Rare and de novo variants in 827 congenital diaphragmatic hernia probands implicate LONP1 and ALYREF as new candidate risk genes

Lu Qiao,1,2 Le Xu,3 Lan Yu,1 Julia Wynn,1 Rebecca Hernan,1 Xueya Zhou,1,2 Christiana Farkouh-Karoleski,1 Usha S. Krishnan,1 Julie Khlevner,1 Aliva De,1 Annette Zygmunt,1 Timothy Crombleholme,4 Foong-Yen Lim,5 Howard Needelman,6 Robert A. Cusick,6 George B. Mychaliska,7 Brad W. Warner,8 Amy J. Wagner,9 Melissa E. Danko,10 Dai Chung,10 Douglas Potoka,11 Przemyslaw Kosiński,12 David J. McCulley,13 Mahmoud Elfiky,14 Kenneth Azarow,15 Elizabeth Fialkowski,15 David Schindel,16 Samuel Z. Soffer,17 Jane B. Lyon,18 Jill M. Zalieckas,19 Badri N. Vardarajan,20 Gudrun Aspelund,1 Vincent P. Duron,1 Frances A. High,19,21,22 Xin Sun,3 Patricia K. Donahoe,21,23 Yufeng Shen,2,24,25,* and Wendy K. Chung1,26,*

1Department of Pediatrics, Columbia University Irving Medical Center, New York, NY
2Department of Systems Biology, Columbia University Irving Medical Center, New York, NY 10032, USA; 3Department of Pediatrics, University of California, San Diego Medical School, San Diego, CA 92092, USA; 4Medical City Children’s Hospital, Dallas, TX 75230, USA; 5Cincinnati Children's Hospital Medical Center, Cincinnati, OH 45229, USA;
6University of Nebraska Medical Center College of Medicine, Omaha, NE 68114, USA; 7University of Michigan Health System, Ann Arbor, MI 48109, USA; 8Washington University School of Medicine, St. Louis, MO 63110, USA; 9Children’s Hospital of Wisconsin, Medical College of Wisconsin, Milwaukee, WI 53226, USA; 10Monroe Carell Jr. Children’s Hospital at Vanderbilt, Nashville, TN 37232, USA; 11University of Pittsburgh, Pittsburgh, PA 15224, USA; 12Medical University of Warsaw, 02-091 Warsaw, Poland;
13Department of Pediatrics, University of Wisconsin-Madison, Madison, WI 52726, USA;
14Cairo University, Cairo 11432, Egypt; 15Oregon 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NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
Congenital diaphragmatic hernia (CDH) is a severe congenital anomaly that is often accompanied by other anomalies. Although the role of genetics in the pathogenesis of CDH has been established, only a small number of disease genes have been identified. To further investigate the genetics of CDH, we analyzed de novo coding variants in 827 proband-parent trios and confirmed an overall significant enrichment of damaging de novo variants, especially in constrained genes. We identified LONP1 (Lon Peptidase 1, Mitochondrial) and ALYREF (Aly/REF Export Factor) as novel candidate CDH genes based on de novo variants at a false discovery rate below 0.05. We also performed ultra-rare variant association...
analyses in 748 cases and 11,220 ancestry-matched population controls and identified

*LONP1* as a risk gene contributing to CDH through both *de novo* and ultra-rare inherited
largely heterozygous variants clustered in the core of the domains and segregating with CDH
in familial cases. Approximately 3% of our CDH cohort was heterozygous with ultra-rare predicted damaging variants in *LONP1* who have a range of clinical phenotypes including other anomalies in some individuals and higher mortality and requirement for extracorporeal membrane oxygenation. Mice with lung epithelium specific deletion of *Lonp1* die immediately after birth and have reduced lung growth and branching that may at least partially explain the high mortality in humans. Our findings of both *de novo* and inherited rare variants in the same gene may have implications in the design and analysis for other genetic studies of congenital anomalies.

**Introduction**

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

The complexity of the phenotypes associated with CDH is mirrored by the complexity of the genetics, which are heterogeneous with approximately 30% of CDH cases having an
identifiable major genetic contributor. Typically, each gene or copy number variant (CNV) associated with CDH accounts for at most 1-2% of cases. The full spectrum of genomic variants has been associated with CDH, including chromosome aneuploidies (10%), copy number variants (CNVs) (3-10%), monogenic conditions (10-22%), and emerging evidence for oligogenic causes (CNVs and individual genes).

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

A genetic diagnosis for probands with CDH can inform prognosis and guide medical management. Some genetic conditions associated with CDH are associated with an increased risk for additional anomalies, increased mortality, and increased morbidity including neurocognitive disabilities that may benefit from early intervention. Over the past decade, advances in genomic sequencing technology have helped to define the genes associated with CDH. We and others have shown that de novo variants with large effect size contribute to 10-22% of CDH cases with enrichment of de novo likely damaging variants in CDH cases with an additional anomaly (complex CDH). We also demonstrated a higher burden of de novo likely damaging (LD) variants in females compared to males supporting a “female protective model.” Most recently, in a cohort with long term developmental outcome data, we demonstrated that de novo likely damaging (LD) variants are associated with poorer...
neurodevelopmental outcomes as well as a higher prevalence of pulmonary hypertension (PH).

To expand upon our knowledge of the diverse genetic etiologies of CDH, we performed whole genome (WGS) or exome sequencing of 827 CDH proband-parent trios. We confirmed an overall enrichment of damaging de novo variants in constrained genes, and identified LONP1 (Lon Peptidase 1, Mitochondrial [MIM: 605490]) and ALYREF (Aly/REF Export Factor [MIM: 604171]) as new candidate CDH genes with recurrent ultra-rare and de novo variants.

Materials and methods

Participant recruitment and control datasets

Study participants were enrolled as fetuses, neonates, children and adults with a radiologically confirmed diaphragm defect by the DHREAMS study16 (Diaphragmatic Hernia Research & Exploration; Advancing Molecular Science) or Boston Children’s Hospital/Massachusetts General Hospital (BCH/MGH) as described previously14. Clinical data were prospectively collected from medical records and entered into a central Research Electronic Data Capture (REDCap) database17. Probands and both parents provided a blood, skin biopsy, or saliva specimen for trio genetic analysis. All studies were approved by the Columbia University institutional review board (IRB), serving as the central site. Each participating site also procured approval from their local IRB and signed informed consent was obtained. Ethical approval was obtained from the following participating institutions: Boston Children’s Hospital/Massachusetts General Hospital, Washington University, Cincinnati Children's Hospital Medical Center, Children's Hospital & Medical Center of Omaha, University of Michigan, Monroe Carell Jr. Children's Hospital, Northwell Health,
Oregon Health & Science University, Legacy Research Institute, University of Texas Southwestern, Children's Hospital of Wisconsin, and Children's Hospital of Pittsburgh.

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

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

The control group consisted of unaffected parents from the Simons Powering Autism Research for Knowledge (SPARK) study\(^18\) (exomes) and Latinx samples from Washington Heights-Hamilton Heights-Inwood Community Aging Project (WHICAP) study\(^19\) (exomes).

**WGS and exome data analysis**

There are 233 CDH trios processed using whole genome sequencing (WGS) that were not included in previous studies\(^3,9\) (Table S1). Of these 233 previously unpublished trios, 1 trio was processed at Baylor College of Medicine Human Genome Sequencing Center and 232 trios at Broad Institute Genomic Services. The genomic libraries of 219 cases were prepared by TruSeq DNA PCR-Free Library Prep Kit (Illumina), while 14 were TruSeq DNA PCR-
Plus Library Prep Kit (Illumina), with average fragment length about 350 bp, and sequenced as paired-end of 150-bp on Illumina HiSeq X platform. Exome sequencing was performed in 20 CDH trios that were not previously published\textsuperscript{3,9}. Among these, the coding exons of 9 trios were captured using Agilent Sure Select Human All Exon Kit v2 (Agilent Technologies), 10 trios using NimbleGen SeqCap EZ Human Exome V3 kit (Regeneron NimbleGen), 1 trio using NimbleGen SeqCap EZ Human Exome V2 kit (Roche NimbleGen). Exomes of SPARK cohort were captured using a slightly modified version of the IDT xGen Exome Research Panel v.1.0 identical to the previous study\textsuperscript{20}. Whole-exome sequencing of the WHICAP cohort was performed at Columbia University using the Roche SeqCap EZ Exome Probes v3.0 Target Enrichment Probes\textsuperscript{21}.

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

De novo variants were defined as a variant present in the offspring with homozygous reference genotypes in both parents. Here, we limited WGS to coding regions based on coding sequences and canonical splice sites of all GENCODE v27 coding genes. We took a
series of stringent filters to identify de novo variants as described previously: VQSR tranche ≤99.8 for SNVs and ≤99.0 for indels; GATK’s Fisher Strand ≤25, quality by depth ≥2. We required the candidate de novo variants in probands to have ≥5 reads supporting the alternative allele, ≥20% alternative allele fraction, Phred-scaled genotype likelihood ≥60 (GQ), and population allele frequency ≤0.01% in gnomAD v2.1.1; both parents to have ≥10 reference reads, <5% alternative allele fraction, and GQ ≥30. We applied DeepVariant to all candidate de novo variants for in silico confirmation and only included the ones with PASS from DeepVariant for downstream analysis.

To reduce batch effects in combined datasets from different sources in analysis of rare variants, for non-Latinx population we targeted ultra-rare variants located in xGen-captured protein coding regions and for Latinx population in regions targeted by xGen and SeqCap EZ v3.0. We used the following criteria to minimize technical artifacts and select ultra-rare variants: cohort AF <0.5% and population cohort <1×10⁻⁵ across all genomes in gnomAD v3.0; mappability=1; >90% target region with depth ≥10; overlapped with segmental duplication regions <95%; genotype quality >30, allele balance >20% and depth >10 in cases.

We used Ensembl Variant Effect Predictor (VEP, Ensemble 102) and ANNOVAR to annotate variant function, variant population frequencies and in silico predictions of deleteriousness. All coding SNVs and indels were classified as synonymous, missense, inframe, or likely-gene-disrupting (LGD, which includes frameshift indels, canonical splice site, or nonsense variants). We defined predicted damaging missense (D-mis) based on CADD score v1.3. All de novo variants and inherited variants in candidate risk genes were manually inspected in the Integrative Genome Viewer (IGV). A total of 179 variants were selected for validation using Sanger sequencing; all of them were confirmed (Table S2).
compare the clinical outcomes between cases with deleterious variants in candidate genes and cases with likely damaging (LD) variants, we defined likely damaging variants as in our previous study\(^3\): (a) de novo LGD or deleterious missense variants in genes that are constrained (ExAC pLI \(\geq 0.9\)) and highly expressed in developing diaphragm\(^34\), or (b) de novo LGD or deleterious missense variants in known risk genes for CDH or commonly comorbid disorders (congenital heart disease [CHD] and neurodevelopmental delay [NDD]), or (c) plausible deleterious missense variants in known risk genes for CDH or commonly comorbid disorders (CHD and NDD), or (d) deletions in constrained (ExAC pLI \(\geq 0.9\)) or haploinsufficient genes from ClinGen genome dosage map\(^35\), or (e) CNVs implicated in known syndromes. We classified CDH cases into two genetic groups: (1) LD, if the case carried at least one de novo LD variant; (2) non-LD, if the case carried no such variants.

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

**Quantitative PCR**

We performed experimental validation of putative *de novo* genic CNVs using quantitative PCR (qPCR). All PCR primers were designed for the selected genes located within the *de novo* CNVs and synthesized by IdtDNA. All qPCR reactions were performed in a total of 10 \(\mu\)l
volume, comprising 5 μl 2x SYBR Green I Master Mix (Promega), 1 μl 10nM of each primer and 2 μl of 1:20 diluted cDNA in 96-well plates using CFX Connect Real-Time PCR Detection System (Bio-Rad). All reactions were performed in triplicate, and the conditions were 5 minutes at 95°C, then 40 cycles of 95°C at 15 seconds and 60°C at 30 seconds. The relative copy numbers were calculated using the standard curve method relative to the β-actin housekeeping gene. Five-serial 4-fold dilutions of DNA samples were used to construct the standard curves for each primer.

Statistical analysis

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

extTADA analysis. To identify risk genes based on de novo variants, we used an empirical Bayesian method, extTADA (Extended Transmission and de novo Association). The extTADA model was developed based on a previous integrated empirical Bayesian model TADA and estimates mean effect sizes and risk-gene proportions from the genetic data using MCMC (Markov Chain Monte Carlo) process (details see supplemental note).
inform the parameter estimation with prior knowledge of developmental disorders, we
stratify the genes into constrained genes (ExAC pLI score >0.5) and non-constrained genes
(other genes), followed by estimating the parameters using the extTADA model to each
group of genes. After estimating posterior probability of association (PPA) of individual
genes in each group, we combined both groups to calculate a final false discovery rate (FDR)
for each gene using extTADA’s procedure.

Gene-based case-control association analysis of ultra-rare variants. To identify novel risk
genes based both on de novo and rare inherited variants, we performed a gene-based
association test comparing the frequency of ultra-rare deleterious variants in CDH cases with
controls, without considering de novo status. Samples with read depth coverage ≥10x for
80% in exome cases and 90% in genome cases of the targeted regions were included in the
analysis (Figure S1). Relatedness was checked using KING\textsuperscript{27}, and only unrelated cases were
included in the association tests (Figure S2). To control for confounding from genetic
ancestry, we selected ancestry-matched controls using SPARK exomes and Latinx WHICAP
exomes to reach a fixed case/control ratio in each population ancestry inferred by peddy\textsuperscript{28}
(Figure S3). Specifically, for a specific ancestry ($i$), consider $x_i$ number of cases, $y_i$ number
of controls, $n_i$ the fold controls to cases ($y_i/x_i$). We chose the minimized $n_{\text{min}}$ among all
ancestries. In each genetic-ancestry group controls ($y_i$), we ranked the Euclidean distance
between each case and controls which were calculated from top 3 PCA eigenvectors and
selected $n_{\text{min}}x_i$ controls from $y_i$ controls to ensure the same proportions in cases and
controls. After filtering to reduce the impact of false positive variants, we tested for similarity
of the ultra-rare synonymous variant rate among cases and controls in specific genetic-ancestry groups, assuming that ultra-rare synonymous variants are mostly neutral with
respect to disease status.
To identify CDH risk genes, we tested the burden of ultra-rare deleterious variants (AF $< 1 \times 10^{-5}$ across all gnomAD v3.0 genomes, LGD or D-Mis) in each protein-coding gene in cases compared to controls. To improve statistical power, we searched for a gene-specific CADD score threshold for defining D-Mis that maximized the burden of ultra-rare deleterious variants in cases compared to controls and used permutations to calculate statistical significance with the variable threshold test\textsuperscript{22,45}. For the binomial tests in each permutation, we used binom.test function in R to calculate p values. We performed two association tests, one with LGD and D-Mis variants combined and the other with D-Mis variants alone, to account for different modes of action. We defined the threshold for genome-wide significance by Bonferroni correction for multiple testing (as two tests for each gene with 20,000 protein-coding genes, threshold p-value=$1.25 \times 10^{-6}$). We checked for inflation using a quantile-quantile (Q-Q) plot and calculated the genomic control factor (lambda [$\lambda$]) using QQperm in R. Lambda equal to 1 indicates no deviation from the expected distribution.

**Protein modeling**

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

**Mice**

All mice were housed in American Association for Accreditation of Laboratory Animal Care accredited facilities and laboratories at University of California, San Diego. All animal
Experiments were conducted under approved guidelines for the Care and Use of Laboratory Animals. Lmp1Lβ and Shhcre mice have all been described previously47 (International Mouse Strain Resource J:204812). All mice were bred on a C57BL/6J background, and littermates were used as controls to minimize potential genetic background effects.

Results

Cohort characteristics

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

In the cohort, there were 486 (59%) male probands (Table 1), consistent with a higher prevalence of CDH in males9,48,49. The genetically determined ancestries (Figure S3A) were European (73.4%), admixed American (hereafter referred to as Latinx; 18.5%), African (3.7%), East Asian (1.8%), and South Asian (2.5%). Among the 277 (33.5%) complex cases, the most frequent additional anomalies were congenital heart disease (n=144), NDD (n=54), skeletal anomalies (n=46), genitourinary anomalies (n=46) and gastrointestinal anomalies (n=42). A total of 533 (64.4%) probands had isolated CDH without additional anomalies at the time of last follow up. The most common type of CDH was left-sided Bochdalek (Table 1).

Burden of de novo coding variants

We identified 1153 de novo protein-coding variants in 619 (74.8%) cases including 1058 single nucleotide variants (SNVs) and 95 indels (Table S2). The average number of de novo
coding variants per proband is 1.39. The number of *de novo* coding variants across probands closely follows a Poisson distribution (Figure S4). Transition-to-transversion ratio of *de novo* SNVs was 2.75. We classified variants that were likely gene disruptive (LGD) or predicted damaging missense (“D-mis” with CADD score ≥ 25) as damaging variants. A total of 418 damaging variants (126 LGD and 292 D-mis) were identified in 318 (38.4%) cases, including 83 (10%) cases harboring two or more such variants.

We analyzed the burden of *de novo* variants in CDH cases by comparing the observed number of variants to the expected number based on the background mutation rate. Consistent with previous studies on CDH⁹ and other developmental disorders⁵⁰-⁵², both *de novo* LGD (0.15 per case) and D-mis variants (0.35 per case) were significantly enriched in cases (relative risk [RR] = 1.5, P = 3.6 × 10⁻⁵ for LGD; RR = 1.3, P = 3.1 × 10⁻⁶; Figures 1A and B; Table S3) while the frequency of synonymous variants (0.30 per case) closely matches the expectation (RR = 0.9, P = 0.12; Table S3). The burden of LGD variants is mostly located in constrained (ExAC³⁹ pLI > 0.5) genes (RR = 2.2, P = 1.8 × 10⁻⁸). It is marginally higher in female cases than male cases (RR = 3.0 vs 1.36, P = 0.012) and marginally higher in complex cases than isolated cases (RR = 3.1 vs 1.75, P = 0.024; Figure 1C; Table S3).

To identify new CDH risk genes by *de novo* variants, we applied extTADA⁴³ to the data of 827 CDH trios. ExtTADA assumes a model of genetic architecture compatible with the observed burden and recurrence of *de novo* damaging variants and estimates a false discovery rate (FDR) for each gene using MCMC. From the burden analysis of *de novo* variants in CDH and previous studies⁵², we reasoned that the constrained genes (ExAC pLI > 0.5) drive the higher burden of *de novo* damaging variants and are more likely to be plausible risk genes. We stratified the data into the constrained gene set and the non-constrained gene set (Table S4) and estimated extTADA priors (mean relative risk and prior probability of being a
risk gene) in these two gene sets separately. Constrained genes had a higher prior of risk
genes than non-constrained genes (0.037 vs 0.006). Meanwhile, both LGD and D-mis had
higher relative risks in constrained genes than non-constrained gene (18.30 vs 5.24 for LGD;
10.01 vs 3.81 for D-mis). We estimated Bayes Factor of individual genes within each gene
group and then combined the genes from two groups together to calculate FDR. We
identified 3 genes with FDR <0.05: MYRF (Myelin Regulatory Factor [MIM: 608329]),
LONP1, and ALYREF. Five of 6 MYRF de novo variants were described in our previous
study. We identified 3 participants harboring de novo D-mis variants in LONP1 and 2
participants for de novo LGD variants in ALYREF. Of two participants with an ALYREF LGD
variant, one had an isolated left-side CDH and the other had right-side CDH and ventricular
septal defect. There were nine additional genes with ≥2 de novo predicted deleterious
variants (HSD17B10 [MIM: 300256], GATA4 [MIM: 600576], SYMPK [MIM: 602388],
PTPN11 [MIM: 176876], WT1 [MIM: 607102], FAM83H [MIM: 611927], CACNA1H
[MIM: 607904], SEPSECS [MIM: 613009], and ZFYVE26 [MIM: 612012]) (Table 2). Of
these, three are known CDH genes (MYRF, GATA4, WT1). All de novo variants in these
genes are heterozygous.

**Recurrent genes in de novo CNVs**

We applied CNVnator to call CNVs from WGS data and used customized filters to identify
de novo CNVs. We performed experimental validation of 25 putative de novo genic CNVs
including all 9 small CNVs (<5kb) using quantitative PCR (qPCR). 22 of 25 (88%) reported
de novo CNV in cases were confirmed by qPCR. Removing the 3 false positive CNVs, there
were 87 de novo CNVs identified in 734 CDH cases with WGS with an average of 0.12 per
case (Table S5). Among them, there were 54 (62%) deletions ranging from 2,096 bp to 33.7
Mb and 33 (38%) duplications ranging from 1,165 bp to 24.9 Mb. Seven samples carried
known syndromic CNVs in DECIPHER\textsuperscript{41} dataset, one of which was heterozygous for a 16p13.11 microduplication, two heterozygous for a 17q12 deletion associated with renal cysts and diabetes (RCAD), three heterozygous for 21q22 duplication in the critical region for Down syndrome, and one heterozygous for 22q11 deletion associated with DiGeorge syndrome. No recurrent genes were identified between \textit{de novo} SNVs and CNVs. Four CNVs were recurrent (Table 3), two of which encompass single genes \textit{CSMD1} (CUB And Sushi Multiple Domains 1 [MIM: 608397]) and \textit{GPHN} (Gephyrin [MIM: 603930]).

Candidate gene \textit{LONP1} contributes to CDH risk through both \textit{de novo} and rare inherited variants

To identify additional risk genes that may contribute through rare inherited variants, we performed a gene-based, case-control association analysis of ultra-rare variants. Specifically, we used exome data from the SPARK (unaffected parents) and Latinx WHICAP samples as controls. Quality control procedures included at least 10x depth of sequence coverage across the target regions (Figure S1) and detection of cryptic relatedness amongst all CDH participants and controls (Figure S2). To prevent confounding by genetic ancestry, we performed principal component analysis (PCA) by peddy to infer genetic ancestry of all cases and controls and selected matching controls (15-fold of cases numbers in each specific genetic-ancestry group) to reach a fixed case/control ratio. With the same genetic-ancestry proportion in cases and controls (77% Europeans, 14.8% Latinx, 4.1% Africans, 2% East Asians, 2.1% South Asians; Figure S3; Table S6), we selected 748 cases and 11,220 controls for downstream analysis. We filtered the ultra-rare variant call sets of cases and controls in each genetic-ancestry group by empirical filters to reduce false positive calls and minimize technical batch effects across data sets. After filtering, the average numbers of ultra-rare (AF<1×10\textsuperscript{-5} across all gnomAD v3.0 genomes) synonymous variants per subject in cases and controls are presented in Table 4.
controls are nearly identical in everyone (enrichment rate=1, P=1) and specific ancestral groups (Table S7). Furthermore, a gene-level burden test confined to ultra-rare synonymous variants was consistent with a global null model in Q-Q plot (Figure S5), indicating that technical batch effects would likely have minimal impact on genetic analyses. We then performed a variable threshold association test\textsuperscript{22,45} to identify new risk genes based on enrichment of ultra-rare damaging variants in individual genes. For each gene, we tested enrichment of LGD and D-mis variants together or just D-mis variants, in order to account for potential different biological modes of action. In the variable threshold test, we determined a gene-specific optimal CADD score threshold to define D-mis in order to maximize the power of the association test and then estimated type I error rate by permutations. The overall result from the case-control association did not show inflation from the null model ($\lambda=1.09$; Figure 2A). The association of $LONP1$ (P=$1\times10^{-7}$; Figure 2) exceeded the Bonferroni-corrected significance threshold ($1.25\times10^{-6}$, account for two tests in each gene). Three of the 24 ultra-rare deleterious variants in $LONP1$ were known \textit{de novo} variants. Two known CDH risk genes, $ZFPM2$ (Zinc Finger Protein, FOG Family Member 2 [MIM: 603693]) and $MYRF$, fell just below the cutoff for genome wide significance.

The association of $LONP1$ is due to both LGD and D-mis variants. We screened the whole cohort (Figure 3 and Table 4), including CDH relatives (n=1) and exome sequencing singletons (n=2), for ultra-rare damaging missense (CADD $\geq 25$) and LGD in $LONP1$ (NM_004793.3). A total of 23 CDH cases in 829 cases (2.8%) carry 24 $LONP1$ variants, including 10 LGD and 14 D-mis variants. Among 22 $LONP1$ variants excluding 2 of unknown inheritance variants in singletons, there are 3 (13.6\%) \textit{de novo} variants (all D-mis) and 19 (86.4\%) inherited variants, 36.8\% of which are from mothers (n=7). Of 19 inherited variants, 8 parents carrying $LONP1$ variants have a family history of CDH or diaphragm eventration (n=4) or other congenital anomaly (n=4; brain abnormality, cerebral palsy, cleft
palate, skeletal abnormality) segregating with the LONP1 variant. Three inherited variants
c.1913C>T [p.638M], c.2122G>A [p.G708S] and c.2263C>G [p.R755G]) are each observed
twice in the cohort on different probands. Familial segregation was established in six familial
CDH cases for c.398C>G (p.P133R), c.6391G>T (p.X213_splice), c.1262delG
(p.F421Lfs*87), c.1574C>T (p.T638M) and c.2719dupG
(p.V907Gfs*73). One case (01-1279) carries biallelic heterozygous variants with c.1574C>T
(p.P525L) inherited from one parent and c.2263C>G (p.R755G) inherited from another
parent. The participant with biallelic heterozygous variants required ECMO and died at 8-9
hours after birth with severe bilateral CDH with near complete diaphragm agenesis, bilateral
lung hypoplasia, and no additional anomalies (Figure 4). All other cases are heterozygous
variants.

Previous studies reported biallelic variants in LONP1 in cerebral, ocular, dental, auricular,
and skeletal (CODAS) syndrome\textsuperscript{53,54} (MIM: 600373). We compared the locations of the
predicted-damaging missense positions in CDH cases and CODAS syndrome cases (Figures
3 and 5). No variants overlap between CDH cases and CODAS syndrome. LONP1 contains
three functional domains. CDH damaging variants are concentrated at the core of the
domains. Biallelic variants in CODAS syndrome are located on the junction of ATP-binding
and proteolytic domains (Figures 3 and 5). The 23 CDH cases with LONP1 variants didn’t
have features of CODAS syndrome.

\textbf{Phenotype of CDH probands with LONP1 variants}

We identified 24 ultra-rare heterozygous variants in 23 sporadic or familial CDH participants
(Table 4). The majority (n=17; 73.9\%) are of European ancestry and 13 (56.5\%) are female
(Table 4). Sixteen (70\%) were enrolled as neonates. Fourteen of the 23 have a family history
of congenital anomalies (Table 4), 6 of whom had a family history of CDH. Nine (39.1\%) are
complex cases. Six of 9 complex cases have CHD in addition to CDH. We compared the
clinical outcomes or phenotypes in CDH cases with LONPI damaging variants and other
CDH cases (Table 5). Compared to CDH cases without LONPI ultra-rare damaging variants,
LONPI damaging variant carriers are associated with higher neonatal mortality rate prior to
initial hospital discharge (69% vs 16%, $P=6.4\times10^{-6}$) and greater need for ECMO (56% vs
28%, $P=2.3\times10^{-2}$). Compared to CDH cases with other likely damaging variants defined in
our previous study$^3$, LONPI damaging variant carriers had higher neonatal mortality rate
prior to discharge (69% vs 24%, $P=1.8\times10^{-3}$) and trended towards greater need for ECMO
(56% vs 30%, $P=0.077$).

**Inactivation of Lonp1 in mouse embryonic lung epithelium leads to disrupted lung
development and full lethality at birth**

The high rate of mortality and need for ECMO in cases with CDH is predominantly due to
abnormal lung and pulmonary vascular development causing lung hypoplasia and pulmonary
hypertension. Our hypothesis was that impaired or partial loss of LONPI function in cases
with CDH might contribute directly to abnormal lung development, independent of its role in
diaphragm formation. To test this hypothesis, we inactivated Lonp1 in the embryonic lung
epithelium in mice. This was achieved by generating $Shh^{cre/+};Lonp1^{fl/fl}$ (hereafter Lonp1 cKO
for conditional knockout) embryos using existing alleles$^{47}$ (International Mouse Strain
Resource J:204812). In the mutant the cre recombinase expressed specifically in the
epithelium drove Lonp1 inactivation at the onset of lung initiation (Figure 6A). This led to
100% lethality of the mutants at birth with normal body size (Figures 6B and C). Upon
dissection, the mutant lung was composed of large fluid-filled sacs, unlike the controls with
normal airways and alveoli (Figure 6D). The lung defect likely contributed to embryonic
lethality at birth in these mutant mice.
Discussion

In the current study of 827 CDH trios, we confirmed there is an overall enrichment of damaging \textit{de novo} variants, particularly in constrained genes. We identified \textit{LONP1} and \textit{ALYREF} as novel candidate genes based on enrichment of \textit{de novo} variants. By case-control association, we also confirmed \textit{LONP1} as a genome-wide significant candidate gene contributing to CDH risk through both \textit{de novo} and inherited damaging variants. We demonstrated segregation of the \textit{LONP1} variant with diaphragm defect in five families. We found that CDH individuals with heterozygous ultra-rare damaging variants in \textit{LONP1} have clinical phenotypes frequently including CHD or skeletal anomalies, frequently requiring ECMO, and having a higher mortality than the rest of our CDH cohort. In addition, we confirmed \textit{MYRF} and \textit{ZFPM2} as genes previously associated with CDH\textsuperscript{9,14,55,56}. In a mouse model with knock out of \textit{Lonp1} only in the embryonic lung epithelium with an intact diaphragm, we demonstrated reduced pulmonary growth and branching, resulting in perinatal lethality that suggests that the higher mortality rate and need for ECMO in human is due to a primary effect of \textit{LONP1} on pulmonary development in addition to diaphragm development.

The burden of damaging \textit{de novo} variants in CDH is consistent with previous studies\textsuperscript{9,14,15}, and damaging \textit{de novo} variants are more frequent in complex CDH compared to isolated CDH cases. Similar patterns have been observed in complex congenital heart disease with other congenital anomalies or neurodevelopmental disorders compared with isolated congenital heart disease\textsuperscript{50} and autism with/without intellectual disability\textsuperscript{57}. Deleterious \textit{de novo} variants are more frequent in many severe early-onset diseases with reduced reproductive fitness compared to the general population\textsuperscript{58}. The higher frequency of \textit{de novo} LGD variants in female relative to male CDH cases supports the “female protective model” similar to autism\textsuperscript{52,59,60}, which means that risk variants have larger effects in males than in
females so that females require a higher burden to reach the same diagnostic threshold as males.

Both de novo and rare inherited variant analyses highlight LONP1 as a novel CDH candidate gene. Approximately 3% of individuals in our CDH cohort are heterozygous for LONP1 rare variants. Three variants (p. T638M, p.G708S and p.R755G) are recurrently and independently found in unrelated families. CDH cases with LONP1 variants had higher mortality in the neonatal period compared with other CDH cases. Biallelic variants in LONP1 have been reported in CODAS, a multi-system developmental disorder characterized by cerebral, ocular, dental, auricular, and skeletal anomalies. The Lonp1 holoenzyme is a homohexamer with six identical subunits. Each subunit consists of a mitochondrial-targeting sequence (MTS), a substrate recognition and binding (N) domain, an ATPase (AAA+) domain, and a proteolytic (P) domain. Biallelic missense variants reported in CODAS individuals are mostly located in the junction of ATP-binding and proteolytic domains of LONP1 while the heterozygous variants identified in CDH individuals are located in the main domains of LONP1. Notably, there are no overlapping variants between CDH and CODAS individuals. Most of the variants in CODAS are located in the alpha-helix and may affect the interactions of subunits. Variants in CDH may interrupt the proteolytic and ATP binding domains, resulting in the dysfunction of LONP1. Homozygous deletion of LONP1 in mice is embryonic lethal, due to progressive loss of mtDNA with subsequent failure to meet energy requirements for embryonic development. Heterozygous Lonp1+/- mice develop normally without obvious abnormalities, but lonp1 expression decreased in both RNA and protein levels. Analysis of Lonp1 expression in heterozygous mice indicated a 50% reduction at both RNA and protein levels in these animals. These data suggest different mechanisms of LONP1 in diseases with biallelic and monoallelic variants. Of note, one CDH individual
carried biallelic variants (p.P525L and p.R755G). No additional phenotypes were noted, perhaps because the baby died at 8-9 hours after birth with severe bilateral CDH (Figure 4).

Lonp1 is a nuclear-encoded mitochondrial protease. Besides binding of mtDNA, Lonp1 was discovered as an ATP-dependent protease involved in the degradation of misfolded or damaged proteins. Accumulation of misfolded proteins has been observed in the impaired lungs of developing mice with deletion of other ATP-dependent proteins. The immature lung development and neonatal respiratory failure of our Lonp1 cKO mice could be due to the inactivation of Lon protease, which results in the accumulation of misfolded proteins and activation of the unfolded protein response (UPR) pathway. UPR activation during development could lead to reduced cell proliferation and cause other congenital anomalies including congenital heart disease.

Lonp1 also acts as a chaperone that interacts with other mitochondrial proteins to regulate several cellular processes. Lon expression may stimulate cell proliferation and Lon downregulation may impair mitochondrial structure and function and cause apoptosis. Alterations in cell proliferation, differentiation and migration can all lead to CDH. Myogenic cell differentiation and migration are essential during formation of the diaphragm. Myogenic differentiation requires increased expression of mitochondrial biogenesis-related genes including Lon. The variants could cause an increased probability of failure of myogenesis during embryonic development, consequently resulting in the hernia.

The neonatal mortality of probands with LONP1 deleterious variants is much higher than CDH neonates without LONP1 deleterious variants or CDH neonates with likely damaging variants in genes other than LONP1. CDH neonates with LONP1 deleterious variants frequently required ECMO. In mice with Lonp1 knock out at the onset of lung development, 100% newborn pups died shortly after birth, with severe pulmonary defects. Thus, LONP1...
could represent a class of CDH genes with high mortality due to primary developmental
effects on the lung, resulting in more severe pulmonary defects than would occur secondary
to lung compression by herniated abdominal viscera alone. This suggests that we should try
to differentiate primary from secondary developmental effects on the lung as we phenotype
newborns with CDH and as we investigate the mechanisms action of CDH candidate genes.

The RNA-binding protein ALYREF plays a key role in nuclear export through binding to the
5’ and the 3’ regions of mRNA. It acts as an RNA 5-methylcytosine (m^5C) adaptor to
regulate the m^5C modification. Disruption of ALYREF could affect the m^5C
modification, resulting in abnormal cell proliferation and migration. Previous studies
identified several RNA binding proteins (RBPs) playing essential roles in autism and
congenital birth defects including CHD. RBFOX2, an RBP that regulates alternative splicing,
is critical for zebrafish heart development and de novo variants in RBFOX2 are associated
with congenital heart defects. Dozens of RBPs have established roles in autism spectrum
disorder. RBFOX1, an RNA splicing factor, regulates expression of large genetic
networks during early neuronal development including autism. The other RBPs such as
FMRP, CELF4, CELF6, have also been implicated in autism. As an RBP, ALYREF may
play a similar role in congenital anomalies and neurodevelopmental disorders. Two de novo
LGDs in ALYREF were identified in our CDH cohort. One had an isolated CDH and the other
had CDH and a ventricular septal defect. Similarly, two CDH cases carried de novo variants
in SYMPK, another RBP identified with FDR<0.1 in extTADA. One had a de novo predicted
deleterious missense variant and isolated CDH and the other had a de novo LGD with
complex CDH with congenital heart disease, central nervous system anomaly, and
genitourinary anomaly.
We found further support for the previously reported CDH genes ZFPM2 and MYRF. We have identified six ultra-rare LGD variants in ZFPM2 in our CDH cohort, accounting for 0.7% of our cases (Figure S6). Three were complex cases, all with minor cardiac malformations. Specifically, two females had atrial septal defects and 1 male had an enlarged aortic root. The other three heterozygotes had isolated CDH. ZFPM2 is expressed in the septum transversum of the diaphragm during early development, and Fog2−/− mice generated through chemical mutagenesis have been shown to have diaphragmatic eventration and pulmonary hypoplasia. ZFPM2 physically interacts with NR2F2 and GATA4, two other components of the retinoid signaling pathway implicated in diaphragm and lung development. Our results further support the pleiotropic role of ZFPM2 in the development of CDH.

MYRF was implicated in our previous de novo variant report as a gene for cardiac-urogenital syndrome (MIM: 618280), and we identified one more additional de novo variant in this cohort (Figure S7). There are now more than 10 variants implicated in CDH with additional anomalies (HGMD® professional 2021.1). MYRF is highly expressed in epithelial cells. Diaphragm is composed of epithelial-like mesothelial cells derived from the mesoderm of the pleuroperitoneal folds (PPFs) through cell proliferation, migration, and epithelial-to-mesenchymal transition. Single cell analysis in fetal gonads suggests the cells that highly express MYRF also express WT1 and NR2F2, two genes associated with diaphragmatic hernia. Previously, we also demonstrated that individuals with pathogenic variants in MYRF have decreased expression of GATA4. WT1, NR2F2 and GATA4 are all important in RA signaling in the developing diaphragm. Therefore, the damaging variants in MYRF may affect the RA signaling pathway, leading to diaphragmatic hernia and other anomalies.
Among the 734 CDH trios with WGS data, we identified a total of 87 de novo CNVs and 4 of them are recurrent genes or CNVs. Given the rarity of de novo CNVs and small sample size, there were limited data to analyze the differential burden between cases and controls in this study. Future studies with larger sample sizes will improve the power to analyze CNVs and structural variants in CDH.

In summary, our analysis of de novo and ultra-rare inherited variants identified two new CDH candidate genes LONP1 and ALYREF and confirmed previous associations of MYRF and ZFPM2 with CDH. The identification of specific highly risk genes would enhance prenatal or early postnatal counseling and decision making, especially with rapid turnaround of WGS or exome sequencing results. It is likely that transmitted rare variants also contribute to other cases in our cohort, but we require a larger sample size to identify these genes confidently.

Future studies will also leverage data from other developmental disorders and integrating genomic data during development.

Supplemental Data

Supplemental Data include notes, 7 figures and 7 tables.

Acknowledgements

We would like to thank the patients and their families for their generous contribution. We are grateful for the technical assistance provided by Na Zhu, Patricia Lanzano, Jiangyuan Hu, Jiancheng Guo, Suying Bao, Charles LeDuc, Liyong Deng, Donna Garey, and Anketil Abreu from Columbia University, Jennifer Lyu at Boston Children’s Hospital, and Caroline Coletti at Massachusetts General Hospital. We thank our clinical coordinators across the DHREAMS centers: Jessica Conway at Washington University School of Medicine, Melissa Reed, Elizabeth Erickson, and Madeline Peters at Cincinnati Children's Hospital, Sheila Horak and
Evan Roberts at Children's Hospital & Medical Center of Omaha, Jeannie Kreutzman and Irene St. Charles at CS Mott Children's Hospital, Tracy Perry at Monroe Carell Jr. Children's Hospital, Dr. Michelle Kallis at Northwell Health, Andrew Mason and Alicia McIntire at Oregon Health and Science University, Gentry Wools and Lorrie Burkhalter at Children's Medical Center Dallas, Elizabeth Jehle at Hassenfeld Children’s Hospital, Michelle Knezevich and Cheryl Kornberg at Medical College of Wisconsin, Min Shi at Children's Hospital of Pittsburgh. We would also like to acknowledge Terry Buchmiller at Boston Children's Hospital, and the other pediatric surgeons and clinicians who referred patients to our studies.

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

Declaration of Interests

The authors declare no competing interests.
Web Resources

DHREAMS study, http://www.cdhgenetics.com/

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

ClinGen genome dosage map, https://dosage.clinicalgenome.org

DECIPHER, https://www.deciphergenomics.org

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


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

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

PyMOL molecular viewer, https://pymol.org/2/

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

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

Reference


variants are associated with congenital diaphragmatic hernia. J Med Genet 49, 650-659. 10.1136/jmedgenet-2012-101135.


Bayesian analysis of rare exonic variants to identify risk genes for schizophrenia and
44. He, X., Sanders, S.J., Liu, L., De Rubeis, S., Lim, E.T., Sutcliffe, J.S., Schellenberg,
novo and inherited genetic variants yields greater power to identify risk genes. PLoS
Genet 9, e1003671. 10.1371/journal.pgen.1003671.
45. Price, A.L., Kryukov, G.V., de Bakker, P.I., Purcell, S.M., Staples, J., Wei, L.J., and
46. Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R.,
homology modelling of protein structures and complexes. Nucleic Acids Res 46,
W296-W303. 10.1093/nar/gky427.
function is essential for lung epithelium morphogenesis. Proc Natl Acad Sci U S A
103, 2208-2213. 10.1073/pnas.0510839103.
10.1002/bdr2.1015.
49. Leeuwen, L., Mous, D.S., van Rosmalen, J., Olieman, J.F., Andriessen, L., Gischler,
Congenital Diaphragmatic Hernia and Growth to 12 Years. Pediatrics 140.
10.1542/peds.2016-3659.
50. Homsy, J., Zaidi, S., Shen, Y., Ware, J.S., Samocha, K.E., Karczewski, K.J.,
mutations in congenital heart disease with neurodevelopmental and other congenital
10.1038/ng.3970.
52. Satterstrom, F.K., Kosmicki, J.A., Wang, J., Breen, M.S., De Rubeis, S., An, J.Y.,
Sequencing Study Implicates Both Developmental and Functional Changes in the
associated with mutations of LONP1, encoding mitochondrial AAA+ Lon protease.
10.1002/ajmg.1320400118.
55. Ackerman, K.G., Herron, B.J., Vargas, S.O., Huang, H., Tevosian, S.G., Kochilas, L.,
10.1371/journal.pgen.0010010.
56. Bleyl, S.B., Moshrefi, A., Shaw, G.M., Saijoh, Y., Schoenwolf, G.C., Pennachio,
hernia from animal models: sequencing of FOG2 and PDGFRalpha reveals rare
10.1038/sj.ejhg.5201872.

57. Iossifov, I., O’Roak, B.J., Sanders, S.J., Ronemus, M., Krumm, N., Levy, D.,
contribution of de novo coding mutations to autism spectrum disorder. Nature 515,
216-221. 10.1038/nature13908.

58. Kosmicki, J.A., Samocha, K.E., Howrigan, D.P., Sanders, S.J., Slowikowski, K., Lek,
M., Karczewski, K.J., Cutler, D.J., Devlin, B., Roeder, K., et al. (2017). Refining the
role of de novo protein-truncating variants in neurodevelopmental disorders by using
population reference samples. Nat Genet 49, 504-510. 10.1038/ng.3789.

59. Jacquemont, S., Coe, B.P., Hersch, M., Duyzend, M.H., Krumm, N., Bergmann, S.,
burden in females supports a "female protective model" in neurodevelopmental

highlights a sex differentiation in parental contribution to de novo and inherited
mutational burdens. Sci Rep 6, 25954. 10.1038/srep25954.

61. Gibellini, L., De Gaetano, A., Mandrioli, M., Van Tongeren, E., Bortolotti, C.A.,
10.1016/bs ircmb.2020.02.005.

62. Quiros, P.M., Espanol, Y., Acin-Perez, R., Rodriguez, F., Barcena, C., Watanabe, K.,
dependent Lon protease controls tumor bioenergetics by reprogramming
mitochondrial activity. Cell Rep 8, 542-556. 10.1016/j.celrep.2014.06.018.

regulates mitochondrial DNA copy number and transcription by selective degradation
of mitochondrial transcription factor A (TFAM). Proc Natl Acad Sci U S A 107,
18410-18415. 10.1073/pnas.1008924107.


Multiple domains of bacterial and human Lon proteases define substrate selectivity.
Emerg Microbes Infect 7, 149. 10.1038/s41426-018-0148-4.

degradation of unstructured protein by Escherichia coli Lon occurs via the slow,
sequential delivery of multiple scissile sites followed by rapid and synchronized
peptide bond cleavage events. Biochemistry 52, 5629-5644. 10.1021/bi4008319.

Endoplasmic Reticulum Homeostasis and Distal Epithelial Cell Survival during Lung
Development. Am J Respir Cell Mol Biol 55, 135-149. 10.1165/rcmb.2015-0327OC.

between mitochondrial unfolded protein stress and mitochondrial translation

69. Shi, H., O’Reilly, V.C., Moreau, J.L., Bewes, T.R., Yam, M.X., Chapman, B.E.,
stress induces the unfolded protein response, resulting in heart defects. Development
143, 2561-2572. 10.1242/dev.136820.


### Table 1. Clinical summary of 827 CDH probands

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

\(^a\)Neurodevelopmental conditions include congenital abnormalities in central nervous system, and developmental delay or neuropsychiatric disorders based on the follow-up developmental evaluations.

\(^b\)does not include pulmonary hypoplasia or hypertension.
Table 2. Top CDH associated genes predicted by pLI-stratified extTADA with \( \geq 2 \) de novo predicted deleterious variant.

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

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

\textsuperscript{a}: known CDH risk genes
Table 3. Recurrent genes or regions in *de novo* CNVs

<table>
<thead>
<tr>
<th>Recurrent</th>
<th>ID</th>
<th>Cytoband</th>
<th>Start</th>
<th>End</th>
<th>Size (kb)</th>
<th>Type</th>
<th>Known risk CDH/CHD/NDD genes</th>
<th>qPCR confirm</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>CSMD1</em> (CUB And Sushi Multiple Domains 1)</td>
<td>CDH1162</td>
<td>8p23.2</td>
<td>3846934</td>
<td>4073105</td>
<td>226</td>
<td>DEL</td>
<td>-</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>CDH12-0009</td>
<td>8p23.3p23.1</td>
<td>191301</td>
<td>7355200</td>
<td>7164</td>
<td>DEL</td>
<td>FBXO25</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>CDH863</td>
<td>8p23.3p23.1</td>
<td>200601</td>
<td>7155000</td>
<td>6954</td>
<td>DUP</td>
<td>FBXO25</td>
<td>Yes</td>
</tr>
<tr>
<td><em>GPHN</em> (Gephyrin)</td>
<td>C1235FSL_169</td>
<td>14q23.3</td>
<td>66559001</td>
<td>66630200</td>
<td>71</td>
<td>DEL</td>
<td>GPHN</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>CDH14-0009</td>
<td>14q23.3</td>
<td>66636783</td>
<td>66668074</td>
<td>31</td>
<td>DEL</td>
<td>GPHN</td>
<td>Yes</td>
</tr>
<tr>
<td>17q12</td>
<td>h1237LPLa1</td>
<td>17q12</td>
<td>36441801</td>
<td>37892100</td>
<td>1450</td>
<td>DEL</td>
<td>GGNB2</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>CDH05-0040</td>
<td>17q12</td>
<td>36442521</td>
<td>37963800</td>
<td>1521</td>
<td>DEL</td>
<td>GGNB2</td>
<td>Yes</td>
</tr>
<tr>
<td>21q</td>
<td>CDH10-0022</td>
<td>21q</td>
<td>13000000</td>
<td>46700000</td>
<td>33700</td>
<td>DUP</td>
<td>SIM2;SON;HMGN1;SIK1;COL6A1;DYRK1A;DSCAM;DIP2A;KCNJ6</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>CDH10-0038</td>
<td>21q</td>
<td>13188001</td>
<td>46700000</td>
<td>33512</td>
<td>DUP</td>
<td>SIM2;SON;HMGN1;SIK1;COL6A1;DYRK1A;DSCAM;DIP2A;KCNJ6</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>CDH10-0042</td>
<td>21q</td>
<td>13192901</td>
<td>46684100</td>
<td>33492</td>
<td>DUP</td>
<td>SIM2;SON;HMGN1;SIK1;COL6A1;DYRK1A;DSCAM;DIP2A;KCNJ6</td>
<td>-</td>
</tr>
</tbody>
</table>
Table 4. Phenotypes of CDH cases with ultra-rare deleterious variants in *LONP1*. Deleterious heterozygous variants include LGD and missense with CADD ≥25 with minor allele frequency (MAF) <1e-5 across all the gnomAD v3.0 genomes.

<table>
<thead>
<tr>
<th>cDNA change</th>
<th>Protein Change</th>
<th>Sample ID</th>
<th>Sex</th>
<th>Genetic ancestry</th>
<th>Inheritance</th>
<th>Family history of other birth defects</th>
<th>Familial CDH</th>
<th>1M PH</th>
<th>3M PH</th>
<th>Vital status</th>
<th>ECMO</th>
<th>Complex</th>
<th>Neuro-related</th>
<th>Other Congenital Anomalies/medical problems</th>
</tr>
</thead>
<tbody>
<tr>
<td>c.296dupA</td>
<td>p.S100QfsX46</td>
<td>01-0794</td>
<td>Female</td>
<td>EUR</td>
<td>paternal</td>
<td>No</td>
<td></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>No</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>c.398C&gt;G</td>
<td>p.P133R</td>
<td>01-0672</td>
<td>Female</td>
<td>AFR</td>
<td>paternal</td>
<td>No</td>
<td>Affected sibling (+)</td>
<td>Unk</td>
<td>Unk</td>
<td>Alive Yes Yes No</td>
<td>Yes</td>
<td>Yes</td>
<td>GI anomaly, GU anomaly</td>
<td>congenital cataracts</td>
</tr>
<tr>
<td>c.398C&gt;G</td>
<td>p.P133R</td>
<td>01-0670</td>
<td>Male</td>
<td>EUR</td>
<td>AFR</td>
<td>paternal</td>
<td>Affected sibling (+)</td>
<td>-</td>
<td>-</td>
<td>Deceased Yes Yes</td>
<td>No</td>
<td>No</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>c.520G&gt;A</td>
<td>p.G181E</td>
<td>01-0670</td>
<td>Male</td>
<td>EUR</td>
<td>de novo</td>
<td>No</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>Deceased Yes No</td>
<td>No</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>c.639G&gt;T</td>
<td>p.X213_splice</td>
<td>04-0022</td>
<td>Female</td>
<td>EUR</td>
<td>paternal</td>
<td>Paternal half-brother with Klippel-Feil PH (N/T)</td>
<td>Affected sibling (N/T), Affected paternal grandmother (+, de novo)</td>
<td>Severe</td>
<td>-</td>
<td>Deceased Yes No</td>
<td>No</td>
<td>No</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>c.792delA</td>
<td>p.P264RfsX5</td>
<td>09-0003</td>
<td>Female</td>
<td>EUR</td>
<td>maternal</td>
<td>Maternal great-uncle with suspected cerebral palsy</td>
<td>No</td>
<td>Severe</td>
<td>-</td>
<td>Deceased Yes No</td>
<td>No</td>
<td>-</td>
<td>Seizures</td>
<td>No</td>
</tr>
<tr>
<td>c.851delA</td>
<td>p.G284HfsX1</td>
<td>1428</td>
<td>Female</td>
<td>EUR</td>
<td>maternal</td>
<td>No</td>
<td>No</td>
<td>Mild</td>
<td>Mild</td>
<td>Alive No No No short stature</td>
<td>No</td>
<td>No</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>c.1123C&gt;A</td>
<td>p.L375M</td>
<td>04-0045</td>
<td>Male</td>
<td>EUR</td>
<td>paternal</td>
<td>Paternal uncle: neonatal death due to brain abnormalities (thrombophlebitis)?</td>
<td>No</td>
<td>None</td>
<td>None</td>
<td>Alive No No No</td>
<td>No</td>
<td>No</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>c.1262delA</td>
<td>p.R421LfsX7</td>
<td>1733</td>
<td>Female</td>
<td>EUR</td>
<td>maternal</td>
<td>No</td>
<td>Maternal uncle with suspected CDH (N/T)</td>
<td>-</td>
<td>-</td>
<td>Deceased Yes Yes global encephalopathy, seizures</td>
<td>CHD</td>
<td>No</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>c.1574T&gt;C</td>
<td>p.P525L</td>
<td>01-1279</td>
<td>Male</td>
<td>EUR</td>
<td>maternal*</td>
<td>No</td>
<td>Affected sibling (+)</td>
<td>-</td>
<td>-</td>
<td>Deceased Yes Yes No</td>
<td>No</td>
<td>No</td>
<td>-</td>
<td>No, bilateral CDH</td>
</tr>
<tr>
<td>c.1624C&gt;T</td>
<td>p.R542*</td>
<td>01-0113</td>
<td>Male</td>
<td>EUR</td>
<td>paternal</td>
<td>Mother with Klippel-Feil syndrome, Sphenopalatine deformity of scapula, crossed fused ectopia (Arnold-Chiari malformation I)</td>
<td>No</td>
<td>Severe</td>
<td>Severe</td>
<td>Deceased Yes Yes No</td>
<td>No</td>
<td>-</td>
<td>Pyloric stenosis</td>
<td>-</td>
</tr>
<tr>
<td>c.1629delF</td>
<td>p.E543delE</td>
<td>04-0077</td>
<td>Female</td>
<td>EUR</td>
<td>maternal</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Severe</td>
<td>-</td>
<td>Deceased Yes Yes</td>
<td>CHD</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>c.1709C&gt;T</td>
<td>p.P950L</td>
<td>04-0031</td>
<td>Female</td>
<td>EUR</td>
<td>unknown (singleton)</td>
<td>Father with residual post axial polydactyly</td>
<td>No</td>
<td>Severe</td>
<td>-</td>
<td>Deceased Yes No</td>
<td>No</td>
<td>No</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>c.1733G&gt;C</td>
<td>p.E591D</td>
<td>1511</td>
<td>Male</td>
<td>EUR</td>
<td>de novo</td>
<td>No</td>
<td>No</td>
<td>Mild</td>
<td>Mild</td>
<td>Alive No No No</td>
<td>No</td>
<td>No</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>c.1796C&gt;T</td>
<td>p.R597*</td>
<td>01-0582</td>
<td>Male</td>
<td>AFR</td>
<td>unknown (singleton)</td>
<td>No</td>
<td>No</td>
<td>None</td>
<td>None</td>
<td>Deceased No Yes</td>
<td>CHD</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>c.1801G&gt;T</td>
<td>p.X632_splice</td>
<td>1449</td>
<td>Female</td>
<td>EUR</td>
<td>maternal</td>
<td>Unknown</td>
<td>Unknown</td>
<td>-</td>
<td>-</td>
<td>Deceased No No</td>
<td>No</td>
<td>No</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>c.1913C&gt;T</td>
<td>p.T638M</td>
<td>01-0657</td>
<td>Female</td>
<td>EUR</td>
<td>AMR</td>
<td>de novo</td>
<td>Unknown</td>
<td>Severe</td>
<td>Moderate</td>
<td>Alive Yes No Yes</td>
<td>CHD, PH</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>c.1913C&gt;T</td>
<td>p.T638M</td>
<td>01-0513</td>
<td>Female</td>
<td>EUR</td>
<td>paternal</td>
<td>No</td>
<td>Affected sibling (N/T), Father with R evagination (+), paternal grandfather with R evagination (+)</td>
<td>-</td>
<td>-</td>
<td>Deceased Yes No</td>
<td>No</td>
<td>No</td>
<td>-</td>
<td>No</td>
</tr>
<tr>
<td>c.2122G&gt;A</td>
<td>p.G706S</td>
<td>04-0025</td>
<td>Male</td>
<td>EUR</td>
<td>paternal</td>
<td>Father with cleft palate</td>
<td>No</td>
<td>Severe</td>
<td>-</td>
<td>Deceased Yes No</td>
<td>No</td>
<td>-</td>
<td>Seizures</td>
<td>No</td>
</tr>
<tr>
<td>c.2122G&gt;A</td>
<td>p.G706S</td>
<td>m.1021LElit</td>
<td>Female</td>
<td>EUR</td>
<td>maternal</td>
<td>No</td>
<td>No</td>
<td>-</td>
<td>-</td>
<td>Deceased Yes Yes</td>
<td>CHD</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>c.2263C&gt;G</td>
<td>p.R785G</td>
<td>01-1279</td>
<td>Male</td>
<td>EUR</td>
<td>paternal*</td>
<td>No</td>
<td>Affected sibling (+)</td>
<td>-</td>
<td>-</td>
<td>Deceased No No</td>
<td>No</td>
<td>No</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>c.2263C&gt;G</td>
<td>p.R785G</td>
<td>09-0024</td>
<td>Male</td>
<td>EUR</td>
<td>paternal</td>
<td>Maternal great-aunt with CHD</td>
<td>No</td>
<td>Unk</td>
<td>Unk</td>
<td>Alive No Yes No</td>
<td>CHD</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>c.2461G&gt;C</td>
<td>p.A821P</td>
<td>03-0008</td>
<td>Male</td>
<td>EUR</td>
<td>maternal</td>
<td>3rd degree maternal history: unilateral arm agenesis</td>
<td>No</td>
<td>-</td>
<td>-</td>
<td>Deceased Yes No</td>
<td>No</td>
<td>No</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>c.2711dupA</td>
<td>p.Y909CfsX7</td>
<td>01-4752</td>
<td>Female</td>
<td>EUR</td>
<td>paternal</td>
<td>No</td>
<td>Paternal aunt with possible CDH (N/T)</td>
<td>Unk</td>
<td>Unk</td>
<td>Alive Yes No No</td>
<td>No</td>
<td>No</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

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 *LONP1* variant
- negative for familial *LONP1* variant
N/T = not tested for familial *LONP1* variant
*cases carried ballelic heterozygous variants
Table 5. *LONP1* deleterious rare variants carriers are associated with higher mortality and need for ECMO.

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

The bold p-values highlight significance. ECMO: extracorporeal membrane oxygenation; PH: pulmonary hypertension.
Figure 1. Burden of *de novo* coding variants in CDH compared to expectation. (A) LGD among all genes; (B) D-mis among all genes; (C) LGD among constrained genes; (D) D-mis among constrained genes. P values between cases and expectation by Poisson test are labeled for each bar. P values between females and males, complex and isolated cases by binormal test are labeled for each pair. Significant P values are highlighted in bold.
Figure 2. Gene-based association analysis using 748 CDH cases and 11,220 controls across all populations. (A) Results of a binomial test confined to ultra-rare LGD and D-Mis variants or D-Mis only variants in 18,939 protein-coding genes. Horizontal blue line indicates the Bonferroni-corrected threshold for significance. (B) Complete list of top association genes with permutation P values <1×10^{-4}. *: a gene-specific CADD score threshold for defining D-Mis that maximized the burden of ultra-rare deleterious variants in cases compared to controls; #: numbers of deleterious variants; a: MIM 600539; b: no MIM number.
Figure 3. Variant locations in LONPI (GenBank: NM_004793.3) of CDH and CODAS syndrome. There are three main domains in LONP1, N-terminal Lon domain, ATP binding domain and proteolytic domain. Positions indicated at upper structure are variants in CDH. Deleterious heterozygous variant such as LGD and missense with CADD ≥25 and allele frequency (AF) <1e-5 across all gnomAD genomes in CDH are presented. Deleterious missense is presented in purple, LGD in yellow, inframe variant in pink. Inheritance pattern were labelled in circles of variants (P: paternal; M: maternal; D: de novo; U: singleton unknown). Positions at lower structure are variants in published CODAS syndrome samples. CODAS syndrome is caused by biallelic variants in LONPI, including homozygous (H) or compound heterozygous (C) variants in the diamonds.
Figure 4. Fetal MRI images of bilateral CDH. (A) Sagittal view shows dorsal herniation of the stomach, ventral herniation of the liver, and anterior displacement of lung remnant. (B) Coronal view shows bilateral herniation of the fetal liver filling both the right and left hemithorax and no lung tissue.
Figure 5. Predicted 3D structure of LONP1 protein using SWISS-Model. (A) Variants in ATPase domain (gray) of CDH (red) and CODAS (blue). CODAS variants (p.A670-pA724) are clustered at alpha-helix in ATPase domain. (B) Variants in Protease domain (yellow) of CDH (red) and CODAS (blue). CDH variants p.A821, S866 and CODAS variants p.A927 are located at alpha-helix.
Figure 6. Inactivation of Lonp1 in mice led to disrupted lung development and lethality at birth. A. Gene structure of mouse Lonp1\(^{fl}\) conditional allele before and after cre-mediated recombination of the loxP sites (red triangles). Recombination led to a premature stop codon (arrow) in the second exon. B. Number of embryos genotyped at perinatal stage, showing 100% lethality of the mutant embryos. C. Representative mutant and control embryos at embryonic day (E)18.5, the day of birth. D. Representative mutant and control lungs at E18.5. Scale bars as indicated.