SUPPLEMENTARY INFORMATION (SI) APPENDIX for:

Systematic analysis of copy number variation associated with congenital diaphragmatic hernia

Qihui Zhu[1*], Frances A. High[2,3,4*], Chengsheng Zhang[1*], Eliza Cerveira[1], Meaghan Russell[2], Mauro Longoni[2,4], Maliackal P. Joy[2], Mallory Ryan[1], Adam Mil-homens [1], Lauren Bellfy[1], Caroline Coletti[2], Pooja Bhayani[2], Regis Hila[2], Jay Wilson[3,4], Patricia K. Donahoe[2,4**], Charles Lee[1,5**]

1 The Jackson Laboratory for Genomic Medicine, Farmington, Connecticut 06032,
2 Pediatric Surgical Research Laboratories, Massachusetts General Hospital, Boston, MA 02114
3 Department of Surgery, Boston Children's Hospital, Boston, MA 02115
4 Harvard Medical School, Boston MA 02115
5 Department of Life Sciences, Ewha Womans University, Seoul, South Korea
*:co-first authors
**: Co-senior authors

Corresponding Author: Patricia K. Donahoe

Email: pdonahoe@mgh.harvard.edu

This PDF includes:

Supplemental Materials and Methods
Figures S1, S2, S3, S4, S5
Tables S1, S2, S3, S4, S5, S6

Supplemental Materials and Methods

Custom array CGH design

Each target region was covered by 9-700 probes with an additional 20 probes outside of the target region. The remaining probes were distributed uniformly across the genome to create a "backbone". Overall, a total of 37,907 probes were placed to cover targeted regions, 2,934 probes to cover flanking regions and 8,170 probes to cover backbones in this array. The average density of the probe coverage was ~51 kb at the genome level, and the average density of targeted region was ~8 kb (0.1-200 kb, Fig. 1B). Array design was tested for reproducibility and quality before use. Overall, this format allowed for the detection of CNVs of 500 bp (or larger) within the 140 targeted genomic regions, CNVs of 6 kb (or larger) flanking 10 kb of a targeted region, and 1.75 Mb (or larger) for the rest of the genome.

Array Comparative Genomic (aCGH) Hybridization

For each aCGH hybridization, we aliquoted 500-750 ng of each genomic DNA, of either the experimental sample or the corresponding reference sample (Promega Male or Female), with TE buffer (pH 8.0) in a total volume of 20.2 µL. Samples were digested using the aCGH Labeling kit and protocol (Agilent Technologies) and quality was assessed by electrophoresis (1% agarose gel). Samples were then fluorescently labeled using a modified protocol including both the BioPrime Labeling Kit (ThermoFisher) and the Agilent aCGH labeling kit. In this modified procedure, 20 µL of 2.5X Random Primer Mix (ThermoFisher) was added to each sample and then denatured at 98°C for 5 minutes. Samples were cooled and mixed with 5 µL of 10X dUTP Mix (ThermoFisher), 3 µL of either Cy5-dUTP (for the experimental sample) or Cy3-dUTP (for the reference sample), respectively, as well as 1 µL of Exo-Klenow (Agilent

Technologies). Labeling reaction conditions, subsequent quality control steps, and review of passing criteria (i.e. specific activity calculations) were followed using the Agilent aCGH protocol. Generally, a specific activity value of greater than 30 pmol dyes per µg gDNA and a yield greater than 3 µg was considered acceptable. However, in certain situations (e.g. limited stock DNA) this QC criterion could be waived on a case-by-case basis. For hybridization, 9.5 µL of labeled experimental gDNA was combined with 9.5 µL of reference gDNA, 5 µg of human Cot-1 DNA (CHIMERx), 6 µL of 10X blocking agent (Agilent Oligo aCGH Hybridization Kit), and 30 µL of 2X hybridization buffer (Agilent Technologies) in a final volume of 60 µL. Hybridization reaction conditions were followed by the Agilent aCGH protocol. Following hybridization, the arrays were washed using either the standard or extended washing protocol as described by Agilent. Arrays were loaded into ozone-protected covers and then scanned using an Agilent G2505C DNA microarray scanner in an ozone-protected hood. The final feature extraction files were used for data analysis.

aCGH data analysis

The raw data from the scanner were normalized using the Feature Extraction Software version 10.5.1.1 and QC files were manually inspected for quality assurance. aCGH performance was evaluated based on the following QC metrics: derivative $\log_2$ ratio spread (DLRSpread) <0.3; Signal-to-noise ratio (STNR) >30; signal intensity >50 for both the red and the green channels; background noisy <15 and reproducibility < 20. Only arrays that passed the QC parameters continued to further data analysis.

The CNV calling and visualization was performed using Genomic Workbench version 7.0.4.0 and our customized scripts. We applied the Aberration Detection Methods 2 (ADM-2) statistical algorithm. This algorithm finds the change point that maximizes the *t*-test of comparing the averages between change points to 0 (Agilent Technologies). When a segment is kept, it is median centered and the procedure is repeated on the three new segments. This effectively combines the segmentation and calling process into one step. A fuzzy zero method is applied to incorporate quality information about each probe measurement. A threshold with a minimum of 5 consecutive probes, and a log ratio higher than 0.5 for loss or higher than 0.25 for gain was used. We also developed a customized script to compare each call with refseq genes, OMIM genes, Decipher database, common CNVs, etc., and visualize it in a custom-designed user-friendly interface.

The functional enrichment analysis of genes within the significant CNVs was carried out using the Database for Annotation, Visualization and Integrated Discovery (DAVID) (1), and the WebGestalt software (2). P < 0.05 was set as the cut-off. Next, networks of protein-protein interaction involving genes from each functional cluster were constructed using the STRING V10 software (3). For this analysis, we identified any interaction from i) known interaction including experimentally determined and curated databases, ii) predicted interactions including gene neighborhood, gene fusions and gene-concordance, iii) other interactions including text mining, co-expression and protein homology. The minimum required interaction score was set at median confidence >=0.4.

Digital droplet PCR for CNV validation

Customized primers and probes were designed for each target region using Primer3Plus (4). All primer pairs were tested for their uniqueness across the human genome using the In-Silico PCR component of the UCSC Genome Browser. The ddPCR assays were performed following the Bio-Rad QX200™ system manufacturer's protocol. A total of 10 ng of DNA template was mixed with 2X ddPCR SuperMix for probes (no dUTP) (BioRad), *Hind*III-HF enzyme (2U/reaction) (New England BioLabs), 20X primer/probe (both FAM and HEX-labeled probes), and water to a

final volume of 20 µL. Each reaction mixture was then loaded into the sample well of an eight-channel droplet generator cartridge. PCR amplification was performed using a C1000 Touch thermal cycler with the following conditions for CNV detection: enzyme activation at 95°C for 10 minutes, denaturation and extension at 94°C for 30 seconds and 60° C for 1 minute for a total of 40 cycles, enzyme deactivation at 98° C for 10 minutes, finished with a 4° C hold. Once completed, the 96-well PCR plate was loaded on the QX200™ Droplet Reader. All experiments had at least two normal controls (NA12878 and NA10851), and a no-template control (NTC) with water. Analysis of the ddPCR data was performed using the QuantaSoft™ software. Data from any well with less than 8,000 droplets was treated as failed QC and excluded from downstream analysis.

SI References:

1. Huang DW, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 4(1):44–57.
2. Zhang B, Kirov S, Snoddy J (2005) WebGestalt: an integrated system for exploring gene sets in various biological contexts. *Nucleic Acids Res* 33(suppl 2):W741--W748.
3. Szklarczyk D, et al. (2014) STRING v10: protein--protein interaction networks, integrated over the tree of life. *Nucleic Acids Res*:gku1003.
4. Huang C-C, Orvis GD, Kwan KM, Behringer RR (2014) Lhx1 is required in M{ü}llerian duct epithelium for uterine development. Dev Biol 389(2):124–136.

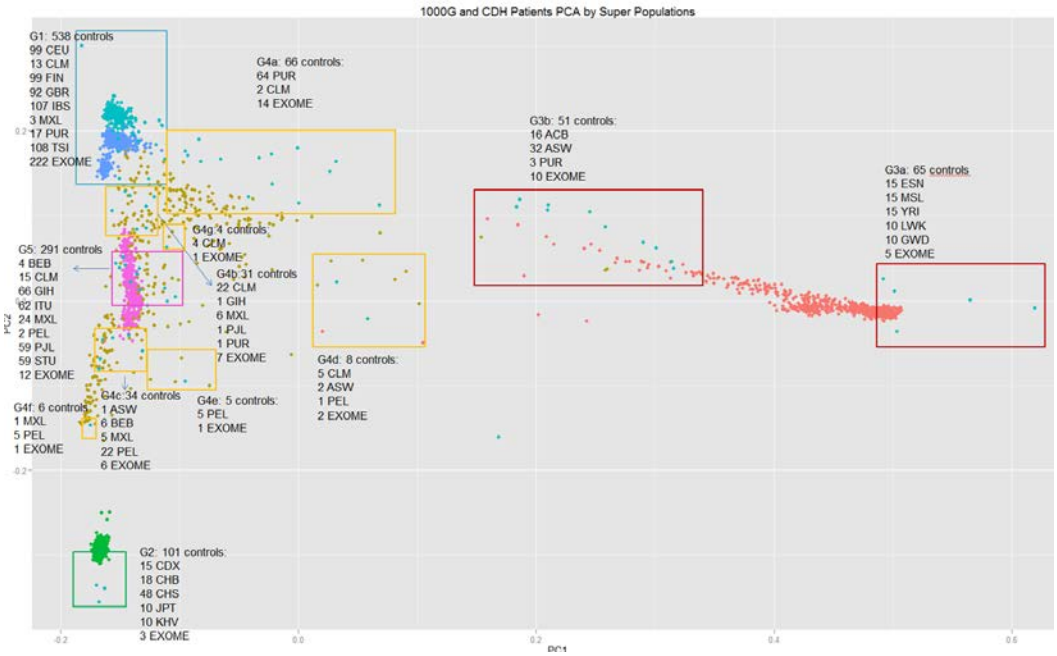


Fig. S1. Principal component analysis on CDH cohort exomes and 1000 Genomes populations. The rectangle boxes indicate samples we selected as controls for our study. Population codes are as follows: CEU: Utah Residents (CEPH) with Northern and Western European Ancestry; CHB: Han Chinese Beijing, China; CHS: Han Chinese South; ACB: African Caribbean in Barbados; ASW: African Ancestry in SW USA; BEB: Bengali in Bangladesh; CDX: Chinese Dai in Xishuangban, China; CLM: Colombian from

Medellin, Colombia; ESN: Esan from Nigeria; FIN: Finnish in Finland; GBR: British from England and Scotland; GIH: Gujarati Indian in TX, USA; GWD: Gambian in Western Divison; IBS: Iberian in Spain; ITU: Indian Telugu in the UK; JPT: Japanese in Tokyo, Japan; KHV: Kinh in Ho Chi Minh City, Vietnam; LWK: Luhya in Webuye, Kenya; MSL: Mende in Sierra Leone; MXL: Mexican from LA, USA; PEL: Peruvian in Lima, Peru; PJL: Punjabi in Lahore, Pakistan; PUR: Puerto Rican from Puerto Rico; STU: Sri Lankan Tamil in UK; TSI: Toscani in Italy; YRI: Yoruba in Ibadan, Nigeria. The numbers before the population codes indicate the number of samples we selected for our study

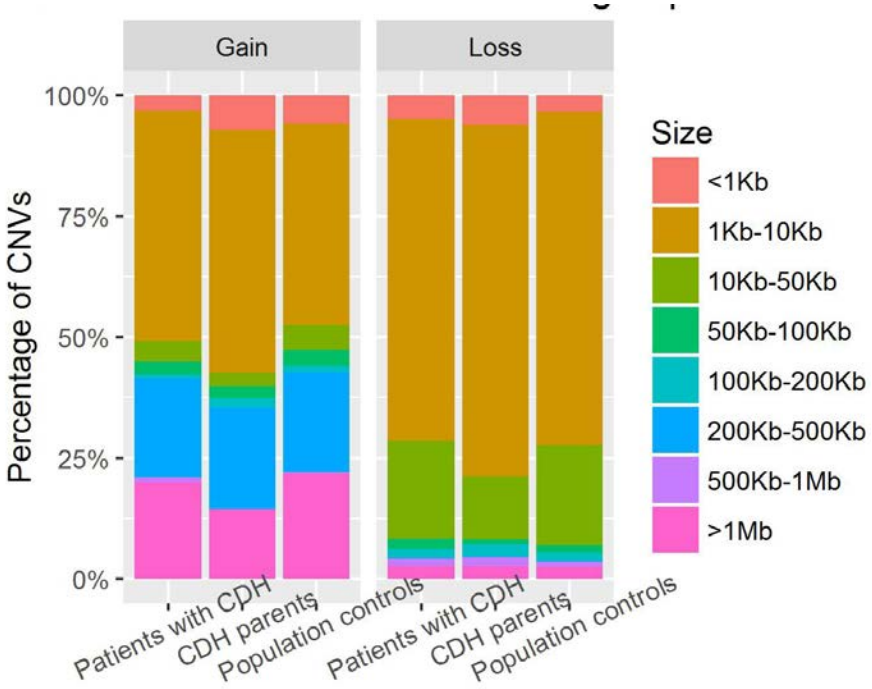
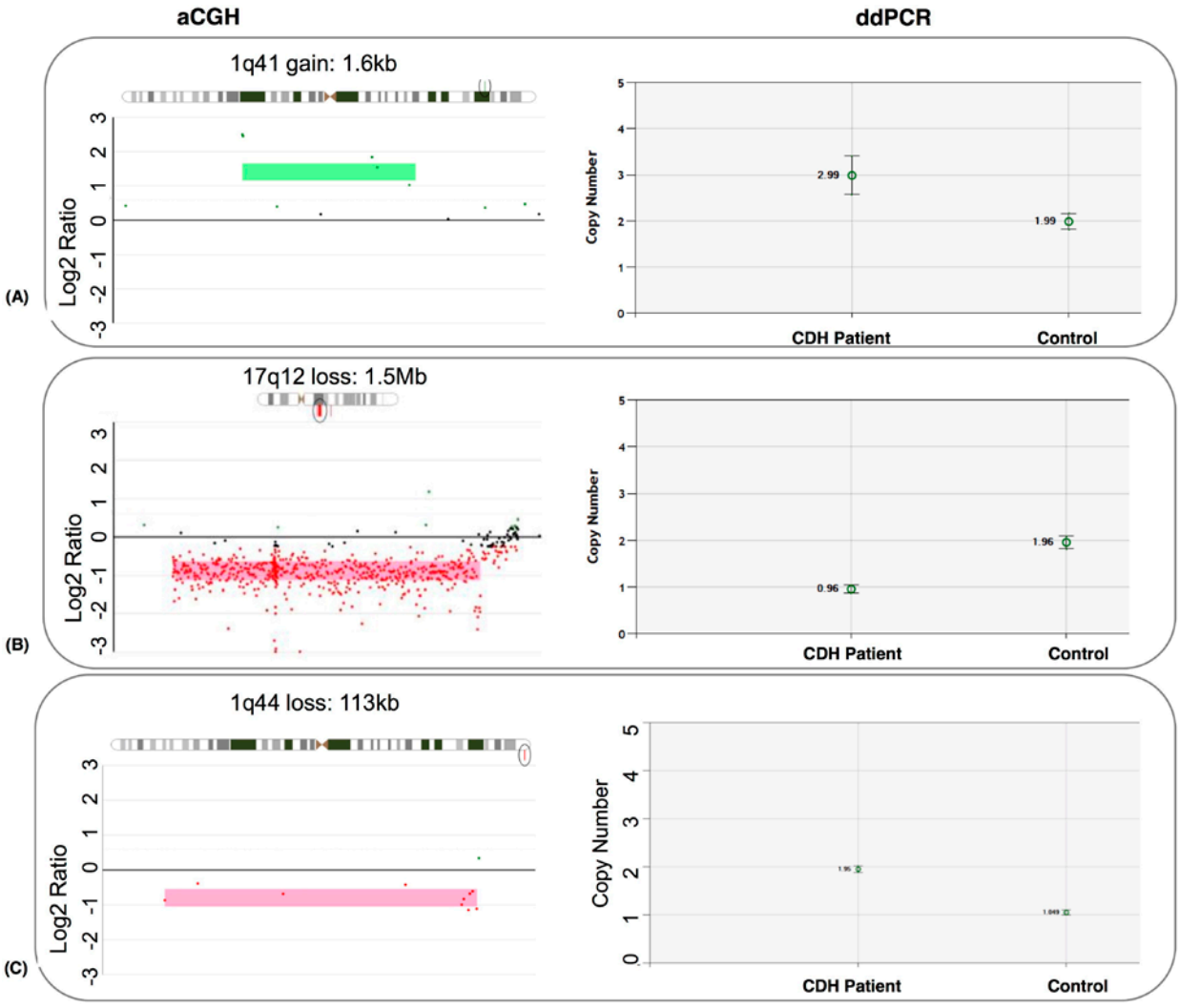


Fig. S2. CNV size distribution in CDH patients and controls


aCGH
ddPCR
(A)
1q41 gain: 1.6kb
Log2 Ratio
Copy Number
2.99
1.99
CDH Patient
Control
(B)
17q12 loss: 1.5Mb
Log2 Ratio
Copy Number
0.96
1.96
CDH Patient
Control
(C)
1q44 loss: 113kb
Log2 Ratio
Copy Number
CDH Patient
Control

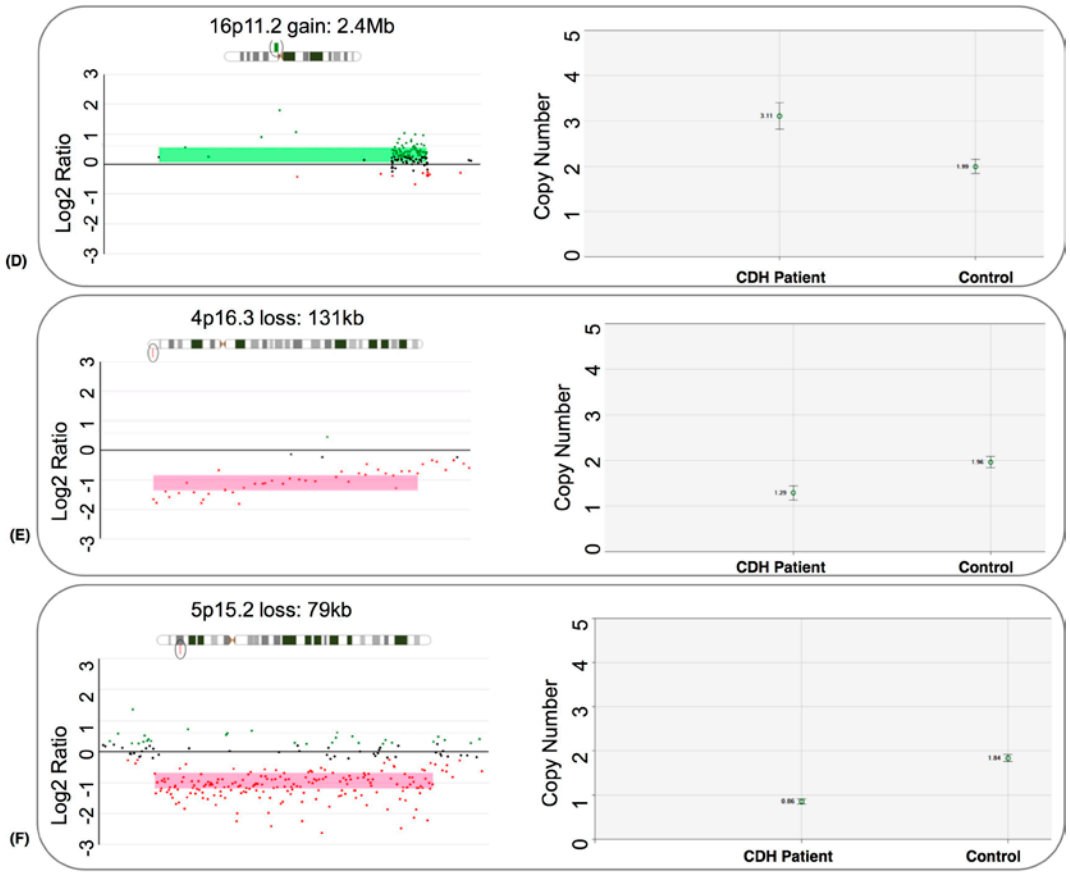


Fig. S3. Validation of 6 significant CNVs by ddPCR. The left panel is the aCGH profile. The green rectangle region represents a gain and the red rectangle region indicates a loss. The dots are the probes designed for the customized array. The right panel is the corresponding ddPCR profile. The significant CNVs detected by customized array were able to be successfully validated by ddPCR. A) A 1.6kb copy number (CN) gain at 1q41. B) a 1.5Mb CN loss at 17q12; C) a 113kb CN loss at 1q44; D) A 2.4Mb CN gain at 16p11.2; E) a 131kb CN loss at 4p16.3; and F) a 79kb CN loss at 5p15.2.

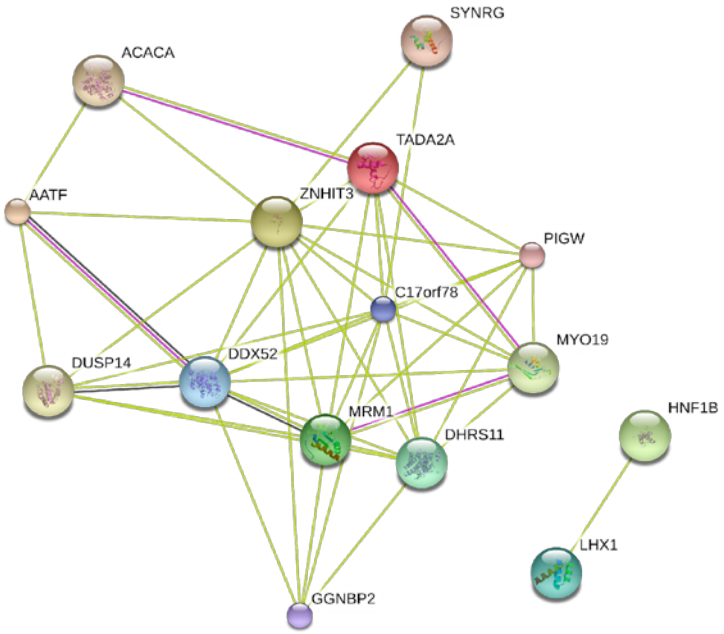


Fig. S4. Functional interaction network generated by the STRING analysis for the 41 genes in significant CDH-associated CNVs. 26 genes with known proteins in the STRING database were included in this analysis. 13 genes formed an interaction network (Left side) and another two genes interacted with each other (Right side) with a minimum required interaction score of medium confidence 0.4. Colored nodes represent genes. Small nodes indicate protein of unknown 3D structure and large nodes indicate 3D structure is known or predicted. Colors of the graph edges indicate the types of interaction evidence for relationship between genes: Red line - fusion evidence; Purple line - experimental evidence; Black line - co-expression evidence; Yellow line - text mining evidence.

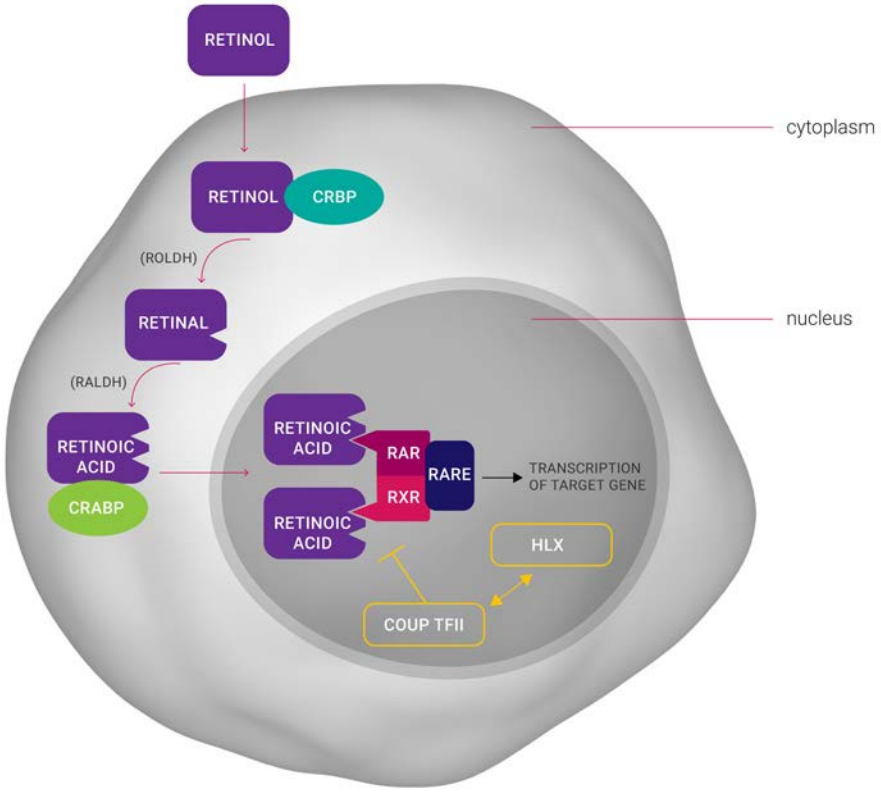


Fig. S5. Retinoic acid (RA) signaling pathway and CDH candidate genes. Retinol is taken up from the blood and bound to CRBP (cellular retinol-binding protein) in the cytoplasm. The retinol is converted to retinal by retinol dehydrogenases (ROLDHs), and then retinal is metabolized to RA by the retinaldehyde dehydrogenases (RALDHs). RA is bound in the cytoplasm by CRABP (cellular RA-binding protein). When RA enters the nucleus, it binds to the RA receptors (RARs) and the retinoid X receptors (RXRs), RARs and RXRs heterodimerize and bind to RA-response element (RAREs), which activates transcription of the target gene. *COUP-TFII* can act as a repressor of this pathway by inhibiting the heterodimerisation of RAR/RXR, thus inhibiting gene transcription. *HLX* was predicted to interact with *COUP-TFII* and could modulate the activity of *COUP-TFII* transcription factor, and therefore disturb this pathway.

Table S1. Clinical information of CDH patients

Patient ID	Ethnicity (self-reported)	Sex	CDH SIDE	CDH TYPE	Complex CDHUnknown	PHENOTYPE
C1	White	M	left	Bochdaleck	NO	isolated CDH
C2	White	M	bilateral	Morgagni	YES	vascular ring, large atrial septal defect, ventriculomegaly, hypospadias, caudal regression, polyda
C3	White	F	right	Bochdaleck	YES	abdominal situs inversus
C4	White	F	left	Bochdaleck	NO	isolated CDH
C5	White	F	right	Morgagni	YES	cleft palate, brachydactyly, dysmorphic features, moderate developmental delay
C6	White	F	left	Bochdaleck	NO	isolated CDH
C7	White	M	left	Bochdaleck	YES	ventricular septal defect, cleft lip and palate, tethered cord, cryptorchidism
C8	White	F	left	Bochdaleck	NO	isolated CDH
C9	White	M	left	Bochdaleck	YES	vascular abnormality of descending aorta
C10	White	M	left	Bochdaleck	NO	isolated CDH
C11	White	F	left	Not otherwise specified	NO	isolated CDH
C12	White	M	left	Bochdaleck	NO	isolated CDH
C13		F	left	Bochdaleck	YES	ventricular septal defect
C14	White	M	left	Bochdaleck	NO	isolated CDH
C15	White	M	left	Bochdaleck	NO	isolated CDH
C16	White	M	left	Bochdaleck	YES	ventricular septal defect, patent foramen ovale
C17	White	F	left	Bochdaleck	NO	isolated CDH
C18	White	M	right	Bochdaleck	NO	isolated CDH
C19	White	M	left	eventration	NO	isolated CDH
C20	White	M	right	Bochdaleck	YES	ventriculomegaly, absence of renal-hepatic inferior vena cava, 2-3 toe syndactyly
C21	White	M	left	Bochdaleck	NO	isolated CDH
C22	White	M	left	Bochdaleck	YES	aortic steNot otherwise specifiedis
C23	White	F	left	Bochdaleck	YES	atrial septal defect
C24	White	M	left	Not otherwise specified	YES	atrial septal defect
C25	White	M	left	Bochdaleck	NO	isolated CDH
C26	Unknown	F	Unknown	Unknown	UNKNOWN	limited phenotype information available
C27	Asian	M	left	Bochdaleck	YES	microtia, thyroglossal duct cyst, kidney abnormality, cryptorchidism
C28	White	M	right	Bochdaleck	YES	hypoplastic right pulmonary artery
C29	White	M	left	Bochdaleck	YES	hydrocephalus, hypotonia, left duplex kidney
C30	White	M	right	Morgagni	NO	isolated CDH
C31	White	F	left	Bochdaleck	YES	renal cystic dysplasia
C32	White	M	left	Bochdaleck	YES	macrocephaly, seizures, hydronephrosis
C33	Black	F	left	Not otherwise specified	YES	autism, intra-uterine growth restriction, post-natal growth failure, Brown's syndrome of right ey
C34	White	M	Unknown	Unknown	NO	isolated CDH
C35	White	M	left	Bochdaleck	NO	isolated CDH
C36	Black	M	anterior	Morgagni	NO	isolated CDH
C37	White	M	left	Bochdaleck	NO	isolated CDH
C38	White	M	left	Bochdaleck	NO	isolated CDH
C39	White	M	left	Bochdaleck	NO	isolated CDH
C40	Hispanic	F	left	Bochdaleck	NO	isolated CDH
C41	Hispanic	M	right	Bochdaleck	NO	isolated CDH
C42	White	F	left	eventration	YES	atrial septal defect, ventricular septal defect, persistent ductus arteriosus
C43	White	M	left	Bochdaleck	YES	ventricular septal defect, persistent ductus arteriosus, ear anomaly, polydactyly
C44	White	F	left	Morgagni	NO	isolated CDH
C45	White	M	left	Bochdaleck	NO	isolated CDH
C46	White	M	left	Bochdaleck	NO	isolated CDH
C47	White	F	left	Not otherwise specified	NO	isolated CDH
C48	Asian	M	left	Bochdaleck	NO	isolated CDH
C49	White	M	left	Bochdaleck	NO	isolated CDH
C50	White	M	left	Not otherwise specified	YES	atrial septal defect, ventricular septal defect, enlarged aortic root
C51	Hispanic	F	left	Not otherwise specified	YES	ventricular septal defect, aberrant right subclavian artery
C52	Black	M	left	agenesis	YES	atrial septal defect, ventricular septal defect, hypoplastic left pulmonary artery
C53	Unknown	M	Unknown	Unknown	UNKNOWN	limited phenotype information available
C54	White	M	left	Bochdaleck	NO	isolated CDH
C55	White	M	right	Morgagni	NO	isolated CDH
C56	Unknown	F	right	Not otherwise specified	YES	spondylocostal dysplasia, horseshoe kidney
C57	Unknown	M	left	Unknown	NO	isolated CDH
C58	White	F	left	Not otherwise specified	YES	right microtia, bilateral hearing loss
C59	White	F	right	agenesis	NO	isolated CDH
C60	White	M	left	Unknown	YES	atrial septal defect, ventricular septal defect
C61	White	F	left	agenesis	YES	coarctation of aorta, hypoplastic genitalia, left arm reduction defect, lumbosacral vertebral anon
C62	White	M	left	Bochdaleck	YES	dysmorphic features, inguinal hernia
C63	Asian	F	left	Bochdaleck	NO	isolated CDH
C64	White	F	left	agenesis	YES	coarctation of aorta, atrial septal defect, hypoplastic right pulmonary artery
C65	White	M	left	Not otherwise specified	NO	isolated CDH
C66	White	F	right	Not otherwise specified	NO	isolated CDH
C67	White	M	left	agenesis	NO	isolated CDH
C68	Hispanic	F	left	lateral	NO	isolated CDH
C69	White	M	left	Unknown	YES	developmental delay, hypotonia, persistent ductus arteriosus, decreased visual acuity, hypertelc
C70	White	M	Unknown	Unknown	YES	ventricular septal defect
C71	White	M	right	eventration	YES	developmental delay, hepatopulmonary fusion, sacral dimple, ankyloglossia
C72	White	M	left	Bochdaleck	YES	coarctation/hypoplastic aortic arch, absent corpus callosum, hypospadias, dysmorphic features

C73	White	M	central	Unknown	NO	isolated CDH
C74	White	F	left	Bochdaleck	YES	dysmorphic features
C75	Unknown	F	left	Bochdaleck	NO	isolated CDH
C76	White	F	left	agenesis	YES	interrupted inferior vena cava with azygous continuation
C77	White	M	left	Bochdaleck	NO	isolated CDH
C78	Unknown	F	left	Bochdaleck	NO	isolated CDH
C79	Unknown	F	left	agenesis	YES	Fryns syndrome (polyhydramnios, hydronephrosis/hyrdoureter, cleft lip and palate, right externa
C80	White	M	right	Bochdaleck	NO	isolated CDH
C81	White	M	left	Bochdaleck	NO	isolated CDH
C82	White	F	left	Bochdaleck	YES	coarctation of aorta, atrial septal defect
C83	Black	M	left	Bochdaleck	NO	isolated CDH
C84	B/AI	F	right	Bochdaleck	NO	isolated CDH
C85	White	F	left	Bochdaleck	NO	isolated CDH
C86	White	M	left	Bochdaleck	YES	seizures, partial agenesis of corpus callosum, dysmorphic features
C87	White	M	central	Morgagni	YES	Pentalogy of Cantrell (ectopia cordis, omphalocele, renal ectopia)
C88	Asian	M	left	Bochdaleck	NO	isolated CDH
C89	White	M	left	Unknown	YES	bilateral sensorineural hearing loss treated with cochlear implants, radioulnar syNot otherwise s
C90	White	M	left	Bochdaleck	NO	isolated CDH
C91	Unknown	M	left	Bochdaleck	NO	isolated CDH
C92	White	M	left	Bochdaleck	NO	isolated CDH
C93	Hispanic	M	left	Bochdaleck	YES	autism
C94	Unknown	M	left	agenesis	NO	isolated CDH
C95	White	F	right	Unknown	YES	ventricular septal defect, mildly dilated aortic root, cholesteatoma, incisional hernia
C96	Unknown	M	Unknown	Unknown	UNKNOWN	limited phenotype information available
C97	White	F	left	Not otherwise specified	YES	atrial septal defect
C98	Unknown	M	Unknown	Unknown	YES	Cornelia de Lange syndrome
C99	White	M	left	Bochdaleck	NO	isolated CDH
C100	Asian	M	left	Not otherwise specified	NO	isolated CDH
C101	White	M	left	lateral	NO	isolated CDH
C102	Unknown	F	Unknown	Unknown	UNKNOWN	limited phenotype information available
C103	Hispanic	M	right	Bochdaleck	NO	isolated CDH
C104	White	F	left	Bochdaleck	NO	isolated CDH
C105	Hispanic	M	Unknown	Unknown	UNKNOWN	limited phenotype information available
C106	White	M	left	Bochdaleck	NO	isolated CDH
C107	Unknown	F	left	Bochdaleck	YES	atrial septal defect, corpus callosum abnormality, severe developmental delay, bilateral club fee
C108	Unknown	M	left	Bochdaleck	YES	multiple minor anomalies
C109	White	F	Unknown	Unknown	UNKNOWN	limited phenotype information available
C110	Unknown	M	Unknown	Unknown	UNKNOWN	limited phenotype information available
C111	Unknown	F	Unknown	eventration	YES	gastroschesis
C112	Unknown				YES	colobomas, dysmorphic features
C113	Hispanic	F	Unknown	Unknown	UNKNOWN	limited phenotype information available
C114	White	F	bilateral	Unknown	NO	isolated CDH
C115	White	F	left	Not otherwise specified	NO	isolated CDH
C116	White	F	right	Morgagni	Yes	atrial septal defect, ligmentous laxity, extra renal pelvis, hypertelorism
C117	White	M	left	Bochdaleck	YES	hypospadias
C118	White	M	left	Unknown	NO	isolated CDH
C119	White	F	Unknown	Unknown	YES	hypoplastic left heart syndrome, stillbirth
C120	White	F	left	Bochdaleck	YES	atrial septal defect, congenital hip dysplasia (severe)
C121	White	F	left	Not otherwise specified	NO	isolated CDH
C122	White	M	left	Morgagni/eventration	Yes	umbilical hernia, bilateral hydronephrosis, multiple dysmorphic features
C123	White	M	left	Not otherwise specified	YES	vesicouretral reflux
C124	White	M	Unknown	Unknown	NO	isolated CDH
C125	White	M	left	agenesis	NO	isolated CDH
C126	White	M	bilateral	Unknown	NO	isolated CDH
C127	White	F	left	other	NO	isolated CDH
C128	White	M	left	Not otherwise specified	YES	laryngeal cleft, cleft palate, microcephaly, hypotonia, severe developmental delay, hypothyroidis
C129	Mid E	M	left	Not otherwise specified	YES	microcephaly, intellectual disability, brain malformation, dysmorphic features
C130	White	M	left	agenesis	NO	isolated CDH
C131	White	M	Unknown	Unknown	YES	microcephaly, coloboma, hyperopia
C132	White	M	left	Morgagni	YES	severe scoliosis and pectus deformity, dysmorphic features
C133	Unknown	M	Unknown	Unknown	YES	Fryns syndrome
C134	Unknown	F	left	Bochdaleck	NO	isolated CDH
C135	White	F	Unknown	Unknown	YES	microcephaly
C136	Unknown	Unknown	Unknown	Unknown	UNKNOWN	limited phenotype information available
C137	Hispanic	F	Unknown	Unknown	YES	Fryns syndrome
C138	Unknown	F	Unknown	Unknown	UNKNOWN	limited phenotype information available
C139	White	F	left	other	NO	isolated CDH
C140	White	F	Unknown	Unknown	UNKNOWN	limited phenotype information available
C141	Unknown	F	Unknown	Unknown	UNKNOWN	limited phenotype information available
C142	Unknown	M	Unknown	Unknown	UNKNOWN	limited phenotype information available
C143	Unknown	F	left	Unknown	NO	isolated CDH
C144	Unknown	M	left	Unknown	UNKNOWN	limited phenotype information available
C145	White	M	bilateral	other (left hernia, right eventrati	YES	anterior body wall defect (diastasis recti, short cleft sternum)
C146	Unknown	F	Unknown	Unknown	UNKNOWN	multiple congenital anomalies, no additional information available
C147	White	M	Unknown	Morgagni	YES	ventricular septal defect, coarctation of aorta, high palate, bilateral inguinal hernias, developmer
C148	White	F	Unknown	Unknown	YES	multiple anomalies (limited phenotype information)
C149	White	F	left	Unknown	NO	isolated CDH
C150	Unknown	Unknown	Unknown	Unknown	UNKNOWN	limited phenotype information available
C151	Unknown	F	Unknown	Unknown	YES	Rett syndrome (confirmed MECP2 mutation), colobomas
C152	White	M	right	Morgagni	YES	microcephaly, siezures, abnormal ears, micropenis, facial dysmorphism, optic nerve hypoplasia
C153	Hispanic	F	left	Bochdaleck	Yes	double outlet right ventricle, multiple dysmorphic features, hypoplastic genetalia, hypoplastic to
C154	White	M	left	Not otherwise specified	YES	hydronephrosis
C155	White	M	Unknown	Unknown	YES	hypospadias, cryptorchidism, Unknownandrogen insensitivity syndrome
C156	Unknown	M	Unknown	Unknown	YES	Fryns syndrome (CDH, TOF, abnormal fingers and toes)
C157	White	M	left	eventration	NO	isolated CDH
C158	Unknown	M	Unknown	Unknown	UNKNOWN	limited phenotype information available
C159	White	M	right	Unknown	NO	isolated CDH
C160	White	M	left	Bochdaleck	NO	isolated CDH
C161	White	F	left	Unknown	NO	isolated CDH
C162	White	M	left	Unknown	NO	isolated CDH

C163	Unknown	F	Unknown	Unknown	UNKNOWN	limited phenotype information available
C164	White	M	left	Bochdaleck	NO	isolated CDH
C165	White	F	left	Bochdaleck	YES	atrial septal defect, patent ductus arteriosus
C166	Unknown	M	Unknown	Unknown	YES	Duane radial ray anomaly
C167	White	M	left	Not otherwise specified	YES	GI anomaly, adrenal insufficiency
C168	Unknown	M	Unknown	Unknown	YES	imperforate anus
C169	White	F	Right	Bochdaleck	Yes	Small atrial septal defect, left pulmonary artery steNot otherwise specifiedis
C170	Unknown	M	left	Unknown	NO	isolated CDH
C171	White	M	left	Unknown	NO	isolated CDH
C172	White	M	right	Morgagni	NO	isolated CDH
C173	White	M	left	Unknown	YES	atrial septal defect
C174	White	F	left	Unknown	NO	isolated CDH
C175	White	M	Unknown	Unknown	UNKNOWN	limited phenotype information available
C176	Hispanic	F	left	Bochdaleck	NO	isolated CDH
C177	Hispanic	F	right	Bochdaleck	NO	isolated CDH
C178	Asian/White	F	right	Bochdaleck	NO	isolated CDH
C179	White	F	left	Bochdaleck	NO	isolated CDH
C180	White	M	right	Not otherwise specified	NO	isolated CDH
C181	White	M	left	Bochdaleck	YES	agenesis corpus callosum, white matter loss, cataract
C182	Black	F	left	eventration	NO	isolated CDH
C183	Unknown	F	left	Bochdaleck	YES	hypoplastic left heart syndrome
C184	White	M	left	Bochdaleck	NO	isolated CDH
C185	Asian	M	right	Bochdaleck	NO	isolated CDH
C186	Black	M	left	Bochdaleck	NO	isolated CDH
C187	White	F	left	agenesis	YES	hypoplastic aortic arch, peristant left superior vena cava, 2 vessel umbilical cord
C188	White	M	left	agenesis	NO	isolated CDH
C189	White	M	right	other	NO	isolated CDH
C190	White	F	left	Bochdaleck	NO	isolated CDH
C191	White	M	right	Not otherwise specified	NO	isolated CDH
C192	White	M	left	Bochdaleck	YES	hepatosplenomegaly, hemolytic anemia
C193	Unknown	M	left	Unknown	NO	isolated CDH
C194	Hispanic	F	left	Bochdaleck	YES	atrial septal defect, pulmonary vein steNot otherwise specifiedis, occipital polymicrogyria
C195	White	F	left	Not otherwise specified	NO	isolated CDH
C196	Unknown	Unknown			UNKNOWN	limited phenotype information available, fetal sample

Table S2. Target genes and genomic regions

1. Previously-published recurrent copy number variations in patients with CDH

Cyto band	Coordinates	Candidate ge	Associated syndrome
1q21.1-q44	chr1:1439825	*CRABP2, PBX1, HMCN1, MYOG*	
2q36-qter	chr2:221,500,	*EPHA4, PAX3*	
4p16.3	chr4:1-23366	*FGFRL1*	Wolf-Hirschorn syndrome
6p25	chr7:1-7100000		
8p23.3-p22	chr8:202262-	*SOX7, GATA4, NEIL2*	
8q22.3-q24.23	chr8:1045080	*ZFPM2*	
11q23.1-q25	chr11:112347	*HYLS1*	
12p13.33-p11.22	chr12:150430-28730836		Pallister-Killian syndrome
13q11-q13.1	chr13:191680	*FGF9*	
13q21.33	chr13:70666491-71594524		
15q25.2	chr15:82635583-85055945		
15q26.1-qter	chr15:891000	*KIF7, NR2F2, ARRDC4, IGF1R, MEF2A*	
16p11.2	chr16:29350831-30332522		

2. Genes identified through human genomic studies on CDH and associated syndromes

Coordinates (hg19)	Candidate ge	Associated syndrome
chr2:169983619-170219122	*LRP2*	Donnai-Barrow syndrome
chr2:189839099-189877472	*COL3A1*	Ehlers Danlos syndrome type IV
chr3:25469754-25639422	*RARB*	CDH with microophthalmia
chr5:36876861-37065921	*NIPBL*	Cornelia de Lange syndrome
chr8:61591324-61780586	*CHD7*	CHARGE syndrome
chr10:123237844-123356159	*FGFR2*	Apert syndrome
chr15:48700503-48937985	*FBN1*	Marfan syndrome
chr11:32409322-32457081	*WT1*	Denys-Drash, Frasier, Meacham syndromes
chr15:74471808-74502046	*STRA6*	Matthew-Wood syndrome
chr17:8377523-8534079	*MYH10*	CDH with brain defect (case report)
chr17:42634812-42638630	*FZD2*	Pentalogy of Cantrell (case report)
chr18:19749404-19782491	*GATA6*	Familial CDH
chr19:39989557-39999121	*DLL3*	Spondylocostal dysostosis
chrX:11129406-11141204	*HCCS*	Linear skin defects with multiple congenital anomalies 1
chrX:68048840-68062006	*EFNB1*	Craniofrontonasal syndrome
chrX:132669776-133119673	*GPC3*	Simpson-Golabi-Behmel syndrome

3. Genomic regions from unpublished studies on patients with CDH from our laboratories*

* preliminary copy number variation data derived from Affymetrix 6.0 SNP arrays and/or preliminary sequencing variation data derived from whole exome sequencing on multiplex families

Cyto band	Coordinates	Candidate gene(s)
1p36.22	chr1: 9,326,658-9,939,698	
3p26.1	chr3:6651128-6653332	
3p22.2	chr3:37982108-37986928	

3q25.1	chr3:151511085-151555731
4q13.3-q22.3	chr4:7439112 *FRAS1*
5p15.2	chr5:12669547-12669547
6p12.3	chr6:49431569-49447295
7p21.1	chr1:16700447-16701844
7q34	chr7:141757591-141784496
7q34	chr7:142825843-142892017
7q35	chr7:145949912-146406228
10p13	chr10:13056263-13058840
10q23.1	chr10:84406349-84432355
12q13.13	chr12:52688684-52782836
14q11.2	chr14:22889777-22978775
16p11.2-p11.1	chr16:34459037-34757071
17p13.2	chr17:3505485-3560005
17q12	chr17:33682510-33758578
17q21.31	chr17:44394412-44752300
17q25.3	chr17:77365534-77389101
18q12.3	chr18:40056569-40057767
19q13.42	chr19:53938333-54015178
20q12	chr20:41178858-41243499
20q13.12	chr20:44350361-44378094
Xp11.23	chrX:47879024-47988177
Xq23	chrX:1147951 *PLS3*

4. Genes that cause diaphragm defects and/or lung hypoplasia in mouse models

Coordinates (hg19)	Candidate gene(s)
chr1:18957500-19075360	*PAX7*
chr2:121493199-121709339	*GLI2*
chr2:220283099-220291461	*DES*
chr3:78646388-79068609	*ROBO1*
chr4:55095264-55164412	*PDGFRA*
chr4:126676418-126849624	*CTBP1*
chr4:144257983-144395718	*GAB1*
chr5:44305097-44388784	*FGF10*
chr5:121398890-121414055	*LOX*
chr5:168088738-168728133	*SLIT3*
chr6:134210259-134216675	*TCF21*
chr7:42000548-42276618	*GLI3*
chr7:116312459-116438440	*MET*
chr7:155595558-155604967	*SHH*
chr8:72109668-72274467	*EYA1*
chr8:72753777-72756731	*MSC*
chr9:14734664-14910993	*FREM1*
chr10:50817141-50873150	*CHAT*

chr10:126676418-126849624	*CTBP2*
chr11:2904448-2906995	*CDKN1C*
chr11:17741110-17743678	*MYOD1*
chr11:65633912-65640405	*EFEMP2*
chr12:114791735-114846247	*TBX5*
chr13:41048131-41185264	*FOXO1*
chr14:23305793-23316803	*MMP14*
chr14:61111417-61116155	*SIX1*
chr14:61176256-61190852	*SIX4*
chr15:56119122-56209329	*NEDD4*
chr16:55513081-55540586	*MMP2*
chr17:38474473-38513895	*RARA*
chr17:46652869-46655743	*HOXB4*
chr19:10764937-10803095	*ILF3*
chr19:12986025-12992335	*DNASE2*
chr20:45523263-45817492	*EYA2*
chr21:38071991-38122510	*SIM2*
chr22:46316248-46373008	*WNT7B*

5. Prioritized candidate genes from analysis of gene expression profiles from mouse embryonic diaphragm, protein-protein interaction analyses with known CDH genes and pathways, and preliminary human genomic data

Coordinates (hg19)	Candidate gene(s)
chr1:23037331-23241823	*EPHB2*
chr1:32479295-32509482	*KHDRBS1*
chr2:11321778-11484711	*ROCK2*
chr2:176964530-176965488	*HOXD12*
chr2:176981492-176984670	*HOXD10*
chr2:176987413-176989645	*HOXD9*
chr2:177016113-177017949	*HOXD4*
chr3:47627378-47823405	*SMARCC1*
chr3:89156674-89531284	*EPHA3*
chr5:92919043-92929786	*NR2F1*
chr5:106712590-107006596	*EFNA5*
chr6:90539619-90584155	*CASP8AP2*
chr6:93949740-94129300	*EPHA7*
chr7:19155091-19157295	*TWIST1*
chr7:27139973-27142394	*HOXA2*
chr7:27180671-27183287	*HOXA5*
chr7:27193338-27196296	*HOXA7*
chr7:27202057-27205149	*HOXA9*
chr7:27220776-27224835	*HOXA11*
chr7:73442427-73484236	*ELN*
chr7:83587659-83824217	*SEMA3A*

chr7:148504464-148581441	*EZH2*
chr7:156797547-156803347	*MNX1*
chr8:28351722-28431785	*FZD3*
chr8:41119476-41166990	*SFRP1*
chr8:49830239-49833999	*SNAI2*
chr10:72972292-73062635	*UNC5B*
chr10:96305574-96361856	*HELLS*
chr10:98757795-98945683	*SLIT1*
chr10:102986733-102988717	*LBX1*
chr11:45950870-46142985	*PHF21A*
chr11:46402334-46405387	*MDK*
chr12:48366748-48398285	*COL2A1*
chr12:66218240-66360071	*HMGA2*
chr12:85674036-85695561	*ALX1*
chr14:24630422-24635774	*IRF9*
chr15:37183222-37393500	*MEIS2*
chr17:7788123-7816075	*CHD3*
chr18:3412072-3458406	*TGIF1*
chr19:4360364-4400565	*SH3GL1*
chr19:4909510-4962165	*UHRF1*
chr19:14491956-14519537	*CD97*
chr19:16435651-16438339	*KLF2*
chr20:22561642-22565101	*FOXA2*
chr20:46286150-46415360	*SULF2*
chr21:36160098-36421595	*RUNX1*
chrX:128580478-128657460	*SMARCA1*

Table S3. Summary of CNVs detected from 1,292 samples

	No. of samples	Gains	Losses
Patients with CDH	196	85	182
CDH parents	109	49	144
Population Controls	987	192	390
Total	1,292	234	437

Table S4. Significant CNVs detected in multiple patients with CDH - with phenotype and inheritance information

Region (hg19)	Size (bp)	Cyto-band	CNV Type	Number of samples affected: Pro-band (n= 196)	Number of samples affected: Pop-ulation (n= 987)	p-value	Description/Gene(s)	Patient/Phenotype	Inheritance
1) CNVs detected in two or more patients but not in population controls									
1:221052740-221054346	1,606	1q41	Gain	5	0	0.00058**	most of exon 1 and part of intron 1 of the ***HLX*** gene, as well as non-coding RNA gene ***HLX-AS1***	Pt. C72: coarctation/hypoplastic aortic arch, absent corpus callosum, hypospadias, dysmorphisms	unknown
								Pt. C41: isolated CDH	unknown
								Pt. C80: isolated CDH	unknown
								Pt. C81: isolated CDH	unknown
								Pt. C131: isolated CDH	unknown
17:34813719-36278623	1,464,904	17q12	Loss	2	0	0.06	***AATF*** , ***ACACA*** , ***DDX52*** , ***DUSP14*** , ***GGNBP2*** , ***HNF1B*** , ***LHX1*** , ***MYO19*** , ***DHRS11*** , ***MRM1*** , ***c17orf78*** , ***PIGW*** , ***SYNRG*** , ***TADA2A*** , and	Pt. C29: hydrocephalus, hypotonia, duplex kidney, pectus excavatum	unknown
								Pt. C161: isolated CDH	unknown
1:249126046-249238916	112,870	1q44	Loss	3	0	0.016*	***ZNF672*** , ***ZNF692*** , and ***PGBD2*** (overlaps terminal region of 1q21.1-q44 duplication)	Pt. C112: colobomas	unknown
								Pt. C139: isolated CDH	*de novo*
								Pt. C151: colobomas, dev delay (pt also has Rett syndrome w/ confirmed MECP2 mutation)	unknown
2) CNVs found in multiple patients with CDH but at frequency higher than ethnically matched control populations									
16:32403182-34759850	2,356,668	16p11.2	Gain	4	4	0.047*	***TP53TG3E, TP53TG3B, TP53TG3F, TP53TG3C***	Pt. C7: VSD, CLP, tethered cord, cryptorchidism	unknown
								Pt. C40 : isolated CDH	*de novo*
								Pt. C156: Fryns syndrome (CDH, TOF, abnormal fingers and toes)	*de novo*
								Pt. C133: possible Fryns syndrome	unknown
4:11942-143314	131,372	4p16.3	Loss	5	1	0.001**	***ZNF595*** and ***ZNF718*** . Overlaps with Wolf-Hirschhorn critical region.	Pt. C123: isolated CDH	unknown
								Pt. C151: Rett syndrome (confirmed MECP2 mutation), coloboma	unknown
								Pt. C139: isolated CDH	inherited
								Pt. C112: colobomas, dysmorphisms	unknown
								Pt. C104: isolated CDH	unknown
5:12674767-12754177	79,410	5p15.2	Loss	3	3	0.036*	***LINC01194*** (non-coding)	Pt. C19: isolated CDH	unknown
								Pt. C34: isolated CDH	unknown
								Pt. C129: consanguineous, microcephaly, MR, brain malform, dysmorphic	unknown

[a] Significance level: * (P<0.05) and ** (P<0.01)

Table S5. Pathways analysis of genes present in the significant CNVs

Pathways and Pathway ID	No. of gene	Percentage (involved gene/total gene)	Fold enrich	P-value	Overlapped Genes
DNA binding/ gene transcription regulation					
GO:0006355	8	19.05	4.47	0.0009	*HLX, LHX1, SIM2, ZNF672, ZNF595, ZNF692, ZNHIT3, ZNF718*
GO:0003700	6	14.29	5.55	0.0027	*LHX1, HNF1B, TADA2A, HNF1B, SIM2, ZNF595*
GO:0006351	7	16.7	3.01	0.0175	*HLX, HNF1B, SIM2, ZNF692, ZNF672, ZNF595, ZNF718*
GO:0003676	5	11.09	3.18	0.0183	*DDX52, ZNF595, ZNF692, ZNF718, ZNF672*
GO:0003677	6	14.29	3.18	0.0272	*HNF1B, SIM2, ZNF692, ZNF672, ZNF595, TADA2A*
Embryonic organ development					
INTERPRO: IPR009057	4	9.52	11.05	0.0046	*HLX, TADA2A, LHX1, HNF1B*
INTERPRO: IPR001356	3	7.14	10.87	0.0278	*HLX; LHX1; HNF1B*
Biotin and lipid metabolic pathways					
GO:0004075	1	0.01	217.87	0.0046	*ACACA*
KEGG:hsa00061	1	0.01	178.95	0.0056	*ACACA*
KEGG:hsa00563	1	0.01	93.05	0.0107	*PIGW*

Table S6. Sequence variants in genes from CNV regions (from 275 patients with CDH studied by WES)

Gene Symbol	Chromosome	Position (hg19)	Variant type	Transcript Variant	Protein Variant	SIFT Function Prediction	PolyPhen-2 Function Prediction	phyloP p-value	dbSNP ID	ExAc control population frequency
HLX*	1	2.21E+08	missense	c.1172C>A	p.T391K	Damaging	Benign	0.145	199521070	0.00001655
AATF	17	35345955	missense	c.1085A>G	p.Y362C	Damaging	Probably Damaging	7.26E-03		0
DDX52	17	35986016	missense	c.1061C>T	p.A354V	Damaging	Probably Damaging	9.68E-07		0.000008237
DDX52	17	35986035	missense	c.1042C>G	p.R348G	Damaging	Possibly Damaging	2.98E-03		0
DDX52	17	36002175	in-frame dup	c.247_249dupAGG	p.R83dup					0.0005354
DDX52	17	36003436	missense	c.14A>G	p.D5G	Damaging	Possibly Damaging	2.62E-03	140497637	0.0007969
GGNBP2	17	34942587	in-frame del	c.1609_1611delAAG	p.K538del			3.37E-05		0.000414
MYO19	17	34854121	missense	c.2146G>A	p.A916T	Damaging	Probably Damaging	1.93E-04	139565052	0
MYO19	17	34861135	splice site	c.1905+1G>A				1.04E-05	200572125	0.0001827
MYO19	17	34871802	missense	c.446A>G	p.Y149C	Damaging	Probably Damaging	3.05E-05	187710120	0.001402
MYO19	17	34871823	missense	c.425C>T	p.S142F	Damaging	Probably Damaging	3.53E-06	375068557	0.00005375
DHRS11	17	34951458	missense	c.205T>A	p.C69S	Damaging	Probably Damaging	2.32E-05	143529065	0.008163
PIGW	17	34893655	missense	c.705C>G	p.H235Q	Damaging	Probably Damaging	5.37E-03	61755368	0.006977
PIGW	17	34893833	missense	c.883C>T	p.R295W	Damaging	Probably Damaging		367592728	0.00006599
SYNRG	17	35902204	missense	c.2835T>G	p.F1024L	Damaging	Probably Damaging			0.0001153
SYNRG	17	35937711	missense	c.590G>A	p.G197D	Damaging	Possibly Damaging	7.11E-06		0.00003319
TADA2A	17	35787059	missense	c.143G>A	p.R48Q	Tolerated	Possibly Damaging	2.87E-05		0.00002537
ZNF692	1	249149756	splice site	c.974+1G>T						0
PGBD2	1	2.49E+08	missense	c.1447T>C	p.Y483H	Damaging	Probably Damaging			0
ZNF595	4	86655	non-coding transcript	G>A			Possibly Damaging			0.000008377
ZNF595	4	86823	non-coding transcript	G>A			Probably Damaging	3.54E-04	373753380	0.0001331
ZNF595	4	87064	non-coding transcript	T>C			Probably Damaging			0

* HLX variant published previously in Longoni et al., Proc Natl Acad Sci USA 2014.Aug 26; 111(34):12450-5 (PMID: 25107291)