

1 **Rare and *de novo* variants in 827 congenital diaphragmatic hernia probands implicate**  
2 ***LONPI* and *ALYREF* as new candidate risk genes**

3 Lu Qiao,<sup>1,2</sup> Le Xu,<sup>3</sup> Lan Yu,<sup>1</sup> Julia Wynn,<sup>1</sup> Rebecca Hernan,<sup>1</sup> Xueya Zhou,<sup>1,2</sup> Christiana  
4 Farkouh-Karoleski,<sup>1</sup> Usha S. Krishnan,<sup>1</sup> Julie Khlevner,<sup>1</sup> Aliva De,<sup>1</sup> Annette Zygmunt,<sup>1</sup>  
5 Timothy Crombleholme,<sup>4</sup> Foong-Yen Lim,<sup>5</sup> Howard Needelman,<sup>6</sup> Robert A. Cusick,<sup>6</sup>  
6 George B. Mychaliska,<sup>7</sup> Brad W. Warner,<sup>8</sup> Amy J. Wagner,<sup>9</sup> Melissa E. Danko,<sup>10</sup> Dai  
7 Chung,<sup>10</sup> Douglas Potoka,<sup>11</sup> Przemyslaw Kosiński,<sup>12</sup> David J. McCulley,<sup>13</sup> Mahmoud  
8 Elfiky,<sup>14</sup> Kenneth Azarow,<sup>15</sup> Elizabeth Fialkowski,<sup>15</sup> David Schindel,<sup>16</sup> Samuel Z. Soffer,<sup>17</sup>  
9 Jane B. Lyon,<sup>18</sup> Jill M. Zalieckas,<sup>19</sup> Badri N. Vardarajan,<sup>20</sup> Gudrun Aspelund,<sup>1</sup> Vincent P.  
10 Duron,<sup>1</sup> Frances A. High,<sup>19,21,22</sup> Xin Sun,<sup>3</sup> Patricia K. Donahoe,<sup>21,23</sup> Yufeng Shen,<sup>2,24,25,\*</sup> and  
11 Wendy K. Chung<sup>1,26,\*</sup>

12 <sup>1</sup>Department of Pediatrics, Columbia University Irving Medical Center, New York, NY  
13 10032, USA; <sup>2</sup>Department of Systems Biology, Columbia University Irving Medical Center,  
14 New York, NY 10032, USA; <sup>3</sup>Department of Pediatrics, University of California, San Diego  
15 Medical School, San Diego, CA 92092, USA; <sup>4</sup>Medical City Children's Hospital, Dallas, TX  
16 75230, USA; <sup>5</sup>Cincinnati Children's Hospital Medical Center, Cincinnati, OH 45229, USA;  
17 <sup>6</sup>University of Nebraska Medical Center College of Medicine, Omaha, NE 68114, USA;  
18 <sup>7</sup>University of Michigan Health System, Ann Arbor, MI 48109, USA; <sup>8</sup>Washington  
19 University School of Medicine, St. Louis, MO 63110, USA; <sup>9</sup>Children's Hospital of  
20 Wisconsin, Medical College of Wisconsin, Milwaukee, WI 53226, USA; <sup>10</sup>Monroe Carell Jr.  
21 Children's Hospital at Vanderbilt, Nashville, TN 37232, USA; <sup>11</sup>University of Pittsburgh,  
22 Pittsburgh, PA 15224, USA; <sup>12</sup>Medical University of Warsaw, 02-091 Warsaw, Poland;  
23 <sup>13</sup>Department of Pediatrics, University of Wisconsin-Madison, Madison, WI 52726, USA;  
24 <sup>14</sup>Cairo University, Cairo 11432, Egypt; <sup>15</sup>Oregon Health & Science University, Portland, OR

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

25 97239, USA; <sup>16</sup>UT Southwestern Medical Center, Dallas, TX 75390, USA; <sup>17</sup>Northwell  
26 Health, New York, NY 11040, USA; <sup>18</sup>Department of Radiology, University of Wisconsin-  
27 Madison, Madison, WI 53792, USA; <sup>19</sup>Department of Surgery, Boston Children's Hospital,  
28 Boston, MA 02115, USA; <sup>20</sup>Department of Neurology, Taub Institute for Research on  
29 Alzheimer's Disease and the Aging Brain and the Gertrude H. Sergievsky Center, Columbia  
30 University in the City of New York, NY 10032, USA; <sup>21</sup>Pediatric Surgical Research  
31 Laboratories, Massachusetts General Hospital, Boston, MA 02114, USA; <sup>22</sup>Department of  
32 Pediatrics, Massachusetts General Hospital, Boston, MA 02114, USA; <sup>23</sup>Department of  
33 Surgery, Harvard Medical School, Boston, MA 02115, USA; <sup>24</sup>Department of Biomedical  
34 Informatics, Columbia University Irving Medical Center, New York, NY 10032, USA; <sup>25</sup>JP  
35 Sulzberger Columbia Genome Center, Columbia University Irving Medical Center, New  
36 York, NY 10032, USA; <sup>26</sup>Department of Medicine, Columbia University Irving Medical  
37 Center, New York, NY 10032, USA

38 \*Correspondence: [ys2411@cumc.columbia.edu](mailto:ys2411@cumc.columbia.edu) (Y.S.), [wkc15@cumc.columbia.edu](mailto:wkc15@cumc.columbia.edu)  
39 (W.K.C.).

## 40 **Abstract**

41 Congenital diaphragmatic hernia (CDH) is a severe congenital anomaly that is often  
42 accompanied by other anomalies. Although the role of genetics in the pathogenesis of CDH  
43 has been established, only a small number of disease genes have been identified. To further  
44 investigate the genetics of CDH, we analyzed *de novo* coding variants in 827 proband-parent  
45 trios and confirmed an overall significant enrichment of damaging *de novo* variants,  
46 especially in constrained genes. We identified *LONPI* (Lon Peptidase 1, Mitochondrial) and  
47 *ALYREF* (Aly/REF Export Factor) as novel candidate CDH genes based on *de novo* variants  
48 at a false discovery rate below 0.05. We also performed ultra-rare variant association

49 analyses in 748 cases and 11,220 ancestry-matched population controls and identified  
50 *LONPI* as a risk gene contributing to CDH through both *de novo* and ultra-rare inherited  
51 largely heterozygous variants clustered in the core of the domains and segregating with CDH  
52 in familial cases. Approximately 3% of our CDH cohort was heterozygous with ultra-rare  
53 predicted damaging variants in *LONPI* who have a range of clinical phenotypes including  
54 other anomalies in some individuals and higher mortality and requirement for extracorporeal  
55 membrane oxygenation. Mice with lung epithelium specific deletion of *Lonpl* die  
56 immediately after birth and have reduced lung growth and branching that may at least  
57 partially explain the high mortality in humans. Our findings of both *de novo* and inherited  
58 rare variants in the same gene may have implications in the design and analysis for other  
59 genetic studies of congenital anomalies.

## 60 **Introduction**

61 Congenital diaphragmatic hernia (CDH) affects approximately 3 per 10,000 neonates<sup>1,2</sup>.  
62 Approximately 40% of CDH cases occur with additional congenital anomalies besides  
63 common secondary anomalies (dextrocardia and lung hypoplasia)<sup>3</sup>. The most common  
64 additional anomalies<sup>4,5</sup> are structural heart defects (11-15%), musculoskeletal malformations  
65 (15-20%) including limb deficiency, club foot, and omphalocele<sup>6</sup>. However, anomalies of  
66 almost every organ have been described in association with CDH. Despite advances in care  
67 including improved prenatal diagnosis, fetal interventions, extracorporeal membrane  
68 oxygenation (ECMO) and gentle ventilation, CDH continues to be associated with at least  
69 20% mortality and significant long-term morbidity including feeding difficulties, pulmonary  
70 hypertension and other respiratory complications, and neurocognitive deficits<sup>3,7,8</sup>.

71 The complexity of the phenotypes associated with CDH is mirrored by the complexity of the  
72 genetics, which are heterogeneous with approximately 30% of CDH cases having an

73 identifiable major genetic contributor. Typically, each gene or copy number variant (CNV)  
74 associated with CDH accounts for at most 1-2% of cases<sup>9</sup>. The full spectrum of genomic  
75 variants has been associated with CDH, including chromosome aneuploidies (10%), copy  
76 number variants (CNVs) (3-10%), monogenic conditions (10-22%), and emerging evidence  
77 for oligogenic causes<sup>1</sup> (CNVs and individual genes<sup>10</sup>).

78 While familial cases have been described, CDH most commonly occurs in individuals  
79 without a family history of CDH, and sibling recurrence risk in isolated cases is less than  
80 1%<sup>11</sup>. Likely due to the historically high mortality and low reproductive fitness, CDH is often  
81 due to *de novo* CNVs and single gene variants. However, dominant inheritance has been  
82 described with transmission of an incompletely penetrant variant from an unaffected parent  
83 or parent with a subclinical diaphragm defect<sup>12</sup>. CDH has also been described in individuals  
84 with biallelic variants such as Donnai-Barrow syndrome<sup>13</sup>. The occurrence of discordant  
85 monogenic twins suggests a role for stochastic events after fertilization<sup>11</sup>.

86 A genetic diagnosis for probands with CDH can inform prognosis and guide medical  
87 management. Some genetic conditions associated with CDH are associated with an increased  
88 risk for additional anomalies, increased mortality, and increased morbidity including  
89 neurocognitive disabilities that may benefit from early intervention<sup>3</sup>. Over the past decade,  
90 advances in genomic sequencing technology have helped to define the genes associated with  
91 CDH. We and others have shown that *de novo* variants with large effect size contribute to 10-  
92 22% of CDH cases with enrichment of *de novo* likely damaging variants in CDH cases with  
93 an additional anomaly (complex CDH)<sup>9,14,15</sup>. We also demonstrated a higher burden of *de*  
94 *nov*o likely damaging (LD) variants in females compared to males supporting a “female  
95 protective model”<sup>9</sup>. Most recently, in a cohort with long term developmental outcome data<sup>3</sup>,  
96 we demonstrated that *de novo* likely damaging (LD) variants are associated with poorer

97 neurodevelopmental outcomes as well as a higher prevalence of pulmonary hypertension  
98 (PH).

99 To expand upon our knowledge of the diverse genetic etiologies of CDH, we performed  
100 whole genome (WGS) or exome sequencing of 827 CDH proband-parent trios. We confirmed  
101 an overall enrichment of damaging *de novo* variants in constrained genes, and identified  
102 *LONPI* (Lon Peptidase 1, Mitochondrial [MIM: 605490]) and *ALYREF* (Aly/REF Export  
103 Factor [MIM: 604171]) as new candidate CDH genes with recurrent ultra-rare and *de novo*  
104 variants.

## 105 **Materials and methods**

### 106 **Participant recruitment and control datasets**

107 Study participants were enrolled as fetuses, neonates, children and adults with a  
108 radiologically confirmed diaphragm defect by the DHREAMS study<sup>16</sup> (Diaphragmatic Hernia  
109 Research & Exploration; Advancing Molecular Science) or Boston Children's  
110 Hospital/Massachusetts General Hospital (BCH/MGH) as described previously<sup>14</sup>. Clinical  
111 data were prospectively collected from medical records and entered into a central Research  
112 Electronic Data Capture (REDCap) database<sup>17</sup>. Probands and both parents provided a blood,  
113 skin biopsy, or saliva specimen for trio genetic analysis. All studies were approved by the  
114 Columbia University institutional review board (IRB), serving as the central site. Each  
115 participating site also procured approval from their local IRB and signed informed consent  
116 was obtained. Ethical approval was obtained from the following participating institutions:  
117 Boston Children's Hospital/Massachusetts General Hospital, Washington University,  
118 Cincinnati Children's Hospital Medical Center, Children's Hospital & Medical Center of  
119 Omaha, University of Michigan, Monroe Carell Jr. Children's Hospital, Northwell Health,

120 Oregon Health & Science University, Legacy Research Institute, University of Texas  
121 Southwestern, Children's Hospital of Wisconsin, and Children's Hospital of Pittsburgh.

122 A total of 827 cases and their parents had whole genome (WGS) or exome sequencing in the  
123 current study. A subset of trios (n=574) has been described in our previous study<sup>3,9</sup>.

124 Participants with only a diaphragm defect were classified as isolated CDH while participants  
125 with at least one additional major congenital anomaly (*e.g.* congenital heart defect, central  
126 nervous system anomaly, gastrointestinal anomaly, skeletal anomaly, genitourinary anomaly,  
127 cleft lip/palate), moderate to severe developmental delay, or other neuropsychiatric  
128 phenotypes at last contact were classified as complex CDH. Pulmonary hypoplasia, cardiac  
129 displacement and intestinal herniation were considered to be part of the diaphragm defect  
130 sequence and were not considered independent malformations. Data on the child's current  
131 and past health including family history of congenital anomalies, postoperative pulmonary  
132 hypertension, mortality or survival status prior to initial discharge, extracorporeal membrane  
133 oxygenation (ECMO) intake were gathered as described previously<sup>3</sup>.

134 The control group consisted of unaffected parents from the Simons Powering Autism Research  
135 for Knowledge (SPARK) study<sup>18</sup> (exomes) and Latinx samples from Washington Heights-  
136 Hamilton Heights-Inwood Community Aging Project (WHICAP) study<sup>19</sup> (exomes).

### 137 **WGS and exome data analysis**

138 There are 233 CDH trios processed using whole genome sequencing (WGS) that were not  
139 included in previous studies<sup>3,9</sup> (Table S1). Of these 233 previously unpublished trios, 1 trio  
140 was processed at Baylor College of Medicine Human Genome Sequencing Center and 232  
141 trios at Broad Institute Genomic Services. The genomic libraries of 219 cases were prepared  
142 by TruSeq DNA PCR-Free Library Prep Kit (Illumina), while 14 were TruSeq DNA PCR-

143 Plus Library Prep Kit (Illumina), with average fragment length about 350 bp, and sequenced  
144 as paired-end of 150-bp on Illumina HiSeq X platform. Exome sequencing was performed in  
145 20 CDH trios that were not previously published<sup>3,9</sup>. Among these, the coding exons of 9 trios  
146 were captured using Agilent Sure Select Human All Exon Kit v2 (Agilent Technologies), 10  
147 trios using NimbleGen SeqCap EZ Human Exome V3 kit (Regeneron NimbleGen), 1 trio  
148 using NimbleGen SeqCap EZ Human Exome V2 kit (Roche NimbleGen). Exomes of  
149 SPARK cohort were captured using a slightly modified version of the IDT xGen Exome  
150 Research Panel v.1.0 identical to the previous study<sup>20</sup>. Whole-exome sequencing of the  
151 WHICAP cohort was performed at Columbia University using the Roche SeqCap EZ Exome  
152 Probes v3.0 Target Enrichment Probes<sup>21</sup>.

153 Exome and WGS data of cases and controls were processed using a pipeline implementing  
154 GATK Best Practice v4.0 as previously described<sup>9,22</sup>. Specifically, reads of exome cases were  
155 mapped to human genome GRCh37 reference using BWA-MEM<sup>23</sup>, while reads of WGS  
156 cases, SPARK and WHICAP controls were mapped to GRCh38; duplicated reads were  
157 marked using Picard<sup>24</sup>; variants were called using GATK<sup>25</sup> (v4.0) HaplotypeCaller to  
158 generate gVCF files for joint genotyping. All samples within the same batch (Table S1) were  
159 jointly genotyped and variant quality score recalibration (VQSR) was performed using  
160 GATK. To combine all cases for further analysis, we lifted over the GRCh37 variants to  
161 GRCh38 using CrossMap<sup>26</sup> (v0.3.0). Common SNP genotypes within exome regions were  
162 used to validate familial relationships using KING<sup>27</sup> and ancestries using peddy<sup>28</sup> (v0.4.3) in  
163 cases, SPARK controls and WHICAP controls.

164 *De novo* variants were defined as a variant present in the offspring with homozygous  
165 reference genotypes in both parents. Here, we limited WGS to coding regions based on  
166 coding sequences and canonical splice sites of all GENCODE v27 coding genes. We took a

167 series of stringent filters to identify *de novo* variants as described previously<sup>9</sup>: VQSR tranche  
168  $\leq 99.8$  for SNVs and  $\leq 99.0$  for indels; GATK's Fisher Strand  $\leq 25$ , quality by depth  $\geq 2$ . We  
169 required the candidate *de novo* variants in probands to have  $\geq 5$  reads supporting the  
170 alternative allele,  $\geq 20\%$  alternative allele fraction, Phred-scaled genotype likelihood  $\geq 60$   
171 (GQ), and population allele frequency  $\leq 0.01\%$  in gnomAD v2.1.1; both parents to have  $\geq 10$   
172 reference reads,  $< 5\%$  alternative allele fraction, and GQ  $\geq 30$ . We applied DeepVariant<sup>29</sup> to all  
173 candidate *de novo* variants for in silico confirmation and only included the ones with PASS  
174 from DeepVariant for downstream analysis.

175 To reduce batch effects in combined datasets from different sources<sup>30</sup> in analysis of rare  
176 variants, for non-Latinx population we targeted ultra-rare variants located in xGen-captured  
177 protein coding regions and for Latinx population in regions targeted by xGen and SeqCap EZ  
178 v3.0. We used the following criteria to minimize technical artifacts and select ultra-rare  
179 variants<sup>22</sup>: cohort AF  $< 0.5\%$  and population cohort  $< 1 \times 10^{-5}$  across all genomes in gnomAD  
180 v3.0; mappability=1;  $> 90\%$  target region with depth  $\geq 10$ ; overlapped with segmental  
181 duplication regions  $< 95\%$ ; genotype quality  $> 30$ , allele balance  $> 20\%$  and depth  $> 10$  in  
182 cases.

183 We used Ensembl Variant Effect Predictor<sup>31</sup> (VEP, Ensemble 102) and ANNOVAR<sup>32</sup> to  
184 annotate variant function, variant population frequencies and *in silico* predictions of  
185 deleteriousness. All coding SNVs and indels were classified as synonymous, missense,  
186 inframe, or likely-gene-disrupting (LGD, which includes frameshift indels, canonical splice  
187 site, or nonsense variants). We defined predicted damaging missense (D-mis) based on  
188 CADD<sup>33</sup> score v1.3. All *de novo* variants and inherited variants in candidate risk genes were  
189 manually inspected in the Integrative Genome Viewer (IGV). A total of 179 variants were  
190 selected for validation using Sanger sequencing; all of them were confirmed (Table S2). To

191 compare the clinical outcomes between cases with deleterious variants in candidate genes and  
192 with likely damaging (LD) variants, we defined likely damaging variants as in our previous  
193 study<sup>3</sup>: (a) *de novo* LGD or deleterious missense variants in genes that are constrained (ExAC  
194 pLI  $\geq$ 0.9) and highly expressed in developing diaphragm<sup>34</sup>, or (b) *de novo* LGD or  
195 deleterious missense variants in known risk genes for CDH or commonly comorbid disorders  
196 (congenital heart disease [CHD] and neurodevelopmental delay [NDD]), or (c) plausible  
197 deleterious missense variants in known risk genes for CDH or commonly comorbid disorders  
198 (CHD and NDD), or (d) deletions in constrained (ExAC pLI  $\geq$ 0.9) or haploinsufficient genes  
199 from ClinGen genome dosage map<sup>35</sup>, or (e) CNVs implicated in known syndromes. We  
200 classified CDH cases into two genetic groups: (1) LD, if the case carried at least one *de novo*  
201 LD variant; (2) non-LD, if the case carried no such variants.

202 *De novo* copy number variants (CNVs) were identified using an inhouse pipeline of read  
203 depth-based algorithm based on CNVnator<sup>36</sup> v0.3.3 in WGS trios as described in our previous  
204 study<sup>3</sup>. The *de novo* CNV segments were validated by the additional pair-end/split-read  
205 (PE/SR) evidence using Lumpy<sup>37</sup> v0.2.13 and SVtyper<sup>38</sup> v0.1.4. Only the CNVs supported by  
206 both read depth (RD) and PE/SR were included in downstream analysis. We mapped *de novo*  
207 CNVs on GENCODE v29 protein coding genes with at least 1bp in the shared interval. The  
208 GENCODE genes were annotated with variant intolerance metric by ExAC pLI<sup>39</sup>,  
209 haploinsufficiency metric by Episcor<sup>40</sup>, haploinsufficiency and triplosensitivity of genes  
210 from ClinGen genome dosage map<sup>35</sup>, and CNV syndromes from DECIPHER<sup>41</sup> v11.1.

## 211 **Quantitative PCR**

212 We performed experimental validation of putative *de novo* genic CNVs using quantitative PCR  
213 (qPCR). All PCR primers were designed for the selected genes located within the *de novo*  
214 CNVs and synthesized by IdtDNA. All qPCR reactions were performed in a total of 10  $\mu$ l

215 volume, comprising 5  $\mu$ l 2x SYBR Green I Master Mix (Promega), 1  $\mu$ l 10nM of each primer  
216 and 2  $\mu$ l of 1:20 diluted cDNA in 96-well plates using CFX Connect Real-Time PCR Detection  
217 System (Bio-Rad). All reactions were performed in triplicate, and the conditions were 5  
218 minutes at 95 °C, then 40 cycles of 95 °C at 15 seconds and 60 °C at 30 seconds. The relative  
219 copy numbers were calculated using the standard curve method relative to the  $\beta$ -actin  
220 housekeeping gene. Five-serial 4-fold dilutions of DNA samples were used to construct the  
221 standard curves for each primer.

## 222 **Statistical analysis**

223 *Burden of de novo variants.* The baseline mutation rates for different classes of *de novo*  
224 variants were calculated in each GENCODE coding gene using the published trinucleotide  
225 sequence context<sup>42</sup>, and we calculated the rate in protein-coding regions that are uniquely  
226 mappable as previously described mutation model<sup>9,18</sup>. The observed number of variants of  
227 various types (*e.g.* synonymous, missense, LGD) in each gene set and case group was  
228 compared with the baseline expectation using Poisson test. In all analyses, constrained genes  
229 were defined by ExAC pLI<sup>39</sup> score of >0.5, and all remaining genes were treated as other  
230 genes. We used a less stringent pLI threshold than previously suggested<sup>39</sup> for defining  
231 constrained genes, because it captures more known haploinsufficient genes important for  
232 heart and diaphragm development. We compared the observed number of variants in female  
233 versus male cases and complex versus isolated cases using the binormal test.

234 *extTADA analysis.* To identify risk genes based on *de novo* variants, we used an empirical  
235 Bayesian method, extTADA<sup>43</sup> (Extended Transmission and *de novo* Association). The  
236 extTADA model was developed based on a previous integrated empirical Bayesian model  
237 TADA<sup>44</sup> and estimates mean effect sizes and risk-gene proportions from the genetic data  
238 using MCMC (Markov Chain Monte Carlo) process (details see supplemental note). To

239 inform the parameter estimation with prior knowledge of developmental disorders, we  
240 stratify the genes into constrained genes (ExAC pLI score >0.5) and non-constrained genes  
241 (other genes), followed by estimating the parameters using the extTADA model to each  
242 group of genes. After estimating posterior probability of association (PPA) of individual  
243 genes in each group, we combined both groups to calculate a final false discovery rate (FDR)  
244 for each gene using extTADA's procedure.

245 *Gene-based case-control association analysis of ultra-rare variants.* To identify novel risk  
246 genes based both on *de novo* and rare inherited variants, we performed a gene-based  
247 association test comparing the frequency of ultra-rare deleterious variants in CDH cases with  
248 controls, without considering *de novo* status. Samples with read depth coverage  $\geq 10x$  for  
249 80% in exome cases and 90% in genome cases of the targeted regions were included in the  
250 analysis (Figure S1). Relatedness was checked using KING<sup>27</sup>, and only unrelated cases were  
251 included in the association tests (Figure S2). To control for confounding from genetic  
252 ancestry, we selected ancestry-matched controls using SPARK exomes and Latinx WHICAP  
253 exomes to reach a fixed case/control ratio in each population ancestry inferred by peddy<sup>28</sup>  
254 (Figure S3). Specifically, for a specific ancestry ( $i$ ), consider  $x_i$  number of cases,  $y_i$  number  
255 of controls,  $n_i$  the fold controls to cases ( $y_i/x_i$ ). We chose the minimized  $n_{min}$  among all  
256 ancestries. In each genetic-ancestry group controls ( $y_i$ ), we ranked the Euclidean distance  
257 between each case and controls which were calculated from top 3 PCA eigenvectors and  
258 selected  $n_{min}x_i$  controls from  $y_i$  controls to ensure the same proportions in cases and  
259 controls. After filtering to reduce the impact of false positive variants, we tested for similarity  
260 of the ultra-rare synonymous variant rate among cases and controls in specific genetic-  
261 ancestry groups, assuming that ultra-rare synonymous variants are mostly neutral with  
262 respect to disease status.

263 To identify CDH risk genes, we tested the burden of ultra-rare deleterious variants (AF  
264  $<1 \times 10^{-5}$  across all gnomAD v3.0 genomes, LGD or D-Mis) in each protein-coding gene in  
265 cases compared to controls. To improve statistical power, we searched for a gene-specific  
266 CADD<sup>33</sup> score threshold for defining D-Mis that maximized the burden of ultra-rare  
267 deleterious variants in cases compared to controls and used permutations to calculate  
268 statistical significance with the variable threshold test<sup>22,45</sup>. For the binomial tests in each  
269 permutation, we used binom.test function in R to calculate p values. We performed two  
270 association tests, one with LGD and D-Mis variants combined and the other with D-Mis  
271 variants alone, to account for different modes of action. We defined the threshold for  
272 genome-wide significance by Bonferroni correction for multiple testing (as two tests for each  
273 gene with 20,000 protein-coding genes, threshold p-value= $1.25 \times 10^{-6}$ ). We checked for  
274 inflation using a quantile-quantile (Q-Q) plot and calculated the genomic control factor  
275 ( $\lambda$ ) using QQperm in R. Lambda equal to 1 indicates no deviation from the  
276 expected distribution.

### 277 **Protein modeling**

278 We searched the LONP1 canonical sequence (identifier: P36776-1) in UniProt and obtained  
279 the structural model of the human mitochondrial LONP1 monomer (encompassing only the  
280 residue range 413–951) using SWISS-MODEL server<sup>46</sup> with SMTL ID 6u5z.1 as template.  
281 The 3D structure was visualized using PyMOL molecular viewer (The PyMOL Molecular  
282 Graphics System, Version 1.2r3pre, Schrödinger, LLC).

### 283 **Mice**

284 All mice were housed in American Association for Accreditation of Laboratory Animal Care  
285 accredited facilities and laboratories at University of California, San Diego. All animal

286 experiments were conducted under approved guidelines for the Care and Use of Laboratory  
287 Animals. *Lonp1<sup>fl</sup>* and *Shh<sup>cre</sup>* mice have all been described previously<sup>47</sup> (International Mouse  
288 Strain Resource J:204812). All mice were bred on a C57BL/6J background, and littermates  
289 were used as controls to minimize potential genetic background effects.

## 290 **Results**

### 291 **Cohort characteristics**

292 Participants were recruited as part of the multi-site DHREAMS study (n=748) and from the  
293 Boston Children's Hospital/Massachusetts General Hospital (n=79). We performed WGS on  
294 734 proband-parent trios and exome sequencing on 93 trios. In total, we analyzed 827 trios  
295 with WGS or exome sequencing.

296 In the cohort, there were 486 (59%) male probands (Table 1), consistent with a higher  
297 prevalence of CDH in males<sup>9,48,49</sup>. The genetically determined ancestries (Figure S3A) were  
298 European (73.4%), admixed American (hereafter referred to as Latinx; 18.5%), African  
299 (3.7%), East Asian (1.8%), and South Asian (2.5%). Among the 277 (33.5%) complex cases,  
300 the most frequent additional anomalies were congenital heart disease (n=144), NDD (n=54),  
301 skeletal anomalies (n=46), genitourinary anomalies (n=46) and gastrointestinal anomalies  
302 (n=42). A total of 533 (64.4%) probands had isolated CDH without additional anomalies at  
303 the time of last follow up. The most common type of CDH was left-sided Bochdalek (Table  
304 1).

### 305 **Burden of *de novo* coding variants**

306 We identified 1153 *de novo* protein-coding variants in 619 (74.8%) cases including 1058  
307 single nucleotide variants (SNVs) and 95 indels (Table S2). The average number of *de novo*

308 coding variants per proband is 1.39. The number of *de novo* coding variants across probands  
309 closely follows a Poisson distribution (Figure S4). Transition-to-transversion ratio of *de novo*  
310 SNVs was 2.75. We classified variants that were likely gene disruptive (LGD) or predicted  
311 damaging missense (“D-mis” with CADD score $\geq$ 25) as damaging variants. A total of 418  
312 damaging variants (126 LGD and 292 D-mis) were identified in 318 (38.4%) cases, including  
313 83 (10%) cases harboring two or more such variants.

314 We analyzed the burden of *de novo* variants in CDH cases by comparing the observed  
315 number of variants to the expected number based on the background mutation rate.  
316 Consistent with previous studies on CDH<sup>9</sup> and other developmental disorders<sup>50-52</sup>, both *de*  
317 *nov*o LGD (0.15 per case) and D-mis variants (0.35 per case) were significantly enriched in  
318 cases (relative risk [RR]=1.5, P=3.6 $\times$ 10<sup>-5</sup> for LGD; RR=1.3, P=3.1 $\times$ 10<sup>-6</sup>; Figures 1A and B;  
319 Table S3) while the frequency of synonymous variants (0.30 per case) closely matches the  
320 expectation (RR=0.9, P=0.12; Table S3). The burden of LGD variants is mostly located in  
321 constrained (ExAC<sup>39</sup> pLI >0.5) genes (RR=2.2, P=1.8 $\times$ 10<sup>-8</sup>). It is marginally higher in  
322 female cases than male cases (RR=3.0 vs 1.36, P=0.012) and marginally higher in complex  
323 cases than isolated cases (RR=3.1 vs 1.75, P=0.024; Figure 1C; Table S3).

324 To identify new CDH risk genes by *de novo* variants, we applied extTADA<sup>43</sup> to the data of  
325 827 CDH trios. ExtTADA assumes a model of genetic architecture compatible with the  
326 observed burden and recurrence of *de novo* damaging variants and estimates a false discovery  
327 rate (FDR) for each gene using MCMC. From the burden analysis of *de novo* variants in  
328 CDH and previous studies<sup>52</sup>, we reasoned that the constrained genes (ExAC pLI >0.5) drive  
329 the higher burden of *de novo* damaging variants and are more likely to be plausible risk  
330 genes. We stratified the data into the constrained gene set and the non-constrained gene set  
331 (Table S4) and estimated extTADA priors (mean relative risk and prior probability of being a

332 risk gene) in these two gene sets separately. Constrained genes had a higher prior of risk  
333 genes than non-constrained genes (0.037 vs 0.006). Meanwhile, both LGD and D-mis had  
334 higher relative risks in constrained genes than non-constrained gene (18.30 vs 5.24 for LGD;  
335 10.01 vs 3.81 for D-mis). We estimated Bayes Factor of individual genes within each gene  
336 group and then combined the genes from two groups together to calculate FDR. We  
337 identified 3 genes with FDR <0.05: *MYRF* (Myelin Regulatory Factor [MIM: 608329]),  
338 *LONP1*, and *ALYREF*. Five of 6 *MYRF* *de novo* variants were described in our previous  
339 study<sup>9</sup>. We identified 3 participants harboring *de novo* D-mis variants in *LONP1* and 2  
340 participants for *de novo* LGD variants in *ALYREF*. Of two participants with an *ALYREF* LGD  
341 variant, one had an isolated left-side CDH and the other had right-side CDH and ventricular  
342 septal defect. There were nine additional genes with  $\geq 2$  *de novo* predicted deleterious  
343 variants (*HSD17B10* [MIM: 300256], *GATA4* [MIM: 600576], *SYMPK* [MIM: 602388],  
344 *PTPN11* [MIM: 176876], *WT1* [MIM: 607102], *FAM83H* [MIM: 611927], *CACNA1H*  
345 [MIM: 607904], *SEPSECS* [MIM: 613009], and *ZFYVE26* [MIM: 612012]) (Table 2). Of  
346 these, three are known CDH genes (*MYRF*, *GATA4*, *WT1*). All *de novo* variants in these  
347 genes are heterozygous.

#### 348 **Recurrent genes in *de novo* CNVs**

349 We applied CNVnator to call CNVs from WGS data and used customized filters to identify  
350 *de novo* CNVs. We performed experimental validation of 25 putative *de novo* genic CNVs  
351 including all 9 small CNVs (<5kb) using quantitative PCR (qPCR). 22 of 25 (88%) reported  
352 *de novo* CNV in cases were confirmed by qPCR. Removing the 3 false positive CNVs, there  
353 were 87 *de novo* CNVs identified in 734 CDH cases with WGS with an average of 0.12 per  
354 case (Table S5). Among them, there were 54 (62%) deletions ranging from 2,096 bp to 33.7  
355 Mb and 33 (38%) duplications ranging from 1,165 bp to 24.9 Mb. Seven samples carried

356 known syndromic CNVs in DECIPHER<sup>41</sup> dataset, one of which was heterozygous for a  
357 16p13.11 microduplication, two heterozygous for a 17q12 deletion associated with renal  
358 cysts and diabetes (RCAD), three heterozygous for 21q22 duplication in the critical region  
359 for Down syndrome, and one heterozygous for 22q11 deletion associated with DiGeorge  
360 syndrome. No recurrent genes were identified between *de novo* SNVs and CNVs. Four CNVs  
361 were recurrent (Table 3), two of which encompass single genes *CSMD1* (CUB And Sushi  
362 Multiple Domains 1 [MIM: 608397]) and *GPHN* (Gephyrin [MIM: 603930]).

363 **Candidate gene *LONPI* contributes to CDH risk through both *de novo* and rare**  
364 **inherited variants**

365 To identify additional risk genes that may contribute through rare inherited variants, we  
366 performed a gene-based, case-control association analysis of ultra-rare variants. Specifically,  
367 we used exome data from the SPARK (unaffected parents) and Latinx WHICAP samples as  
368 controls. Quality control procedures included at least 10x depth of sequence coverage across  
369 the target regions (Figure S1) and detection of cryptic relatedness amongst all CDH  
370 participants and controls (Figure S2). To prevent confounding by genetic ancestry, we  
371 performed principal component analysis (PCA) by peddy to infer genetic ancestry of all cases  
372 and controls and selected matching controls (15-fold of cases numbers in each specific  
373 genetic-ancestry group) to reach a fixed case/control ratio. With the same genetic-ancestry  
374 proportion in cases and controls (77% Europeans, 14.8% Latinx, 4.1% Africans, 2% East  
375 Asians, 2.1% South Asians; Figure S3; Table S6), we selected 748 cases and 11,220 controls  
376 for downstream analysis. We filtered the ultra-rare variant call sets of cases and controls in  
377 each genetic-ancestry group by empirical filters to reduce false positive calls and minimize  
378 technical batch effects across data sets. After filtering, the average numbers of ultra-rare  
379 ( $AF < 1 \times 10^{-5}$  across all gnomAD v3.0 genomes) synonymous variants per subject in cases and

380 controls are nearly identical in everyone (enrichment rate=1, P=1) and specific ancestral  
381 groups (Table S7). Furthermore, a gene-level burden test confined to ultra-rare synonymous  
382 variants was consistent with a global null model in Q-Q plot (Figure S5), indicating that  
383 technical batch effects would likely have minimal impact on genetic analyses. We then  
384 performed a variable threshold association test<sup>22,45</sup> to identify new risk genes based on  
385 enrichment of ultra-rare damaging variants in individual genes. For each gene, we tested  
386 enrichment of LGD and D-mis variants together or just D-mis variants, in order to account  
387 for potential different biological modes of action. In the variable threshold test, we  
388 determined a gene-specific optimal CADD score threshold to define D-mis in order to  
389 maximize the power of the association test and then estimated type I error rate by  
390 permutations. The overall result from the case-control association did not show inflation from  
391 the null model ( $\lambda=1.09$ ; Figure 2A). The association of *LONPI* ( $P=1\times 10^{-7}$ ; Figure 2)  
392 exceeded the Bonferroni-corrected significance threshold ( $1.25\times 10^{-6}$ , account for two tests in  
393 each gene). Three of the 24 ultra-rare deleterious variants in *LONPI* were known *de novo*  
394 variants. Two known CDH risk genes, *ZFPM2* (Zinc Finger Protein, FOG Family Member 2  
395 [MIM: 603693]) and *MYRF*, fell just below the cutoff for genome wide significance.

396 The association of *LONPI* is due to both LGD and D-mis variants. We screened the whole  
397 cohort (Figure 3 and Table 4), including CDH relatives (n=1) and exome sequencing  
398 singletons (n=2), for ultra-rare damaging missense ( $CADD \geq 25$ ) and LGD in *LONPI*  
399 (NM\_004793.3). A total of 23 CDH cases in 829 cases (2.8%) carry 24 *LONPI* variants,  
400 including 10 LGD and 14 D-mis variants. Among 22 *LONPI* variants excluding 2 of  
401 unknown inheritance variants in singletons, there are 3 (13.6%) *de novo* variants (all D-mis)  
402 and 19 (86.4%) inherited variants, 36.8% of which are from mothers (n=7). Of 19 inherited  
403 variants, 8 parents carrying *LONPI* variants have a family history of CDH or diaphragm  
404 eventration (n=4) or other congenital anomaly (n=4; brain abnormality, cerebral palsy, cleft

405 palate, skeletal abnormality) segregating with the *LONPI* variant. Three inherited variants  
406 (c.1913C>T [p.638M], c.2122G>A [p.G708S] and c.2263C>G [p.R755G]) are each observed  
407 twice in the cohort on different probands. Familial segregation was established in six familial  
408 CDH cases for c.398C>G (p.P133R), c.6391G>T (p.X213\_splice), c.1262delG  
409 (p.F421Lfs\*87), c.1574C>T (p.P525L), c.1913C>T (p.T638M) and c.2719dupG  
410 (p.V907Gfs\*73). One case (01-1279) carries biallelic heterozygous variants with c.1574C>T  
411 (p.P525L) inherited from one parent and c.2263C>G (p.R755G) inherited from another  
412 parent. The participant with biallelic heterozygous variants required ECMO and died at 8-9  
413 hours after birth with severe bilateral CDH with near complete diaphragm agenesis, bilateral  
414 lung hypoplasia, and no additional anomalies (Figure 4). All other cases are heterozygous  
415 variants.

416 Previous studies reported biallelic variants in *LONPI* in cerebral, ocular, dental, auricular,  
417 and skeletal (CODAS) syndrome<sup>53,54</sup> (MIM: 600373). We compared the locations of the  
418 predicted-damaging missense positions in CDH cases and CODAS syndrome cases (Figures  
419 3 and 5). No variants overlap between CDH cases and CODAS syndrome. *LONPI* contains  
420 three functional domains. CDH damaging variants are concentrated at the core of the  
421 domains. Biallelic variants in CODAS syndrome are located on the junction of ATP-binding  
422 and proteolytic domains (Figures 3 and 5). The 23 CDH cases with *LONPI* variants didn't  
423 have features of CODAS syndrome.

#### 424 **Phenotype of CDH probands with *LONPI* variants**

425 We identified 24 ultra-rare heterozygous variants in 23 sporadic or familial CDH participants  
426 (Table 4). The majority (n=17; 73.9%) are of European ancestry and 13 (56.5%) are female  
427 (Table 4). Sixteen (70%) were enrolled as neonates. Fourteen of the 23 have a family history  
428 of congenital anomalies (Table 4), 6 of whom had a family history of CDH. Nine (39.1%) are

429 complex cases. Six of 9 complex cases have CHD in addition to CDH. We compared the  
430 clinical outcomes or phenotypes in CDH cases with *LONPI* damaging variants and other  
431 CDH cases (Table 5). Compared to CDH cases without *LONPI* ultra-rare damaging variants,  
432 *LONPI* damaging variant carriers are associated with higher neonatal mortality rate prior to  
433 initial hospital discharge (69% vs 16%,  $P=6.4\times 10^{-6}$ ) and greater need for ECMO (56% vs  
434 28%,  $P=2.3\times 10^{-2}$ ). Compared to CDH cases with other likely damaging variants defined in  
435 our previous study<sup>3</sup>, *LONPI* damaging variant carriers had higher neonatal mortality rate  
436 prior to discharge (69% vs 24%,  $P=1.8\times 10^{-3}$ ) and trended towards greater need for ECMO  
437 (56% vs 30%,  $P=0.077$ ).

#### 438 **Inactivation of *Lonpl* in mouse embryonic lung epithelium leads to disrupted lung** 439 **development and full lethality at birth**

440 The high rate of mortality and need for ECMO in cases with CDH is predominantly due to  
441 abnormal lung and pulmonary vascular development causing lung hypoplasia and pulmonary  
442 hypertension. Our hypothesis was that impaired or partial loss of *LONPI* function in cases  
443 with CDH might contribute directly to abnormal lung development, independent of its role in  
444 diaphragm formation. To test this hypothesis, we inactivated *Lonpl* in the embryonic lung  
445 epithelium in mice. This was achieved by generating *Shh<sup>cre/+</sup>;Lonpl<sup>fl/fl</sup>* (hereafter *Lonpl* cKO  
446 for conditional knockout) embryos using existing alleles<sup>47</sup> (International Mouse Strain  
447 Resource J:204812). In the mutant the cre recombinase expressed specifically in the  
448 epithelium drove *Lonpl* inactivation at the onset of lung initiation (Figure 6A). This led to  
449 100% lethality of the mutants at birth with normal body size (Figures 6B and C). Upon  
450 dissection, the mutant lung was composed of large fluid-filled sacs, unlike the controls with  
451 normal airways and alveoli (Figure 6D). The lung defect likely contributed to embryonic  
452 lethality at birth in these mutant mice.

## 453 Discussion

454 In the current study of 827 CDH trios, we confirmed there is an overall enrichment of  
455 damaging *de novo* variants, particularly in constrained genes. We identified *LONPI* and  
456 *ALYREF* as novel candidate genes based on enrichment of *de novo* variants. By case-control  
457 association, we also confirmed *LONPI* as a genome-wide significant candidate gene  
458 contributing to CDH risk through both *de novo* and inherited damaging variants. We  
459 demonstrated segregation of the *LONPI* variant with diaphragm defect in five families. We  
460 found that CDH individuals with heterozygous ultra-rare damaging variants in *LONPI* have  
461 clinical phenotypes frequently including CHD or skeletal anomalies, frequently requiring  
462 ECMO, and having a higher mortality than the rest of our CDH cohort. In addition, we  
463 confirmed *MYRF* and *ZFPM2* as genes previously associated with CDH<sup>9,14,55,56</sup>. In a mouse  
464 model with knock out of *Lonpl* only in the embryonic lung epithelium with an intact  
465 diaphragm, we demonstrated reduced pulmonary growth and branching, resulting in perinatal  
466 lethality that suggests that the higher mortality rate and need for ECMO in human is due to a  
467 primary effect of *LONPI* on pulmonary development in addition to diaphragm development.

468 The burden of damaging *de novo* variants in CDH is consistent with previous studies<sup>9,14,15</sup>,  
469 and damaging *de novo* variants are more frequent in complex CDH compared to isolated  
470 CDH cases. Similar patterns have been observed in complex congenital heart disease with  
471 other congenital anomalies or neurodevelopmental disorders compared with isolated  
472 congenital heart disease<sup>50</sup> and autism with/without intellectual disability<sup>57</sup>. Deleterious *de*  
473 *novo* variants are more frequent in many severe early-onset diseases with reduced  
474 reproductive fitness compared to the general population<sup>58</sup>. The higher frequency of *de novo*  
475 LGD variants in female relative to male CDH cases supports the “female protective model”  
476 similar to autism<sup>52,59,60</sup>, which means that risk variants have larger effects in males than in

477 females so that females require a higher burden to reach the same diagnostic threshold as  
478 males.

479 Both *de novo* and rare inherited variant analyses highlight *LONPI* as a novel CDH candidate  
480 gene. Approximately 3% of individuals in our CDH cohort are heterozygous for *LONPI* rare  
481 variants. Three variants (p. T638M, p.G708S and p.R755G) are recurrently and  
482 independently found in unrelated families. CDH cases with *LONPI* variants had higher  
483 mortality in the neonatal period compared with other CDH cases. Biallelic variants in *LONPI*  
484 have been reported in CODAS, a multi-system developmental disorder characterized by  
485 cerebral, ocular, dental, auricular, and skeletal anomalies<sup>61</sup>. The Lonp1 holoenzyme is a  
486 homohexamer with six identical subunits. Each subunit consists of a mitochondrial-targeting  
487 sequence (MTS), a substrate recognition and binding (N) domain, an ATPase (AAA+)  
488 domain, and a proteolytic (P) domain. Biallelic missense variants reported in CODAS  
489 individuals are mostly located in the junction of ATP-binding and proteolytic domains of  
490 *LONPI* while the heterozygous variants identified in CDH individuals are located in the main  
491 domains of *LONPI*. Notably, there are no overlapping variants between CDH and CODAS  
492 individuals. Most of the variants in CODAS are located in the alpha-helix and may affect the  
493 interactions of subunits<sup>61</sup>. Variants in CDH may interrupt the proteolytic and ATP binding  
494 domains, resulting in the dysfunction of *LONPI*. Homozygous deletion of *LONPI* in mice is  
495 embryonic lethal, due to progressive loss of mtDNA with subsequent failure to meet energy  
496 requirements for embryonic development<sup>62</sup>. Heterozygous *Lonp1*<sup>+/-</sup> mice develop normally  
497 without obvious abnormalities, but *lonp1* expression decreased in both RNA and protein  
498 levels<sup>62</sup>. Analysis of *Lonp1* expression in heterozygous mice indicated a 50% reduction at  
499 both RNA and protein levels in these animals. These data suggest different mechanisms of  
500 *LONPI* in diseases with biallelic and monoallelic variants. Of note, one CDH individual

501 carried biallelic variants (p.P525L and p.R755G). No additional phenotypes were noted,  
502 perhaps because the baby died at 8-9 hours after birth with severe bilateral CDH (Figure 4).

503 Lonp1 is a nuclear-encoded mitochondrial protease. Besides binding of mtDNA<sup>63</sup>, Lonp1 was  
504 discovered as an ATP-dependent protease involved in the degradation of misfolded or  
505 damaged proteins<sup>64-66</sup>. Accumulation of misfolded proteins has been observed in the impaired  
506 lungs of developing mice with deletion of other ATP-dependent proteins<sup>67</sup>. The immature  
507 lung development and neonatal respiratory failure of our *Lonp1* cKO mice could be due to  
508 the inactivation of Lon protease, which results in the accumulation of misfolded proteins and  
509 activation of the unfolded protein response (UPR) pathway<sup>68</sup>. UPR activation during  
510 development could lead to reduced cell proliferation and cause other congenital anomalies  
511 including congenital heart disease<sup>69</sup>.

512 Lonp1 also acts as a chaperone that interacts with other mitochondrial proteins to regulate  
513 several cellular processes<sup>70</sup>. Lon expression may stimulate cell proliferation<sup>71</sup> and Lon  
514 downregulation may impair mitochondrial structure and function and cause apoptosis<sup>72,73</sup>.

515 Alterations in cell proliferation, differentiation and migration can all lead to CDH. Myogenic  
516 cell differentiation and migration are essential during formation of the diaphragm<sup>74</sup>.

517 Myogenic differentiation requires increased expression of mitochondrial biogenesis-related  
518 genes including Lon<sup>75</sup>. The variants could cause an increased probability of failure of  
519 myogenesis during embryonic development, consequently resulting in the hernia.

520 The neonatal mortality of probands with *LONP1* deleterious variants is much higher than  
521 CDH neonates without *LONP1* deleterious variants or CDH neonates with likely damaging  
522 variants in genes other than *LONP1*. CDH neonates with *LONP1* deleterious variants  
523 frequently required ECMO. In mice with *Lonp1* knock out at the onset of lung development,  
524 100% newborn pups died shortly after birth, with severe pulmonary defects. Thus, *LONP1*

525 could represent a class of CDH genes with high mortality due to primary developmental  
526 effects on the lung, resulting in more severe pulmonary defects than would occur secondary  
527 to lung compression by herniated abdominal viscera alone. This suggests that we should try  
528 to differentiate primary from secondary developmental effects on the lung as we phenotype  
529 newborns with CDH and as we investigate the mechanisms action of CDH candidate genes.

530 The RNA-binding protein ALYREF plays a key role in nuclear export through binding to the  
531 5' and the 3' regions of mRNA<sup>76,77</sup>. It acts as an RNA 5-methylcytosine (m<sup>5</sup>C) adaptor to  
532 regulate the m<sup>5</sup>C modification<sup>78,79</sup>. Disruption of ALYREF could affect the m<sup>5</sup>C  
533 modification, resulting in abnormal cell proliferation and migration<sup>79</sup>. Previous studies<sup>50</sup>  
534 identified several RNA binding proteins (RBPs) playing essential roles in autism and  
535 congenital birth defects including CHD. RBFOX2, an RBP that regulates alternative splicing,  
536 is critical for zebrafish heart development<sup>80</sup> and *de novo* variants in *RBFOX2* are associated  
537 with congenital heart defects<sup>50</sup>. Dozens of RBPs have established roles in autism spectrum  
538 disorder. RBFOX1<sup>81,82</sup>, an RNA splicing factor, regulates expression of large genetic  
539 networks during early neuronal development including autism. The other RBPs such as  
540 FMRP<sup>83</sup>, CELF4, CELF6<sup>84</sup>, have also been implicated in autism. As an RBP, ALYREF may  
541 play a similar role in congenital anomalies and neurodevelopmental disorders. Two *de novo*  
542 LGDs in *ALYREF* were identified in our CDH cohort. One had an isolated CDH and the other  
543 had CDH and a ventricular septal defect. Similarly, two CDH cases carried *de novo* variants  
544 in *SYMPK*, another RBP identified with FDR<0.1 in extTADA. One had a *de novo* predicted  
545 deleterious missense variant and isolated CDH and the other had a *de novo* LGD with  
546 complex CDH with congenital heart disease, central nervous system anomaly, and  
547 genitourinary anomaly.

548 We found further support for the previously reported CDH genes *ZFPM2* and *MYRF*. We  
549 have identified six ultra-rare LGD variants in *ZFPM2* in our CDH cohort, accounting for  
550 0.7% of our cases (Figure S6). Three were complex cases, all with minor cardiac  
551 malformations. Specifically, two females had atrial septal defects and 1 male had an  
552 enlarged aortic root. The other three heterozygotes had isolated CDH. *ZFPM2* is expressed  
553 in the septum transversum of the diaphragm during early development, and *Fog2*<sup>-/-</sup> mice  
554 generated through chemical mutagenesis have been shown to have diaphragmatic  
555 eventration and pulmonary hypoplasia<sup>55</sup>. *ZFPM2* physically interacts with *NR2F2*<sup>85</sup> and  
556 *GATA4*<sup>86</sup>, two other components of the retinoid signaling pathway implicated in diaphragm  
557 and lung development<sup>87</sup>. Our results further support the pleiotropic role of *ZFPM2* in the  
558 development of CDH.

559 *MYRF* was implicated in our previous *de novo* variant report<sup>9</sup> as a gene for cardiac-urogenital  
560 syndrome (MIM: 618280), and we identified one more additional *de novo* variant in this  
561 cohort (Figure S7). There are now more than 10 variants implicated in CDH with additional  
562 anomalies (HGMD® professional 2021.1). *MYRF* is highly expressed in epithelial cells.  
563 Diaphragm is composed of epithelial-like mesothelial cells derived from the mesoderm of the  
564 pleuroperitoneal folds (PPFs) through cell proliferation, migration, and epithelial-to-  
565 mesenchymal transition<sup>88</sup>. Single cell analysis<sup>89</sup> in fetal gonads suggests the cells that highly  
566 express *MYRF* also express *WT1* and *NR2F2*, two genes associated with diaphragmatic  
567 hernia. Previously, we also demonstrated<sup>9</sup> that individuals with pathogenic variants in *MYRF*  
568 have decreased expression of *GATA4*. *WT1*, *NR2F2* and *GATA4* are all important in RA  
569 signaling in the developing diaphragm<sup>1</sup>. Therefore, the damaging variants in *MYRF* may  
570 affect the RA signaling pathway, leading to diaphragmatic hernia and other anomalies.

571 Among the 734 CDH trios with WGS data, we identified a total of 87 *de novo* CNVs and 4 of  
572 them are recurrent genes or CNVs. Given the rarity of *de novo* CNVs and small sample size,  
573 there were limited data to analyze the differential burden between cases and controls in this  
574 study. Future studies with larger sample sizes will improve the power to analyze CNVs and  
575 structural variants in CDH.

576 In summary, our analysis of *de novo* and ultra-rare inherited variants identified two new CDH  
577 candidate genes *LONP1* and *ALYREF* and confirmed previous associations of *MYRF* and  
578 *ZFPM2* with CDH. The identification of specific highly risk genes would enhance prenatal or  
579 early postnatal counseling and decision making, especially with rapid turnaround of WGS or  
580 exome sequencing results. It is likely that transmitted rare variants also contribute to other  
581 cases in our cohort, but we require a larger sample size to identify these genes confidently.  
582 Future studies will also leverage data from other developmental disorders and integrating  
583 genomic data during development.

#### 584 **Supplemental Data**

585 Supplemental Data include notes, 7 figures and 7 tables.

#### 586 **Acknowledgements**

587 We would like to thank the patients and their families for their generous contribution. We are  
588 grateful for the technical assistance provided by Na Zhu, Patricia Lanzano, Jiangyuan Hu,  
589 Jiancheng Guo, Suying Bao, Charles LeDuc, Liyong Deng, Donna Garey, and Anketil Abreu  
590 from Columbia University, Jennifer Lyu at Boston Children's Hospital, and Caroline Coletti  
591 at Massachusetts General Hospital. We thank our clinical coordinators across the DHREAMS  
592 centers: Jessica Conway at Washington University School of Medicine, Melissa Reed,  
593 Elizabeth Erickson, and Madeline Peters at Cincinnati Children's Hospital, Sheila Horak and

594 Evan Roberts at Children's Hospital & Medical Center of Omaha, Jeannie Kreutzman and  
595 Irene St. Charles at CS Mott Children's Hospital, Tracy Perry at Monroe Carell Jr. Children's  
596 Hospital, Dr. Michelle Kallis at Northwell Health, Andrew Mason and Alicia McIntire at  
597 Oregon Health and Science University, Gentry Wools and Lorrie Burkhalter at Children's  
598 Medical Center Dallas, Elizabeth Jehle at Hassenfeld Children's Hospital, Michelle  
599 Knezevich and Cheryl Kornberg at Medical College of Wisconsin, Min Shi at Children's  
600 Hospital of Pittsburgh. We would also like to acknowledge Terry Buchmiller at Boston  
601 Children's Hospital, and the other pediatric surgeons and clinicians who referred patients to  
602 our studies.

603 The whole genome sequencing data were generated through NIH Gabriella Miller Kids First  
604 Pediatric Research Program (X01HL132366, X01HL136998, X01HL155060). This work  
605 was supported by NIH grants R01HD057036 (L.Y., J.W., W.K.C.), R03HL138352 (A.K.,  
606 W.K.C., Y.S.), R01GM120609 (H.Q., Y.S.), UL1 RR024156 (W.K.C.) 1P01HD068250  
607 (P.K.D, F.A.H., J.M.W., W.K.C., Y.S., J.M.Z, D.J.M, X.S.) and NSFC81501295 (L.Y.).  
608 Additional funding support was provided by grants from CHERUBS, CDHUK, and the  
609 National Greek Orthodox Ladies Philoptochos Society, Inc. and generous donations from the  
610 Williams Family, Wheeler Foundation, Vanech Family Foundation, Larsen Family, Wilke  
611 Family and many other families. Whole genome sequencing data can be obtained from  
612 dbGAP through accession phs001110. WHICAP study is supported by funding from NIA  
613 RF1AG054023 (B.N.V.). Biogen Inc provided support for whole-exome sequencing for the  
614 WHICAP cohort.

#### 615 **Declaration of Interests**

616 The authors declare no competing interests.

617 **Web Resources**

618 DHREAMS study, <http://www.cdhgenetics.com/>

619 Integrative Genome Viewer (IGV), <http://software.broadinstitute.org/software/igv>

620 ClinGen genome dosage map, <https://dosage.clinicalgenome.org>

621 DECIPHER, <https://www.deciphergenomics.org>

622 Combined Annotation Dependent Depletion (CADD), <https://cadd.gs.washington.edu/>

623 GenBank, <https://www.ncbi.nlm.nih.gov/genbank/>

624 Genome Aggregation Database (gnomAD), <https://gnomad.broadinstitute.org/>

625 Online Mendelian Inheritance in Man (OMIM), <https://www.omim.org/>

626 PyMOL molecular viewer, <https://pymol.org/2/>

627 Mouse Genome Informatics (MGI), <http://www.informatics.jax.org>

628 The Human Protein Atlas, <https://www.proteinatlas.org/>

629 **Reference**

- 630 1. Yu, L., Hernan, R.R., Wynn, J., and Chung, W.K. (2020). The influence of genetics in  
631 congenital diaphragmatic hernia. *Semin Perinatol* 44, 151169.  
632 10.1053/j.semperi.2019.07.008.
- 633 2. Kardon, G., Ackerman, K.G., McCulley, D.J., Shen, Y., Wynn, J., Shang, L.,  
634 Bogenschutz, E., Sun, X., and Chung, W.K. (2017). Congenital diaphragmatic  
635 hernias: from genes to mechanisms to therapies. *Dis Model Mech* 10, 955-970.  
636 10.1242/dmm.028365.
- 637 3. Qiao, L., Wynn, J., Yu, L., Hernan, R., Zhou, X., Duron, V., Aspelund, G., Farkouh-  
638 Karoleski, C., Zygmunt, A., Krishnan, U.S., et al. (2020). Likely damaging de novo  
639 variants in congenital diaphragmatic hernia patients are associated with worse clinical  
640 outcomes. *Genet Med* 22, 2020-2028. 10.1038/s41436-020-0908-0.

- 641 4. Montalva, L., Lauriti, G., and Zani, A. (2019). Congenital heart disease associated  
642 with congenital diaphragmatic hernia: A systematic review on incidence, prenatal  
643 diagnosis, management, and outcome. *J Pediatr Surg* 54, 909-919.  
644 10.1016/j.jpedsurg.2019.01.018.
- 645 5. Lin, A.E., Pober, B.R., and Adata, I. (2007). Congenital diaphragmatic hernia and  
646 associated cardiovascular malformations: type, frequency, and impact on  
647 management. *Am J Med Genet C Semin Med Genet* 145C, 201-216.  
648 10.1002/ajmg.c.30131.
- 649 6. Kosinski, P., and Wielgos, M. (2017). Congenital diaphragmatic hernia: pathogenesis,  
650 prenatal diagnosis and management - literature review. *Ginekol Pol* 88, 24-30.  
651 10.5603/GP.a2017.0005.
- 652 7. Wynn, J., Aspelund, G., Zygmunt, A., Stolar, C.J., Mychaliska, G., Butcher, J., Lim,  
653 F.Y., Gratton, T., Potoka, D., Brennan, K., et al. (2013). Developmental outcomes of  
654 children with congenital diaphragmatic hernia: a multicenter prospective study. *J*  
655 *Pediatr Surg* 48, 1995-2004. 10.1016/j.jpedsurg.2013.02.041.
- 656 8. Wynn, J., Krishnan, U., Aspelund, G., Zhang, Y., Duong, J., Stolar, C.J., Hahn, E.,  
657 Pietsch, J., Chung, D., Moore, D., et al. (2013). Outcomes of congenital  
658 diaphragmatic hernia in the modern era of management. *J Pediatr* 163, 114-119 e111.  
659 10.1016/j.jpeds.2012.12.036.
- 660 9. Qi, H., Yu, L., Zhou, X., Wynn, J., Zhao, H., Guo, Y., Zhu, N., Kitaygorodsky, A.,  
661 Hernan, R., Aspelund, G., et al. (2018). De novo variants in congenital diaphragmatic  
662 hernia identify MYRF as a new syndrome and reveal genetic overlaps with other  
663 developmental disorders. *PLoS Genet* 14, e1007822. 10.1371/journal.pgen.1007822.
- 664 10. Bogenschutz, E.L., Fox, Z.D., Farrell, A., Wynn, J., Moore, B., Yu, L., Aspelund, G.,  
665 Marth, G., Yandell, M., Shen, Y., et al. (2020). Deep whole-genome sequencing of  
666 multiple proband tissues and parental blood reveals the complex genetic etiology of  
667 congenital diaphragmatic hernias. *HGG Adv* 1. 10.1016/j.xhgg.2020.100008.
- 668 11. Pober, B.R., Lin, A., Russell, M., Ackerman, K.G., Chakravorty, S., Strauss, B.,  
669 Westgate, M.N., Wilson, J., Donahoe, P.K., and Holmes, L.B. (2005). Infants with  
670 Bochdalek diaphragmatic hernia: sibling precurrence and monozygotic twin  
671 discordance in a hospital-based malformation surveillance program. *Am J Med Genet*  
672 *A* 138A, 81-88. 10.1002/ajmg.a.30904.
- 673 12. Yu, L., Wynn, J., Cheung, Y.H., Shen, Y., Mychaliska, G.B., Crombleholme, T.M.,  
674 Azarow, K.S., Lim, F.Y., Chung, D.H., Potoka, D., et al. (2013). Variants in GATA4  
675 are a rare cause of familial and sporadic congenital diaphragmatic hernia. *Hum Genet*  
676 *132*, 285-292. 10.1007/s00439-012-1249-0.
- 677 13. Kantarci, S., Al-Gazali, L., Hill, R.S., Donnai, D., Black, G.C., Bieth, E., Chassaing,  
678 N., Lacombe, D., Devriendt, K., Teebi, A., et al. (2007). Mutations in LRP2, which  
679 encodes the multiligand receptor megalin, cause Donnai-Barrow and facio-oculo-  
680 acoustico-renal syndromes. *Nat Genet* 39, 957-959. 10.1038/ng2063.
- 681 14. Longoni, M., High, F.A., Qi, H., Joy, M.P., Hila, R., Coletti, C.M., Wynn, J.,  
682 Loscertales, M., Shan, L., Bult, C.J., et al. (2017). Genome-wide enrichment of  
683 damaging de novo variants in patients with isolated and complex congenital  
684 diaphragmatic hernia. *Hum Genet* 136, 679-691. 10.1007/s00439-017-1774-y.
- 685 15. Yu, L., Sawle, A.D., Wynn, J., Aspelund, G., Stolar, C.J., Arkovitz, M.S., Potoka, D.,  
686 Azarow, K.S., Mychaliska, G.B., Shen, Y., and Chung, W.K. (2015). Increased  
687 burden of de novo predicted deleterious variants in complex congenital diaphragmatic  
688 hernia. *Hum Mol Genet* 24, 4764-4773. 10.1093/hmg/ddv196.
- 689 16. Yu, L., Wynn, J., Ma, L., Guha, S., Mychaliska, G.B., Crombleholme, T.M., Azarow,  
690 K.S., Lim, F.Y., Chung, D.H., Potoka, D., et al. (2012). De novo copy number

- 691 variants are associated with congenital diaphragmatic hernia. *J Med Genet* *49*, 650-  
692 659. [10.1136/jmedgenet-2012-101135](https://doi.org/10.1136/jmedgenet-2012-101135).
- 693 17. Harris, P.A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., and Conde, J.G. (2009).  
694 Research electronic data capture (REDCap)--a metadata-driven methodology and  
695 workflow process for providing translational research informatics support. *J Biomed*  
696 *Inform* *42*, 377-381. [10.1016/j.jbi.2008.08.010](https://doi.org/10.1016/j.jbi.2008.08.010).
- 697 18. Feliciano, P., Zhou, X., Astrovskaya, I., Turner, T.N., Wang, T., Brueggeman, L.,  
698 Barnard, R., Hsieh, A., Snyder, L.G., Muzny, D.M., et al. (2019). Exome sequencing  
699 of 457 autism families recruited online provides evidence for autism risk genes. *NPJ*  
700 *Genom Med*, 4-19. [10.1038/s41525-019-0093-8](https://doi.org/10.1038/s41525-019-0093-8).
- 701 19. Tang, M.X., Cross, P., Andrews, H., Jacobs, D.M., Small, S., Bell, K., Merchant, C.,  
702 Lantigua, R., Costa, R., Stern, Y., and Mayeux, R. (2001). Incidence of AD in  
703 African-Americans, Caribbean Hispanics, and Caucasians in northern Manhattan.  
704 *Neurology* *56*, 49-56. [10.1212/wnl.56.1.49](https://doi.org/10.1212/wnl.56.1.49).
- 705 20. Van Hout, C.V., Tachmazidou, I., Backman, J.D., Hoffman, J.D., Liu, D., Pandey,  
706 A.K., Gonzaga-Jauregui, C., Khalid, S., Ye, B., Banerjee, N., et al. (2020). Exome  
707 sequencing and characterization of 49,960 individuals in the UK Biobank. *Nature*  
708 *586*, 749-756. [10.1038/s41586-020-2853-0](https://doi.org/10.1038/s41586-020-2853-0).
- 709 21. Raghavan, N.S., Brickman, A.M., Andrews, H., Manly, J.J., Schupf, N., Lantigua, R.,  
710 Wolock, C.J., Kamalakaran, S., Petrovski, S., Tosto, G., et al. (2018). Whole-exome  
711 sequencing in 20,197 persons for rare variants in Alzheimer's disease. *Ann Clin*  
712 *Transl Neurol* *5*, 832-842. [10.1002/acn3.582](https://doi.org/10.1002/acn3.582).
- 713 22. Zhu, N., Swietlik, E.M., Welch, C.L., Pauciulo, M.W., Hagen, J.J., Zhou, X., Guo, Y.,  
714 Karten, J., Pandya, D., Tilly, T., et al. (2021). Rare variant analysis of 4241  
715 pulmonary arterial hypertension cases from an international consortium implicates  
716 FBLN2, PDGFD, and rare de novo variants in PAH. *Genome Med* *13*, 80.  
717 [10.1186/s13073-021-00891-1](https://doi.org/10.1186/s13073-021-00891-1).
- 718 23. Li, H., Ruan, J., and Durbin, R. (2008). Mapping short DNA sequencing reads and  
719 calling variants using mapping quality scores. *Genome Res* *18*, 1851-1858.  
720 [10.1101/gr.078212.108](https://doi.org/10.1101/gr.078212.108).
- 721 24. DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C.,  
722 Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., et al. (2011). A framework  
723 for variation discovery and genotyping using next-generation DNA sequencing data.  
724 *Nat Genet* *43*, 491-498. [10.1038/ng.806](https://doi.org/10.1038/ng.806).
- 725 25. Van der Auwera, G.A., Carneiro, M.O., Hartl, C., Poplin, R., Del Angel, G., Levy-  
726 Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., et al. (2013). From  
727 FastQ data to high confidence variant calls: the Genome Analysis Toolkit best  
728 practices pipeline. *Curr Protoc Bioinformatics* *43*, 11.10.11-11.10.33.  
729 [10.1002/0471250953.bi1110s43](https://doi.org/10.1002/0471250953.bi1110s43).
- 730 26. Zhao, H., Sun, Z., Wang, J., Huang, H., Kocher, J.P., and Wang, L. (2014).  
731 CrossMap: a versatile tool for coordinate conversion between genome assemblies.  
732 *Bioinformatics* *30*, 1006-1007. [10.1093/bioinformatics/btt730](https://doi.org/10.1093/bioinformatics/btt730).
- 733 27. Manichaikul, A., Mychaleckyj, J.C., Rich, S.S., Daly, K., Sale, M., and Chen, W.M.  
734 (2010). Robust relationship inference in genome-wide association studies.  
735 *Bioinformatics* *26*, 2867-2873. [10.1093/bioinformatics/btq559](https://doi.org/10.1093/bioinformatics/btq559).
- 736 28. Pedersen, B.S., and Quinlan, A.R. (2017). Who's Who? Detecting and Resolving  
737 Sample Anomalies in Human DNA Sequencing Studies with Peddy. *Am J Hum*  
738 *Genet* *100*, 406-413. [10.1016/j.ajhg.2017.01.017](https://doi.org/10.1016/j.ajhg.2017.01.017).
- 739 29. Poplin, R., Chang, P.C., Alexander, D., Schwartz, S., Colthurst, T., Ku, A.,  
740 Newburger, D., Dijamco, J., Nguyen, N., Afshar, P.T., et al. (2018). A universal SNP

- 741 and small-indel variant caller using deep neural networks. *Nat Biotechnol* 36, 983-  
742 987. 10.1038/nbt.4235.
- 743 30. Tom, J.A., Reeder, J., Forrest, W.F., Graham, R.R., Hunkapiller, J., Behrens, T.W.,  
744 and Bhangale, T.R. (2017). Identifying and mitigating batch effects in whole genome  
745 sequencing data. *BMC Bioinformatics* 18, 351. 10.1186/s12859-017-1756-z.
- 746 31. McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R., Thormann, A., Flicek, P.,  
747 and Cunningham, F. (2016). The Ensembl Variant Effect Predictor. *Genome Biol* 17,  
748 122. 10.1186/s13059-016-0974-4.
- 749 32. Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of  
750 genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 38, e164.  
751 10.1093/nar/gkq603.
- 752 33. Kircher, M., Witten, D.M., Jain, P., O'Roak, B.J., Cooper, G.M., and Shendure, J.  
753 (2014). A general framework for estimating the relative pathogenicity of human  
754 genetic variants. *Nat Genet* 46, 310-315. 10.1038/ng.2892.
- 755 34. Russell, M.K., Longoni, M., Wells, J., Maalouf, F.I., Tracy, A.A., Loscertales, M.,  
756 Ackerman, K.G., Pober, B.R., Lage, K., Bult, C.J., and Donahoe, P.K. (2012).  
757 Congenital diaphragmatic hernia candidate genes derived from embryonic  
758 transcriptomes. *Proc Natl Acad Sci U S A* 109, 2978-2983.  
759 10.1073/pnas.1121621109.
- 760 35. Rehm, H.L., Berg, J.S., Brooks, L.D., Bustamante, C.D., Evans, J.P., Landrum, M.J.,  
761 Ledbetter, D.H., Maglott, D.R., Martin, C.L., Nussbaum, R.L., et al. (2015). ClinGen-  
762 the Clinical Genome Resource. *N Engl J Med* 372, 2235-2242.  
763 10.1056/NEJMSr1406261.
- 764 36. Abyzov, A., Urban, A.E., Snyder, M., and Gerstein, M. (2011). CNVnator: an  
765 approach to discover, genotype, and characterize typical and atypical CNVs from  
766 family and population genome sequencing. *Genome Res* 21, 974-984.  
767 10.1101/gr.114876.110.
- 768 37. Layer, R.M., Chiang, C., Quinlan, A.R., and Hall, I.M. (2014). LUMPY: a  
769 probabilistic framework for structural variant discovery. *Genome Biol* 15, R84.  
770 10.1186/gb-2014-15-6-r84.
- 771 38. Chiang, C., Layer, R.M., Faust, G.G., Lindberg, M.R., Rose, D.B., Garrison, E.P.,  
772 Marth, G.T., Quinlan, A.R., and Hall, I.M. (2015). SpeedSeq: ultra-fast personal  
773 genome analysis and interpretation. *Nat Methods* 12, 966-968. 10.1038/nmeth.3505.
- 774 39. Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T.,  
775 O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., et al. (2016). Analysis  
776 of protein-coding genetic variation in 60,706 humans. *Nature* 536, 285-291.  
777 10.1038/nature19057.
- 778 40. Teschendorff, A.E., Zhu, T., Breeze, C.E., and Beck, S. (2020). EPISCORE: cell type  
779 deconvolution of bulk tissue DNA methylomes from single-cell RNA-Seq data.  
780 *Genome Biol* 21, 221. 10.1186/s13059-020-02126-9.
- 781 41. Firth, H.V., Richards, S.M., Bevan, A.P., Clayton, S., Corpas, M., Rajan, D., Van  
782 Vooren, S., Moreau, Y., Pettett, R.M., and Carter, N.P. (2009). DECIPHER: Database  
783 of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources.  
784 *Am J Hum Genet* 84, 524-533. 10.1016/j.ajhg.2009.03.010.
- 785 42. Samocha, K.E., Robinson, E.B., Sanders, S.J., Stevens, C., Sabo, A., McGrath, L.M.,  
786 Kosmicki, J.A., Rehnstrom, K., Mallick, S., Kirby, A., et al. (2014). A framework for  
787 the interpretation of de novo mutation in human disease. *Nat Genet* 46, 944-950.  
788 10.1038/ng.3050.
- 789 43. Nguyen, H.T., Bryois, J., Kim, A., Dobbyn, A., Huckins, L.M., Munoz-Manchado,  
790 A.B., Ruderfer, D.M., Genovese, G., Fromer, M., Xu, X., et al. (2017). Integrated

- 791 Bayesian analysis of rare exonic variants to identify risk genes for schizophrenia and  
792 neurodevelopmental disorders. *Genome Med* 9, 114. 10.1186/s13073-017-0497-y.
- 793 44. He, X., Sanders, S.J., Liu, L., De Rubeis, S., Lim, E.T., Sutcliffe, J.S., Schellenberg,  
794 G.D., Gibbs, R.A., Daly, M.J., Buxbaum, J.D., et al. (2013). Integrated model of de  
795 novo and inherited genetic variants yields greater power to identify risk genes. *PLoS*  
796 *Genet* 9, e1003671. 10.1371/journal.pgen.1003671.
- 797 45. Price, A.L., Kryukov, G.V., de Bakker, P.I., Purcell, S.M., Staples, J., Wei, L.J., and  
798 Sunyaev, S.R. (2010). Pooled association tests for rare variants in exon-resequencing  
799 studies. *Am J Hum Genet* 86, 832-838. 10.1016/j.ajhg.2010.04.005.
- 800 46. Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R.,  
801 Heer, F.T., de Beer, T.A.P., Rempfer, C., Bordoli, L., et al. (2018). SWISS-MODEL:  
802 homology modelling of protein structures and complexes. *Nucleic Acids Res* 46,  
803 W296-W303. 10.1093/nar/gky427.
- 804 47. Harris, K.S., Zhang, Z., McManus, M.T., Harfe, B.D., and Sun, X. (2006). Dicer  
805 function is essential for lung epithelium morphogenesis. *Proc Natl Acad Sci U S A*  
806 103, 2208-2213. 10.1073/pnas.0510839103.
- 807 48. Hinton, C.F., Siffel, C., Correa, A., and Shapira, S.K. (2017). Survival Disparities  
808 Associated with Congenital Diaphragmatic Hernia. *Birth Defects Res* 109, 816-823.  
809 10.1002/bdr2.1015.
- 810 49. Leeuwen, L., Mous, D.S., van Rosmalen, J., Olieman, J.F., Andriessen, L., Gischler,  
811 S.J., Joosten, K.F.M., Wijnen, R.M.H., Tibboel, D., H, I.J., and Spoel, M. (2017).  
812 Congenital Diaphragmatic Hernia and Growth to 12 Years. *Pediatrics* 140.  
813 10.1542/peds.2016-3659.
- 814 50. Homsy, J., Zaidi, S., Shen, Y., Ware, J.S., Samocha, K.E., Karczewski, K.J.,  
815 DePalma, S.R., McKean, D., Wakimoto, H., Gorham, J., et al. (2015). De novo  
816 mutations in congenital heart disease with neurodevelopmental and other congenital  
817 anomalies. *Science* 350, 1262-1266. 10.1126/science.aac9396.
- 818 51. Jin, S.C., Homsy, J., Zaidi, S., Lu, Q., Morton, S., DePalma, S.R., Zeng, X., Qi, H.,  
819 Chang, W., Sierant, M.C., et al. (2017). Contribution of rare inherited and de novo  
820 variants in 2,871 congenital heart disease probands. *Nat Genet* 49, 1593-1601.  
821 10.1038/ng.3970.
- 822 52. Satterstrom, F.K., Kosmicki, J.A., Wang, J., Breen, M.S., De Rubeis, S., An, J.Y.,  
823 Peng, M., Collins, R., Grove, J., Klei, L., et al. (2020). Large-Scale Exome  
824 Sequencing Study Implicates Both Developmental and Functional Changes in the  
825 Neurobiology of Autism. *Cell* 180, 568-584 e523. 10.1016/j.cell.2019.12.036.
- 826 53. Strauss, K.A., Jinks, R.N., Puffenberger, E.G., Venkatesh, S., Singh, K., Cheng, I.,  
827 Mikita, N., Thilagavathi, J., Lee, J., Sarafianos, S., et al. (2015). CODAS syndrome is  
828 associated with mutations of LONP1, encoding mitochondrial AAA+ Lon protease.  
829 *Am J Hum Genet* 96, 121-135. 10.1016/j.ajhg.2014.12.003.
- 830 54. Shebib, S.M., Reed, M.H., Shuckett, E.P., Cross, H.G., Perry, J.B., and Chudley, A.E.  
831 (1991). Newly recognized syndrome of cerebral, ocular, dental, auricular, skeletal  
832 anomalies: CODAS syndrome--a case report. *Am J Med Genet* 40, 88-93.  
833 10.1002/ajmg.1320400118.
- 834 55. Ackerman, K.G., Herron, B.J., Vargas, S.O., Huang, H., Tevosian, S.G., Kochilas, L.,  
835 Rao, C., Pober, B.R., Babiuk, R.P., Epstein, J.A., et al. (2005). Fog2 is required for  
836 normal diaphragm and lung development in mice and humans. *PLoS Genet* 1, 58-65.  
837 10.1371/journal.pgen.0010010.
- 838 56. Bleyl, S.B., Moshrefi, A., Shaw, G.M., Sajjoh, Y., Schoenwolf, G.C., Pennacchio,  
839 L.A., and Slavotinek, A.M. (2007). Candidate genes for congenital diaphragmatic  
840 hernia from animal models: sequencing of FOG2 and PDGFRalpha reveals rare

- 841 variants in diaphragmatic hernia patients. *Eur J Hum Genet* 15, 950-958.  
842 10.1038/sj.ejhg.5201872.
- 843 57. Iossifov, I., O'Roak, B.J., Sanders, S.J., Ronemus, M., Krumm, N., Levy, D.,  
844 Stessman, H.A., Witherspoon, K.T., Vives, L., Patterson, K.E., et al. (2014). The  
845 contribution of de novo coding mutations to autism spectrum disorder. *Nature* 515,  
846 216-221. 10.1038/nature13908.
- 847 58. Kosmicki, J.A., Samocha, K.E., Howrigan, D.P., Sanders, S.J., Slowikowski, K., Lek,  
848 M., Karczewski, K.J., Cutler, D.J., Devlin, B., Roeder, K., et al. (2017). Refining the  
849 role of de novo protein-truncating variants in neurodevelopmental disorders by using  
850 population reference samples. *Nat Genet* 49, 504-510. 10.1038/ng.3789.
- 851 59. Jacquemont, S., Coe, B.P., Hersch, M., Duyzend, M.H., Krumm, N., Bergmann, S.,  
852 Beckmann, J.S., Rosenfeld, J.A., and Eichler, E.E. (2014). A higher mutational  
853 burden in females supports a "female protective model" in neurodevelopmental  
854 disorders. *Am J Hum Genet* 94, 415-425. 10.1016/j.ajhg.2014.02.001.
- 855 60. Wang, B., Ji, T., Zhou, X., Wang, J., Wang, X., Wang, J., Zhu, D., Zhang, X., Sham,  
856 P.C., Zhang, X., et al. (2016). CNV analysis in Chinese children of mental retardation  
857 highlights a sex differentiation in parental contribution to de novo and inherited  
858 mutational burdens. *Sci Rep* 6, 25954. 10.1038/srep25954.
- 859 61. Gibellini, L., De Gaetano, A., Mandrioli, M., Van Tongeren, E., Bortolotti, C.A.,  
860 Cossarizza, A., and Pinti, M. (2020). The biology of Lonp1: More than a  
861 mitochondrial protease. *Int Rev Cell Mol Biol* 354, 1-61.  
862 10.1016/bs.ircmb.2020.02.005.
- 863 62. Quiros, P.M., Espanol, Y., Acin-Perez, R., Rodriguez, F., Barcena, C., Watanabe, K.,  
864 Calvo, E., Loureiro, M., Fernandez-Garcia, M.S., Fueyo, A., et al. (2014). ATP-  
865 dependent Lon protease controls tumor bioenergetics by reprogramming  
866 mitochondrial activity. *Cell Rep* 8, 542-556. 10.1016/j.celrep.2014.06.018.
- 867 63. Matsushima, Y., Goto, Y., and Kaguni, L.S. (2010). Mitochondrial Lon protease  
868 regulates mitochondrial DNA copy number and transcription by selective degradation  
869 of mitochondrial transcription factor A (TFAM). *Proc Natl Acad Sci U S A* 107,  
870 18410-18415. 10.1073/pnas.1008924107.
- 871 64. Gur, E., and Sauer, R.T. (2008). Recognition of misfolded proteins by Lon, a AAA(+)   
872 protease. *Genes Dev* 22, 2267-2277. 10.1101/gad.1670908.
- 873 65. He, L., Luo, D., Yang, F., Li, C., Zhang, X., Deng, H., and Zhang, J.R. (2018).  
874 Multiple domains of bacterial and human Lon proteases define substrate selectivity.  
875 *Emerg Microbes Infect* 7, 149. 10.1038/s41426-018-0148-4.
- 876 66. Mikita, N., Cheng, I., Fishovitz, J., Huang, J., and Lee, I. (2013). Processive  
877 degradation of unstructured protein by *Escherichia coli* Lon occurs via the slow,  
878 sequential delivery of multiple scissile sites followed by rapid and synchronized  
879 peptide bond cleavage events. *Biochemistry* 52, 5629-5644. 10.1021/bi4008319.
- 880 67. Flodby, P., Li, C., Liu, Y., Wang, H., Marconett, C.N., Laird-Offringa, I.A., Minoo,  
881 P., Lee, A.S., and Zhou, B. (2016). The 78-kD Glucose-Regulated Protein Regulates  
882 Endoplasmic Reticulum Homeostasis and Distal Epithelial Cell Survival during Lung  
883 Development. *Am J Respir Cell Mol Biol* 55, 135-149. 10.1165/rcmb.2015-0327OC.
- 884 68. Pareek, G., and Pallanck, L.J. (2018). Inactivation of Lon protease reveals a link  
885 between mitochondrial unfolded protein stress and mitochondrial translation  
886 inhibition. *Cell Death Dis* 9, 1168. 10.1038/s41419-018-1213-6.
- 887 69. Shi, H., O'Reilly, V.C., Moreau, J.L., Bewes, T.R., Yam, M.X., Chapman, B.E.,  
888 Grieve, S.M., Stocker, R., Graham, R.M., Chapman, G., et al. (2016). Gestational  
889 stress induces the unfolded protein response, resulting in heart defects. *Development*  
890 143, 2561-2572. 10.1242/dev.136820.

- 891 70. Kao, T.Y., Chiu, Y.C., Fang, W.C., Cheng, C.W., Kuo, C.Y., Juan, H.F., Wu, S.H.,  
892 and Lee, A.Y. (2015). Mitochondrial Lon regulates apoptosis through the association  
893 with Hsp60-mtHsp70 complex. *Cell Death Dis* 6, e1642. 10.1038/cddis.2015.9.
- 894 71. Luciakova, K., Sokolikova, B., Chloupkova, M., and Nelson, B.D. (1999). Enhanced  
895 mitochondrial biogenesis is associated with increased expression of the mitochondrial  
896 ATP-dependent Lon protease. *FEBS Letters* 444, 186-188. 10.1016/s0014-  
897 5793(99)00058-7.
- 898 72. Gibellini, L., Pinti, M., Boraldi, F., Giorgio, V., Bernardi, P., Bartolomeo, R., Nasi,  
899 M., De Biasi, S., Missiroli, S., Carnevale, G., et al. (2014). Silencing of mitochondrial  
900 Lon protease deeply impairs mitochondrial proteome and function in colon cancer  
901 cells. *FASEB J* 28, 5122-5135. 10.1096/fj.14-255869.
- 902 73. Bota, D.A., Ngo, J.K., and Davies, K.J. (2005). Downregulation of the human Lon  
903 protease impairs mitochondrial structure and function and causes cell death. *Free  
904 Radic Biol Med* 38, 665-677. 10.1016/j.freeradbiomed.2004.11.017.
- 905 74. Babiuk, R.P., Zhang, W., Clugston, R., Allan, D.W., and Greer, J.J. (2003).  
906 Embryological origins and development of the rat diaphragm. *J Comp Neurol* 455,  
907 477-487. 10.1002/cne.10503.
- 908 75. Bota, D.A., and Davies, K.J. (2016). Mitochondrial Lon protease in human disease  
909 and aging: Including an etiologic classification of Lon-related diseases and disorders.  
910 *Free Radic Biol Med* 100, 188-198. 10.1016/j.freeradbiomed.2016.06.031.
- 911 76. Shi, M., Zhang, H., Wu, X., He, Z., Wang, L., Yin, S., Tian, B., Li, G., and Cheng, H.  
912 (2017). ALYREF mainly binds to the 5' and the 3' regions of the mRNA in vivo.  
913 *Nucleic Acids Res* 45, 9640-9653. 10.1093/nar/gkx597.
- 914 77. Fan, J., Wang, K., Du, X., Wang, J., Chen, S., Wang, Y., Shi, M., Zhang, L., Wu, X.,  
915 Zheng, D., et al. (2019). ALYREF links 3'-end processing to nuclear export of non-  
916 polyadenylated mRNAs. *EMBO J* 38. 10.15252/embj.201899910.
- 917 78. Yang, X., Yang, Y., Sun, B.F., Chen, Y.S., Xu, J.W., Lai, W.Y., Li, A., Wang, X.,  
918 Bhattarai, D.P., Xiao, W., et al. (2017). 5-methylcytosine promotes mRNA export -  
919 NSUN2 as the methyltransferase and ALYREF as an m(5)C reader. *Cell Res* 27, 606-  
920 625. 10.1038/cr.2017.55.
- 921 79. Chen, Y.S., Yang, W.L., Zhao, Y.L., and Yang, Y.G. (2021). Dynamic transcriptomic  
922 m(5) C and its regulatory role in RNA processing. *Wiley Interdiscip Rev RNA*,  
923 e1639. 10.1002/wrna.1639.
- 924 80. Gallagher, T.L., Arribere, J.A., Geurts, P.A., Exner, C.R., McDonald, K.L., Dill,  
925 K.K., Marr, H.L., Adkar, S.S., Garnett, A.T., Amacher, S.L., and Conboy, J.G.  
926 (2011). Rbfox-regulated alternative splicing is critical for zebrafish cardiac and  
927 skeletal muscle functions. *Dev Biol* 359, 251-261. 10.1016/j.ydbio.2011.08.025.
- 928 81. Bill, B.R., Lowe, J.K., Dybuncio, C.T., and Fogel, B.L. (2013). Orchestration of  
929 neurodevelopmental programs by RBFOX1: implications for autism spectrum  
930 disorder. *Int Rev Neurobiol* 113, 251-267. 10.1016/B978-0-12-418700-9.00008-3.
- 931 82. Lee, J.A., Damianov, A., Lin, C.H., Fontes, M., Parikshak, N.N., Anderson, E.S.,  
932 Geschwind, D.H., Black, D.L., and Martin, K.C. (2016). Cytoplasmic Rbfox1  
933 Regulates the Expression of Synaptic and Autism-Related Genes. *Neuron* 89, 113-  
934 128. 10.1016/j.neuron.2015.11.025.
- 935 83. Fernandez, E., Rajan, N., and Bagni, C. (2013). The FMRP regulon: from targets to  
936 disease convergence. *Front Neurosci* 7, 191. 10.3389/fnins.2013.00191.
- 937 84. Dougherty, J.D., Maloney, S.E., Wozniak, D.F., Rieger, M.A., Sonnenblick, L.,  
938 Coppola, G., Mahieu, N.G., Zhang, J., Cai, J., Patti, G.J., et al. (2013). The disruption  
939 of Celf6, a gene identified by translational profiling of serotonergic neurons, results in

- 940 autism-related behaviors. *J Neurosci* 33, 2732-2753. 10.1523/JNEUROSCI.4762-  
941 12.2013.
- 942 85. Huggins, G.S., Bacani, C.J., Boltax, J., Aikawa, R., and Leiden, J.M. (2001). Friend  
943 of GATA 2 physically interacts with chicken ovalbumin upstream promoter-TF2  
944 (COUP-TF2) and COUP-TF3 and represses COUP-TF2-dependent activation of the  
945 atrial natriuretic factor promoter. *J Biol Chem* 276, 28029-28036.  
946 10.1074/jbc.M103577200.
- 947 86. Svensson, E.C., Tufts, R.L., Polk, C.E., and Leiden, J.M. (1999). Molecular cloning  
948 of FOG-2: a modulator of transcription factor GATA-4 in cardiomyocytes. *Proc Natl*  
949 *Acad Sci U S A* 96, 956-961. 10.1073/pnas.96.3.956.
- 950 87. Goumy, C., Gouas, L., Marceau, G., Coste, K., Veronese, L., Gallot, D., Sapin, V.,  
951 Vago, P., and Tchirkov, A. (2010). Retinoid pathway and congenital diaphragmatic  
952 hernia: hypothesis from the analysis of chromosomal abnormalities. *Fetal Diagn Ther*  
953 28, 129-139. 10.1159/000313331.
- 954 88. Carmona, R., Canete, A., Cano, E., Ariza, L., Rojas, A., and Munoz-Chapuli, R.  
955 (2016). Conditional deletion of WT1 in the septum transversum mesenchyme causes  
956 congenital diaphragmatic hernia in mice. *Elife* 5. 10.7554/eLife.16009.
- 957 89. Hamanaka, K., Takata, A., Uchiyama, Y., Miyatake, S., Miyake, N., Mitsunashi, S.,  
958 Iwama, K., Fujita, A., Imagawa, E., Alkanaq, A.N., et al. (2019). MYRF  
959 haploinsufficiency causes 46,XY and 46,XX disorders of sex development:  
960 bioinformatics consideration. *Hum Mol Genet* 28, 2319-2329. 10.1093/hmg/ddz066.

961

962 **Tables**

963 **Table 1. Clinical summary of 827 CDH probands**

		Number	Percent
Sex	Male	486	58.8%
	Female	341	41.2%
Genetic ancestry	African	31	3.7%
	Latinx	153	18.5%
	European	607	73.4%
	East Asian	15	1.8%
	South Asian	21	2.5%
CDH classification	Isolated	533	64.4%
	Complex	277	33.5%
	Unknown	17	2.1%
CDH side	Left	645	78.0%
	Right	119	14.4%
	Bilateral/Center/Eventration/Other	38	4.6%
	Unknown	25	3.0%
Timing of enrollment	Fetal	53	6.4%
	Neonatal	464	56.1%
	Child	285	34.5%
	Adult	2	0.2%
	Not specified	23	2.8%
Additional anomalies in complex cases (n=277)	Cardiovascular	144	52.0%
	Neurodevelopmental <sup>a</sup>	54	19.5%
	Skeletal	46	16.6%
	Genitourinary	46	16.6%
	Gastrointestinal	42	15.2%
	Pulmonary defects <sup>b</sup>	18	6.5%
	Cleft lip or palate and/or micrognathia	11	4.0%

964 <sup>a</sup>Neurodevelopmental conditions include congenital abnormalities in central nervous system, and developmental  
 965 delay or neuropsychiatric disorders based on the follow-up developmental evaluations.

966 <sup>b</sup>does not include pulmonary hypoplasia or hypertension

967 **Table 2. Top CDH associated genes predicted by pLI-stratified extTADA with  $\geq 2$  *de***  
 968 ***novo* predicted deleterious variant.**

Gene	Gene name	#D-mis	#LGD	PPA	FDR	pLI
<i>MYRF<sup>a</sup></i>	Myelin Regulatory Factor	3	3	1.00	3.97E-06	1
<i>LONP1</i>	Lon Peptidase 1, Mitochondrial	3	0	0.97	0.014	1
<i>ALYREF</i>	Aly/REF Export Factor	0	2	0.93	0.033	0.83
<i>HSD17B10</i>	Hydroxysteroid 17-Beta Dehydrogenase 10	1	1	0.87	0.056	0.89
<i>GATA4<sup>a</sup></i>	GATA Binding Protein 4	1	1	0.86	0.072	0.8
<i>SYMPK</i>	Symplekin	1	1	0.82	0.090	1
<i>PTPN11</i>	Protein Tyrosine Phosphatase Non-Receptor Type 11	2	0	0.79	0.11	1
<i>WT1<sup>a</sup></i>	WT1 Transcription Factor	2	0	0.78	0.12	1
<i>FAM83H</i>	Family With Sequence Similarity 83 Member H	2	0	0.75	0.13	0.89
<i>CACNA1H</i>	Calcium Voltage-Gated Channel Subunit Alpha1 H	2	0	0.63	0.16	0
<i>SEPSECS</i>	Sep (O-Phosphoserine) TRNA:Sec (Selenocysteine) TRNA Synthase	0	2	0.23	0.66	0
<i>ZFYVE26</i>	Zinc Finger FYVE-Type Containing 26	2	0	0.09	0.72	0

969 #D-mis: number of *de novo* D-mis; #LGD: number of *de novo* LGD; PPA: posterior probability of association; FDR: false discovery rate

970 <sup>a</sup>: known CDH risk genes

971 **Table 3. Recurrent genes or regions in *de novo* CNVs**

Recurrent	ID	Cytoband	Start	End	Size(kb)	Type	Known risk CDH/CHD/NDD genes	qPCR confirm
<i>CSMD1</i> (CUB And Sushi Multiple Domains 1)	CDH1162	8p23.2	3846934	4073105	226	DEL	-	Yes
	CDH12-0009	8p23.3p23.1	191301	7355200	7164	DEL	FBXO25	Yes
	CDH863	8p23.3p23.1	200601	7155000	6954	DUP	FBXO25	Yes
<i>GPHN</i> (Gephyrin)	C1235FSL_169	14q23.3	66559001	66630200	71	DEL	GPHN	-
	CDH14-0009	14q23.3	66636783	66668074	31	DEL	GPHN	Yes
17q12	h1237LPLa1	17q12	36441801	37892100	1450	DEL	GGNBP2	Yes
	CDH05-0040	17q12	36442521	37963800	1521	DEL	GGNBP2	Yes
21q	CDH10-0022	21q	13000000	46700000	33700	DUP	SIM2;SON;HMGN1;SIK1;COL6A1;DYRK1A;DSCAM;DIP2A;KCNJ6	-
	CDH10-0038	21q	13188001	46700000	33512	DUP	SIM2;SON;HMGN1;SIK1;COL6A1;DYRK1A;DSCAM;DIP2A;KCNJ6	-
	CDH10-0042	21q	13192901	46684100	33492	DUP	SIM2;SON;HMGN1;SIK1;COL6A1;DYRK1A;DSCAM;DIP2A;KCNJ6	-

972

973 **Table 4. Phenotypes of CDH cases with ultra-rare deleterious variants in *LONPI*.** Deleterious heterozygous variants include LGD and missense with CADD  $\geq 25$  with minor allele frequency (MAF)  $< 1e-5$  across  
974 all the gnomAD v3.0 genomes.

cDNA change	Protein Change	Sample ID	Sex	Genetic ancestry	Inheritance	Family history of other birth defects	Familial CDH	1M PH	3M PH	Vital status	ECMO	Complex	Neuro-related	Other Congenital Anomalies/medical problems
c.296dup	p.S100Qfs*46	01-0794	Female	EUR	paternal	No	No	-	-	Deceased	No	No		No
c.398C>G	p.P133R	01-0672	Female	AFR	paternal	No	Affected sibling (+)	Unk	Unk	Alive	Yes	Yes	No	congenital cataracts
c.398C>G	p.P133R	01-0670	Male	AFR	paternal	No	Affected sibling (+)	-	-	Deceased	Yes	Yes		GI anomaly, GU anomaly
c.629G>A	p.G210E	01-0070	Male	EUR	<i>de novo</i>	No	No	-	-	Deceased	Yes	No		No
c.639-1G>T	p.X213_splice	04-0022	Female	EUR	paternal	Paternal half-brother with idiopathic PH (N/T)	Affected sibling (N/T), Affected paternal grandmother (+, <i>de novo</i> )	Severe	-	Deceased	Yes	No		No
c.792del	p.P264Rfs*5	09-0003	Female	EUR	maternal	Maternal great uncle with suspected cerebral palsy	No	Severe	-	Deceased	Yes	No	Seizures	No
c.851del	p.Q284Hfs*61	1428	Female	EUR	maternal	No	No	Mild	Mild	Alive	No	No	No	short stature
c.1123C>A	p.L375M	04-0045	Male	EUR	paternal	Paternal uncle: neonatal death due to brain abnormality (hydrocephalus?)	No	None	None	Alive	No	No	No	No
c.1262del	p.F421Lfs*87	1733	Female	EUR	maternal	No	Maternal uncle with suspected CDH (N/T)	-	-	Deceased	Yes	Yes	global encephalopathy, seizures	CHD
c.1574C>T	p.P525L	01-1279	Male	EUR	maternal <sup>a</sup>	No	Affected sibling (+)	-	-	Deceased	Yes	No		No, bilateral CDH
c.1624C>T	p.R542*	01-0113	Male	EUR	paternal	Mother with Klippel Feil syndrome, Sprengel deformity of scapula, crossed fused ectopia (kidneys), Arnold Chiari malformation I	No	Severe	Severe	Deceased	Yes	Yes	No	Pyloric stenosis
c.1629delT	p.E543del	04-0077	Female	EUR	maternal	Unknown	Unknown	Severe	-	Deceased	Yes	Yes		CHD
c.1709C>T	p.P570L	04-0031	Female	EUR	unknown (singleton)	Father with residual post axial polydactyly	No	Severe	-	Deceased	Yes	No		No
c.1773G>C	p.E591D	1511	Male	EUR	<i>de novo</i>	No	No	Mild	None	Alive	No	No	No	No
c.1789C>T	p.R597*	01-0582	Male	AMR	unknown (singleton)	No	No	None	None	Deceased	No	Yes		CHD
c.1895-1G>T	p.X632_splice	1449	Female	EUR	maternal	Unknown	Unknown	-	-	Deceased	No	No		
c.1913C>T	p.T638M	01-0057	Female	AMR	<i>de novo</i>	Unknown	Unknown	Severe	Moderate	Alive	No	Yes	No	CHD, PH
c.1913C>T	p.T638M	01-0513	Female	EUR	paternal	No	Affected sibling (N/T), Father with R eventration (+), paternal grandfather with R eventration (+)	-	-	Alive	No	No	No	No
c.2122G>A	p.G708S	04-0025	Male	EUR	paternal	Father with cleft palate	No	Severe	-	Deceased	Yes	No	Seizures	No
c.2122G>A	p.G708S	m1021LEMa	Female	EUR	maternal	No	No	-	-	Deceased	Yes	Yes		CHD
c.2263C>G	p.R755G	01-1279	Male	EUR	paternal <sup>a</sup>	No	Affected sibling (-)	-	-	Deceased	No	No		No, bilateral CDH
c.2263C>G	p.R755G	09-0028	Male	EUR	paternal	Maternal great-aunt with CHD	No	Unk	Unk	Alive	No	Yes	No	CHD
c.2461G>C	p.A821P	03-0008	Male	EUR	maternal	>3rd degree maternal history: unilateral arm agenesis	No	-	-	Deceased	Yes	No		No
c.2719dup	p.V907Gfs*73	01-0732	Female	EUR	paternal	No	Paternal aunt with possible CDH (N/T)	Unk	Unk	Alive	Yes	No	No	No

ECMO = extracorporeal membrane oxygenation, PH = pulmonary hypertension, CHD = congenital heart disease, GU = genitourinary, GI = gastrointestinal

1M PH = pulmonary hypertension status at 1 month, 3M PH = pulmonary hypertension status at third month, - in 1PH and 3PH = deceased before 1 or 3 months

+ positive for familial *LONPI* variant

- negative for familial *LONPI* variant

N/T = not tested for familial *LONPI* variant

<sup>a</sup> cases carried biallelic heterozygous variants

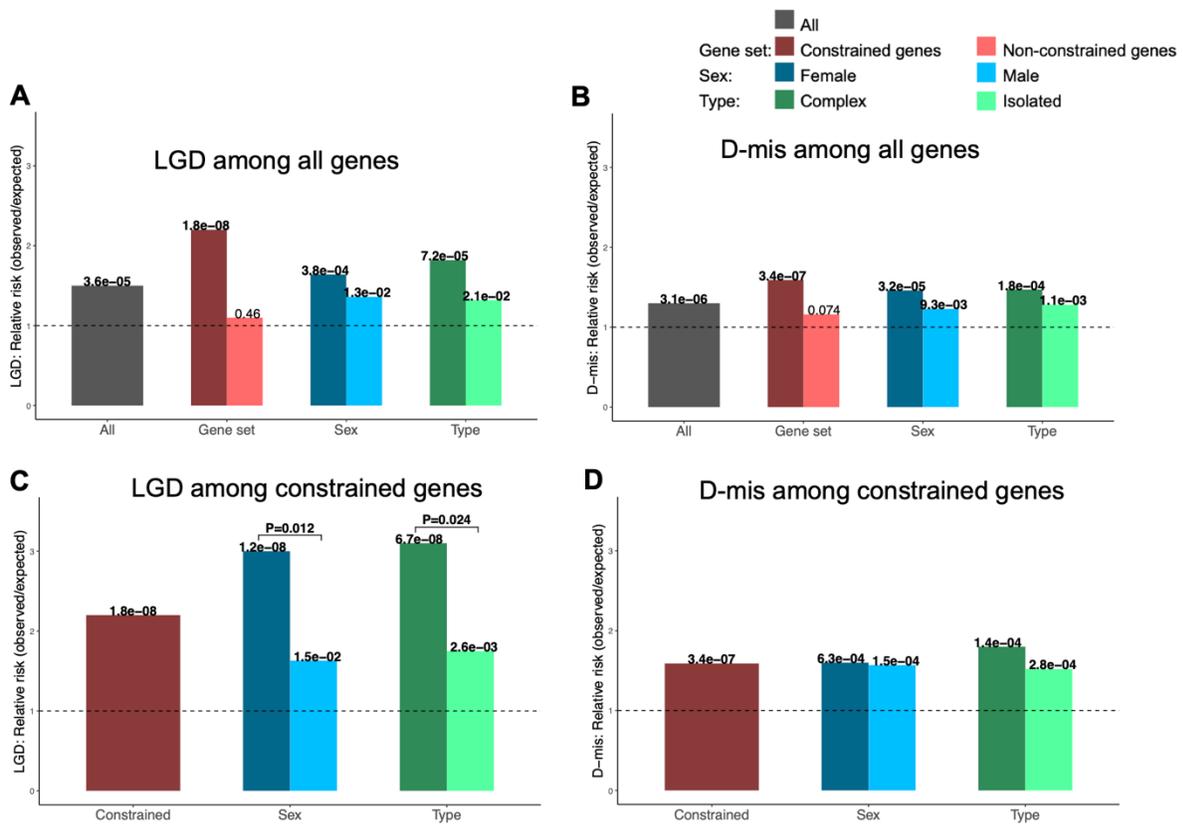
976 **Table 5. *LONPI* deleterious rare variants carriers are associated with higher mortality and**  
 977 **need for ECMO.**

	CDH w/ <i>LONPI</i> deleterious variants (n=23)			CDH w/o <i>LONPI</i> deleterious variants (n=806)			w/ <i>LONPI</i> vs. w/o <i>LONPI</i> deleterious variants	CDH w/ likely damaging variants (n=98)			w/ <i>LONPI</i> deleterious variants vs. w/ likely damaging variants
	Case N	n	%	Control N	n	%	P value	Control N	n	%	P value
<b>Male</b>	23	10	43%	806	477	59%	0.14	98	47	48%	0.82
<b>Complex</b>	23	9	39%	789	269	34%	0.66	96	50	52%	0.35
<b>Familial CDH</b>	19	6	32%	806	61	8%	<b>2.7E-02</b>	98	4	4%	<b>1.2E-03</b>
<b>Neonatal death prior to discharge</b>	16	11	69%	450	72	16%	<b>6.4E-06</b>	55	13	24%	<b>1.8E-03</b>
<b>ECMO</b>	16	9	56%	442	124	28%	<b>2.3E-02</b>	53	16	30%	0.077
<b>PH at 1m</b>	11	7	64%	340	188	55%	0.76	41	29	71%	0.72
<b>PH at 3m</b>	6	2	33%	260	100	39%	1	29	16	55%	0.4

978 The bold p-values highlight significance. ECMO: extracorporeal membrane oxygenation; PH: pulmonary hypertension

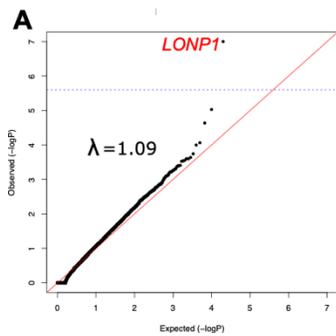
979 **Figures**

980 **Figure 1. Burden of *de novo* coding variants in CDH compared to expectation.** (A) LGD  
 981 among all genes; (B) D-mis among all genes; (C) LGD among constrained genes; (D) D-mis  
 982 among constrained genes. P values between cases and expectation by Poisson test are labeled for  
 983 each bar. P values between females and males, complex and isolated cases by binormal test are  
 984 labeled for each pair. Significant P values are highlighted in bold.



985

986 **Figure 2. Gene-based association analysis using 748 CDH cases and 11,220 controls across**  
 987 **all populations.** (A) Results of a binomial test confined to ultra-rare LGD and D-Mis variants or  
 988 D-Mis only variants in 18,939 protein-coding genes. Horizontal blue line indicates the  
 989 Bonferroni-corrected threshold for significance. (B) Complete list of top association genes with  
 990 permutation P values  $<1 \times 10^{-4}$ . \*: a gene-specific CADD score threshold for defining D-Mis that  
 991 maximized the burden of ultra-rare deleterious variants in cases compared to controls; #:  
 992 numbers of deleterious variants; a: MIM 600539; b: no MIM number.

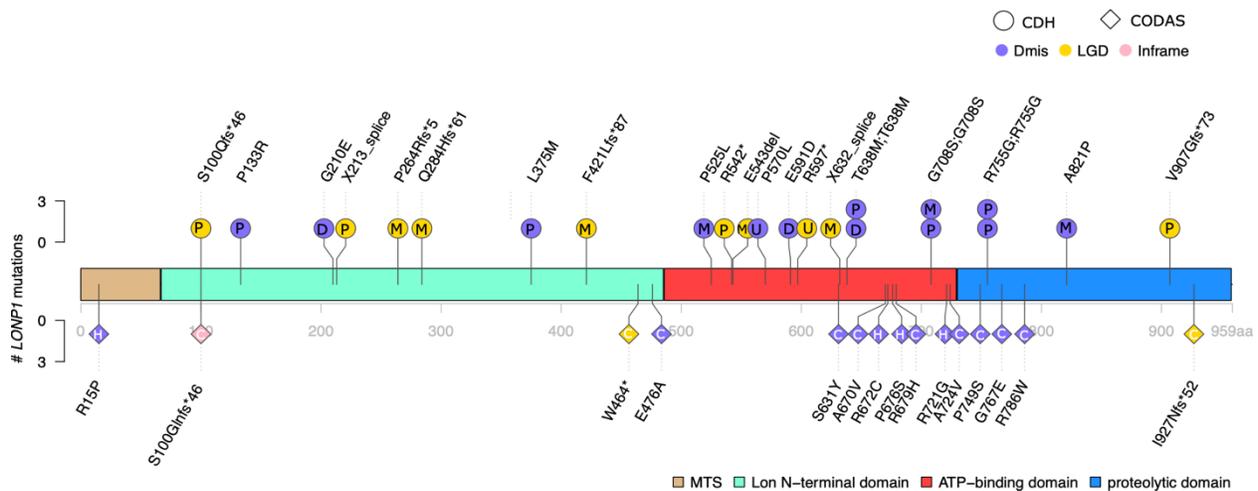


**B**

Gene	Gene name	CADD cut-off*	Case (n=748)		Controls (n=11,220)		Original P	Permutation times	Permutation P	Variant type
			# variants	Rate	# variants	Rate				
<i>LONP1</i>	Lon Peptidase 1, Mitochondrial	23	24	0.032	28	0.0025	9.5E-16	10,000,000	1.00E-07	LGD+D-mis
<i>ZFPM2</i>	Zinc Finger Protein, FOG Family Member 2	33	6	0.008	2	0.0002	1.50E-06	10,000,000	9.30E-06	LGD+D-mis
<i>MYRF</i>	Myelin Regulatory Factor	27	6	0.008	4	0.0004	1.00E-05	10,000,000	2.30E-05	LGD+D-mis
<i>PRKC<sup>a</sup></i>	Protein Kinase C Iota	19	7	0.009	12	0.0011	9.59E-05	1,000,000	8.60E-05	LGD+D-mis
<i>ZNF830<sup>b</sup></i>	Zinc Finger Protein 830	21	4	0.005	1	0.0001	7.25E-05	1,000,000	9.60E-05	D-mis

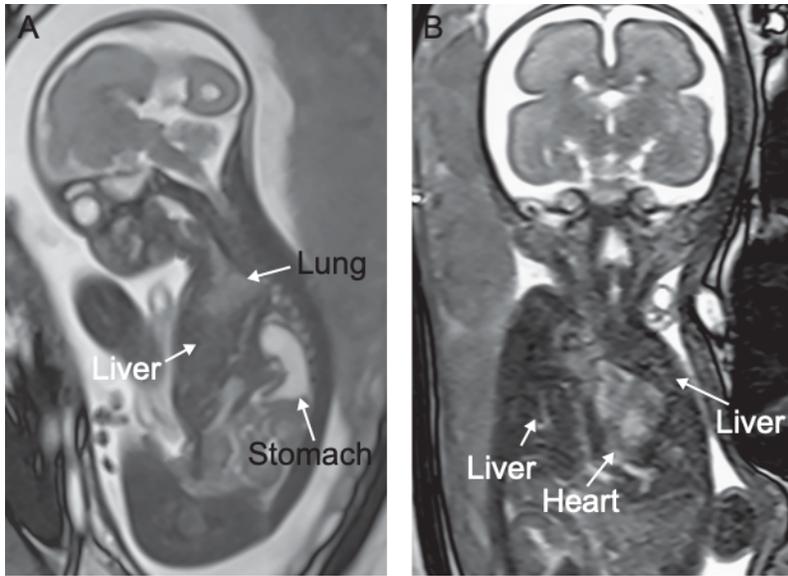
993

994 **Figure 3. Variant locations in *LONP1* (GenBank: NM\_004793.3) of CDH and CODAS**  
 995 **syndrome.** There are three main domains in *LONP1*, N-terminal Lon domain, ATP binding  
 996 domain and proteolytic domain. Positions indicated at upper structure are variants in CDH.  
 997 Deleterious heterozygous variant such as LGD and missense with CADD  $\geq 25$  and allele  
 998 frequency (AF)  $< 1e-5$  across all gnomAD genomes in CDH are presented. Deleterious missense  
 999 is presented in purple, LGD in yellow, inframe variant in pink. Inheritance pattern were labelled  
 1000 in circles of variants (P: paternal; M: maternal; D: *de novo*; U: singleton unknown). Positions at  
 1001 lower structure are variants in published CODAS syndrome samples. CODAS syndrome is  
 1002 caused by biallelic variants in *LONP1*, including homozygous (H) or compound heterozygous  
 1003 (C) variants in the diamonds.



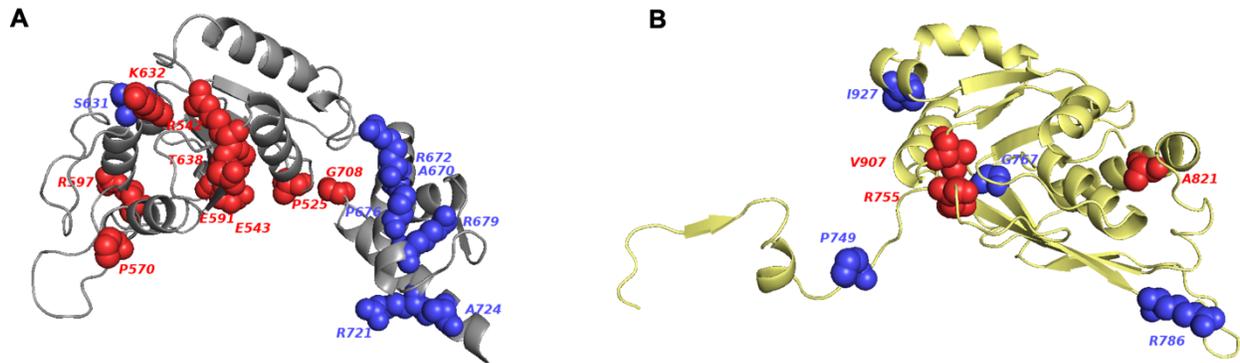
1004

1005 **Figure 4. Fetal MRI images of bilateral CDH.** (A) Sagittal view shows dorsal herniation of the  
1006 stomach, ventral herniation of the liver, and anterior displacement of lung remnant. (B) Coronal  
1007 view shows bilateral herniation of the fetal liver filling both the right and left hemithorax and no  
1008 lung tissue.

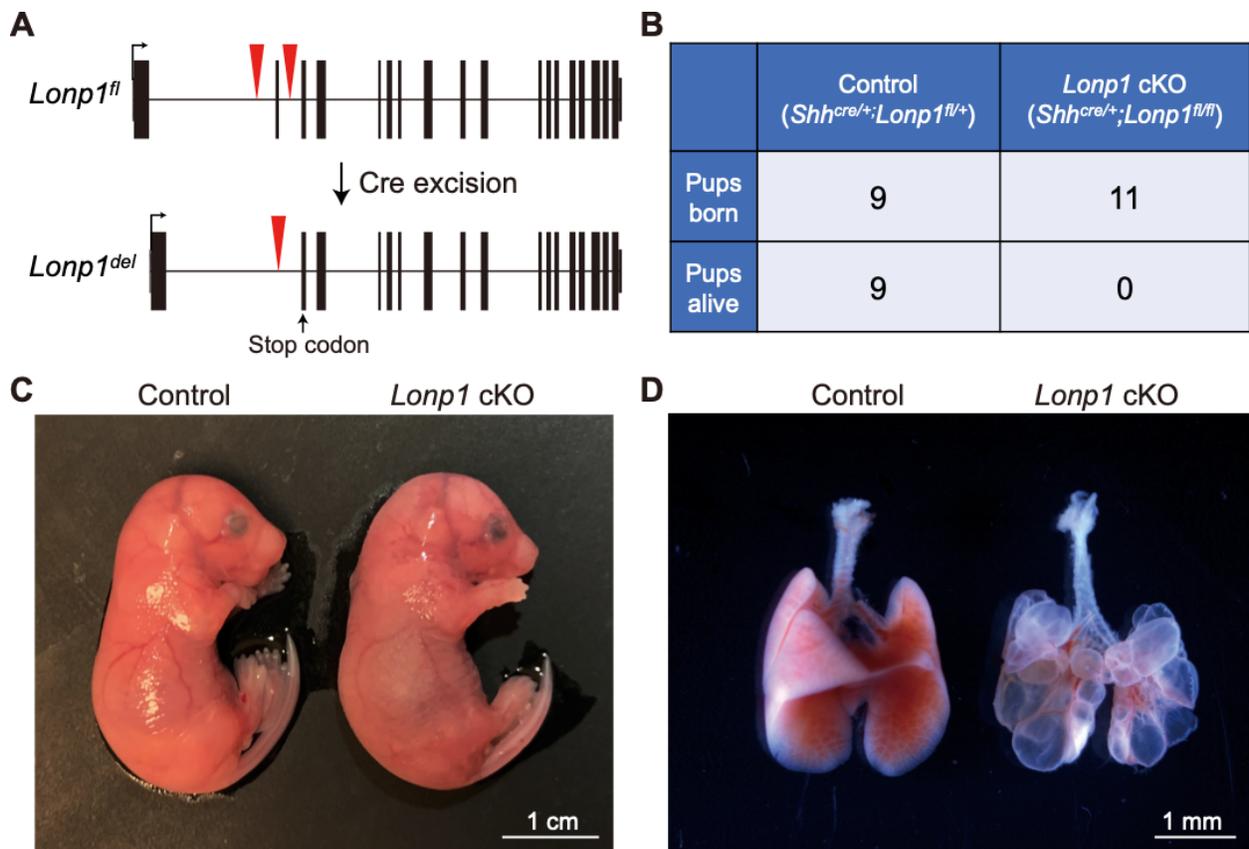


1009

1010 **Figure 5. Predicted 3D structure of *LONP1* protein using SWISS-Model.** (A) Variants in  
1011 ATPase domain (gray) of CDH (red) and CODAS (blue). CODAS variants (p.A670-pA724) are  
1012 clustered at alpha-helix in ATPase domain. (B) Variants in Protease domain (yellow) of CDH  
1013 (red) and CODAS (blue). CDH variants p.A821, S866 and CODAS variants p.A927 are located  
1014 at alpha-helix.



1016 **Figure 6. Inactivation of *Lonp1* in mice led to disrupted lung development and lethality at**  
 1017 **birth.** A. Gene structure of mouse *Lonp1<sup>fl</sup>* conditional allele before and after cre-mediated  
 1018 recombination of the *loxP* sites (red triangles). Recombination led to a premature stop codon  
 1019 (arrow) in the second exon. B. Number of embryos genotyped at perinatal stage, showing 100%  
 1020 lethality of the mutant embryos. C. Representative mutant and control embryos at embryonic day  
 1021 (E)18.5, the day of birth. D. Representative mutant and control lungs at E18.5. Scale bars as  
 1022 indicated.



1023

1024