, language and fine motor milestones were unremarkable. He received physical therapy since 1 y/o for reduced strength and gross motor delays and wore supramalleolar orthoses from 11 months to 4 y/o. Gross motor delays continued during childhood with poor balance and excessive clumsiness and falls as well as fine motor clumsiness. Gross and fine motor function are currently age-appropriate. He presently has intermittently poor balance and sustains occasional falls, mostly when he is very tired or hot, and has intermittent unexplained left leg collapse. At about 9.5 years, he developed myoclonic jerks of the extremities that occur when ill or very tired, as well as an excessive startle response to unexpected bright light. He has longstanding difficulty with visual tracking that, in turn, causes difficulty with reading. Visual acuity and hearing are unremarkable. There are no behavioral concerns and social skills are normal. He has an individualized educational plan for his schooling. He is in inclusion classes for music, science, social studies, and language arts. His math/calculation skills are not at age level and are felt to be several grades lower; overall, his neuropsychological profile is complex with splintered skills. It has been noted that his cognitive function can vary from day-to-day and that he

sometimes has loss of an established intellectual competency that is regained a few days later.

The family history is notable for a 16 y/o brother who has a diagnosis of high functioning autism, a normal IQ, learning difficulty and a past history of speech delay; he has had little diagnostic testing. The proband's mother, 43 y/o, has type I diabetes since 16 y/o, celiac disease, nephrolithiasis, a bout of lymphocytic colitis at 35 y/o and a past history of a learning disability. She has normal stature and upslanting palpebral fissures. Two of three maternal uncles have normal stature and past histories of speech delays and learning disabilities. A maternal aunt has mild short stature and a history of a learning disability. The maternal grandmother has mild short stature, hypothyroidism and a history of a learning disability. The proband's father repeated one year of elementary school (unknown reason) and his family has largely non-contributory clinical histories except for a sister who is deceased at one day of life (unknown etiology) and a niece with speech delay and scoliosis; he has an ~209 kb deletion of 16p12.2.

Diagnostic testing for the proband includes the following normal studies: MRI pituitary (2.5 y/o), MRI brain (11 y/o), skeletal survey at 21 months (except for widely open anterior fontanelle), EEG (9.5 y/o), ECG, thyroid function tests, serum cortisol, blood lactate and pyruvate, blood ammonia, serum urate, serum CK and aldolase, plasma amino acids, plasma acylcarnitines, and urine organic acids. H19 methylation analysis showed normal methylation at DMR1. Uniparental disomy analysis showed biparental inheritance of chromosome 7. Sequencing of the mitochondrial genome did not reveal any pathogenic mutations or deletions. Exome analysis was unrevealing except, possibly, for maternally inherited hemizygosity for a variant of uncertain significance of ZMYM3, c.721G>A (p.Glu241Lys). No other variants were identified from ES.

A chromosomal microarray analysis showed arr 15q11.2(20,224,751-20,852,202)x1 which was maternally-inherited, and 16p12.2(21,441,805-21,650,621)x1, which was paternally-inherited. Coordinates are based on human genome build 36.3. This indicated a maternally inherited copy number loss of 15q11.2 of about 627 kb and a paternally inherited copy number loss of 16p12.2 of about 209 kb. The loss of 15q11.2 contains 20 genes, 4 of which have OMIM entries (TUBGCP5, CYFIP1, NIPA2, NIPA1). This region is between breakpoint 1 and breakpoint 2 of the Prader Willi/Angelman region; these 4 genes are nonimprinted and conserved. The loss of 16p12.2 contains 3 OMIM gene entries (METTL9, IGSF6, OTOA).

Individual 3, p.Arg302His

Individual 3 is an 8 year old male. He was the product of a normal pregnancy, delivered by cesarean section at 36 weeks and 6 days. Birth weight was 3700 g., APGAR 9/10. The proband was breast fed for 5 months, with normal growth, and teething was observed as normal at this time.

At 4 months old, the proband had a diagnosis of pyelonephritis due to complete bilateral vesicoureteric reflux. The proband sat independently at 9 months of age and walked independently at 18 months of age. The proband exhibited lallation, but otherwise had an absence of language development. He was enrolled in speech therapy and speaks in simple sentences but has persistent pronunciation difficulties.

The proband acquired urinary continence at 5 years old with several episodes of incontinence at 8 years of age. He also has fecal incontinence.

The proband began first grade at age 7 years and has special education classes. He has little relationship with others and shows selective alimentation; he was diagnosed with autism at 2.5 years. The proband exhibits behavioral disturbances, including intolerance to frustration, episodes of aggressivity, and bruxism. He has displayed motor stereotypies since age 5 years, and he has had regular nocturnal awakenings since age 6 years. The proband also has persistent gastric reflux. He has strabismus in his right eye, and lenses were prescribed but he wasn't compliant. Hearing is normal. No MRI or EEG studies have been performed yet.

The proband has a paternally-inherited duplication identified by arrayCGH (arrayCGH: 203-972 Kb duplication in 2q13, paternal arr[GRCh37]2q13(110427255x2,110841715_111044815x3,111399242x2) arr[GRCh37](XY)x). Fragile X testing indicated he has 30 CGG repeats. No other variants were identified by ES other than the *ZMYM3* variant described here.

The proband has two healthy brothers, and no relevant diseases or consanguinity reported in the family.

Individuals 4a, 4b, p.Arg395Ser

Individuals 4a and 4b are full siblings. Individual 4a was born at term and required resuscitation and intensive care after delivery (Apgar Scores of 1,6). In infancy, he was followed up for transient neutropenia and recurrent vomiting. Although he walked independently at one year of age, some motor delay was present that required rehabilitation. Speech was delayed, and communitive speech was not present until three years of age. His behavior was often reported to be stereotypic, marked with aggressivity and throwing things. He was reported to have a lack of need for social interaction at an early age and later was diagnosed with high-functioning autism. His intelligent quotient is low normal (IQ 81). Individual 4a underwent surgery for hypospadias (at 6 years) and bruises easily. He is 24 years old, now. His aggression is controlled by boxing with a punching bag; however, his level of self-care is low.

Individual 4b had autistic features notable in infancy including little eye contact, lack of interaction/affection with his mother, and had poor sleep. Although he was reported to have some single syllable vocalizations at the age of one, loss of the speech occurred after a febrile infection. He has no functional speech. He can say some words, but these appear to represent echolalia. He was also reported to have stereotypic behaviors and suffered from aggression and auto-aggression in the form of biting. Individual 4b has a diagnosis of low-functioning autism and is reported to have severe intellectual disability (IQ ~35). He also bruises easily. Other than the *ZMYM3* variant, no other potentially causal variants were identified by data analyses.

Individual 5, p.Pro398Ser

A male patient was seen in the outpatient Genetics clinic at the age of 9 years old due to global developmental delay. Speech was more delayed than motor skills, but all were behind and noted at 2 years of age. He was also diagnosed with ADD. He required special education classes in school. He was evaluated by a developmental pediatrician who noted he did not have Autism, but he was diagnosed with sensory processing

disorder. Overall health was good. An Autism/ID NGS panel at GeneDx was done revealing a VUS in the X-linked *ZMYM3* gene (c.1192C>T, p.(Pro398Ser)), with no additional variants reported. Mom was found to carry this same genetic variant and has no learning or health problems. This individual has two older brothers, one of whom has ADHD.

Individual 6, p.Arg441Gln

Individual 6 is a now 10 yo male followed by Genetics due to dysmorphic features, trigonocephaly and autism. He has a complex neurobehavioral history which includes global developmental delays, intellectual impairment, autistic spectrum disorder problems and impulse control. In the past he had problems being a very selective eater with GI difficulties, problems swallowing and feeding which have markedly improved over time. He has a history of exotropia treated with bilateral recessions. His other medical issues include mild thinning of the splenium of the corpus callosum, mild-moderate persistent asthma, eczema, chronically low WBCs and neutropenia, chronic functional constipation and a single renal cyst. The proband's mother had history of ADHD and had an IEP during school years. The proband has a paternal half-sister with ADHD and depression and a maternal half-sister with autism spectrum disorder, ADD, anxiety and possibly dyslexia. There is some distant maternal and paternal family history of autism spectrum disorder and possible ADHD. Other than the *ZMYM3* variant discussed here, no other variants were reported.

Individual 7, p.Arg441Gln

Individual 7 is a 13-year-old boy, a fourth child in a family of European ancestry. He has two healthy sisters. His third sister has severe intellectual deficiency due to a major chromosomal anomaly (maternal isodicentric 15q11.2-q13.1 supernumerary chromosome resulting in tetrasomy). His parents have normal karyotypes. The proband had a normal peri- and postnatal period, but global developmental delay was detected in infancy. The proband exhibits hypotonia, and held his head at 1 year, started to sit at 5 years. He never developed fine motor skills. When put in his hand, he holds a toy but rarely moves it to the other hand. He has a short attention span and is frequently upset but does make eye contact. On clinical examination were also noted hypotrophy of the muscles, and diminished reflexes. The proband never acquired toilet training. His height and weight are deeply below the 3rd percentile. He also has cryptorchidism, and a brain MRI showed enlarged ventricles. Dysmorphic features included coarse and triangular face, widow's peak, long forehead, thick eyebrows, long palpebral fissures, deeply set eyes, blue sclerae, and flashy ears with cupped formed ear lobes. The proband has normally functioning heart and kidneys, and basic biochemical analyses are routinely normal. Genetics testing included FMR1 repeat expansion (Fragile X syndrome) and array-CGH which were negative.

In addition to the ZMYM3 variant discussed here, we also found two compound heterozygous variants (NM_000512.5:c.714del, p.(V239Sfs*80) and c.499T>G, p.(F167V)) in *GALNS* (*612222), a gene associated with Mucopolysaccharidosis IVA (*253000), an autosomal recessive disease. The gnomAD v.2.1.1 database reports at least one p.(F167V) homozygous subject and ClinVar (VCV000321204.9) reports a conflicting interpretation (likely pathogenic/likely benign). However, this variant has been

shown to have reduced enzyme activity (PMID: 1786718). Mucopolysaccharidosis IVA is a severe disease, incompatible with our patient's phenotype; however, we concluded that the biallelic combination of a strong LOF and p.(F167V) in may lead to mild features of Mucopolysaccharidosis IVA, such as the skeletal phenotype of our proband.

Individual 8, p.Arg441Gln

Individual 8 is a now 17y old boy born at 39w3d weighing 2690gr(-1.5 SD); OFC 33.5 cm (-1.5SD); Apgar 7 and 9. Ambiguous genitalia were noted with rugated labioscrotal walls, phallus of 1.4 cm, soft gonads palpable; chromosomes were 46,XY and on ultrasound no Müllerian structures . He was assigned the male gender with scrotal hypospadias. He had a small VSD and ASD type II. Dysmorphisms were fleshy nose, retrognathia, square ears, and simian creases. He was hypotonic with severe delayed development, but with good hearing and vision. He began walking at age 3y. He could not tolerate solid food at age 8 y, had no speech at 15, but was starting to be toilet trained. Dysmorphisms include a hypoplastic midface, protruding ears, thin lips, small hands. At age 15, height was 138.5cm (-5.12 SD), weight 34kg (W/L -1.12 SD). OFC at age 8y was 49cm (-2SD). The proband had a normal multiple congenital anomalies (MCA) sequencing panel, followed by trio exome sequencing which identified the *ZMYM3* variant. The proband's mother has short stature and had a mild learning problem, although she does not need help for her daily functioning.

Individuals 9a, 9b, p.Glu731Asp

Individual 9a and 9b represent two full brothers born from consanguineous parents of Iranian origin. Both affected individuals present with autism and mild intellectual disability. They are reported with normal motor activity. No MRI anomalies have been detected. Both boys have mild facial dysmorphism including long face, high anterior hairline, deep eyes, long philtrum and thin lips. A hemizygous missense variant in *ZMYM3* was identified in both affected brothers. The variant is heterozygous in the mother but absent in the father and the unaffected sister.

Individual 10, p.Ile932Val

Individual 10 was born to a G8P2 mother with dyslexia. The pregnancy was complicated by hypertension, gestational insulin-dependent diabetes, breech presentation, a history of premature births (reportedly caused by low progesterone), and maternal obesity. There were normal ultrasound exams. The proband was delivered at 32 weeks' gestation by c-section delivery due to maternal hypertension and required full resuscitation in the delivery room. At birth he weighed 2020 grams, was 42 cm long, and had an OFT of 31.6 cm. He was discharged home after 2 months in the NICU. At 4 months he was noted to have slightly downslanting palpebral fissures, low set, prominent ears, widely spaced nipples, and mild hypotonia. At three years of age, he has plagiocephaly, torticollis, recurrent acute otitis media, feeding disorder of early childhood, gastro-esophageal reflux disease without esophagitis, constipation, and expressive language delay.

Individual 11, p.Arg1124Gln

Individual 11 is a now 20yo male who first was referred to genetics at age 16 yo due to scoliosis. He was born at term to a 22yo G1 mother, birth weight of 6 pounds, 5 ounces, with no pregnancy or delivery complications. At 3.5 years he was diagnosed with severe delays in receptive and expressive language and poor interaction skills. He was homeschooled after age 11 due to learning and behavior concerns. He left school after the 9th grade and is now attempting to obtain a GED (general educational development) degree. He has oppositional defiant disorder (ODD) and attention deficit hyperactivity disorder (ADHD).

Scoliosis was noted since 9 years of age, which progressed to a left thoracic curve from T2-T7 of 60 degrees, right thoracic curve from T7-L3 of 77 degrees, and left lumbar curve L3-L5 of 38 degrees with kyphosis of 70 degrees. He required spinal fusion surgery at age 16. He needed a tonsillectomy/adenoidectomy at 4 years of age. Brain MRI was normal. He has irregular heartbeats which were evaluated by cardiology with long-term home monitoring, but no treatment was needed.

His physical exam at age 18 years was significant for short stature, long palpebral fissures (3.5 cm (>99%)) which are narrow in height and upslanting, small ears (5.2 cm (1%)) which are posteriorly rotated and low-set, a depressed nasal bridge with wide, thickened and wide nasal alae. He has slight macroglossia with inability to view base of uvula, prominent lips with upturned upper vermillion border, and prognathism.

Prior testing included a microarray which demonstrated a variant of uncertain significance deletion on the X chromosome (Xq26.3-q27.1) which involves *FGF13*. Lysosomal enzyme testing was normal. He was referred for research whole genome sequencing, which revealed the maternally-inherited hemizygous *ZMYM3* variant (c.3371G>A,p.(R1124Q)).

The proband has a maternal aunt with a history of intellectual disability who was in special education classes, but no additional information is known.

Individual 12, p.Tyr1137Asn

Proband 12 is a male born at 38 5/7 weeks gestation by emergency C-section for abnormal fetal positioning after a pregnancy complicated by severe maternal nausea requiring repeated infusions. Measurements at birth were 2780g, 50cm length and 32cm OFC. Initial feeding problems required naso-gastric tube feedings and two weeks hospitalization. Gross motor delays were noted with independent sitting at age 8 months and independent walking at 19 months. Speech development was initially very delayed but progressed more rapidly after tonsillectomy at age 2 years. At the last clinic visit at age 5 years, he was speaking in full sentences albeit with some pronunciation difficulties. He goes to regular Kindergarten where he receives special support. His behavioral profile includes a low tolerance to frustration and some degree of aggressivity but no autism. No health issues besides frequent respiratory infections in infancy and as a toddler. He has no epilepsy and an EEG performed because of a questionable seizure episode during an infection was normal. Family history is unremarkable except for a two-year-old brother with mild speech delay; physical appearance, behavior and overall development of this brother are otherwise unremarkable and very different from the proband's. The brother does not carry the *ZMYM3* variant, which we determined to have occurred *de novo* in the proband's

mother, who is healthy but shares facial features with the proband. No additional variants were reported along with the *ZMYM3* variant described here.

Individual 13, p.Val1202Asp

This individual is a 3-year-old male subject from Nigeria. He is the third-born of four children from healthy non-consanguineous parents. The family history was not contributive except for a record of mother's first-degree cousin with short stature and deafness. The pregnancy was unmonitored. He was born at 35 weeks of gestational age by induced delivery. Swallowing difficulties were reported in the first days of life.

At last clinical evaluation at the age of 3 years and 5 months, he displayed severe intellectual disability with delayed acquisition of all motor milestones such as head control (15 months), sitting position (24 months), and independent walking (36 months). Speech was always absent. He also displayed short stature and rhizo-mesomelia of major upper limbs, microcephaly with trigonocephalic appearance, and some facial dysmorphisms, which include hypertelorism, ptosis major left eyelid, ears with low cup-shaped implantation. Bilateral cryptorchidism was also documented and surgically corrected. Brain MRI revealed global enlargement of the subarachnoid spaces of the posterior cranial fossa. The skeletal X-ray scan documented several skeletal abnormalities. The spine presented with schisis of entire vertebral soma of L5, butterfly vertebrae at T3, median schisis of the posterior arch in T12, and schisis of the posterior arch in L5. Notably, at this level, the hemi lamina right does not merge with the left one but blends with the overlying posterior arch. Humerus and ulna bones were short and dysmorphic, radius was bowed, and Madelung deformity was also noticed bilaterally. Lower limbs presented with dysplastic epiphysis of both proximal fibulae. Biochemical testing and routine laboratory assessments were normal. Karyotype analysis was 46,XY, and a pericentric inversion of the chromosome 9, inv(9)(p11q13) was identified but also considered benign. From ES, the *ZMYM3* variant described here was the only clinically relevant variant reported. No other private/rare (MAF<0.01 in gnomAD) clinically relevant variants involving OMIM genes were identified.

Individual 14, p.Met1213Thr

Individual 14 is a 60 year old gentleman who presented to the Genetics Clinic given a history of cognitive deficits, progressive weakness, and family history of "muscular dystrophy".

His cognitive deficits appeared to represent a mild intellectual disability; however a formal assessment was not available. He attended a "special needs" school. He was unemployed and on social assistance. He describes always having a degree of weakness but that it has become progressively worse over the last few years (distal>proximal) and he was experiencing an increasing number of falls. He previously used a walker and had started to use a scooter. Associated with the weakness, he also experiences a numbness to hands and feet. He does have other health issues including scoliosis, diabetes, hypercholesterolemia, hypertension, asthma, and cataracts. He was estranged from his family for many years, and he does not know where they live or their current health status. Both parents were deceased. He has 4 brothers and 3 sisters. One of his sisters has a diagnosis of "muscular dystrophy" or possibly multiple sclerosis and requires a wheelchair. He is not aware of any genetic testing performed for his

sister. There is no family history of intellectual disability or learning disability in the family.

On examination his OFC was 53cm (-2SD), height was 178cm and weight was 98.1kg. Headshape had a mild plagiocephaly but otherwise normal. He had frontal balding and wore glasses. Eyes were mildly deep-set. Nose, philtrum, mouth, and palate were normally shaped. Ears were normally placed with normal architecture. Strength was mildly reduced (4+/5) to proximal upper extremities and to distal upper extremities (4/5). Hip flexion was diminished 4/5 bilaterally as was knee flexion and extension 4/5, and inversion, eversion, and dorsi- and plantar flexion (4/5).

Brain MRI was normal. Given the family history, he was tested for several genes associated with progressive weakness including myotonic dystrophy 1 and 2, OPMD and FSHD. CK was within normal limits. Given his intellectual disability he was also tested with a microarray (normal) and Fragile X testing (normal). A subsequent comprehensive intellectual disability panel was performed (Fulgent) and was non-diagnostic. However, two variants of uncertain significance and one pathogenic variant were observed. They included *ZMYM3*:c.3638T>C, p.(M1213T); *IGF1R*: NM_000875.4:c.3988G>A, p.(G1330S); and BBS1:NM_024649.4:c.1169T>G, p.(M390R). While this individual was found to be a *carrier of a BBS1 pathogenic variant* he had no retinal dystrophy, no polydactyly or hypogonadism, and no second mutation was observed. He was also found to have a VUS in IGFR1. This variant is seen 15x in gnomAD and the gene is associated with recessive condition. Further, his normal height and that his diabetes is well-controlled and late-onset. Given the lack of diagnosis he was enrolled in the Care4Rare research program. The Care4Rare research program was able to reanalyze the sequence data provided from Fulgent and the research group also identified the variant at *ZMYM3* as a potential candidate.

Individual 15, p.Arg1274Trp

Individual 15 was last evaluated at 16 years 3 months of age. He is a male with a history of global developmental delay, autism, and intellectual disability. Since the start of puberty, the major concern has been with his behavior: he has emotional outbursts, is quick to anger, and has impulsive and aggressive behaviors. He has profound microcephaly and dysmorphic features including angled palpebral fissures, protuberant ears, and a large nose. His extremities are thin with decreased muscle mass, knuckles on his hands are knobby, and he has right concave thoracic scoliosis. He wears glasses to treat myopia. Recently identified mildly dilated aortic root with no cardiac symptoms. General health is fine. No additional variants were reported other than the *ZMYM3* variant described here.

Individual 16, p.Arg1294Cys

The proband was a 26-week estimated gestational age (EGA) fetus of healthy unrelated parents. This was the first pregnancy of the couple. The first trimester ultrasound showed normal nuchal translucency (1.2 mm for a fetal crown-rump length of 60 mm). Estimated risk for trisomy 21 was 1/10,000. The 22-week scan showed a voluminous open dysraphism with significant hydrocephalus with septal rupture and a cardiac defect. Fetal karyotype and CGH-array on amniotic fluid were normal 46, XY.

The parents requested termination of pregnancy, and a male fetus was delivered at 26 weeks. Post-mortem examination showed a male fetus appropriate for gestational age. He had facial dysmorphism with high forehead, synophris, widely spaced eyes, low-set ears and microretrognathia. He had spinal dysraphism with malposition of the lower limbs and feet. External genitalia were abnormal with a median genital tubercle with urethral orifice at the base of the tubercle and unfused genital swelling. Autopsy showed pulmonary abnormal lobulation, thymic hypoplasia and a complex conotruncal heart defect with right aortic arch, overriding aorta, pulmonary stenosis and bicuspidism, atresia of the ductus arteriosus, atrial septal defect, cono-ventricular septal defect, and moderate right ventricular hypertrophy. The brain examination showed ventricular dilatation, right arhinencephaly and neuronal migration abnormalities (heterotopias).

Trio ES identified a *de novo* missense variant in *ZMYM3* and no additional variants.

Individual 17, p.Met1343Ile

Individual 17 was the product of a singleton gestation to a then 36 y/o G4O3->4 mom following a pregnancy that was complicated by hyperemesis gravidarum. He was delivered via vaginal delivery at 38 weeks' gestation with normal growth parameters (BW=3.997 kg; L=48.3 cm). He did well in the newborn period. He presented with severe hypoglycemia in the setting of a viral illness at 13 months of age and was subsequently diagnosed with ketotic hypoglycemia by Endocrinology. He still requires continuous GJ-tube feeds during the day due to rapid drop in glucose levels when off feeds. He also has oral aversion and GI dysmotility. At ~2.5 years of age, he was noted to have to have a R-sided facial palsy and his articulation and oromotor function declined at that time. Brain MRI and EMG/NCS were subsequently unremarkable. He has chronic joint pain and swelling for which he takes naproxen and is followed by Rheumatology. He is also followed by Cardiology due to history of dysautonomic symptoms that manifest as dizziness, dehydration, diaphoresis, overheating, headache, and fatigue/decreased endurance. The proband's history is otherwise significant for ocular tracking issues, recurrent otitis media s/p tympanostomy tubes, asthma, frequent UTIs, right distal femur osteochondroma (s/p resection), allergic rhinitis, and non-IgE mediated food allergies; diagnosis of mast cell activation syndrome raised but not confirmed.

Given his joint laxity and concern for possible connective tissue disorder with clinical diagnosis if hypermobile EDS in mother and three maternal half-sisters, he has undergone a Connective Tissue clinic evaluation and did not meet criteria for a primary connective tissues disorder.

He sat at 6-7 mo; walked at 10-11 mo, but never learned to ride a tricycle. He has issues with motor planning and coordination affecting gross and fine motor skills, balance issues; he uses a stroller for distance due to decreased endurance. His first single words and use of short phrases was on-time with some regression of language skills after emergence of R-sided facial palsy. He developed a stutter at ~5 years of age without an inciting event. He has age-level vocabulary but with a pragmatic language disorder and articulation issues (s/p normal palatal evaluation for VPI). He was diagnosed with autism spectrum disorder at 3 yo as well as general anxiety disorder and separation anxiety disorder. At 8 yr of age, he was reported to have age-level math

skills with below grade-level reading and writing skills. Formal neuropsychological testing data not available.

No variants were reported other than the *ZMYM3* variant discussed here.

Individual 18, p.Leu226TrpfTer8

Individual 18 is a female who was born at 38 weeks' gestation to a 37 year old primigravida Caucasian mother and 47 year old African American father. Both are cognitively normal, and their family histories negative for intellectual disabilities, learning disabilities, or fetal loss. At 34 weeks, mild IUGR and short femurs were noted. Delivery was by C/S and Apgars 8/8. Her birth weight was 2450 g (-2.13 SD), length 44.5 cm (-2.61 SD) and OFC 33 cm (-1.90 SD). She had some problems feeding and had symptoms of GERD in the neonatal period. She had early plagiocephaly and torticollis that resolved with helmet treatment. At 2.5 months she had a brief resolved unexplained episode (BRUE) leading to a short ICU admission.

Her growth has been normal; at age 2 her weight was 12.2 kg (0.02 SD), length 86.5 cm (0.25 SD) and OFC 46.5 cm (-0.94 SD). Development has been mildly delayed with sitting at 8 months, crawling at 12 months, and walking at age 2 years. She had delayed receptive language on assessment at 19 months. She has unprecipitated episodes of hand flapping.

The proband has dysmorphic features noted, including deep set eyes, flat nasal bridge with epicanthal folds, full cheeks, a broad nasal tip, long philtrum, small mouth, mild micrognathia and thin vermillion border of the upper lip and a pouting lower lip. Her hands and feet are normal. She has small nipples.

Testing has included a normal SNP microarray, an echocardiogram showing prominence of the papillary muscles, normal sleep and awake EEGs, and a normal swallow study. Exome sequencing studies revealed the maternally-inherited variant in *ZMYM3*. X-inactivation studies at one lab revealed 99% skewing in the child using the AR probe (whole blood sample). Mom's studies were not successful as she is homozygous for the *AR* alleles that were tested. Studies at a second lab using the *RP2* locus revealed skewing with a ratio of 94:6 in both mom and child. Both mom and child have alleles of the same size (366/362), and both individuals appear to be inactivating the same allele (366). No other variants were identified by ES.

Individual 19, p.Tyr752Cys

Individual 19 was a female infant born AGA at 37 3/7 weeks with uncomplicated pregnancy. Her newborn period was complicated by murmur on exam, and she was found to have atrial septal defect that closed spontaneously. Echocardiogram at 7 mo showed small PDA, PFO, and LSVC to coronary sinus, with ASD no longer present. Patient was first evaluated at 17 mo in genetics clinic for hypotonia, poor weight gain, and gross motor delays. History notable for sitting up at 8 mo, walking at 16 mo, and persistent central hypotonia. She initially had delays in speech development, but after therapies, she has now graduated from speech therapy. She has mild textural sensitivities, but overall, there are no concerns for autism spectrum disorder. On exam, patient had apparent telecanthus, bilateral epicanthal folds, flat midface, short nose, soft skin, and central hypotonia. She was observed to be unbalanced while walking and sitting in the tripod position throughout the visit. Growth chart showed her weight in 1st

percentile, length at 18th percentile, and OFC at 6th percentile. Initiated genetic testing with chromosomal microarray, which was negative. Reflexed to the GeneDx Autism/ID Xpanded Panel (trio), which found two variants of uncertain significance: *SCN2A* c.1421T>A, heterozygous, paternally inherited (not thought to be contributing to her features given lack of clinical correlation and unaffected father); and *ZMYM3* c.2255A>G, heterozygous, *de novo*.

Family history is significant for many paternal family members with stretchy skin, joint issues, and hypermobile joints. Cousins of father: one died in infancy of aortic complication, one 6'3" and "lanky", and one girl with low tone and atrial septal defect. Paternal great-great-aunt had twins who both had poor growth and low tone.

Individual 20, p.Arg1294Cys

Individual 20 is a 5 yo girl presenting with global developmental delay. Dysmorphic features include mild microcephaly, plagiocephaly, synophris, thin lips, protruding and badly hemmed ears. Hands and feet are marked by tapered fingers, bilateral single transverse palmar crease, bilateral camptodactyly of third and fourth toes. She sat at 16 month and walked at two years old. Language development is delayed. Other clinical features include volvulus of midgut, pyelectasis, pancreatic cysts. X-inactivation studies using the HUMARA assay indicated 97% skewing in this proband. Parental XCI testing not performed due to the *de novo* nature of her *ZMYM3* variant. No additional variants of interest were identified other than the *ZMYM3* variant discussed here.

Individual 21, p.Arg169Ser

This individual presented for an initial genetic evaluation at 10 years of age for moderate intellectual disability. He was born at 38 weeks gestation. Hypospadias was reported. He walked at 12 months, said his first words around 12 months, spoke in phrases at 3 years, and was toilet trained around 2 years. There was no developmental regression. His mother first noted significant concerns when he started school. He required repeating kindergarten and started attending special-education classes in the 1st grade. At the time of visit, he knew some letters but was not reading and had significant difficulties in all subjects. He was diagnosed with ADHD at 8 years of age. Additional behavior concerns included being easily frustrated and hitting himself. Around age 11, he was diagnosed with autism spectrum. Parents had one daughter born prematurely at 6-1/2 months' gestation; she died at 9 days of age. The child was not known to have any birth defects or other anomalies. No other family members were reported with intellectual disability or birth defects. There was no known history of consanguinity. Genetic testing including chromosomal microarray analysis and testing for Fragile X were normal. An X-linked Intellectual Disability Gene Panel was ordered, which revealed sequence variants in *ZMYM3* and *IL1RAPL1*. The *IL1RAPL1* variant (NM_014271.3:c.1039G>A (p.Val347Ile)) will now be classified as Likely Benign (BS2, BP4). There are 8 hets/2 hemizygotes in gnomAD v2.1.1, in BRAVO there are 13 hets/2 hemizygotes. One report in ClinVar as Likely benign.

Individual 22, p.Arg441Trp

This individual is a now 4 year old male referred to genetics for developmental delay. He had prenatal drug exposure, but there is no other information regarding pregnancy, delivery, or early developmental milestones. He has mild global

developmental delay, mild short stature, a triangular face, prominent forehead, long eyelashes, and single transverse palmar crease. At 4 years old, he is not toiled trained. He only has a few words but he seems to understand more than he can express. He points and has decreased eye contact. He was in a 3K special needs program last year. He does not sleep well. He has one sister with developmental delay, 2 brothers with autism, and one healthy brother. His mother is reported to have bipolar disorder and schizophrenia. There is no known information about his father or other family members. Other than the *ZMYM3* variant, no other variants were reported for this case.

Individual 23, p.Cys454Arg

Individual 23 is a 16-year-old male with global developmental delay, cognitive disorder not otherwise specified, ADHD, short stature, microcephaly, gastrointestinal dysmotility, myopia, and retinopathy. He was born at 36 weeks gestational age. The pregnancy was complicated by intrauterine growth restriction and polyhydramnios. The neonatal period was complicated by jaundice requiring phototherapy.

He has had feeding difficulties since early infancy and has been diagnosed with gastroparesis, ultimately requiring placement of a GJ tube. He underwent a nissen fundoplication for GERD. He also has chronic constipation. His development was globally delayed since birth. He walked at 25 months. Speech was also significantly delayed. He has been diagnosed with cognitive disorder not otherwise specified and ADHD. He also has emotional lability. He has had progressive myopia, and pigmentary abnormalities of the retina concerning for a retinopathy. He has short stature and growth hormone deficiency, on treatment with growth hormone. He also has microcephaly. He has outgoing toes secondary to femoral retroversion/external tibial torsion. He has a history of left-sided vesicoureteric reflux. Investigations include normal audiometry. Brain MRI's have demonstrated mild parenchymal volume loss that has been stable over time (last performed at age 11 years). Electroencephalogram (EEG) performed at 4 years of age was normal. Echocardiograms have been normal.

In terms of family history, he has a sister with articulation difficulties, and a brother with learning disability, dyslexia, and ADHD. There is no other contributory family history. Clinical trio whole exome sequencing reported a *de novo* hemizygous variant in the ZMYM3 gene (c.1360T>C, p.C454R). He also had a heterozygous VUS in the VPS13B gene and a heterozygous known pathogenic variant in the FLG gene.

Individual 24, p.Ser1173Asn

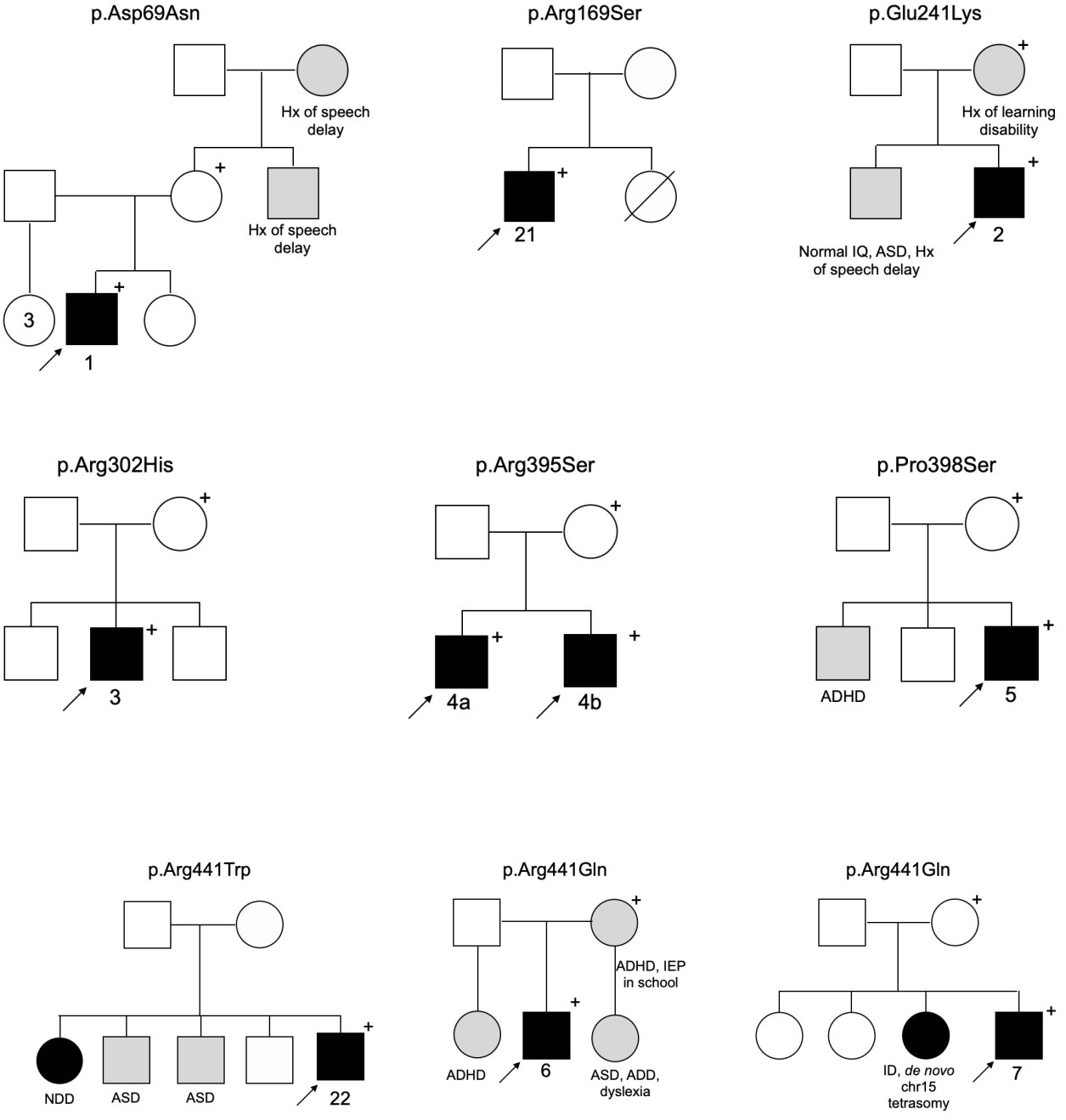
Individual 24 is an 8-year-old male with a history of autism and developmental delay. The proband had normal prenatal screening and was born to a then 17-year-old G1P0 mother via induced vaginal delivery at 40 weeks gestation. Birth weight was 7 pounds and 11 ounces with a birth length of 18 inches. Initial development was delayed, with sitting at 18 months, crawling at 2 years, walking at 2.5 years, and running at 4 years. The individual's first words were at age 2 years, with use of sentences beginning at age 5. He continues to make progress with speech, occupational, physical, and behavioral therapies. Upon examination, the proband was found to have ear size above the 99th percentile. No other dysmorphisms were noted. Family history was significant for ADHD and developmental delays in the proband's mother. A maternal uncle is also reported to

have autism, developmental delays, mild ID, reportedly large ears, and ADHD. Limited paternal history was available, though autism and learning issues were suspected. Initial genetic testing for the proband included chromosomal microarray and Fragile X testing at GeneDx, both of which were negative. Subsequent whole genome sequencing was pursued through the HudsonAlpha Clinical Services Lab, and the *ZMYM3* p.S1173N variant was identified. Sanger testing confirmed maternal inheritance of the variant. This individual is also carrier for a likely pathogenic variant in *ASPM* (NM_018136.5:c.77delG, p.Gly26AlafsTer42) and a likely pathogenic variant in *RARS2* (NM_020320.5:c.35A>G, p.Gln12Arg=). Neither of these are considered a strong phenotypic fit, and no second hits were observed for either gene.

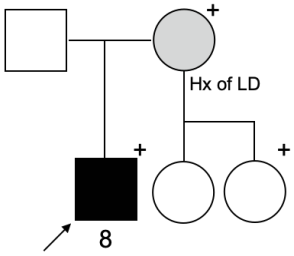
Individual 25, p.Arg1324Trp

This individual is an 8-year-old male first seen by genetics age 3. He was born full term to a 34-year-old female. He was referred to genetics for developmental delay and autism. He had aggressive behaviors as well. He walked at 15 months of age and had very limited language development. He was difficult to examine due to hyperkinesis but did not have major dysmorphic features. His weight was greater than 99th percentile.

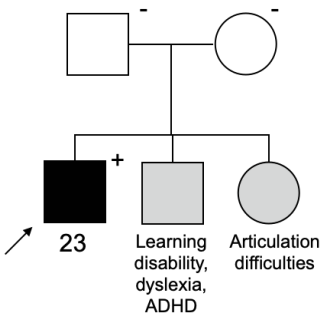
Figure S1. Pedigrees for each family in the study. Variants are shown above the pedigree, with *de novo* variation in red. Affected individuals presented here are indicated with arrows and labeled with IDs matching the text (1-23). Individuals affected with an NDD are shown in black, while mildly-affected individuals, those with a single feature of an NDD, or a history of such are shown in gray. Presence of the variant is indicated with +, while absence of the variant (if tested) is indicated by a -. Numbers in unaffected individuals' shapes indicate the number of unaffected siblings (or half-siblings) when there are more than two.



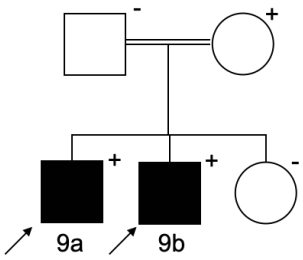
p.Arg441Gln



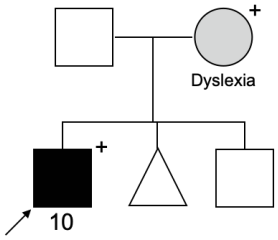
p.Cys454Arg



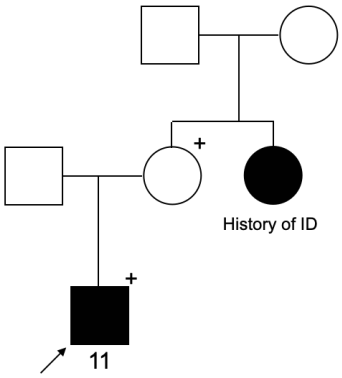
p.Glu731Asp



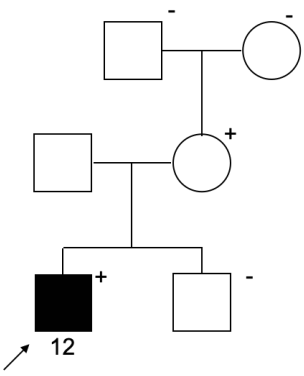
p.Ile932Val



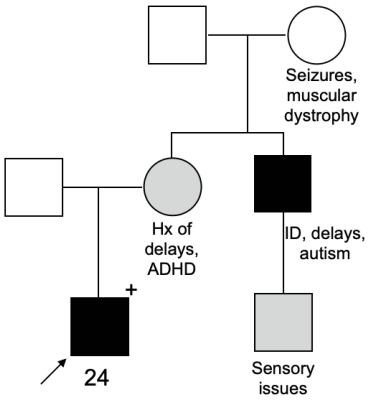
p.Arg1124Gln



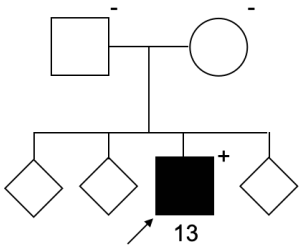
p.Tyr1137Asn



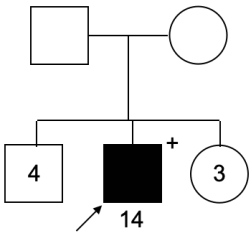
p.Ser1173Asn



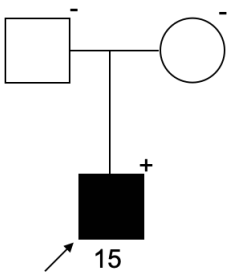
p.Val1202Asp



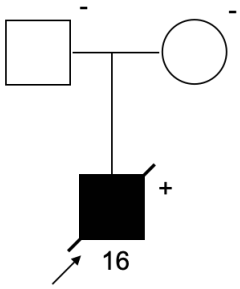
p.Met1213Thr



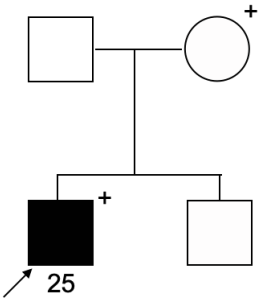
p.Arg1274Trp



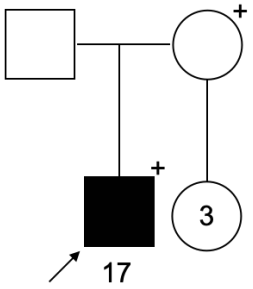
p.Arg1294Cys



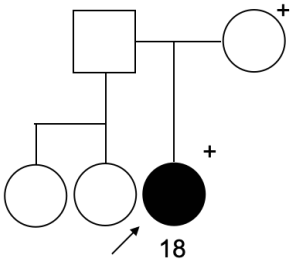
p.Arg1324Trp



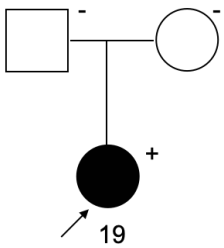
p.Met1343Ile



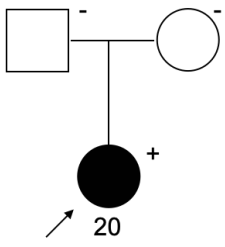
p.Leu226TrpfsTer8



p.Tyr752Cys



p.Arg1294Cys



NDD

Mildly-affected with NDD, single feature of an NDD, or a history of such

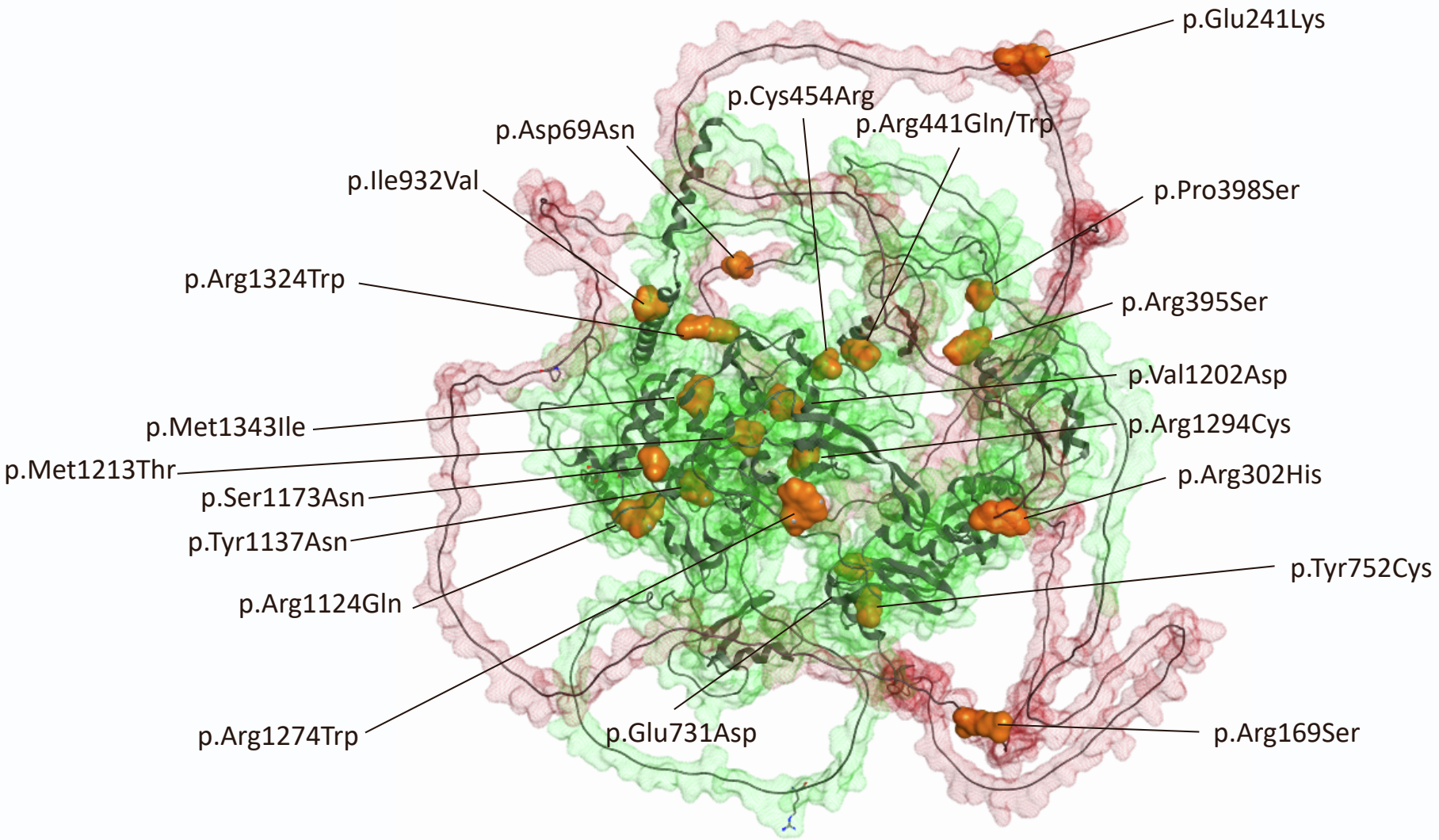


Figure S2. Missense variants in ZMYM3 mainly lie in ordered regions. Visual representation of the variant residues on a 3D model of ZMYM3. Disordered regions are in red, while structured regions are shown in green. Individual variants are shown in orange.

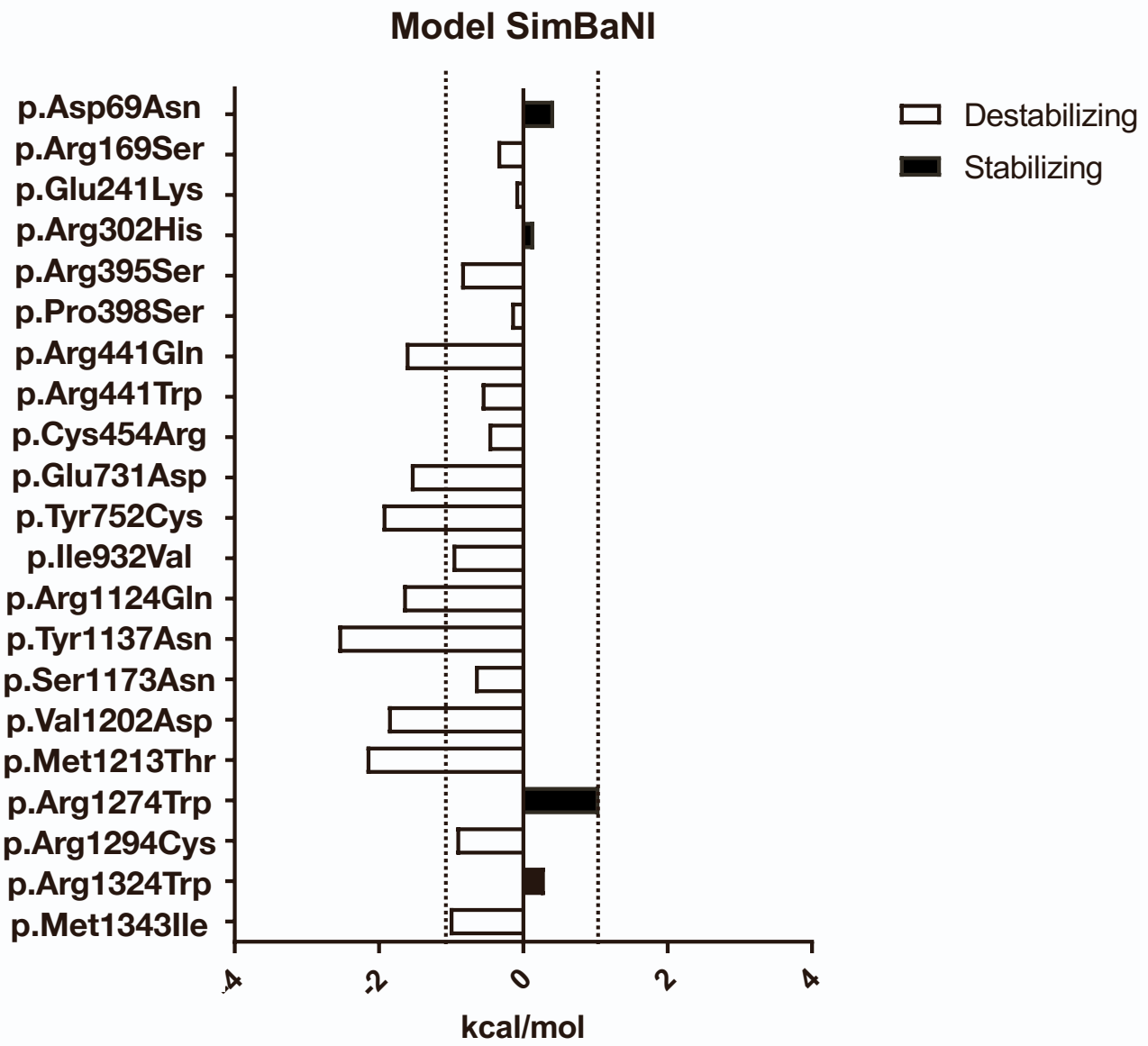


Figure S3. Analysis of the free energy variation upon amino acid substitution suggests a destabilization for the majority of mutants with a major impact of mutants p.Arg441Gln, p.Glu731Asp, p.Tyr752Cys, p.Arg1124Gln, p.Tyr1137Asn, p.Val1202Asp, and p.Met1213Thr . Conversely, p.Arg1274Trp is predicted to stabilize the protein structure. Vertical dotted line indicates threshold of +/- 1 kcal/mol.

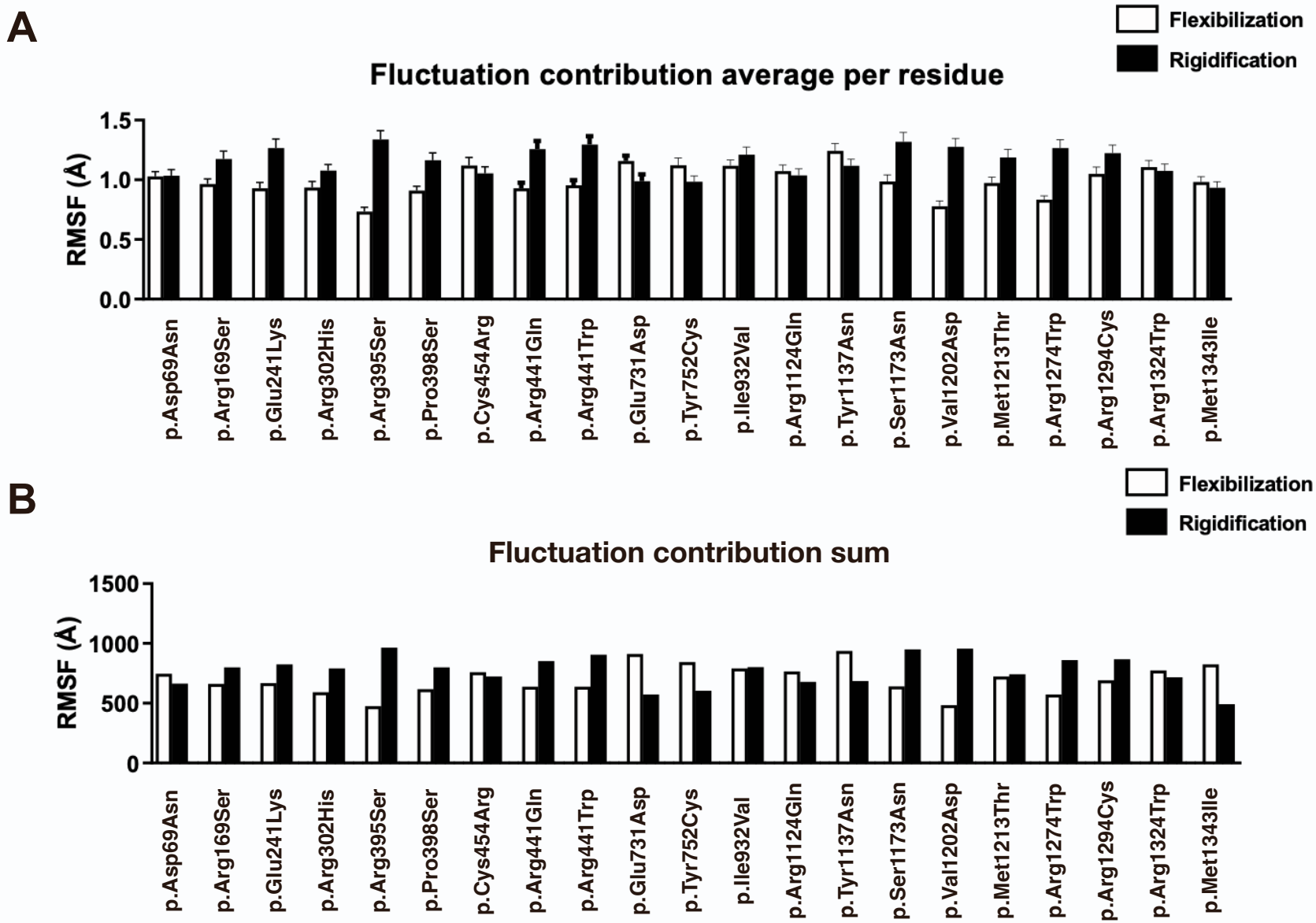


Figure S4: Coarse-Grained molecular dynamics run. Assessment of the flexibility contribution across the entire length of the protein, calculated by comparing the difference of RMSF of the WT minus RMSF of the mutant. Contribution to flexibility is shown in white bars, contribution to rigidity in black bars **A.** Average RMSF per residue for both contributing factors. **B.** Sum of RMSF values for both contributing factors.

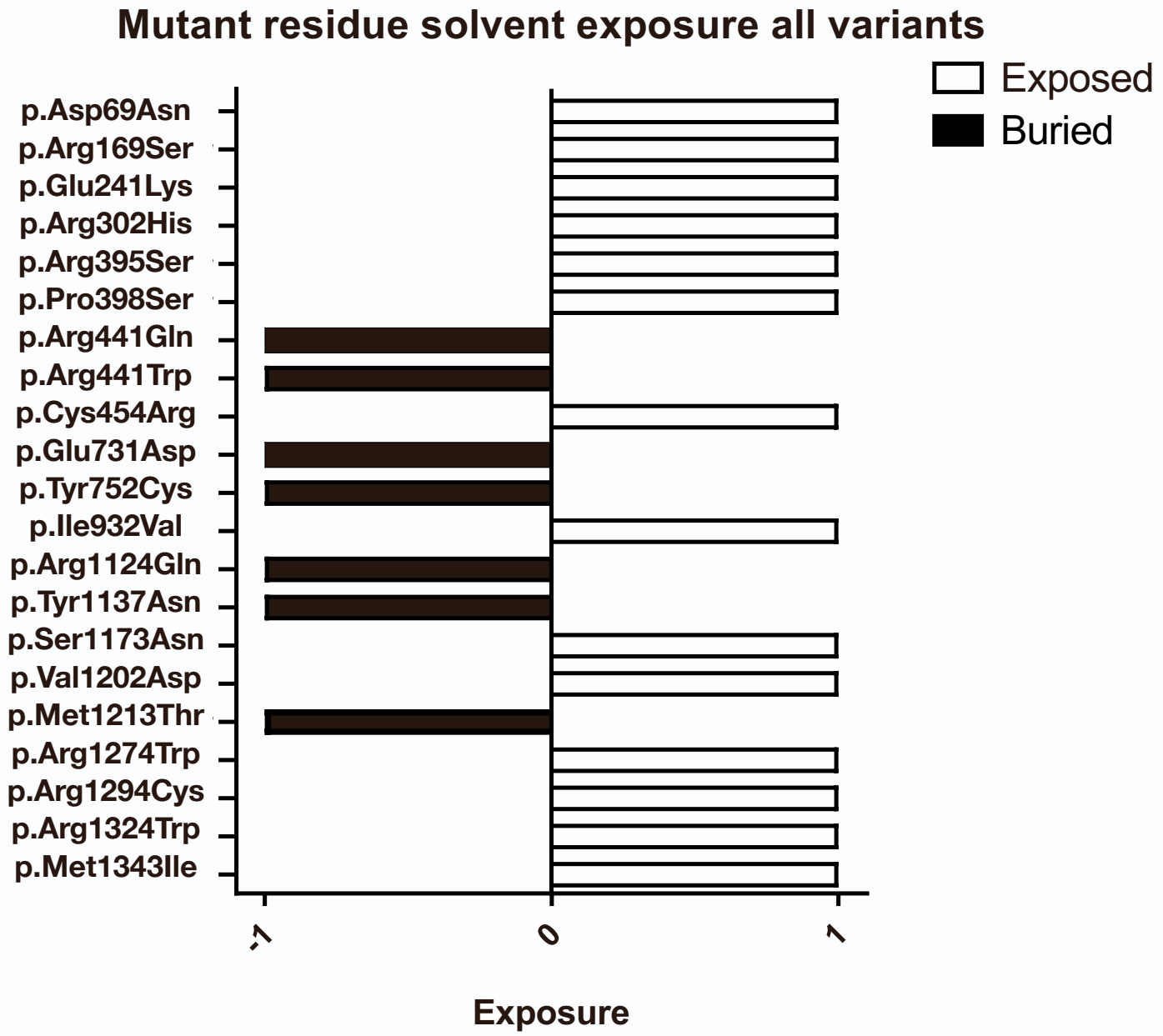


Figure S5: Difference in solvent accessibility of mutant residues. Exposure is expressed as -1 (black bars) for buried and +1 (white bars) for exposed residues. Seven residues are buried, while 14 are exposed.

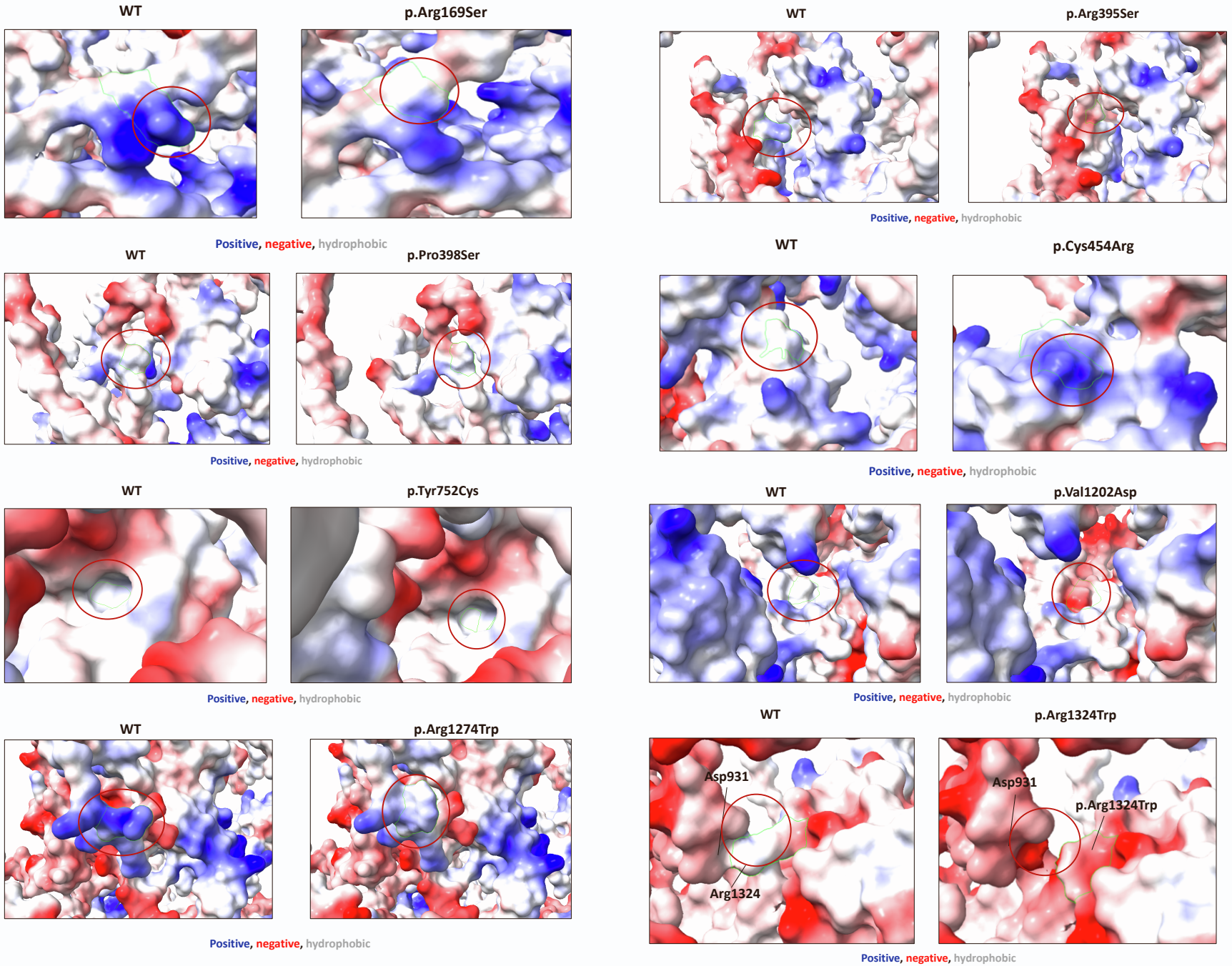


Figure S6. Surface analysis of several selected substitutions.

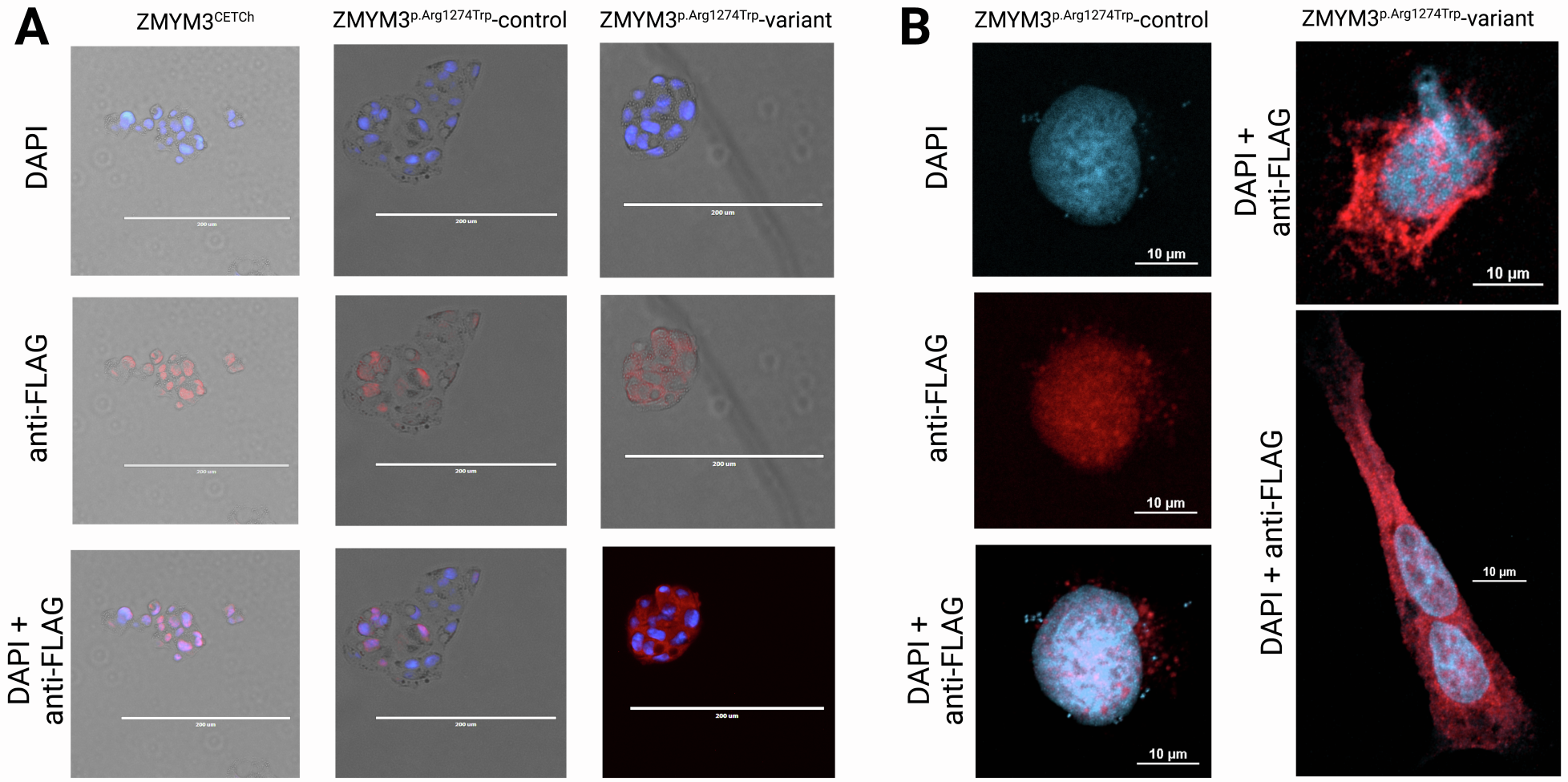


Figure S7. Immunocytochemistry shows localization of ZMYM3CETCh and ZMYM3$^{p.Arg1274Trp}$-control proteins primarily in nuclei of cells, and ZMYM3$^{p.Arg1274Trp}$-variant protein primarily in cytoplasm of cells. A. Images visualized on Evos FL microscope at 20X. Columns of images are labeled with cell line visualized. Separate spectral visualizations of the same field of view are shown in each column; these are (top row) DAPI, showing nuclei; (middle row) secondary antibody spectrum, showing FLAG-tagged protein localization; and (bottom row) both DAPI and anti-FLAG, showing relative localizations. **B.** Images visualized on Nikon AX confocal microscope at 100X. Columns of images are labeled with cell line visualized. First column is organized in labeled rows as in A showing the same field of view. In last column, with ZMYM3$^{p.Arg1274}$Trp-variant only cells, both DAPI and secondary (anti-FLAG) spectra are shown for two separate fields of view; top shows one cell, bottom shows two adjacent cells. Created with www.BioRender.com. See methods for additional details.

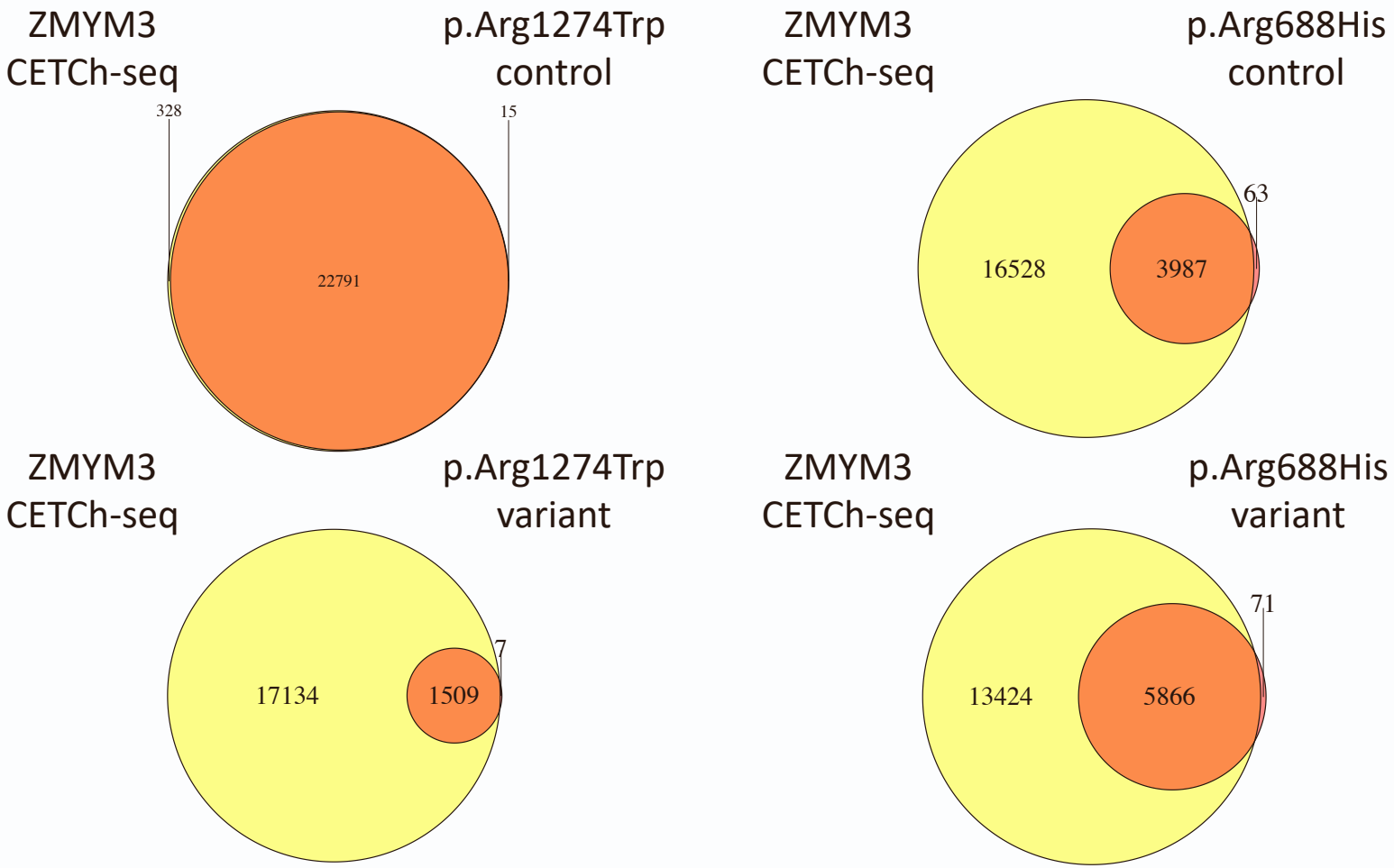


Figure S8. Intersection of csaw analysis and peak calls. For each pairwise comparison, the union of peaks was determined (downsampled to 20 million reads per replicate, IDR 0.05, merged peaks) and intersected with all csaw analyzed regions. Regions with significantly higher reads in the ZMYM3 CETCh-seq experiment are colored in yellow; regions with significantly higher reads in the comparison experiment are colored in red; and regions that are not significantly different are colored in orange (the overlap set; FDR cutoff for significance is 0.05). p.Arg1274Trp control is highly similar to ZMYM3 CETCh-seq, with the vast majority of regions not significantly differential. The other compared experiments are subsets of the ZMYM3 CETCh-seq experiment with very few differentially higher regions.

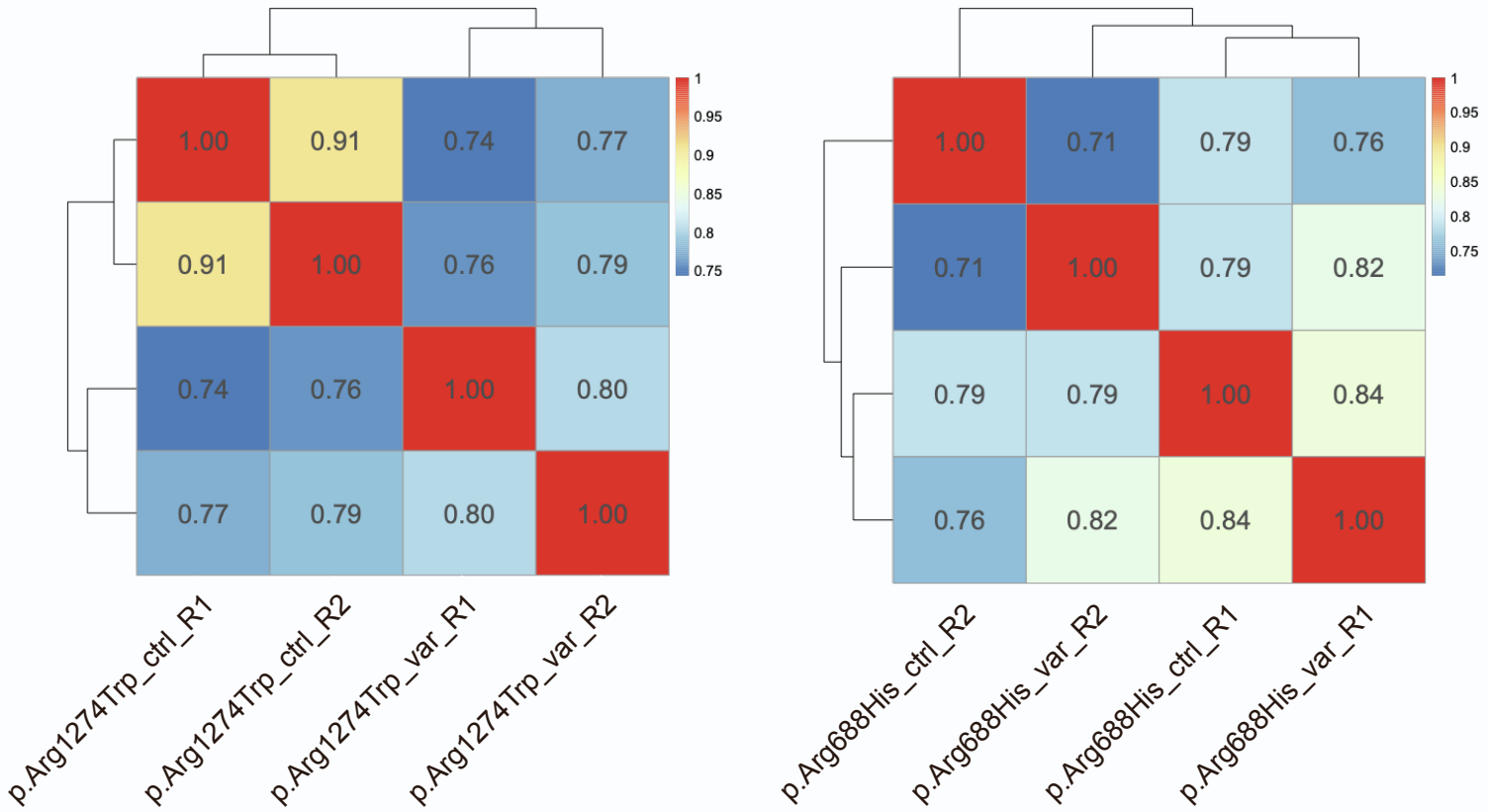


Figure S9. Read count correlations. The union of peaks between all experiments (ZMYM3^{CETCh}, ZMYM3$^{p.Arg1274Trp}$, and ZMYM3$^{p.Arg688His}$) was calculated, and reads from the .bam file for each replicate (downsampled to 20 million reads) were determined at each of these genomic positions. These data were correlated for each pairwise comparison and Pearson correlation coefficient was determined and plotted with clustering by pheatmap v. 1.0.12 in R v. 4.2.1.

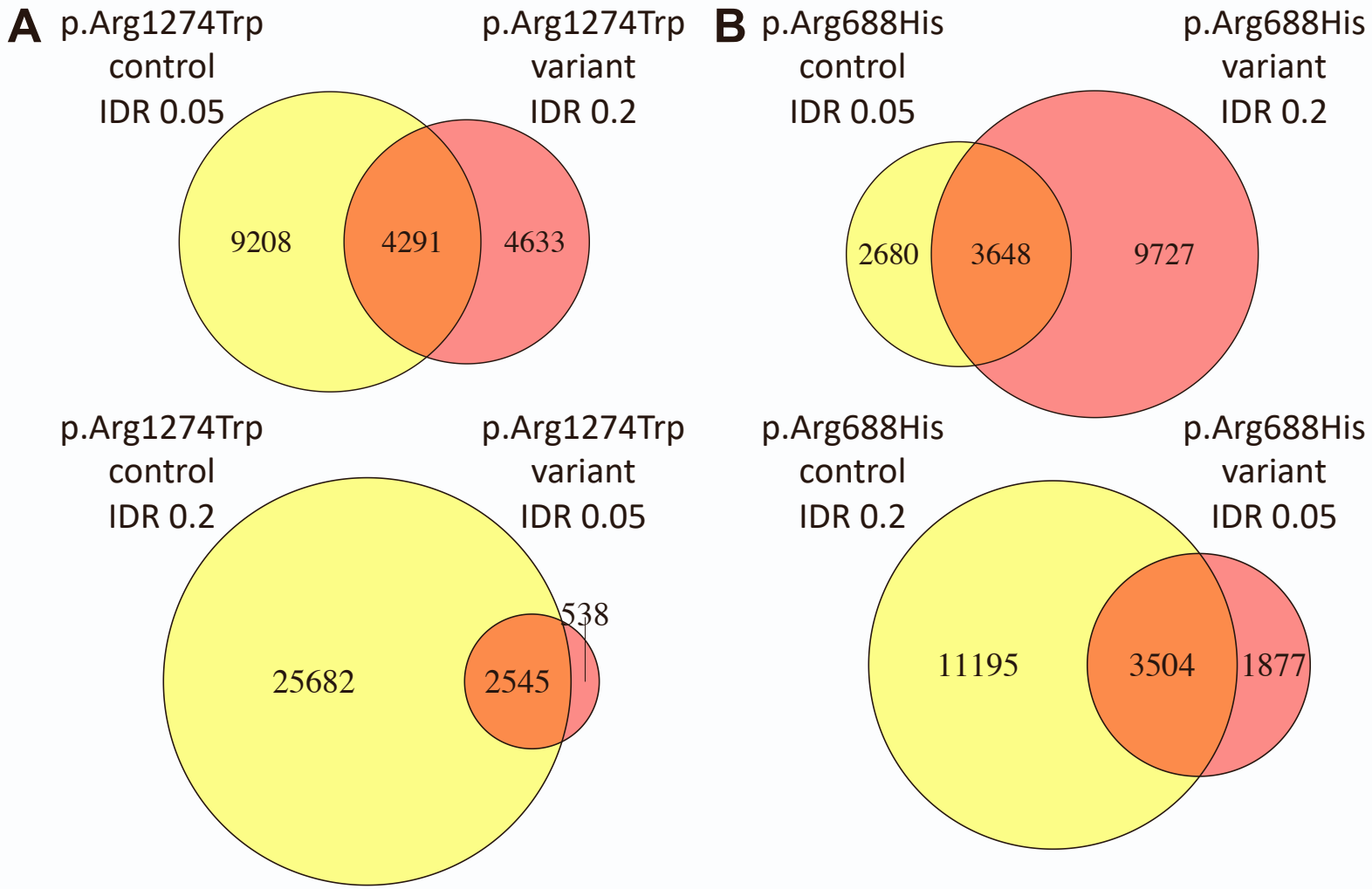


Figure S10. Standard peak overlaps with relaxed peaks. Peaks were called between replicate experiments downsampled to 20 million reads at the standard IDR threshold of 0.05 and also at a relaxed IDR threshold of 0.2. This analysis attempts to find peak overlaps missed by comparing standard peak calls due to enriched regions near peak-calling threshold. For p.Arg1274Trp (A), many regions in the variant experiment called as peaks in control are just under threshold and found by this approach, indicating a general reduced binding by the variant. For p.Arg688His (B), an experiment with greater noise, peaks are gained in both comparisons.

Supplemental Materials and Methods

SEQUENCING

Several probands had either exome sequencing (ES) or the Autism/ID Xpanded Panel (AIDX) through GeneDx, and details are listed in the table below. Using genomic DNA from the proband and parents (if submitted), from the tissues listed below, the exonic regions and flanking splice junctions of the genome were captured using the IDT xGen Exome Research Panel v1.0 (Integrated DNA Technologies, Coralville, IA). Massively parallel (NextGen) sequencing was done on an Illumina system with 100 bp or greater paired-end reads. Reads were aligned to human genome build GRCh37/UCSC hg19 and analyzed for sequence variants using a custom-developed analysis tool. Reported variants were confirmed, if necessary, by an appropriate orthogonal method in the proband and, if submitted, in selected relatives. Additional sequencing technology and variant interpretation protocol has been previously described[1]. The general assertion criteria for variant classification are publicly available on the GeneDx ClinVar submission page (http://www.ncbi.nlm.nih.gov/clinvar/submitters/26957/).

Individual	Protein Effect	Testing details	Tissue Tested
1	p.Asp69Asn	AIDX trio	Buccal
2	p.Glu241Lys	ES duo with mom	Blood
5	p.Pro398Ser	AIDX trio	Blood
6	p.Arg441Gln	ES trio	Buccal
15	p.Arg1274Trp	AIDX duo with mom	Blood
18	p.Leu226TrpfsTer8	ES trio	Buccal
19	p.Tyr752Cys	AIDX trio, Sanger confirmation for proband but parents had high quality NGS data and Sanger confirmation was not necessary	Buccal
23	p.Arg454Cys	ES trio	Blood

Individual 3, p.Arg302His and Individual 7, p.Arg441Gln

Exome sequencing was performed on Individuals 3 and 7 as described[2]. Genomic DNA was isolated from blood. The patients were enrolled for trio exome sequencing (ES) as part of the international network of Autism Sequencing Consortium (ASC) (https://asc.broadinstitute.org/). ES was performed at the Broad Institute on Illumina HiSeq sequencers[3]. ES raw data of the trios were processed and analyzed using an in-house implemented pipeline[4,5].

Individuals 4a, 4b, p.Arg395Ser
Genomic DNA extracted from leukocytes of both patients and their parents was used for whole-exome sequencing. Exome enrichment was performed on individually barcoded samples using SeqCap EZ Exome Probes v3.0 (Roche) and sequencing was performed on HiSeq 2500 (Illumina) with 100bp paired-end reads. Reads were aligned to the hg19 reference genome using Novoalign version 3.02.13 (Novocraft) with default parameters. After genome alignment, conversion of SAM format to BAM and duplicate removal was performed using Picard Tools (2.20.8). The Genome Analysis Toolkit, GATK (3.8)[6] was used for local realignment around indels, base recalibration, variant recalibration, and variant calling. Variants were annotated using the GEMINI framework[7] and filtered based on the population frequencies using several public databases and an in-house database of population-specific variants. Identification of candidate variants was performed for autosomal dominant (*de novo* variants) and autosomal recessive inheritance patterns. Variants were further prioritized according to the functional impact and conservation score. Sanger sequencing confirmed presence of the variants in the patients.

Individual 8, p.Arg441Gln
Trio exome sequencing was performed for Proband 8. DNA was enriched using Agilent SureSelect Clinical Research Exome V2 capture and paired-end sequencing using the Illumina platform (outsourced). The aim was to obtain 8.1 Giga base pairs per exome with a mapped fraction of 0.99. The average coverage of the exome was ~50x. Duplicate reads were excluded. Data were demultiplexed with bcl2fastq Conversion Software from Illumina. Reads were mapped to the genome using the BWA-MEM, and variants were called using GATK HaplotypeCaller. Detected variants were filtered and annotated with Cartagenia software and classified with Alamut Visual.

Individuals 9a, 9b, p.Glu731Asp
ES was performed on gDNA of both Individuals 9a and 9b. The exome was captured using the xGen Exome Research Panel v2 (Integrated DNA Technologies) and sequenced using the Illumina HiSeq4000 platform according to the manufacturer's protocols. The overall mean-depth base coverage was 136- and 125-fold, while on average 93% and 92% of the targeted region was covered at least 20-fold, respectively for V.2 and V.3. Read mapping and variant calling were performed as described[8] using the Varapp software[9]. Sanger sequencing confirmed the segregation of the potentially causative variant.

Individual 10, p.Ile932Val
Proband 10 had exome sequencing, and both short-read and long-read genome sequencing as described[10]. The variant was first observed in ES.

Individual 11, p.Arg1124Gln
For Individual 11, patient blood was drawn into an EDTA blood collection tube. Isolation of DNA from whole blood was performed using the QIAsymphony (Qiagen). Sequencing libraries were constructed from patient whole blood genomic DNA using the HudsonAlpha Clinical Services Lab's custom whole genome library preparation protocol.

Patient DNA was sequenced on the Illumina HiSeqX sequencer. DNA library fragments were sequenced from both ends (paired) with a read length of 150 base pairs. Patient genomes were sequenced at an approximate depth of 30X, with at least 80% of base positions reaching 20X coverage. Sequence variants were called using GATK3 and loaded into a custom software analysis application for interpretation. All sequence variants were annotated with relevant information from established data sources to provide support for variant interpretation. Variant pathogenicity was determined using ACMG criteria[11].

Individual 12, p.Tyr1137Asn

For Individual 12, Genomic DNA was extracted from EDTA blood of the patient and his parents. Whole Exome Sequencing (ES) on the patient was performed using the xGen® Exome Research Panel v1.0 (IDT) with paired-end sequencing (HiSeq SBS Kit v4, 125 Fwd-125 Rev, Q30-value: 84) on a HiSeq System (Illumina Inc.). Raw fastQ files were aligned to the hg19 reference genome using NextGene (Softgenetics). The average depth of coverage was 225x and 99.4% of the targeted bases were assessed by ≥20 independent sequence reads. By applying filters for known and candidate ID genes (SYSID and In-House) and Minor Allele Frequency ≤ 2% (gnomAD, ExAC) a total of 37 variants were observed in at least 16% of reads with sufficient quality level. Variants were investigated computationally for deleterious effects, by associations of the affected gene with proband's phenotype and by literature search for functional information. The candidate *ZMYM3* mutation from the ES approach was re-sequenced in the index, his mother and maternal grandparents after PCR amplification by Sanger sequencing using an ABI Genetic Analyzer 3730 (Applied Biosystems, Foster City, California.

Individual 13, p.Val1202Asp

The proband had ES as described[12,13]. Genomic DNA was isolated from circulating leukocytes. The *de novo* status of the *ZMYM3* variant was confirmed by Sanger sequencing.

Individual 14, p.Met1213Thr

A clinical comprehensive intellectual disability panel (555 genes) was ordered through Fulgent Genetics, and testing was done on genomic DNA from blood. This testing resulted in identification of the *ZMYM3* variant described here. Follow-up reanalysis through the Care4Rare project[14,15] also identified this *ZMYM3* variant as the top hit.

Individual 16, p.Arg1294Cys

Individual 16 had trio exome sequencing. DNA was extracted from fetal muscle using the Prepito automate machine. Exome DNA library was prepared with the Agilent Focused Exome preparation kit. High-throughput sequencing was performed on a NextSeq550 sequencer (Illumina) with a 2x75 bp paired-end running method. The BWA-MEM algorithm was used to map the reads on the reference genome (GRCh37/hg19). The variant calling was performed according to GATK and FreeBayes best practices. The ANNOVAR and ALAMUT (Interactive Biosoftware) tools were used for variant annotation.

Individual 17, p.Met1343Ile
For Individual 17, a CHOP Medical Exome was performed on the proband and mother. Genomic DNA was extracted from peripheral blood or other patient tissues following standard DNA extraction protocols. After extraction of genomic DNA, targeted exons are captured with the Agilent SureSelect XT Clinical Research Exome kit (per manufacturer's protocol) and sequenced on the Illumina HiSeq 2000 or 2500 platform with 100bp paired-end reads. Mapping and analysis were based on the human genome build UCSC hg19 reference sequence. Sequencing data is processed using an in-house custom-built bioinformatics pipeline. The bioinformatics protocol utilized for this evaluation is version CWES-2.2. The exome sequencing protocol utilized for this evaluation is version 3.1. Coding exons and splice sites targeted with the exome kit are analyzed and reported. The following pathogenic variants are detectable: single nucleotide variants, small deletions and small insertions.

Individual 20, p.Arg1294Cys
Individual 20 had exome sequencing on DNA extracted from blood. ES was performed using Nimblegen SeqCap Ez MedExome Target Enrichment Kit (Roche Sequencing Solutions, Pleasanton, CA, USA) and an Illumina NextSeq500 (Illumina Inc., CA, USA) as paired-end 150 bp reads. Sequences were analyzed with the SeqOne platform (Montpellier, France). Sanger sequencing was performed for variant confirmation and segregation. X-chromosome inactivation study was performed by methylation analyses using the HUMARA assay[16].

Individual 21, p.Arg169Ser
An Agilent Sure Select Target Enrichment System for the Clinical Research Exome was used to capture the regions of interest using genomic DNA isolated from blood. This method allows for analysis of greater than 98% of the targeted sequence. Analysis of data was performed using NextGENe software (SoftGenetics, State College, PA) along with an in-house bioinformatics pipeline. The data was reviewed with emphasis on novel alterations and those reported in the Human Gene Mutation Database (HGMD). All alterations of potential clinical relevance were confirmed by Sanger sequencing. Routinely observed drop-out and low coverage regions of the NGS data were also Sanger sequenced.

Individual 22, p.Arg441Trp
Exome sequencing for Individual 21 was done on DNA isolated from saliva swab. The Agilent SureSelect^{XT} Clinical Research Exome kit was used to target known disease-associated exonic regions of the genome (coding sequences and splice junctions of known protein-coding genes associated with disease, as well as an exomic backbone) using genomic DNA isolated from the patient. The targeted regions were sequenced using the Illumina NovaSeq^{TM} 6000 System with 150 bp paired-end reads. Using Illumina DRAGEN Bio-IT Platform® software, the DNA sequence was aligned and compared to the human genome build 19 (hg19/NCBI build 37). The average depth of coverage was calculated to be approximately 105X across all targeted regions. The emedgene® software was used to filter and analyze sequence variants identified in the patient.

Individual 24, p.Ser1173Asn
Patient blood was drawn into an EDTA blood collection tube for preparation by the HudsonAlpha Clinical Services Lab. Isolation of DNA from whole blood was performed using the QIAsymphony (Qiagen). Sequencing library was constructed from patient genomic DNA using the Illumina TruSeq PCR-free library preparation protocol and sequenced on the Illumina NovaSeq 6000 sequencer. DNA library fragments were sequenced from both ends (paired) with a read length of 150 base pairs. The patient genome was sequenced to a mean depth of 30X. Raw sequence data was demultiplexed and aligned to reference genome GRCh38 using Sentieon. Sequence variants were called using Sentieon DNAscope and loaded into a custom software analysis application for interpretation. All sequence variants were annotated with relevant information from established data sources to provide support for variant interpretation. Variant pathogenicity was determined using modified ACMG criteria.

Individual 25, p.Arg1324Trp
The Illumina TruSeq Nano DNA Library Prep Kit was used to prepare library for genome sequencing using the genomic DNA isolated from blood. Sequencing was performed on the Illumina NovaSeq™ 6000 System with 150 bp paired-end reads. Using Illumina DRAGEN Bio-IT Platform® software, the DNA sequence was aligned and compared to the human genome build 19 (hg19/NCBI build 37). The average depth of coverage across all genomic regions was calculated to be approximately 40X. The emedgene® software was used to filter and analyze sequence variants identified within the patient's genome sequencing data and to compare variants identified in the patient to the sequences of family members.

p.Arg1294Cys Cases
We note that p.Arg1294Cys has been observed in two individuals here (probands 16 and 20), each confirmed to be *de novo*. Each of these cases were submitted by sites in Europe, and we note that a ClinVar submission (SCV000297052.2) was made by the Children's Hospital of Philadelphia, USA. While this does appear to be a unique case from those described here, we cannot confirm this.

ACMG EVIDENCE CODES
As ACMG evidence codes are only applied to variants in established disease genes, we did not explicitly apply them here. If evidence codes were applied, most variants presented here would likely remain VUSs. All variants presented here are rare (PM2), and most are predicted damaging (PP3). The combination of these two codes would result in an overall status of VUS. For a few variants, additional evidence codes may apply. Functional studies in support of pathogenicity for p.Arg1274Trp (PS3) or in support of a benign status for p.Arg688His (BS3) may also apply. Additionally, while six variants are de novo (PS2), this code can only be applied if "The phenotype in the patient matches the genes' disease association with reasonable specificity."

COMPUTATIONAL MODELING

The wild-type 3D protein structure was downloaded from AlphaFoldDB (https://alphafold.ebi.ac.uk/)[17], which was included with the reference from UniProt (Accession number: Q14202). When not possible online, structures were visualized, colored and the sequence was mutated with Chimera version 1.15, rotamer builder tool[18]. Specifically, 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[19]. Depiction of molecular surfaces was defined as VdW surface and colored according to the electrostatic potential.

Wild-Type Protein model analysis: The distinction between organized and disordered regions was based upon the uniprot reference. The pLDDT value for each residue was extracted from the pdb file (opened as text file) from the B-factor field as prescribed on https://alphafold.ebi.ac.uk/about. The PAE matrix was downloaded from the main page of AlphaFoldDB and analyzed with the web-interface built-in tool.

Free folding energy estimation: the ΔΔG was calculated according to the SimBaNI model developed by Caldarau and colleagues[20]. For the SASA term, the RSA obtained with FreeSASA and naccess parameters were considered. Variations >1.0 kcal/mol in module were considered as significant.

Flexibility inspection: The flexibility analysis was performed by submission of pdb structures to CABS-flex2 with no restraints[21]. The results, expressed as RMSF were downloaded and a cutoff of 1 Å was considered as relevant.

Normal mode analysis: The normal mode analysis was performed by submission of the WT pdb file to WebNMA[22] and run in a comparative mode by separately setting the mutation parameters. The deformation energy per-residue was then downloaded and plotted. We applied a cutoff of 2 kcal/mol.

Surfaces calculation: the molecular and solvent accessible surfaces were calculated with two methods. First, both molecular (VdW radii) and solvent-accessible (SASA, probe radius 1.4 Å) total, polar and (by difference) non-polar surfaces were computed to calculate the variations. To this aim, VegaZZ suite was employed[23]. Then, relative-solvent-accessible surface area (RSA) was calculated by submitting the pdb files to FreeSASA with the parameters derived from naccess[24]. Burial-exposure mutant residue classification: the classification into buried/exposed mutant residues was based on the computed RSA values. By comparing the RSAs with the tabulated values[25] and selecting a cutoff of 0.20.

Eukaryotic Linear Motif identification: the UniProt accession (Q14202) was submitted to the online server ELM (http://elm.eu.org/) with standard settings (100 as probability cutoff, species Homo sapiens). The results identified about 107 ELMs. The results were exported as .csv file and the positions were intersected with the mutation sites.

CHIP-SEQ EXPERIMENTS

We edited the genomic DNA at the *ZMYM3* endogenous locus in HepG2 cells to introduce the variant of interest 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[26]. HepG2 cells were sourced from ATCC (HB-8065) and cultured using the recommended protocol. We identified a CRISPR/Cas9 sgRNA targeting a DNA cleavage site near each variant

(Table S5) using established methods and cloned this sgRNA into pX330-U6-Chimeric_BB-CBh-hSpCas9 (Addgene plasmid #42230)[27]. We designed homology-directed repair donor templates composed of 400 bp of genomic sequence upstream (relative to coding direction of *ZMYM3*) of the cleavage site, exonic sequence surrounding the variant of interest, all exons downstream of the exon harboring the variant without the stop codon, the in-frame FLAG/P2A/NeoR cassette, and 400 bp of genomic sequence downstream of the cleavage site (Table S5). For control experiments, we repaired the genomic DNA with reference sequence at the variant position; for variant experiments, we used the variant nucleotide. For both experiments, an additional mutation was inserted to abolish the PAM site and block cleavage of edited sequence by Cas9. This mutation was a synonymous substitution for PAM sites in coding sequence. The donor templates were synthesized and cloned into the BamHI site of pUC19 by GenScript (Piscataway, NJ, USA). Genomic coordinates for the variant positions, the PAM site mutation positions, and the exons included in the super-exon for each experiment are shown in the table below.

Variant	**hg38 substitution**	**super-exon**	**PAM mutation**
p.Arg441Trp	chrX:71,249,610 G>A	chrX:71,249,461-71,249,679 chrX:71,249,020-71,249,169 chrX:71,248,686-71,248,799 chrX:71,248,438-71,248,524 chrX:71,248,163-71,248,312 chrX:71,247,734-71,247,904 chrX:71,247,346-71,247,510 chrX:71,246,595-71,246,690 chrX:71,246,354-71,246,512 chrX:71,245,986-71,246,098 chrX:71,245,669-71,245,842 chrX:71,245,340-71,245,483 chrX:71,244,790-71,244,891 chrX:71,244,303-71,244,470 chrX:71,243,829-71,243,978 chrX:71,242,971-71,243,084 chrX:71,242,171-71,242,422 chrX:71,241,229-71,241,342 chrX:71,240,919-71,241,107	chrX:71,249,611 G>A
p.Arg688His	chrX:71,247,819 C>T	chrX:71,247,734-71,247,904 chrX:71,247,346-71,247,510 chrX:71,246,595-71,246,690 chrX:71,246,354-71,246,512 chrX:71,245,986-71,246,098 chrX:71,245,669-71,245,842 chrX:71,245,340-71,245,483 chrX:71,244,790-71,244,891 chrX:71,244,303-71,244,470	chrX:71,247,824 G>A

		chrX:71,243,829-71,243,978 chrX:71,242,971-71,243,084 chrX:71,242,171-71,242,422 chrX:71,241,229-71,241,342 chrX:71,240,919-71,241,107	
p.Arg1274Trp	chrX:71,241,327 G>A	chrX:71,241,229-71,241,342 chrX:71,240,919-71,241,107	chrX:71,241,351 G>A

For each of the control and variant experiments, we nucleofected two million HepG2 cells with 5 ug total plasmid DNA (2.5 ug sgRNA/Cas9 plasmid and 2.5 ug respective donor plasmid) using a Lonza Nucleofector Kit V with an Amaxa Nucleofector 2. Immediately after nucleofection, each experiment was split into two wells of a 6-well plate, and these two replicates were recovered and grown separately. Two days post-nucleofection, we began selection with Geneticin (Invitrogen 10131) at 300 ug/mL. Cells were grown under selection for two weeks, and afterwards continued to expand in non-selective media for another three to four weeks. The p.Arg441Trp experiment did not survive selection, and no further work was performed on this variant. Genomic DNA was purified from cells and used as template for PCR and Sanger sequence validation of edits using a Qiagen DNeasy Blood & Tissue Kit (Table S5). Cells (20 million for each replicate) were crosslinked and harvested, immunoprecipitation with M2 FLAG monocolonal antibody (Sigma F1804) was performed, and sequencing libraries were constructed, all as previously described[28]. The libraries were pooled with other CETCh-seq libraries and sequenced on a NovaSeq S2 flowcell yielding total aligned read counts as shown in Table S4. We performed peak calling using SPP[29] and Irreproducible Discovery Rate (IDR)[30] using ENCODE-standardized pipelines for analysis and quality-control[31]. We also downsampled all replicate bam files to 20 million reads and performed peak calling, using either the standard IDR cutoff of 0.05 or a relaxed IDR cutoff of 0.2. We performed additional differential binding analyses using the R package csaw v1.28.0[32], using window widths of 10 nucleotides and background bins of 10,000 nucleotides, and using downsampled (to 20 million reads) bam files for all replicates. We generated Activity-by-Contact (ABC) v0.2 loop calls[33] using hg38 reference sequence and gene coordinates downloaded from the UCSC genome table browser, the ENCODE datasets ENCSR149XIL (DNase-seq), ENCSR000AMO (H3K27Ac), and ENCFF356LFX (blacklist), RNA-seq for HepG2 downloaded from Expression Atlas, and processed HepG2 Hi-C data from 4D Nucleome (4DNESC2DEQIJ).

The overlaps of peaks called from downsampled bam files (20 million reads) were 67.8% between ZMYM3$^{p.Arg1274Trp}$-control and ZMYM3$^{p.Arg1274Trp}$-variant, and 46.9% between ZMYM3$^{p.Arg688His}$-control and ZMYM3$^{p.Arg688His}$-variant. Knowing that peak overlaps suffer from missed calls at regions near threshold in experiments, we expanded the peak overlap analysis to use peaks called at the standard IDR cutoff of 0.05 in one experiment and peaks called at a relaxed IDR cutoff of 0.2 in the other experiment. This increased overlaps to 82.5% and 65.1%, respectively (Figure S10). To examine experiment similarity agnostic to peak calls, we performed read count correlations using each separate replicate for each experiment. Rather than performing read count correlations across the entire genome (an analysis that suffers from the

majority of regions being near background level of read counts), we filtered the analysis space to ZMYM3-specific regions. We determined the union of all peaks called in ZMYM3^{CETCh}, ZMYM3$^{p.Arg688His}$, and ZMYM3$^{p.Arg1274Trp}$, and used this set of regions for read counts. Reads from each bam file were determined at each genomic location and the Pearson correlation coefficient was calculated for each pairwise comparison (Figure S9). For p.Arg1274Trp, replicate experiments were highly correlated (r=0.80-0.91), and the correlation between control and variant experiments was high (r=0.74-0.79). For p.Arg688His, replicate experiments correlated similarly to correlations between control and variant (r=0.71-0.84). These results are consistent with the p.Arg688His experiments being of somewhat lower quality than p.Arg1274Trp experiments, and with no distinguishable difference between p.Arg688His control and variant beyond the noise of the assay.

IMMUNOCYTOCHEMISTRY EXPERIMENTS

We grew ZMYM3^{CETCh}, ZMYM3$^{p.Arg1274Trp}$-control, and ZMYM3$^{p.Arg1274Trp}$-variant cells in 6-well plates, seeding 50,000 cells per well, and in Millicell EZ Slides (Millipore), seeding 2,000 cells per chamber. After 24 hours, we performed fixation and permeabilization of cells on plates and slides using the Image-iT kit (Invitrogen R37602) following the manufacturer's protocol. For primary antibody, we used Sigma F1804 mouse anti-FLAG. For secondary (labeled) antibodies, we used Invitrogen T6390 goat anti-mouse IgG Texas Red-X, Abcam ab150116 goat anti-mouse IgG Alexa Fluor 594, Abcam ab6787 goat anti-mouse IgG Texas Red, or Invitrogen A21236 goat anti-mouse IgG Alexa Fluor 647. Each cell line/secondary antibody combination was performed in duplicate, and the complete experiment (from seeding to visualization) was performed twice. We imaged 6-well plates on an Evos FL microscope (ThermoFisher), and Millicell EZ Slides on a Nikon Eclipse Ti2 AX confocal microscope. In both replicates and both experiments, >95% of cells showed the localization patterns in Figure S7: ZMYM3^{CETCh} and ZMYM3$^{p.Arg1274Trp}$-control cells showed primarily nuclear localization of protein, and ZMYM3$^{p.Arg1274Trp}$-variant cells showed primarily cytoplasmic localization of protein.

SUPPLEMENTAL REFERENCES

1. Retterer, K., Juusola, J., Cho, M.T., Vitazka, P., Millan, F., Gibellini, F., Vertino-Bell, A., Smaoui, N., Neidich, J., Monaghan, K.G., et al. (2016). Clinical application of whole-exome sequencing across clinical indications. Genet Med *18*, 696–704.
2. Pavinato, L., Trajkova, S., Grosso, E., Giorgio, E., Bruselles, A., Radio, F.C., Pippucci, T., Dimartino, P., Tartaglia, M., Petlichkovski, A., et al. (2021). Expanding the clinical phenotype of the ultra-rare Skraban-Deardorff syndrome: Two novel individuals with WDR26 loss-of-function variants and a literature review. Am. J. Med. Genet. A *185*, 1712–1720.
3. Satterstrom, F.K., Kosmicki, J.A., Wang, J., Breen, M.S., De Rubeis, S., An, J.Y., Peng, M., Collins, R., Grove, J., Klei, L., et al. (2020). Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism. Cell *180*, 568-584.e23.
4. Bauer, C.K., Calligari, P., Radio, F.C., Caputo, V., Dentici, M.L., Falah, N., High, F., Pantaleoni, F., Barresi, S., Ciolfi, A., et al. (2018). Mutations in KCNK4 that Affect Gating Cause a Recognizable Neurodevelopmental Syndrome. Am. J. Hum. Genet. *103*, 621–630.
5. Flex, E., Martinelli, S., Van Dijck, A., Ciolfi, A., Cecchetti, S., Coluzzi, E., Pannone, L., Andreoli, C., Radio, F.C., Pizzi, S., et al. (2019). Aberrant Function of the C-Terminal Tail of HIST1H1E Accelerates Cellular Senescence and Causes Premature Aging. Am. J. Hum. Genet. *105*, 493–508.
6. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., et al. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res *20*, 1297–1303.
7. Paila, U., Chapman, B.A., Kirchner, R., and Quinlan, A.R. (2013). GEMINI: integrative exploration of genetic variation and genome annotations. PLoS Comput Biol *9*, e1003153.
8. Alfaiz, A.A., Micale, L., Mandriani, B., Augello, B., Pellico, M.T., Chrast, J., Xenarios, I., Zelante, L., Merla, G., and Reymond, A. (2014). TBC1D7 Mutations are Associated with Intellectual Disability, Macrocrania, Patellar Dislocation, and Celiac Disease. Hum. Mutat. *35*, 447–451.
9. Delafontaine, J., Masselot, A., Liechti, R., Kuznetsov, D., Xenarios, I., and Pradervand, S. (2016). Varapp: A reactive web-application for variants filtering. BioRxiv 060806.
10. Cohen, A.S.A., Farrow, E.G., Abdelmoity, A.T., Alaimo, J.T., Amudhavalli, S.M., Anderson, J.T., Bansal, L., Bartik, L., Baybayan, P., Belden, B., et al. (2022). Genomic answers for children: Dynamic analyses of >1000 pediatric rare disease genomes. Genet. Med. *24*, 1336–1348.
11. Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E., Spector, E., et al. (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med *17*, 405–424.
12. Motta, M., Fasano, G., Gredy, S., Brinkmann, J., Bonnard, A.A., Simsek-Kiper, P.O.,

Gulec, E.Y., Essaddam, L., Utine, G.E., Guarnetti Prandi, I., et al. (2021). SPRED2 loss-of-function causes a recessive Noonan syndrome-like phenotype. Am. J. Hum. Genet. *108*, 2112–2129.
13. Lin, Y.C., Niceta, M., Muto, V., Vona, B., Pagnamenta, A.T., Maroofian, R., Beetz, C., van Duyvenvoorde, H., Dentici, M.L., Lauffer, P., et al. (2021). SCUBE3 loss-of-function causes a recognizable recessive developmental disorder due to defective bone morphogenetic protein signaling. Am. J. Hum. Genet. *108*, 115–133.
14. Beaulieu, C.L., Majewski, J., Schwartzentruber, J., Samuels, M.E., Fernandez, B.A., Bernier, F.P., Brudno, M., Knoppers, B., Marcadier, J., Dyment, D., et al. (2014). FORGE Canada consortium: Outcomes of a 2-year national rare-disease gene-discovery project. Am. J. Hum. Genet. *94*, 809–817.
15. Hamilton, A., Tétreault, M., Dyment, D.A., Zou, R., Kernohan, K., Geraghty, M.T., FORGE Canada Consortium, Care4Rare Canada Consortium, Hartley, T., and Boycott, K.M. (2016). Concordance between whole-exome sequencing and clinical Sanger sequencing: implications for patient care. Mol. Genet. Genomic Med. *4*, 504–512.
16. Bertelsen, B., Tümer, Z., and Ravn, K. (2011). Three new loci for determining x chromosome inactivation patterns. J. Mol. Diagn. *13*, 537–540.
17. 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.
18. 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.
19. 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.
20. Caldararu, O., Blundell, T.L., and Kepp, K.P. (2021). Three Simple Properties Explain Protein Stability Change upon Mutation. J. Chem. Inf. Model. *61*, 1981–1988.
21. Kuriata, A., Gierut, A.M., Oleniecki, T., Ciemny, M.P., Kolinski, A., Kurcinski, M., and Kmiecik, S. (2018). CABS-flex 2.0: a web server for fast simulations of flexibility of protein structures. Nucleic Acids Res. *46*, W338–W343.
22. Tiwari, S.P., Fuglebakk, E., Hollup, S.M., Skjærven, L., Cragnolini, T., Grindhaug, S.H., Tekle, K.M., and Reuter, N. (2014). WEBnm@ v2.0: Web server and services for comparing protein flexibility. BMC Bioinformatics *15*,.
23. Pedretti, A., Mazzolari, A., Gervasoni, S., Fumagalli, L., and Vistoli, G. (2021). The VEGA suite of programs: an versatile platform for cheminformatics and drug design projects. Bioinformatics *37*, 1174–1175.
24. Mitternacht, S. (2016). FreeSASA: An open source C library for solvent accessible surface area calculations. F1000Research *5*,.
25. Creighton, T.E. (2013). Proteins : structures and molecular properties (New York, NY, USA: Freeman).
26. 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.
27. Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., et al. (2013). Multiplex genome engineering using

CRISPR/Cas systems. Science *339*, 819–823.
28. 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.
29. Kharchenko, P. V., Tolstorukov, M.Y., and Park, P.J. (2008). Design and analysis of ChIP-seq experiments for DNA-binding proteins. Nat. Biotechnol. 2008 2612 *26*, 1351–1359.
30. Li, Q., Brown, J.B., Huang, H., and Bickel, P.J. (2011). Measuring reproducibility of high-throughput experiments. Https://Doi.Org/10.1214/11-AOAS466 *5*, 1752–1779.
31. 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.
32. 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.
33. 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.