

1 **Systematic analysis and prediction of genes associated with disorders**

2 **on chromosome X**

3 Elsa Leitão,¹ Christopher Schröder,¹ Ilaria Parenti,¹ Carine Dalle,² Agnès Rastetter,² Theresa
4 Kühnel,¹ Alma Kuechler,¹ Sabine Kaya,¹ Bénédicte Gérard,³ Elise Schaefer,⁴ Caroline Nava,²
5 Nathalie Drouot,⁵ Camille Engel,⁵ Juliette Piard,^{6,15} Bénédicte Duban-Bedu,⁷ Laurent Vil-
6 lard,^{8,9} Alexander P.A. Stegmann,^{10,11} Els K. Vanhoutte,¹¹ Job A.J Verdonshot^{11,12}, Frank J.
7 Kaiser,¹ Frédéric Tran Mau-Them,^{13,14,15} Marcello Scala,^{16,17} Pasquale Striano,^{16,17} Suzanna
8 G.M. Frints,^{11,18} Emanuela Argilli,^{19,20} Elliott H. Sherr,^{19,20} Fikret Elder,²¹ Julien Buratti,²¹ Boris
9 Keren,²¹ Cyril Mignot,^{2,21,22} Delphine Héron,^{21,22} Jean-Louis Mandel,^{3,5,23,24,25} Jozef Gecz,^{26,27}
10 Vera M. Kalscheuer,²⁸ Bernhard Horsthemke,¹ Amélie Piton,^{3,5,24,24,25} Christel Depienne^{1*}

11 ¹ Institute of Human Genetics, University Hospital Essen, University Duisburg-Essen, Essen,
12 Germany

13 ² Institut du Cerveau et de la Moelle épinière (ICM), Sorbonne Université, UMR S 1127, In-
14 serm U1127, CNRS UMR 7225, F-75013 Paris, France

15 ³ Unité de Génétique Moléculaire, IGMA, Hôpitaux Universitaire de Strasbourg, Strasbourg,
16 France

17 ⁴ Service de Génétique Médicale, IGMA, Hôpitaux Universitaires de Strasbourg, Strasbourg,
18 France

19 ⁵ Institut de Génétique et de Biologie Moléculaire et Cellulaire, Illkirch 67400, France

20 ⁶ Centre de Génétique Humaine, Université de Franche-Comté, Besançon, France.

21 ⁷ Centre de génétique chromosomique, Hôpital Saint-Vincent de Paul, Lille, France.

22 ⁸ Aix-Marseille University, INSERM, MMG, UMR-S 1251, Faculté de médecine, Marseille,
23 France

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

24 ⁹ Département de Génétique Médicale, APHM, Hôpital d'Enfants de La Timone, Marseille,
25 France

26 ¹⁰ Department of Human Genetics, Radboud University Medical Center, 6500 HB, Nijmegen,
27 the Netherlands

28 ¹¹ Department of Clinical Genetics, Maastricht University Medical Center⁺, Maastricht, The
29 Netherlands

30 ¹² Cardiovascular Research Institute (CARIM), Departments of Cardiology, Maastricht Univer-
31 sity Medical Center, the Netherlands

32 ¹³ Unité Fonctionnelle Innovation en Diagnostic génomique des maladies rares, CHU Dijon
33 Bourgogne, Dijon, France

34 ¹⁴ Centre de Génétique, Centre de Référence Anomalies du Développement et Syndromes
35 Malformatifs, et Centre de Compétence Maladies Mitochondriales, FHU TRANSLAD, Hôpital
36 d'Enfants, CHU de Dijon, France

37 ¹⁵ INSERM UMR1231, Equipe Génétique des Anomalies du Développement, Université de
38 Bourgogne, Dijon, France

39 ¹⁶ Department of Neurosciences, Rehabilitation, Ophthalmology, Genetics, Maternal and
40 Child Health, University of Genoa, 16132 Genoa, Italy

41 ¹⁷ Pediatric Neurology and Muscular Diseases Unit, IRCCS Istituto Giannina Gaslini, 16147
42 Genoa, Italy

43 ¹⁸ Department of Genetics and Cell Biology, Faculty of Health Medicine Life Sciences, Maas-
44 tricht University Medical Center⁺, Maastricht University, Maastricht, The Netherlands

45 ¹⁹ Department of Neurology, University of California, San Francisco, San Francisco, CA, USA.

46 ²⁰ Institute of Human Genetics and Weill Institute for Neurosciences, University of California,
47 San Francisco, San Francisco, CA, USA

48 ²¹ UF de Génomique du Développement, Département de Génétique, Groupe Hospitalier Pi-
49 tié-Salpêtrière, APHP-Sorbonne Université, Paris, France.

50 ²² APHP, Sorbonne Université, Département de Génétique, Centre de Référence Déficiences
51 Intellectuelles de Causes Rares, Groupe Hospitalier Pitié-Salpêtrière and Hôpital Trousseau,
52 Paris, France

53 ²³ Centre National de la Recherche Scientifique, UMR7104, Illkirch 67400, France

54 ²⁴ Institut National de la Santé et de la Recherche Médicale, U964, Illkirch 67400, France

55 ²⁵ Université de Strasbourg, Illkirch 67400, France

56 ²⁶ School of Biological Sciences, School of Medicine and Robinson Research Institute, The
57 University of Adelaide, Adelaide, 5005, SA, Australia

58 ²⁷ South Australian Health and Medical Research Institute, The University of Adelaide, Ade-
59 laide, 5005, SA, Australia

60 ²⁸ Research Group Development and Disease, Max Planck Institute for Molecular Genetics,
61 Berlin, Germany.

62 * Correspondence to: Christel Depienne (christel.depienne@uni-due.de)

63 **Abstract**

64 Disease gene discovery on chromosome (chr) X is challenging owing to its unique modes of
65 inheritance. We undertook a systematic analysis of human chrX genes. We observe a higher
66 proportion of disorder-associated genes and an enrichment of genes involved in cognition,
67 language, and seizures on chrX compared to autosomes. We analyze gene constraints, exon
68 and promoter conservation, expression and paralogues, and report 127 genes sharing one or
69 more attributes with known chrX disorder genes. Using a neural network trained to distin-
70 guish disease-associated from dispensable genes, we classify 235 genes, including 121 of the
71 127, as having high probability of being disease-associated. We provide evidence of an ex-
72 cess of variants in predicted genes in existing databases. Finally, we report damaging variants
73 in *CDK16* and *TRPC5* in patients with intellectual disability or autism spectrum disorders.
74 This study predicts large-scale gene-disease associations that could be used for prioritization
75 of X-linked pathogenic variants.

76 Introduction

77 Sex in mammals is determined by a diverging pair of sex chromosomes (chr). Human fe-
78 males have two copies of the 155-Mb chrX while males have a single X copy and a smaller
79 30-Mb chrY. Compensation of gene dosage in females is achieved through X chromosome
80 inactivation (XCI), a process leading to the epigenetic silencing of an entire chrX, apart from
81 two pseudoautosomal regions (PARs). This process happens during early embryonic develop-
82 ment, is randomly and independently established in each cell, and stably maintained during
83 further cell divisions.^{1,2} As a consequence, female individuals are cell mosaics, each cell ex-
84 pressing genes from either the maternal or paternal X copy.³ A subset of genes, which can be
85 variable between individuals and tissues, escapes X inactivation and continues to be ex-
86 pressed from both X chromosomes.⁴

87

88 The last two decades have revolutionized concepts of X-linked inheritance, by depict-
89 ing its unique but multiple forms.^{1,5} The first and most widely described disorders (>100
90 genes) mainly affect hemizygous males and are transmitted through healthy or mildly symp-
91 tomatic female carriers. Other modes of X-linked inheritance are mainly observed in disor-
92 ders affecting the central nervous system. Variants in X-linked genes such as *MECP2*⁶,
93 *CDKL5*^{7,8} and *DDX3X*⁹ preferentially affect heterozygous females. These variants usually oc-
94 cur *de novo*, and hemizygous males are either not viable or survive only if variants are hypo-
95 morphic or mosaic. Variants in other genes affect hemizygous males and heterozygous fe-
96 males almost equally. The list of X-linked disorders first described as selectively affecting
97 males but turning out to also affect females is continually increasing and include *IQSEC2*¹⁰,
98 *NEXMIF*,¹¹ *KDM5C*,¹² *HUWE1*,¹³ *USP9X*¹⁴ and *CLCN4*.¹⁵ Lastly, two X-linked disorders,
99 related to *PCDH19* and *EFNB1*, affect heterozygous females and postzygotic somatic mosaic
100 males (due to cellular interference), while hemizygous males are spared.^{16,17}

101

102 Disease gene discovery on chrX is thus associated with greater challenges, including
103 male-female patient selection and variant interpretation biases, compared to autosomes. ChrX
104 is often omitted from genome-wide analyses in a research context due to the difficulty of
105 dealing with sex dichotomy in bioinformatics pipelines. Identifying novel gene-disease asso-
106 ciations on chromosome X requires dedicated studies of families with multiple affected
107 males^{18,19} or multiple subjects with matching phenotypes.⁹ The interpretation of X-linked var-
108 iants in sporadic cases and small families remains difficult in absence of extended segrega-
109 tion in the family, which is rarely available. Furthermore, the presence of damaging variants
110 at relatively high frequency in large databases (ExAC, gnomAD) led to question previously
111 established gene-disease associations.²⁰ The Deciphering Developmental Disorders (DDD)
112 consortium recently estimated that X-linked disorders overall affect males and females
113 equally and represent 6% of developmental disorders.²¹ However, despite the remarkable size
114 of the cohort (11,044 affected individuals), this study failed to identify new X-linked disor-
115 der-associated genes.

116

117 In this study, we first undertook a systematic analysis of all coding genes on human
118 chromosome X and compared the proportion and characteristics of associated disorders to
119 those on autosomes. In a second step, we investigated the relevance of multiple variables to
120 predict gene-disorder associations. Lastly, we used these predictions to uncover new disease-
121 gene associations supported by the literature as well as by patient data and functional studies.

122 **Results**

123 **Chromosome X is enriched in disorder genes and in genes relevant to brain function**

124 Chromosome X comprises 829 protein coding genes annotated in HUGO Gene Nomenclature
125 Committee (HGNC), including 205 associated with at least one monogenic disorder (referred
126 to as ‘disorder genes’) in OMIM (Fig. 1a). We used the clinical synopsis to compare the pro-
127 portion of disorder genes and their associated clinical features on chrX (available for 202
128 genes) and autosomes (Supplementary Methods, Fig. 1b). We observed a significant and spe-
129 cific enrichment in disorder genes on chrX (24% versus 12-22%, $p=1.87\times 10^{-3}$; OR=1.5; Fig.
130 1c-d and Supplementary Tables 1-4). Furthermore, genes on chrX were significantly more
131 frequently associated with neurological phenotypes than genes on autosomes (77% versus 55-
132 76%; $p=7.01\times 10^{-3}$; OR=2.0; Fig. 1e-f). More specifically, we observed that chrX is enriched
133 in genes associated with intellectual disability (ID; 58% versus 27-45%; $p=5.92\times 10^{-11}$;
134 OR=2.9), seizures (46% versus 23-38%; $p=1.12\times 10^{-4}$; OR=2.1) and language impairment
135 (32% versus 11-24%; $p=5.59\times 10^{-3}$; OR=2.0; Fig. 1g-l), but not motor development, spastic-
136 ity, or ataxia (Fig. 1m-n; Extended Data Fig. 1a-b; Supplementary Table 5). The difference
137 remained significant when genes associated with provisional gene-phenotype relationships
138 (P), susceptibility to multifactorial disorders (M) or traits (T) (referred to as ‘PMT’ genes)
139 were included in the comparison (Extended Data Fig. 1c-d; Supplementary Tables 4 and 5).
140 In total, 84% of known disease genes on chrX are associated with ID, seizures or language
141 impairment and about 30% are associated with all three clinical outcomes (Fig. 1o).

142

143 **Confirmed disorder-associated genes share specific features**

144 Despite chrX being enriched in disorder-associated genes, 598 genes (71%) are not yet re-
145 lated to any clinical phenotype (referred to as ‘no-disorder genes’; Fig. 1a). We hypothesized
146 that disorder genes share specific common features that dispensable genes do not exhibit and

147 that could be used to predict genes that remain to be associated with human disorders. To test
148 this hypothesis, we retrieved annotations from different sources and/or calculated additional
149 metrics, including: 1) gnomAD gene constraint metrics: LOEUF (rank of intolerance to loss-
150 of-function (LoF) variants), misZ (intolerance to missense variants score) and synZ (intoler-
151 ance to synonymous variants score); 2) coding-sequence (CDS) length; 3) the degree of ex-
152 ons and promoter conservation across 100 species; 4) the promoter CpG density;²² and 5)
153 gene expression data, conveyed as a tissue specificity measure (tau) and brain-related expres-
154 sion levels from GTEx and BrainSpan resources ([Supplementary Table 1](#)).

155

156 We investigated whether the distribution of these variables differ between disorder
157 genes and no-disorder genes. As expected, disorder genes had lower LOEUF values and
158 higher misZ scores but similar synZ scores than no-disorder genes. We observed a significant
159 enrichment of disorder genes in the three lowest LOEUF deciles and in the two highest misZ
160 deciles, with the distributions of LOEUF and misZ metrics in the PMT group showing inter-
161 mediate values between disorder and no-disorder genes ([Fig. 2a-b](#)). Disorder genes were en-
162 riched in the highest decile of exon conservation, promoter conservation scores and CDS
163 length distribution with an overall distribution that was variable in all groups ([Fig. 2c](#); [Ex-
164 tended Data Fig. 2](#); [Supplementary Table 6](#)). Disorder genes tended to present higher pro-
165 moter CpG density than no-disorder genes, although the difference was not significant. Re-
166 markably, we observed that disorder genes are more broadly expressed ($\tau < 0.6$) and show
167 intermediary levels of gene expression in brain-tissues compared to no-disorder genes, for
168 which expression was more often restricted to a few tissues ($\tau > 0.6$) and are less expressed
169 in the brain ([Extended Data Fig. 2](#)).

170

171 We next investigated how the existence of close paralogues and the location in PARs
172 may influence their relationship to human disease. From the 18 PAR protein-coding genes,
173 only two (*SHOX* and *CSF2RA*) have been associated with a medical condition so far. We also
174 observed that disorder genes are significantly depleted in close paralogues compared to
175 PMT/no-disorder genes ($p=3.1\times 10^{-6}$; OR = 0.33; [Extended Data Fig. 3a-d](#)). This suggests that
176 highly similar paralogues might compensate pathogenic variants in corresponding X-linked
177 genes.

178

179 Previous comparisons suggest that LOEUF, misZ, and exon conservation scores,
180 which showed marked differences between the two subgroups, can best differentiate disorder
181 from no-disorder genes. We then focused on the deciles enriched in disorder genes to list the
182 genes in the no-disorder group exhibiting similar characteristics ([Fig. 2d](#)). More specifically,
183 we defined thresholds for LOEUF, misZ, and exon conservation scores as follows: LOEUF \leq
184 0.326 (L), misZ \geq 2.16 (M), and/or exon-score \geq 0.9491 (E). 149 of the 205 (73%) disorder
185 genes met one or more L, M or E (LME) criteria ([Fig. 2e](#)). Among the 205 disorder genes, 56
186 genes failing to meet any of the LOEUF/misZ criteria had a smaller CDS length ($p=4.20\times 10^{-9}$),
187 indicating that the performance of LOEUF and misZ in differentiating disorder from no-
188 disorder genes depends on CDS length ([Extended Data Fig. 3e](#)). Exon conservation alone al-
189 lowed retrieving eight disorder genes with small CDS for which LOEUF and misZ failed to
190 reach the thresholds ([Extended Data Fig. 3e; Supplementary Table 1](#)). When applying the
191 same thresholds to the 598 genes not associated with disease, 127 (21%) genes fulfilled at
192 least one condition and 35 fulfilled at least two criteria ([Fig. 2e; Supplementary Table 1](#)).
193 Half (n=13) of the PMT genes also met at least one criterion and seven at least two criteria
194 ([Fig. 2e](#)). Altogether, 140 genes shared at least one of the LME criteria with known disorder
195 genes.

196

197 **Prediction of novel disorder-associated genes on chrX**

198 The threshold approach was limited by the number of variables that could be taken into ac-
199 count to differentiate disorder from no-disorder genes. We then used machine learning to pre-
200 dict remaining disorder genes in a more systematic and unbiased fashion. As most X-linked
201 disorders are associated with neurological features, we trained a neural network (NN) to dis-
202 tinguish two classes of genes: 1) known disorder genes associated with neurological features
203 on all chromosomes and 2) dispensable genes as defined by Karczewski et al. (2020)²³ (Sup-
204 plementary Methods; [Extended Data Fig. 4](#); [Extended Data Fig. 5a-b](#)). We used additional
205 variables including expression data stratified according to sex and data on paralogues ([Sup-](#)
206 [plementary Tables 7 and 8](#)). Although using data on all chromosomes, the neural network
207 performed remarkably better for genes on chrX, predicting 191 (93%) of the 205 known dis-
208 order genes on this chromosome using $FDR < 0.05$ ([Extended Data Fig. 5c-f](#)). 145 of the pre-
209 dicted genes met the LME thresholds ([Fig. 2f](#)). The NN confirmed 16 of the 26 PMT genes as
210 disease-associated, 13 of which also met LME criteria ([Fig. 2f](#), [Fig. 3a-b](#)). Focusing on the
211 598 no-disorder genes on chrX, the NN predicted 235 as disease-associated ($FDR < 0.05$;
212 [Supplementary Table 9](#); [Extended data Fig. 5a](#)). 121 genes (51%) also met at least one of the
213 LME criteria ([Fig. 2f](#)), with four meeting all three, 30 two and 87 genes one criterion ([Fig.](#)
214 [3c](#)). The 114 remaining genes predicted by the NN did not meet any LME criteria ([Fig. 3d](#)),
215 whereas only six meeting LME criteria were not predicted by the NN ([Fig. 3e](#)). Altogether,
216 these results suggest that the NN is a valid approach to predict genes remaining to be associ-
217 ated with disease with high accuracy.

218

219 **Validation of putative novel disorder genes**

220 To provide additional evidence supporting our predictions, we searched for evidence from
221 databases or the literature that variants in the 235 NN-predicted genes could indeed result in
222 new X-linked disorders. We first used HGMD and DECIPHER to retrieve the number of sin-
223 gle nucleotide variants and indels of unknown significance reported in each gene and com-
224 pared the number of variants in predicted versus non-predicted gene categories ([Supplemen-
225 tary Table 9](#)). This analysis revealed that the 235 NN-predicted disorder genes are enriched in
226 point variants compared to non-predicted genes ($p=5.3\times 10^{-6}$; OR=2.2), suggesting that this
227 excess is due to pathogenic variants. NN-predicted genes meeting at least one LME criterion
228 had more point variants than predicted genes that do not meet any LME criterion ([Fig. 4a-b](#);
229 [Extended Data Fig. 6](#)).

230

231 Second, we used the SysID and Gene2Phenotype (G2P) developmental disorders da-
232 tabases to compare the overlap of predicted and non-predicted genes ([Fig. 4c](#)). Nine of the 16
233 NN-predicted PMT genes were in SysID ([Fig. 3a-b](#)). Three genes (*PTCHD1*, *NLGN3*, and
234 *KIF4A*), indicated as confirmed SysID genes, were also in G2P. *PTCHD1* and *NLGN3* are as-
235 sociated with susceptibility to autism in OMIM, but both were recently confirmed to cause a
236 monogenic ID disorder frequently associated with autism.²⁴⁻²⁶ A splice site variant in *KIF4A*
237 was identified in a family with four affected males,²⁷ and this finding was recently strength-
238 ened by the identification of additional *de novo* or inherited variants causing a range of differ-
239 ent phenotypic manifestations.^{28,29} Focusing on no-disorder genes, 21 (9%) of the NN-pre-
240 dicted genes and 14 (4%) of the non-predicted genes were in SysID ([Fig. 3c-d](#)). Five pre-
241 dicted genes (*CSTF2*, *FGF13*, *PLXNA3*, *OTUD5* and *STEEP1*) were confirmed SysID genes.
242 In addition, six genes (*ARHGEF6*, *CDK16*, *FGF13*, *GSPT2*, *MMGT1* and *ZDHHC15*) had
243 entries in G2P. Pathogenic variants in *OTUD5* and *FGF13* respectively cause a severe neuro-

244 developmental disorder with multiple congenital anomalies and early lethality,^{30,31} and devel-
245 opmental and epileptic encephalopathy,³² both described in 2021. *PLXNA3* have been associ-
246 ated with ID and autism spectrum disorders (ASD),³³ or hypogonadotropic hypogonadism.³⁴
247 Additional evidence from the literature not yet reflected in OMIM or any other database in-
248 clude possibly pathogenic variants described in one or a few families in *ZMYM3*,³⁵ *GLRA2*,³⁶
249 and *GPKOW*.³⁷ Although some of these genes now have an OMIM disorder entry, their asso-
250 ciation with a monogenic disorder was not known at the beginning of our study.

251

252 Third, we retrieved the number of *de novo* predicted damaging (truncating or
253 CADD \geq 25) variants identified in these genes from the Martin et al.²¹ and Kaplanis et al.³⁸
254 studies. In parallel, we examined exome data from two additional cohorts of patients with de-
255 velopmental disorders, containing mainly patients with intellectual disability (6500 from
256 2346 families [Paris-APHP cohort] and 1399 individuals from 463 families [UCSF cohort])
257 and extracted predicted damaging variants in NN-predicted genes meeting at least one LME
258 criterion ([Supplementary Table 10](#)). This led us to select 13 genes containing *de novo* damag-
259 ing variants and additional predicted damaging variants identified in independent cohorts
260 ([Fig. 5](#)). Interestingly, 19 out the 55 missense variants reported in these 13 genes are localized
261 in known functional protein domains ([Fig. 5](#); [Supplementary Table 11](#)). Among those genes,
262 *SMARCA1* has been associated with syndromic intellectual disability and Coffin-Siris-like
263 features by the Clinical Genome Resource (ClinGen) with moderate evidence.³⁹

264

265 ***CDK16* and *TRPC5* are novel genes associated with intellectual disability**

266 Lastly, we focused on *CDK16* and *TRPC5*, both predicted by the NN and threshold ap-
267 proaches but for which genetic evidence was limited ([Fig. 6a-b](#)). *CDK16* encodes a protein
268 kinase involved in neurite outgrowth, vesicle trafficking and cell proliferation.⁴⁰ A deletion of

269 two nucleotides leading to a frameshift (NM_006201.5: c.976_977del, p.(Trp326Valfs*5))
270 segregating in four males with ID, ASD, absence seizures and mild spasticity was reported in
271 this gene by Hu et al.¹⁸ Using exome sequencing, we identified a nonsense variant
272 (c.961G>T, p.(Glu321*)) in an adult patient with ID and spasticity. A missense variant
273 (c.1039G>T, p.(Gly347Cys)) affecting a highly conserved amino acid of the kinase domain
274 (CADD PHRED score: 32) was identified by genome sequencing in a male patient with ID,
275 ASD and epilepsy, whose family history was compatible with X-linked inheritance (Fig. 6a
276 and c; Supplementary Table 12). In addition, a nonsense variant (c.46C>T, p.(Arg16*)) was
277 recently reported in a patient with ASD by Satterstrom et al.⁴¹

278

279 *TRPC5* encodes the short transient receptor potential channel 5, a channel permeable
280 to calcium predominantly expressed in the brain.⁴² We identified a missense variant in this
281 gene (NM_012471.2:c.523C>T, p.(Arg175Cys), CADD PHRED score: 29.8) in three male
282 siblings with ID and ASD. The variant was maternally inherited but absent from the maternal
283 grandparents (Fig. 6b and d; Supplementary Table 12). We investigated the functional impact
284 of the p.(Arg175Cys) variant on the TRPC5 channel using whole-cell patch-clamp. Immedi-
285 ately after break-in, HEK293 cells expressing mutant TRPC5 exhibited an increase in imme-
286 diate current recorded compared with cell expressing wild-type TRPC5 (p= 0.009 for out-
287 ward current at -80mV and p=0.002 for inward current at -80mV; Fig.6 e-g), suggesting a
288 constitutively active current. A nonsense variant (c.965G>A, p.(Trp322*)) was identified by
289 exome sequencing in a patient with high-functioning ASD. An intragenic deletion of the first
290 coding exon of *TRPC5*, encoding conserved ankyrin repeats, was previously reported in a pa-
291 tient with ASD who had a family history compatible with X-linked inheritance.⁴³ In addition,
292 three *de novo* variants in *TRPC5* (p.(Pro667Thr), p.(Arg71Gln), p.(Trp225*)) had been iden-
293 tified in patients with intellectual disability and/or autism disorders in the literature.^{21,44,45}

294

295 Altogether, these results strongly suggest that pathogenic variants altering the func-
296 tions of *CDK16* and *TRPC5* lead to novel X-linked disorders featuring ID and ASD.

297

298 **Discussion**

299 Male and female individuals tolerate pathogenic variants on chrX in different manners. Vari-
300 ants in X-linked genes have to be interpreted taking this complexity into account. We aimed
301 at providing an inventory of disorder genes on chrX and predict genes that remain to be asso-
302 ciated with human disease, assuming that they share similar characteristics. We first used a
303 threshold approach to list genes similarly constrained during evolution, which are the most
304 likely to lead to disease when altered by genetic variants. This revealed 127 novel genes shar-
305 ing at least one constraint metric with known disorder genes and 35 sharing at least two. To
306 avoid bias and limitations linked to this approach, we used a hypothesis-free neural network
307 to differentiate genes associated with brain disorders from genes where damaging variants are
308 tolerated (dispensable genes). The NN predicted 235 genes as putative disorder genes, includ-
309 ing most of the genes uncovered by the threshold approach.

310

311 Our predictions are supported by a higher number of point variants reported in DECI-
312 PHER and HGMD in predicted versus non-predicted genes, which strengthens the probability
313 that variants reported contribute to the patients' phenotypes and prompts the prioritization of
314 these genes in further genetic analyses. We notably highlight 13 genes in which several possi-
315 bly pathogenic variants in functional domains have been identified in patients, likely consti-
316 tuting novel neurodevelopmental disorder genes. Furthermore, a subset of predicted genes

317 (e.g. *OTUD5*, *FGF13*, *GLRA2* and *GPKOW*) have already been associated with human dis-
318 eases, although these associations had no OMIM Gene-Phenotype Relationship entries when
319 we started our study. Furthermore, we provide additional evidence that variants in two pre-
320 dicted genes, *CDK16* and *TRPC5*, likely cause novel X-linked disorders by gathering genetic
321 and clinical data of unrelated families with damaging variants in these genes.

322

323 Altogether, our gene-focused approach therefore suggests that less than half of genes
324 associated with human pathology on chrX are known so far and many more remain to be
325 characterized clinically, which challenges the recent estimation that X-chromosome has been
326 saturated for disease genes.²¹ The differences in conclusions may come from recruitment bias
327 in the DDD study, with inclusion of families with X-linked inheritance in more specific stud-
328 ies, as well as from the patient-driven approach itself, which only allows detecting what is
329 statistically significant in a sample, no matter how large this sample is. It is indeed possible or
330 even likely that phenotypes associated with genes we predict are ultra-rare, lethal *in utero*,
331 difficult to recognize or assess clinically, or associated with atypical modes of X-linked in-
332 heritance, making their identification difficult using classical genetic approaches.

333

334 Focusing on known genes associated with monogenic conditions in the first part of
335 this study, we observed a higher proportion of genes associated with disorders on chrX com-
336 pared to autosomes. This finding suggests that family-driven approaches have been more effi-
337 cient in identifying disorder-associated genes on this chromosome. Nevertheless, considering
338 disease-causing genes independently of their proportion, we also observed an enrichment of
339 genes associated with specific neurological features such as ID, seizures, and language delay
340 on chrX. The specific association of X-linked genes with cognitive functions has largely been

341 conveyed in the literature but, to our knowledge, this study is the first demonstrating the sta-
342 tistical significance of this finding using a systematic unbiased analysis. The reason why so
343 many genes important for cognition and language are on a sex chromosome is fascinating but
344 remains so far mysterious.⁴⁶

345

346 Our study also indicates that disorder genes on chrX are depleted in highly similar pa-
347 ralogues (>95% identity) compared to PMT/no-disorder genes, a finding that remained sig-
348 nificant for predicted genes compared to non-predicted ones ($p < 2.2 \times 10^{-16}$; OR= 0.22). The
349 consequence of the existence of paralogues for disorders is probably different depending on
350 the expression of these paralogues and the ability of the gene product to compensate the func-
351 tion of the original genes. Interestingly, 20% of genes on chrX have paralogues with >95%
352 identity. This list includes copies or retrocopies on autosomes, a functional redundancy that
353 has been attributed at least partly to the transcriptional silencing of chromosome X during
354 spermatogenesis.^{47,48} Although many gene copies or retrocopies are specifically expressed in
355 male germ cells, others are still broadly expressed and could therefore buffer genetic variants
356 in the X-linked paralogous genes, as shown for *UPF3B* and *UPF3A*.⁴⁹ The existence of pa-
357 ralogues is also linked in some ways to the location in PAR regions and escape to XCI. In-
358 deed, genes in PAR have paralogues on chrY and escape inactivation.⁵⁰ Accordingly, 35 of
359 the 231 predicted genes outside PAR also show some degree of escape from X inactivation
360 (Supplementary Table 1). Interestingly, genes escaping XCI outside PAR, including *IQSEC2*
361 and *KDM5C*, may lead to disorders manifesting in both sexes.^{10,12} Our study indicates that
362 genes in PAR and genes with close paralogues are less likely to be associated with a disorder,
363 suggesting that paralogues can compensate pathogenic variants in some X-linked genes and
364 raising the possibility of digenism or oligogenism in genes with redundant functions.

365

366 Our study focused on coding genes and did not include genes encoding long non-cod-
367 ing RNA (lncRNA) or other RNA classes, which are however very abundant on chrX. Only a
368 few non-coding genes have been associated with disease so far and we believe that con-
369 straints and pathological mechanisms applying to non-coding genes are likely different from
370 those of coding genes. In this respect, our study only predicts genes associated with disorders
371 when affected by usual mutation types. Therefore, our findings do not exclude that non-pre-
372 dicted genes could lead to disease when associated with unusual mechanisms, such as gain of
373 a new function, dominant-negative impact on another gene, and ectopic expression of a gene
374 in the wrong tissue or at an abnormal time during development.

375

376 In conclusion, our study provides new insights into the complexity of X-linked disor-
377 ders and indicates that alternative approaches not initially based on patient cohorts are effec-
378 tive to reveal gene-disease associations. We provide a list of genes that are likely to be asso-
379 ciated with human disorders. Further studies are required to delineate these disorders clini-
380 cally and determine whether males and/or females harboring variants in these genes are af-
381 fected.

382 References

- 383 1. Balaton, B.P., Dixon-McDougall, T., Peeters, S.B. & Brown, C.J. The eXceptional
384 nature of the X chromosome. *Hum Mol Genet* **27**, R242-R249 (2018).
- 385 2. Galupa, R. & Heard, E. X-Chromosome Inactivation: A Crossroads Between
386 Chromosome Architecture and Gene Regulation. *Annu Rev Genet* **52**, 535-566 (2018).
- 387 3. Orstavik, K.H. X chromosome inactivation in clinical practice. *Hum Genet* **126**, 363-
388 73 (2009).
- 389 4. Tukiainen, T. *et al.* Landscape of X chromosome inactivation across human tissues.
390 *Nature* **550**, 244-248 (2017).
- 391 5. Migeon, B.R. X-linked diseases: susceptible females. *Genet Med* **22**, 1156-1174
392 (2020).
- 393 6. Kaur, S. & Christodoulou, J. MECP2 Disorders. in *GeneReviews((R))* (eds. Adam,
394 M.P. *et al.*) (Seattle (WA), 1993).
- 395 7. Weaving, L.S. *et al.* Mutations of CDKL5 cause a severe neurodevelopmental
396 disorder with infantile spasms and mental retardation. *Am J Hum Genet* **75**, 1079-93
397 (2004).
- 398 8. Tao, J. *et al.* Mutations in the X-linked cyclin-dependent kinase-like 5
399 (CDKL5/STK9) gene are associated with severe neurodevelopmental retardation. *Am*
400 *J Hum Genet* **75**, 1149-54 (2004).
- 401 9. Snijders Blok, L. *et al.* Mutations in DDX3X Are a Common Cause of Unexplained
402 Intellectual Disability with Gender-Specific Effects on Wnt Signaling. *Am J Hum*
403 *Genet* **97**, 343-52 (2015).
- 404 10. Mignot, C. *et al.* IQSEC2-related encephalopathy in males and females: a
405 comparative study including 37 novel patients. *Genet Med* **21**, 837-849 (2019).
- 406 11. Stamberger, H. *et al.* NEXMIF encephalopathy: an X-linked disorder with male and
407 female phenotypic patterns. *Genet Med* **23**, 363-373 (2021).
- 408 12. Carmignac, V. *et al.* Further delineation of the female phenotype with KDM5C
409 disease causing variants: 19 new individuals and review of the literature. *Clin Genet*
410 **98**, 43-55 (2020).
- 411 13. Moortgat, S. *et al.* HUWE1 variants cause dominant X-linked intellectual disability: a
412 clinical study of 21 patients. *Eur J Hum Genet* **26**, 64-74 (2018).
- 413 14. Jolly, L.A. *et al.* Missense variant contribution to USP9X-female syndrome. *NPJ*
414 *Genom Med* **5**, 53 (2020).
- 415 15. Palmer, E.E. *et al.* De novo and inherited mutations in the X-linked gene CLCN4 are
416 associated with syndromic intellectual disability and behavior and seizure disorders in
417 males and females. *Mol Psychiatry* **23**, 222-230 (2018).
- 418 16. Depienne, C. & LeGuern, E. PCDH19-related infantile epileptic encephalopathy: an
419 unusual X-linked inheritance disorder. *Hum Mutat* **33**, 627-34 (2012).
- 420 17. Wieland, I. *et al.* Dissecting the molecular mechanisms in craniofrontonasal
421 syndrome: differential mRNA expression of mutant EFNB1 and the cellular mosaic.
422 *Eur J Hum Genet* **16**, 184-91 (2008).
- 423 18. Hu, H. *et al.* X-exome sequencing of 405 unresolved families identifies seven novel
424 intellectual disability genes. *Mol Psychiatry* **21**, 133-48 (2016).

- 425 19. Tarpey, P.S. *et al.* A systematic, large-scale resequencing screen of X-chromosome
426 coding exons in mental retardation. *Nat Genet* **41**, 535-43 (2009).
- 427 20. Piton, A., Redin, C. & Mandel, J.L. XLID-causing mutations and associated genes
428 challenged in light of data from large-scale human exome sequencing. *Am J Hum*
429 *Genet* **93**, 368-83 (2013).
- 430 21. Martin, H.C. *et al.* The contribution of X-linked coding variation to severe
431 developmental disorders. *Nat Commun* **12**, 627 (2021).
- 432 22. Boukas, L., Bjornsson, H.T. & Hansen, K.D. Promoter CpG Density Predicts
433 Downstream Gene Loss-of-Function Intolerance. *Am J Hum Genet* **107**, 487-498
434 (2020).
- 435 23. Karczewski, K.J. *et al.* The mutational constraint spectrum quantified from variation
436 in 141,456 humans. *Nature* **581**, 434-443 (2020).
- 437 24. Quartier, A. *et al.* Novel mutations in NLGN3 causing autism spectrum disorder and
438 cognitive impairment. *Hum Mutat* **40**, 2021-2032 (2019).
- 439 25. Chaudhry, A. *et al.* Phenotypic spectrum associated with PTCHD1 deletions and
440 truncating mutations includes intellectual disability and autism spectrum disorder.
441 *Clin Genet* **88**, 224-33 (2015).
- 442 26. Halewa, J. *et al.* Novel missense mutations in PTCHD1 alter its plasma membrane
443 subcellular localization and cause intellectual disability and autism spectrum disorder.
444 *Hum Mutat* **42**, 848-861 (2021).
- 445 27. Willemsen, M.H. *et al.* Involvement of the kinesin family members KIF4A and
446 KIF5C in intellectual disability and synaptic function. *J Med Genet* **51**, 487-94
447 (2014).
- 448 28. Kalantari, S. *et al.* Expanding the KIF4A-associated phenotype. *Am J Med Genet A*
449 **185**, 3728-3739 (2021).
- 450 29. Gowans, L.J.J. *et al.* Missense Pathogenic variants in KIF4A Affect Dental
451 Morphogenesis Resulting in X-linked Taurodontism, Microdontia and Dens-
452 Invaginatus. *Front Genet* **10**, 800 (2019).
- 453 30. Beck, D.B. *et al.* Linkage-specific deubiquitylation by OTUD5 defines an embryonic
454 pathway intolerant to genomic variation. *Sci Adv* **7**(2021).
- 455 31. Tripolszki, K. *et al.* An X-linked syndrome with severe neurodevelopmental delay,
456 hydrocephalus, and early lethality caused by a missense variation in the OTUD5 gene.
457 *Clin Genet* **99**, 303-308 (2021).
- 458 32. Fry, A.E. *et al.* Missense variants in the N-terminal domain of the A isoform of
459 FHF2/FGF13 cause an X-linked developmental and epileptic encephalopathy. *Am J*
460 *Hum Genet* **108**, 176-185 (2021).
- 461 33. Steele, J.L. *et al.* Semaphorin-Plexin Signaling: From Axonal Guidance to a New X-
462 Linked Intellectual Disability Syndrome. *Pediatr Neurol* **126**, 65-73 (2022).
- 463 34. Kotan, L.D. *et al.* Loss-of-function variants in SEMA3F and PLXNA3 encoding
464 semaphorin-3F and its receptor plexin-A3 respectively cause idiopathic
465 hypogonadotropic hypogonadism. *Genet Med* **23**, 1008-1016 (2021).
- 466 35. Philips, A.K. *et al.* X-exome sequencing in Finnish families with intellectual
467 disability--four novel mutations and two novel syndromic phenotypes. *Orphanet J*
468 *Rare Dis* **9**, 49 (2014).

- 469 36. Pilorge, M. *et al.* Genetic and functional analyses demonstrate a role for abnormal
470 glycinergic signaling in autism. *Mol Psychiatry* **21**, 936-45 (2016).
- 471 37. Carroll, R. *et al.* Variant in the X-chromosome spliceosomal gene GPKOW causes
472 male-lethal microcephaly with intrauterine growth restriction. *Eur J Hum Genet* **25**,
473 1078-1082 (2017).
- 474 38. Kaplanis, J. *et al.* Evidence for 28 genetic disorders discovered by combining
475 healthcare and research data. *Nature* **586**, 757-762 (2020).
- 476 39. Strande, N.T. *et al.* Evaluating the Clinical Validity of Gene-Disease Associations:
477 An Evidence-Based Framework Developed by the Clinical Genome Resource. *Am J*
478 *Hum Genet* **100**, 895-906 (2017).
- 479 40. Gillani, S.Q. *et al.* Regulation of PCTAIRE1 protein stability by AKT1, LKB1 and
480 BRCA1. *Cell Signal* **85**, 110032 (2021).
- 481 41. Satterstrom, F.K. *et al.* Large-Scale Exome Sequencing Study Implicates Both
482 Developmental and Functional Changes in the Neurobiology of Autism. *Cell* **180**,
483 568-584 e23 (2020).
- 484 42. Beck, A. *et al.* Conserved gating elements in TRPC4 and TRPC5 channels. *J Biol*
485 *Chem* **288**, 19471-83 (2013).
- 486 43. Mignon-Ravix, C. *et al.* Intragenic rearrangements in X-linked intellectual deficiency:
487 results of a-CGH in a series of 54 patients and identification of TRPC5 and KLHL15
488 as potential XLID genes. *Am J Med Genet A* **164A**, 1991-7 (2014).
- 489 44. Kosmicki, J.A. *et al.* Refining the role of de novo protein-truncating variants in
490 neurodevelopmental disorders by using population reference samples. *Nat Genet* **49**,
491 504-510 (2017).
- 492 45. de Ligt, J. *et al.* Diagnostic exome sequencing in persons with severe intellectual
493 disability. *N Engl J Med* **367**, 1921-9 (2012).
- 494 46. Graves, J.A., Gecz, J. & Hameister, H. Evolution of the human X--a smart and sexy
495 chromosome that controls speciation and development. *Cytogenet Genome Res* **99**,
496 141-5 (2002).
- 497 47. Potrzebowski, L. *et al.* Chromosomal gene movements reflect the recent origin and
498 biology of therian sex chromosomes. *PLoS Biol* **6**, e80 (2008).
- 499 48. Marques, A.C., Dupanloup, I., Vinckenbosch, N., Reymond, A. & Kaessmann, H.
500 Emergence of young human genes after a burst of retroposition in primates. *PLoS Biol*
501 **3**, e357 (2005).
- 502 49. Nguyen, L.S. *et al.* Transcriptome profiling of UPF3B/NMD-deficient
503 lymphoblastoid cells from patients with various forms of intellectual disability. *Mol*
504 *Psychiatry* **17**, 1103-15 (2012).
- 505 50. Raudsepp, T. & Chowdhary, B.P. The Eutherian Pseudoautosomal Region. *Cytogenet*
506 *Genome Res* **147**, 81-94 (2015).

507 **Supplementary Methods**

508 **Protein-coding genes**

509 Annotations of genes in all chromosomes were downloaded from the HUGO Gene Nomen-
510 clature Committee (HGNC) database in November 2020 focusing on protein-coding genes
511 with approved status and present in the reference assembly.⁵¹ *MEDIA4OS* was excluded from
512 chrX-focused analysis due to its encoded protein being curated in Uniprot as "Product of a
513 dubious CDS prediction". Information concerning genes in PARs was also retrieved from
514 HGNC.

515 **Gene-phenotype relationships**

516 Gene associations with disorders and/or traits were retrieved from data files provided by
517 Online Mendelian Inheritance in Man (OMIM). Information concerning the Clinical Synopsis
518 of phenotypes was obtained through OMIM Search. OMIM morbid genes were annotated for
519 1) the availability of Clinical Synopsis data for any of the associated phenotypes, 2) the exist-
520 ence of Clinical Synopsis neurologic features in any of the associated phenotypes and 3) the
521 presence of terms related to intellectual disability, seizures, impaired language development,
522 impaired motor development, spasticity and ataxia among the Clinical Synopsis neurologic
523 features. For the latter and due to the free-text nature of the Clinical Synopsis data, synony-
524 mous terms and sentences of the above-mentioned clinical features were used and their com-
525 plete lists are shown in [Supplementary Table 2](#). Genes were divided into three groups based
526 on the existence and type of their gene-phenotype relationships: 1) “confirmed” when associ-
527 ated with at least one confirmed monogenic disorder; 2) “PMT” for genes either with provi-
528 sional gene-phenotype relationship (P), or associated with susceptibility factors to multifacto-

529 rial disorders (M) or with traits (T), which are labelled in the OMIM Gene Map with a ques-
530 tion mark, braces or brackets, respectively; 3) “no disorder” for genes showing no phenotype
531 associations. For each chromosome, we calculated the fraction of protein-coding genes that
532 are CS-genes and the fraction of CS-genes that contain non-specific or specific neurologic
533 features in the associated Clinical Synopsis data.

534 **Gene intolerance to variation**

535 Metrics related to intolerance of chrX genes to genetic variation (loss-of-function (LoF), mis-
536 sense and synonymous) were retrieved from the Genome Aggregation Database (gnomAD)
537 v2.1.1²³ after conversion of gene identifiers to HGNC approved symbols. For genes with
538 available data for multiple transcripts, the metrics of the one with lowest LoF observed/ex-
539 pected upper bound fraction (LOEUF) were kept.

540 **Gene expression data**

541 Median gene-level transcripts per million (TPM) for 54 tissues from the v8 release were
542 downloaded from the GTEx portal⁵². The robust tissue specificity measure τ ^{53,54} was cal-
543 culated for chrX genes as previously described, aggregating the median expression of each
544 gene for the different brain regions into two values: i) the median of the values for the two
545 cerebellar regions and ii) the median of the values for the other brain tissues.²² Genes with τ
546 below 0.6 are broadly expressed, while τ higher than 0.6 indicates genes expressed in a re-
547 stricted number of tissues.

548

549 Additional expression data was downloaded from the BrainSpan Atlas of the Devel-
550 oping Human Brain.⁵⁵ The Developmental Transcriptome Dataset contains gene summarized
551 RPKM (Reads Per Kilobase of transcript, per Million mapped reads) normalized expression

552 values generated across 13 developmental stages in 8-16 brain structures from a total of 42
553 individuals of both sexes. After conversion of RPKM to TPM and gene identifiers to HGNC
554 approved symbols, we restricted to genes on chrX and calculated the mean TPM according to
555 age groups, ignoring sex and tissue origin: 1) Pre-natal; 2) Post-natal; 2a) Post-natal 1: after
556 birth until four years-old (inclusive). 2b) Postnatal 2: older than four years old.

557

558 For the neural network, the expression data was stratified by sex: 1) we calculated the
559 mean and variance expression of each gene for females and males independently, for brain,
560 cerebellar tissues and nerve, aggregating in one metavalue the data for the two cerebellar tis-
561 sues and in another the data for the other brain regions; 2) tau values were also calculated in-
562 dependently for females and males; 3) Brainspan mean and variance TPM for each age group
563 was calculated separately for females and males.

564 **Canonical transcript selection**

565 The criteria for canonical transcript selection prioritized: i) MANE (Matched Annotation be-
566 tween NCBI and EBI) Select transcripts, which were independently identified by both En-
567sembl and NCBI as the most biologically relevant; and ii) APPRIS-annotated transcripts⁵⁶.

568 TSS, MANE, and APPRIS annotations of transcripts were retrieved from Ensembl Genes 102
569 via BioMart⁵⁷. For most genes (ca. 80%), the canonical transcripts belong to the MANE Se-
570 lect category. The longest transcript with APPRIS annotation was selected for ca. 20% of
571 genes. In one case, the only existing transcript was kept.

572 **Promoter CpG density**

573 CpG density was calculated for the 4kb region surrounding the transcription start site
574 (TSS±2kb) of the canonical transcripts of chrX protein-coding genes following previous pub-
575 lications.^{22,58} Promoter sequences were downloaded through UCSC Table Browser.⁵⁹ The
576 CpG density, defined as the observed-to-expected CpG ratio, was calculated as follows: num-
577 ber of CpG dinucleotides / (number of C x number of G) × N, where N is the total number of
578 nucleotides in the sequence being analyzed.

579 **Exon and promoter conservation across species**

580 Nucleotide conservation across 100 vertebrate species was calculated using the phastCons
581 score obtained with the phastCons100way.UCSC.hg38 R package,⁶⁰ and represents the prob-
582 ability that a given nucleotide is conserved (range 0 to 1). Exon-conservation score was cal-
583 culated for each gene on chrX as the average phastCons of all nucleotides belonging to the
584 gene coding sequence. Coding coordinates for the canonical transcripts were retrieved from
585 Ensembl Genes 102⁵⁷ using biomaRt R package.^{61,62} Promoter-conservation score for each
586 gene on chrX was calculated as the average phastCons score of all nucleotides 4kb around the
587 TSS of the canonical transcript.

588 **Distance to centromere and telomeres**

589 The coordinates from centromeres and telomeres were downloaded through UCSC Table
590 Browser. We calculated the distance from the TSS of the canonical transcripts to the centro-
591 mere and to the telomere in the corresponding chromosomal arm.

592 **Annotation of encoded proteins**

593 Data referring to function, subcellular location, subunit structure, and gene ontology terms for
594 chrX encoded proteins were retrieved from Uniprot.⁶³

595 **Paralogues**

596 All paralogues from chrX genes were retrieve from Ensembl Genes 102⁵⁷ using biomaRt R
597 package.^{61,62} The 90th, 95th, 98th and 99th and 100th percentiles were calculated for i) the per-
598 centage of paralogous sequence matching the query sequence and ii) the percentage of query
599 sequence matching the paralogue sequence. Only paralogues with both metrics above the 95th
600 percentile were considered as close paralogues.

601 **X-chromosome inactivation**

602 The information on genes escaping X-chromosome inactivation (XCI) was obtained from
603 multiple publications.^{4,64-69} After converting gene identifiers from all studies into HGNC ap-
604 proved symbols, genes were divided into seven categories based on the agreement between
605 the various studies: 1) high confidence escapee and 2) high confidence non-escapee, when al-
606 most all studies agreed on one status; 3) low confidence escapee and 4) low confidence non-
607 escapee, whenever some studies disagreed but a higher number of studies reported one status;
608 5) variable escapee, when most studies stated variable escape; 6) discordant, when similar
609 number of studies agreed on both status; 7) not available, when there was not enough data to
610 have reliable evidence of XCI status.

611 **Features shared by disorder-causing genes**

612 To uncover no-disorder genes exhibiting similar characteristics to known disorder-causing
613 genes, we considered metrics showing enrichment of confirmed-disorder genes at one of the
614 extremes of the distribution with marked difference between confirmed-disorder and no-dis-
615 order genes. Then, we applied a threshold approach to categorize genes showing values
616 within the deciles enriched for confirmed disorder-causing genes for each of the considered

617 metrics. The minimum number of criteria showing enrichment for confirmed disorder-caus-
618 ing genes was used as the minimum number of criteria required to consider a PMT or no-dis-
619 order gene as having similar features to disorder-causing genes.

620 **Pre-classification of genes for the neural network**

621 Genes were pre-classified based on 1) the type of associations with disorders and/or traits
622 (confirmed, PMT, no-disorder), 2) the association with a brain disorder and 3) the tolerance
623 to loss-of-function (LoF) variants ([Extended Data Fig. 5](#)). The list of confidently LoF-tolerant
624 genes based on the gnomAD dataset was retrieved from data files from Karczewski et al.
625 (2020) and consists of genes with at least one homozygous LoF variants.¹⁶ Genes showing no
626 homozygous LoF variants were considered LoF-intolerant.

627 **Neural Network**

628 Information regarding the data fed into the neural network is in Supplementary Tables 7 and
629 8. We used a simple multilayer perceptron (MLP) network with 3 inner layers of size 128
630 nodes, 64 nodes and 32 nodes with rectified linear unit (ReLU) activation function and a sin-
631 gle output node with sigmoid activation. The weights were initialized with a random normal
632 distribution with a bias of zero, and batch normalization and a dropout of 0.5 were applied
633 between all layers. Training was performed with a batch size of 5 and data shuffling between
634 epochs using stochastic gradient descent (SGD) with a learning rate of 1e-3, no decay, and a
635 binary crossentropy / log loss as the objective function.

636 Class labels for the training data T are "Cbi" (C1, value 1.0) and "NDt" (C0, value 0.0) ([Ex-
637 tended Data Fig. 5; Supplementary Table 8](#)). Beforehand columns of the training data were
638 separately normalized to [0;1] via min-max-scaling.

639 To validate the model we performed a 10-fold cross validation by scikit-learn K-Fold⁷⁰,

640 which randomly splits T into ten non overlapping parts T_0 to T_9 each containing 10% of the
641 data (Extended Data Fig. 4). Ten networks are trained. Each network N_i is trained on the nine
642 parts $\{P_j | j \neq i\}$. Afterwards N_i is applied to the remaining unseen part P_i to create estimation
643 E_i .

644 We then concatenate all partial estimations to get full estimation $\cup_i E_i = E$ for each gene in
645 T . Together with the known classes C0 and C1 for each gene, we estimate a threshold t for
646 given false discovery rate d . First define functions $C1(t)$ and $C0(t)$ returning all genes of
647 class C1 respectively C0 with estimation score $< t$. For $t > 0.5$ we can consider a gene with
648 label T0 to be false positive. Thereby for a given t the empirical FDR can be measured by
649 $T1(t)/(T0(t) + T1(t))$. By calculating the FDR threshold by estimations on ten different
650 neural networks, each applied on unseen data, we ensure that the threshold is not biased by
651 overfitting. FDR thresholds were calculated as 0.8152034 (FDR 0.05) and 0.9745963
652 (FDR<0.01).

653 Finally, a full network N is trained on the full data T and applied on the complete data set
654 with scaled column by the former min-max-scaling. The inferred score for each gene, repre-
655 sents the estimated probability for an association between the gene and a brain diseases.

656 All computational operations were performed in python with sklearn⁷⁰ for normalization and
657 keras for building and training the neural network.⁷¹ Seeds for numpy, K-Fold and Tensor-
658 flow were set to 1094795585.

659

660 **Known mutations**

661 The number of indels, nonsense, splice site and missense variants reported in PMT and no-
662 disorder genes were retrieved from the Human Gene Mutation Database (HGMD) Profes-
663 sional 2020.3 (Qiagen) and DECIPHER,⁷² and normalized by the CDS length. DECIPHER

664 variants include DDD variants and patient sequence variants that are not benign or likely be-
665 nign.

666 **Developmental disorders gene databases**

667 chrX genes were annotated for their classification in SysID
668 (<https://www.sysid.dbmr.unibe.ch/>)⁷³ and Gene2Phenotype (<https://www.ebi.ac.uk/gene2phe->
669 [notype](https://www.ebi.ac.uk/gene2phenotype/))⁷⁴ developmental disorders databases, which list genes associated with intellectual
670 disability or developmental disorders.

671 **Identification of damaging variants in predicted gene**

672 *De novo* variants in predicted genes were retrieved from Kaplanis et al.³⁸ and Martin et al.²¹
673 In addition, we examined variants in predicted genes identified in unsolved cases with
674 developmental disorders from two different cohorts: a cohort from Hôpital Pitié-Salpêtrière
675 (APHP, Paris, France), which mainly includes patients with intellectual disability or
676 neurological disorders (6500 individuals from 2346 families) and a cohort of patients with
677 dysgenesis of the corpus callosum, associated with learning or intellectual disabilities from
678 University of California San Francisco (UCSF; 1399 individuals from 463 families).
679 Informed consent of the legal representatives and appropriate approval of an ethics
680 committee, according to the French and American laws, have been obtained. Variant data and
681 clinical information were shared anonymously. Only damaging variants (*i.e.* variants leading
682 to a premature termination codon or missense variants with CADD PHRED scores ≥ 25) in
683 NN-predicted genes meeting at least one LME criterion were kept for further analysis.

684 **Functional validation of *TRPC5* missense variant**

685 HEK293 cells were transiently transfected with a plasmid expressing TRPC5-GFP WT or
686 p.Arg175Cys mutant channels. Currents were recorded from green fluorescent cells using the
687 whole-cell configuration of the patch-clamp technique 16 to 24 h after transfection at room
688 temperature. Voltage-clamp recordings were performed using an Axopatch 200B amplifier
689 and a Digidata 1440 Analogue/Digital interface (Axon Instruments). Data were low-pass fil-
690 tered at 2 kHz, digitized at 10 kHz and analysed offline using Clampfit software. Currents
691 were recorded during a 500 ms voltage ramp from -100 mV to 100 mV applied from a hold-
692 ing potential of -100 mV every 5 s. Series resistance was not compensated, and no leak sub-
693 traction was performed. Data were not corrected offline for voltage error and liquid junction
694 potential. The pipette solution contained (in mM): NaCl 8, Cs-methanesulfonate 120, MgCl₂
695 1, CaCl₂ 3.1, EGTA 10, HEPES 10, pH adjusted to 7.3 with CsOH. The extracellular solu-
696 tion contained (in mM): NaCl 140, MgCl₂ 1, CaCl₂ 2, HEPES 10, glucose 10, pH adjusted to
697 7.2 with NaOH. Immediate currents were recorded upon break-in (using patch pipettes that
698 contained 100 nM free Ca²⁺). The mutant currents are readily distinguished even in the ab-
699 sence of agonist stimulation, indeed the current-voltage relationship (IV) was similar to that
700 described in the literature, showing inward and outward ('double') rectification, giving some-
701 thing that is roughly 'Nshaped'. Englerin A (100nM) was applied to the cell expressing WT
702 and mutant TRPC5 without immediate current upon break-in to confirm its functional expres-
703 sion.

704 **Statistics**

705 Fisher's tests were performed for each chromosome to determine associations between: i)
706 genes being present in the specific chromosome and genes having an associated phenotype
707 with Clinical Synopsis data (CS-genes); ii) CS-genes being in the specific chromosome and

708 CS-genes having non-specific or specific neurologic features described in the Clinical Synop-
709 sis data. For chrX, Fisher's tests were performed to compute the enrichment/depletion of con-
710 firmed disorder-associated genes in each decile of the distribution of continuous variables.
711 Odds ratios were log₂ transformed and indicate enrichment or depletion of genes, for positive
712 or negative values, respectively (10^{-4} was added prior to transformation, whenever necessary
713 to deal with log of 0 issue). *P*-values were adjusted for multiple comparisons using Bonfer-
714 roni correction. For gene prediction, Fisher's tests were performed to calculate the association
715 between genes being confirmed-disorder genes and the number of predictors. Mann-Whitney
716 U test followed by Bonferroni correction for multiple testing was used to assess the coding-
717 sequence length difference between groups of genes.

718

719 **Data availability**

720 All data are available in the main text or supplementary materials. We used data from HGNC
721 (<https://www.genenames.org/>), OMIM (<https://www.omim.org/>), gnomAD v2.1.1
722 (<https://gnomad.broadinstitute.org/>), GTEx Portal (<https://gtexportal.org/home/>), BrainSpan
723 Atlas of the Developing Human Brain (<https://www.brainspan.org/>), Ensembl
724 (<https://www.ensembl.org/index.html>), UCSC Table Browser ([https://genome.ucsc.edu/cgi-](https://genome.ucsc.edu/cgi-bin/hgTables)
725 [bin/hgTables](https://genome.ucsc.edu/cgi-bin/hgTables)), Uniprot (<https://www.uniprot.org/>), Human Gene Mutation Database
726 (<http://www.hgmd.cf.ac.uk/ac/index.php>), DECIPHER (<https://www.deciphergenomics.org/>),
727 SysID database (<https://www.sysid.dbmr.unibe.ch/>), Gene2Phenotype
728 (<https://www.ebi.ac.uk/gene2phenotype>) and CADD - Combined Annotation Dependent De-
729 pletion (<https://cadd.gs.washington.edu/>).

730

731 **Code availability**

732 The following publicly available software packages were used to perform analyses: biomaRt
733 (<https://bioconductor.org/packages/release/bioc/html/biomaRt.html>),
734 phastCons100way.UCSC.hg38 (<http://bioconductor.org/packages/release/data/annotation/html/phastCons100way.UCSC.hg38.html>), rstatix (<https://cran.r-project.org/web/packages/rstatix/index.html>), numpy (<https://numpy.org/>), scikit-learn K-Fold
735 (<https://github.com/scikit-learn/scikit-learn>) and Tensorflow (<https://www.tensorflow.org/>).
736
737
738

739 **Supplementary references**

- 740 51. Braschi, B. *et al.* Genenames.org: the HGNC and VGNC resources in 2019. *Nucleic*
741 *Acids Res* **47**, D786-D792 (2019).
- 742 52. Consortium, G.T. The GTEx Consortium atlas of genetic regulatory effects across
743 human tissues. *Science* **369**, 1318-1330 (2020).
- 744 53. Yanai, I. *et al.* Genome-wide midrange transcription profiles reveal expression level
745 relationships in human tissue specification. *Bioinformatics* **21**, 650-9 (2005).
- 746 54. Kryuchkova-Mostacci, N. & Robinson-Rechavi, M. A benchmark of gene expression
747 tissue-specificity metrics. *Brief Bioinform* **18**, 205-214 (2017).
- 748 55. Miller, J.A. *et al.* Transcriptional landscape of the prenatal human brain. *Nature* **508**,
749 199-206 (2014).
- 750 56. Rodriguez, J.M. *et al.* APPRIS: annotation of principal and alternative splice
751 isoforms. *Nucleic Acids Res* **41**, D110-7 (2013).
- 752 57. Yates, A.D. *et al.* Ensembl 2020. *Nucleic Acids Res* **48**, D682-D688 (2020).
- 753 58. Gardiner-Garden, M. & Frommer, M. CpG islands in vertebrate genomes. *J Mol Biol*
754 **196**, 261-82 (1987).
- 755 59. Karolchik, D. *et al.* The UCSC Table Browser data retrieval tool. *Nucleic Acids Res*
756 **32**, D493-6 (2004).
- 757 60. Siepel, A. *et al.* Evolutionarily conserved elements in vertebrate, insect, worm, and
758 yeast genomes. *Genome Res* **15**, 1034-50 (2005).
- 759 61. Durinck, S. *et al.* BioMart and Bioconductor: a powerful link between biological
760 databases and microarray data analysis. *Bioinformatics* **21**, 3439-40 (2005).
- 761 62. Durinck, S., Spellman, P.T., Birney, E. & Huber, W. Mapping identifiers for the
762 integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat*
763 *Protoc* **4**, 1184-91 (2009).
- 764 63. UniProt, C. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res* **47**,
765 D506-D515 (2019).
- 766 64. Cotton, A.M. *et al.* Landscape of DNA methylation on the X chromosome reflects
767 CpG density, functional chromatin state and X-chromosome inactivation. *Hum Mol*
768 *Genet* **24**, 1528-39 (2015).
- 769 65. Carrel, L. & Willard, H.F. X-inactivation profile reveals extensive variability in X-
770 linked gene expression in females. *Nature* **434**, 400-4 (2005).

- 771 66. Wainer Katsir, K. & Linial, M. Human genes escaping X-inactivation revealed by
772 single cell expression data. *BMC Genomics* **20**, 201 (2019).
- 773 67. Balaton, B.P., Cotton, A.M. & Brown, C.J. Derivation of consensus inactivation
774 status for X-linked genes from genome-wide studies. *Biol Sex Differ* **6**, 35 (2015).
- 775 68. Zhang, Y. *et al.* Genes that Escape X-Inactivation in Humans Have High Intraspecific
776 Variability in Expression, Are Associated with Mental Impairment but Are Not Slow
777 Evolving. *Mol Biol Evol* **33**, 302 (2016).
- 778 69. Oliva, M. *et al.* The impact of sex on gene expression across human tissues. *Science*
779 **369**(2020).
- 780 70. Pedregosa, F. *et al.* Scikit-learn: Machine learning in Python. *Journal of Machine*
781 *Learning Research*, 2825–2830 (2011).
- 782 71. Chollet, F. Keras: the Python deep learning API. (2005).
- 783 72. Firth, H.V. *et al.* DECIPHER: Database of Chromosomal Imbalance and Phenotype in
784 Humans Using Ensembl Resources. *Am J Hum Genet* **84**, 524–33 (2009).
- 785 73. Kochinke, K. *et al.* Systematic Phenomics Analysis Deconvolutes Genes Mutated in
786 Intellectual Disability into Biologically Coherent Modules. *Am J Hum Genet* **98**, 149-
787 64 (2016).
- 788 74. Thormann, A. *et al.* Flexible and scalable diagnostic filtering of genomic variants
789 using G2P with Ensembl VEP. *Nat Commun* **10**, 2373 (2019).

790 **Acknowledgements**

791 The authors thank the patients with *CDK16* and *TRPC5* pathogenic variants and their family
792 for their participation in this study. We kindly thank Mrs Céline Cuny for Sanger sequencing
793 analysis of the patient with *TRPC5* missense variant. Electrophysiological experiments were
794 carried out at the electrophysiology core facility of ICM funded from the program “Inves-
795 tissements d’avenir” ANR-10-IAIHU-06. We thank Universitätsklinikum Essen, University
796 Duisburg-Essen, the Deutsche Forschungsgemeinschaft (DFG), the Tom-Wahlig-Stiftung
797 (TWS), the Deutsche Stiftung Neurologie (DSN), and Assistance Publique des Hôpitaux de
798 Paris (APHP) for their financial support to the research studies conducted by the authors. E.L.
799 and F.K. are associated with the FOR 2488 project (DFG, Project number 287074911). This
800 study makes use of data generated by the DECIPHER community. A full list of centres who
801 contributed to the generation of the data is available from [https://decipherge-](https://deciphergenomics.org/about/stats)
802 [nomics.org/about/stats](https://deciphergenomics.org/about/stats) and via email from contact@deciphergenomics.org. Funding for the
803 DECIPHER project was provided by Wellcome.

804

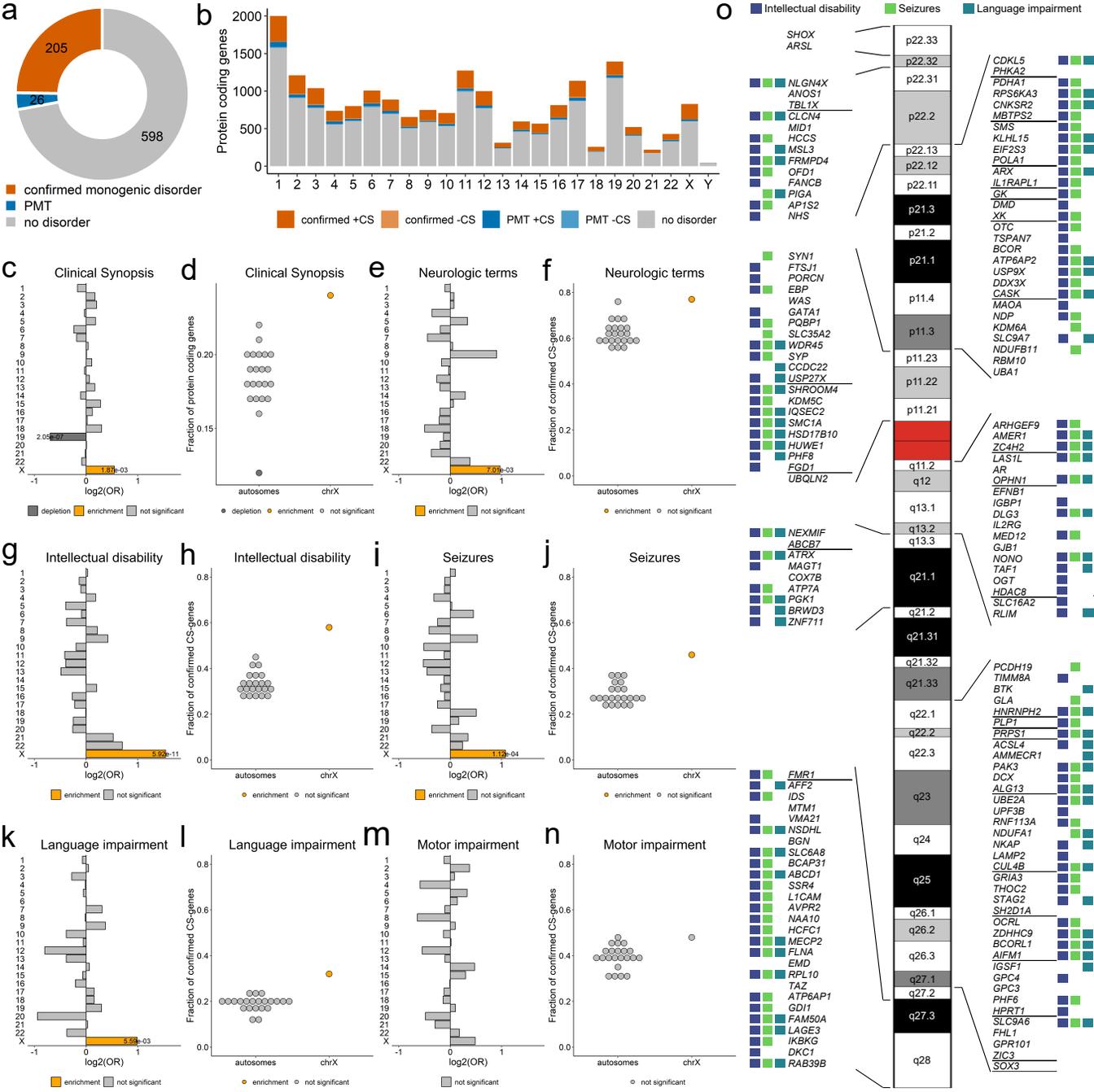
805 **Author Contributions**

806 E.L. and C.De. conceived and supervised the study. E.L. and C.S. performed the computa-
807 tions. E.L., I.P., C.Da., A.R., T.K., S.K., N.D., and C.E. carried out the experiments. A.K.,
808 B.G., E.S., C.N., J.P., B.D-B., L.V., A.P.A.S., E.K.V., J.A.J.V., F.J.K., F.T.M-T., M.S., P.S.,
809 S.G.M.F., E.A., E.H.S., F.E., J.B., B.K., C.M., D.H., J-L.M., J.G., V.M.K., B.H., and A.P.
810 contributed analysis, data and/or critically revised the manuscript for intellectual content.
811 E.L. and C.De. drafted the manuscript. All authors reviewed and approved the manuscript.

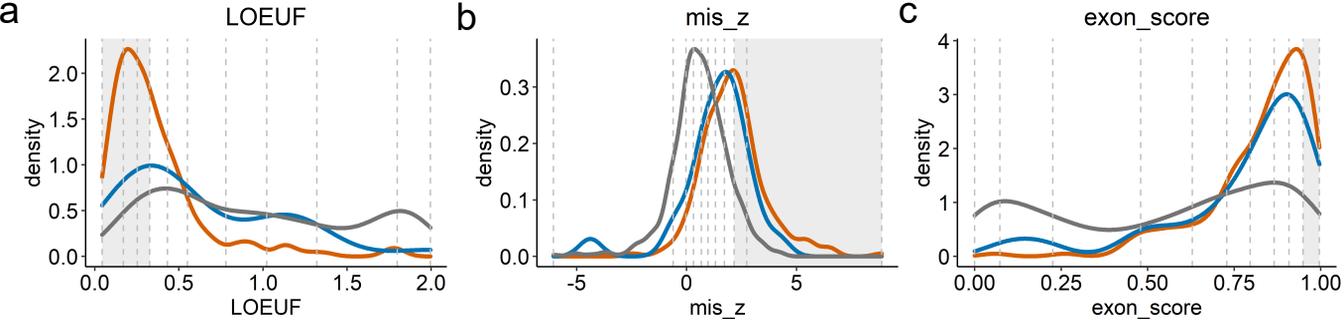
812

813 **Competing Interests statement**

814 The authors declare no competing interests.



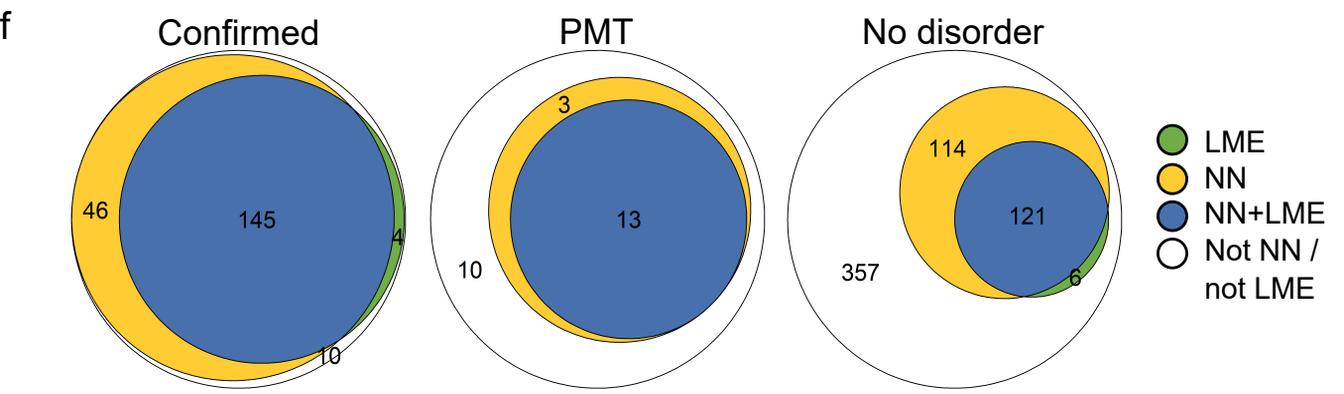
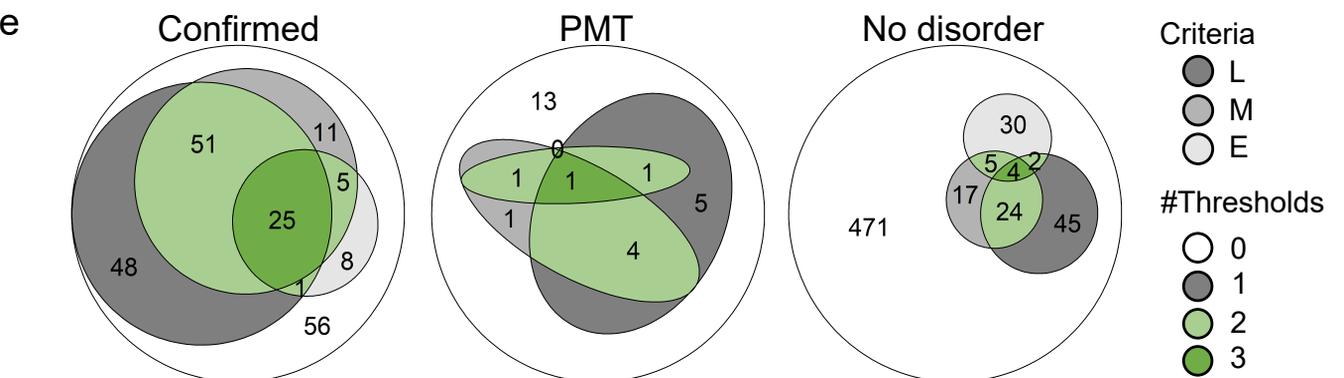
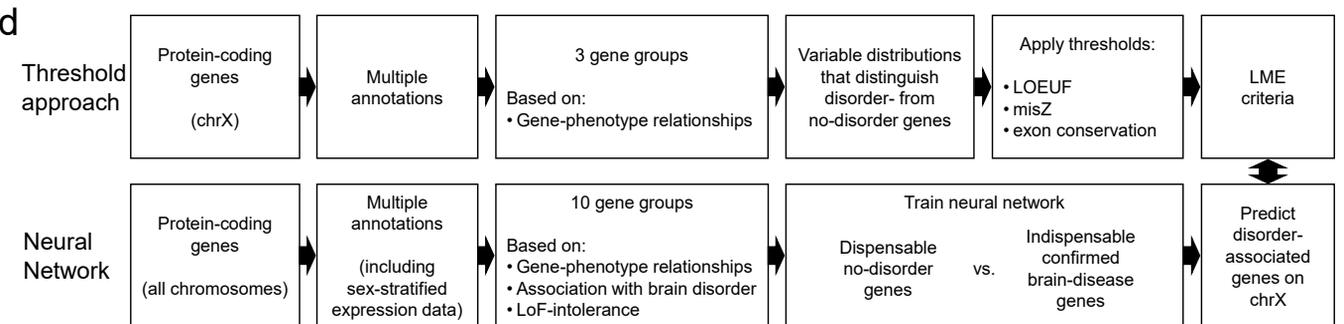
815 **Fig. 1: Disorder genes across chromosomes. a**, Number of protein-coding genes on chrX.
816 Genes associated with at least one monogenic disorder: orange, confirmed genes; blue: genes
817 with provisional associations (P), associated with susceptibility factors to multifactorial disor-
818 ders (M) or traits (T) (altogether: PMTs); grey: genes without known phenotypes (no-disorder
819 genes). **b**, Number of protein-coding genes *per* chromosome. Dark and light colors represent
820 genes with Clinical Synopsis (CS) data for at least one of the associated phenotypes (+CS) or
821 without CS (-CS), respectively. **c**, *Per* chromosome enrichment/depletion of protein-coding
822 genes with at least one associated phenotype comprising Clinical Synopsis data (confirmed
823 CS-genes). **d**, Fraction of confirmed CS-genes among protein-coding genes. **e-n**. Predomi-
824 nance of chrX genes associated with neurologic features. **e, g, i, k and m**, *Per* chromosome
825 enrichment/depletion of genes with non-specific neurologic features (**e**), intellectual disability
826 (**g**), seizures (**i**), language impairment (**k**) or motor development (**m**). **f, h, j, l and n**, Fraction
827 of genes associated with non-specific neurologic features (**f**), intellectual disability (**h**), sei-
828 zures (**j**), language impairment (**l**) or motor development (**n**). **c-n**, Yellow, enrichment; dark-
829 grey, depletion; light grey, not significant. Terms corresponding to the same neurological
830 clinical features were used in OMIM searches ([Supplementary Table 2](#)). Only significant p-
831 values are shown (Fisher's test followed by Bonferroni correction for multiple testing; [Sup-](#)
832 [plementary Tables 4 and 5](#)). **o**, Scheme of genes associated with neurological features on
833 chrX. Squares next to the genes represent association with intellectual disability (blue), sei-
834 zures (green) or language impairment (cyan). A horizontal line separates genes present in dif-
835 ferent chromosome bands.



confirmed PMT no disorder

confirmed PMT no disorder

confirmed PMT no disorder



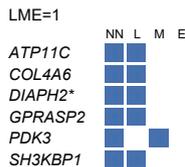
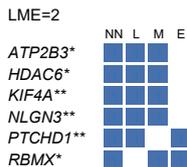
836 **Fig. 2: Prediction of disorder-genes. a-c**, Density plots showing the distribution of LOEUF
837 **(a)**, misZ **(b)** and exon-conservation score **(c)** according to gene group. Confirmed disorder
838 genes (orange), PMT genes (blue), no-disorder genes (grey). Vertical dashed lines separate
839 deciles of the overall distribution. Grey areas depict deciles for which confirmed disorder-as-
840 sociated genes are enriched (related to [Extended Data Fig. 2](#)). **d**, Overview of the approaches
841 used to uncover features shared by disorder genes and to predict new putative disorder-genes
842 (details in Supplementary Methods). **e**, Euler diagrams showing the number of genes ful-
843 filling LOEUF (L), misZ (M) and exon-conservation (E) criteria for confirmed (left), PMT
844 (middle) and no-disorder genes (right). Thresholds: $LOEUF \leq 0.326$, $misZ \geq 2.16$, exon-con-
845 servation score ≥ 0.9491 . Genes fulfilling at least two LME criteria are shown in green (light
846 or dark green for two or three criteria, respectively). Genes meeting only one of the metrics
847 are shown in different shades of grey (dark to light: LOEUF, misZ, exon-conservation). **f**, Eu-
848 ler diagrams showing the number of genes fulfilling LME criteria (LME, green), predicted by
849 the neural network (yellow) or both (blue) for confirmed (left), PMT (middle) and no-disor-
850 der genes (right). **e and f**, Genes neither predicted by the neural network nor meeting LME
851 criteria are shown in white.

a NN + LME≥1

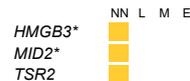


* ID candidate genes

** Current primary ID genes / ID data freeze

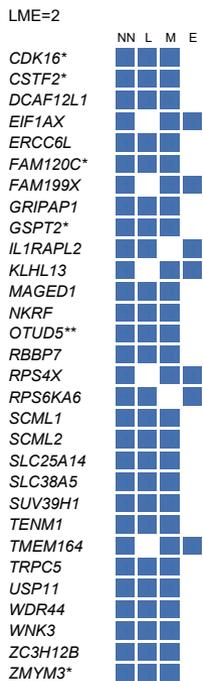
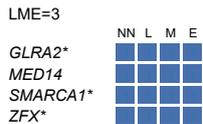


b NN + LME=0



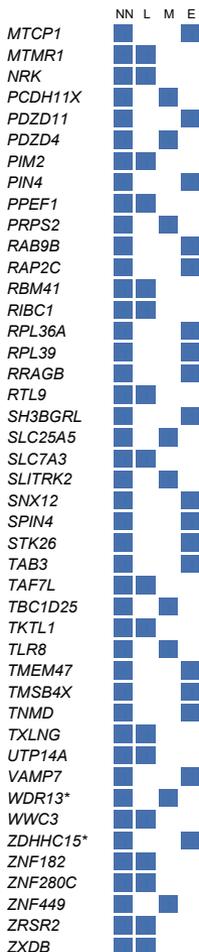
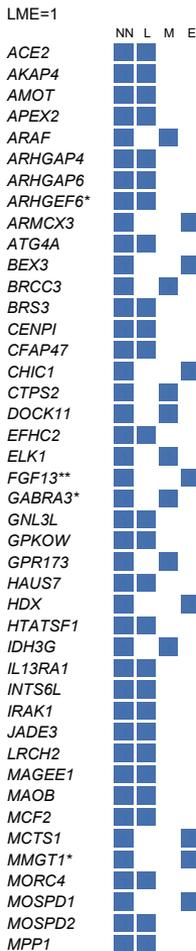
c

NN + LME≥1

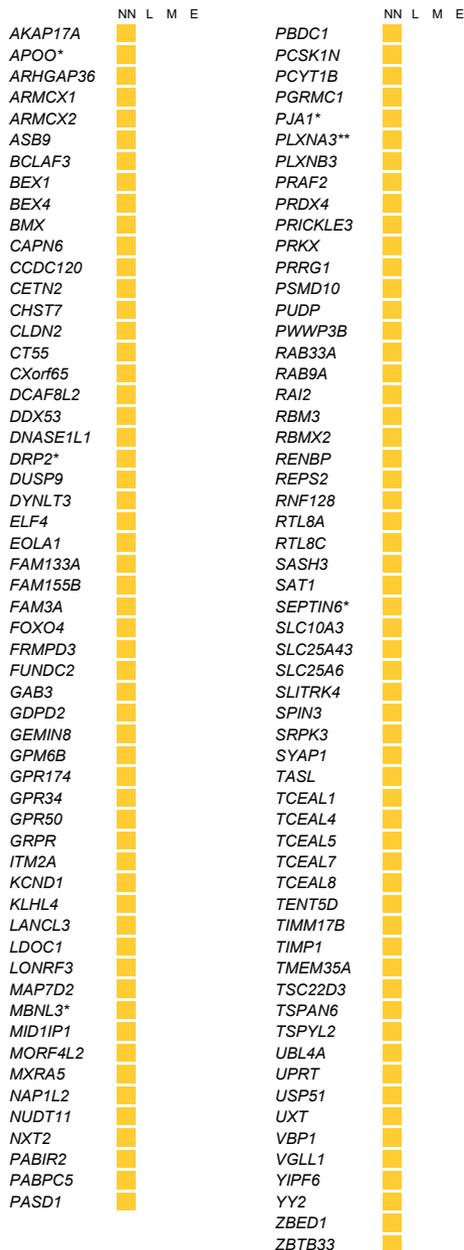


* ID candidate genes

** Current primary ID genes

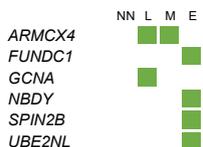


d NN + LME=0

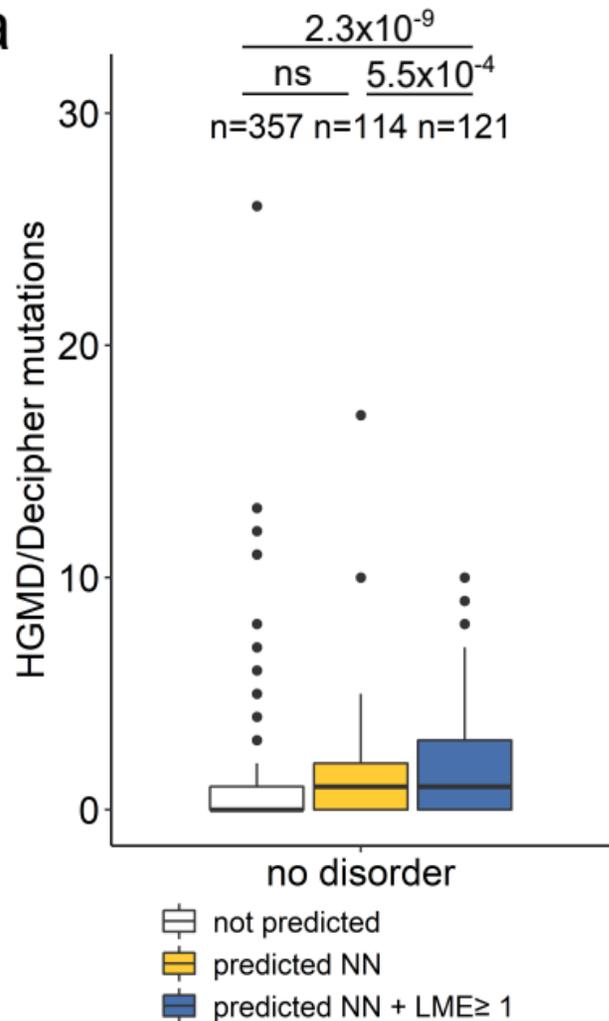
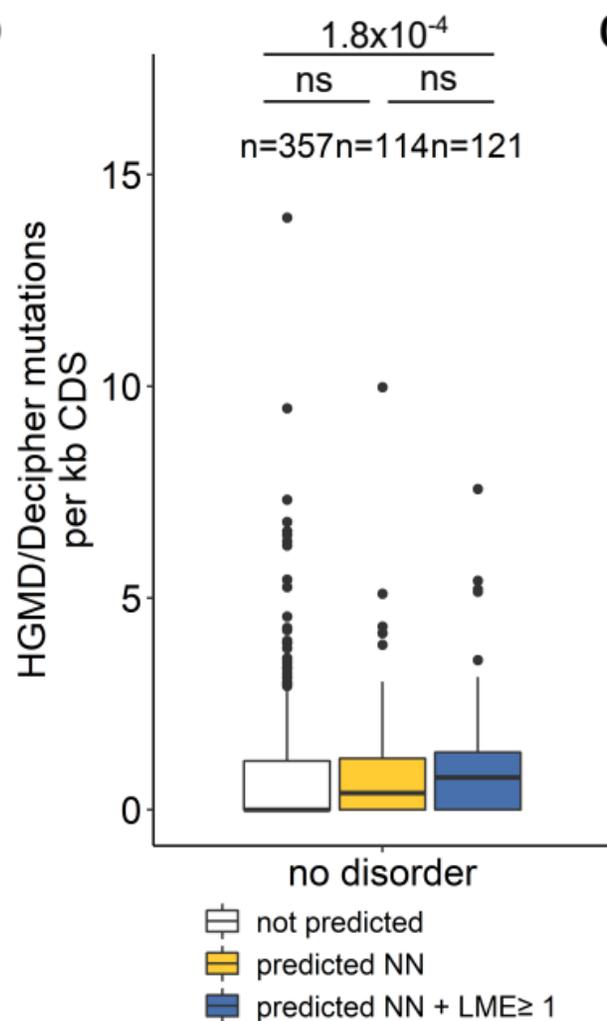
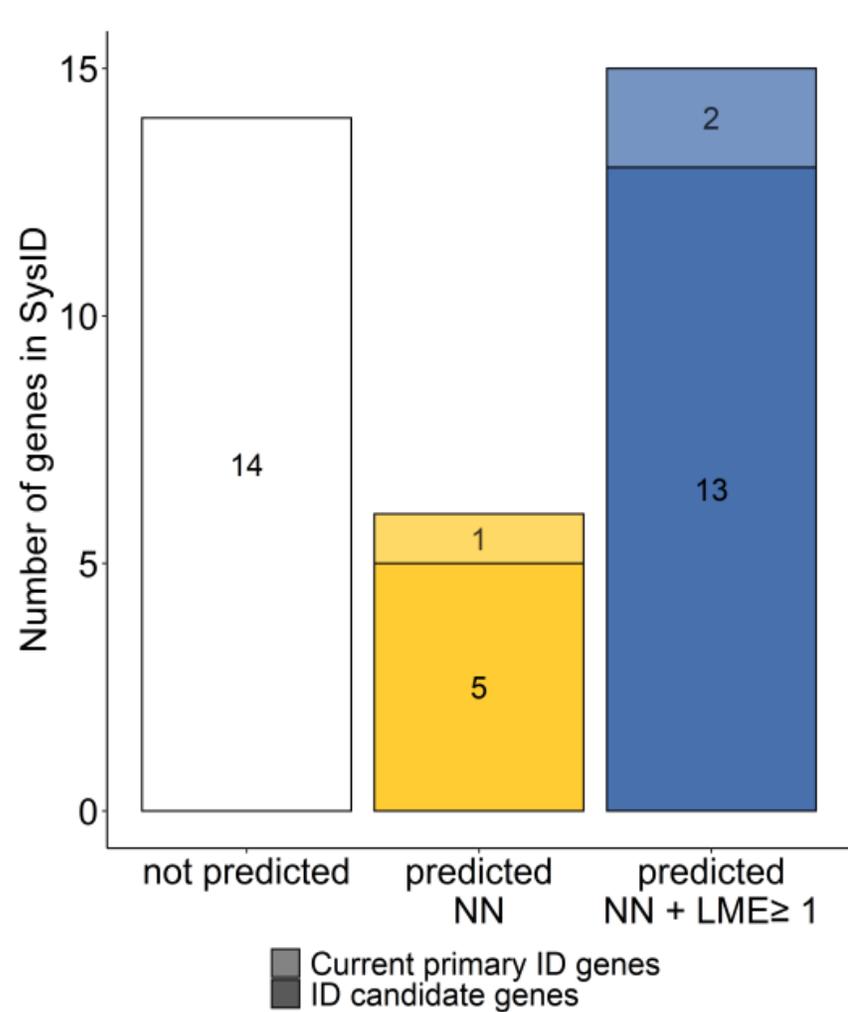


e

LME only



852 **Fig. 3: Different evidence strength for the predicted disorder-associated genes. a-d, PMT**
853 **(a, b)** and no-disorder genes **(c, d)** predicted by the neural network as possibly disease-associ-
854 ated, either fulfilling **(a, c)** or not LME criteria **(b, d)**. Genes present in the SysID database as
855 “ID candidate genes” (*) and “Current primary ID” or “ID data freeze” (**) are marked. **e,**
856 No-disorder genes meeting LME criteria but not predicted by the neural network.

a**b****c**

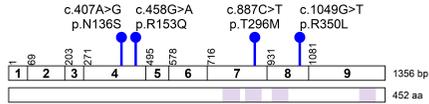
857 **Fig. 4: External supporting evidence for the predicted disorder-associated genes. a and**
858 **b**, Known point mutations in no-disorder genes. Boxplot showing the number of known mu-
859 tations reported in HGMD and DECIPHER (**a**) or their value normalized by coding-sequence
860 (CDS) length (**b**) according to their predicted status. No-disorder genes meeting LME criteria
861 but not predicted by the neural network were omitted from the analysis. Box plot elements
862 are defined as follows: center line: median; box limits: upper and lower quartiles; whiskers:
863 1.5× interquartile range; points: outliers. Related to [Supplementary Table 9](#). **c**, Number of no-
864 disorder genes present in the SysID database according to their predicted status.

Variants: ● LoF ● missense ● splicing

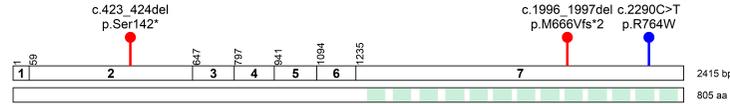
Domain Transmembrane Repeat (NHL, YD or WD40) Coiled coil Zinc finger

NN + LME=3

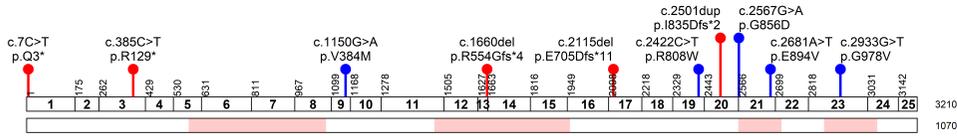
GLRA2/NM_002063.4/P23416



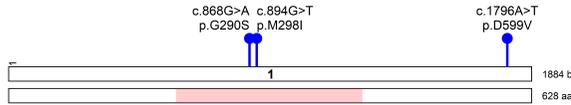
ZFX/NM_003410.4/P17010



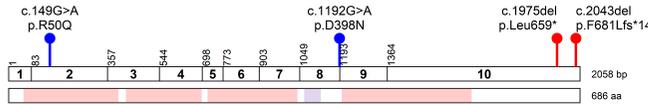
SMARCA1/NM_001282874.2/P28370



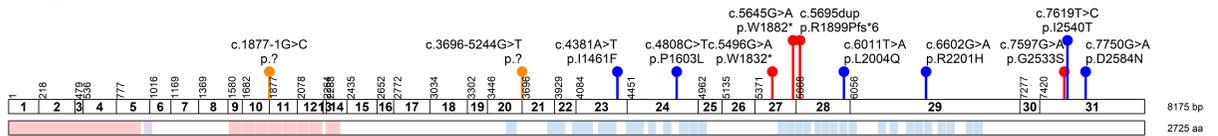
GSPT2/NM_018094.5/Q81YD1



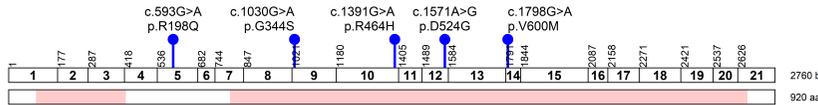
IL1RAPL2/NM_017416.2/Q9NP60



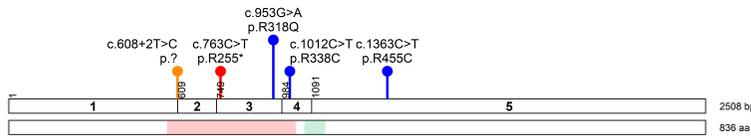
TENM1/NM_014253.3/Q9UKZ4



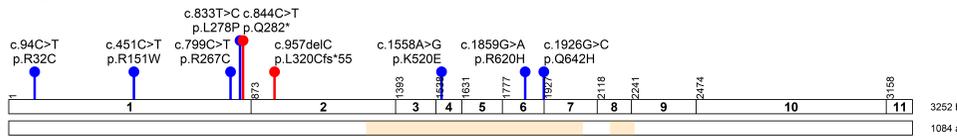
USP11/NM_001371072.1/G5E9A6



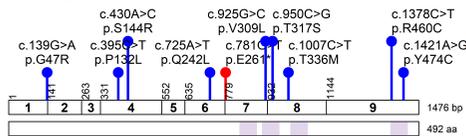
ZC3H12B/NM_001010888.4/Q5HYM0



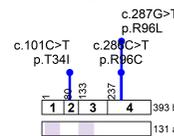
AMOT/NM_001113490.2/Q4VCS5



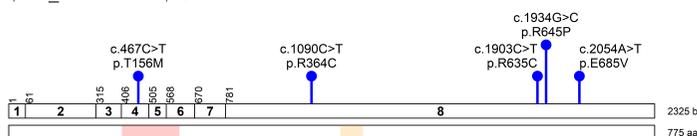
GABRA3/NM_000808.4/P34903



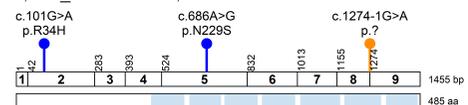
MMGT1/NM_173470.3/Q8N4V1



PDZD4/NM_001303512.2/Q17RL8



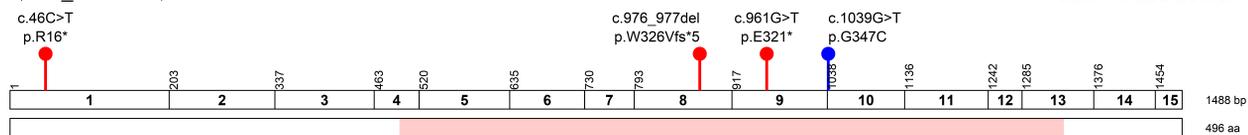
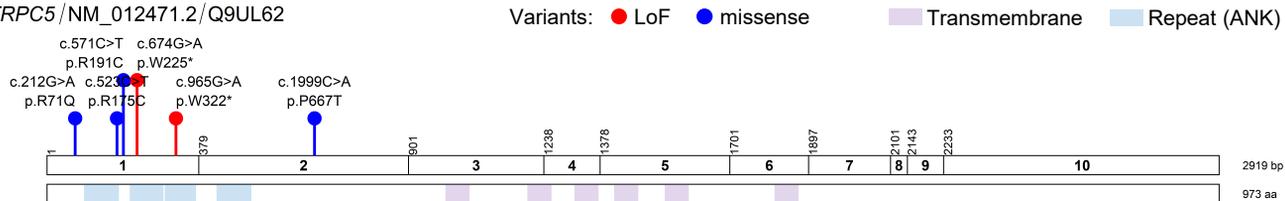
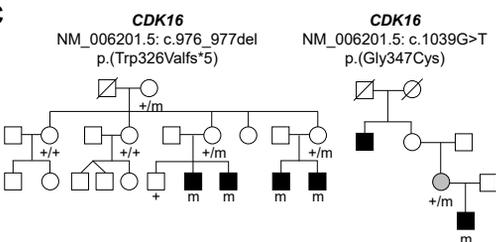
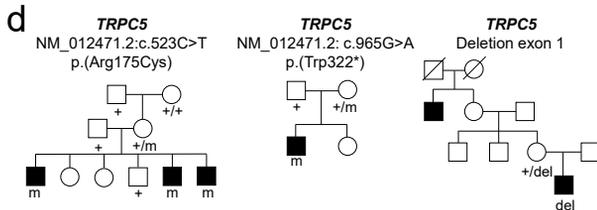
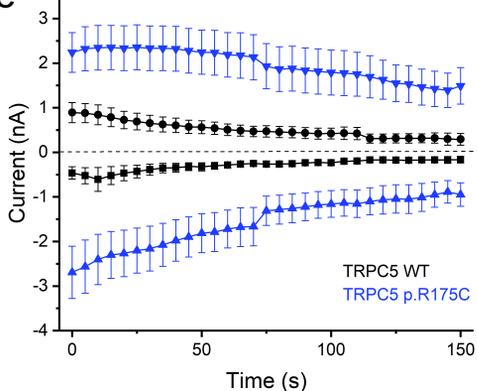
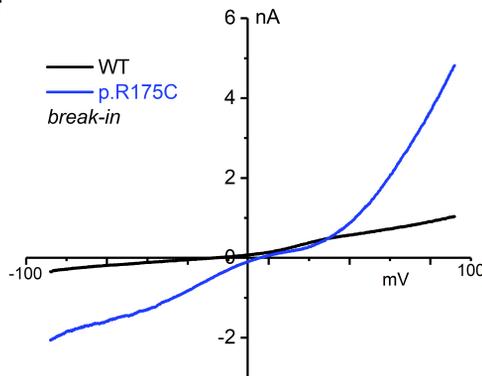
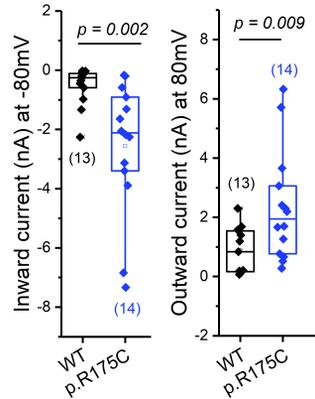
WDR13/NM_001347217.2/Q9H1Z4



NN + LME=2

NN + LME=1

865 **Fig. 5: Damaging variants in selected predicted disorder-associated genes.** Schematic
866 representation of the coding exons, protein domains (when present) and available damaging
867 variants (truncating or $CADD \geq 25$) for each selected predicted gene. Types of variants are
868 shown in different colors: lof-of-function (LoF, red), missense (blue), splicing (orange). Pro-
869 tein functional domains are shown: domains (light red), transmembrane segments (purple),
870 NHL, YD or WD40 repeats (light blue), coiled-coils (yellow), zinc-fingers (green). HGVS
871 cDNA and HGVS protein descriptions are shown. The corresponding RefSeq identifier of the
872 MANE Select transcript and Uniprot identifier are shown for each gene. Details of variants
873 displayed in this figure appear in [Supplementary Table 11](#).

a *CDK16*/NM_006201.5/Q00536**b** *TRPC5*/NM_012471.2/Q9UL62**c****d****e****f****g**

874 **Fig. 6: Validation of selected predicted disorder-associated genes: *CDK16* and *TRPC5*. a**
875 **and b**, Schematic representation of the coding exons, protein domains and variants of *CDK16*
876 **(a)** and *TRPC5* **(b)**. Variant types: loss-of-function (LoF, red), missense (blue). **c**, Pedigrees
877 of the three families with *CDK16* variants. The pedigree of family with c.976_977del,
878 p.(Trp326Valfs*5) was adapted from Hu et al. (2016)¹⁸. The two remaining families are un-
879 published. **d**, Pedigrees of the three families with *TRPC5* variants. The family with an intra-
880 genic exon deletion was reported in Mignon-Ravix et al. (2014)⁴³. The two other families are
881 novel. **e-g**, Functional characterization of *TRPC5* p.R175C by whole-cell patch-clamp record-
882 ings showing that the mutation renders the mutant channel constitutively opened. **e**, Time
883 course of inward and outward current amplitudes measured at +80 mV and -80 mV in
884 HEK293 cells transiently expressing WT (black) and mutant TRPC5 (red) in presence of
885 100nM free Ca²⁺ in the pipette. Values are reported as mean ± SEM. Recordings started few
886 seconds after the rupture. **f**, Representative whole-cell current-voltage (I-V) relationships of
887 WT and mutant TRPC5 channels current obtained shortly after break-in (≤ 10 s) with 100nM
888 free Ca²⁺ in the pipette. **g**, Boxplot of WT and mutant whole-cell current at -80 mV and +80
889 mV after break-in (one-way ANOVA with Dunnett's post hoc test). The number of independ-
890 ent recordings appears in brackets. Details of variants displayed in this figure appear in [Sup-](#)
891 [plementary Table 12](#).