Deleterious, protein-altering variants in the X-linked transcriptional coregulator

2 **ZMYM3** in 22 individuals with a neurodevelopmental delay phenotype

3

Susan M. Hiatt,¹* Slavica Trajkova,² Matteo Rossi Sebastiano,³ E. Christopher 4 Partridge,¹ Fatima E. Abidi,⁴ Ashlyn Anderson,¹ Muhammad Ansar,⁵ Stylianos E. 5 Antonarakis,⁶ Azadeh Azadi,⁷ Ruxandra Bachmann-Gagescu,⁸ Andrea Bartuli,⁹ Caroline 6 Benech,¹⁰ Jennifer L. Berkowitz,¹¹ Michael J. Betti,¹² Alfredo Brusco,² Ashley Cannon,¹³ Giulia Caron,¹⁴ Yanmin Chen,¹¹ Molly M. Crenshaw,¹⁵ Laurence Cuisset,¹⁶ Cynthia J. 7 8 Curry,¹⁷ Hossein Darvish,¹⁸ Serwet Demirdas,¹⁹ Maria Descartes,¹³ Jessica Douglas,²⁰ 9 David A. Dyment,²¹ Houda Zghal Elloumi,¹¹ Giuseppe Ermondi,¹⁴ Marie Faoucher,²² Emily G. Farrow,²³ Stephanie A. Felker,¹ Heather Fisher,²⁴ Anna C. E. Hurst,¹³ Pascal 10 11 Joset,²⁵ Stanislav Kmoch,^{26,27} Benjamin R. Leadem,¹¹ Marina Macchiaiolo,²⁸ Martin 12 Magner,²⁷ Giorgia Mandrile,²⁹ Francesca Mattioli,³⁰ Megan McEown,¹ Sarah K. Meadows,¹ Livija Medne,³¹ Naomi J. L. Meeks,³² Sarah Montgomery,³³ Melanie P. Napier,¹¹ Marvin Natowicz,³⁴ Kimberly M. Newberry,¹ Marcello Niceta,²⁸ Lenka 13 14 15 Noskova,²⁶ Catherine Nowak,²⁰ Amanda G. Noyes,¹¹ Matthew Osmond,²¹ Verdiana Pullano,² Chloé Quélin,³⁵ Simin Rahimi-Aliabadi ,³⁶ Anita Rauch,^{8,37} Sylvia Redon,^{10,38,39} Alexandre Reymond,³⁰ Caitlin R. Schwager,⁴⁰ Elizabeth A. Sellars,⁴¹ Angela 16 17 18 Scheuerle,⁴² Elena Shukarova-Angelovska,⁴³ Cara Skraban,³¹ Bonnie R. Sullivan,⁴⁰ Marco Tartaglia,²⁸ Isabelle Thiffault,²³ Kevin Uguen,^{10,38,39} Luis A. Umaña,⁴² Yolande van Bever,¹⁹ Saskia N. van der Crabben,⁴⁴ Marjon A. van Slegtenhorst,¹⁹ Quinten 19 20 21 Waisfisz,^{45,46} Richard M. Myers,¹ Gregory M. Cooper,^{1*} 22 23 24 ¹HudsonAlpha Institute for Biotechnology, Huntsville, AL, 35806, USA ²Department of Medical Sciences, University of Torino, 10126 Torino, Italy 25 ³Molecular Biotechnology and Health Sciences Dept., Università degli Studi di Torino, 26 27 via Quarello 15, 10135 Torino, Italy ⁴Molecular Diagnostic Laboratory, Greenwood Genetic Center, Greenwood, SC, 29646, 28 29 USA 30 ⁵University of Lausanne, Jules Gonin Eye Hospital, Fondation Asile des Aveugles, 31 Lausanne, Switzerland ⁶Department of Genetic Medicine and Development, University of Geneva, Switzerland 32 ⁷Obestetrics and Gynecology Department, Golestan University of Medical Sciences, 33 34 Gorgan, Iran ⁸Institute of Medical Genetics, University of Zurich, Schlieren 8952, Switzerland 35 ⁹Rare Diseases and Medical Genetics Unit, Bambino Gesù Children's Hospital, IRCCS, 36 37 Rome, Italy ¹⁰Univ Brest, Inserm, EFS, UMR 1078, GGB, F-29200 Brest, France 38

- ¹¹GeneDx, LLC, Gaithersburg, MD, 20877, USA
- 40 ¹²Vanderbilt University Medical Center, Nashville, TN, 37232, USA
- ¹³Department of Genetics, University of Alabama at Birmingham, Birmingham, AL, USA
- 42 ¹⁴Molecular Biotechnology and Health Sciences Dept., Università degli Studi di Torino,
- 43 via Quarello 15, 10135 Torino, Italy
- ¹⁵Pediatrics and Medical Genetics, University of Colorado, Aurora CO, USA

- ⁴⁵ ¹⁶Service de Médecine Génomique des Maladies de Système et d'Organe,
- 46 Département Médico-Universitaire BioPhyGen, Hôpital Cochin, APHP, Université Paris 47 Cité, Paris, France
- 47 Cite, Paris, France
- 48 ¹⁷Genetic Medicine, UCSF/Fresno, Fresno, CA, 93701, USA
- ⁴⁹ ¹⁸Neuroscience Research Center, Faculty of Medicine, Golestan University of Medical
- 50 Sciences, Gorgan, Iran; Nikagene Genetic Diagnostic Laboratory, Gorgan, Golestan, 51 Iran
- ¹⁹Department of Clinical Genetics, Erasmus MC University Medical Center, Rotterdam,
- 53 The Netherlands
- 54 ²⁰Boston Children's, Boston, MA, USA
- ²¹Children's Hospital of Eastern Ontario Research Institute, Ottawa, Ontario, Canada
- ²²Service de Génétique Moléculaire et Génomique, CHU, Rennes, F-35033, France;
- 57 Univ Rennes, CNRS, IGDR, UMR 6290, Rennes, F-35000, France
- ²³Children's Mercy Kansas City, Center for Pediatric Genomic Medicine, Kansas City,
 KS. USA
- 59 KS, USA 24 Children's M
- 60 ²⁴Children's Medical Center, Dallas TX, USA
- ²⁵Medical Genetics, Institute of Medical Genetics and Pathology, University Hospital
- 62 Basel, Basel, Switzerland
- ⁶³²⁶Research Unit for Rare Diseases, 1st Faculty of Medicine, Charles University in
- 64 Prague, Czech Republic
- ⁶⁵²⁷Department of Pediatrics and Inherited Metabolic Disorders, 1st Faculty of Medicine,
- 66 Charles University in Prague, Czech Republic
- ²⁸Genetics and Rare Diseases Research Division, Ospedale Pediatrico Bambino Gesù,
- 68 IRCCS, 00146, Rome, Italy.
- ²⁹Medical Genetics Unit and Thalassemia Center, San Luigi University Hospital,
- 70 University of Torino, Orbassano, Italy
- ³⁰Center for Integrative Genomics, University of Lausanne, Lausanne, Switzerland
- 72 ³¹Childrens Hospital of Philadelphia, Philadelphia PA, USA
- ³²Section of Genetics & Metabolism, Department of Pediatrics, University of Colorado
- 74 Anschutz Medical Campus, Aurora, CO 80045
- ³³Division of Genetics and Metabolism, Children's Health, Dallas, TX, USA
- ³⁴Pathology & Laboratory Medicine, Genomic Medicine, Neurological and Pediatrics
- 77 Institutes, Cleveland Clinic, Cleveland, OH, USA
- ³⁵Service de Génétique Clinique, Centre de Référence Maladies Rares CLAD-Ouest,
- 79 CHU Hôpital Sud, Rennes, France
- ³⁶Department of Pharmacology and Toxicology, College of Pharmacy, University of
- 81 Utah, Salt Lake City, UT 84112, USA
- 82 ³⁷University Children's Hospital Zurich, University of Zurich, Zurich 8032, Switzerland
- 83 ³⁸Service de Génétique Médicale et Biologie de la Reproduction, CHU de Brest, France
- ³⁹Centre de Référence Déficiences Intellectuelles de causes rares, Brest, France
- ⁴⁰Division of Genetics, Children's Mercy Kansas City, Kansas City, MO
- ⁴¹Genetics and Metabolism, Arkansas Children's Hospital, Little Rock, AR, 72202, USA
- ⁴²Department of Pediatrics, Division of Genetics and Metabolism, University of Texas
- 88 Southwestern Medical Center, Dallas TX, USA
- ⁴³Department of Endocrinology and Genetics, University Clinic for Children's Diseases,
- 90 Medical Faculty, University Sv. Kiril i Metodij, Skopje, Republic of Macedonia

It is made available under a CC-BY-NC-ND 4.0 International license .

91 92 03	⁴⁴ Amsterdam University Medical Centers, Department of Clinical Genetics, Amsterdam, The Netherlands ⁴⁵ Department of Human Genetics, Amsterdam University Medical Centers, VII
95 94	University Amsterdam
95	⁴⁶ Amsterdam Neuroscience, Amsterdam, The Netherlands
96	
97	
98	
99	
100	
101	
102	
103	
104	*Correspondence:
105	Susan M. Hiatt, Gregory M. Cooper
106	shiatt@hudsonalpha.org; gcooper@hudsonalpha.org
107	
108	
109	Key Words
110	ZMYM3, X-linked intellectual disability; neurodevelopmental disorder; transcriptional
111	coregulators; chromatin modifiers;

It is made available under a CC-BY-NC-ND 4.0 International license .

113 ABSTRACT

114 Neurodevelopmental disorders (NDDs) often result from highly penetrant variation in 115 one of many genes, including genes not yet characterized. Using the MatchMaker 116 Exchange, we assembled a cohort of 22 individuals with rare, protein-altering variation 117 in the X-linked transcriptional coregulator gene ZMYM3. Most (n=19) individuals were 118 males; 15 males had maternally-inherited alleles, three of the variants in males arose de 119 novo, and one had unknown inheritance. Overlapping features included developmental 120 delay, intellectual disability, behavioral abnormalities, and a specific facial gestalt in a 121 subset of males. Variants in almost all individuals (n=21) are missense, two of which are 122 recurrent. Three unrelated males were identified with inherited variation at R441, a site 123 at which variation has been previously reported in NDD-affected males, and two 124 individuals have de novo variation at R1294. All variants affect evolutionarily conserved 125 sites, and most are predicted to damage protein structure or function. ZMYM3 is 126 relatively intolerant to variation in the general population, is highly expressed in the 127 brain, and encodes a component of the KDM1A-RCOR1 chromatin-modifying complex. ChIP-seq experiments on one mutant, ZMYM3^{R1274W}, indicate dramatically reduced 128 129 genomic occupancy, supporting a hypomorphic effect. While we are unable to perform 130 statistical evaluations to support a conclusive causative role for variation in ZMYM3 in 131 disease, the totality of the evidence, including the presence of recurrent variation, 132 overlapping phenotypic features, protein-modeling data, evolutionary constraint, and 133 experimentally-confirmed functional effects, strongly supports ZMYM3 as a novel NDD 134 gene.

135

It is made available under a CC-BY-NC-ND 4.0 International license .

136

137 **INTRODUCTION**

Neurodevelopmental disorders (NDD) as a group affect 1-3% of children, but individual 138 139 NDDs are typically rare and often result from highly penetrant genetic variation affecting one of many NDD-associated loci^{1,2}. While exome/genome sequencing tests have 140 141 provided molecular diagnoses for many individuals, the diagnostic yield of NDDs with these tests remains below 50%³. Various hypotheses exist to explain this diagnostic 142 143 limitation, one of which is that many NDD-associated genes have yet to be identified. 144 The wide availability of sequencing tests, coupled with data sharing, has allowed identification of many new NDD genes over the last few years⁴. 145 146 X-linked intellectual disability (XLID) is a subset of NDDs that accounts for a 147 portion of the observed increased ratio of male to female probands with NDDs⁵. 148 However, it can be difficult to separate pathogenic and benign variation on the X chromosome since many pathogenic XLID variants are inherited from unaffected 149 heterozygous mothers^{6,7}. For autosomal dominant ID genes, in contrast, pathogenic 150 151 variation frequently arises de novo, and this fact can be used for both robust quantification of global disease enrichments and as a key piece of evidence for clinical 152 interpretation of individual variants⁸. Despite this, data from large sequence databases 153 such as gnomAD⁹ and TopMed/BRAVO (https://bravo.sph.umich.edu/freeze8/hg38/) do 154 155 provide critical frequency data from control populations to support (or refute) new XLID associations⁶. 156

ZMYM3 (MIM: 300061) lies on the X chromosome and encodes a member of a
 transcriptional corepressor complex that includes HDAC1, RCOR1, and KDM1A^{10,11}.

It is made available under a CC-BY-NC-ND 4.0 International license .

159 ZMYM3 was originally identified as an XLID candidate gene in a female with a balanced X;13 translocation affecting the 5' UTR of one isoform of ZMYM3¹². The proband 160 presented with ID, scoliosis, spotty abdominal hypopigmentation, slight facial 161 162 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 163 missense variant in ZMYM3 (R441W)¹³. The brothers displayed developmental delay, a 164 165 sleeping disorder, microcephaly, genitourinary anomalies and facial dysmorphism. 166 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 167 discovery over the last decade¹⁴. Here we describe a cohort of individuals with rare 168 variants in ZMYM3, assembled from submissions to GeneMatcher¹⁵ and 169 PhenomeCentral¹⁶. We provide strong evidence for an X-linked, *ZMYM3*-associated 170 171 NDD based on phenotypic, computational, and experimental analysis of variants 172 observed in 22 individuals. 173 174 SUBJECTS AND METHODS 175 ZMYM3 was submitted to GeneMatcher (https://genematcher.org/) by 176 HudsonAlpha in 2018, and follow-up discussion of cases from either research studies or 177 clinical sequencing was performed via email over the course of four years. Some matches originated from GeneMatcher¹⁵, while others originated from 178 PhenomeCentral¹⁶. Over the course of the collaboration, some affected individuals were 179 180

181 family members, including individuals harboring R688H, a variant that was initially

excluded from the cohort due to segregation of the variant of interest in unaffected male

182 identified as a VUS but later reclassified to likely benign after observation in an

It is made available under a CC-BY-NC-ND 4.0 International license .

unaffected male relative. Additionally, one of the individuals with R688H variation did
not have a phenotype that was similar to those described here; he had normocephaly,
ID, autism, developmental regression and facial dysmorphism dissimilar to those
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

190 panel testing was performed as described in Supplemental Materials and Methods.

191 For protein modeling, the wild-type 3D protein structure was downloaded from

192 AlphaFoldDB (https://alphafold.ebi.ac.uk/)¹⁷, which was included with the reference from

193 UniProt (Accession number: Q14202). When not possible online, structures were

visualized, colored, and the sequence was mutated with Chimera version 1.15, rotamer

¹⁹⁵ builder tool¹⁸. Structure superposition was obtained in Chimera with the tool

196 Matchmaker. Structure refinement was performed with the Chimera tool Dock Prep with

197 standard settings, as previously described¹⁹. Depiction of molecular surfaces was

198 defined as VdW surface and colored according to the electrostatic potential. For

199 additional analyses, see Supplemental Materials and Methods.

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) protocol²⁰. We nucleofected cells and selected for correctly edited cells using neomycin,

It is made available under a CC-BY-NC-ND 4.0 International license .

206	confirmed edits by PCR and Sanger sequencing of genomic DNA, and performed ChIP
207	seq as previously described ²¹ with duplicate experiments for each condition (see
208	Supplemental Materials and Methods). We performed peak calling using SPP ²² and
209	Irreproducible Discovery Rate (IDR) ²³ , using ENCODE-standardized pipelines for
210	analysis and quality-control ²⁴ . We performed additional differential binding analyses
211	using the R package csaw v1.28.0 ²⁵ .
212	As an additional control, we used the standard CETCh-seq ZMYM3 experiment
213	in HepG2 available on the ENCODE portal (ENCSR505DVB), with these data
214	processed to match (i.e., downsampled to 20M reads) the other CETCh-seq
215	experiments described here. See Supplemental Materials and Methods for additional
216	details.
217	
218	RESULTS
219	ZMYM3 Variants

Through a collaboration facilitated by the MatchMaker Exchange¹⁴, we identified 17 unique variants in *ZMYM3* in 22 affected individuals from 20 unrelated families (Figure 1, Table 1). All observed variants had high CADD scores (average 24.2, range 19-32, Table S1), indicating that they rank among the most deleterious SNVs in the human reference assembly, similar to most known highly penetrant NDD-associated variants²⁶.

Nineteen of these 22 individuals are males that harbor hemizygous missense
variants, including two sets of affected brothers. For the majority of males (n=15),
variants were inherited from heterozygous carrier mothers. In three males, the *ZMYM3*

It is made available under a CC-BY-NC-ND 4.0 International license .

229 variant arose *de novo*, while inheritance could not be defined for one. All variants were 230 rare, with three or fewer total alleles and no hemizygous males or homozygous females in anomAD⁹ or TopMed/Bravo (https://bravo.sph.umich.edu/freeze8/hg38/)(Table S1). 231 232 In addition, we identified three heterozygous ZMYM3 variants in three unrelated, 233 affected females (Figure 1, Table 1). All three variants are absent from population 234 databases. Two of these variants arose *de novo*, while one was inherited from an 235 apparently unaffected mother. In two of the three females, X-inactivation testing targeting either the AR locus²⁷ or the RP2 locus²⁸ was performed, and in both, skewed 236 237 X-inactivation was observed. In the case of the maternally-inherited L226WfsX8 variant, >94% skewing was observed in both the proband and her unaffected, heterozygous 238 239 mother at the RP2 locus. Both mother and daughter were heterozygous for two RP2 240 alleles (366/362), and in both, the 366 allele was inactivated (see Supplemental Note: 241 Case Reports for additional details). In Individual 20, a female carrying a de novo 242 R1294C variant, 97% skewing at the AR locus was observed. Skewing of the precise 243 ZMYM3 alleles was not tested in these individuals.

244

245 **Phenotypic Characterization**

Of the nineteen identified males, one died at 26 weeks gestational age with a *de novo* variant in *ZMYM3* (R1294C) 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 18 affected males, all were reported to have developmental delay (18/18), with speech delay (18/18) being more prominent than motor delay (13/18)(Table 1, Supplemental Note: Case Reports). Of those who could be

It is made available under a CC-BY-NC-ND 4.0 International license .

252	assessed, 13/15 showed intellectual disability, and most were diagnosed with autism or
253	were reported to have autistic traits (11/16). The majority of males had behavioral
254	concerns at some point in development (15/18). Almost all affected males were also
255	reported to have at least mild facial dysmorphism (14/18), some of which were highly
256	similar to the individuals reported in Philips, <i>et al.</i> ¹³ (Figure 2). Similarities include thick
257	eyebrows, deeply set eyes, long palpebral fissures, protruding ears, and a high anterior
258	hairline. Other variable features include genitourinary anomalies (9 individuals), short
259	stature (4), microcephaly (4), scoliosis/kyphosis (4), and functional gastrointestinal
260	problems (5) (Table 1). See Supplemental Note: Case Reports for additional clinical
261	features for each case.
262	While most variants in males were inherited from unaffected heterozygous carrier
263	mothers (9/13 mothers), the remaining four heterozygous mothers were reported to
264	have a history of learning disabilities, attention deficit-hyperactivity disorder (ADHD), or
265	dyslexia (Table 1, Supplemental Note: Case Reports, Figure S1).
266	Among the affected females, all three displayed developmental delay and some
267	facial dysmorphism, but many of their additional features were variable and do not lead
268	to a clear syndromic picture (Table 1, Figure 2, Supplemental Note: Case Reports).
269	
270	Protein Modeling
271	ZMYM3 encodes a DNA-binding transcriptional coregulator with multiple
272	protein isoforms, the longest of which is 1370 amino acids (Q14202, NP_005087.1).
273	This isoform has nine MYM-type zinc fingers, a proline-rich region, and a C-terminal
274	Cre-Like domain (Figure 1). As most of the observed variants are missense (16/17

It is made available under a CC-BY-NC-ND 4.0 International license .

unique variants), we performed computational modeling to assess the potential effects
of these changes. Homology-based protein modeling using AlphaFold¹⁷ indicates that
13 of the 16 missense variants lie in ordered regions, and the majority have
intermediate to high predicted Local Distance Difference Test (pLDDT) scores²⁹,
indicating that there is a moderate to high degree of confidence in further computational
predictions (Figure 3, Figure S2, Table S2).

281 We assessed flexibility, stability, solvent exposure, and deformation energy of 282 the mutant protein models (Figures S3-S6). A general trend towards protein 283 destabilization (negative folding energy differential) was observed for several mutants, 284 while R1274W was predicted to be stabilizing (Figure S3). We observed patterns 285 somewhat consistent with solvent exposure across the 16 unique missense variants 286 (Table S2). Six of the seven variants leading to the highest destabilization (R441Q, 287 E731D, Y752C, R1124Q, Y1137N, M1213T) represent buried residues in high 288 confidence regions of the protein. While disruption of each of these rigid residues is 289 predicted to be destabilizing, some are due to likely increased flexibility (E731D, Y752C, 290 Y1137N) while others are predicted to be more rigid (R441Q, R1124Q, M1213T). This 291 result is consistent with the observation that substitutions of amino acids within the 292 protein core are often associated with folding destabilization.

293 Conversely, the remaining 10 variants are exposed residues; six of these lie in 294 low confidence regions or have very low pLDDT values (D69N, E241K, R302H, R395S, 295 P398S, R1274W). These wild type residues are predicted to be flexible, and in most 296 cases the observed mutation is predicted to lead to a more rigid structure. The 297 remaining four residues represent rigid residues that have varying predicted deleterious

It is made available under a CC-BY-NC-ND 4.0 International license .

298 effects. More detailed surface analyses indicated that several variants result in 299 significant changes of polarity, charge, and hydrophobicity (Table S2, Figure S6). In 300 particular, R1274W is predicted to have major effects, resulting in stabilization of an 301 exposed residue through the substitution of a polar, charged and flexible arginine with a 302 neutral, aromatic and hydrophobic tryptophan moiety (Figure S6). 303 In addition to structural analysis, we submitted the sequences to the Eukaryotic Linear Motif (ELM)³⁰. This resource predicts short amino acid motifs either bound by 304 305 other proteins, or sites of post-translational modifications (phosphorylation, cleavage 306 sites, ubiquitination, etc.). Intersecting this information with the position of our mutations 307 suggests that several of the variants alter motifs (Table S2, Table S3) and that 308 modifications of residues R302, V1202, M1213, R1274, and M1343 are predicted to 309 possibly disrupt multiple interactions.

310

Genome-wide occupancy of selected ZMYM3 variant transcription factors

312 A key role of ZMYM3 is to function as a component of the KDM1A/RCOR1 chromatin-modifying complex that regulates gene expression¹⁰. Therefore, we sought to 313 314 measure the impact of variation on ZMYM3 activities genome-wide. Given the time and 315 expense of these experiments, we chose three variants for testing: R441W, a previously reported variant¹³ that affects a residue where we have seen recurrent variation 316 317 (R441Q); R1274W, a *de novo* variant within the Cre-like domain that was found in an 318 individual with notable facial similarities to those individuals with R441 variation; and 319 R688H, which early in our collaboration appeared to be a recurrent variant seen in two 320 affected individuals. Subsequently, segregation studies in one family indicated that the

It is made available under a CC-BY-NC-ND 4.0 International license .

R688H variant was present in an unaffected maternal uncle, suggesting that it is likelybenign.

323 For each of these, we introduced the variant into the ZMYM3 gene in the 324 genomic DNA of cultured HepG2 cells using a modified version of the CRISPR epitope tagging ChIP-seq (CETCh-seq) protocol²⁰; CETCh-seq uses an active Cas9 double-325 326 stranded DNA cleavage followed by homology-directed repair with an introduced donor 327 template. In these experiments, we simultaneously introduced a "super-exon" consisting 328 of all exons of ZMYM3 downstream (relative to coding direction) of the exon in which 329 the variant resides, along with a FLAG epitope tag and selectable resistance gene. 330 These modifications result in cells that express the ZMYM3 protein with the variant 331 residue and a carboxyl-terminus FLAG tag for immunoprecipitation, as well as a 332 neomycin resistance gene product for selection (cells that survive selection contain an 333 in-frame insertion of the neomycin resistance gene that is driven by the endogenous 334 ZMYM3 promoter and indicates the transcription of in-frame ZMYM3 with the FLAG 335 tag). As a control for each super-exon edit, we performed the same protocol on cells but 336 reintroduced the reference sequence in the super-exon along with the FLAG tag and 337 resistance gene. In both the control and variant experiments, an additional single base 338 substitution was introduced, located in the Protospacer Adjacent Motif (PAM) for the 339 single guide RNA (sgRNA) used in the CRISPR editing, in order to abolish Cas9 340 cleavage after cells are correctly edited. The PAM mutation was either synonymous (for the ZMYM3^{R688H} and ZMYM3^{R441W} constructs) or intronic (for the ZMYM3^{R1274W} 341 342 construct). The genomic DNA modifications were confirmed by PCR amplification and 343 Sanger sequencing. The key advantage of this approach is that the control and variant

It is made available under a CC-BY-NC-ND 4.0 International license .

344 ZMYM3 proteins are produced from the endogenous genomic loci, each modified by the 345 same super-exon, and that the antibody used (along with other experimental and 346 analytical steps) is the same: the only difference between the variant and control 347 experiments is the presence of the missense variant of interest. For both R688H and 348 R1274W, we successfully obtained correctly edited cells; however, for R441W, we were 349 unable to obtain edited cells. For both R688H and R1274W, we performed chromatin 350 immunoprecipitation followed by high-throughput sequencing (ChIP-seq) and peak calling as previously described^{21,31}, downsampling all replicate experiments to 20 million 351 352 reads to control for total read-depth (Table S4), and compared the genome-wide binding 353 profiles of the control and variant experiments. As an additional control, we used a standard CETCh-seq experiment on ZMYM3 (ZMYM3^{CETCh}) in HepG2 cells to establish 354 355 baseline genome-wide occupancy of the wild-type ZMYM3 protein without a super-exon 356 or missense variant (ENCODE dataset ENCSR505DVB).

When comparing ZMYM3^{R1274W}-control (i.e. super-exon with reference sequence 357 at the variant position) to ZMYM3^{R1274W}-variant, we observed a large difference in the 358 359 number of peaks called between the experiments, with the control experiment yielding 360 16,214 peaks and the variant only 3,699 peaks (Table S4). Among the variant peaks, 361 2,508 (67.8%) were also called in the control experiment. We know from extensive 362 previous ChIP-seq analyses that many loci exhibit read-depth levels near (above or 363 below) peak-calling thresholds, resulting in situations where experiments are more 364 similar than they appear when only considering peak-call overlaps. Thus, we performed additional, more quantitative comparisons, including read-depth correlations and 365 366 overlaps of peak-calls using both standard and relaxed thresholds; these analyses

It is made available under a CC-BY-NC-ND 4.0 International license .

367 consistently show a large, global reduction of occupancy for R1274W variant 368 experiments (see Supplemental Materials and Methods). The large disparity between the control and variant ZMYM3^{R1274W} experiments was also observed in a differential 369 370 analysis using the R package $csaw^{25}$. Rather than relying on peak calls, csaw performs 371 a sliding window analysis to detect regions that show differing levels of read-depth between the experiments. Using the two replicates of ZMYM3^{R1274W}-control and the two 372 373 replicates of ZMYM3^{R1274W}-variant, *csaw* identified 25,845 genomic regions with 374 sufficient reads for analysis. Of these regions, 13,225 showed differential associations 375 between control and variant at FDR < 0.05. All but 19 (99.9%) of these sites had higher 376 read counts in the control than in the variant. We also intersected the 25,845 csaw 377 regions with the union of peak calls between control and variant experiments 378 (n=17,405), resulting in 11,259 genomic regions; of these, 4,625 were not differential at 379 FDR < 0.05, while 6,631 were significantly higher in control than in variant, and only 380 three were significantly higher in variant than in control (Figure 4A). We performed the *csaw*/peak overlap analyses between ZMYM3^{CETCh} and the R1274W variant and control 381 experiments and observed that ZMYM3^{CETCh} and ZMYM3^{R1274W}-control had very similar 382 genomic association patterns, while ZMYM3^{R1274W}-variant comprised only a small 383 minority of ZMYM3^{CETCh} sites (Figure S7). 384

We similarly analyzed the ZMYM3^{R688H}-control and ZMYM3^{R688H}-variant experiments. Global correlation analyses indicated a high degree of similarity of genomic occupancy between the control and variant experiments (Supplemental Materials and Methods). Analysis of *csaw* regions intersected with peak calls gave 6,416 genomic sites, none of which were significantly different (FDR < 0.05) between

390	control and variant (Figure 4B). Pairwise Pearson correlation coefficients of read counts
391	of each of the two replicates of control and variant ranged from 0.71 to 0.84, indicating a
392	high degree of overall similarity (Figure S8). However, both control and variant R688H
393	experiments yielded less binding and fewer peaks than both ZMYM3 ^{R1274W} -control and
394	ZMYM3 ^{CETCh} experiments (Figure S7). We hypothesize that the super-exon edit itself is
395	affecting the ZMYM3 ^{R688H} function. Importantly, we find no substantial difference
396	between control and variant function, supporting the hypothesis that R688H is a likely
397	benign variant.
398	

It is made available under a CC-BY-NC-ND 4.0 International license .

399 **DISCUSSION**

400 Here we describe 22 NDD-affected individuals with protein-altering variation in 401 ZMYM3. While four of these variants arose de novo, most were present in males in a 402 hemizygous state and inherited from unaffected or mildly affected heterozygous 403 mothers. All variants presented here are rare and predicted to be deleterious. Many of 404 the variants are predicted to interfere with protein structure or function. ZMYM3 is 405 relatively intolerant to both missense variation (gnomAD missense Z=4.31) and loss-offunction variation (RVIS=8.46³², pLOEUF=0.11⁹), further supporting the potential for the 406 407 variants observed here to have phenotypic effects, and suggesting that the gene is likely dosage sensitive. Using ChIP-seg, we have also provided functional analyses 408 409 showing that one deleterious variant, R1274W, acts as a hypomorphic allele, with 410 reduced genome occupancy compared to its control and to wild type, and that one likely 411 benign allele, R688H, has genome occupancy similar to its control experiment. 412 Among the variants in our cohort, there are two sets of alleles affecting the same 413 codon. At R441, a residue that lies within a zinc finger domain that functions in DNA 414 binding, we found substitutions (R441Q) in three unrelated males. Three additional 415 affected males with a distinct allele at this same residue (R441W) in one family have been previously reported¹³. Overlapping phenotypic features of these six individuals 416 417 include developmental delay (mainly speech), nocturnal enuresis, and microcephaly. In

addition, the facial features in these R441 individuals are quite similar. The other
recurrent variant that we observed here is R1294C, observed as *de novo* in a deceased
male and in a female with 97% skewed X-inactivation. R1294C has also been submitted
to ClinVar³³ as a VUS (SCV000297052.2) by a different group than those that identified

It is made available under a CC-BY-NC-ND 4.0 International license .

422 R1294C variation for this study. We thus believe the ClinVar submission represents a

423 third, independent R1294C case, although we are unable to confirm this (see

424 Supplemental Materials and Methods).

425 The biological role of ZMYM3 is also supportive of disease relevance. ZMYM3 is part of a transcriptional corepressor complex that includes HDAC1, RCOR1 and 426 KDM1A^{10,11}. Additional interactors in this complex can include ZMYM2 and REST. 427 428 Variation in two of these five genes has been robustly associated with neurodevelopmental disorders^{34,35}. Additionally, ZMYM3 has since been shown to 429 physically interact with RNASEH2A; variation in RNASEH2A (MIM:606034) has been 430 431 associated with Aicairdi-Gouteres Syndrome 4 (AGS4, MIM:610333). Specifically, a 432 cluster of pathogenic variants found in AGS4 patients have been shown to disrupt binding of RNAseH to ZMYM3¹¹. Residues within the PV-rich domain of ZMYM3 433 434 (specifically 862-943) have been shown to be necessary for this interaction. I932V, 435 observed in our cohort, lies in this region and may disrupt this interaction. 436 Recently, Connaughton et al. demonstrated a connection between loss-offunction variation in ZMYM2 (MIM:602221), a paralog of ZMYM3 with 44% protein-437 identity, to congenital anomalies of the kidney and urinary tract, with extra-renal features 438 or NDD findings (MIM:619522)³⁴. This same publication also reported two male 439 440 probands who had hemizygous variants of uncertain significance in ZMYM3, resulting in 441 G673D and V866M (Figure 1). Phenotypic overlap of individuals with variation in 442 ZMYM2 and ZMYM3 presented here include developmental delay, microcephaly, and ID. Some similarity of facial features is also shared with the ZMYM2 cohort, including 443 444 one proband with protuberant ears. In addition to ZMYM2 and ZMYM3, the ZMYM-

It is made available under a CC-BY-NC-ND 4.0 International license .

family of proteins includes two additional members, ZMYM4, and QRICH1. Variation in

446 *QRICH1* (MIM:617387) has been associated with Ververi-Brady syndrome

447 (MIM:617387), which has features including developmental delay, intellectual disability,

448 non-specific facial dysmorphism, and hypotonia among others 36 .

449 Variants observed in this cohort lie across the length of the protein, and modeling 450 data suggest that while several may affect protein structure, several also likely affect 451 protein-interactions, which are key in the biological function of ZMYM3. ChIP-seg data for ZMYM3^{R1274W} indicate a large reduction in genome-wide occupancy, even though 452 453 the variant is not within any direct DNA-binding domains. Leung et al. have previously 454 shown that this specific residue is necessary for interaction with RAP80, a ubiquitinbinding protein that plays a role in the DNA damage response³⁷. While the observed 455 456 widespread reduction in genomic occupancy indicates a global hypomorphic effect. 457 individual binding event differences may be of particular interest. For example, one of 458 the most significant differential binding events, as determined by csaw, occurs at a 459 regulatory element on chromosome 8 (Figure 4C); this region is annotated as a distal enhancer by the ENCODE Consortium³⁸, and, according to Activity-by-Contact (ABC) 460 analysis³⁹, this region connects to and is likely a regulatory element for the gene 461 462 EFR3A. Pathogenic variants in EFR3A have been associated with autism spectrum disorders⁴⁰, with phenotypes that overlap those described here. 463

A key limitation of this study is the X-linked nature of *ZMYM3*, and the fact that most of the variants observed here are inherited, 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 NDDs⁴¹. Traditional

It is made available under a CC-BY-NC-ND 4.0 International license .

468 association or burden testing also cannot be done given the absence of systematically 469 ascertained and matched cases and controls. However, this is a common limitation of 470 establishing X-linked recessive disease in males. Additionally, while many pedigrees 471 suggest that testing in other family members may be informative for each individual 472 variant's interpretation (Figure S1), we were unable to test all siblings and relatives. This additional information may be particularly useful for flagging any potential benign 473 474 variants within these families (i.e., those present in unaffected male relatives, such as 475 was observed for R688H). X-chromosome inactivation studies in carrier females may 476 also be informative. 477 Despite the above limitations, the totality of the evidence presented here is 478 strong. The presence of recurrent variation, overlapping phenotypic features, protein-479 modeling data, evolutionary constraint, and experimentally-confirmed functional effects

480 strongly support *ZMYM3* as a novel NDD gene. While additional analyses are

481 necessary to ultimately confirm these findings and adjudicate the pathogenicity of each

individual variant, we provide substantial evidence that *ZMYM3* is likely to be a novel

483 NDD-associated gene.

It is made available under a CC-BY-NC-ND 4.0 International license .

485 **Declaration of interests.**

486 JLB, YC, BRL, AGN and HZE are employees of GeneDx, LLC. SEA is a cofounder and

487 CEO of MediGenome, the Swiss Institute of Genomic Medicine. All other authors

- 488 declare no competing interests.
- 489

490 **Acknowledgements**

491 We sincerely thank all the families who participated in this study.

492 SMH, AC, ACEH, MD and GMC were supported by a grant from the Alabama Genomic

- 493 Health Initiative (an Alabama-State earmarked project F170303004) through the
- 494 University of Alabama in Birmingham. LN was supported by the project National

495 Institute for Neurological Research (Programme EXCELES, ID Project No.

- 496 LX22NPO5107), Funded by the European Union, Next Generation EU and by
- 497 institutional program UNCE/MED/007 of Charles University in Prague. SK was

498 supported by grant NV19-07-00136 from the Ministry of Health of the Czech Republic.

499 LN and SK thank the National Center for Medical Genomics (LM2018132) for WES

analyses. Sequencing and analysis of one individual in this study was made possible by

- 501 the generous gifts to Children's Mercy Research Institute and Genomic Answers for
- 502 Kids program at Children's Mercy Kansas City. This work was also supported by the

503 Italian Ministry of Health (Ricerca 5x1000, RCR-2020-23670068_001, and RCR-2021-

504 23671215 to MT), and Italian Ministry of Research (FOE 2019, to MT), and PRIN2020

- 505 (code 20203P8C3X, to AB). Reanalysis of exome sequencing for individual 14 was
- 506 performed on a research basis by the Care4Rare Canada Consortium. This work was
- 507 supported by grants from the Swiss National Science Foundation (31003A_182632 to

It is made available under a CC-BY-NC-ND 4.0 International license .

508	AR) and the Blackswan Foundation (to AR), and the ChildCare Foundation to SEA. This
509	work was also funded through the CRT Foundation (Progam "Erogazioni Ordinarie"
510	2019) and the Italian Ministry of University and Research (Assegni, Tornata 2022,
511	Bando: BMSS.2022.06/XXIV), to MRS, GC and GE. MM was supported by RVO VFN
512	64165, Czech Ministry of Health. This work was also supported by a Swiss National
513	Science Foundation grant 320030_179547 to AR. CN and JD were supported by The
514	Genesis Foundation for Children.
515	
516	
517	
518	Data and code availability
519	The published article includes all variant information pertinent to this study. ChIP-seq
520	data will be made available via the NCBI Gene Expression Omnibus (GEO,
521	https://www.ncbi.nlm.nih.gov/geo/).
522	
523	
504	

It is made available under a CC-BY-NC-ND 4.0 International license .

525 **References**

- 526 1. Ropers, H.H. (2008). Genetics of intellectual disability. Curr Opin Genet Dev *18*, 241–
 527 250.
- 528 2. Cooper, G.M., Coe, B.P., Girirajan, S., Rosenfeld, J.A., Vu, T.H., Baker, C., Williams,
- 529 C., Stalker, H., Hamid, R., Hannig, V., et al. (2011). A copy number variation morbidity
- 530 map of developmental delay. Nat. Genet. *43*, 838–846.
- 531 3. Neu, M.B., Bowling, K.M., and Cooper, G.M. (2019). Clinical utility of genomic
- 532 sequencing. Curr. Opin. Pediatr. 31, 732–738.
- 4. Bamshad, M.J., Nickerson, D.A., and Chong, J.X. (2019). Mendelian Gene
- 534 Discovery: Fast and Furious with No End in Sight. Am. J. Hum. Genet. *105*, 448–455.
- 535 5. Martin, H.C., Gardner, E.J., Samocha, K.E., Kaplanis, J., Akawi, N., Sifrim, A.,
- 536 Eberhardt, R.Y., Tavares, A.L.T., Neville, M.D.C., Niemi, M.E.K., et al. (2021). The
- 537 contribution of X-linked coding variation to severe developmental disorders. Nat.
- 538 Commun. 12,.
- 539 6. Piton, A., Redin, C., and Mandel, J.L. (2013). XLID-causing mutations and associated
- 540 genes challenged in light of data from large-scale human exome sequencing. Am. J.
- 541 Hum. Genet.
- 542 7. Migeon, B.R. (2020). X-linked diseases: susceptible females. Genet. Med. 2020 227
 543 22, 1156–1174.
- 544 8. Deciphering Developmental Disorders, S. (2015). Large-scale discovery of novel
- 545 genetic causes of developmental disorders. Nature *519*, 223–228.
- 546 9. Lek, M., Karczewski, K.J., Minikel, E. V, Samocha, K.E., Banks, E., Fennell, T.,
- 547 O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., et al. (2016). Analysis of

- 548 protein-coding genetic variation in 60,706 humans. Nature 536, 285–291.
- 549 10. Hakimi, M.-A., Dong, Y., Lane, W.S., Speicher, D.W., and Shiekhattar, R. (2003). A
- 550 Candidate X-linked Mental Retardation Gene Is a Component of a New Family of
- 551 Histone Deacetylase-containing Complexes. J. Biol. Chem. 278, 7234–7239.
- 11. Shapson-Coe, A., Valeiras, B., Wall, C., and Rada, C. (2019). Aicardi-Goutières
- 553 Syndrome associated mutations of RNase H2B impair its interaction with ZMYM3 and
- the CoREST histone-modifying complex. PLoS One 14,.
- 555 12. van der Maarel, S.M., Scholten, I.H., Huber, I., Philippe, C., Suijkerbuijk, R.F.,
- 556 Gilgenkrantz, S., Kere, J., Cremers, F.P., and Ropers, H.H. (1996). Cloning and
- 557 characterization of DXS6673E, a candidate gene for X-linked mental retardation in
- 558 Xq13.1. Hum. Mol. Genet. 5, 887–897.
- 13. Philips, A.K., Sirén, A., Avela, K., Somer, M., Peippo, M., Ahvenainen, M., Doagu,
- 560 F., Arvio, M., Kääriäinen, H., Van Esch, H., et al. (2014). X-exome sequencing in
- 561 Finnish families with Intellectual Disability Four novel mutations and two novel
- 562 syndromic phenotypes. Orphanet J. Rare Dis.
- 563 14. 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.
- 565 Hum. Mutat. 43, 659–667.
- 566 15. Hamosh, A., Wohler, E., Martin, R., Griffith, S., Rodrigues, E. da S., Antonescu, C.,
- 567 Doheny, K.F., Valle, D., and Sobreira, N. (2022). The impact of GeneMatcher on
- 568 international data sharing and collaboration. Hum. Mutat. *43*, 668–673.
- 16. Osmond, M., Hartley, T., Johnstone, B., Andjic, S., Girdea, M., Gillespie, M., Buske,
- 570 O., Dumitriu, S., Koltunova, V., Ramani, A., et al. (2022). PhenomeCentral: 7 years of

- 571 rare disease matchmaking. Hum. Mutat. 43, 674–681.
- 572 17. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O.,
- 573 Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. (2021). Highly accurate
- 574 protein structure prediction with AlphaFold. Nature *596*, 583–589.
- 575 18. Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng,
- 576 E.C., and Ferrin, T.E. (2004). UCSF Chimera--a visualization system for exploratory
- research and analysis. J. Comput. Chem. 25, 1605–1612.
- 578 19. Rossi Sebastiano, M., Ermondi, G., Hadano, S., and Caron, G. (2022). Al-based
- 579 protein structure databases have the potential to accelerate rare diseases research:
- 580 AlphaFoldDB and the case of IAHSP/Alsin. Drug Discov. Today 27, 1652–1660.
- 581 20. Savic, D., Partridge, E.C., Newberry, K.M., Smith, S.B., Meadows, S.K., Roberts,
- 582 B.S., Mackiewicz, M., Mendenhall, E.M., and Myers, R.M. (2015). CETCh-seq: CRISPR
- 583 epitope tagging ChIP-seq of DNA-binding proteins. Genome Res. 25, 1581.
- 584 21. Meadows, S.K., Brandsmeier, L.A., Newberry, K.M., Betti, M.J., Nesmith, A.S.,
- 585 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,
- 587 **3–34**.
- 588 22. Kharchenko, P. V., Tolstorukov, M.Y., and Park, P.J. (2008). Design and analysis of
- 589 ChIP-seg experiments for DNA-binding proteins. Nat. Biotechnol. 2008 2612 26, 1351-
- 590 1359.
- 591 23. Li, Q., Brown, J.B., Huang, H., and Bickel, P.J. (2011). Measuring reproducibility of
- 592 high-throughput experiments. Https://Doi.Org/10.1214/11-AOAS466 5, 1752–1779.
- 593 24. Landt, S.G., Marinov, G.K., Kundaje, A., Kheradpour, P., Pauli, F., Batzoglou, S.,

- 594 Bernstein, B.E., Bickel, P., Brown, J.B., Cayting, P., et al. (2012). ChIP-seg guidelines
- and practices of the ENCODE and modENCODE consortia. Genome Res. 22, 1813–1831.
- 597 25. Lun, A.T.L., and Smyth, G.K. (2016). csaw: a Bioconductor package for differential
- 598 binding analysis of ChIP-seq data using sliding windows. Nucleic Acids Res. 44, e45.
- 599 26. Kircher, M., Witten, D.M., Jain, P., O'Roak, B.J., Cooper, G.M., and Shendure, J.
- 600 (2014). A general framework for estimating the relative pathogenicity of human genetic
- 601 variants. Nat Genet *46*, 310–315.
- 602 27. Bertelsen, B., Tümer, Z., and Ravn, K. (2011). Three new loci for determining x
- 603 chromosome inactivation patterns. J. Mol. Diagn. 13, 537–540.
- 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., Chuva De Sousa Lopes, S.M., Serafim Da
- 606 Silveira, L., et al. (2014). 5meCpG epigenetic marks neighboring a primate-conserved
- 607 core promoter short tandem repeat indicate X-chromosome inactivation. PLoS One 9,.
- 29. Mariani, V., Biasini, M., Barbato, A., and Schwede, T. (2013). IDDT: a local
- superposition-free score for comparing protein structures and models using distance
- 610 difference tests. Bioinformatics *29*, 2722–2728.
- 30. Kumar, M., Michael, S., Alvarado-Valverde, J., McRossed D Sign©száros, B.,
- 512 Sámano-Sánchez, H., Zeke, A., Dobson, L., Lazar, T., Örd, M., Nagpal, A., et al. (2022).
- 613 The Eukaryotic Linear Motif resource: 2022 release. Nucleic Acids Res. 50, D497–
- 614 D508.
- 615 31. Johnson, D.S., Mortazavi, A., Myers, R.M., and Wold, B. (2007). Genome-wide
- mapping of in vivo protein-DNA interactions. Science (80-.). *316*, 1497–1502.

- 617 32. Petrovski, S., Wang, Q., Heinzen, E.L., Allen, A.S., and Goldstein, D.B. (2013).
- 618 Genic intolerance to functional variation and the interpretation of personal genomes.
- 619 PLoS Genet 9, e1003709.
- 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
- 622 of clinically relevant variants. Nucleic Acids Res 44, D862-8.
- 623 34. 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).
- 625 Mutations of the Transcriptional Corepressor ZMYM2 Cause Syndromic Urinary Tract
- 626 Malformations. Am. J. Hum. Genet. 107, 727–742.
- 35. 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
- 629 Mendelian conditions via social networking: de novo variants in KDM1A cause
- 630 developmental delay and distinctive facial features. Genet. Med. 18, 788–795.
- 36. 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.
- 634 Mutat. *43*, 266–282.
- 635 37. 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
- 638 Dev. 31, 260–274.
- 639 38. Dunham, I., Kundaje, A., Aldred, S.F., Collins, P.J., Davis, C.A., Doyle, F., Epstein,

It is made available under a CC-BY-NC-ND 4.0 International license .

- 640 C.B., Frietze, S., Harrow, J., Kaul, R., et al. (2012). An integrated encyclopedia of DNA
- 641 elements in the human genome. Nature 489, 57–74.
- 642 39. Nasser, J., Bergman, D.T., Fulco, C.P., Guckelberger, P., Doughty, B.R.,
- 643 Patwardhan, T.A., Jones, T.R., Nguyen, T.H., Ulirsch, J.C., Lekschas, F., et al. (2021).
- 644 Genome-wide enhancer maps link risk variants to disease genes. Nature 593, 238–243.
- 40. Gupta, A.R., Pirruccello, M., Cheng, F., Kang, H.J., Fernandez, T. V., Baskin, J.M.,
- 646 Choi, M., Liu, L., Ercan-Sencicek, A.G., Murdoch, J.D., et al. (2014). Rare deleterious
- 647 mutations of the gene EFR3A in autism spectrum disorders. Mol. Autism 5,.
- 41. Samocha, K.E., Robinson, E.B., Sanders, S.J., Stevens, C., Sabo, A., McGrath,
- L.M., Kosmicki, J.A., Rehnstrom, K., Mallick, S., Kirby, A., et al. (2014). A framework for
- 650 the interpretation of de novo mutation in human disease. Nat Genet 46, 944–950.

651

652

653

It is made available under a CC-BY-NC-ND 4.0 International license .

655 Figure Legends

656 Figure 1. Observed variation along the length of the ZMYM3 protein. The 1370 aa

- 657 ZMYM3 protein (Q14202, NP_005087.1) is annotated with MYM-type zinc fingers (1-9,
- orange), a Proline-Valine Rich region (PV Rich, red), and Cre-like domain (blue) as
- described by UniProt. A. Hemizygous variants observed in males in this study are noted
- above the protein model, with *de novo* variants in red. (a) Note that R441Q was
- observed in three unrelated males. Hemizygous variants that were previously reported
- in males are shown below the protein (from PMID: 32891193, PMID: 24721225). B.
- 663 Maternally-inherited (black) or *de novo* (red) heterozygous variants observed in females
- in this study are noted above the protein model.
- **Figure 2. Facial features of a subset of individuals with** *ZMYM3* **variation.**
- 666 Individual ID and protein effect are noted for each. Note deep-set eyes, long palpebral
- 667 fissures, large/prominent/cupped ears, and tall forehead.

668 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
- 671 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, 241, and 302 lie in
- disordered regions, the remainder of residues lie in structured regions.

676 Figure 4. R1274W is a hypomorphic allele, while R688H has similar genome

677 occupancy to that of wild type. A, B. Genomic regions called by csaw between

It is made available under a CC-BY-NC-ND 4.0 International license .

678 experiments, then overlapped with IDR 0.05 peaks called in either experiment. Yellow 679 color indicates regions determined by csaw to have significantly higher differential 680 binding (at FDR < 0.05) in control, orange indicates regions with no differential binding, 681 and red indicates regions with higher differential binding in variant. For R1274W (A), 682 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 R688H (B), all 6,416 regions 683 684 had no differential binding. C. Genome browser track for ZMYM3-R1274W-variant ChiP-685 seq experiments. Human genome (hq38) chr8:131,561,953-131,925,404 is displayed. 686 Top track is Activity-by-Contact ("ABC loops") showing predicted interaction between 687 enhancer element on left and TSS for the gene EFR3A on right. "Genes" track is 688 RefSeq gene model. "Control Rep 1" and "Control Rep 2" are aligned bam reads from 689 ZMYM3-R1274W-control experiments, "Variant Rep 1" and "Variant Rep 2" are aligned 690 bam reads from ZMYM3-R1274W-variant experiments. All bam files are downsampled 691 to an equal number of reads in each replicate, and all four replicate tracks are scaled 692 from 0-60 vertically. "ENCODE" tracks are shown below the ChIP-seq tracks: "cCREs" 693 represent candidate cis-regulatory elements colored by ENCODE standards, 694 "H3K27Ac" is layered H3K27Ac signal from seven ENCODE cell lines, and "TFBSs" are 695 ENCODE TF clusters (340 factors, 129 cell types). The putative enhancer element 696 identified as the most significant loss of binding in the variant experiment is highlighted 697 in yellow, showing the ENCODE distal enhancer cCRE call, the ABC loop to the TSS of 698 *EFR3A*, and the difference in binding with the two control replicates showing strong 699 signal and the two variant replicates showing substantially less binding. **D.** Protein 700 modeling of R1274 and R1274W.

Individual	Gender	Zygosity	Inheritance	Mother's Phenotype	Variant (NP_005087.1)	Speech Delay	Motor Delay	ID	ASD Traits	Behavioral Problems	Facial Dys- morphism	GU Anomalies	Other
1	Male	Hemizygous	Maternal	None	p.(D69N)	Yes	Yes	NA	NA	No	No	Urinary tract dilatation of left kidney on ultrasound	Congenital heart defects
2	Male	Hemizygous	Maternal	Hx of LD	р.(E241K)	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	Hemizygous	Maternal	None	p.(R302H)	Yes	Yes	NA	Yes	Yes	Yes	Pielonephritis, vesicoureteral reflux	GERD
4a	Male	Hemizygous	Maternal	None	p.(R395S)	Yes	Yes	Yes	Yes	Yes	Yes	Hypospadias	
4b	Male	Hemizvaous	Maternal	None	p.(R395S)	Yes	No	Yes	Yes	Yes	Yes	No	
5	Male	Hemizvaous	Maternal	None	p.(P398S)	Yes	Yes	Yes	Yes	Yes	No	No	Weight <1%ile
6	Male	Hemizygous	Maternal	ADHD	p.(R441Q)	Yes	Yes	Yes	Yes	Yes	Yes	Single renal cyst	Constipation
7	Male	Hemizygous	Maternal	None	p.(R441Q)	Yes	Yes	Yes	Yes	Yes	Yes	Cryptorchidism, enuresis	Short stature
8	Male	Hemizygous	Maternal	Hx of LD	p.(R441Q)	Yes	Yes	Yes	Yes	Yes	Yes	Hypospadias, ambiguous genitalia	Short stature
9a	Male	Hemizygous	Maternal	None	p.(E731D)	Yes	No	Yes	Yes	Yes	Yes	No	
9b	Male	Hemizygous	Maternal	None	p.(E731D)	Yes	No	Yes	Yes	Yes	Yes	No	
10	Male	Hemizygous	Maternal	Dyslexia	p.(I932V)	Yes	No	NA	NA	Yes	Yes	No	GERD, constipation
11	Male	Hemizygous	Maternal	None	p.(R1124Q)	Yes	No	Yes	No	Yes	Yes	Ectopic kidney	Short stature, kyphoscoliosis
12	Male	Hemizygous	Maternal	None	p.(Y1137N)	Yes	Yes	Yes	No	Yes	Yes	No	Microcephaly
13	Male	Hemizygous	De novo	None	p.(V1202D)	Yes	Yes	Yes	No	Yes	Yes	Cryptorchidism	Microcephaly, short stature, weight <3%ile, kyphosis, long bone defects, Madelung deformity
14	Male	Hemizygous	Unknown	None	p.(M1213T)	Yes	Yes	Yes	No	No	No	Enuresis	Microcephaly, scoliosis, reflux
15	Male	Hemizygous	De novo	None	p.(R1274W)	Yes	Yes	Yes	Yes	Yes	Yes	No	Microcephaly, scoliosis
16 17	Male Male	Hemizygous Hemizygous	De novo Maternal	None None	р.(R1294C) р.(M1343I)	NA Yes	NA Yes	NA No	NA Yes	NA Yes	NA No	NA No	Deceased GI dysmotility, joint laxity, pain & swelling, dysautonomic symptoms
					Total	18/18	13/18	13/15	11/16	15/18	14/18	9/18	
18	Female	Heterozygous, skewed XCI	Maternal	None	p.(L226WfsX8)	Yes	Yes	NA	NA	No	Yes	No	GERD
19	Female	Heterozygous	De novo	None	p.(Y752C)	Yes	Yes	NA	NA	No	Yes	No	
20	Female	Heterozygous, skewed XCI	De novo	Unknown	p.(R1294C)	Yes	Yes	NA	NA	NA	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. NA, Not assessed. GERD, Gastroesophageal reflux disease. Hx, history. LD, learning disability. ADHD, attention deficit-hyperactivity disorder.

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



It is made available under a CC-BY-NC-ND 4.0 International license .

Figure 2 has been removed due to medRxiv photograph policy.



It is made available under a CC-BY-NC-ND 4.0 International license .



Figure 4