Common variants increase risk for congenital diaphragmatic hernia within the context of *de novo* variants

Authors

Lu Qiao, Carrie L. Welch, Rebecca Hernan, ..., Erwin Brosens, Yufeng Shen, Wendy K. Chung

Correspondence

ys2411@cumc.columbia.edu (Y.S.), wendy.chung@childrens.harvard.edu (W.K.C.)

Congenital diaphragmatic hernia (CDH) is a severe congenital anomaly. Genetic diagnoses are made in ~30% of people with CDH, but the rest of the cases remain unexplained. We identified rare variants in novel genes and common variants in two loci that support a polygenic model for CDH genetic architecture.

Qiao et al., 2024, The American Journal of Human Genetics *111*, 1–20
November 7, 2024
https://doi.org/10.1016/j.ajhg.2024.08.024

CellPress

Please cite this article in press as: Qiao et al., Common variants increase risk for congenital diaphragmatic hernia within the context of *de novo* variants, The American Journal of Human Genetics (2024), https://doi.org/10.1016/j.ajhg.2024.08.024

ARTICLE

Common variants increase risk for congenital diaphragmatic hernia within the context of *de novo* variants

Lu Qiao,[1,2] Carrie L. Welch,[1] Rebecca Hernan,[3] Julia Wynn,[1] Usha S. Krishnan,[1] Jill M. Zalieckas,[4,5] Terry Buchmiller,[4] Julie Khlevner,[1] Aliva De,[1] Christiana Farkouh-Karoleski,[1] Amy J. Wagner,[6] Andreas Heydweiller,[7] Andreas C. Mueller,[8] Annelies de Klein,[9] Brad W. Warner,[10] Carlo Maj,[11,37] Dai Chung,[12] David J. McCulley,[13] David Schindel,[14] Douglas Potoka,[15] Elizabeth Fialkowski,[16] Felicitas Schulz,[17] Florian Kipfmuller,[8] Foong-Yen Lim,[18] Frank Magielsen,[9] George B. Mychaliska,[19] Gudrun Aspelund,[1] Heiko Martin Reutter,[20] Howard Needelman,[21] J. Marco Schnater,[22] Jason C. Fisher,[23] Kenneth Azarow,[16] Mahmoud Elfiky,[24] Markus M. Nöthen,[25] Melissa E. Danko,[12] Mindy Li,[26] Przemyslaw Kosiński,[27] Rene M.H. Wijnen,[22] Robert A. Cusick,[21] Samuel Z. Soffer,[28] Suzan C.M. Cochius-Den Otter,[29] Thomas Schaible,[30] Timothy Crombleholme,[14] Vincent P. Duron,[31] Patricia K. Donahoe,[32,33] Xin Sun,[13] Frances A. High,[4,32,34] Charlotte Bendixen,[7] Erwin Brosens,[9] Yufeng Shen,[2,35,36,*] and Wendy K. Chung[1,3,*]

Summary

Congenital diaphragmatic hernia (CDH) is a severe congenital anomaly often accompanied by other structural anomalies and/or neurobehavioral manifestations. Rare *de novo* protein-coding variants and copy-number variations contribute to CDH in the population. However, most individuals with CDH remain genetically undiagnosed. Here, we perform integrated *de novo* and common-variant analyses using 1,469 CDH individuals, including 1,064 child-parent trios and 6,133 ancestry-matched, unaffected controls for the genome-wide association study. We identify candidate CDH variants in 15 genes, including eight novel genes, through deleterious *de novo* variants. We further identify two genomic loci contributing to CDH risk through common variants with similar effect sizes among Europeans and Latinx. Both loci are in putative transcriptional regulatory regions of developmental patterning genes. Estimated heritability in common variants is ~19%. Strikingly, there is no significant difference in estimated polygenic risk scores between isolated and complex CDH or between individuals harboring deleterious *de novo* variants and individuals without these variants. The data support a polygenic model as part of the CDH genetic architecture.

Introduction

Congenital anomalies are the leading cause of death during the first year of life and account for half of all pediatric hospitalizations.[1] Congenital diaphragmatic hernia (CDH) (MIM: 142340) is a congenital anomaly characterized by incomplete closure of the diaphragm early in gestation and herniation of fetal abdominal organs into the chest.

[1]Department of Pediatrics, Columbia University Irving Medical Center, New York, NY 10032, USA; [2]Department of Systems Biology, Columbia University Irving Medical Center, New York, NY 10032, USA; [3]Department of Pediatrics, Boston Children's Hospital, Harvard Medical School, Boston, MA 02115, USA; [4]Department of Surgery, Boston Children's Hospital, Harvard Medical School, Boston, MA 02115, USA; [5]Department of Anesthesiology, Boston Children's Hospital, Harvard Medical School, Boston, MA 02115, USA; [6]Children's Hospital of Wisconsin, Medical College of Wisconsin, Milwaukee, WI 53226, USA; [7]Department of General, Visceral, Vascular, and Thoracic Surgery, Unit of Pediatric Surgery, University Hospital Bonn, Bonn, Germany; [8]Department of Neonatology and Pediatric Intensive Care, Children's Hospital, University of Bonn, Bonn, Germany; [9]Department of Clinical Genetics, Erasmus MC Sophia Children's Hospital, Rotterdam, the Netherlands; [10]Washington University School of Medicine, St. Louis, MO 63110, USA; [11]Institute for Genomic Statistics and Bioinformatics, University of Bonn, Bonn, Germany; [12]Monroe Carell Jr. Children's Hospital at Vanderbilt, Nashville, TN 37232, USA; [13]Department of Pediatrics, San Diego Medical School, University of California, San Diego, San Diego, CA 92092, USA; [14]Medical City Children's Hospital, Dallas, TX 75230, USA; [15]University of Pittsburgh, Pittsburgh, PA 15224, USA; [16]Oregon Health and Science University, Portland, OR 97239, USA; [17]Department of Hematology, Oncology and Clinical Immunology, University Hospital Düsseldorf, Düsseldorf, Germany; [18]Cincinnati Children's Hospital Medical Center, Cincinnati, OH 45229, USA; [19]University of Michigan Health System, Ann Arbor, MI 48109, USA; [20]Neonatology and Pediatric Intensive Care, Department of Pediatrics and Adolescent Medicine, University Hospital Erlangen, Erlangen, Germany; [21]University of Nebraska Medical Center College of Medicine, Omaha, NE 68114, USA; [22]Department of Pediatric Surgery, Erasmus MC Sophia Children's Hospital, Rotterdam, the Netherlands; [23]New York University Grossman School of Medicine, Hassenfeld Children's Hospital at NYU Langone, New York, NY 10016, USA; [24]Cairo University, Cairo 11432, Egypt; [25]Institute of Human Genetics, University of Bonn, Bonn, Germany; [26]Rush University Medical Center, Chicago, IL 60612, USA; [27]Department of Obstetrics, Perinatology and Gynecology, Medical University of Warsaw, 02-091 Warsaw, Poland; [28]Northwell Health, New York, NY 11040, USA; [29]Department of Neonatology and Pediatric Intensive Care, Erasmus MC Sophia Children's Hospital, Rotterdam, the Netherlands; [30]Department of Neonatology, University Children's Hospital Mannheim, University of Heidelberg, Mannheim, Germany; [31]Department of Surgery (Pediatrics), Columbia University Irving Medical Center, New York, NY 10032, USA; [32]Pediatric Surgical Research Laboratories, Massachusetts General Hospital, Boston, MA 02114, USA; [33]Department of Surgery, Harvard Medical School, Boston, MA 02115, USA; [34]Department of Pediatrics, Massachusetts General Hospital, Boston, MA 02114, USA; [35]Department of Biomedical Informatics, Columbia University Irving Medical Center, New York, NY 10032, USA; [36]JP Sulzberger Columbia Genome Center, Columbia University Irving Medical Center, New York, NY 10032, USA
[37]Present address: Center for Human Genetics, University of Marburg, Marburg, Germany
*Correspondence: ys2411@cumc.columbia.edu (Y.S.), wendy.chung@childrens.harvard.edu (W.K.C.)
https://doi.org/10.1016/j.ajhg.2024.08.024.

Please cite this article in press as: Qiao et al., Common variants increase risk for congenital diaphragmatic hernia within the context of *de novo* variants, The American Journal of Human Genetics (2024), https://doi.org/10.1016/j.ajhg.2024.08.024

Herniation causes compression of the developing lungs, leading to lung hypoplasia and altered pulmonary vascular development, which clinically manifests as pulmonary hypertension. Additional primary defects in lung development might be pleiotropic effects of genes involved in diaphragm development.[2,3] CDH affects ~1/3,000 live births and is associated with high morbidity and mortality. About 60% of CDH is isolated, and 40% is associated with other congenital anomalies ("complex"). The most common additional anomalies include cardiovascular, neurodevelopmental, gastrointestinal, skeletal, genitourinary, and kidney defects; other pulmonary defects; and cleft lip and/or cleft palate.[4] Despite advances in genetic testing, fetal interventions, postnatal extracorporeal membrane oxygenation (ECMO), and gentle-ventilation strategies, CDH continues to be associated with ~35% mortality worldwide; there is higher mortality for complex CDH.[5]

The common pathology in pulmonary development in the setting of CDH includes reduced airway branching and alveolar surface area, decreased pulmonary vascularization, aberrant vascular smooth-muscle-cell development with increased pulmonary arterial muscularization, and cardiac hypoplasia.[4] There is significant heterogeneity in the degree of pulmonary and vascular remodeling across CDH individuals, and clinical severity is variable.[6,7] The etiology of CDH is most likely multifactorial and probably includes maternal factors (age, micronutrients, and pregestational metabolic disease),[4] fetal microenvironmental influences on gene expression,[8,9] and genetic factors (chromosomal, rare inherited, and *de novo*).[4] Understanding the genetics and associated phenotypes of CDH can help with prognosis and guide clinical care. Genetic diagnoses are identified in ~30% of CDH; variants in individual genes contribute to at most 1–3%. We previously screened for copy-number variants in a subset of the cohort and identified variants in 4.2% (31/734 individuals).[10,11] The most common genetic group of known CDH risk factors (identified in 13%–22% of individuals) is rare *de novo* variants of large effect size, as is typical for neonatal diseases with high mortality and low reproductive fitness.[12,13] We recently identified both *de novo* and ultra-rare inherited variants in *LONP1*, LON peptidase 1 (mitochondrial), as genetic contributors to CDH, suggesting a more complex genetic architecture.[3] Affected individuals with *LONP1* variants had variable clinical phenotypes with a higher requirement for ECMO and higher mortality than those without *LONP1* variants. Lung epithelial-specific deletion of *Lonp1* in mice resulted in perinatal death due to lung hypoplasia. Despite these advances in identification of genes with CDH-causative variants, most CDH in the population remains unexplained.

To identify additional genetic causes of CDH, we performed rare *de novo* variant analysis in an expanded CDH trio cohort. Because CDH-associated comorbidities and outcomes are heterogeneous in nature, we hypothesized that common variants might play a role in CDH risk. Therefore, we also performed a case-control genome-wide association study (GWAS) with common variants (single-nucleotide polymorphisms, SNPs). Using a CDH cohort of 1,469 participants of the Diaphragmatic Hernia Research & Exploration; Advancing Molecular Science (DHREAMS) study[14] combined with CDH individuals from Boston Children's Hospital and Massachusetts General Hospital (Boston cohort),[15] we identified eight novel candidate genes on the basis of ≥ 2 *de novo* variants in constrained genes with high diaphragm expression. We defined novelty by the presence of previous reports of no more than one individual with a rare variant associated with a developmental syndrome and CDH (see references for reviews[16,17,18]). We estimated that these monogenic factors confer relative CDH risk ranging from 10 to 30. We also identified two susceptibility loci of common variants with replication in an independent cohort. Variant annotation indicated that the lead SNPs, SNP rs55705711 (chromosome 3) and SNP rs7777647 (chromosome 7), are most likely located in transcriptional regulatory regions of established developmental patterning genes. Together, the data support the contribution of polygenic risk as part of the CDH genomic architecture.

Methods

Ethics approval and consent to participate

This study was approved by the institutional review boards of Columbia University Irving Medical Center, Boston Children's Hospital, Massachusetts General Hospital, University Hospital Bonn, and Erasmus MC Sophia Children's Hospital and by the individual review boards at each of the DHREAMS enrolling centers' institutions. All CDH individuals have signed consent forms which are on file at Columbia University Irving Medical Center, Boston Children's Hospital, Massachusetts General Hospital, University Hospital Bonn, or Erasmus MC Sophia Children's Hospital. Written informed consent for publication was obtained at enrollment. No protected health information on any individuals enrolled for this study have been forwarded to the data analyzing group. All research using these CDH samples conformed to the principles of the Helsinki Declaration.

Study subjects and controls

We enrolled 1,469 participants who were part of the DHREAMS study[14] or Boston cohort[15] and had a radiological confirmation of a diaphragm defect. Clinical data were collected prospectively from medical records and entered into a central REDCap database.[19] Blood, skin biopsy, or saliva specimens were collected from 1,064 affected child-parent trios and 405 singletons for genetic analyses (Table S1).

Participants were classified as having isolated or complex CDH on the basis of the presence of a diaphragm defect alone (isolated) or at least one additional structural congenital anomaly (i.e., congenital heart defect, central nervous system anomaly, gastrointestinal anomaly, skeletal anomaly, genitourinary anomaly, or cleft or lip palate), moderate to severe developmental delay, or other neuropsychiatric phenotypes (complex). Pulmonary hypoplasia, cardiac displacement, intestinal herniation, and pulmonary hypertension were considered to be part of the diaphragm defect sequence and not independent malformations. Participant

health history data, including postoperative pulmonary hypertension; mortality or survival status prior to initial discharge; and ECMO intake were collected as described previously.[10] The study was approved by the institutional review boards of all participating centers and Columbia University Irving Medical Center, and signed informed consent was obtained.

We conducted *de novo* variant analysis by using 1,064 trios with genome or exome sequencing and a discovery genetic association analysis by using 1,443 affected individuals with genome sequencing. A subset of participants ($n = 827$) were described in our previous studies.[10,11,20] The discovery control group consisted of 6,133 unaffected parents with genome sequencing data (Table S1) from Simons Powering Autism Research for Knowledge (SPARK)[21] (data available under managed access from Simons Foundation Autism Research Initiative [SFARI]).

The replication cohort consisted of 450 German and Dutch CDH-affected individuals recruited at the University Hospital of Bonn, University Hospital of Mannheim, or Erasmus Medical Center, Rotterdam. Radiological confirmation of diaphragm defects showed that 79.6% (358/450) had left-sided hernias and that 71.3% ($n = 321$) were isolated vs. 25.1% ($n = 113$) complex CDH (3.6% unknown). Of the individuals with complex CDH, 37.2% had congenital heart disease (CHD), 16.8% neurodegenerative disease (NDD), 12.4% genitourinary anomalies, 6.2% skeletal anomalies, and 2.7% gastrointestinal anomalies and/or cleft lip or palate. Blood or saliva specimens were obtained from all participants for genetic analyses. The study was approved by the local ethics boards of each institution (protocol no. 193.948/2000/159, addendum numbers 1 and 2) and adhered to guidelines of the Declaration of Helsinki. Written informed consent was obtained from all participants or their guardians. The control cohort consisted of 4,144 unaffected German individuals from the Heinz Nixdorf Recall Study[22] and 671 unaffected Dutch parents of children with various conditions (developmental delay, intellectual disability, or major congenital anomalies) admitted to Erasmus Medical Center and referred for trio copy-number variant analysis.

Genome sequencing

Of 642 new participants with CDH in the discovery cohort, 637 were sequenced after genomic libraries were prepared with the Illumina TruSeq DNA PCR-Free Library Prep Kit (Illumina, San Diego, CA. USA). Five were sequenced after libraries were prepared with the TruSeq DNA PCR-Plus Library Prep Kit. All CDH samples had an average fragment length of ~350 bp and were sequenced on the Illumina HiSeq X platform, generating 150 bp paired-end reads. SPARK control samples were sequenced on the Illumina NovaSeq 6000 system[21] (Table S1).

Sequencing data of affected and control individuals were processed on a pipeline implementing GATK Best Practice v.4.0 as described.[20,23] In brief, we used BWA-MEM[24] to map and align paired-end reads to human reference genome GRCh38/hg38, Picard v.1.93[25] MarkDuplicates to identify and flag PCR duplicates, and GATK v.4.1[26] HaplotypeCaller to generate individual-level gVCF files from the aligned sequence data. We then performed joint calling of variants subdivided into batch 1 ($n = 762$ cases and 4,031 controls, primarily from our previous report[11]) or batch 2 (681 cases and 2,102 controls, new samples) (Table S1). Variant quality score recalibration was performed with GATK. Sample sex was validated by comparison of the ratio of heterozygous to homozygous genotypes on X chromosomes, which also identified chromosomal aneuploidies (Figure S1A).

De novo variant quality control and annotation

Of 642 CDH-affected participants not previously sequenced,[11] 237 were complete proband-parent trios and were used for calling *de novo* variants that were present in proband genomes but absent in both parents. We limited genome sequencing data to coding regions on the basis of coding sequences and canonical splice sites of all GENCODE v.27 coding genes (https://www.gencodegenes.org/human/release_27.html). We used a series of heuristic filters to identify *de novo* variants as previously described.[11] Inclusion criteria were that variants passed the VQSR filter with tranche $\leq$99.8 for SNVs and $\leq$99.0 for indels, had GATK's FisherStrand $\leq$25, and had quality by depth $\geq$2. We required candidate *de novo* variants in probands to have $\geq$5 reads supporting the alternative alleles, $\geq$20% alternative allele fraction, minimum Phred-scaled genotype quality = 60, and population allele frequency $\leq$0.01% in gnomAD v.2.1.1 (https://gnomad.broadinstitute.org/downloads#v2-liftover), and we required both parents to have $\geq$10 reference reads, <5% alternative allele fraction, and genotype quality $\geq$30. DeepVariant[27] was applied to all candidate *de novo* variants for *in silico* confirmation and only included the ones with PASS filter. We manually inspected all coding *de novo* variants in the Integrated Genomics Viewer[28] to rule out strand bias, end read, and repeat region artifacts. We aligned sequencing reads from the proband, mother, and father to confirm the *de novo* nature of the variants.

We used Ensembl Variant Effect Predictor[29] (VEP, Ensemble 102) and ANNOVAR[30] to annotate functional consequences, variant population frequencies, and in silico predictions of deleteriousness. All coding *de novo* variants were classified as synonymous, missense, in-frame, or likely gene-disrupting (LGD, including frameshift indels, canonical splice sites, and nonsense variants). We defined deleterious missense variants (D-mis) by phred-scaled combined annotation dependent depletion score $\geq$20 (CADD version 1.4).[31] Constrained genes were defined by gnomAD estimated probability of loss-of-function intolerance (pLI) $\geq$0.5,[32] and all remaining genes were treated as other genes. We used a less stringent pLI threshold than previously suggested[33] for defining constrained genes because it captured more known haploinsufficient genes important for heart and diaphragm development. To compare the characteristics between CDH individuals with and without likely damaging variants for common variant-related downstream analysis, we defined likely damaging variants as *de novo* LGD or D-mis in CDH known risk genes (see references for reviews[16,17,18]) or constrained genes.

Common-variant discovery: Quality control and imputation

We used KING[34] to resolve cryptic relatedness up to the third degree (Figure S1B) by keeping one sample of each pair. Using principal-component analysis (PCA), we compared individuals from discovery affected and control individuals to population panels from the 1000 Genomes Project (https://www.internationalgenome.org) to identify individuals who were more genetically similar to reference panels and evaluate population structure as implemented by Peddy v.0.4.8[35] (Figure S2). Unrelated affected and control individuals in the same ancestry groups were merged into one dataset for quality control and downstream analysis. We excluded SNPs with <20 reads, phred-scaled genotype quality < 20, and minor-allele frequency (MAF) < 0.01, that were successfully genotyped in <95% of individuals, and that had a Hardy-Weinberg equilibrium test p value $< 1 \times 10^{-6}$. Then, we performed more rigorous quality-control measures to remove erroneous genotype calls. SNPs were excluded on the basis of missing-genotype rate differences between

affected and control individuals; location in repeat masker regions (https://www.repeatmasker.org); no call in TOPMed Freeze 8[36]; and MAF differences between discovery control individuals and ancestry-matched gnomAD3.1.1 non-neuro individuals.[37]

Genome-wide imputation for non-genotyped autosomal SNPs was carried out with Michigan Imputation Server v.1.7.1.[38] Phase 3 with 30× coverage of 1000 Genomes[39] was used as the reference for imputing genotypes after phasing with Eagle v.2.4.[40] Imputation was based on genotyped SNPs that met genotyping quality-control criteria. Imputed SNPs with imputation r^2 (i.e., estimated squared correlation between imputed and actual genotypes) < 0.8 were excluded. Data for imputed SNPs with MAF $< 1\%$ or for SNPs imputed in $<90\%$ of individuals were also excluded from all statistical analyses.

Statistical analysis

De novo variant burden and extTADA analyses

We used a previously described model[41] to calculate baseline mutation rates for different classes of *de novo* variants in each GENCODE coding gene (https://www.gencodegenes.org/human/release_27.html). The observed number of variants in each gene set and group of affected individuals was then compared with the baseline expectation via a Poisson test. To identify risk genes on the basis of *de novo* variants, we used an empirical Bayesian method, extended transmission, and *de novo* association (extTADA).[42] The extTADA model was developed on the basis of a previous integrated empirical Bayesian model TADA[43] and estimates mean effect sizes and risk-gene proportions from genetic data by using the Markov Chain Monte Carlo process. To inform the parameter estimation with prior knowledge of developmental disorders, we stratified the genes into "constrained" and "other" genes and then estimated the parameters by applying the extTADA model to each group of genes as described previously.[11] After estimating the posterior probability of association of individual genes in each group, we combined both groups to calculate a final false-discovery rate (FDR) for each gene using the extTADA procedure.

Genome-wide association analysis

The association of CDH with alternate allele dosage (0, 1, or 2) was assessed with a logistic genetic model (Equation 1), including both genotyped and imputed SNPs. Within each ancestry case-control merged group, we performed principal-component analysis to control for population stratification by using Peddy.[35] In the discovery dataset, we adjusted the association analysis for covariates, sex and the first four principal components, by using a logistic regression implemented in PLINK 1.07[44]:

$$phenotype = 1 / \left(1 + e^{-(genotype + covariates)}\right) \qquad \text{(Equation 1)}$$

Genome-wide significance was set at $p < 5 \times 10^{-8}$. Manhattan and quantile-quantile plots were generated with the qqman package[45] in R v.4.1.0. LocusZoom v.1.4[46] was used for creating regional association plots. Genome-wide inflation values were calculated with Cochran's Q statistic χ^2 to test for heterogeneity between groups in stratified analyses.

Replication cohort and meta-analysis

Replication case and control genotyping was performed with the Illumina Infinium Global Screening Array v.1.0 (Dutch controls) or v.2.0 (all affected individuals and German control individuals). Samples were excluded if the computed sex did not match the sex reported in medical records or the Illumina-based genotype call rates were $<98\%$. Relatedness was calculated with KING, and only one sample was included if kinship coefficients suggested a 3rd-degree or higher relationship (threshold > 0.0442).[34] Principal components of affected-individual, control-indivdiual, and reference samples from the 1,000 Genomes Project Phase 3[47] were calculated with PLINK. Individuals with CDH and control samples clustering with non-European populations from the 1,000 Genomes Project were removed. Principal-component analysis was performed on the remaining affected and control individuals, and samples that deviated ≥ 6 standard deviations from the first or second principal components were excluded. Genotypes were phased with SHAPEIT v.2[48] and then imputed via IMPUTE2 v.2 (hg19)[49] with the 1,000 Genomes Project data as a reference. Variants with IMPUTE2 score <0.6 or minor-allele count <20 were excluded. Finally, 216 German and 178 Dutch CDH-affected individuals and 4,144 German and 675 Dutch control individuals were included in the genetic analyses. From these samples, 12,137,928 variants were successfully genotyped or imputed.

To merge the discovery and replication datasets, we performed a standard fixed-effects meta-analysis, implemented in METAL, on the basis of the estimated log odds ratios (ORs) and their standard errors.[50]

For association loci reaching genome-wide significance, we performed conditional analyses in PLINK,[44] by conditioning on the lead SNP at each locus, to test for the number of independent signals. To compare the role of common genetic variants by presence of isolated versus complex CDH, sex, and the presence or absence of damaging *de novo* variants, we used Fisher's exact test. The numbers of affected individuals with different genotypes for genome-wide significant SNPs were tested for nonrandom association with the three CDH categorical subset groupings.

Power estimation

To determine the power and sample size required for replication, we tested the probability of rejecting the null hypothesis, H_0 (here, effect size $\beta = 0$), at given significance thresholds when the data follow a specific alternative hypothesis, H_1. Here, H_1 is specified by counts of affected and control individuals, MAF, and effect size. We set the significance threshold for GWAS at $p < 5 \times 10^{-8}$ and replication at $p < 0.05$.

SNP heritability and polygenic risk score

SNP-based heritability

We used the linkage disequilibrium score regression software tool[51] to estimate SNP-based heritability of CDH with discovery summary statistics after adjusting for sex and the top four principal components. Because CDH-affected individuals were more prevalent in the discovery dataset than in the general population, SNP heritability estimates are given on a liability scale that assumes a prevalence of 1/3000[14] and a range of prevalence of 0.01%–0.05% to test the effect of prevalence by additivity.

Variance explained by significant SNPs

To estimate the variance explained by independent genome-wide-significant SNPs, we estimated R^2 as a goodness-of-fit measure in logistic full Equation 1 and the null formula (Equation 2) separately:

$$phenotype = 1 / \left(1 + e^{-covariates}\right) \qquad \text{(Equation 2)}$$

The variance explained per independent SNP was calculated as the R^2 difference (ΔR^2) between the full and null models.

For calculation of polygenic risk scores (PRSs), we randomly separated the discovery affected and control individuals from the European cohort into training ($n = 502$ affected individuals; $n = 2{,}282$ control individuals) and test ($n = 503$ affected individuals; $n = 2{,}283$ control individuals) sets while maintaining similar

Please cite this article in press as: Qiao et al., Common variants increase risk for congenital diaphragmatic hernia within the context of *de novo* variants, The American Journal of Human Genetics (2024), https://doi.org/10.1016/j.ajhg.2024.08.024

Table 1. Clinical phenotypes of 1,469 participants with CDH from the discovery cohort

Characteristics	Number	Percent
Sex		
Male	867	59.0%
Female	602	41.0%
Ancestry[a]		
European	1,047	71.3%
African American	104	7.1%
Latinx	206	14.0%
East Asian	26	1.8%
South Asian	27	1.8%
Other	59	4.0%
CDH side		
Left	1,047	71.3%
Right	209	14.2%
Bilateral/Center/Eventration/Other	68	4.6%
Unknown	145	9.9%
Timing of enrollment		
Fetal	83	5.7%
Neonatal	863	58.7%
Child	493	33.6%
Not specified	30	2.0%
CDH classification		
Isolated	924	62.9%
Complex	506	34.4%
Unknown	39	2.7%
Additional anomalies in complex CDH (n = 506)		
Cardiovascular	275	54.3%
Neurodevelopmental[b]	125	24.7%
Gastrointestinal	88	17.4%
Skeletal	73	14.4%
Genitourinary	62	12.3%
Kidney	32	6.3%
Pulmonary defects[c]	30	5.9%
Cleft lip or palate and/or micrognathia	24	4.7%
Discharge vital status (n = 1345)		
Alive	1,113	82.8%
Deceased	232	17.2%
Extracorporeal membrane oxygenation (ECMO; n = 1,210)		
Yes	352	29.1%
No	858	70.9%

(*Continued on next page*)

Please cite this article in press as: Qiao et al., Common variants increase risk for congenital diaphragmatic hernia within the context of *de novo* variants, The American Journal of Human Genetics (2024), https://doi.org/10.1016/j.ajhg.2024.08.024

Table 1. ***Continued***

Characteristics	Number	Percent
Pulmonary hypertension at 1 month (PH1; *n* = 446)		
None or mild	244	54.7%
Moderate or severe	202	45.3%
Pulmonary hypertension at 3 months (PH3; *n* = 326)		
None or mild	257	78.8%
Moderate or severe	69	21.2%

[a]Ancestry was determined by PCA and comparison of individuals from discovery affected and control individuals to population panels from the 1000 Genomes Project.
[b]Neurodevelopmental conditions include congenital abnormalities in the central nervous system and developmental delay or neuropsychiatric disorders as determined on the basis of follow-up developmental evaluations.
[c]Does not include pulmonary hypoplasia or hypertension.

trio/singleton ratios. We performed logistic regression adjusting for sex and the top four principal components in the training set to provide summary statistics of genotyped and imputed SNPs. In total, 1,326,338 common SNPs residing in independent linkage disequilibrium blocks were used in the analysis. We used the LDpred2 algorithm implemented in the "bigsnpr" package of R v.4.1.3[52] to generate PRSs in the test set with manual hyperparameter threshold set at $p < 5 \times 10^{-3}$ and liability scale heritability estimated as above. We used a Student's t test to compare the PRS in CDH test set with different characteristics.

Polygenic transmission disequilibrium

We used individual-level genotyped CDH trios in the test set ($n = 371$ trios) to generate proband, father, and mother PRSs by using the above PRS model. The mid-parent PRSs (PRS_{MP})[53] were defined (Equation 3) as follows:

$$PRS_{MP} = \frac{PRS_{father} + PRS_{mother}}{2} \quad \text{(Equation 3)}$$

We inferred polygenic transmission disequilibrium by comparing the difference between the means of the offspring and mid-parent PRS distributions by using a Student's t test.

Annotation of GWAS loci

We identified all genes within the same topologically associated domain (TAD) as candidate genes because the likelihood that an SNP acts on its target gene is higher when they are in close proximity in three-dimensional space.[54] Genomic coordinates of TADs were estimated from human induced-pluripotent-stem-cell-derived cardiomyocyte (iPSC-CM) promoter capture Hi-C data.[55] To visualize chromatin interaction data,[56] we queried Hi-C interactions around candidate genes in an H1 human embryonic stem cell line[57] (http://3dgenome.fsm.northwestern.edu/chic.php). We annotated the loci on the basis of chromatin immunoprecipitation sequencing data curated by ReMap 2022[58] for transcription regulator binding sites from the University of California Santa Clara Genome Browser.[59,60]

Single nucleus ATAC-sequencing from normal human lung and murine skeletal muscle

Single-nucleus assays for transposase-accessible chromatin sequencing (snATAC-seq) datasets were described previously;[61] data were generated by the University of California San Diego Lung Research Center under the National Heart, Lung, and Blood Institute LungMAP consoritium. In brief, snATAC-seq data were generated from three newborn, three ~3-year-old and three 30-year-old normal human lungs. Cell-type-resolved chromatin-accessible regions were identified and displayed as open-source data on LungMAP.net. Detailed methods have been described,[61] and data are available for query at LungMAP.net. For skeletal muscle, we extracted single-nucleus ATAC-seq data reported from mouse tissue collected at multiple stages of development: embryonic E14.5, fetal E18.5, neonatal P5, and adult 2 months of age.[62] We used the UCSC genome browser liftover tool[63] to align our top GWAS SNPS from reference genome hg38 (chr3:55559903–55572080) to mouse genome mm10 (chr14:28417328–28430803). We note that human *ERC2* aligned to mouse *Erc2*.

Results

Cohort characteristics

Demographic and clinical details of 1,469 participants from the DHREAMS and Boston cohorts are shown in Table 1 (see also Figures S1 and S2). The cohort was 59% male and 41% female. Fifty-nine percent were enrolled during the neonatal period, 34% as young children, and 5.7% prior to birth. 71% carried genotypes most genetically similar by PCA to the European, 14% to the Latinx, 7.1% to the African American, 1.8% to the East Asian, and 1.8% to the South Asian panels of the 1000 Genomes dataset. 71% of participants had left-sided hernias, and 63% had isolated CDH without other congenital anomalies. Of those with complex CDH, cardiovascular anomalies were the most prevalent (54%) and were followed by neurodevelopmental sequelae (25%), gastrointestinal anomalies (17%), and other anomalies. Thirty percent of individuals with CDH received extracorporeal membrane oxygenation while in the hospital, and 45% had a diagnosis of pulmonary hypertension at 1 month of age. 17% of individuals died before discharge from the hospital.

De novo variant analysis for CDH gene discovery

De novo variant analysis was performed on 1,064 child-parent trios, including 827 described in our previous report[3] and 237 new trios (Table S1). In total, 1,504 coding variants were identified in 794 (75%) individuals with CDH; these variants included 1,383 single-nucleotide variants (transition-transversion ratio = 2.71) and 121 short

insertions or deletions. The number of *de novo* variants per proband followed a Poisson distribution with a mean of 1.41 variants/proband (Figure S3).

Filtering for rare damaging variants identified 194 LGD and 692 D-mis variants. Consistent with previous studies,[11] we observed a 1.4-fold enrichment ($p = 1.40 \times 10^{-24}$) of rare damaging variants, especially of constrained genes, among participants with CDH when the mutation rate for these genes was compared to the background mutation rate (Table 2). Enrichment was observed for both LGD and D-mis variants, in both sexes, as well as isolated and complex CDH (Table 2). However, *de novo* LGD variants showed greater enrichment among females than males and among complex vs. isolated CDH.

Constrained genes are genes with fewer than expected LGD variants in the general population. The pLI-stratified extTADA analysis predicted a significant increase in the proportion of risk genes as constrained genes (0.072) compared to all genes (0.041) or non-constrained genes (0.04), and average relative risks ranged from 4.7 (D-mis variants) to 16.5 (LGD variants) for constrained genes (Table S2). We identified 15 genes with FDR < 0.1; these included four new candidate genes with FDR < 0.05: *STAG2* (STAG2 cohesin complex component), *SIN3A*, (SIN3 transcription regulator family member A), *POGZ* (pogo transposable element derived with ZNF domain), and *ZNF462* (zinc finger protein 462) (Table 3). The new candidates are all constrained genes with pLI ≥ 0.9, are expressed during diaphragm,[64] heart, or brain development[13] (Table S3), and have functional roles in chromatin remodeling, transcription regulation, mitosis, and protein phosphorylation. Three of the genes—*SIN3A*, *STAG2*, and *POGZ*—are genes with known causal variants for genetic syndromes that include CDH as a rare manifestation.[65–69] We note that *GPC3* (glypican 3, FDR = 0.072) has been reported with causal variants for Simpson-Golabi-Behmel syndrome type 1 (MIM: 312870), which also includes CDH as a rare manifestation.[70–73] The population attributable risk of causal *de novo* variants in CDH is estimated at 18.3% for D-mis and 6.9% for LGD variants (Figure S4). The complete list of genes with two or more LGD or D-mis *de novo* variants identified in individuals with CDH is provided in Table S3.

Variant and associated clinical phenotype details for the top 15 associated genes (extTADA FDR < 0.1) are provided in Table S4. We note that none of the individuals with rare *de novo* variants had CDH causal copy-number variants identified in our previous study.[10,11] Although most gene variants were associated with left-sided hernias, *MYRF*, *ALYREF*, *SIN3A*, and *STAG2* variants were associated with left- or right-sided hernias. Haploinsuffiency of *MYRF* (myelin regulatory factor) is known to cause cardiac urogenital syndrome (CUGS [MIM: 618280]) with abnormal development of the heart, genitourinary system, diaphragm, and lungs[20,74]; five out of eight variant heterozygotes identified in our cohort have complex CDH including congenital heart disease and genitourinary anomalies. The variant types include LGD and D-mis; the missense variants are located in conserved protein domains or local regions of constraint.[20] Rare variants in *GATA4* (GATA binding protein 4) and *GATA6* (GATA binding protein 6) are well-established causal factors in CDH congenital heart disease,[75] and six out of seven participants with CDH and variants in *GATA4* or *GATA6* (LGD and D-mis) have diagnoses of congenital heart disease. We note that *GATA4* c.848G>A (p.Arg283His) (GenBank: NM_002052.5) was recurrent in two unrelated participants. *GATA6* c.1366C>T (p.Arg456Cys) (GenBank: NM_005257.6) is an inactivating allele[76] and a hotspot; it is associated with congenital heart disease, pancreatic agenesis, diabetes, and CDH in four probands reported in ClinVar. We recently reported rare inherited and *de novo* *LONP1* variants associated with isolated and complex CDH.[3] Of four infants identified with *LONP1* (D-mis) variants in this study, three were isolated (two on ECMO during a hospital stay,) and one was complex CDH with congenital heart disease. Heterozygous *STAG2* variants cause Mellegama-Klein-Martinez syndrome (MIM: 301022), a cohesinopathy including NDD, CDH, and skeletal issues.[66] Two CDH infants identified in this study had LGD variants and comorbidities of skeletal anomalies and NDD or CHD. Chromosomal region 15q24 microdeletion (including *SIN3A*) syndrome includes CDH as a rare manifestation,[65] and gene-specific *SIN3A* variants are causal for Witteveen-Kolk syndrome (MIM: 616364), an NDD with skeletal and genital anomalies. Two infants with unique frameshift variants identified in this study have overlapping features of the chromosome 15 syndrome and Witteveen-Kolk syndrome. Heterozygous *POGZ* variants are causal for White-Sutton syndrome (MIM: 616364), an NDD with a phenotypic spectrum including ophthalmic and genitourinary anomalies and CDH, and the two CDH participants with LGD *POGZ* variants have complex congenital anomalies overlapping this spectrum. Heterozygous *ZNF462* LGD variants are causal for Weiss-Kruszka syndrome (MIM: 618619), a multiple-congenital-anomaly syndrome including NDD (with autism spectrum disorder and facial dysmorphism) and CHD. Most of the Weiss-Kruszka syndrome LGD variants are clustered in exon 3 (13/17 variants).[77] The nonsense variant identified in this study is also located in exon 3, and the participant had CHD and genitourinary anomalies. The participant with a *ZNF462* canonical splicing variant in intron 3 was diagnosed with CDH as a fetus and died before birth. A child with CDH and a *GPC3* nonsense variant identified in this study had clinical features of Simpson-Golabi-Behmel syndrome type 1 (MIM: 312870).

GWAS for CDH susceptibility loci

We performed common variant analysis by using a discovery cohort of 1,443 affected individuals and 6,133 unaffected parents from SPARK. We performed joint calling of variants initially subdivided into batch 1 (primarily from our previous report[11]) and batch 2 (new samples) and then combined these (batch 1 + batch 2). To control for

Table 2. Burden of *de novo* coding variants

	All ($n = 1{,}064$)				Female ($n = 433$)				Male ($n = 631$)				Complex ($n = 379$)				Isolated ($n = 665$)			
Variant class	**#Obs.**	**#Exp.**	**Fold**	***p* value**	**#Obs.**	**#Exp.**	**Fold**	***p* value**	**#Obs.**	**#Exp.**	**Fold**	***p* value**	**#Obs.**	**#Exp.**	**Fold**	***p* value**	**#Obs.**	**#Exp.**	**Fold**	***p* value**
All genes																				
Synonymous	316	346	0.9	0.1	129	142	0.9	0.3	187	204	0.9	0.3	115	123	0.9	0.5	194	216	0.9	0.1
Missense	906	771	1.2	2.2×10^{-6}	364	317	1.2	0.01	539	454	1.2	9.7×10^{-5}	322	275	1.2	5.0×10^{-3}	573	482	1.2	5.5×10^{-5}
D-mis	692	497	1.4	1.3×10^{-16}	279	204	1.4	6.4×10^{-7}	411	293	1.4	6.5×10^{-11}	245	177	1.4	1.2×10^{-6}	439	310	1.4	5.9×10^{-12}
LGD	181	108	1.7	1.2×10^{-10}	83	44	1.9	2.5×10^{-7}	98	64	1.5	5.5×10^{-5}	83	38	2.2	4.3×10^{-10}	96	67	1.4	9.8×10^{-4}
Damaging	873	605	1.4	1.4×10^{-24}	362	249	1.5	1.5×10^{-11}	509	356	1.4	2.5×10^{-14}	328	215	1.5	9.0×10^{-13}	535	378	1.4	2.8×10^{-14}
Constrained genes (pLI $\geq$ 0.5)																				
Synonymous	102	108	0.9	0.6	42	45	0.9	0.8	60	63	1.0	0.8	39	39	1	0.9	61	68	0.9	0.5
Missense	292	238	1.2	7.5×10^{-4}	122	99	1.2	0.02	168	139	1.2	0.02	93	85	1.1	0.4	193	149	1.3	5.7E-04
D-mis	264	172	1.5	7.6×10^{-11}	112	71	1.6	8.0×10^{-6}	150	101	1.5	4.2×10^{-6}	83	61	1.4	7.2×10^{-3}	175	107	1.6	2.4×10^{-9}
LGD	87	33	2.6	4.6×10^{-15}	44	14	3.2	6.6×10^{-11}	43	19	2.2	3.1×10^{-6}	45	12	3.8	1.3×10^{-13}	41	21	2	6.0×10^{-7}
Damaging	351	205	1.7	1.7×10^{-20}	156	85	1.8	5.2×10^{-12}	193	120	1.6	7.9×10^{-10}	128	73	1.8	5.3×10^{-9}	216	128	1.7	1.6×10^{-12}
Other genes																				
Synonymous	214	238	0.9	0.1	87	97	0.9	0.3	127	140	0.9	0.3	76	85	0.9	0.4	133	149	0.9	0.2
Missense	614	533	1.2	5.7×10^{-4}	242	218	1.1	0.1	371	314	1.2	1.7×10^{-3}	229	190	1.2	0.01	380	333	1.1	0.01
D-mis	428	325	1.3	4.4×10^{-8}	167	133	1.3	4.2×10^{-3}	261	192	1.4	2.2×10^{-6}	162	116	1.4	4.2×10^{-5}	264	203	1.3	3.9×10^{-5}
LGD	94	75	1.3	0.03	39	31	1.3	0.2	55	44	1.2	0.1	38	27	1.4	0.03	55	47	1.2	0.2
Damaging	522	400	1.3	4.7×10^{-9}	206	164	1.3	1.3×10^{-3}	316	236	1.3	7.3×10^{-7}	200	142	1.4	4.7×10^{-6}	319	250	1.3	2.5×10^{-5}

D-mis: missense variant with CADD $\geq$ 20. Damaging: damaging variants, including D-mis and LGD. #Obs: number of observed variants. #Exp: number of expected variants. RR: relative risk. Constrained genes: genes with pLI>0.5.

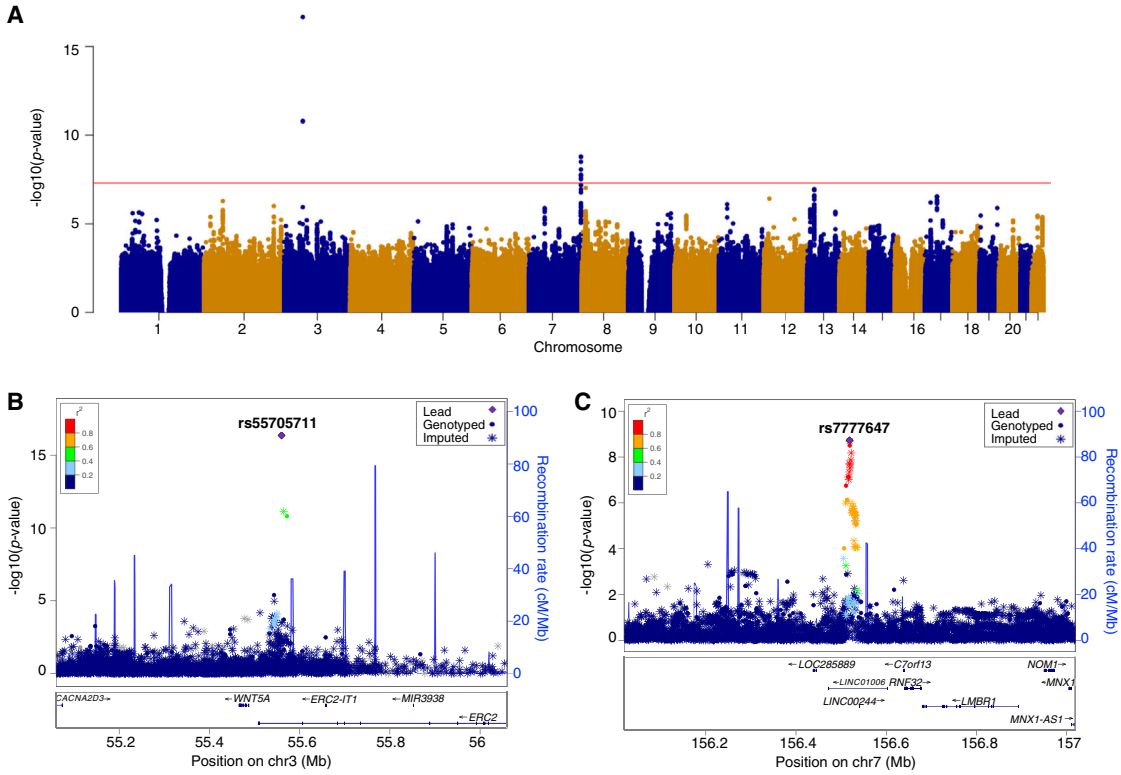
Please cite this article in press as: Qiao et al., Common variants increase risk for congenital diaphragmatic hernia within the context of *de novo* variants, The American Journal of Human Genetics (2024), https://doi.org/10.1016/j.ajhg.2024.08.024

Table 3. Known and candidate CDH genes with pLI-stratified extTADA

Gene	Gene Name	#Dmis	#LGD	PPA	FDR	pLI	OMIM
MYRF[a]	myelin regulatory Factor	4	4	1.0	1.8×10^{-7}	1.00	cardiac urogenital syndrome (MIM: 618280); encephalitis/ encephalopathy (MIM: 618113); AD
GATA6[a]	GATA binding protein 6	1	3	1.0	3.4×10^{-4}	1.00	atrial septal defect (MIM: 614475); atrioventricular septal defect (MIM: 614474); pancreatic agenesis and congenital heart disease (MIM: 600001); persistent truncus arteriosus (MIM: 217095); tetralogy of fallot (MIM: 187500); AD
LONP1[a]	Lon peptidase 1, mitochondrial	4	0	1.0	6.2×10^{-3}	1.00	CODAS syndrome (MIM: 600373); AD
ALYREF[a]	Aly/REF export factor	0	2	1.0	0.017	0.93	46XY sex reversal (MIM: 616067); diaphragmatic hernia (MIM: 610187); tetralogy of fallot (MIM: 187500); AD
ZFPM2[a]	Zinc finger protein, FOG family member 2	0	2	0.9	0.025	1.00	atrial septal defect (MIM: 614475); atrioventricular septal defect (MIM: 614474); tetralogy of fallot (MIM: 187500); ventricular septal defect (MIM: 614429); AD
SIN3A	SIN3 transcription regulator family member A	0	2	0.9	0.032	1.00	Witteveen-Kolk syndrome (MIM: 613406); included in Chr 15q24 microdeletion syndrome.
STAG2	stromal antigen 2	0	2	0.9	0.037	1.00	Holoprosencephaly 13 X-linked (MIM: 301043); Mellegama-Klein-Martinez syndrome (MIM: 301022); XL
POGZ	Pogo transposable element derived with ZNF domain	0	2	0.9	0.041	1.00	White-Sutton syndrome (MIM: 616364); AD
ZNF462	zinc finger protein 462	0	2	0.9	0.047	1.00	Weiss-Kruszka syndrome (MIM: 618619); AD
KALRN	Kalirin RhoGEF kinase	2	1	0.9	0.053	1.00	none
HSD17B10	hydroxysteroid 17-beta dehydrogenase 10	1	1	0.8	0.063	0.90	HSD10 mitochondrial disease (MIM: 300438); AD
GPC3[a]	glypican 3	1	1	0.8	0.072	1.00	Simpson-Golabi-Behmel syndrome type 1 (MIM: 312870); XLR
SRGAP2	SLIT-ROBO Rho GTPase-activating protein 2	1	1	0.8	0.081	0.98	none
SYMPK	symplekin	1	1	0.8	0.091	1.00	none
GATA4[a]	GATA binding protein 4	2	1	0.8	0.10	0.49	atrial septal defect (MIM: 614475); atrioventricular septal defect (MIM: 614474); tetralogy of fallot (MIM: 187500); ventricular septal defect (MIM: 614429); AD

#D-mis, number of *de novo* D-mis; #LGD, number of *de novo* LGD; PPA, posterior probability of association; FDR, false-discovery rate; OMIM, Online Mendelian Inheritance in Man. AD, autosomal dominant; XL, X linked; XLR, X-linked recessive.
All genes with extTADA FDR <0.1.
[a]Denotes genes that have been reported previously.

population stratification (Table S5) and relatedness (Figure S1B), we confined the initial analysis to the largest ancestry group (1,005 European affected individuals; 4,565 European controls). The quantille-quantile plot of *p* values for genotyped and imputed SNPs, after adjustment for sex and the top four principle components, shows negligible genomic inflation (Figure S5). Two loci reached genome-wide significance: chromosomal region 3p14.3/lead SNP rs55705711 ($p = 5.1 \times 10^{-17}$, OR = 1.65) and chromosomal region 7q36.3/lead SNP rs7777647 ($p = 1.9 \times 10^{-9}$, OR = 1.27) (Figure 1; Table S6). The loci were supported by multiple significant SNPs at each locus. Manhattan plots are provided in Figures 1A–1C. rs55705711 resides in an intron of *ERC2* (ELK2/RAB6-interacting/ CAST family member 2) and adjacent to *WNT5A* (Wnt family member 5A), a developmental patterning gene.[78] rs7777647 resides in a region containing five protein-coding genes and a long-range *cis*-acting regulatory domain upstream of *SHH* (sonic hedgehog). We note that none of the rare copy-number variants identified in our previous CDH study[10,11] overlapped these novel common-variant loci. To confirm the findings, we tested the associations in a European replication cohort of 389 affected individuals and 4,815 control individuals. Replication was



D rs55705711

Phase	*n* cases	*n* controls	AF cases	AF controls	OR (95% CI)	*p* value
Discovery	1,005	4,565	0.25	0.18	1.65 (1.53-1.76)	5.12e-17
Replication	389	4,815	0.20	0.18	1.18 (0.97-1.40)	5.60e-02
Meta-analysis	**1,394**	**9,380**			**1.54 (1.44-1.65)**	**1.43e-16**

Odds ratio

E rs3907611

Phase	*n* cases	*n* controls	AF cases	AF controls	OR (95% CI)	*p* value
Discovery	1,005	4,565	0.13	0.09	1.69 (1.54-1.85)	1.16e-11
Replication	389	4,815	0.10	0.08	1.38 (1.09-1.66)	0.031
Meta-analysis	**1,394**	**9,380**			**1.63 (1.49-1.77)**	**2.25e-12**

Odds ratio

F rs7777647

Phase	*n* cases	*n* controls	AF cases	AF controls	OR (95% CI)	*p* value
Discovery	1,005	4,565	0.69	0.62	1.27 (1.19-1.35)	1.87e-09
Replication	389	4,815	0.68	0.61	1.22 (1.05-1.39)	3.88e-03
Meta-analysis	**1,394**	**9,380**			**1.26 (1.17-1.35)**	**3.38e-11**

Odds ratio

Figure 1. Genome-wide association results in CDH

(A) Manhattan plot of SNP-based association results in the discovery cohort. The x axis shows genomic coordinates, and the y axis shows the $-\log_{10}(p$ value) obtained by logistic regression adjusting for sex and the first four principal components.

(legend continued on next page)

observed for both loci and with the same direction of effects (Table S6). Meta-analysis in the combined cohort replicated all SNPs at both loci and with increased significance (Table S6). Conditional analyses for both loci revealed an independent significant signal at 7q36.3 (Figure S6) but no additional signal at 3p14.3 (rs75263535, $p = 2.2 \times 10^{-5}$ conditional analysis; $p = 2 \times 10^{-3}$ discovery phase). Forest plots depicting the GWAS results in the discovery, replication, and combined cohorts are provided in Figures 1D–1F.

Power analyses indicated that our study was well powered for new gene discovery in the European ancestry group and for replication in the European and Latinx groups but that it was less well-powered for other ancestry groups (Figure S7). Testing for association by ancestry group for lead SNPs at 3p14.3 and 7q36.3 demonstrated significant association at the threshold for replication for the Latinx group (Table S7), and there were similar effect sizes in both Europeans and Latinx (OR $\approx$ 1.7 for 3p14.3 and $\approx$ 1.3 for 7q36.3) (Figures 1D–1F). No replication was observed for other ancestry groups, at least in part because of small sample sizes. Pairwise linkage disequilibrium heatmap plots of GWAS signals by genetic ancestry are provided for 3p14.3 (Figure S8) and 7q36.3 (Figure S9).

We performed additional cohort subgroup comparisons of allele frequencies at rs55705711 and rs7777647. For individuals with CDH, with and without likely pathogenic rare variants in genes with known CDH causal variants or constrained genes, we observed no differences in allele frequencies at either locus (Table S8). Similarly, for isolated vs. complex CDH, alive vs. deceased discharge status, requirement for ECMO treatment during hositalization, or diagnosis of pulmonary hypertension at one or three months of age, we observed no differences (Table S8).

Most disease-associated common variants map to noncoding regions of the genome, and many have chromatin signatures of *cis*-acting regulatory elements, suggesting a role for the variants in transcriptional regulation.[57,79,80] The regulatory elements may regulate adjacent or nonadjacent target genes, and chromosome-conformation capture methods have been developed to map regional chromatin-interaction sites. We performed functional annotation of the CDH GWAS loci by aligning epigenomic signatures, transcriptional regulatory domains, and protein three-dimensional structural data with the locations of GWAS signals at both loci. Multiple lines of evidence support a role in *WNT5A* enhancer function for the 3p14.3 locus (Figure 2A). The locus resides in the same TAD as *WNT5A* and *ERC2*.[55] The locus contains binding sites for multiple transcriptional regulators important in development and/or implicated in CDH; these include *CTCF*, *SIN3A*, and *GATA6*.[58] The top SNPs are located in an open chromatin region, as indicated by snATAC-seq data from human lung[61] and mouse skeletal muscle[62] (Figure S10), and are predicted on the basis of chromosome conformation capture (Hi-C) data to interact with the the *WNT5A* promoter.[57] The 7q36.3 locus resides in a TAD including five protein-coding genes and a long-range *cis*-acting regulatory domain upstream of *SHH* (Figure 2B). Similar to the 3p14.3 locus, the locus contains binding sites for *CTCF* and *GATA6*. The 7q36.3 locus is predicted on the basis of capture Hi-C data to interact with the the *RNF32* promoter.[57] The long-range *SHH* regulatory domain is conserved between humans and mice, and studies in mice indicate that the *Shh* enhancer extends 900 kb upstream and is composed of tissue-specific short enhancer regions distributed throughout the full enhancer[81] including at least one specific for lung epithelial linings.[82]

SNP heritability and polygenic risk scores

Genome-wide SNP-based heritability estimates were calculated under the assumption of a CDH prevalence of 1/3,000 and over a range of prevalence from 1/1,000–1/5,000. The estimated heritability from common variants for CDH was about 19% across the range of prevalence (Table S9). The heritability from the two genome-wide-significant loci was 1.3% (3p14.3/rs55705711) and 0.7% (7q36.3/rs7777647) (Table S10).

The polygenic risk score was calculated from a total of 1,326,338 common SNPs with independent linkage disequilibrium blocks. Elevated polygenic risk was observed for participants with CDH compared to controls across all subsets of CDH-affected individuals—isolated, complex, male, female, with damaging variants, and without damaging variants (Figure 3A). No significant differences in PRS were observed for isolated vs. complex CDH, male vs. female CDH, or CDH with damaging rare variants vs. CDH without damaging rare variants (Figure 3A). There was no significant difference in complex CDH with *de novo* variants in NDD genes vs. complex CDH without *de novo* variants in NDD genes (Table S11). Comparison of proband PRSs to midparent PRSs showed overtransmission of polygenic risk for CDH to offspring (Figure 3B). Finally, individuals with a PRS in the top decile had an odds ratio of 2 for CDH, whereas individuals with a PRS in the bottom decile had an odds ratio of ~1.1 (Figure 3C; Table S12).

(B and C) Regional association plots of the 3p14.3 (B) and 7q36.3 (C) loci with surrounding genes. The y axis shows the $-\log_{10}(p \text{ value})$. Round points indicate genotyped SNPs, and star points represent imputed SNPs, colored by degree of linkage disequilibrium with the top association SNPs (purple points): rs55705711at 3p14.3 (B) and rs7777647 at 7q36.3 (C).
(D–F) Forest plots demonstrating replication of the lead SNP at 3p14.3 rs55705711 (D), a second SNP at 3p14.3 rs3907611 (E), and the lead SNP at 7q36.3 rs7777647 (F) in an independent European cohort and the combined cohort. The x axis shows odds ratios (ORs) and 95% confidence intervals (CIs). *n*, sample size; AF, allele frequency of the risk allele.

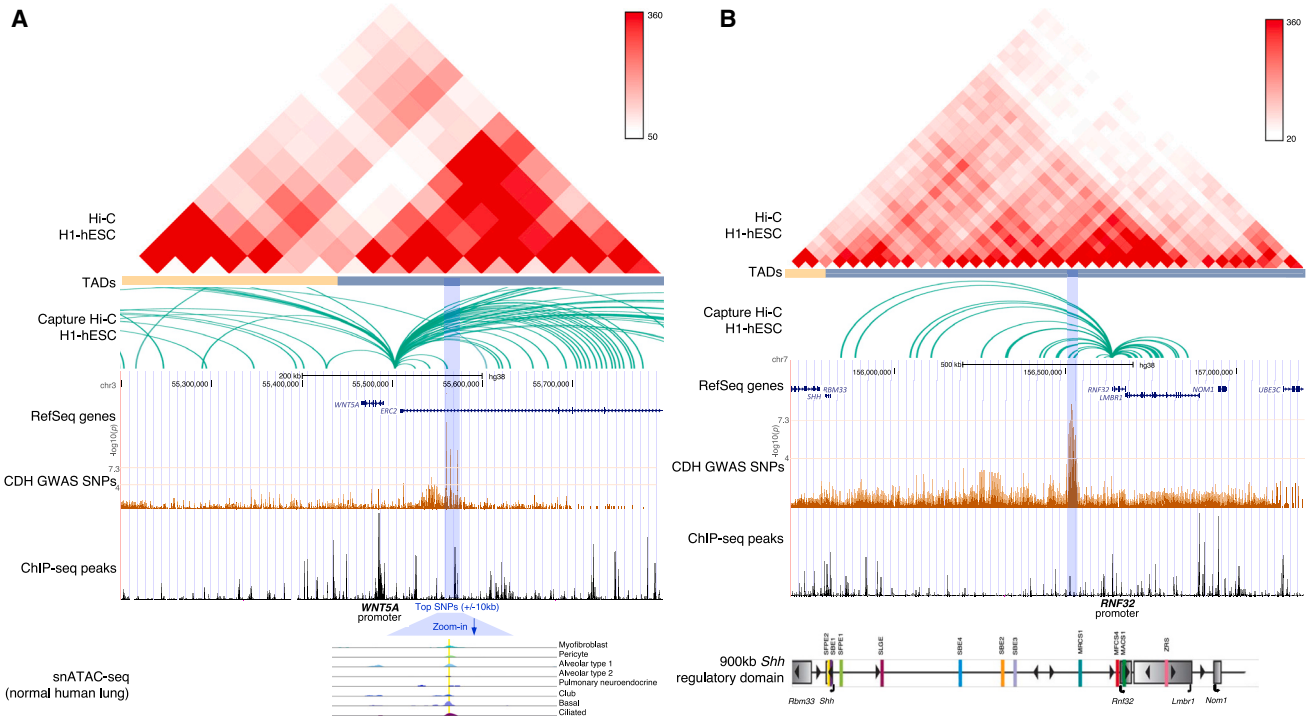


Figure 2. Functional annotation of genome-wide-significant CDH loci

(A and B) Alignment of functional data at 3p14.3 (A) and 7q36.3 (B). From top to bottom: Hi-C data derived from H1 human embryonic stem cell line (H1-hESC)[55]; TADs derived from H1-hESC; capture Hi-C data derived from H1-hESC[57]; RefSeq genes; CDH GWAS SNP data from the combined cohort shown on a $-\log_{10}(p\ \text{value})$ scale; transcription regulators from ReMap ChIP-seq in heart, brain, lung, and stem-cell-related-tissue track[58]; single-nucleus ATAC-sequencing (snATAC-seq) data from normal human lung[61] (A) and the 900 kb murine *Shh* distal enhancer containing multiple small enhancer regions with tissue specificity[81] (B).

Discussion

We report genetic findings from an expanded CDH trio cohort including 1,469 DHREAMS and Boston CDH probands. *De novo* LGD and D-mis variants were enriched in individuals with CDH, and there was greater enrichment for complex vs. isolated CDH and in females vs. males. We identified fifteen candidate CDH genes that were highly constrained for loss-of-function variants, including variants in four novel genes (*SIN3A*, *STAG2*, *POGZ*, and *ZNF462*) with FDR < 0.05. Together, variants in these genes are estimated to confer relative risks for CDH ranging from 10 to 30 but still explain only 3.9% of participants with CDH. To expand the understanding of the range of genetic contributors, we performed common-variant association analysis and identified two genome-wide-significant loci, with replication, that st molikely contribute to CDH risk: these two loci are 3p14.3 (lead SNP rs55705711) and 7q36.3 (rs7777647). The effect size for each locus was similar among Europeans and Latinx, but the study lacked power to draw conclusions for African American and Asian populations (Figure S7). Both susceptibility loci are located in putative transcriptional regulatory regions for established developmental patterning genes. The estimated attributable risk from common variants for CDH was 19%, and PRS estimates demonstrated increased PRSs for both isolated and complex CDH, and risk overtransmission to offspring with CDH. Together, the data support a polygenic model as part of the CDH genetic architecture for all individuals with CDH.

The burden of rare predicted deleterious *de novo* variants in our current CDH cohort, including the enrichment of variants in complex vs. isolated CDH and in females vs males, is consistent with our previous reports.[3,15,20,83] Similar patterns have been observed in autism for individuals with vs. without intellectual disability[84] and in females vs. males.[85] Hirschprung disease is another example of a congenital anomaly with complex genetics.[86,87] The enrichment of *de novo* variants in females vs. males is consistent with a female-protective model in which females have a higher liability threshold than males.

Our analysis confirms the associations of *MYRF*,[20,74] *GATA6*,[75,88] *GATA4*,[89,90] *LONP1*,[3] *ZFPM2* (zinc finger protein, FOG family member 2),[91–93] and *GPC3*[70–73] with CDH. We also confirm the association of *ALYREF* with CDH from our earlier report based on an overlapping but smaller cohort (DHREAMS and Boston).[3] *ALYREF* is an RNA-binding protein that regulates 5′-methylcytosine modification, which if left unregulated results in abnormal cell proliferation and migration.[94] There is an expanding body of evidence implicating RNA-binding proteins with congenital and developmental diseases.[95–97]

Please cite this article in press as: Qiao et al., Common variants increase risk for congenital diaphragmatic hernia within the context of *de novo* variants, The American Journal of Human Genetics (2024), https://doi.org/10.1016/j.ajhg.2024.08.024

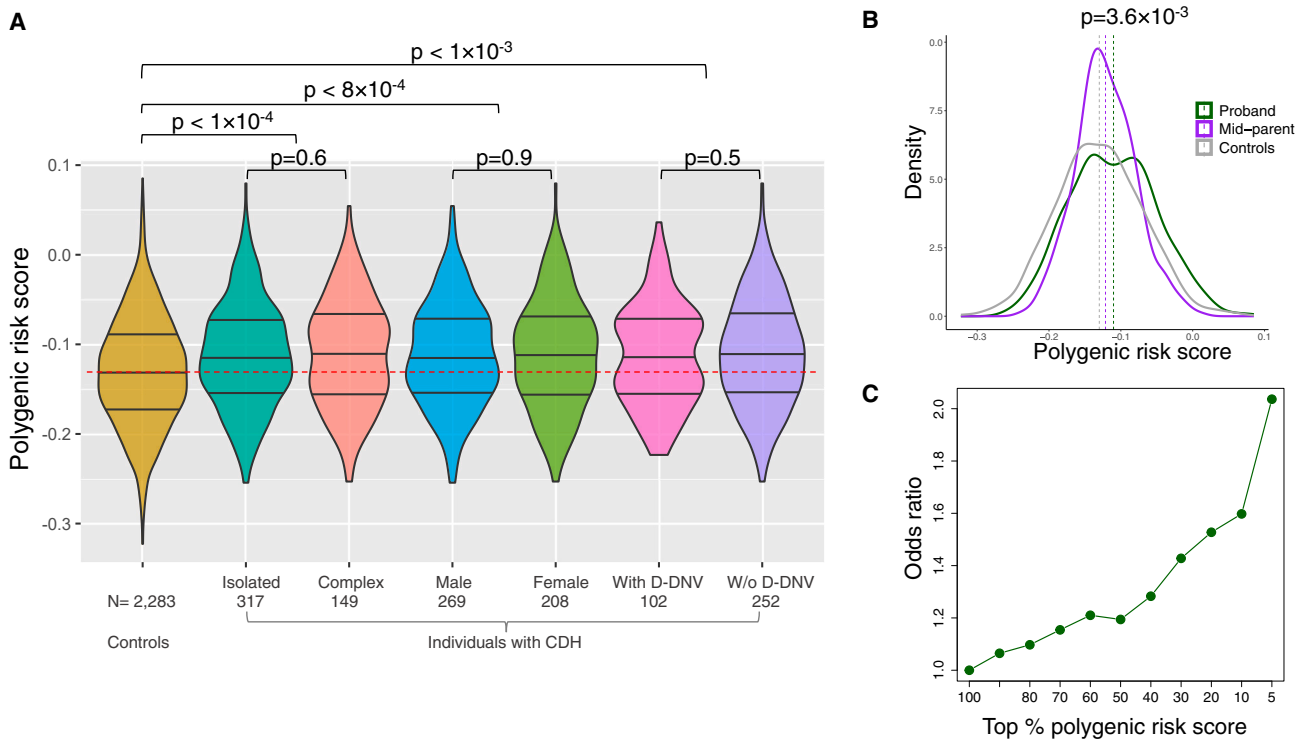


Figure 3. Polygenic risk scores in different subsets of the CDH testing cohort

(A) PRSs in participants with CDH-affected individuals vs. control individuals, individuals with isolated vs. complex CDH, male vs. female individuals, or individuals with vs. without *de novo* damaging variants (D-DNV). *p* values were determined with a Student's t test.

(B) Polygenic risk for CDH in CDH-affected children (probands) compared to mid-parents (i.e., the average of mother and father risk scores) or control individuals. *p* values were determined with a Student's t test.

(C) Odds ratio of case-control proportions for PRSs at multiple thresholds. Exact *p* values by Fisher's exact test are listed in Table S12.

De novo analysis of the expanded DHREAMS/Boston cohort allowed the identification of variants in novel genes with diaphragm, heart or brain developmental expression patterns and known roles in chromatin remodeling, transcriptional regulation, mitosis, and protein phosphorylation. We defined novelty as previous reports of no more than one individual with a rare variant associated with a developmental syndrome and CDH (see references for reviews[16,17,18]). Most of the genes have known roles in embryonic developmental processes, including tissue specification and patterning. As such, variants in these genes are often associated with rare multi-organ syndromes.[20,66,74,98–100] Novel CDH candidate *SIN3A* encodes a scaffold protein of the SIN3/histone deacetylase transcriptional repressor complex that regulates cell-cycle progression. *SIN3A* is widely expressed in embryonic development, and germline deletion causes early embryonic lethality in mice.[101,102] We recently reported that *Sin3a* conditional deletion in mouse skeletal muscle results in a thin, membranous diaghragm; in the mesothelium; results in left posterior-lateral CDH; and in the lung mesenchyme, causes lung hypoplasia, failure of alveologenesis, and pulmonary hypertension.[103] Embryonic treatment of *Sin3a* mutant mice with an inhibitor of histone acetyltransferase improved lung and pulmonary vasculature development and reduced pulmonary hypertension.[103] *SIN3A* expression is dysregulated in human pulmonary arterial hypertension; overexpression increased *BMPR2* (bone morphogenetic protein receptor 2) expression, inhibited human pulmonary arterial smooth muscle cell proliferation, and ameliorated experimental pulmonary hypertension.[104] *SIN3A* LGD variants are also rare causes of mild cognitive deficits and facial dysmorphism.[100] Genetic diagnoses in CDH can inform clinical surveillance of comorbidities, disease management, and prognosis.

To test the role of common variants in CDH, we performed a case-control GWAS and identified two susceptibility loci, with replication, at 3p14.3 (lead SNP rs55705711) and 7q36.3 (rs7777647). rs55705711 resides in a TAD including two protein-coding genes, *WNT5A* and *ERC2*. *WNT5A* encodes a widely expressed and essential regulator of embryonic developmental pathways. *ERC2* encodes a regulator of neurotransmitter release and is highly expressed in the brain and to a lesser extent in the gut and lung. *WNT5A* is unique to other *WNT* signaling molecules in that it predominantly signals through beta-catenin-independent pathways, including the planar cell polarity pathway. *Wnt5a*-null embryos exhibit decreased extension of the anterior-posterior axis, resulting in severely decreased outgrowth of diverse embryonic structures and perinatal lethality.[105] *Wnt5a* is expressed in the embryonic gut mesoderm, and germline

deficiency results in incomplete closure of the primitive gut tube at E10 and impaired elongation of the small intestine.[106] Although diaphragm-specific studies have not been reported, the planar cell polarity pathway is involved in multiple developmental tissue closure processes.[107] In the developing lung, *Wnt5a* regulation of epithelial-mesenchymal interactions leads to alterations in airway branching and mesenchymal growth.[108] CDH-associated rs55705711, and several neighboring SNPs with significant CDH association signals, are located in an open chromatin region[61] predicted to interact with the *WNT5A* promoter.[57] Whether these 3p14.3 SNPs define, or are closely linked to, a *WNT5A* transcriptional regulatory element warrants further investigation. We note that a nearby but unique region of 3p14.3 was recently identified to harbor a common variant increasing risk for dextro-transposition of the great arteries, a severe congenital heart defect.[109] *In vitro* and *in vivo* data suggested *cis* regulation of *WNT5A* expression as a potential pathogenetic mechanism.[109] The associated region also contains putative binding sites for other transcriptional regulators important in development (*CTCF*,[110] *SIN3A*; [111] *GATA6*[112]) and implicated in CDH (*SIN3A* [this report], *GATA6*[4]). However, lung-mesenchyme-specific *Sin3a* deletion mice do not exhibit differential *Wnt5a* mRNA levels in lung or recombined mesenchyme.[103]

The CDH 7q36.3 locus resides in a TAD including a long-range *SHH* enhancer and five protein-coding genes. The *SHH* enhancer comprises ~900 kb across a gene desert and extending into two genes in the neighboring gene-rich region, *RNF32* (ring finger protein 32) and *LMBR1* (limb development membrane protein 1).[81] *SHH* encodes a ubiquitously expressed ligand of the hedgehog developmental signaling receptor that is essential for lung branching morphogenesis.[113] Decreased *SHH* expression in CDH lungs is associated with small lung size and fewer airway branches than in normal lungs.[113] Null mutations in mice are embryonic lethal and result in severe morphological defects in multiple organs.[114] Regulation of spatial and temporal *Shh* expression is coordinated by tissue-specific elements within the long-range enhancer,[81] and mutations in these regions cause congenital anomalies (reviewed in Anderson et al.[81]). Rare variants in an *LMBR1* intronic *SHH* enhancer element are associated with congenital limb defects.[115] At least one lung-specific enhancer, MACS1, located in *RNF32* intron 8, has been reported,[82] but the effects of genetic variation in this element are unknown. Other genes in the TAD include *NOM1* (nucleolar protein with MIF4G domain 1), *MNX1* (motor neuron and pancreas homeobox1), and *UBE3C* (ubiquitin protein ligase E3C). There is a single report of *Nom1* involvement in pancreas development in zebrafish.[116] *MNX1* is primarily expressed in the pancreas, and mutations are a cause of Currarino syndrome (MIM: 176450), a rare congenital malformation syndrome involving anorectal, sacral, and presacral anomalies. *Mnx1*-null mice exhibit immature pancreas development and function.[117] *UBE3C* is ubiquitously expressed, and mutations cause a rare speech, motor, and behavioral neuropathy (MIM: 620270). Functional studies of the CDH 7q36.3 variants will be necessary for identification of the pathogenic mechanism, but transcriptional regulation of *SHH* expression, via the long-range enhancer including the *RNF32* intronic element, is a viable candidate mechanism.

We note that several of the genes with causative and candidate CDH variants may function along the same pathway or regulate the same biological processes. ZFPM2 heterodimerizes with GATA6/GATA4 to regulate transcription of downstream target gene,s[118] and GPC3 is a regulator of *SHH* and *WNT5A*[119–121] candidate genes at the 3p14.3 and 7q36.3 GWAS loci.

Genetic contributors understood to cause developmental disorders are largely rare variants in single genes. However, monogenic diseases can exhibit highly variable expressivity. It has become increasingly clear that common variants contribute to disease penetrance and severity and that individuals with higher PRS exhibit increased severity of disease.[122] In our study, SNP-based heritability analysis indicated that 19% of variance in susceptibility to CDH can be explained by common variants. Similarly, it has been estimated that common variants might explain ~25% of heritability for a rare form of congenital heart disease.[109] We detected elevated PRSs in both isolated- and complex-CDH-affected individuals compared to unaffected control individuals. We found no difference in the contribution of common variants between individuals with and without rare variants in genes with known CDH causal variants. The combination of clinically significant monogenic variants and PRS can increase risk prediction.[122] The effect size of common variants can be relatively large for rare conditions compared to common conditions. Although common variants for common diseases typically confer minimal disease risk individually, the risk allele at CDH-associated rs55705711 confers an odds ratio of 1.65, which is substantial.

The finding of no significant PRS differences between individuals with isolated and complex CDH, as well as between individuals with or without damaging *de novo* variants, suggests that *de novo* and common variants contribute additive genetic risk for CDH. This is in contrast to that reported for autism.[123] Autism is common in the population, such that either elevated polygenic risk from common variants or a rare *de novo* pathogenic variant is sometimes sufficient to push an individual over the liability threshold. CDH is much rarer, with a much higher liability threshold. Reaching the higher liability threshold often requires the combination of both elevated polygenic risk from common variants and rare or *de novo* pathogenic variants.

The current genetic diagnostic yield of genome sequencing is ~30% of individuals with CDH. As such, genomic testing (exome or genome sequencing) is recommended for all individuals with CDH and with parent-child

trios when possible. Return of results to families should include rare deleterious variants in well-established genes and genes with rare *de novo* variants in ≥ 2 probands. The phenotypes for many of these recently described genetic conditions is still evolving, and CDH can be an uncommon but consistent feature of many developmental disorders. Genetic diagnoses clarify prognosis and support tailored surveillance of associated medical and neurodevelopmental features to identify and treat associated conditions early. Genetic data also assists families in clarifying the risk of recurrence.

In conclusion, we identified candidate risk variants in at least four new high-priority genes for CDH and two susceptibility loci associated with common variants. We estimate that *de novo* damaging variants most likely explain 25% of the population attributable risk, with a 2-fold enrichment of LGD variants among complex vs. isolated CDH. The estimated heritability from common variants is 19%. Identification of common variant risk loci and elevated polygenic risk scores in CDH support a polygenic model as part of the CDH genomic architecture at least in European and Latinx populations.

Data and code availability

Data for identified damaging variants are provided in Table S4. The whole-genome sequencing and exome sequencing CDH data used in this study are available at the Database of Genotypes and Phenotypes (dbGaP: phs001110.v3.p1). Access to the summary stats of the Dutch/German cohort is available on request and stored in the Digital Research Environment of the Netherlands (https://support.mydre.org/portal/en/home), an Azure-based cloud system optimized for collaborations.

Acknowledgments

We thank the participants and their families for their generous contribution. We are grateful for technical assistance provided by Na Zhu, Patricia Lanzano, Amanda McPartland, Jiangyuan Hu, Charles LeDuc, and Liyong Deng (Columbia University); Jennifer Lyu (Boston Children's Hospital); Caroline Coletti (Massachusetts General Hospital); and Nadine Fricker (Life & Brain GmbH). We thank our clinical coordinators across the DHREAMS centers: Jessica Conway (Washington University School of Medicine); Melissa Reed, Elizabeth Erickson, and Madeline Peters (Cincinnati Children's Hospital); Sheila Horak and Evan Roberts (Children's Hospital & Medical Center of Omaha); Jeannie Kreutzman and Irene St. Charles (CS Mott Children's Hospital); Tracy Perry (Monroe Carell Jr. Children's Hospital); Michelle Kallis (Northwell Health); Andrew Mason and Alicia McIntire (Oregon Health and Science University); Gentry Wools and Lorrie Burkhalter (Children's Medical Center Dallas); Elizabeth Jehle (Hassenfeld Children's Hospital); Michelle Knezevich and Cheryl Kornberg (Medical College of Wisconsin); and Min Shi (Children's Hospital of Pittsburgh). 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 (J.W., W.K.C.), R03HL138352 (W.K.C., Y.S.), R01GM120609 (Y.S.), R03HL161595 (Y.S.), R35GM149527 (Y.S.), UL1 RR024156 (W.K.C.), P01HD0 68250 (P.K.D., F.A.H., J.M.W., W.K.C., Y.S., J.M.Z., D.J.M., X.S.), and R01HL146859 (D.J.M.). Additional grant funding was from CHERUBS/Congenital Diaphragmatic Hernia (CDH) International (https://cdhi.org), CDH UK (https://cdhuk.org.uk), and the National Greek Orthodox Ladies Philoptochos Society and generous donations from the Williams family, Wheeler Foundation, Vanech Family Foundation, Larsen family, Wilke family, and many other families. C.B. is supported by BONFOR stipend O-112.0062 from the Medical Faculty of the University of Bonn.

Author contributions

W.K.C. and Y.S. had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Concept and design: W.K.C. and Y.S. conceived of the study.

Acquisition, analysis, or interpretation of data: L.Q., C.L.W., R.H., J.W., U.S.K., J.M.Z., T.B., J.K., A.D., C.F.K., A.J.W., A.H., A.C.M.., A.d.K., B.W.W., D.C., D.J.M., D.S., D.P., E.F., F.S., F.K., F.-Y.L., G.B.M., G.A., H.M.R., H.N., J.M.S., J.C.F., K.A., M.E., M.E.D., M.L., P.K., R.M.H.W., R.A.C., S.Z.S., S.C.M.C.-D.O., T.S., T.C., V.P.D., P.K.D., X.S., F.A.H., C.B., E.B., Y.S., and W.K.C..

Drafting of the manuscript: L.Q., C.L.W., C.B., E.B., Y.S., and W.K.C..

Critical revision of the manuscript for important intellectual content: L.Q., C.L.W., R.H., J.W., U.S.K., J.M.Z., T.B., J.K., A.D., C.F.K., A.J.W., A.H., A.C.M., A.d.K., B.W.W., C.M., D.C., D.J.M., D.S., D.P., E.F., F.S., F.K., F.-Y.L., F.M., G.B.M., G.A., H.M.R., H.N., J.M.S., J.C.F., K.A., M.E., M.M.N., M.E.D., M.L., P.K., R.M.H.W., R.A.C., S.Z.S., S.C.M.C.-D.O., T.S., T.C., V.P.D., P.K.D., X.S., F.A.H., C.B., E.B., Y.S., W.K.C..

Statistical analysis: L.Q., C.L.W., C.M., F.M., H.M.R., M.M.N., C.B., E.B., Y.S., W.K.C..

Supervision: P.K.D., C.B., E.B., Y.S., W.K.C..

Declaration of interests

W.K.C. is on the Board of Directors of Prime Medicine. The authors declare no additional competing interests.

Supplemental information

Supplemental information can be found online at https://doi.org/10.1016/j.ajhg.2024.08.024.

Received: April 23, 2024
Accepted: August 30, 2024
Published: September 26, 2024

References

1. Sattolo, M.L., Arbour, L., Bilodeau-Bertrand, M., Lee, G.E., Nelson, C., and Auger, N. (2022). Association of Birth Defects With Child Mortality Before Age 14 Years. JAMA Netw. Open *5*, e226739.
2. Holden, K.I., and Harting, M.T. (2023). Recent advances in the treatment of complex congenital diaphragmatic hernia-a narrative review. Transl. Pediatr. *12*, 1403–1415.
3. Qiao, L., Xu, L., Yu, L., Wynn, J., Hernan, R., Zhou, X., Farkouh-Karoleski, C., Krishnan, U.S., Khlevner, J., De, A., et al. (2021). Rare and de novo variants in 827 congenital

diaphragmatic hernia probands implicate LONP1 as candidate risk gene. Am. J. Hum. Genet. *108*, 1964–1980.

4. Zani, A., Chung, W.K., Deprest, J., Harting, M.T., Jancelewicz, T., Kunisaki, S.M., Patel, N., Antounians, L., Puligandla, P.S., and Keijzer, R. (2022). Congenital diaphragmatic hernia. Nat. Rev. Dis. Primers *8*, 37.
5. Politis, M.D., Bermejo-Sanchez, E., Canfield, M.A., Contiero, P., Cragan, J.D., Dastgiri, S., de Walle, H.E.K., Feldkamp, M.L., Nance, A., Groisman, B., et al. (2021). Prevalence and mortality in children with congenital diaphragmatic hernia: a multicountry study. Ann. Epidemiol. *56*, 61–69.e63.
6. Geggel, R.L., Murphy, J.D., Langleben, D., Crone, R.K., Vacanti, J.P., and Reid, L.M. (1985). Congenital diaphragmatic hernia: arterial structural changes and persistent pulmonary hypertension after surgical repair. J. Pediatr. *107*, 457–464.
7. Byrne, F.A., Keller, R.L., Meadows, J., Miniati, D., Brook, M.M., Silverman, N.H., and Moon-Grady, A.J. (2015). Severe left diaphragmatic hernia limits size of fetal left heart more than does right diaphragmatic hernia. Ultrasound Obstet. Gynecol. *46*, 688–694.
8. Song, R., Hu, X.Q., and Zhang, L. (2019). Glucocorticoids and programming of the microenvironment in heart. J. Endocrinol. *242*, T121–T133.
9. Liu, Y., Chen, Q., Jeong, H.W., Koh, B.I., Watson, E.C., Xu, C., Stehling, M., Zhou, B., and Adams, R.H. (2022). A specialized bone marrow microenvironment for fetal haematopoiesis. Nat. Commun. *13*, 1327.
10. Qiao, L., Wynn, J., Yu, L., Hernan, R., Zhou, X., Duron, V., Aspelund, G., Farkouh-Karoleski, C., Zygumunt, A., Krishnan, U.S., et al. (2020). Likely damaging de novo variants in congenital diaphragmatic hernia patients are associated with worse clinical outcomes. Genet. Med. *22*, 2020–2028.
11. Qiao, L., Xu, L., Yu, L., Wynn, J., Hernan, R., Zhou, X., Farkouh-Karoleski, C., Krishnan, U.S., Khlevner, J., De, A., et al. (2021). Rare and de novo variants in 827 congenital diaphragmatic hernia probands implicate LONP1 as candidate risk gene. Am. J. Hum. Genet. *108*, 1964–1980.
12. Sanders, S.J., Murtha, M.T., Gupta, A.R., Murdoch, J.D., Raubeson, M.J., Willsey, A.J., Ercan-Sencicek, A.G., DiLullo, N.M., Parikshak, N.N., Stein, J.L., et al. (2012). De novo mutations revealed by whole-exome sequencing are strongly associated with autism. Nature *485*, 237–241.
13. Homsy, J., Zaidi, S., Shen, Y., Ware, J.S., Samocha, K.E., Karczewski, K.J., DePalma, S.R., McKean, D., Wakimoto, H., Gorham, J., et al. (2015). De novo mutations in congenital heart disease with neurodevelopmental and other congenital anomalies. Science *350*, 1262–1266.
14. Yu, L., Wynn, J., Ma, L., Guha, S., Mychaliska, G.B., Crombleholme, T.M., Azarow, K.S., Lim, F.Y., Chung, D.H., Potoka, D., et al. (2012). De novo copy number variants are associated with congenital diaphragmatic hernia. J. Med. Genet. *49*, 650–659.
15. Longoni, M., High, F.A., Qi, H., Joy, M.P., Hila, R., Coletti, C.M., Wynn, J., Loscertales, M., Shan, L., Bult, C.J., et al. (2017). Genome-wide enrichment of damaging de novo variants in patients with isolated and complex congenital diaphragmatic hernia. Hum. Genet. *136*, 679–691.
16. Scott, D.A., Gofin, Y., Berry, A.M., and Adams, A.D. (2022). Underlying genetic etiologies of congenital diaphragmatic hernia. Prenat. Diagn. *42*, 373–386.
17. Wild, K.T., Schindewolf, E., Hedrick, H.L., Rintoul, N.E., Hartman, T., Gebb, J., Moldenhauer, J.S., Zackai, E.H., and Krantz, I.D. (2022). The Genomics of Congenital Diaphragmatic Hernia: A 10-Year Retrospective Review. J. Pediatr. *248*, 108–113.e2.
18. Longoni, M., Pober, B.R., and High, F.A. (2020). Congenital Diaphragmatic Hernia Overview. In GeneReviews, M.P. Adam, J. Feldman, and G.M. Mirzaa, eds. (University of Washington, Seattle, WA). https://www.ncbi.nlm.nih.gov/books/NBK1359/.
19. Harris, P.A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., and Conde, J.G. (2009). Research electronic data capture (REDCap)-A metadata-driven methodology and workflow process for providing translational research informatics support. J. Biomed. Inform. *42*, 377–381.
20. Qi, H., Yu, L., Zhou, X., Wynn, J., Zhao, H., Guo, Y., Zhu, N., Kitaygorodsky, A., Hernan, R., Aspelund, G., et al. (2018). De novo variants in congenital diaphragmatic hernia identify MYRF as a new syndrome and reveal genetic overlaps with other developmental disorders. PLoS Genet. *14*, e1007822.
21. Zhou, X., Feliciano, P., Shu, C., Wang, T., Astrovskaya, I., Hall, J.B., Obiajulu, J.U., Wright, J.R., Murali, S.C., Xu, S.X., et al. (2022). Integrating de novo and inherited variants in 42,607 autism cases identifies mutations in new moderate-risk genes. Nat. Genet. *54*, 1305–1319.
22. Schmermund, A., Möhlenkamp, S., Stang, A., Grönemeyer, D., Seibel, R., Hirche, H., Mann, K., Siffert, W., Lauterbach, K., Siegrist, J., et al. (2002). Assessment of clinically silent atherosclerotic disease and established and novel risk factors for predicting myocardial infarction and cardiac death in healthy middle-aged subjects: rationale and design of the Heinz Nixdorf RECALL Study. Risk Factors, Evaluation of Coronary Calcium and Lifestyle. Am. Heart J. *144*, 212–218.
23. Zhu, N., Swietlik, E.M., Welch, C.L., Pauciulo, M.W., Hagen, J.J., Zhou, X., Guo, Y., Karten, J., Pandya, D., Tilly, T., et al. (2021). Rare variant analysis of 4241 pulmonary arterial hypertension cases from an international consortium implicates FBLN2, PDGFD, and rare de novo variants in PAH. Genome Med. *13*, 80.
24. Li, H., Ruan, J., and Durbin, R. (2008). Mapping short DNA sequencing reads and calling variants using mapping quality scores. Genome Res. *18*, 1851–1858.
25. DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. *43*, 491–498.
26. Van der Auwera, G.A., Carneiro, M.O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., et al. (2013). From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr. Protoc. Bioinformatics *43*, 11.10.1–11.10.33.
27. Poplin, R., Chang, P.C., Alexander, D., Schwartz, S., Colthurst, T., Ku, A., Newburger, D., Dijamco, J., Nguyen, N., Afshar, P.T., et al. (2018). A universal SNP and small-indel variant caller using deep neural networks. Nat. Biotechnol. *36*, 983–987.
28. Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat. Biotechnol. *29*, 24–26.
29. McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R.S., Thormann, A., Flicek, P., and Cunningham, F. (2016). The Ensembl Variant Effect Predictor. Genome Biol. *17*, 122.

30. Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. *38*, e164.
31. Kircher, M., Witten, D.M., Jain, P., O'Roak, B.J., Cooper, G.M., and Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. Nat. Genet. *46*, 310–315.
32. Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P., et al. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. Nature *581*, 434–443.
33. Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., et al. (2016). Analysis of protein-coding genetic variation in 60,706 humans. Nature *536*, 285–291.
34. Manichaikul, A., Mychaleckyj, J.C., Rich, S.S., Daly, K., Sale, M., and Chen, W.M. (2010). Robust relationship inference in genome-wide association studies. Bioinformatics *26*, 2867–2873.
35. Pedersen, B.S., and Quinlan, A.R. (2017). Who's Who? Detecting and Resolving Sample Anomalies in Human DNA Sequencing Studies with Peddy. Am. J. Hum. Genet. *100*, 406–413.
36. Taliun, D., Harris, D.N., Kessler, M.D., Carlson, J., Szpiech, Z.A., Torres, R., Taliun, S.A.G., Corvelo, A., Gogarten, S.M., Kang, H.M., et al. (2021). Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. Nature *590*, 290–299.
37. Chen, S., Francioli, L.C., Goodrich, J.K., Collins, R.L., Kanai, M., Wang, Q., Alföldi, J., Watts, N.A., Vittal, C., Gauthier, L.D., et al. (2024). A genomic mutational constraint map using variation in 76,156 human genomes. Nature *625*, 92–100.
38. Das, S., Forer, L., Schönherr, S., Sidore, C., Locke, A.E., Kwong, A., Vrieze, S.I., Chew, E.Y., Levy, S., McGue, M., et al. (2016). Next-generation genotype imputation service and methods. Nat. Genet. *48*, 1284–1287.
39. Byrska-Bishop, M., Evani, U.S., Zhao, X., Basile, A.O., Abel, H.J., Regier, A.A., Corvelo, A., Clarke, W.E., Musunuri, R., Nagulapalli, K., et al. (2022). High-coverage whole-genome sequencing of the expanded 1000 Genomes Project cohort including 602 trios. Cell *185*, 3426–3440.e19.
40. Loh, P.R., Danecek, P., Palamara, P.F., Fuchsberger, C., A Reshef, Y., K Finucane, H., Schoenherr, S., Forer, L., McCarthy, S., Abecasis, G.R., et al. (2016). Reference-based phasing using the Haplotype Reference Consortium panel. Nat. Genet. *48*, 1443–1448.
41. Samocha, K.E., Robinson, E.B., Sanders, S.J., Stevens, C., Sabo, A., McGrath, L.M., Kosmicki, J.A., Rehnström, K., Mallick, S., Kirby, A., et al. (2014). A framework for the interpretation of de novo mutation in human disease. Nat. Genet. *46*, 944–950.
42. Nguyen, H.T., Bryois, J., Kim, A., Dobbyn, A., Huckins, L.M., Munoz-Manchado, A.B., Ruderfer, D.M., Genovese, G., Fromer, M., Xu, X., et al. (2017). Integrated Bayesian analysis of rare exonic variants to identify risk genes for schizophrenia and neurodevelopmental disorders. Genome Med. *9*, 114.
43. He, X., Sanders, S.J., Liu, L., De Rubeis, S., Lim, E.T., Sutcliffe, J.S., Schellenberg, G.D., Gibbs, R.A., Daly, M.J., Buxbaum, J.D., et al. (2013). Integrated model of de novo and inherited genetic variants yields greater power to identify risk genes. PLoS Genet. *9*, e1003671.
44. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., and Sham, P.C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. *81*, 559–575.
45. Turner, S, D. (2018). qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. J. Open Source Softw. *3*.
46. Pruim, R.J., Welch, R.P., Sanna, S., Teslovich, T.M., Chines, P.S., Gliedt, T.P., Boehnke, M., Abecasis, G.R., and Willer, C.J. (2010). LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics *26*, 2336–2337.
47. 1000 Genomes Project Consortium, Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., and Abecasis, G.R. (2015). A global reference for human genetic variation. Nature *526*, 68–74.
48. Delaneau, O., Howie, B., Cox, A.J., Zagury, J.F., and Marchini, J. (2013). Haplotype estimation using sequencing reads. Am. J. Hum. Genet. *93*, 687–696.
49. Howie, B.N., Donnelly, P., and Marchini, J. (2009). A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet. *5*, e1000529.
50. Willer, C.J., Li, Y., and Abecasis, G.R. (2010). METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics *26*, 2190–2191.
51. Bulik-Sullivan, B.K., Loh, P.R., Finucane, H.K., Ripke, S., Yang, J., Schizophrenia Working Group of the Psychiatric Genomics Consortium, Patterson, N., Daly, M.J., Price, A.L., and Neale, B.M. (2015). LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat. Genet. *47*, 291–295.
52. Prive, F., Arbel, J., and Vilhjalmsson, B.J. (2021). LDpred2: better, faster, stronger. Bioinformatics *36*, 5424–5431.
53. Weiner, D.J., Wigdor, E.M., Ripke, S., Walters, R.K., Kosmicki, J.A., Grove, J., Samocha, K.E., Goldstein, J.I., Okbay, A., Bybjerg-Grauholm, J., et al. (2017). Polygenic transmission disequilibrium confirms that common and rare variation act additively to create risk for autism spectrum disorders. Nat. Genet. *49*, 978–985.
54. McArthur, E., and Capra, J.A. (2021). Topologically associating domain boundaries that are stable across diverse cell types are evolutionarily constrained and enriched for heritability. Am. J. Hum. Genet. *108*, 269–283.
55. Dixon, J.R., Jung, I., Selvaraj, S., Shen, Y., Antosiewicz-Bourget, J.E., Lee, A.Y., Ye, Z., Kim, A., Rajagopal, N., Xie, W., et al. (2015). Chromatin architecture reorganization during stem cell differentiation. Nature *518*, 331–336.
56. Wang, Y., Song, F., Zhang, B., Zhang, L., Xu, J., Kuang, D., Li, D., Choudhary, M.N.K., Li, Y., Hu, M., et al. (2018). The 3D Genome Browser: a web-based browser for visualizing 3D genome organization and long-range chromatin interactions. Genome Biol. *19*, 151.
57. Jung, I., Schmitt, A., Diao, Y., Lee, A.J., Liu, T., Yang, D., Tan, C., Eom, J., Chan, M., Chee, S., et al. (2019). A compendium of promoter-centered long-range chromatin interactions in the human genome. Nat. Genet. *51*, 1442–1449.
58. Hammal, F., de Langen, P., Bergon, A., Lopez, F., and Ballester, B. (2022). ReMap 2022: a database of Human, Mouse,

Drosophila and Arabidopsis regulatory regions from an integrative analysis of DNA-binding sequencing experiments. Nucleic Acids Res. *50*, D316–D325.

59. Nassar, L.R., Barber, G.P., Benet-Pagès, A., Casper, J., Clawson, H., Diekhans, M., Fischer, C., Gonzalez, J.N., Hinrichs, A.S., Lee, B.T., et al. (2023). The UCSC Genome Browser database: 2023 update. Nucleic Acids Res. *51*, D1188–D1195.
60. Raney, B.J., Dreszer, T.R., Barber, G.P., Clawson, H., Fujita, P.A., Wang, T., Nguyen, N., Paten, B., Zweig, A.S., Karolchik, D., and Kent, W.J. (2014). Track data hubs enable visualization of user-defined genome-wide annotations on the UCSC Genome Browser. Bioinformatics *30*, 1003–1005.
61. Wang, A., Chiou, J., Poirion, O.B., Buchanan, J., Valdez, M.J., Verheyden, J.M., Hou, X., Kudtarkar, P., Narendra, S., Newsome, J.M., et al. (2020). Single-cell multiomic profiling of human lungs reveals cell-type-specific and age-dynamic control of SARS-CoV2 host genes. Elife *9*, e62522.
62. Dos Santos, M., Shah, A.M., Zhang, Y., Bezprozvannaya, S., Chen, K., Xu, L., Lin, W., McAnally, J.R., Bassel-Duby, R., Liu, N., and Olson, E.N. (2023). Opposing gene regulatory programs governing myofiber development and maturation revealed at single nucleus resolution. Nat. Commun. *14*, 4333.
63. Hinrichs, A.S., Karolchik, D., Baertsch, R., Barber, G.P., Bejerano, G., Clawson, H., Diekhans, M., Furey, T.S., Harte, R.A., Hsu, F., et al. (2006). The UCSC Genome Browser Database: update 2006. Nucleic Acids Res. *34*, D590–D598.
64. Russell, M.K., Longoni, M., Wells, J., Maalouf, F.I., Tracy, A.A., Loscertales, M., Ackerman, K.G., Pober, B.R., Lage, K., Bult, C.J., and Donahoe, P.K. (2012). Congenital diaphragmatic hernia candidate genes derived from embryonic transcriptomes. Proc. Natl. Acad. Sci. USA *109*, 2978–2983.
65. Magoulas, P.L., and El-Hattab, A.W. (2012). Chromosome 15q24 microdeletion syndrome. Orphanet J. Rare Dis. *7*, 2.
66. Yuan, B., Neira, J., Pehlivan, D., Santiago-Sim, T., Song, X., Rosenfeld, J., Posey, J.E., Patel, V., Jin, W., Adam, M.P., et al. (2019). Clinical exome sequencing reveals locus heterogeneity and phenotypic variability of cohesinopathies. Genet. Med. *21*, 663–675.
67. Murch, O., Jain, V., Benneche, A., Metcalfe, K., Hobson, E., Prescott, K., Chandler, K., Ghali, N., Carmichael, J., Foulds, N.C., et al. (2022). Further delineation of the clinical spectrum of White-Sutton syndrome: 12 new individuals and a review of the literature. Eur. J. Hum. Genet. *30*, 95–100.
68. Merriweather, A., Murdock, D.R., Rosenfeld, J.A., Dai, H., Ketkar, S., Emrick, L., Nicholas, S., Lewis, R.A., Undiagnosed Diseases Network, and Bacino, C.A., et al. (2022). A novel, de novo intronic variant in POGZ causes White-Sutton syndrome. Am. J. Med. Genet. *188*, 2198–2203.
69. Yelagandula, R., Stecher, K., Novatchkova, M., Michetti, L., Michlits, G., Wang, J., Hofbauer, P., Vainorius, G., Pribitzer, C., Isbel, L., et al. (2023). ZFP462 safeguards neural lineage specification by targeting G9A/GLP-mediated heterochromatin to silence enhancers. Nat. Cell Biol. *25*, 42–55.
70. Kehrer, C., Hoischen, A., Menkhaus, R., Schwab, E., Müller, A., Kim, S., Kreiß, M., Weitensteiner, V., Hilger, A., Berg, C., et al. (2016). Whole exome sequencing and array-based molecular karyotyping as aids to prenatal diagnosis in fetuses with suspected Simpson-Golabi-Behmel syndrome. Prenat. Diagn. *36*, 961–965.
71. Yano, S., Baskin, B., Bagheri, A., Watanabe, Y., Moseley, K., Nishimura, A., Matsumoto, N., and Ray, P.N. (2011). Familial Simpson-Golabi-Behmel syndrome: studies of X-chromosome inactivation and clinical phenotypes in two female individuals with GPC3 mutations. Clin. Genet. *80*, 466–471.
72. Zimmermann, N., and Stanek, J. (2017). Perinatal Case of Fatal Simpson-Golabi-Behmel Syndrome with Hyperplasia of Seminiferous Tubules. Am. J. Case Rep. *18*, 649–655.
73. Peng, H.H., Yu, C.J., Chen, Y.C., Hsu, C.C., Chang, S.D., Chueh, H.Y., Chang, Y.L., Cheng, P.J., and Lee, Y.C. (2023). Prenatal diagnosis of Simpson-Golabi-Behmel syndrome type 1 with an 814 kb Xq26.2 deletion with the initial presentation of a thick nuchal fold. Taiwan. J. Obstet. Gynecol. *62*, 163–166.
74. Rossetti, L.Z., Glinton, K., Yuan, B., Liu, P., Pillai, N., Mizerik, E., Magoulas, P., Rosenfeld, J.A., Karaviti, L., Sutton, V.R., et al. (2019). Review of the phenotypic spectrum associated with haploinsufficiency of MYRF. Am. J. Med. Genet. *179*, 1376–1382.
75. Yu, L., Bennett, J.T., Wynn, J., Carvill, G.L., Cheung, Y.H., Shen, Y., Mychaliska, G.B., Azarow, K.S., Crombleholme, T.M., Chung, D.H., et al. (2014). Whole exome sequencing identifies de novo mutations in GATA6 associated with congenital diaphragmatic hernia. J. Med. Genet. *51*, 197–202.
76. Allen, H.L., Flanagan, S.E., Shaw-Smith, C., De Franco, E., Akerman, I., Caswell, R., International Pancreatic Agenesis Consortium, Ferrer, J., Hattersley, A.T., and Ellard, S. (2011). GATA6 haploinsufficiency causes pancreatic agenesis in humans. Nat. Genet. *44*, 20–22.
77. Kruszka, P., Hu, T., Hong, S., Signer, R., Cogné, B., Isidor, B., Mazzola, S.E., Giltay, J.C., van Gassen, K.L.I., England, E.M., et al. (2019). Phenotype delineation of ZNF462 related syndrome. Am. J. Med. Genet. *179*, 2075–2082.
78. Loscertales, M., Mikels, A.J., Hu, J.K.H., Donahoe, P.K., and Roberts, D.J. (2008). Chick pulmonary Wnt5a directs airway and vascular tubulogenesis. Development *135*, 1365–1376.
79. Maurano, M.T., Humbert, R., Rynes, E., Thurman, R.E., Haugen, E., Wang, H., Reynolds, A.P., Sandstrom, R., Qu, H., Brody, J., et al. (2012). Systematic localization of common disease-associated variation in regulatory DNA. Science *337*, 1190–1195.
80. Roadmap Epigenomics Consortium, Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., et al. (2015). Integrative analysis of 111 reference human epigenomes. Nature *518*, 317–330.
81. Anderson, E., Devenney, P.S., Hill, R.E., and Lettice, L.A. (2014). Mapping the Shh long-range regulatory domain. Development *141*, 3934–3943.
82. Sagai, T., Amano, T., Tamura, M., Mizushina, Y., Sumiyama, K., and Shiroishi, T. (2009). A cluster of three long-range enhancers directs regional Shh expression in the epithelial linings. Development *136*, 1665–1674.
83. Yu, L., Sawle, A.D., Wynn, J., Aspelund, G., Stolar, C.J., Arkovitz, M.S., Potoka, D., Azarow, K.S., Mychaliska, G.B., Shen, Y., and Chung, W.K. (2015). Increased burden of de novo predicted deleterious variants in complex congenital diaphragmatic hernia. Hum. Mol. Genet. *24*, 4764–4773.
84. Iossifov, I., O'Roak, B.J., Sanders, S.J., Ronemus, M., Krumm, N., Levy, D., Stessman, H.A., Witherspoon, K.T., Vives, L., Patterson, K.E., et al. (2014). The contribution of de novo coding mutations to autism spectrum disorder. Nature *515*, 216–221.

85. Turner, T.N., Wilfert, A.B., Bakken, T.E., Bernier, R.A., Pepper, M.R., Zhang, Z., Torene, R.I., Retterer, K., and Eichler, E.E. (2019). Sex-Based Analysis of De Novo Variants in Neurodevelopmental Disorders. Am. J. Hum. Genet. *105*, 1274–1285.
86. Tilghman, J.M., Ling, A.Y., Turner, T.N., Sosa, M.X., Krumm, N., Chatterjee, S., Kapoor, A., Coe, B.P., Nguyen, K.D.H., Gupta, N., et al. (2019). Molecular Genetic Anatomy and Risk Profile of Hirschsprung's Disease. N. Engl. J. Med. *380*, 1421–1432.
87. Kuil, L.E., MacKenzie, K.C., Tang, C.S., Windster, J.D., Le, T.L., Karim, A., de Graaf, B.M., van der Helm, R., van Bever, Y., Sloots, C.E.J., et al. (2021). Size matters: Large copy number losses in Hirschsprung disease patients reveal genes involved in enteric nervous system development. PLoS Genet. *17*, e1009698.
88. Gaisl, O., Konrad, D., Joset, P., and Lang-Muritano, M. (2019). A novel GATA6 variant in a boy with neonatal diabetes and diaphragmatic hernia: a familial case with a review of the literature. J. Pediatr. Endocrinol. Metab. *32*, 1027–1030.
89. Longoni, M., Lage, K., Russell, M.K., Loscertales, M., Abdul-Rahman, O.A., Baynam, G., Bleyl, S.B., Brady, P.D., Breckpot, J., Chen, C.P., et al. (2012). Congenital diaphragmatic hernia interval on chromosome 8p23.1 characterized by genetics and protein interaction networks. Am. J. Med. Genet. *158A*, 3148–3158.
90. Yu, L., Wynn, J., Cheung, Y.H., Shen, Y., Mychaliska, G.B., Crombleholme, T.M., Azarow, K.S., Lim, F.Y., Chung, D.H., Potoka, D., et al. (2013). Variants in GATA4 are a rare cause of familial and sporadic congenital diaphragmatic hernia. Hum. Genet. *132*, 285–292.
91. Brady, P.D., Van Houdt, J., Callewaert, B., Deprest, J., Devriendt, K., and Vermeesch, J.R. (2014). Exome sequencing identifies ZFPM2 as a cause of familial isolated congenital diaphragmatic hernia and possibly cardiovascular malformations. Eur. J. Med. Genet. *57*, 247–252.
92. Longoni, M., Russell, M.K., High, F.A., Darvishi, K., Maalouf, F.I., Kashani, A., Tracy, A.A., Coletti, C.M., Loscertales, M., Lage, K., et al. (2015). Prevalence and penetrance of ZFPM2 mutations and deletions causing congenital diaphragmatic hernia. Clin. Genet. *87*, 362–367.
93. Nicholas, T.J., Al-Sweel, N., Farrell, A., Mao, R., Bayrak-Toydemir, P., Miller, C.E., Bentley, D., Palmquist, R., Moore, B., Hernandez, E.J., et al. (2022). Comprehensive variant calling from whole-genome sequencing identifies a complex inversion that disrupts ZFPM2 in familial congenital diaphragmatic hernia. Mol. Genet. Genomic Med. *10*, e1888.
94. Yang, X., Yang, Y., Sun, B.F., Chen, Y.S., Xu, J.W., Lai, W.Y., Li, A., Wang, X., Bhattarai, D.P., Xiao, W., et al. (2017). 5-methylcytosine promotes mRNA export - NSUN2 as the methyltransferase and ALYREF as an m(5)C reader. Cell Res. *27*, 606–625.
95. Richter, J.D., and Zhao, X. (2021). The molecular biology of FMRP: new insights into fragile X syndrome. Nat. Rev. Neurosci. *22*, 209–222.
96. Gan, P., Wang, Z., Morales, M.G., Zhang, Y., Bassel-Duby, R., Liu, N., and Olson, E.N. (2022). RBPMS is an RNA-binding protein that mediates cardiomyocyte binucleation and cardiovascular development. Dev. Cell *57*, 959–973.e7.
97. Prashad, S., and Gopal, P.P. (2021). RNA-binding proteins in neurological development and disease. RNA Biol. *18*, 972–987.
98. Piche, J., Van Vliet, P.P., Puceat, M., and Andelfinger, G. (2019). The expanding phenotypes of cohesinopathies: one ring to rule them all. Cell Cycle *18*, 2828–2848.
99. Deciphering Developmental Disorders Study (2015). Large-scale discovery of novel genetic causes of developmental disorders. Nature *519*, 223–228.
100. Balasubramanian, M., Dingemans, A.J.M., Albaba, S., Richardson, R., Yates, T.M., Cox, H., Douzgou, S., Armstrong, R., Sansbury, F.H., Burke, K.B., et al. (2021). Comprehensive study of 28 individuals with SIN3A-related disorder underscoring the associated mild cognitive and distinctive facial phenotype. Eur. J. Hum. Genet. *29*, 625–636.
101. Cowley, S.M., Iritani, B.M., Mendrysa, S.M., Xu, T., Cheng, P.F., Yada, J., Liggitt, H.D., and Eisenman, R.N. (2005). The mSin3A chromatin-modifying complex is essential for embryogenesis and T-cell development. Mol. Cell Biol. *25*, 6990–7004.
102. Dannenberg, J.H., David, G., Zhong, S., van der Torre, J., Wong, W.H., and Depinho, R.A. (2005). mSin3A corepressor regulates diverse transcriptional networks governing normal and neoplastic growth and survival. Genes Dev. *19*, 1581–1595.
103. Stokes, G., Li, Z., Talaba, N., Genthe, W., Brix, M.B., Pham, B., Wienhold, M.D., Sandok, G., Hernan, R., Wynn, J., et al. (2024). Rescuing lung development through embryonic inhibition of histone acetylation. Sci. Transl. Med. *16*, eadc8930.
104. Bisserier, M., Mathiyalagan, P., Zhang, S., Elmastour, F., Dorfmüller, P., Humbert, M., David, G., Tarzami, S., Weber, T., Perros, F., et al. (2021). Regulation of the Methylation and Expression Levels of the BMPR2 Gene by SIN3a as a Novel Therapeutic Mechanism in Pulmonary Arterial Hypertension. Circulation *144*, 52–73.
105. Yamaguchi, T.P., Bradley, A., McMahon, A.P., and Jones, S. (1999). A Wnt5a pathway underlies outgrowth of multiple structures in the vertebrate embryo. Development *126*, 1211–1223.
106. Cervantes, S., Yamaguchi, T.P., and Hebrok, M. (2009). Wnt5a is essential for intestinal elongation in mice. Dev. Biol. *326*, 285–294.
107. Yu, H., Smallwood, P.M., Wang, Y., Vidaltamayo, R., Reed, R., and Nathans, J. (2010). Frizzled 1 and frizzled 2 genes function in palate, ventricular septum and neural tube closure: general implications for tissue fusion processes. Development *137*, 3707–3717.
108. Vladar, E.K., and Königshoff, M. (2020). Noncanonical Wnt planar cell polarity signaling in lung development and disease. Biochem. Soc. Trans. *48*, 231–243.
109. Skoric-Milosavljevic, D., Tadros, R., Bosada, F.M., Tessadori, F., van Weerd, J.H., Woudstra, O.I., Tjong, F.V.Y., Lahrouchi, N., Bajolle, F., Cordell, H.J., et al. (2022). Common Genetic Variants Contribute to Risk of Transposition of the Great Arteries. Circ. Res. *130*, 166–180.
110. Wu, D., and Dean, J. (2020). Maternal factors regulating preimplantation development in mice. Curr. Top. Dev. Biol. *140*, 317–340.
111. Laugesen, A., and Helin, K. (2014). Chromatin repressive complexes in stem cells, development, and cancer. Cell Stem Cell *14*, 735–751.
112. Bessonnard, S., De Mot, L., Gonze, D., Barriol, M., Dennis, C., Goldbeter, A., Dupont, G., and Chazaud, C. (2014). Gata6, Nanog and Erk signaling control cell fate in the inner cell mass through a tristable regulatory network. Development *141*, 3637–3648.

113. Belgacemi, R., Danopoulos, S., Deutsch, G., Glass, I., Dormoy, V., Bellusci, S., and Al Alam, D. (2022). Hedgehog Signaling Pathway Orchestrates Human Lung Branching Morphogenesis. Int. J. Mol. Sci. *23*, 5265.
114. Chiang, C., Litingtung, Y., Lee, E., Young, K.E., Corden, J.L., Westphal, H., and Beachy, P.A. (1996). Cyclopia and defective axial patterning in mice lacking Sonic hedgehog gene function. Nature *383*, 407–413.
115. Ushiki, A., Zhang, Y., Xiong, C., Zhao, J., Georgakopoulos-Soares, I., Kane, L., Jamieson, K., Bamshad, M.J., Nickerson, D.A., et al.; University of Washington Center for Mendelian Genomics (2021). Deletion of CTCF sites in the SHH locus alters enhancer-promoter interactions and leads to acheiropodia. Nat. Commun. *12*, 2282.
116. Qin, W., Chen, Z., Zhang, Y., Yan, R., Yan, G., Li, S., Zhong, H., and Lin, S. (2014). Nom1 mediates pancreas development by regulating ribosome biogenesis in zebrafish. PLoS One *9*, e100796.
117. Harrison, K.A., Thaler, J., Pfaff, S.L., Gu, H., and Kehrl, J.H. (1999). Pancreas dorsal lobe agenesis and abnormal islets of Langerhans in Hlxb9-deficient mice. Nat. Genet. *23*, 71–75.
118. Cantor, A.B., and Orkin, S.H. (2005). Coregulation of GATA factors by the Friend of GATA (FOG) family of multitype zinc finger proteins. Semin. Cell Dev. Biol. *16*, 117–128.
119. Ng, A., Wong, M., Viviano, B., Erlich, J.M., Alba, G., Pflederer, C., Jay, P.Y., and Saunders, S. (2009). Loss of glypican-3 function causes growth factor-dependent defects in cardiac and coronary vascular development. Dev. Biol. *335*, 208–215.
120. Li, F., Shi, W., Capurro, M., and Filmus, J. (2011). Glypican-5 stimulates rhabdomyosarcoma cell proliferation by activating Hedgehog signaling. J. Cell Biol. *192*, 691–704.
121. Song, H.H., Shi, W., Xiang, Y.Y., and Filmus, J. (2005). The loss of glypican-3 induces alterations in Wnt signaling. J. Biol. Chem. *280*, 2116–2125.
122. Goodrich, J.K., Singer-Berk, M., Son, R., Sveden, A., Wood, J., England, E., Cole, J.B., Weisburd, B., Watts, N., Caulkins, L., et al. (2021). Determinants of penetrance and variable expressivity in monogenic metabolic conditions across 77,184 exomes. Nat. Commun. *12*, 3505.
123. Antaki, D., Guevara, J., Maihofer, A.X., Klein, M., Gujral, M., Grove, J., Carey, C.E., Hong, O., Arranz, M.J., Hervas, A., et al. (2022). A phenotypic spectrum of autism is attributable to the combined effects of rare variants, polygenic risk and sex. Nat. Genet. *54*, 1284–1292.