Integrated genetic analysis of transcriptome sequencing data in congenital diaphragmatic hernia

Lu Qiao,[1,2] Haoquan Zhao,[2] Rebecca Hernan,[1] Eve Byington,[2] Guojie Zhong,[2] Julia Wynn,[1] Lan Yu,[1] Xueya Zhou,[1,2] Douglas Potoka,[3] Amy J. Wagner,[4] Brad W. Warner,[5] George B. Mychaliska,[6] Melissa Ellen Danko,[7] Dai Chung,[7] Robert Cusick,[8] Timothy Crombleholme,[9] Foong Yen Lim,[10] Gudrun Aspelund,[11] Vincent P. Duron,[11] Yufeng Shen,[2,12,13,*] and Wendy K. Chung[1,14,*]

[1]Department of Pediatrics, Columbia University Irving Medical Center, New York, NY 10032, USA;[2]Department of Systems Biology, Columbia University Irving Medical Center, New York, NY 10032, USA; [3]Division of Pediatric Surgery, Department of Surgery, University of Pittsburgh, Pittsburgh, PA 15224, USA; [4]Division of Pediatric Surgery, Children's Hospital of Wisconsin, Medical College of Wisconsin, Milwaukee, WI 53226, USA; [5]Division of Pediatric Surgery, Washington University School of Medicine, St. Louis, MO 63110, USA; [6]Section of Pediatric Surgery, Fetal Diagnosis and Treatment Center, University of Michigan Health System, Ann Arbor, MI 48109, USA; [7]Department of Pediatric Surgery, Monroe Carell Jr. Children's Hospital at Vanderbilt, Nashville, TN 37232, USA; [8]Boys Town National Research Hospital, Omaha, NE 68114, USA; [9]Medical City Children's Hospital, Dallas, TX 75230, USA; [10]Division of Pediatric General and Thoracic Surgery, Cincinnati Children's Fetal Care Center, Cincinnati Children's Hospital Medical Center, Cincinnati, OH 45229, USA; [11]Department of Pediatric Surgery, Columbia University Medical Center, New York, NY 10032, USA; [12]Department of Biomedical Informatics, Columbia University Irving Medical Center, New York, NY 10032, USA; [13]JP Sulzberger Columbia Genome Center, Columbia University Irving Medical Center, New York, NY 10032, USA; [14]Department of Medicine, Columbia University Irving Medical Center, New York, NY 10032, USA

*Correspondence: ys2411@cumc.columbia.edu (Y.S.), wkc15@cumc.columbia.edu (W.K.C.).

Abstract:

Congenital diaphragmatic hernia (CDH) is a severe birth defect that is often accompanied by other congenital anomalies. Although the role of genetics in the pathogenesis of CDH has been established and have identified a number of candidate risk genes, the genetic causes of most CDH cases remain unexplained. Here, we performed RNA sequencing (RNA-seq) of diaphragm cells of 153 CDH cases to quantify cis- and trans-effects of putative pathogenic variants and confirm their functional role at the cellular level. Fifty cases (32.6%) have putative pathogenic *de novo* variants. Most genes with likely gene disrupting variants have nonsense mediated decay (NMD) that affects expression levels for that gene. We identified the expression of *ALYREF* and *CSTF2* in CDH subjects were significantly lower and functionally null due to allele-specific expression with nonsense variants. Their inferred targets had altered 3' UTR length within the corresponding carriers. We found the sample from a *HNRNPC* deleterious missense variant carrier had the highest global rate of intron retention among all samples. Meanwhile, the inferred *HNRNPC* targets had altered expression and splicing mode. Further, the potential convergent pathways linked to CDH pathogenesis were mainly involved in DNA replication and

transcription regulation by RNA polymerase II. In conclusion, we demonstrated that RNA-seq data of patient tissues can confirm the functional impact of putative pathogenic variants at the cellular levels and proposed *ALYREF*, *CSTF2* and *HNRNPC* as especially strong candidate CDH genes.

Keywords:

RNA-seq, congenital diaphragmatic hernia, multi-omics study, functional effect of genetic variation