COLD SPRING HARBOR
Molecular Case Studies

RAPID COMMUNICATION

Novel inherited *CDX2* variant segregating in a family with diverse congenital malformations of the genitourinary system

Swetha Ramadesikan,[1] Caitlyn M. Colwell,[1] Rachel Supinger,[1] Jesse Hunter,[1] Jessica Thomas,[2] Elizabeth Varga,[1] Elaine R. Mardis,[1,3] Richard J. Wood,[2,4] and Daniel C. Koboldt[1,3]

[1]Steve and Cindy Rasmussen Institute for Genomic Medicine, [2]Department of Pediatric Colorectal & Pelvic Reconstructive Surgery, Nationwide Children's Hospital, Columbus, Ohio 43205, USA; [3]Department of Pediatrics, [4]Department of Surgery, The Ohio State University College of Medicine, Columbus, Ohio 43210, USA

Abstract Anorectal malformations (ARMs) constitute a group of congenital defects of the gastrointestinal and urogenital systems. They affect males and females, with an estimated worldwide prevalence of 1 in 5000 live births. These malformations are clinically heterogeneous and can be part of a syndromic presentation (syndromic ARM) or as a nonsyndromic entity (nonsyndromic ARM). Despite the well-recognized heritability of nonsyndromic ARM, the genetic etiology in most patients is unknown. In this study, we describe three siblings with diverse congenital anomalies of the genitourinary system, anemia, delayed milestones, and skeletal anomalies. Genome sequencing identified a novel, paternally inherited heterozygous Caudal type Homeobox 2 (*CDX2*) variant (c.722A > G (p.Glu241Gly)), that was present in all three affected siblings. The variant identified in this family is absent from population databases and predicted to be damaging by most in silico pathogenicity tools. So far, only two other reports implicate variants in *CDX2* with ARMs. Remarkably, the individuals described in these studies had similar clinical phenotypes and genetic alterations in *CDX2*. *CDX2* encodes a transcription factor and is considered the master regulator of gastrointestinal development. This variant maps to the homeobox domain of the encoded protein, which is critical for interaction with DNA targets. Our finding provides a potential molecular diagnosis for this family's condition and supports the role of *CDX2* in anorectal anomalies. It also highlights the clinical heterogeneity and variable penetrance of ARM predisposition variants, another well-documented phenomenon. Finally, it underscores the diagnostic utility of genomic profiling of ARMs to identify the genetic etiology of these defects.

[Supplemental material is available for this article.]

Corresponding author: Daniel.Koboldt@nationwidechildrens.org

Ontology terms: abnormality of the rectum; rectoperineal fistula

Published by Cold Spring Harbor Laboratory Press

doi:10.1101/mcs.a006294

CASE PRESENTATION

Our proband (Individual II-1; Fig. 1A), at the time of enrollment, was a 4-yr-old male with perineal fistula, anorectal malformation (ARM), mixed receptive–expressive language disorder, bilateral hydronephrosis with resolved right hydronephrosis and minimal left dilation, and anemia. The proband subsequently underwent posterior sagittal anorectoplasty.

The proband's sister (Individual II-2; Fig. 1A), a 3-yr-old female, presented with symptoms including a cloacal variant (urogenital sinus with perineal fistula), prenatal hydrometrocolpos,

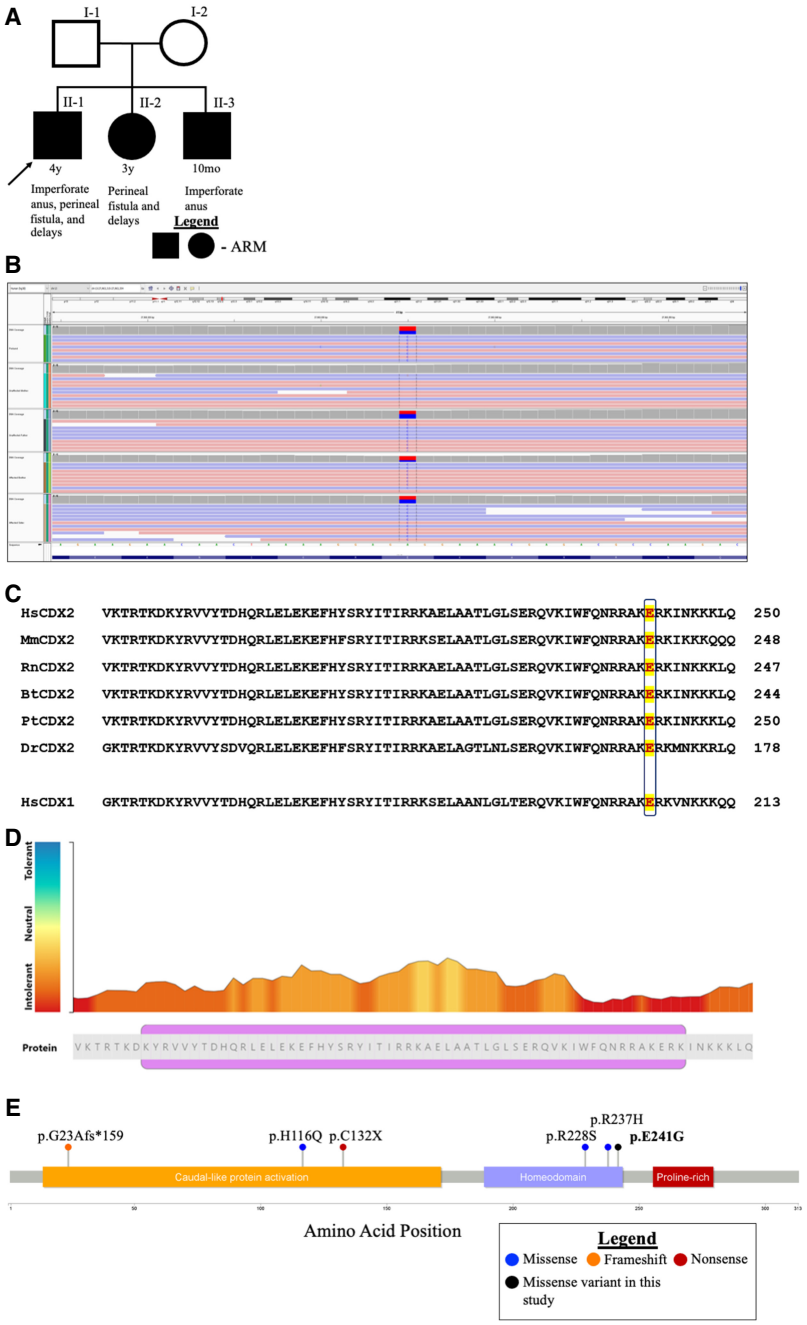

Figure 1. Novel Caudal type Homeobox 2 (CDX2) variant identified by genome sequencing (GS) in a kindred presenting with diverse anorectal malformations (ARMs). (*A*) Pedigree of the family described in the study; the proband is indicated by an arrow. Filled circles and squares indicate affected individuals. (*B*) Aligned GS reads supporting the paternally inherited CDX2 variant as seen in Integrated Genome Viewer (IGV). (*C*) Clustal ω alignment demonstrating the conservation of human CDX2 homeodomain (amino acids 180–250 in hsCDX2) across other vertebrate species and with its paralog, human CDX1. Highlighted in red is the residue affected by the p.Glu241Gly variant. (Hs) *Homo sapiens*, (Mm) *Mus musculus*, (Rn) *Rattus norvegicus*, (Bt) *Bos taurus*, (Pt) *Pan troglodytes*, (Dr) *Danio rerio*. (*D*) Protein intolerance landscape of human CDX2 homeodomain (amino acids 180–250) as predicted by MetaDome. (*E*) Lollipop plot of all reported CDX2 variants (including this study) associated with ARMs.

Table 1. Clinical features of all affected siblings

Clinical phenotype	Individual II-1	Individual II-2	Individual II-3
Perineal fistula	+	+	+
Imperforate anus	+	+	–
Prenatal hydrometrocolpos	N/A	N/A	+
Anteriorly displaced rectum	–	–	+
Shortened perineum	–	–	+
Abnormality of vaginal cavity	N/A	N/A	+
Didelphys uterus	N/A	N/A	+
Bilateral hydronephrosis	+	+	–
Anemia	+	–	+
Developmental delay	–	+	+
Language disorder	+	+	–
Skeletal anomalies	–	+	–

"+" indicates presence, "–" indicates absence of phenotype, "N/A" indicates not applicable.

anteriorly displaced rectum, shortened perineum, two separate vaginal cavities, uterus didelphys, and developmental delay. Given the constellation of symptoms including delays, she had previously been evaluated with a chromosomal microarray (CMA) that was nondiagnostic and later underwent surgery to repair the ARM.

The proband also had a brother (Individual II-3; Fig. 1A), a 10-mo-old male, who also presented with ARM, 13 ribs, mild bilateral hydronephrosis (noted at birth; it has since resolved), speech delay, and a small filum terminale cyst on spine ultrasound. Given the family history of ARMs (the genetics of which are unclear), he was evaluated using exome sequencing (ES) as the standard of care testing. At the time, his ES was reported to be nondiagnostic.

Because of the strong positive family history comprising overlapping caudal anomalies and the siblings' previous nondiagnostic testing, we enrolled the family in a multidisciplinary, IRB-approved translational research study. This study uses genome sequencing (GS) to elucidate the molecular landscape of ARM through comprehensive molecular profiling of patients, family members, and (when available) surgically resected tissue. At that time, parents (Individuals I-1 and I-2) self-reported that they were unaffected by any of the manifestations reported in the children. A full family history and comprehensive clinical history (including extra genitourinary findings) of the affected children was recorded and a family history was taken (Table 1; Fig. 1A). The parents were not clinically evaluated, but according to the expertise of the clinician (R.J.W.), their self-reported status and absence of a surgical history obviate a physical examination.

TECHNICAL ANALYSES AND METHODS

The three children and their parents underwent GS (2 × 150 bp). Library preparation was done using NEBNextUltra II (New England Biolabs) and sequencing was performed using the NovaSeq6000 instrument (Illumina Inc.). Reads were mapped to the GRCh38/UCSC hg38 reference sequence. Secondary data analyses were performed using Churchill (Kelly et al. 2015), which implements the Genome Analysis Toolkit (GATK) best practices workflow to allow for a computationally efficient analysis of GS data. We generated ~141.5 Gbp of uniquely mapped reads per individual, achieving ~35.5× haploid coverage on average. Sequencing metrics are provided in Supplemental Table S1. SnpEff, ANNOVAR, and custom

in-house scripts were used to annotate single-nucleotide polymorphisms (SNPs)/indels with gene, transcript, function class, damaging scores, and population allele frequencies.

Our approach to variant annotation and prioritization has been described previously (Koboldt et al. 2018). Briefly, multisample variant calling was performed with GATK HaplotypeCaller. Common variants (with minor allele frequency [MAF] > 0.001 in the gnomAD and ExAC database) were eliminated and remaining missense, frameshift, nonsense, and splice site variants that were predicted to be damaging by a majority of in silico tools available through VarSome (Kopanos et al. 2019) were prioritized for further analyses. Given the presence of multiple affected offspring, we first searched for variants consistent with a Mendelian inheritance pattern—autosomal recessive or X-linked recessive with expression in carrier females—but no variants in recognized disease genes met such patterns. Next, we identified rare coding variants shared by all three affected children, independent of inheritance. This approach resulted in a shortlist of seven candidate heterozygous gene variants shared among the three affected siblings, all of which were inherited from an apparently unaffected parent (Supplemental Table S2). Of these, only two were associated with human diseases according to the Online Mendelian Inheritance in Man (OMIM) database and were ruled out either because of poor phenotypic fit and/or inconsistent mode of inheritance. Of the remaining five candidates, four genes did not have a known physiological role in caudal/gastrointestinal development. The only remaining candidate was a novel missense variant in the *CDX2* gene. Given the published evidence of the role of *CDX2* in caudal tissue patterning, along with the segregation of this variant in this family, this candidate was pursued further. Incidentally, the male sibling had previously undergone clinical ES testing that, despite reporting the paternally inherited *CDX2* variant, had classified it as a variant of uncertain significance (VUS) and was considered nondiagnostic.

VARIANT CLASSIFICATION

Genome sequencing of the family identified a paternally inherited variant in the gene *CDX2* shared by all three affected children (Table 2; Fig. 1B). The nonsynonymous variant (NM_001265.6:c.722A > G) is predicted to cause a missense change at amino acid codon 241 (p.Glu241Gly) of CDX2. It maps to the well-established homeodomain of CDX2, particularly in the region that binds methylated CpG regions of DNA (Yin et al. 2017) where other *CDX2* variants associated with ARM have been identified (Hsu et al. 2018; Lecoquierre et al. 2020; Stevens et al. 2022). This study reports the third disease-causing variant in residues 228–242, which compose the key motif of the homeobox domain that interacts with 5-mCpG DNA (UniProt). This region is also free of benign variation: There are no missense variants in gnomAD v2 or v3 affecting residues 228–242. Thus, we consider this a well-established functional domain free of benign variation (evidence PM1_strong). The c.722A > G variant is absent from population databases including gnomAD and dbSNP (evidence PM2_supporting) and is predicted to be damaging by a majority of in silico prediction tools including meta predictors such as REVEL (0.86) and SNPred (0.987) (Molotkov I, Artomov M,

Table 2. Genomic findings and variant interpretation

Genomic location	HGVS cDNA	HGVS protein	Zygosity	Origin	Interpretation
Chr 13:27963335	NM_001265.6: c.722A > G	*CDX2*: p.Glu241Gly	Het	Paternal	Likely pathogenic (PM1_Strong, PM2_Supporting, PP3_Moderate)

Genomic coordinates reflect build GRCh38.

under review). Under recently updated guidelines for computational predictions of pathogenicity (Pejaver et al. 2022), this is moderate evidence of pathogenicity (evidence PP3_moderate; refer to Supplemental Table S3 for detailed in silico prediction tool scores and interpretations).

Incomplete penetrance of phenotypes is a well-documented phenomenon in syndromic and nonsyndromic ARM (Schwoebel et al. 1984; Dworschak et al. 2021; Stevens et al. 2022). Although *CDX2* is not currently associated with human disease according to the OMIM database (MIM #600297), there is growing evidence of its critical role in gastrointestinal development (Suh et al. 1994; Beck 2004; Chawengsaksophak et al. 2004; Beck and Stringer 2010). MetaDome, a protein intolerance landscape prediction tool (Wiel et al. 2019), predicts the homeodomain to be highly intolerant to changes at the protein level (Fig. 1D), but globally the *CDX2* gene shows only modest constraint for missense changes ($z = 1.06$), so evidence code PP2 was not applied. Last, multiple recent studies have reported *CDX2* variants in patients with ARMs (Fig. 1E), but this gene is not definitively associated with disease according to OMIM. Therefore, evidence code PP1 (cosegregation of variant with disease in the affected patients in a family in a gene definitively known to cause the disease) was not applied. Altogether, under ACMG guidelines (Richards et al. 2015), the *CDX2* variant was classified as "Likely Pathogenic" (Table 2).

SUMMARY

ARMs constitute a group of congenital anomalies that are estimated to occur in one out of 5000 live births (Levitt and Peña 2007). They encompass diverse anomalies of varying severity, primarily affecting the genitourinary system. However, involvement of other organs including spine (and spinal cord), gastrointestinal tracts, musculoskeletal system (limbs/digits), and orofacial system have also been reported, particularly in syndromic ARM such as VACTERL (Dworschak et al. 2021).

CDX2 is located on Chr 13q.12.3 and encodes a transcription factor (belonging to the *Parahox* family of transcription factors) that is indispensable for caudal development (Suh et al. 1994; Beck 2004; Chawengsaksophak et al. 2004; Beck and Stringer 2010). The encoded protein comprises an amino-terminal caudal-like protein transactivation region that facilitates transcriptional activation, a central homeobox domain that interacts with methylated DNA targets, and terminal poly(Q) and poly(P) repeat motifs (Trinh et al. 1999; Stevens et al. 2022). Functional studies have demonstrated that loss of *CDX2* leads to aberrant transcriptional programming, affecting tissue patterning and differentiation in the gut of the developing embryo, thereby leading to hindgut anomalies (Lorentz et al. 1997; Chawengsaksophak et al. 2004; Tang et al. 2014; Freund et al. 2015). Abnormal expression of *CDX2* has also been documented in colorectal cancer (Salari et al. 2012; Dalerba et al. 2016; Graule et al. 2018).

The novel *CDX2* variant identified in this family (p.Gln241Gly) maps to a conserved methylated DNA-binding motif within the homeobox domain that is critical for its function (Yin et al. 2017). The residue Gln241 is conserved across *CDX2* in higher vertebrates and in human *CDX1*, a paralog of human *CDX2* (Fig. 1C), and is highly intolerant to change (Fig. 1D). So far, five out of six (~83%) *CDX2* variants reported in patients with genitourinary tract defects (including this study) are missense alterations, most of which cluster in the homeobox domain (Fig. 1E). Further, the reported variants in literature have been either de novo or inherited from a mildly affected parent with relatively fewer clinical manifestations (Lecoquierre et al. 2020; Stevens et al. 2022), suggestive of phenotypic variability. Ours is the first study that reports multiple affected patients in a family who harbor a pathogenic *CDX2* variant that was inherited from a father (Individual I-1) who reported no such anomalies. This raises the possibility incomplete penetrance or variable expressivity of the phenotype, a phenomenon

well-documented in this group of disorders (Dworschak et al. 2021). We have not excluded the possibility that the father is affected—as he was not examined by a clinician—but he self-reports as healthy and had no history of surgeries. It is also possible that the father could be mosaic for the *CDX2* variant and may not be expressing the variant in the relevant gastrointestinal tissues. Ultimately, further studies investigating the penetrance and expressivity of this condition are warranted.

All individuals with *CDX2* variants reported to date have presented with diverse malformations of variable severity with occasional multiorgan involvement (Hsu et al. 2018; Lecoquierre et al. 2020; Stevens et al. 2022), suggesting phenotypic variability. Interestingly, two of the three affected children in this study were also reported to have mild developmental delays (Table 1). However, given the limited literature reports of individuals reported to have variants in *CDX2*, it cannot be determined whether the *CDX2*-associated condition also includes developmental delays in its clinical spectrum and warrants further investigation. Nonetheless, many of the individuals exhibited clinical overlap comprising of imperforate anus, cloacal variants, and renal anomalies, suggesting a shared molecular pathogenesis of disease in the genitourinary system. In summary, this study contributes to growing evidence supporting the role of *CDX2* variants as disease-associated in which heterozygous alterations are associated with variable ARM (with phenotypic variability and possible incomplete penetrance of phenotypes).

ADDITIONAL INFORMATION

Data Deposition and Access

The *CDX2* variant and our classification are available on the ClinVar database (https://www.ncbi.nlm.nih.gov/clinvar/) under accession number VCV002429738.1 (NM_001265.6: c.722A > G).

Ethics Statement

Written informed consent was obtained for all participants in this study under a research protocol approved by the Institutional Review Board at Nationwide Children's Hospital (IRB11-00215 Study: Using Genome Sequencing to Identify Causes of Rare Birth Defects and Rare Disorders).

Competing Interest Statement

The authors have declared no competing interest.

Referees

Paul Lasko
Anonymous

Received May 19, 2023; accepted in revised form September 12, 2023.

Acknowledgments

We thank the family for participation in this research.

Author Contributions

All authors contributed to scientific discussion, variant interpretation, and manuscript review.

Funding

This work was supported by The Research Institute at Nationwide Children's Hospital.

REFERENCES

Beck F. 2004. The role of *Cdx* genes in the mammalian gut. *Gut* **53:** 1394–1396. doi:10.1136/gut.2003.038240

Beck F, Stringer EJ. 2010. The role of *Cdx* genes in the gut and in axial development. *Biochem Soc Trans* **38:** 353–357. doi:10.1042/BST0380353

Chawengsaksophak K, de Graaff W, Rossant J, Deschamps J, Beck F. 2004. *Cdx2* is essential for axial elongation in mouse development. *Proc Natl Acad Sci* **101:** 7641–7645. doi:10.1073/pnas.0401654101

Dalerba P, Sahoo D, Paik S, Guo X, Yothers G, Song N, Wilcox-Fogel N, Forgó E, Rajendran PS, Miranda SP, et al. 2016. CDX2 as a prognostic biomarker in stage II and stage III colon cancer. *N Engl J Med* **374:** 211–222. doi:10.1056/NEJMoa1506597

Dworschak GC, van Rooij IALM, Reutter HM. 2021. The role of *de novo* variants in formation of human anorectal malformations. *Genes (Basel)* **12:** 1298. doi:10.3390/genes12091298

Freund JN, Duluc I, Reimund JM, Gross I, Domon-Dell C. 2015. Extending the functions of the homeotic transcription factor Cdx2 in the digestive system through nontranscriptional activities. *World J Gastroenterol* **21:** 1436–1443. doi:10.3748/wjg.v21.i5.1436

Graule J, Uth K, Fischer E, Centeno I, Galván JA, Eichmann M, Rau TT, Langer R, Dawson H, Nitsche U, et al. 2018. CDX2 in colorectal cancer is an independent prognostic factor and regulated by promoter methylation and histone deacetylation in tumors of the serrated pathway. *Clin Epigenetics* **10:** 120. doi:10.1186/s13148-018-0548-2

Hsu JSJ, So M, Tang CSM, Karim A, Porsch RM, Wong C, Yu M, Yeung F, Xia H, Zhang R, et al. 2018. *De novo* mutations in Caudal Type Homeo Box transcription Factor 2 (CDX2) in patients with persistent cloaca. *Hum Mol Genet* **27:** 351–358. doi:10.1093/hmg/ddx406

Kelly BJ, Fitch JR, Hu Y, Corsmeier DJ, Zhong H, Wetzel AN, Nordquist RD, Newsom DL, White P. 2015. Churchill: an ultra-fast, deterministic, highly scalable and balanced parallelization strategy for the discovery of human genetic variation in clinical and population-scale genomics. *Genome Biol* **16:** 6. doi:10.1186/s13059-014-0577-x

Koboldt DC, Mihalic Mosher T, Kelly BJ, Sites E, Bartholomew D, Hickey SE, McBride K, Wilson RK, White P. 2018. In-frame *de novo* mutation in *BICD2* in two patients with muscular atrophy and arthrogryposis. *Cold Spring Harb Mol Case Stud* **4:** a003160. doi:10.1101/mcs.a003160

Kopanos C, Tsiolkas V, Kouris A, Chapple CE, Albarca Aguilera M, Meyer R, Massouras A. 2019. VarSome: the human genomic variant search engine. *Bioinformatics* **35:** 1978–1980. doi:10.1093/bioinformatics/bty897

Lecoquierre F, Brehin AC, Coutant S, Coursimault J, Bazin A, Finck W, Benoist G, Begorre M, Beneteau C, Cailliez D, et al. 2020. Exome sequencing identifies the first genetic determinants of sirenomelia in humans. *Hum Mutat* **41:** 926–933. doi:10.1002/humu.23998

Levitt MA, Peña A. 2007. Anorectal malformations. *Orphanet J Rare Dis* **2:** 33. doi:10.1186/1750-1172-2-33

Lorentz O, Duluc I, Arcangelis AD, Simon-Assmann P, Kedinger M, Freund JN. 1997. Key role of the *Cdx2* homeobox gene in extracellular matrix–mediated intestinal cell differentiation. *J Cell Biol* **139:** 1553–1565. doi:10.1083/jcb.139.6.1553

Pejaver V, Byrne AB, Feng BJ, Pagel KA, Mooney SD, Karchin R, O'Donnell-Luria A, Harrison SM, Tavtigian SV, Greenblatt MS, et al. 2022. Calibration of computational tools for missense variant pathogenicity classification and ClinGen recommendations for PP3/BP4 criteria. *Am J Hum Genet* **109:** 2163–2177. doi:10.1016/j.ajhg.2022.10.013

Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, et al. 2015. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med* **17:** 405–424. doi:10.1038/gim.2015.30

Salari K, Spulak ME, Cuff J, Forster AD, Giacomini CP, Huang S, Ko ME, Lin AY, van de Rijn M, Pollack JR. 2012. *CDX2* is an amplified lineage-survival oncogene in colorectal cancer. *Proc Natl Acad Sci* **109:** E3196–E3205. doi:10.1073/pnas.1206004109

Schwoebel MG, Hirsig J, Schinzel A, Stauffer UG. 1984. Familial incidence of congenital anorectal anomalies. *J Pediatr Surg* **19:** 179–182. doi:10.1016/S0022-3468(84)80444-3

Stevens SJC, Stumpel CTRM, Diderich KEM, van Slegtenhorst MA, Abbott MA, Manning C, Balciuniene J, Pyle LC, Leonard J, Murrell JR, et al. 2022. The broader phenotypic spectrum of congenital caudal abnormalities associated with mutations in the caudal type homeobox 2 gene. *Clin Genet* **101:** 183–189. doi:10.1111/cge.14076

Suh E, Chen L, Taylor J, Traber PG. 1994. A homeodomain protein related to caudal regulates intestine-specific gene transcription. *Mol Cell Biol* **14:** 7340–7351. doi:10.1128/mcb.14.11.7340-7351.1994

Tang XB, Zhang T, Wang WL, Yuan ZW, Bai YZ. 2014. Temporal and spatial expression of caudal-type homeobox gene-2 during hindgut development in rat embryos with ethylenethiourea-induced anorectal malformations. *Cell Tissue Res* **357:** 83–90. doi:10.1007/s00441-014-1858-0

Trinh KY, Jin T, Drucker DJ. 1999. Identification of domains mediating transcriptional activation and cytoplasmic export in the caudal homeobox protein Cdx-3. *J Biol Chem* **274:** 6011–6019. doi:10.1074/jbc.274.9.6011

Wiel L, Baakman C, Gilissen D, Veltman JA, Vriend G, Gilissen C. 2019. MetaDome: pathogenicity analysis of genetic variants through aggregation of homologous human protein domains. *Hum Mutat* **40:** 1030–1038. doi:10.1002/humu.23892

Yin Y, Morgunova E, Jolma A, Kaasinen E, Sahu B, Khund-Sayeed S, Das PK, Kivioja T, Dave K, Zhong F, et al. 2017. Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science* **356:** eaaj2239. doi:10.1126/science.aaj2239