# Congenital diaphragmatic hernia candidate genes derived from embryonic transcriptomes

Meaghan K. Russell[a,1], Mauro Longoni[a,1], Julie Wells[b], Faouzi I. Maalouf[a], Adam A. Tracy[a], Maria Loscertales[a], Kate G. Ackerman[c], Barbara R. Pober[a,d,e], Kasper Lage[a,f,g,h,i], Carol J. Bult[b,2], and Patricia K. Donahoe[a,f,g,2,3]

[a]Pediatric Surgical Research Laboratories, Massachusetts General Hospital, Boston, MA 02114; [b]The Jackson Laboratory, Bar Harbor, ME 04609; [c]Departments of Pediatrics and Biomedical Genetics, School of Medicine and Dentistry, University of Rochester, Rochester, NY 14642; [d]Children's Hospital Boston, Boston, MA 02115; Departments of [e]Pediatrics and [f]Surgery, Harvard Medical School, Boston, MA 02115; [g]Broad Institute, Cambridge, MA 02142; [h]Center for Biological Sequence Analysis, Technical University of Denmark, DK-2800 Lyngby, Denmark; and [i]Center for Protein Research, University of Copenhagen, DK-2200 Copenhagen, Denmark

Contributed by Patricia K. Donahoe, January 6, 2012 (sent for review October 28, 2011)

**Congenital diaphragmatic hernia (CDH) is a common (1 in 3,000 live births) major congenital malformation that results in significant morbidity and mortality. The discovery of CDH loci using standard genetic approaches has been hindered by its genetic heterogeneity. We hypothesized that gene expression profiling of developing embryonic diaphragms would help identify genes likely to be associated with diaphragm defects. We generated a time series of whole-transcriptome expression profiles from laser captured embryonic mouse diaphragms at embryonic day (E)11.5 and E12.5 when experimental perturbations lead to CDH phenotypes, and E16.5 when the diaphragm is fully formed. Gene sets defining biologically relevant pathways and temporal expression trends were identified by using a series of bioinformatic algorithms. These developmental sets were then compared with a manually curated list of genes previously shown to cause diaphragm defects in humans and in mouse models. Our integrative filtering strategy identified 27 candidates for CDH. We examined the diaphragms of knockout mice for one of the candidate genes, pre–B-cell leukemia transcription factor 1 (*Pbx1*), and identified a range of previously undetected diaphragmatic defects. Our study demonstrates the utility of genetic characterization of normal development as an integral part of a disease gene identification and prioritization strategy for CDH, an approach that can be extended to other diseases and developmental anomalies.**

pleuroperitoneal folds | expression profiling | diaphragm development

Congenital diaphragmatic hernia (CDH) encompasses a broad spectrum of developmental defects including complete absence (diaphragmatic aplasia), well-circumscribed defects or "holes" (posterolateral and central CDH), and thinning and/or muscularization defects (eventration and sac type CDH). Despite advances in surgical techniques and neonatal management, limited improvements in survival and long-term morbidity have been realized over the past two decades (1–3). Among CDH cases, 40% have a phenotype that also includes other malformations or associations with genetic syndromes (complex CDH) (2, 4). Such extreme phenotypic variability, likely reflecting multiple etiologies, and limited knowledge of diaphragm developmental processes have slowed the pace of gene discovery (5). Mice, however, provide a surrogate system to elucidate the molecular mechanisms active during diaphragm morphogenesis and those perturbed in CDH (6, 7).

The diaphragm develops between embryonic day (E)10.5 to E15.5 in mice, corresponding to the 4th to 10th weeks of human gestation (6). Among the different components of the primordial diaphragm, the pleuroperitoneal folds (PPFs) are the anlagen of its lateral muscular component where ≈80% of human defects occur (4, 8). PPFs taper medially from the lateral cervical body wall and fuse ventrally with the septum transversum and the posthepatic mesodermal plate (9–12).

Previous whole-transcriptome studies in other organ systems, such as the mouse lung and heart, have demonstrated connections between transcriptional programs in normal development and those perturbed in disease processes (13–15). Similar correlations have been described between diaphragm morphogenesis and CDH only at the level of individual genes (8, 16–18). In this study, we analyzed global trends in gene expression in the PPFs (at E11.5 and E12.5) compared with in the mature diaphragm (at E16.5) by using Short Time-series Expression Miner (STEM) (19) to cluster genes by similarities in temporal expression patterns, as well as Gene Set Enrichment Analysis (GSEA) (20, 21) to assess the expression data in the context of biological processes and signaling pathways. Based on the assumption that CDH candidate genes would have similar temporal expression trends and would be part of the same biological pathways as those already implicated in CDH, we used known CDH-associated genes as "baits" to filter gene sets and to identify novel candidates. One such gene, pre–B-cell leukemia transcription factor 1 (*Pbx1*), validated this strategy by demonstrating previously undetected diaphragm defects in knockout mice.

## Results

**Generation of Expression Data from Laser-Captured Developing Diaphragms.** Transcriptional profiles were generated from laser-capture microdissection (LCM) in C57BL/6J mouse diaphragms at three timepoints during normal embryonic development. Gene expression trends were characterized within the PPF tissue at E11.5 and E12.5 (Fig. 1 *A* and *B*) and within the mature diaphragm at E16.5 (Fig. 1*C*).

There were no statistically significant changes in gene expression between E11.5 and E12.5 (Welch's $t$ test, $P < 0.01$). However, a comparison of early (combined E11.5 and E12.5) versus late (E16.5) diaphragm transcript levels revealed 871 probes corresponding to 770 genes listed in the Mouse Genome Informatics (MGI database; March 2011) that were differentially expressed (Welch's $t$ test; $P < 0.01$) (Dataset S1). Unsupervised hierarchical clustering of the differentially expressed probes resulted in two distinct groups (1), genes up-regulated within the fully muscularized diaphragm and (2) genes up-regulated within the primordial diaphragm; these groups are further described in Fig. S1. Five genes within the second cluster were known to be

Author contributions: M.K.R., M. Longoni, B.R.P., C.J.B., and P.K.D. designed research; M.K.R., M. Longoni, J.W., F.I.M., A.A.T., M. Loscertales, and K.G.A. performed research; M.K.R., M. Longoni, K.L., and C.J.B. analyzed data; and M.K.R., M. Longoni, B.R.P., C.J.B., and P.K.D. wrote the paper.

The authors declare no conflict of interest.

Data deposition: The array data reported in this paper have been deposited in the Gene Expression Omnibus (GEO) database, www.ncbi.nlm.nih.gov/geo (accession no. GSE35243).

[1]M.K.R. and M. Longoni contributed equally to this work.

[2]C.J.B. and P.K.D. contributed equally to this work.

[3]To whom correspondence should be addressed. E-mail: pdonahoe@partners.org.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1121621109/-/DCSupplemental.

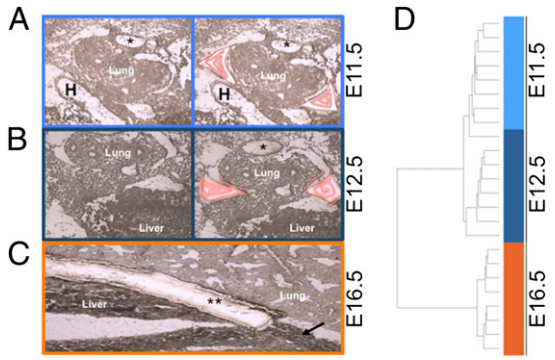


**Fig. 1.** Expression study of microdissected tissue. (*A* and *B*) Unstained and unfixed transverse cryosections of E11.5 (*A*) and E12.5 (*B*) mouse embryo before (*Left*) and after (*Right*) laser capture microdissection (LCM) (H, heart; *, fused dorsal aorta). Captured tissue is outlined in red. (*C*) Sagittal cryosection of an E16.5 mouse embryo showing an anatomically mature diaphragm (arrow) (**, microdissected diaphragm). (*D*) Unsupervised hierarchical clustering of 871 differentially expressed probes (Welch modified *t* test) identifies two distinct clusters [the early (E11.5, E12.5) and late (E16.5)] Additionally, E11.5 and E12.5 separate in two subclusters, except for one outlier.

associated with abnormal diaphragm development in murine models of gene inactivation (*Slit3*, *Pdgfrα*, *Mmp2*, *Gli3*, and *Ilf3*), and two others are orthologs of genes associated with syndromic forms of human CDH (*STRA6* and *GPC3*) (22–30) (cumulative enrichment $P = 2.9 \times 10^{-4}$), suggesting that CDH-associated genes are active during early diaphragm formation.

Genes above the median of normalized hybridization intensities for intronic probes were interpreted as expressed (Dataset S2). It is, however, possible that genes below the thresholds are expressed, but below the detection capacity of the array.

**Temporal Trends in Gene Expression and Molecular Pathways Active During Diaphragm Development.** To identify significant temporal expression trends (profiles) during diaphragm development and the genes associated with each profile, we used the STEM clustering system. STEM analysis revealed five gene expression profiles; three (8, 13, and 12) (Fig. 2*A*) showed upward trends in gene expression at E16.5 when the diaphragm is fully muscularized, whereas two (7 and 2) (Fig. 2*A*) showed highest expression during early development (E11.5 and E12.5). Their functional and phenotypic signatures were analyzed by using Visual Annotation Display (VLAD) to test for enrichment of terms in the Gene Ontology (GO) (31) and Mammalian Phenotype (MP) ontology (32). Genes clustering within STEM profiles 8, 13, and 12 were associated with muscle development and metabolism (GO), and with abnormal muscle and cardiovascular phenotypes (MP). Genes in STEM profiles 7 and 2 were associated with transcriptional regulation and cell proliferation (among others) (GO), and with embryonic lethality, as well as developmental abnormalities of the body wall (among others) (MP) (Dataset S3).

Because the biological significance of gene expression differences can often be more readily appreciated in the context of pathways, we performed GSEA. Seventy-five GO and 33 canonical pathway gene sets correlated with late diaphragmatic development, i.e., members of this list tended to occur among the genes up-regulated at E16.5. The highest ranking sets (with an acceptable false discovery rate of <25%) were associated with structural muscle proteins and mitochondria. One hundred thirteen GO gene sets and 16 canonical pathways correlated with the E11.5 and E12.5 timepoints; among them 12 GO gene sets and 15 canonical pathways had a false discovery rate of <25% (Dataset S4). These pathways include the WNT, TGFβ, ALK, and axonal guidance, all of which are important in mesodermal differentiation, a key component of the early developing diaphragm (33–35).

**Gene Selection and Prioritization.** To prioritize candidate genes from the analysis of diaphragm development transcriptional profiles, we used a strategy based on the hypothesis that novel CDH candidate genes have similar temporal expression trends and participate in the same biological pathways as those already implicated in CDH. Accordingly, all five STEM gene sets were assessed (by hypergeometric distribution after correction for multiple testing) for enrichment of 48 genes previously shown to be associated with mouse models of CDH and/or in human genetic conditions that have CDH as part of their phenotype (Dataset S5). Similarly all 31 biological pathway gene sets obtained by GSEA were also examined for overrepresentation of the same 48 genes.

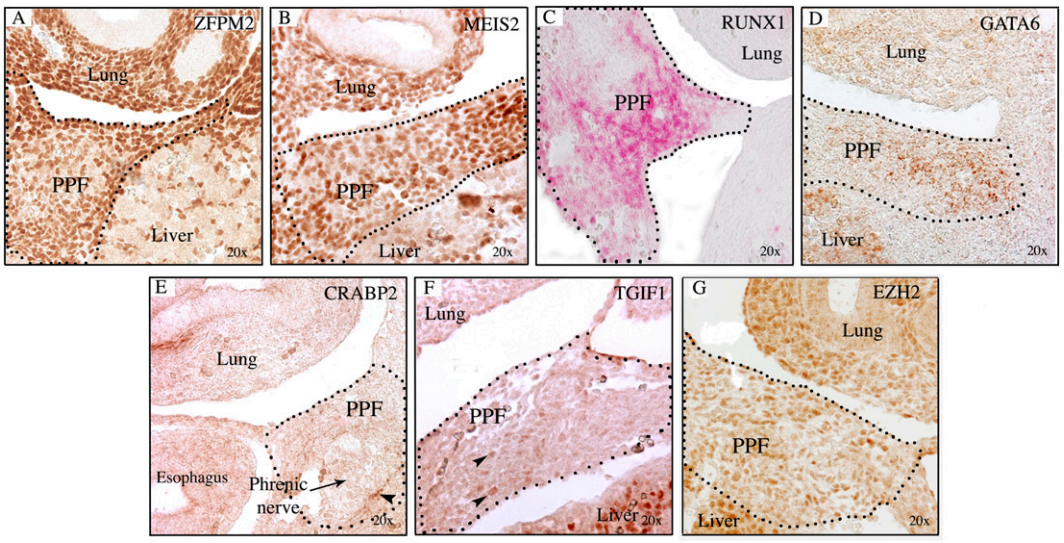


**Fig. 2.** Expression of CDH candidate genes in E12.5 transverse sections. ZFPM2 (*A*) and MEIS2 (*B*) are expressed in diaphragmatic and lung bud mesenchyme. RUNX1 (*C*) and GATA6 (*D*) expression are restricted to the most lateral primordial diaphragmatic mesenchyme and are not expressed in the lung or body wall. CRABP2 (*E*), TGIF1 (*F*), and EZH2 (*G*) are expressed in the PPFs.

**Table 1. Enrichment analysis**

| Profiles and gene sets | Program (database) | *P* value | Overlap with CDH associated genes |
|---|---|---|---|
| Profile 7 | STEM | $4.46 \times 10^{-7}$ | *Gli3, Gpc3, Ilf3, Mmp2, Pdgfra, Robo1, Stra6, Wt1, Zfpm2* |
| Regulation of transcription, DNA dependent | GSEA (CP) | $1.21 \times 10^{-5}$ | *Gli2, Gata4, Zfpm2, Rarb, Nr2f2, Ilf3* |
| Regulation of RNA metabolic process | GSEA (CP) | $1.50 \times 10^{-5}$ | *Gli2, Gata4, Zfpm2, Rarb, Nr2f2, Ilf3* |
| Axon guidance | GSEA (CP) | $2.76 \times 10^{-3}$ | *Efnb1, Slit3, Robo1* |
| Transcription repressor activity | GSEA (CP) | $6 \times 10^{-3}$ | *Zfpm2, Nr2f2, Ilf3* |
| Basal cell carcinoma | GSEA (CP) | 0.035 | *Gli2, Gli3* |

As a result, this filtering strategy revealed a single STEM profile (profile 7) (301 genes) and five GSEA gene sets (totaling 216 genes) (Table 1) that passed the significance threshold. Interestingly, all genes in these gene sets were expressed at highest levels in the primordial diaphragm compared with the mature diaphragm. The top priority CDH candidates were predicted to belong to the intersecting subset of genes that overlapped both biological (GSEA) and temporal (STEM) contexts (Fig. S2 *B* and *C*). This approach resulted in a list of 31 prioritized genes: *Alx1*, *Casp8ap2*, *Chd3*, *Crabp2*, *Efna5*, *Epha3*, *Epha4*, *Epha7*, *Ephb2*, *Ezh2*, *Fzd2*, *Gata6*, *Gli3*, *Hells*, *Hmga2*, *Ilf3*, *Khdrbs1*, *Meil2*, *Pbx1*, *Phf21a*, *Robo1*, *Runx1*, *Sema3a*, *Smarca1*, *Smarcc1*, *Snai2*, *Tgif1*, *Twist1*, *Uhrf1*, *Unc5c*, *Zfpm2*. Among these prioritized genes, 27 were considered candidates as four [*Gli3*, *Ilf3*, *Zfpm2* (*Fog2*), and *Robo1*] had already been causally associated with diaphragm defects (Dataset S5).

**Expression of Prioritized Genes in Primordial Diaphragm Mesenchyme.** To delineate expression patterns and to validate the prioritized genes in the developing diaphragm, we performed immunohistochemistry (IHC) on E11.5 and/or E12.5 PPF tissue of C57BL/6J mouse embryos on seven candidates [*Zfpm2*, *Meis2*, *Runx1*, *Gata6*, *Crabp2*, *Tgif1*, *Ezh2* (Fig. 2), and *Pbx1* (Fig. 3*A*)] for which appropriate antibodies were available and hybridization was successful. ZFPM2 expression was detected throughout the midlung field, as well as the PPF mesenchyme. PBX1 was expressed in the PPFs and other mesenchymal tissues, as was MEIS2, a PBX1 binding partner. RUNX1 expression was detected in the primordial diaphragm just medial to the body wall and in the liver; no expression was seen in the lungs, body wall, or limb buds at this stage. GATA6 expression was restricted to the lateral PPF regions, whereas CRABP2, TGIF1, and EZH2 were expressed diffusely in the PPF. Additionally, expression of *Pbx1* was confirmed by reverse transcription–quantitative PCR (RT-qPCR) (Fig. 3*B*).

***Pbx1* Knockout Mice Have Diaphragmatic Defects.** To determine whether the 27 genes were worthy candidates, *Pbx1* was selected for functional validation of the prioritization strategy. The choice of *Pbx1* was based purely on feasibility, namely embryonic survival up to E15.5; it was not biased by biological preconceptions. Whereas wild-type embryos expressed PBX1 in the mesenchymal tissues of the PPF, lungs, and body wall by IHC analysis (Fig. 3*A*), *Pbx1*$^{-/-}$ embryos (36) did not (Fig. 3*A, Inset*). These mice were generated by insertion of a PGK-neo cassette in exon 3, upstream of the homeodomain (36). Three knockout litters composed of four homozygote mutant embryos, eight heterozygotes, and nine wild types, were harvested at E15.5. Of these embryos, three homozygote mutants were suitable for examination of whole diaphragms; homozygotes, but not heterozygotes, showed diaphragmatic and muscle patterning defects (Fig. 4 *D–F*). Two mutants had left-sided posterolateral defects with an intact membrane but absent musculature. In one, the fundus of the stomach was herniated into the thorax. This mutant also had abnormal muscular patterning with a region of enhanced vascularity adjacent to the circumscribed posterolateral defect (Fig. 3*F*). Expression of muscle markers (PAX3, MYOD, MYOG) was abnormal in E11.5 *Pbx1* mutants (details in Fig. S3).

## Discussion

To determine whether the genes important for development are those perturbed in congenital anomalies affecting the diaphragm, we generated gene expression datasets that represent a global transcriptional survey of genes and biological pathways active during early (E11.5/E12.5) and late (E16.5) diaphragm embryogenesis. The molecular environment of the early diaphragm is characterized by elevated expression of embryonic "master regulator" genes involved in a variety of signaling pathways critical for early developmental processes (i.e., tissue specification and patterning) (37, 38), whereas in the late diaphragm, enhanced expression of genes associated with differentiated skeletal muscle system processes, such as muscle contraction, maintenance, and energy metabolism, was observed.

We identified 27 CDH candidate genes by using a prioritization strategy to filter the transcriptome datasets. Many of the 27 genes play cooperative roles in critical embryonic signaling pathways including retinoic acid (RA) signaling (*Crabp2*, *Tgif1*, *Pbx1*, *Gata6, Ezh2*), WNT canonical signaling (*Fzd2*, *Pbx1*, *Gata6*), TGFβ signaling (*Tgfβ2*, *Tgif1*, *Pbx1*), and Ephrin signaling/axon guidance (*Epha3*, *Efna5*, *Epha4*, *Epha7*, *Ephb2*, *Sema3a*, *Unc5c*) (39–47).

The human orthologs of these 27 genes do not map to previously identified CDH "hotspots" (i.e., recurring deletions or duplications that confer risk for CDH) (5), although a cytogenetic duplication that potentially involves the *PBX1* locus may confer risk for CDH (48). Taken as a whole, these observations suggest our approach to gene discovery is complementary to genomic studies.

Among the 27 candidates, *Pbx1* was selected to validate the prioritization strategy. Examination of *Pbx1* knockout embryos at E15.5 revealed a range of diaphragmatic muscularization and tissue patterning defects (Fig. 4). *Pbx1* encodes a three amino acid loop extension (TALE) homeodomain protein that functions as a transcriptional cofactor central to numerous regulatory networks that modulate cell specification, segmental patterning of gene expression, and mesenchymal precursor cell migration (49, 50).

PBX1 is known to interact with HOXA5 and HOXB5 (among others) (51, 52), which were identified as being expressed in the primordial diaphragm (as per STEM and GSEA). Because PBX1 directs local RA production by transcriptional regulation of *Aldh1a2* in murine hindbrain mesoderm (46), we extrapolate that loss of PBX1 in the primordial diaphragm may lead to decreased RA. This decrease in RA, in turn, may affect transcription the RA regulated *Hoxa5* and/or *Hoxb5* genes (53, 54) (Fig. 4). Furthermore, rodents treated with the RA inhibitor nitrofen, a common model of CDH-associated pulmonary hypoplasia, showed abnormal expression patterns of HOXA5 in the developing lung buds (55).

In addition to RA, other signaling pathways, such as SHH, WNT, and TGFβ (49, 50, 56, 57), are considered important for proper diaphragm formation (42), because genes involved in the SHH pathway (e.g., *Gli2*, *Gli3*), the WNT pathway (e.g., *Ctbp1*, *Ctbp2*), and the TGFβ pathway (e.g., *Fbn1*) are associated with

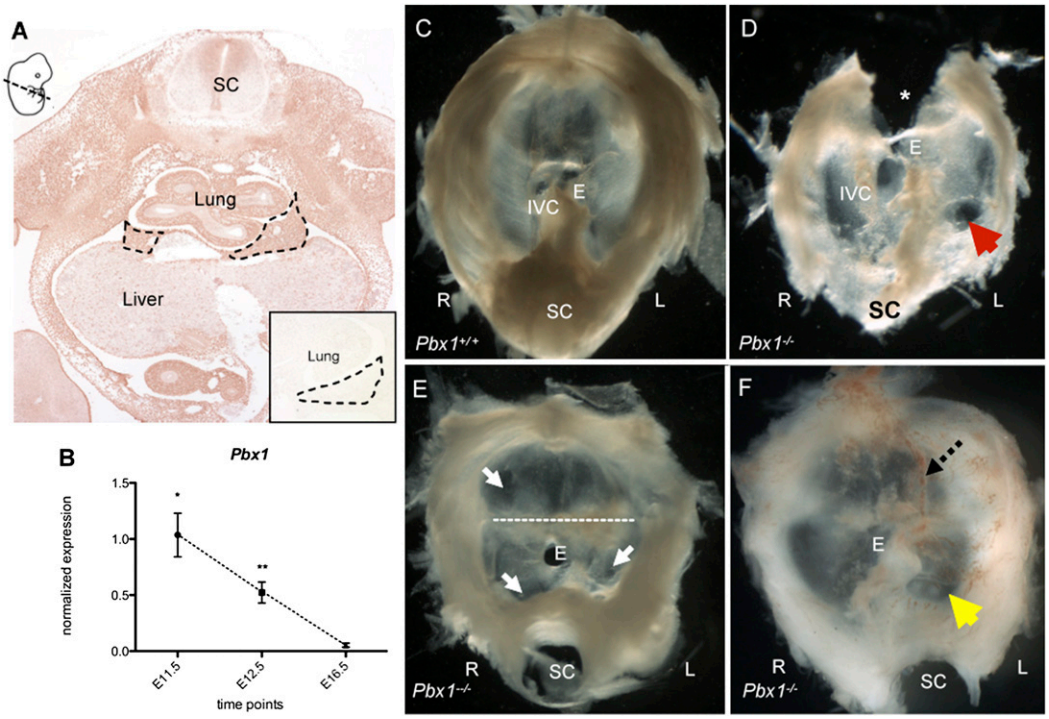


**Fig. 3.** PBX1 expression and characterization of *Pbx1* null mutant diaphragms. (*A*) IHC labeling of PBX1 shows expression in the E11.5 primordial diaphragm. The cartoon (*Upper Left*) indicates the plane of section. (*Inset*) PBX1 is not expressed in the mutant mouse PPF. (*B*) RT-qPCR validation of *Pbx1* expression levels indicate significant down-regulation (±SEM) E11.5–E16.5 (*) and E12.5–E16.5 (**) ($P < 0.05$, two-tailed Welch *t* test). (*C–F*) Dissected diaphragms of E15.5 *Pbx1* mutants (*D–F*) and control embryo (*C*) viewed from the abdomen. (*D*) Circumscribed muscularization defect of the left posterolateral aspect of the diaphragm (indicated by a red arrow) and diffuse decreased muscularization in the lateral regions bilaterally. *, Due to abnormal fusion of the heart to the anterior aspect of diaphragm and thoracic wall, a section of diaphragm was removed during heart dissection. (*E*) Multiple areas of abnormal muscle patterning in the anterior and posterior regions of the diaphragm (white arrows), with an abnormal band of tissue (dotted line). We observed reduced musculature in the crural region, without an independent opening for the IVC. (*F*) Herniated stomach fundus (yellow arrow). Abnormally increased vascularity (dotted arrow). E, esophagus; IVC, inferior vena cava; SC, spinal cord.

CDH phenotypes in knockout mice and/or in humans (24, 58, 59). PBX1 has been shown to influence SHH signaling during early embryonic patterning of the limb and neural tube with its MEIS and PKNOX1/HOX binding partners via *Gli3* (60, 61). Furthermore, PBX1, in a complex with TGIF1/PKNOX1 and SMAD proteins, modulates transcription of TGFβ pathway target genes in pituitary cells (56), and also up-regulates genes involved in WNT signaling via interaction with PKNOX1/MEIS2 proteins in murine midface ectoderm (62). Expression of these PBX1 interactors was identified in the primordial diaphragm (as per STEM and GSEA). Therefore, we surmise that abnormal *Pbx1*-mediated regulation of transcription factors involved in one or several diaphragm-associated signaling pathway(s) (RA, SHH, WNT, and/or TGFβ) may lead to the tissue patterning and muscularization defects observed in diaphragms of E15.5 knockout mice. Abnormal expression patterns of MYOD1 in the developing diaphragm anlagen of *Pbx1* mutants at E11.5 may result from impaired initiation of myogenesis. The homeodomain proteins PBX and MEIS have been shown to be necessary for MYOD1 binding of the *Myog* promoter and, possibly, other downstream targets, thus promoting the myogenic fate (63–65).

Among the 27 candidates, 8 are associated with human or mouse phenotypes, in which CDH was not among the spectrum of defects according to MGI or OMIM. The failure to detect CDH does not negate our prioritization approach, because CDH may be a subtle defect only visible upon careful scrutiny or occur as a low penetrant component of the overall phenotype. Additionally, mutations in some of the prioritized genes may not be compatible with life.

In conclusion, the unbiased transcriptome data generated herein have significantly expanded the number of high quality candidates for CDH. These candidates warrant follow-up in resequencing studies of human CDH cohorts, where risk alleles

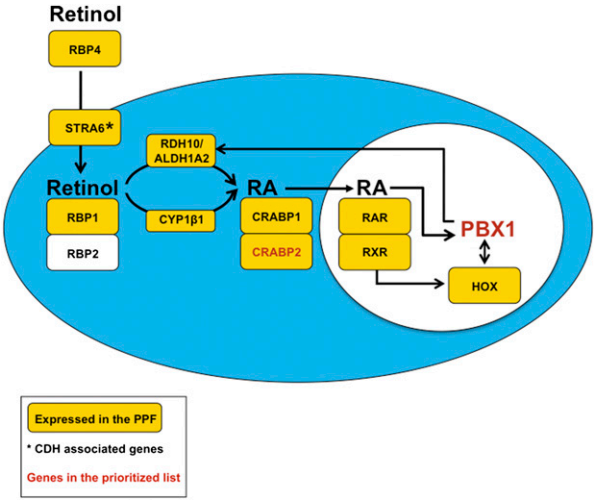


**Fig. 4.** PBX1 transcriptional regulation in the PPFs. PBX1 may function downstream of the RA pathway and act via a feedback loop involving *Aldh1a2* (46) (details in the text). Yellow boxes represent genes in significant STEM and/or GSEA gene sets; *, CDH-associated genes; red letters, members of the prioritized list of candidates.

may be detected by mutational burden tests in case-control studies. The diaphragmatic defects identified in *Pbx1* knockout mice provide proof of principle that prioritization strategies are capable of predicting CDH-associated genes and illustrate the power of integrating normal developmental expression data from the mouse with previous knowledge of mouse and human disease-causing mutations. Furthermore, genes involved in the causation of severe diaphragm anomalies, such as CDH, become candidates for causing less severe diaphragm-related phenotypes, for example gastroesophageal reflux. We predict that this approach can be generalized to a wide array of congenital anomalies that exhibit genetic heterogeneity.

## Materials and Methods

**Animals.** Tissue collections for LCM, IHC, and for diaphragm examination were performed as detailed in *SI Materials and Methods*.

**LCM.** LCM, RNA extraction, purification, and amplification were performed as detailed in *SI Materials and Methods*.

**Expression Arrays.** Gene expression experiments were performed on Affymetrix Mouse Gene 1.0_ST microarrays. The RNA from multiple laser captured tissues was pooled for each embryo, but each embryo was analyzed separately. Hybridization was performed according to the manufacturer's instructions.

**Data Analyses.** The microarray data from E11.5, E12.5, and E16.5 embryos were normalized in the R statistical programming environment (version 2.9.2) by using Robust Multichip Averaging [RMA; ref. 66; implemented in the Affymetrix package for R (http://bioconductor.org/packages/2.1/bioc/html/affy.html)]. Differentially expressed genes were identified by using Welch's *t* test, assuming unequal variances, with a Bonferroni correction. Significance was set at $P < 0.01$. To determine genes expressed at each time-point, the median of the normalized hybridization intensities for 5,222 probes designed to correspond to introns was calculated by using an R script and used as a threshold (for E11.5 and E12.5, the median was 6.3; for E16.5, the median was 6.2). Groups of genes with similar temporal expression patterns were identified by using STEM (http://gene.ml.cmu.edu/stem/) (19), a modified *k*-means clustering algorithm (a method which partitions data in a predefined number of groups, based on their mean) (67). In GSEA, a continuous phenotype label and a Pearson metric was applied to rank genes and gene sets enriched at a nominal $P < 0.05$ and a false discovery rate <0.25, as discussed by Subramanian et al. (20).

**Candidate Gene Prioritization.** A list of 48 known diaphragm defect genes (Dataset S5) was generated by manually curating data from several sources including PubMed; Online Mendelian Inheritance in Man, OMIM, McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore), and the National Center for Biotechnology Information, National Library of Medicine (Bethesda, MD) [last accession November 2010]; and the Mouse Genome Informatics (MGI) database, The Jackson Laboratory [last accession November 2010]. STEM and GSEA both partition the expression data into subsets of genes that participate in a common pathway (GSEA) or specific temporal expression profiles (STEM). To identify the subsets most likely to contain unique CDH genes, we reasoned that unique CDH causing genes are likely to belong to the pathways or expression profiles most enriched for known disease genes, used as "baits". We measured the representation of CDH-associated genes in the different GSEA pathways, or STEM-derived expression datasets, using a hypergeometric distribution (a probability test describing the number of successes in a number of draws from a finite population) (68), and considered only gene sets that were significantly enriched ($P < 0.05$ after adjustment for multiple hypothesis testing). For example, if an enriched GSEA-defined pathway contained 10 genes of which 8 were known to cause CDH, the two remaining genes would be likely to be involved in CDH themselves. Although some genes occur in more then one pathway, each enrichment calculation is independent, thus preventing bias.

**RT-qPCR.** qPCR experiments and data normalization were performed as detailed in *SI Materials and Methods*.

**IHC.** IHC was performed as described (69). Sections were visualized with the ABComplex/HRP Vectastain Detection System according to manufacturer's protocols (Vector Laboratories) after 3,3′-diaminobenzidine (DAB) enhancement (Sigma) (CRABP2, EZH2, GATA6, MEIS2, MYOD1, MYOG, PAX3, PBX1b, TGIF1, ZFPM2), or by the VECTASTAIN ABC-AP KIT (AP-1000) (Vector Laboratories) (RUNX1). The combinations of primary and biotinylated secondary antibodies are listed in Table S1.

**Note.** During the preparation of the manuscript, another gene among the novel candidates (*Fzd2*) was implicated as causal of CDH (47), further substantiating the effectiveness of our gene discovery approach.

**ACKNOWLEDGMENTS.** We thank the Jackson Laboratory Scientific Services for valuable expertise and Dr. Jim Denegre, Dr. Eric Antoniou, Sonya Kamdar, Karen Moore, Ellen Akeson, and Dr. Leslie Bechtold. *Pbx1* knockout embryos were provided by Dr. Michael Cleary and Carmencita Nicholas (Stanford University). We also thank Mark Daly and David MacLaughlin for helpful suggestions. This work was submitted by M.K.R. in partial fulfillment for her doctoral degree.

1. Stege G, Fenton A, Jaffray B (2003) Nihilism in the 1990s: The true mortality of congenital diaphragmatic hernia. *Pediatrics* 112:532–535.
2. Pober BR (2007) Overview of epidemiology, genetics, birth defects, and chromosome abnormalities associated with CDH. *Am J Med Genet C Semin Med Genet* 145C: 158–171.
3. Dott MM, Wong LY, Rasmussen SA (2003) Population-based study of congenital diaphragmatic hernia: Risk factors and survival in Metropolitan Atlanta, 1968-1999. *Birth Defects Res A Clin Mol Teratol* 67:261–267.
4. Pober BR, Russell MK, Ackerman KG (2010) Congenital diaphragmatic hernia overview. *GeneReviews*, eds Pagon RABT, Dolan CR, Stephens K (University of Washington, Seattle).
5. Holder AM, et al. (2007) Genetic factors in congenital diaphragmatic hernia. *Am J Hum Genet* 80:825–845.
6. Ackerman KG, Greer JJ (2007) Development of the diaphragm and genetic mouse models of diaphragmatic defects. *Am J Med Genet C Semin Med Genet* 145C:109–116.
7. Fisher JC, Bodenstein L (2006) Computer simulation analysis of normal and abnormal development of the mammalian diaphragm. *Theor Biol Med Model* 3:9.
8. Babiuk RP, Zhang W, Clugston R, Allan DW, Greer JJ (2003) Embryological origins and development of the rat diaphragm. *J Comp Neurol* 455:477–487.
9. Greer JJ, Allan DW, Martin-Caraballo M, Lemke RP (1999) An overview of phrenic nerve and diaphragm muscle development in the perinatal rat. *J Appl Physiol* 86:779–786.
10. Greer JJ, et al. (2000) Structure of the primordial diaphragm and defects associated with nitrofen-induced CDH. *J Appl Physiol* 89:2123–2129.
11. Mayer S, Metzger R, Kluth D (2011) The embryology of the diaphragm. *Semin Pediatr Surg* 20:161–169.
12. Iritani I (1984) Experimental study on embryogenesis of congenital diaphragmatic hernia. *Anat Embryol (Berl)* 169:133–139.
13. Kho AT, et al. (2004) Conserved mechanisms across development and tumorigenesis revealed by a mouse development perspective of human cancers. *Genes Dev* 18: 629–640.
14. Naxerova K, et al. (2008) Analysis of gene expression in a developmental context emphasizes distinct biological leitmotifs in human cancers. *Genome Biol* 9:R108.
15. Lage K, et al. (2008) A large-scale analysis of tissue-specific pathology and gene expression of human disease genes and complexes. *Proc Natl Acad Sci USA* 105: 20870–20875.
16. Ackerman KG, et al. (2005) Fog2 is required for normal diaphragm and lung development in mice and humans. *PLoS Genet* 1:58–65.
17. Clugston RD, Zhang W, Greer JJ (2008) Gene expression in the developing diaphragm: Significance for congenital diaphragmatic hernia. *Am J Physiol Lung Cell Mol Physiol* 294:L665–L675.
18. Dingemann J, Doi T, Ruttenstock E, Puri P (2011) The role of primary myogenic regulatory factors in the developing diaphragmatic muscle in the nitrofen-induced diaphragmatic hernia. *Pediatr Surg Intl* 27:579–582.
19. Ernst J, Bar-Joseph Z (2006) STEM: A tool for the analysis of short time series gene expression data. *BMC Bioinformatics* 7:191.
20. Subramanian A, et al. (2005) Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* 102: 15545–15550.
21. Subramanian A, Kuehn H, Gould J, Tamayo P, Mesirov JP (2007) GSEA-P: A desktop application for Gene Set Enrichment Analysis. *Bioinformatics* 23:3251–3253.
22. Bleyl SB, et al. (2007) Candidate genes for congenital diaphragmatic hernia from animal models: Sequencing of FOG2 and PDGFRalpha reveals rare variants in diaphragmatic hernia patients. *Eur J Hum Genet* 15:950–958.
23. Golzio C, et al. (2007) Matthew-Wood syndrome is caused by truncating mutations in the retinol-binding protein receptor gene STRA6. *Am J Hum Genet* 80:1179–1187.
24. Kim PC, Mo R, Hui Cc C (2001) Murine models of VACTERL syndrome: Role of sonic hedgehog signaling pathway. *J Pediatr Surg* 36:381–384.
25. Liu J, et al. (2003) Congenital diaphragmatic hernia, kidney agenesis and cardiac defects associated with Slit3-deficiency in mice. *Mech Dev* 120:1059–1070.
26. Neri G, Gurrieri F, Zanni G, Lin A (1998) Clinical and molecular aspects of the Simpson-Golabi-Behmel syndrome. *Am J Med Genet* 79:279–283.
27. Oh J, et al. (2004) Mutations in two matrix metalloproteinase genes, MMP-2 and MT1-MMP, are synthetic lethal in mice. *Oncogene* 23:5041–5048.

28. Shi L, et al. (2005) NF90 regulates cell cycle exit and terminal myogenic differentiation by direct binding to the 3′-untranslated region of MyoD and p21WAF1/CIP1 mRNAs. *J Biol Chem* 280:18981–18989.
29. Yano S, et al. (2010) Familial Simpson-Golabi-Behmel syndrome: Studies of X-chromosome inactivation and clinical phenotypes in two female individuals with GPC3 mutations. *Clin Genet* 80:466–471.
30. Pasutto F, et al. (2007) Mutations in STRA6 cause a broad spectrum of malformations including anophthalmia, congenital heart defects, diaphragmatic hernia, alveolar capillary dysplasia, lung hypoplasia, and mental retardation. *Am J Hum Genet* 80: 550–560.
31. Ashburner M, et al.; The Gene Ontology Consortium (2000) Gene ontology: Tool for the unification of biology. *Nat Genet* 25:25–29.
32. Smith CL, Eppig JT (2009) The mammalian phenotype ontology: Enabling robust annotation and comparative analysis. *Wiley Interdiscip Rev Syst Biol Med* 1:390–399.
33. Grigoryan T, Wend P, Klaus A, Birchmeier W (2008) Deciphering the function of canonical Wnt signals in development and disease: conditional loss- and gain-of-function mutations of beta-catenin in mice. *Genes Dev* 22:2308–2341.
34. Harland RM (1994) The transforming growth factor beta family and induction of the vertebrate mesoderm: bone morphogenetic proteins are ventral inducers. *Proc Natl Acad Sci USA* 91:10243–10246.
35. Winnier G, Blessing M, Labosky PA, Hogan BL (1995) Bone morphogenetic protein-4 is required for mesoderm formation and patterning in the mouse. *Genes Dev* 9: 2105–2116.
36. Selleri L, et al. (2001) Requirement for Pbx1 in skeletal patterning and programming chondrocyte proliferation and differentiation. *Development* 128:3543–3557.
37. Hunt P, Krumlauf R (1992) Hox codes and positional specification in vertebrate embryonic axes. *Annu Rev Cell Biol* 8:227–256.
38. Martínez L, et al. (2007) [The etiology of congenital diaphragmatic hernia and esophageal atresia: The Hox genes]. *Cir Pediatr* 20:223–228.
39. Bertolino E, Reimund B, Wildt-Perinic D, Clerc RG (1995) A novel homeobox protein which recognizes a TGT core and functionally interferes with a retinoid-responsive motif. *J Biol Chem* 270:31178–31188.
40. Qin P, Cimildoro R, Kochhar DM, Soprano KJ, Soprano DR (2002) PBX, MEIS, and IGF-I are potential mediators of retinoic acid-induced proximodistal limb reduction defects. *Teratology* 66:224–234.
41. Mey J, Babiuk RP, Clugston R, Zhang W, Greer JJ (2003) Retinal dehydrogenase-2 is inhibited by compounds that induce congenital diaphragmatic hernias in rodents. *Am J Pathol* 162:673–679.
42. Bielinska M, et al. (2007) Molecular genetics of congenital diaphragmatic defects. *Ann Med* 39:261–274.
43. Kuijper S, Turner CJ, Adams RH (2007) Regulation of angiogenesis by Eph-ephrin interactions. *Trends Cardiovasc Med* 17:145–151.
44. Clugston RD, Zhang W, Greer JJ (2010) Early development of the primordial mammalian diaphragm and cellular mechanisms of nitrofen-induced congenital diaphragmatic hernia. *Birth Defects Res A Clin Mol Teratol* 88:15–24.
45. Hogue J, et al. (2010) A novel EFNB1 mutation (c.712delG) in a family with craniofrontonasal syndrome and diaphragmatic hernia. *Am J Med Genet A* 152A: 2574–2577.
46. Vitobello A, et al. (2011) Hox and Pbx factors control retinoic acid synthesis during hindbrain segmentation. *Dev Cell* 20:469–482.
47. Wat MJ, et al. (2011) Genomic alterations that contribute to the development of isolated and non-isolated congenital diaphragmatic hernia. *J Med Genet* 48:299–307.
48. Otake K, et al. (2009) Congenital diaphragmatic hernia with a pure duplication of chromosome 1q: Report of the first surviving case. *Pediatr Surg Int* 25:827–831.
49. Capellini TD, et al. (2006) Pbx1/Pbx2 requirement for distal limb patterning is mediated by the hierarchical control of Hox gene spatial distribution and Shh expression. *Development* 133:2263–2273.
50. Laurent A, Bihan R, Omilli F, Deschamps S, Pellerin I (2008) PBX proteins: much more than Hox cofactors. *Int J Dev Biol* 52:9–20.
51. Lu Q, Kamps MP (1996) Structural determinants within Pbx1 that mediate cooperative DNA binding with pentapeptide-containing Hox proteins: Proposal for a model of a Pbx1-Hox-DNA complex. *Mol Cell Biol* 16:1632–1640.
52. Shen WF, et al. (1996) Hox homeodomain proteins exhibit selective complex stabilities with Pbx and DNA. *Nucleic Acids Res* 24:898–906.
53. Kim C, Nielsen HC (2000) Hoxa-5 in mouse developing lung: Cell-specific expression and retinoic acid regulation. *Am J Physiol Lung Cell Mol Physiol* 279:L863–L871.
54. Mahony S, et al. (2011) Ligand-dependent dynamics of retinoic acid receptor binding during early neurogenesis. *Genome Biol* 12:R2.
55. Volpe MV, Wang KT, Nielsen HC, Chinoy MR (2008) Unique spatial and cellular expression patterns of Hoxa5, Hoxb4, and Hoxb6 proteins in normal developing murine lung are modified in pulmonary hypoplasia. *Birth Defects Res A Clin Mol Teratol* 82: 571–584.
56. Bailey JS, Rave-Harel N, McGillivray SM, Coss D, Mellon PL (2004) Activin regulation of the follicle-stimulating hormone beta-subunit gene involves Smads and the TALE homeodomain proteins Pbx1 and Prep1. *Mol Endocrinol* 18:1158–1170.
57. Capellini TD, et al. (2008) Pbx1/Pbx2 govern axial skeletal development by controlling Polycomb and Hox in mesoderm and Pax1/Pax9 in sclerotome. *Dev Biol* 321:500–514.
58. Hildebrand JD, Soriano P (2002) Overlapping and unique roles for C-terminal binding protein 1 (CtBP1) and CtBP2 during mouse development. *Mol Cell Biol* 22:5296–5307.
59. Pereira L, et al. (1999) Pathogenetic sequence for aneurysm revealed in mice underexpressing fibrillin-1. *Proc Natl Acad Sci USA* 96:3819–3823.
60. Vokes SA, Ji H, Wong WH, McMahon AP (2008) A genome-scale analysis of the cis-regulatory circuitry underlying sonic hedgehog-mediated patterning of the mammalian limb. *Genes Dev* 22:2651–2663.
61. Coy S, Caamaño JH, Carvajal J, Cleary ML, Borycki AG (2011) A novel Gli3 enhancer controls the Gli3 spatiotemporal expression pattern through a TALE homeodomain protein binding site. *Mol Cell Biol* 31:1432–1443.
62. Ferretti E, et al. (2011) A conserved Pbx-Wnt-p63-Irf6 regulatory module controls face morphogenesis by promoting epithelial apoptosis. *Dev Cell* 21:627–641.
63. Berkes CA, et al. (2004) Pbx marks genes for activation by MyoD indicating a role for a homeodomain protein in establishing myogenic potential. *Mol Cell* 14:465–477.
64. Maves L, et al. (2007) Pbx homeodomain proteins direct Myod activity to promote fast-muscle differentiation. *Development* 134:3371–3382.
65. Heidt AB, Rojas A, Harris IS, Black BL (2007) Determinants of myogenic specificity within MyoD are required for noncanonical E box binding. *Mol Cell Biol* 27: 5910–5920.
66. Irizarry RA, et al. (2003) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* 4:249–264.
67. MacQueen J (1967) Some methods for classification and analysis of multivariate observations. *Fifth Berkeley Symposium on Mathematical Statistics and Probability*, eds Le Cam LM, Neyman J (Univ of California Press, Berkeley, CA), pp 281–297.
68. Rice JA (2007) *Mathematical Statistics and Data Analysis* (Duxbury, Belmont, CA).
69. Loscertales M, Mikels AJ, Hu JK, Donahoe PK, Roberts DJ (2008) Chick pulmonary Wnt5a directs airway and vascular tubulogenesis. *Development* 135:1365–1376.