Prioritization of Candidate Genes for Congenital Diaphragmatic Hernia in a Critical Region on Chromosome 4p16 using a Machine-Learning Algorithm

Danielle A. Callaway¹ Ian M. Campbell² Samantha R. Stover³ Andres Hernandez-Garcia³
Shalini N. Jhangiani^{3,4} Jaya Punetha³ Ingrid S. Paine³ Jennifer E. Posey³ Donna Muzny^{3,4}
Kevin P. Lally⁵ James R. Lupski^{3,4,6} Chad A. Shaw³ Caraciolo J. Fernandes⁶ Daryl A. Scott^{3,7}

¹Department of Pediatrics, Baylor College of Medicine, Houston, Texas, United States

²Division of Human Genetics, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, United States

³Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas, United States

⁴Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas, United States

⁵Department of Pediatric Surgery, McGovern Medical School at UT Health, Houston, Texas, United States

⁶Division of Neonatology, Department of Pediatrics, Baylor College of Medicine, Houston, Texas, United States

⁷Department of Molecular Physiology and Biophysics, Baylor College of Medicine, Houston, Texas, United States

Address for correspondence: Danielle A. Callaway, MD, PhD, Department of Pediatrics, Baylor College of Medicine, One Baylor Plaza, BCM 30, Houston, TX 77030, United States (e-mail: Danielle.callaway@bcm.edu).

J Pediatr Genet 2018;7:164–173.

Abstract

Keywords

- Wolf–Hirschhorn syndrome
- congenital diaphragmatic hernia
- machine-learning algorithm

Wolf–Hirschhorn syndrome (WHS) is caused by partial deletion of the short arm of chromosome 4 and is characterized by dysmorphic facies, congenital heart defects, intellectual/developmental disability, and increased risk for congenital diaphragmatic hernia (CDH). In this report, we describe a stillborn girl with WHS and a large CDH. A literature review revealed 15 cases of WHS with CDH, which overlap a 2.3-Mb CDH critical region. We applied a machine-learning algorithm that integrates large-scale genomic knowledge to genes within the 4p16.3 CDH critical region and identified *FGFRL1*, *CTBP1*, *NSD2*, *FGFR3*, *CPLX1*, *MAEA*, *CTBP1-AS2*, and *ZNF141* as genes whose haploinsufficiency may contribute to the development of CDH.

Introduction

Wolf–Hirschhorn syndrome (WHS; OMIM 194190) is a contiguous gene deletion syndrome involving a group of genes physically clustered in the short arm of chromosome 4. This clinically recognizable syndrome was described independently by Drs. Ulrich Wolf and Kurt Hirschhorn, and has a minimum birth incidence of 1 in 95,896 and infant mortality

rate of 17%.^{1–3} WHS is characterized by a constellation of dysmorphic facial features, including a “Greek warrior helmet” appearance caused by a combination of a broad, flat nasal bridge and a high forehead, hypertelorism, abnormally formed ears with preauricular tags and pits, a short philtrum, micrognathia, and microcephaly. Individuals with WHS typically have pre- and postnatal growth retardation, developmental delay, hypotonia, variable intellectual disability, and,

received

December 1, 2017

accepted after revision

April 10, 2018

published online

May 30, 2018

Copyright © 2018 by Georg Thieme
Verlag KG, Stuttgart · New York

DOI <https://doi.org/10.1055/s-0038-1655755>.
ISSN 2146-4596.

often, seizures.⁴ Several structural birth defects are also commonly observed in individuals with WHS, including central nervous system anomalies (33%), cleft lip and/or cleft palate (25–50%), eye defects (25–50%), congenital heart defects (50%), scoliosis and kyphosis (60–70%), and genitourinary tract anomalies (25%).⁴ A small subset of individuals with WHS also have congenital diaphragmatic hernia (CDH).^{5–18}

The WHS critical region has been defined as a 1.5- to 1.6-Mb region on chromosome 4p16.3 encompassing a region between ~0.4 and 1.9 Mb from the 4p terminus that includes *LETM1*, *SLBP*, *NSD2* (previously known as *WHSC1*), and *NELFA* (previously known as *WHSC2*).^{17,19–21} However, it remains unclear whether deletion of this region is sufficient to cause CDH and which genes on 4p16.3 potentially contribute to the development of the diaphragm. Here, we present a novel case of CDH associated with WHS, review the literature to define a CDH critical region on chromosome 4p16.3, and use a machine-learning algorithm that integrates large-scale genomic knowledge sources to identify candidate genes within the WHS CDH critical region that may contribute to the development of CDH.

Case Report

Our patient was a stillborn girl who was delivered vaginally after induction of labor for gestational hypertension and polyhydramnios at 36^{6/7} weeks of gestation. This was the first pregnancy of this nonconsanguineous couple, but the mother had a history of two prior first trimester miscarriages. Pregnancy was complicated by advanced maternal age of 38 years and advanced paternal age of 43 years as well as maternal obesity. MaterniT21 cell-free DNA screening did not reveal increased risk for genetic abnormalities. At 32 weeks of gestation, ultrasound examinations revealed intrauterine growth restriction with an estimated fetal weight of 1,235 g, (<3rd percentile), polyhydramnios with an amniotic fluid index of 30+, bilateral cleft lip and palate, a large left-sided CDH with herniation of the stomach, multiple bowel loops, and 20% of the liver, and dextroposition of the heart secondary to the CDH. At delivery, the fetus demonstrated no signs of life and was pronounced stillborn. Fetal autopsy including imaging with computed tomography (CT) and magnetic resonance imaging (MRI) and an external visual evaluation confirmed the in utero findings. The confinement of the herniated organs and their separation from pleural fluid suggested that the CDH was covered by a membranous sac.

A chromosome analysis performed on amniocytes at Baylor Genetics revealed a 46,XX,del(4)(p15.3) chromosomal complement. A chromosomal microarray analysis (CMA Version 8.3.1) revealed a de novo ~18.9-Mb deletion of the distal end of chromosome 4p (minimal deletion chr4:85,743–18,953,893; maximal deletion chr4:1–18,984,868; hg19) consistent with a molecular diagnosis of WHS. A gain on chromosome 15q13.1q13.2 (minimum gain chr15:29,213,743–30,300,265; maximum gain chr15:28,525,505–30,349,558; hg19) involving four genes—*APBA2*, *FAM189A1*, *NSMCE3*,

and *TJP1*—was also identified. This gain is located between recurrent breakpoint 3a (BP3A) and recurrent breakpoint 4 (BP4) and was inherited from the asymptomatic father. None of the genes involved are known to play a role in the development of CDH.

A review of the literature revealed 15 other cases of CDH associated with WHS. Clinical and cytogenetic/molecular data for these individuals are summarized in ▶Table 1 and ▶Fig. 1, respectively. In all cases, the phenotypes of these individuals were consistent with those previously seen in individuals with WHS. The cytogenetic and molecularly defined deletions in these cases overlap an ~2.3-Mb deletion carried by an individual described by Casaccia et al.¹⁵ This CDH critical region contains 47 RefSeq genes.

To identify genes in the critical region that may contribute to the development of CDH, we adapted a machine-learning approach that had previously been developed to identify candidate genes with regard to pathogenicity in epilepsy.²² The core function of the algorithm is to integrate large-scale genomic knowledge to develop a phenotype-specific pathogenicity score to predict which genes may predispose to a specified phenotype. Detailed methods and validation of the approach can be found in the original manuscript.²² A description of how we applied this approach to identify CDH-related genes is described below.

Briefly, 31 training genes known to be associated with CDH and diaphragm development in humans and/or mice (*CHAT*, *DNASE2*, *EFEMP2*, *EFNB1*, *FBN1*, *FGFR1*, *FREM1*, *FZD2*, *GATA4*, *GLI2*, *GLI3*, *HLX*, *HOXB4*, *LOX*, *LRP2*, *MET*, *MSC*, *NIPBL*, *NR2F2*, *PAX3*, *PBX1*, *PDGFRA*, *RARA*, *RARB*, *ROBO1*, *SLIT3*, *SOX7*, *STRA6*, *TCF21*, *WT1*, and *ZFPM2*) were manually curated based on a review of the literature and data obtained from the Mouse Genome Informatics database (<http://www.informatics.jax.org/>).²³ We determined training patterns for these CDH genes based on data from each of the following knowledge sources: Gene Ontology (GO),²⁴ Mouse Genome Informatics (MGI) phenotype annotation,²³ protein–protein integration networks,²⁵ the Kyoto Encyclopedia of Genes and Genomics (KEGG) molecular interaction network data,²⁶ micro RNA (miRNA) targeting,²⁷ GeneAtlas expression distribution,²⁸ transcription factor binding, and epigenetic histone modifications.²⁹ We then compared the patterns of all RefSeq genes to the CDH-specific pattern in each knowledge area and calculated an omnibus score for each RefSeq gene. Leave-one-out cross-validation revealed that our scoring approach identified the CDH training genes more efficiently than random chance (▶Fig. 2).

In reviews authored by Donahoe et al and Kardon et al, we identified 35 CDH-related genes/CDH candidates that were not among the genes selected for use in the training set.^{30,31} As an additional test of the algorithm, we determined the percentiles of these genes compared with all RefSeq genes based on their omnibus scores (▶Supplemental Table S1, available in online version only). The median percentile of these genes was 95.5 (▶Supplemental Fig. S1, available in online version only). All but one gene had percentiles > 50. *DSEL*, which had a percentile of 47, is considered a candidate gene for CDH based on its disruption by a maternally

Table 1 Clinical and molecular data from individuals with WHS and CDH. Individual cases are listed based on the year of their publication. Data presented includes phenotypic information, CDH type, and underlying genetic defects with method of analysis other than karyotype, if applicable. Only cases with phenotype information and/or genetic details are included. CDH and WHS cases not present in the table include Pober et al (2005) with a described deletion at 4p16 without inclusion of further phenotypic or genetic information

Report	Sex	Clinical features	Deletion (hg19)	Genetic origin/size/analysis
This report	F; stillbirth at 36-wk gestation	<ul style="list-style-type: none"> –Dysmorphic facial features, bilateral cleft lip and palate –Incomplete ossification of C3 and C4 vertebral bodies, dysgenesis of the corpus callosum –Left-sided saccular CDH containing 20% liver 	Minimum: chr4:85,743–18,953,893 Maximum: chr4:1–18,984,868	De novo, ~19 Mb Chromosomal microarray
Tautz et al (2010)	F; alive	<ul style="list-style-type: none"> –Dysmorphic facial features –Right-sided double-outlet right ventricle of the Fallot type, with a pulmonary artery stenosis, ventricular septal defect, and a persistent ductus arteriosus –Hypoplasia of the corpus callosum and immature gyration –Left-sided CDH 	chr4:1–9,945,373	De novo, ~9.9 Mb SNP array, FISH
Slavotinek et al (2006)	M	–Left-sided CDH	Minimum:chr4:1–6,762,449 Maximum: chr4:1–7,978,364	De novo, ~8 Mb FISH, array comparative genomic hybridization
Basgul et al (2006)	F; terminated at 27-wk gestation	<ul style="list-style-type: none"> –Dysmorphic facial features, scaphoid abdomen –Fetal hypotonia on ultrasound, cystic dilation of nuchal area –Left-sided CDH 	46,XX, del(4)(p15.2)	De novo
Casaccia et al (2006)	F; alive	<ul style="list-style-type: none"> –Dysmorphic facial features, cleft soft palate –Mildly retarded psychomotor development appreciated with age –Left-sided CDH 	chr4:1–2,336,628 (FISH)	De novo, ~2.3 Mb FISH
Van Bug-genhout et al (2004)	F; alive	<ul style="list-style-type: none"> –Cleft palate –Pulmonary stenosis –Impaired growth parameters appreciated at age 1 –Left-sided CDH 	Minimum:chr4:1–3,679,582 Maximum:chr4:1–4,119,652	Unspecified, ~4.1 Mb FISH, comparative genomic hybridization
van Dooren et al (2004)	M; alive	<ul style="list-style-type: none"> –Unilateral cleft palate, right-sided ear tag –Mildepispadias –Severe epilepsy at 1 year of life –Left-sided CDH 	46,XY,ish del(4)(p16.1)	De novo FISH
Tapper et al (2002)	F; terminated in utero	<ul style="list-style-type: none"> –Bilateral cleft lip, clubbed feet –Increased nuchal translucency, cystic hygroma, cervical hemivertebrae, lumbar meningocele –Ventricular septal defect –Left-sided CDH 	46,XX,der(4)t(4;13)(p16;q32). ish der(4)t(4;13)(p16;q32)(WHS–)	Paternal unbalanced segregation of t(4;13)(p16;q32) FISH
Sergi et al (1998)	F; deceased following birth at 31-wk gestation	<ul style="list-style-type: none"> –Dysmorphic facial features, microcephaly, bilateral cleft palate –Multiple renal cortical cysts –Left-sided CDH 	46,XX,del(4)(p13)	De novo
Howe et al (1996)	M; deceased following birth	<ul style="list-style-type: none"> –Ventricular septal defect, transposition of the great vessels, dilated third cerebral ventricle –CDH, unspecified side 	46,XYdel(4)(p16)	De novo
	F; stillbirth	<ul style="list-style-type: none"> –Dandy–Walker malformation, facial cleft –CDH, unspecified side 	46,XXdel(4)(p16)	De novo

Table 1 (Continued)

Report	Sex	Clinical features	Deletion (hg19)	Genetic origin/size/analysis
Kobori et al (1993)	M; deceased following birth at 27-wk gestation	–Dysmorphic facial features, bilateral cleft lip, unilateral cleft palate, clubbed foot –Undescended testicles, hypospadias, horseshoe kidney –Truncus arteriosus –CDH, unspecified side	46,XY,rec(4)dup(4q)inv(4)(p15.2q25)	Meiotic recombinant chromosome derived from paternal 46,XY,inv(4)(p15.2q25)
Tachdjian et al (1992)	M; terminated at 37-wk gestation	–Dysmorphic facial features, cleft palate –Major hypospadias, anteposed anus –Left-sided CDH	46,XYdel(4)(p16)	De novo
Lazjuk et al (1980)	F; stillborn at 35-wk gestation	–Dysmorphic facial features, exophthalmos, coloboma of the right iris –Dislocation of the hips and bilateral club feet –Left-sided CDH	Not obtained, sibling with 46,XY,del(4)(p15)	Maternal balanced translocation 46,XX,t(4;22)(p15.2;p11)
Lurie et al (1980)	F; deceased following birth	–Dysmorphic facial features, hard palate cleft –Bilateral club feet, hypotonia –Left-sided CDH	46,XX,4p-	Unspecified

Abbreviations: SNP, single nucleotide polymorphism; FISH, fluorescence in situ hybridization.

inherited 2.7-Mb deletion of 18q22.1 that was identified in a patient with a late-presenting, right-sided diaphragmatic hernia and microphthalmia and a p.Met14Ile variant identified in an unrelated individual with a late-presenting, anterior diaphragmatic hernia.³² Overall, this analysis provides additional evidence that the algorithm is able to detect CDH-related/CDH candidate genes more efficiently than random chance.

Having tested the machine-learning algorithm, we then prioritized the RefSeq genes within the CDH critical region of

approximately 2.3 Mb based on the omnibus scores generated by the algorithm. The top eight genes identified—*FGFR3*, *FGFR1*, *CTBP1-AS2*, *NSD2*, *ZNF141*, *MAEA*, *CPLX1*, and *CTBP1*—are described in ►Table 2. All of these genes had omnibus scores whose percentile was > 85 when compared with all RefSeq genes and all genes within the CDH critical region.

Using exome sequencing, we screened a previously described cohort of 68 individuals with CDH for rare (<1% allele frequency), putatively deleterious sequence changes in the seven protein-coding genes in this list, as

Fig. 1 Gene mapping of chromosome deletions in individuals with WHS and CDH reveals a ~2.3-Mb CDH critical region. The maximum deletions identified in individuals with WHS and CDH are represented by blue bars mapped using the UCSC Genome Browser (GRCh37/hg19). The order of their presentation is the same as in ►Table 1. All deletions overlap an ~2.3-Mb deletion carried by an individual described by Casaccia et al¹⁵ This CDH critical region contains 47 RefSeq genes.

Fig. 2 Leave-one-out cross-validation demonstrates that our machine-learning algorithm is able to identify CDH training genes more efficiently than random chance. Each of the 31 CDH-associated training genes was removed from the set and genome-wide pathogenicity scores were recalculated. The colored lines represent the efficiency of each knowledge source to score the training gene when left out. The bold black line represents the mean of all other scores. Each of the knowledge sources, as well as the omnibus score, identified the CDH training genes better than random chance.

previously described.³³ This cohort consists of 39 males and 29 females—29 European Americans, 25 Hispanics, 5 African Americans, 1 Asian, 1 Asian Indian, 1 Filipino, 1 Middle Easterner, 1 European American/Hispanic, and 4 individuals of undeclared ancestry.³³ None of the individuals in this cohort are known to be related and the molecular cause of their CDH has not been determined. All of the rare, putatively deleterious sequence changes identified in *FGFR2*, *FGFRL1*, *NSD2*, and *ZNF141* were missense changes (—Table 3). In all cases in which parental samples were available, these sequence variants were inherited from a parent without CDH and most have been previously described in the ExAC database (<http://exac.broadinstitute.org/>) or in gnomAD (<http://gnomad.broadinstitute.org/>).³⁴ No rare, putatively deleterious sequence changes were identified in *MAEA*, *CPLX1*, or *CTBP1*.

Discussion

Congenital diaphragmatic hernia is a rare but recurrently identified feature of WHS. Our patient is the first identified case of diaphragmatic hernia with a sac in association with WHS. It is likely, possibly through a haploinsufficiency model, that one or more genes mapping to the ~2.3-Mb CDH critical region on chromosome 4p16 are sufficient to cause CDH. Using a machine-learning algorithm, we identified *FGFR3*, *FGFRL1*, *CTBP1-AS2*, *NSD2*, *ZNF141*, *MAEA*, *CPLX1*, and *CTBP1* as genes in this region that may contribute to the development of CDH (—Table 2).

FGFRL1, *CTBP1*, and *NSD2* have been previously suggested as CDH candidate genes. In mouse embryos, *Fgfr1l*, *Ctbp1*, and *Nsd2* transcripts have been detected in the pleuroperitoneal folds (PPF) at E11.5 and E12.5 as well as the developing diaphragm at E16.5.^{16,35–38}

FGFRL1 is thought to act as a decoy receptor that can bind and sequester fibroblast growth factor ligands and may be involved in cell–cell adhesion.^{39,40} Initial studies by Trueb and Taeschler showed that *FGFRL1* expression increased during development, particularly within the diaphragm.⁴¹ Two independently generated knockout models confirmed the importance of *FGFRL1* in the diaphragm as these mice displayed abnormally muscularized diaphragms that resulted in lung hypoinflation and death shortly after birth.^{35,36} These descriptions did not identify any abnormalities in innervation or in myofiber types, although a more recent paper indicates that loss of *FGFRL1* leads to a decrease in slow muscle fibers during late embryological development secondary to apoptosis.⁴² In addition, the expression of *Fgfr1l* is decreased in the nitrofen CDH model of Bochdalek-type CDH.⁴³ Decreased levels of *FGFRL1* during late gestation could contribute to abnormal diaphragm muscularization and eventual hernia formation. Although these studies clearly establish a role for *FGFRL1* in diaphragm development, none have indicated that haploinsufficiency of *FGFRL1* is sufficient to cause diaphragmatic hernia in humans.⁴⁴

CTBP1 is a transcriptional coregulator identified as a component in complexes containing DNA-binding transcription factors that are involved in many different biological pathways such as cell–cell adhesion, apoptosis, tumor suppression, neurodevelopment, myogenesis, and vascularization.⁴⁵ Loss of *Ctbp1* in a mouse knockout model resulted in mice that were 30% smaller at birth and was associated with 23% mortality rate by postnatal day 20 due to unspecified causes. Interestingly, loss of *CTBP1* and *CTBP2* in *Ctbp1*^{−/−}; *Ctbp2*^{+/−} mice resulted in embryonic lethality and defective myofiber formation in the diaphragm.³⁷

NSD2 encodes a histone methyltransferase that is ubiquitously expressed during early development.^{46–48} *NSD2* functions together with developmental transcription factors to repress abnormal transcription. *NSD2*-deficient mice demonstrate WHS-related phenotypes including growth deficiencies, craniofacial defects, and cardiac abnormalities.⁴⁷ In addition, *NSD2* has been associated with various forms of cancer and may contribute to development of cancer in WHS patients.^{46–49} Although there were no diaphragm abnormalities appreciated in *NSD2*-deficient mice, *NSD2*'s function as a transcriptional regulator during early development makes it a strong candidate.

FGFR3, *CPLX1*, *MAEA*, *CTBP1-AS2*, and *ZNF141* have not been previously suggested as possible CDH candidate genes. Although *Fgfr3*, *Cplx1*, and *Maea* are expressed in the PPF at E11.5 and E12.5 and in the developing mouse diaphragm at E16.5, diaphragmatic defects have not been documented in *FGFR3*-, *CPLX1*-, and *MAEA*-deficient mice.^{38,50–53} *CTBP1-AS2* is a noncoding RNA gene located in a head-to-head configuration with *CTBP1*. *ZNF141* encodes a zinc finger-containing

Table 2 Descriptions of top CDH candidate genes located in the CDH critical region on 4p16 as selected by the machine-learning algorithm. Genes from the CDH critical region on 4p16 whose omnibus percentile scores were >85 when compared with all RefSeq genes and all genes within the CDH critical region. Genes are ranked by omnibus score. Data presented includes known or postulated function including human disease associations, mouse phenotype, gene location, and ExAC pLI score

Gene	Omnibus score percentile (RefSeq/CDH critical region)	Function/human disease associations	Knockout mouse phenotypes	Location (hg19)	pLIscore
<i>FGFR3</i> (OMIM no. 134934)	99.9/100	-Cell determination and proliferation, wound healing, angiogenesis, embryo development, ossification -Achondroplasia (OMIM no. 100800), hypochondroplasia (OMIM no. 146000), crouzonodermoskeletal syndrome (OMIM no. 612247)	-Abnormal chondrocytes, decreased bone mineral density ⁵⁰ -Abnormal long bone morphology and structure ^{51,56}	chr4:1,793,299–1,808,872	0
<i>FGFR1</i> (OMIM no. 605830)	98.9/97.9	-Not fully understood, postulated to act as a decoy receptor that can bind and sequester fibroblast growth factor ligands and may be involved in cell–cell adhesion ^{39,40} -WHS (OMIM no. 194190)	-Cranial abnormalities, diaphragmatic hernia, abnormal kidney development, neonatal lethality ³⁶ -Abnormal vascular development, VSD, abnormal cranial morphology, anemia, abnormal diaphragm development, abnormal skeletal morphology ³⁵	chr4:1,011,822–1,026,898	0.58
<i>CTBP1-AS2</i> (C4orf42)	98/95.7	-Unknown	None available	chr4:1,249,300–1,288,291	NA
<i>NSD2</i> (WHSC1; OMIM no. 602952)	97.3/93.6	-Expressed ubiquitously in early development, histone methylation -WHS (OMIM no. 194190)	-ASD, VSD, cleft palate, abnormal sternum ossification ⁴⁷	chr4: 1,871,424–1,982,207	1
<i>ZNF141</i> (OMIM no. 194648)	97/91.5	-Limb development, possibly transcriptional regulation ⁵⁵ -Postaxial polydactyly type A6 (OMIM no. 615226)	None available	chr4: 337,814–384–864	0.16
<i>MAEA</i> (OMIM no. 606801)	91/89.4	-Erythroblast attachment to macrophages, erythroblast maturation	-Increased NK T cell number, abnormal coat appearance (International Knockout Mouse Consortium, mousephenotype.org)	chr4: 1,289,851–1,340,147	0.84
<i>CPLX1</i> (OMIM no. 605032)	89/87.2	-Synaptic vesicle exocytosis -WHS (OMIM no. 194190)	-Seizures, abnormal hearing, abnormal neuron physiology ⁵²	chr4:784,957–826,198	0.8
<i>CTBP1</i> (OMIM no. 602618)	85.6/85.1	-Cell–cell adhesion and apoptosis, tumor suppression, neurodevelopment, and myogenesis and vascularization ⁴⁵ -WHS (OMIM no. 194190)	-Small size and increased mortality ³⁷	chr4:1,211,448–1,249,953	0.98

Abbreviations: OMIM, Online Mendelian Inheritance in Man phenotype number; pLI, probability of loss-of-function intolerance in the ExAC database (<http://exac.broadinstitute.org/>).

Table 3 Rare, putatively deleterious sequencing changes in CDH candidate genes identified in a cohort of 48 individuals with CDH. We screened a cohort of 48 individuals with CDH for rare, putatively deleterious sequencing changes in the top seven protein-coding genes identified by the machine-learning algorithm—*FGFR3*, *FGFRL1*, *NSD2*, *ZNF141*, *MAEA*, *CPLX1*, and *CTBP1*. For each gene, patient identifier information including ethnicity and sex, change in transcript, mode of inheritance, in silico predictions, ExAC/gnomAD information, and clinical description are provided. No rare, putatively deleterious sequence changes were identified in *MAEA*, *CPLX1*, or *CTBP1*

Gene	Identifier, ethnicity, sex	Change [transcript]	Inheritance	In silico predictions	ExAC/gnomAD	Clinical description
<i>FGFR3</i>	TX73, Filipino male	c.985G > A p.Val329Ile [NM_000142.4]	Maternal	CADD = 16.9 SIFT = damaging PolyPhen-2 = probably damaging MT = disease causing	0 homo, 13 het in 120,218 alleles; 0 homo, 29 het in 276,896 alleles	Left-sided CDH, after repair developed a paraesophageal hernia, sister died of CDH
<i>FGFRL1</i>	TX51, Middle Eastern male	c.1474G > A p.Val492Met [NM_001004356]	Unknown	CADD = 21.5 SIFT = tolerated PolyPhen-2 = benign MT = disease causing	22 homo, 532 het in 26,252 alleles; 50 homo, 1,459 het in 193,304 alleles	Agensis of the left diaphragm, consanguinity due to conception with donor egg from father's family, triplet gestation, PFO, PDA, left pulmonary artery hypoplasia, inguinal hernia, dysmorphic features
<i>FGFRL1</i>	TX57, Asian Indian female	c.1474G > A p.Val492Met [NM_001004356]	Paternal	CADD = 21.5 SIFT = tolerated PolyPhen-2 = benign MT = disease causing	22 homo, 532 het in 26,252 alleles; 50 homo, 1,459 het in 193,304 alleles	Agensis of the left diaphragm, PDA, ASD
<i>NSD2</i>	TX12, Northern European male	c.904C > T p.Pro302Ser [NM_133330]	Maternal	CADD = 14.4 SIFT = tolerated PolyPhen-2 = benign MT = disease causing	0 homo, 18 het in 120,646 alleles; 0 homo, 40 het in 275,674 alleles	Left-sided CDH
<i>NSD2</i>	TX13, European male	c.2807T > C p.Val936Ala [NM_133330]	Paternal	CADD = 21.2 SIFT = tolerated PolyPhen-2 = benign MT = disease causing	Not found Not found	Left-sided sac CDH, right-sided eventration
<i>NSD2</i>	TX80, Northern European male	c.229G > A p.Asp77Asn [NM_133330]	Maternal	CADD = 24.3 SIFT = damaging PolyPhen-2 = benign MT = disease causing	0 homo, 12 het in 121,272 alleles; 0 homo, 17 het in 276,304 alleles	Left-sided posterolateral CDH, ASD
<i>ZNF141</i>	TX28, African American female	c.487C > T p.Arg163Cys [NM_003441]	Unknown	CADD = 26.5 SIFT = damaging PolyPhen-2 = possibly damaging MT = disease causing	1 homo, 217 het in 121,106 alleles; 1 homo, 456 het in 276,920 alleles	Agensis of the left diaphragm, large VSD
<i>ZNF141</i>	TX29, Northern European male	c.487C > T p.Arg163Cys [NM_003441]	Unknown	CADD = 26.5 SIFT = damaging PolyPhen-2 = possibly damaging MT = disease causing	1 homo, 217 het in 121,106 alleles; 1 homo, 456 het in 276,920 alleles	Right-sided sac CDH

Abbreviations: ASD, atrial septal defect; CADD, Combined Annotation Dependent Depletion; EEG, electroencephalogram; MT, MutationTaster; PDA, patent ductus arteriosus; PolyPhen-2, HumVar predictions reported; SIFT, Sorting Intolerant From Tolerant.

protein ubiquitously expressed at a low level in all tissues tested.⁵⁴ Postaxial polydactyly in a consanguineous Pakistani family has been attributed to a homozygous c.1420C > T, p. Thr474Ile variant in *ZNF141*.⁵⁵ Mouse models of *ZNF141* deficiency have not been generated. Phenotypes associated with the aforementioned genes such as achondroplasia for *FGFR3* and polydactyly for *ZNF141* were not reported in any of the 15 cases of CDH in WHS. However, some patients did display limb or skeletal defects, including club foot and incomplete ossification of cervical vertebrae.

In all cases where parental samples were available, the rare, putatively deleterious sequence changes identified in *FGFR3*, *FGFRL1*, *NSD2*, *ZNF141*, *MAEA*, *CPLX1*, and *CTBP1* in a cohort of 68 individuals with CDH were found to be inherited from an unaffected parent. In all but one case, these changes were also documented among control individuals in the ExAC database or in gnomAD (►Table 3). While we cannot rule out the possibility that these changes may confer some level of increased risk for the development of CDH, it is unlikely that they are sufficient to cause CDH in isolation.

In leave-one-out cross-validation studies, we have shown that the machine-learning algorithm used to identify *FGFR3*, *FGFRL1*, *CTBP1-AS2*, *NSD2*, *ZNF141*, *MAEA*, *CPLX1*, and *CTBP1* as candidate genes in the WHS CDH critical region is able to identify CDH training genes more efficiently than random chance. We also demonstrated that the algorithm was able to detect 35 CDH-related/CDH candidate genes not included in the training list more efficiently than random chance. This suggests that this machine-learning algorithm could be used in future studies to prioritize CDH candidate genes in other CDH critical regions. It is also possible that this algorithm could be used to prioritize putatively deleterious changes found in individuals with CDH for further analysis based on the likelihood that the gene(s) they affect are CDH-related.

One strength of this machine-learning algorithm is its ability to incorporate data from a wide variety of knowledge sources.²² At the same time, its reliance on previously generated data stored in these sources is one of its limitations. This reliance may bias predictions against CDH genes whose mode of action is unlike those of previously reported CDH genes. A similar bias can also be introduced by the choice of the training genes. We also note that the scores generated for each gene are only as effective as the a priori knowledge for that gene. If little or no information is known about a gene, or if a gene is not annotated in the RefSeq, the algorithm will not be able to accurately calculate a score.

In conclusion, CDH associated with WHS can be attributed to haploinsufficiency of one or more genes located in the ~2.3-Mb critical region on chromosome 4p16.3. Using a machine-learning algorithm trained using a set of CDH-associated genes, we identified *FGFRL1*, *CTBP1*, *NSD2*, *FGFR3*, *CPLX1*, *MAEA*, *CTBP1-AS2*, and *ZNF141* as genes whose haploinsufficiency may contribute to the development of CDH.

Funding

This project was supported by the National Institutes of Health/Eunice Kennedy Shriver National Institute of Child Health and Human Development (R01 HD064667 to DAS),

National Institutes of Health/National Institute of Neurological Disease and Stroke (F30 NS083159 to IMC), the United States National Human Genome Research Institute/National Heart Blood and Lung Institute (UM1 HG006542 to the Baylor-Hopkins Center for Mendelian Genomics), the National Human Genome Research Institute (K08 HG008986 to JEP), and the Ting Tsung and Wei Fong Chao Foundation (Physician-Scientist Award to JEP).

Conflict of Interest

None.

Acknowledgments

We thank the patient and her family for allowing us to present this interesting case.

References

- 1 Wolf U, Reinwein H, Porsch R, Schröter R, Baitsch H. Deficiency on the short arms of a chromosome No. 4 [in German]. *Human-genetik* 1965;1(05):397–413
- 2 Cooper H, Hirschhorn K. Apparent deletion of short arms of one chromosome (4 or 5) in a child with defects of midline fusion. *Mamm Chrom Nwsl* 1961;4(14):14–16
- 3 Hirschhorn K, Cooper HL, Firschein IL. Deletion of short arms of chromosome 4-5 in a child with defects of midline fusion. *Humangenetik* 1965;1(05):479–482
- 4 Battaglia A, Carey JC, South ST. Wolf-Hirschhorn syndrome. In: Pagon RA, Adam MP, Ardinger HH, et al, eds. *GeneReviews*(®). Seattle, WA: University of Washington, Seattle; 2015
- 5 Lazjuk GI, Lurie IW, Ostrowskaja TI, et al. The Wolf-Hirschhorn syndrome. II. Pathologic anatomy. *Clin Genet* 1980;18(01):6–12
- 6 Lurie IW, Lazjuk GI, Ussova YI, Presman EB, Gurevich DB. The Wolf-Hirschhorn syndrome. I. Genetics. *Clin Genet* 1980;17(06):375–384
- 7 Tachdjian G, Fondacci C, Tapia S, Hutten Y, Blot P, Nessmann C. The Wolf-Hirschhorn syndrome in fetuses. *Clin Genet* 1992;42(06):281–287
- 8 Kobori J, Seto-Donlon S, Gregory T, Bangs D, Hsieh C. A case of monosomy 4p and trisomy 4q derived from a meiotic recombination. *Am J Hum Genet Suppl* 1993;55:1578
- 9 Howe DT, Kilby MD, Sirry H, et al. Structural chromosome anomalies in congenital diaphragmatic hernia. *Prenat Diagn* 1996;16(11):1003–1009
- 10 Sergi C, Schulze BR, Hager HD, et al. Wolf-Hirschhorn syndrome: case report and review of the chromosomal aberrations associated with diaphragmatic defects. *Pathologica* 1998;90(03):285–293
- 11 Tapper JK, Zhang S, Harirah HM, et al. Prenatal diagnosis of a fetus with unbalanced translocation (4;13)(p16;q32) with overlapping features of Patau and Wolf-Hirschhorn syndromes. *Fetal Diagn Ther* 2002;17(06):347–351
- 12 van Dooren MF, Brooks AS, Hoozeboom AJM, et al. Early diagnosis of Wolf-Hirschhorn syndrome triggered by a life-threatening event: congenital diaphragmatic hernia. *Am J Med Genet A* 2004;127A(02):194–196
- 13 Pober BR, Lin A, Russell M, et al. Infants with Bochdalek diaphragmatic hernia: sibling precurrence and monozygotic twin discordance in a hospital-based malformation surveillance program. *Am J Med Genet A* 2005;138A(02):81–88
- 14 Basgul A, Kavak ZN, Akman I, Basgul A, Gokaslan H, Elcioglu N. Prenatal diagnosis of Wolf-Hirschhorn syndrome (4p-) in association with congenital diaphragmatic hernia, cystic hygroma and IUGR. *Clin Exp Obstet Gynecol* 2006;33(02):105–106

- 15 Casaccia G, Mobili L, Braguglia A, Santoro F, Bagolan P. Distal 4p microdeletion in a case of Wolf-Hirschhorn syndrome with congenital diaphragmatic hernia. *Birth Defects Res A Clin Mol Teratol* 2006;76(03):210–213
- 16 Tautz J, Veenma D, Eussen B, et al. Congenital diaphragmatic hernia and a complex heart defect in association with Wolf-Hirschhorn syndrome. *Am J Med Genet A* 2010;152A(11):2891–2894
- 17 Van Buggenhout G, Melotte C, Dutta B, et al. Mild Wolf-Hirschhorn syndrome: micro-array CGH analysis of atypical 4p16.3 deletions enables refinement of the genotype-phenotype map. *J Med Genet* 2004;41(09):691–698
- 18 Slavotinek AM, Moshrefi A, Davis R, et al. Array comparative genomic hybridization in patients with congenital diaphragmatic hernia: mapping of four CDH-critical regions and sequencing of candidate genes at 15q26.1–15q26.2. *Eur J Hum Genet* 2006;14(09):999–1008
- 19 South ST, Bleyl SB, Carey JC. Two unique patients with novel microdeletions in 4p16.3 that exclude the WHS critical regions: implications for critical region designation. *Am J Med Genet A* 2007;143A(18):2137–2142
- 20 Hammond P, Hannes F, Suttie M, et al. Fine-grained facial phenotype-genotype analysis in Wolf-Hirschhorn syndrome. *Eur J Hum Genet* 2012;20(01):33–40
- 21 Battaglia A, Carey JC, South ST. Wolf-Hirschhorn syndrome: A review and update. *Am J Med Genet C Semin Med Genet* 2015;169(03):216–223
- 22 Campbell IM, Rao M, Arredondo SD, et al. Fusion of large-scale genomic knowledge and frequency data computationally prioritizes variants in epilepsy. *PLoS Genet* 2013;9(09):e1003797
- 23 Blake JA, Bult CJ, Kadin JA, Richardson JE, Eppig JT; Mouse Genome Database Group. The Mouse Genome Database (MGD): premier model organism resource for mammalian genomics and genetics. *Nucleic Acids Res* 2011;39(Database issue):D842–D848
- 24 Ashburner M, Ball CA, Blake JA, et al; The Gene Ontology Consortium. Gene ontology: tool for the unification of biology. *Nat Genet* 2000;25(01):25–29
- 25 Cowley MJ, Pinese M, Kassahn KS, et al. PINA v2.0: mining interactome modules. *Nucleic Acids Res* 2012;40(Database issue):D862–D865
- 26 Kanehisa M, Goto S, Furumichi M, Tanabe M, Hirakawa M. KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res* 2010;38(Database issue):D355–D360
- 27 Lewis BP, Burge CB, Bartel DP. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* 2005;120(01):15–20
- 28 Su AI, Wiltshire T, Batalov S, et al. A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc Natl Acad Sci U S A* 2004;101(16):6062–6067
- 29 Bernstein BE, Stamatoyannopoulos JA, Costello JF, et al; The NIH Roadmap Epigenomics Mapping Consortium. The NIH Roadmap Epigenomics Mapping Consortium. *Nat Biotechnol* 2010;28(10):1045–1048
- 30 Donahoe PK, Longoni M, High FA. Polygenic causes of congenital diaphragmatic hernia produce common lung pathologies. *Am J Pathol* 2016;186(10):2532–2543
- 31 Kardon G, Ackerman KG, McCulley DJ, et al. Congenital diaphragmatic hernias: from genes to mechanisms to therapies. *Dis Model Mech* 2017;10(08):955–970
- 32 Zayed H, Chao R, Moshrefi A, et al. A maternally inherited chromosome 18q22.1 deletion in a male with late-presenting diaphragmatic hernia and microphthalmia-evaluation of DSEL as a candidate gene for the diaphragmatic defect. *Am J Med Genet A* 2010;152A(04):916–923
- 33 Beck TF, Campeau PM, Jhangiani SN, et al. FBN1 contributing to familial congenital diaphragmatic hernia. *Am J Med Genet A* 2015;167A(04):831–836
- 34 Lek M, Karczewski KJ, Minikel EV, et al; Exome Aggregation Consortium. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 2016;536(7616):285–291
- 35 Catela C, Bilbao-Cortes D, Slonimsky E, Kratsios P, Rosenthal N, Te Welscher P. Multiple congenital malformations of Wolf-Hirschhorn syndrome are recapitulated in Fgfr1 null mice. *Dis Model Mech* 2009;2(5–6):283–294
- 36 Baertschi S, Zhuang L, Trueb B. Mice with a targeted disruption of the Fgfr1 gene die at birth due to alterations in the diaphragm. *FEBS J* 2007;274(23):6241–6253
- 37 Hildebrand JD, Soriano P. Overlapping and unique roles for C-terminal binding protein 1 (CtBP1) and CtBP2 during mouse development. *Mol Cell Biol* 2002;22(15):5296–5307
- 38 Russell MK, Longoni M, Wells J, et al. Congenital diaphragmatic hernia candidate genes derived from embryonic transcriptomes. *Proc Natl Acad Sci U S A* 2012;109(08):2978–2983
- 39 Trueb B, Zhuang L, Taeschler S, Wiedemann M. Characterization of FGFR1, a novel fibroblast growth factor (FGF) receptor preferentially expressed in skeletal tissues. *J Biol Chem* 2003;278(36):33857–33865
- 40 Rieckmann T, Kotevic I, Trueb B. The cell surface receptor FGFR1 forms constitutive dimers that promote cell adhesion. *Exp Cell Res* 2008;314(05):1071–1081
- 41 Trueb B, Taeschler S. Expression of FGFR1, a novel fibroblast growth factor receptor, during embryonic development. *Int J Mol Med* 2006;17(04):617–620
- 42 Amann R, Wyder S, Slavotinek AM, Trueb B. The Fgfr1 receptor is required for development of slow muscle fibers. *Dev Biol* 2014;394(02):228–241
- 43 Dingemann J, Doi T, Rutenstock EM, Puri P. Downregulation of FGFR1 contributes to the development of the diaphragmatic defect in the nitrofen model of congenital diaphragmatic hernia. *Eur J Pediatr Surg* 2011;21(01):46–49
- 44 Lopez Jimenez N, Gerber S, Popovici V, et al. Examination of FGFR1 as a candidate gene for diaphragmatic defects at chromosome 4p16.3 shows that Fgfr1 null mice have reduced expression of Tpm3, sarcomere genes and Lrtm1 in the diaphragm. *Hum Genet* 2010;127(03):325–336
- 45 Stankiewicz TR, Gray JJ, Winter AN, Linseman DA. C-terminal binding proteins: central players in development and disease. *Biomol Concepts* 2014;5(06):489–511
- 46 Stec I, Wright TJ, van Ommen GJ, et al. WHSC1, a 90 kb SET domain-containing gene, expressed in early development and homologous to a Drosophila dysmorphia gene maps in the Wolf-Hirschhorn syndrome critical region and is fused to IgH in t(4;14) multiple myeloma. *Hum Mol Genet* 1998;7(07):1071–1082
- 47 Nimura K, Ura K, Shiratori H, et al. A histone H3 lysine 36 trimethyltransferase links Nkx2-5 to Wolf-Hirschhorn syndrome. *Nature* 2009;460(7252):287–291
- 48 Vougiouklakis T, Hamamoto R, Nakamura Y, Saloura V. The NSD family of protein methyltransferases in human cancer. *Epigenomics* 2015;7(05):863–874
- 49 Ozcan A, Acer H, Ciraci S, et al. Neuroblastoma in a child with Wolf-Hirschhorn syndrome. *J Pediatr Hematol Oncol* 2017;39(04):e224–e226
- 50 Su N, Xu X, Li C, et al. Generation of Fgfr3 conditional knockout mice. *Int J Biol Sci* 2010;6(04):327–332
- 51 Colvin JS, Bohne BA, Harding GW, McEwen DG, Ornitz DM. Skeletal overgrowth and deafness in mice lacking fibroblast growth factor receptor 3. *Nat Genet* 1996;12(04):390–397
- 52 Reim K, Mansour M, Varoqueaux F, et al. Complexins regulate a late step in Ca²⁺-dependent neurotransmitter release. *Cell* 2001;104(01):71–81
- 53 Soni S, Bala S, Gwynn B, Sahr KE, Peters LL, Hanspal M. Absence of erythroblast macrophage protein (Emp) leads to failure of erythroblast nuclear extrusion. *J Biol Chem* 2006;281(29):20181–20189

- 54 Tommerup N, Aagaard L, Lund CL, et al. A zinc-finger gene ZNF141 mapping at 4p16.3/D4S90 is a candidate gene for the Wolf-Hirschhorn (4p-) syndrome. *Hum Mol Genet* 1993;2(10):1571–1575
- 55 Kalsoom UE, Klopocki E, Wasif N, et al. Whole exome sequencing identified a novel zinc-finger gene ZNF141 associated with autosomal recessive postaxial polydactyly type A. *J Med Genet* 2013;50(01):47–53
- 56 Eswarakumar VP, Schlessinger J. Skeletal overgrowth is mediated by deficiency in a specific isoform of fibroblast growth factor receptor 3. *Proc Natl Acad Sci U S A* 2007;104(10):3937–3942