Supporting Information

Russell et al. 10.1073/pnas.1121621109

SI Materials and Methods

Animals. ***Tissue collection for LCM.*** Embryos were collected from timed-pregnant C57BL/6J mice E11.5 (4 litters, 29 embryos), E12.5 (4 litters, 24 embryos), and E16.5 (2 litters, 11 embryos) (Jackson Laboratory Institutional Animal Care and Use Committee protocol 01011). The embryos were embedded in OCT (Tissue-Tek) in an isopentane slurry.

Tissue collection for IHC. E11.5 and E12.5 embryos were fixed overnight at 4 °C [with 4% (vol/vol) PFA in PBS], dehydrated, and embedded in paraffin.

Diaphragm examination. Three litters of timed-pregnant *Pbx1*$^{tm1Mlc/tm1Mlc}$ mice at E15.5 (generated by insertion of a PGK-neo cassette in exon 3, upstream of the homeodomain) (1) were obtained from the Cleary Laboratory at Stanford University (Protocol APLAC 9839). The diaphragms of mutants and wild-type littermates were dissected from fixed whole embryos and photographed from a caudal orientation by using a Leica MZ 12.5 stereomicroscope with a transilluminating base and a Leica DFC 480 camera.

Laser Capture Microdissection (LCM) and RNA Extraction. LCM was performed on 10-μm transverse (E11.5 and E12.5) or sagittal (E16.5) midthoracic sections on PEN membrane glass slides (LCM0522; Applied Biosystems) maintained at −80 °C (maximum storage 36 h). Frozen sections were rapidly thawed and fixed in 75% vol/vol ethanol (30 s). Excess OCT was removed with two washes in RNase-free water (30 s each). The sections were then dehydrated with 75%, 95%, 100% vol/vol ethanol, and xylene (30 s each). The target tissue was identified on a Nikon Eclipse Ti-Emicroscope and microdissected by using the Arcturus XT instrument on CapSure HS LCM Caps (LCM0215) (Applied Biosystems). Total RNA was immediately extracted by using the PicoPure RNA extraction reagent (KIT0204) (Applied Biosystems) and incubated at 42 °C for 30 min.

Purification and Amplification. RNA extracted from serial sections of the same embryo were pooled after independent reviews of the microdissected slides (MKR, ML). The RNA samples were then purified on preconditioned columns by using the PicoPure Kit (Applied Biosystems), including DNase treatment with Qiagen RNase-Free DNase Set. Sample quality was determined by Agilent 2100 Bioanalyzer (≈200 pg per assay). Samples were retrotranscribed by the WT-Ovation Pico System (NuGEN) and purified by using Agencourt RNAClean Beads. The double-stranded cDNA was subsequently amplified by using the SPIA amplification protocol (NuGEN), and the supernatant was cleared by using the Zymo clean-up kit (Zymo Research). Next, the Encore Biotin Module (NuGEN) was used for fragmentation and labeling. To assess quality, 1-uL aliquots of each fragmented/labeled cDNA sample were run on the Agilent mRNA Nano Assay. Unless otherwise specified, all products were used according to the manufacturers' instructions.

RT-qPCR. To confirm expression, RT-qPCR was performed on 10 ng of template cDNA from E12.5 and E15.5 diaphragms by using iQ SYBR Green Supermix (Bio-Rad) on a C1000TM Thermal Cycler (Bio-Rad) and a CFX96TM Real-Time System (Bio-Rad). The data were analyzed by using the ΔΔCt method and normalized against *Sdha* (geNorm Kit; PrimerDesign Ltd).

Immunohistochemistry (IHC). Transverse thoracic tissue sections (7 μm) were deparaffinized, hybridized with primary antibodies as described (2), and visualized with the ABComplex/HRP Vectastain Detection System according to manufactures protocols (Vector Laboratories) after 3,3′-diaminobenzidine (DAB) enhancement (Sigma) (PBX1b, MYOD1, MEIS2), or by the VECTASTAIN ABC-AP KIT (AP-1000) (Vector Laboratories) (RUNX1). The combinations of primary and biotinylated secondary antibodies used in this study are listed in Table S1.

1. Selleri L, et al. (2001) Requirement for Pbx1 in skeletal patterning and programming chondrocyte proliferation and differentiation. *Development* 128:3543–3557.

2. Loscertales M, Mikels AJ, Hu JK, Donahoe PK, Roberts DJ (2008) Chick pulmonary Wnt5a directs airway and vascular tubulogenesis. *Development* 135:1365–1376.

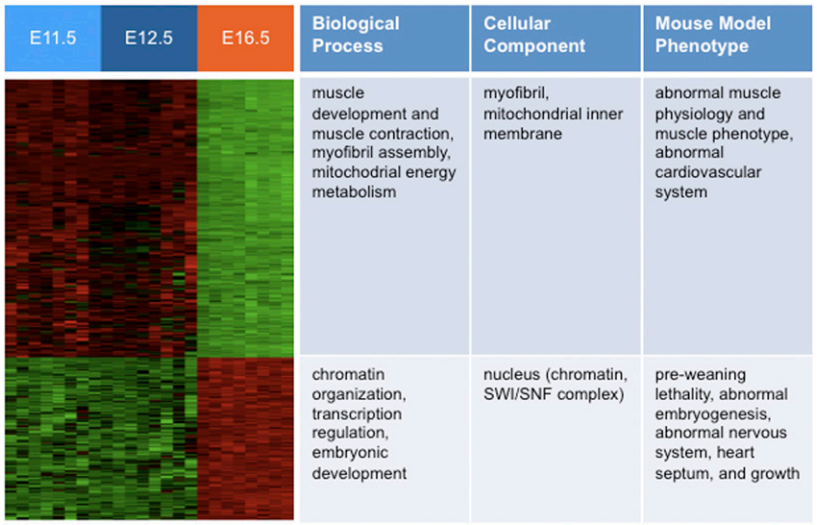
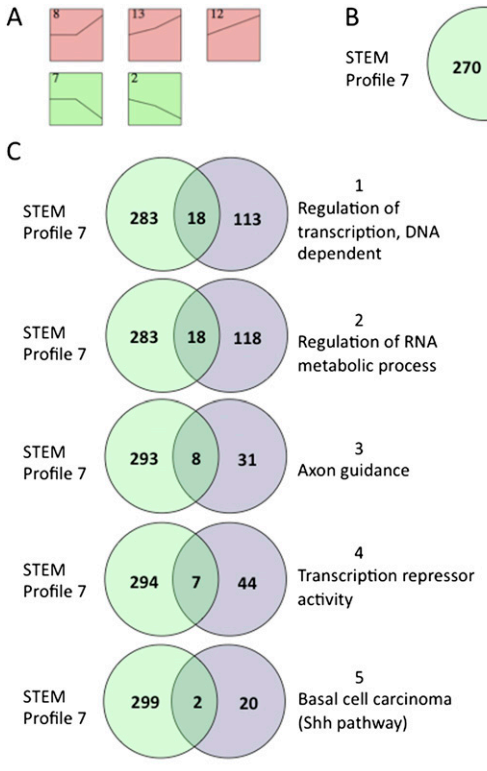


Fig. S1. Heatmap of the differentially expressed probes determined by Welch *t* test. Significance was set at *P* values < 0.01 after Bonferroni correction. The heatmap shows a marked difference between the transcriptional profiles of early and late time-points. The functional and phenotypic signatures associated with members of each cluster were analyzed by using Visual Annotation Display (VLAD; http://tolfalas.informatics.jax.org/vlad/) from MGI to test for enrichment of terms in the Gene Ontology (GO) and Mammalian Phenotype (MP) ontology. On the right, VLAD annotations of cluster 1 (genes up-regulated at E16.5) and cluster 2 (genes up-regulated at E11.5 and E12.5) are summarized. The cluster showing increased expression at E16.5 contained 550 probes corresponding to 477 genes, of which 463 had functional GO annotations. Among these genes, VLAD identified an overrepresentation of GO Biological Process terms related to muscle development, muscle contraction, and generation of precursor metabolites and energy (see details in Dataset S1). Additionally, 210 genes were associated with one or more terms in the MP ontology, showing overrepresentation of abnormal muscle and cardiovascular phenotypes (see details in Dataset S1). The second cluster with increased expression at E11.5 and E12.5 contained 311 probes corresponding to 293 genes, among which 283 had functional GO annotations, representing chromatin regulation and cell proliferation (see details in Dataset S1). In addition, 157 (of 293) genes were associated with one or more terms in the MP ontology, e.g., developmental abnormalities of heart septum morphology and the nervous system (see details in Dataset S1).



Gene Symbol	Associated GSEA Pathway
Alx1	4
Casp8ap2	4
Chd3	1; 2
Crabp2	1; 2
Efna5	3
Epha3	3
Epha4	3
Epha7	3
Ephb2	3
Ezh2	1; 2
Fzd2	5
Gata6	1; 2
Gli3	5
Hells	1; 2
Hmga2	1; 2
Ilf3	1; 2; 4
Khdrbs1	4
Meis2	1; 2; 4
Pbx1	1; 2
Phf21a	1; 2
Robo1	3
Runx1	1; 2
Sema3a	3
Smarca1	1; 2
Smarcc1	1; 2
Snai2	1; 2
Tgif1	1; 2; 4
Twist1	1; 2
Uhrf1	1; 2
Unc5c	3
Zfpm2	1; 2; 4

Fig. S2. Candidate gene prioritization. (*A*) Gene temporal profiles E11.5 to E16.5 (black lines) identified by STEM. Red boxes indicate profiles (8, 13, and 12) with increasing expression. Green boxes indicate profiles (7 and 2) with decreasing expression. (*B*) Thirty-one of 301 genes belonging to STEM Cluster 7 (up-regulated expression early in diaphragm development at E11.5/E12.5) intersect with genes from the combined "bait" list enriched GSEA gene set. (*C*) Overlap partitioned by GSEA gene set(s) 1–5. (*D*) The 31 overlapping genes, and corresponding GSEA gene set(s) (1-5), are listed.

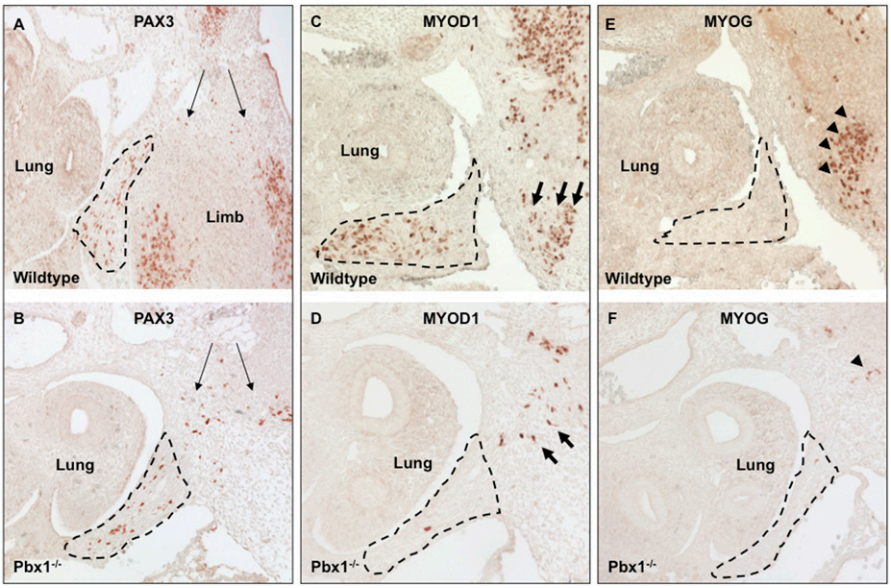


Fig. S3. Expression patterns of myogenic markers. To explore whether the loss of PBX1 interfered with myogenic processes in the early diaphragm, we characterized expression patterns of muscle precursor cell migration (PAX3), muscle specification (MYOD1), and differentiation (MYOG) markers on transverse sections from wild-type E11.5 C57BL/6J mice and *Pbx1* mutant embryos [PAX3 (*A* and *B*), MYOD1 (*C* and *D*), and MYOG (*E* and *F*) expression at E11.5; primordial diaphragmatic mesenchymal regions is outlined by dotted lines]. Muscle precursor cells expressing PAX3 (*A* and *B*, long arrows), and cells expressing MYOD1, followed the same migratory route from the cervical myotome to the PPF tissue (*C* and *D*, short arrows) both in *Pbx1* mutants and wild-type animals. However, fewer cells expressing MYOD1 were observed in the body wall and in the PPF of the mutants (*D*, short arrows). MYOG (*E* and *F*) expression was observed in the muscle bundles of the body wall and limb buds at E11.5 and is decreased in the mutant compared with wild-type embryos. MYOG expression, however, was not detected in the PPF tissue of mutant or wild-type embryos at this stage of development, indicating that these cells have not begun to terminally differentiate. Expression of all these markers was reduced in the *Pbx1* mutants. (Magnification: 10×).

Table S1. Antibodies for IHC and hybridization conditions

Table S1 (XLS)

Dataset S1. Results of Welch *t* test and annotations

Dataset S1 (XLS)

Sheet 1: Results of Welch *t* test. probe_id, Affymetrix 1.0_ST microarray probe IDs; stat, *t* test statistic; pv, *P* value; bonf, Bonferroni corrected *P* values. Sheets 2 and 3: List of genes upregulated at E16.5 (group 1, see text) and E11.5/E12.5 (group 2, see text), according to Welch *t* test (P value < 0.01). Sheets 4–7. Results of VLAD summary annotations of GO (Biological Processes) and MP databases in group 1 genes (sheets 4 and 6, respectively), and in group 2 genes (sheets 5 and 7, respectively). Database terms are listed in VLAD format from the most overrepresented. Term ID, database annotation number; Term, database annotation name; Pval, VLAD computed p-value; *k*, number of queried genes with the specific annotation; *n*, total number of queried genes; *K*, number of genes with the specific annotation in the database; *N*, total number of genes in the database; *k*/*n*, ratio of queried genes with the specific annotation; *K*/*N*, ratio of genes with the specific annotation in the database; *k*/*K*, ratio of queried to database genes with the specific annotation; *n*/*N*, ratio of queried to database genes; Qset, name of the queried dataset; Symbols, genes with the specific annotation among the queried genes.

Dataset S2. RMA normalized expression levels at each time-point

Dataset S2 (XLS)

Expression levels above the threshold of the median for intronic probes are indicated in green (for E11.5 and E12.5 the median was 6.3; for E16.5 the median was 6.2). MGI ID, Mouse Genome Informatics database IDs; Probe ID, Affymetrix 1.0_ST microarray probe IDs; Symbol, common gene nomenclature; Time-points, RMA normalized intensity values at each time-point.

Dataset S3. Results of STEM analysis

Dataset S3 (XLS)

Sheets 1-5: Statistically significant STEM generated profiles and genes contained therein. Profiles 8, 13, and 12 contain genes upregulated in late development (E16.5); profiles 7, and 2 contain genes upregulated in early development (E11.5, and E12.5).

Dataset S4. Results of GSEA analysis

Dataset S4 (XLS)

Sheets 1 and 2: Gene Ontology (GO) Molecular Signature Database (MSigDB) Gene Sets (collection C5) enriched during late (E16.5, sheet 1) and early (E11.5 and E12.5, combined, sheet 2) diaphragmatic development. Sheets 3 and 4: Canonical Pathway Molecular Signature Database (MSigDB) Gene Sets (collection C2) enriched during late (E16.5, sheet 3) and early (E11.5 and E12.5, combined, sheet 4) diaphragmatic development. *ES*, enrichment score; FDR q-val, false discovery rate *q* value; FWER p-val, family wise error rate *P* value; LEADING EDGE, standard GSEA leading edge statistics (see GSEA User Guide, http://www.broadinstitute.org/gsea/doc/GSEAUserGuideFrame.html, for details); NAME, pathway name; NES, normalized enrichment score; NOM p-val, nominal *P* value; *RANK AT MAX*, position of the maximum enrichment score; SIZE, number of genes in the pathway.

Dataset S5. Manually curated list of CDH associated genes, used as "baits" in the enrichment analysis

Dataset S5 (XLS)

Citation for Mouse Phenotypes, reference paper(s) for mouse phenotype; Entrez Gene ID, Entrez Gene ID; Gene Symbol, common gene nomenclature; Herniated Diaphragm, Yes/No; Human Diaphragm Defect, Yes/No; Human OMIM Term, MIM descriptive terms; Human Phenotype Citations, reference paper(s) for human phenotype; Mouse Diaphragm Defect, Yes/No; MGI Gene ID, Mouse Genome Informatics database IDs; Mouse Phenotype, type of diaphragmatic defect described in mouse; Mouse Model, name of mouse model; Mouse Model (MGI) Allelic Composition, as derived from the MGI database; Notes, anatomical location of the defect (if reported) or details about the phenotype; OMIM ID, MIM accession ID; Thin Diaphragm Muscle, Yes/No; #, row number.