

# UCSF

## UC San Francisco Previously Published Works

### Title

Long noncoding RNAs are spatially correlated with transcription factors and regulate lung development

### Permalink

<https://escholarship.org/uc/item/6p21j2vq>

### Journal

Genes & Development, 28(12)

### ISSN

0890-9369

### Authors

Herriges, Michael J  
Swarr, Daniel T  
Morley, Michael P  
[et al.](#)

### Publication Date

2014-06-15

### DOI

10.1101/gad.238782.114

Peer reviewed

# Long noncoding RNAs are spatially correlated with transcription factors and regulate lung development

Michael J. Herriges,<sup>1,7</sup> Daniel T. Swarr,<sup>2,3,7</sup> Michael P. Morley,<sup>4</sup> Komal S. Rathi,<sup>4</sup> Tien Peng,<sup>4</sup> Kathleen M. Stewart,<sup>4</sup> and Edward E. Morrisey<sup>1,4,5,6,8</sup>

<sup>1</sup>Department of Cell and Developmental Biology, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA; <sup>2</sup>Division of Neonatology, <sup>3</sup>Division of Human Genetics, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania 19104, USA; <sup>4</sup>Department of Medicine, <sup>5</sup>Institute for Regenerative Medicine, <sup>6</sup>Cardiovascular Institute, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA

Long noncoding RNAs (lncRNAs) are thought to play important roles in regulating gene transcription, but few have well-defined expression patterns or known biological functions during mammalian development. Using a conservative pipeline to identify lncRNAs that have important biological functions, we identified 363 lncRNAs in the lung and foregut endoderm. Importantly, we show that these lncRNAs are spatially correlated with transcription factors across the genome. In-depth expression analyses of lncRNAs with genomic loci adjacent to the critical transcription factors *Nkx2.1*, *Gata6*, *Foxa2* (forkhead box a2), and *Foxf1* mimic the expression patterns of their protein-coding neighbor. Loss-of-function analysis demonstrates that two lncRNAs, LL18/NANCI (*Nkx2.1*-associated noncoding intergenic RNA) and LL34, play distinct roles in endoderm development by controlling expression of critical developmental transcription factors and pathways, including retinoic acid signaling. In particular, we show that LL18/NANCI acts upstream of *Nkx2.1* and downstream from Wnt signaling to regulate lung endoderm gene expression. These studies reveal that lncRNAs play an important role in foregut and lung endoderm development by regulating multiple aspects of gene transcription, often through regulation of transcription factor expression.

[Keywords: *Nkx2.1*; Wnt signaling; development; lncRNA; lung; retinoic acid]

Supplemental material is available for this article.

Received January 23, 2014; revised version accepted May 13, 2014.

The advent of large-scale sequencing efforts has resulted in the identification of thousands of small (<200 nucleotides [nt]) and large (>200 nt) noncoding RNAs that potentially play critical roles in regulating gene expression. Small noncoding RNAs such as microRNAs (miRNAs), piwi RNAs, and snoRNAs have been shown to be important players in regulating gene expression in many contexts, including tissue development (Dieci et al. 2009; Wapinski and Chang 2011; Luteijn and Ketting 2013; Yates et al. 2013). In comparison, long noncoding RNAs (lncRNAs) are less conserved, and their function is less well understood. lncRNAs can be transcribed from a unique

transcriptional locus or a locus overlapping a protein-coding gene (PCG) in either the sense or antisense direction. Moreover, while many reported lncRNAs are polyadenylated similar to mRNAs, other lncRNAs are not, and, on average, lncRNAs are less stable than mRNA (Clark et al. 2012). lncRNAs are also spliced less efficiently than mRNAs, further suggesting that mRNAs and lncRNAs are processed differently (Tilgner et al. 2012). lncRNAs often act as scaffolds that help localize chromatin-modifying complexes important for activating or repressing gene transcription in *cis* or in *trans* (Zhao et al.

<sup>7</sup>These authors contributed equally to this work.

<sup>8</sup>Corresponding author

E-mail [emorrise@mail.med.upenn.edu](mailto:emorrise@mail.med.upenn.edu)

Article is online at <http://www.genesdev.org/cgi/doi/10.1101/gad.238782.114>.

© 2014 Herriges et al. This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see <http://genesdev.cshlp.org/site/misc/terms.xhtml>). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

2008; Khalil et al. 2009; Wang et al. 2011; Ulitsky and Bartel 2013).

The development of the respiratory system is a highly complex process involving the interaction of foregut-derived Nkx2.1<sup>+</sup> endoderm with surrounding cardiopulmonary mesoderm (Goss et al. 2009; Peng et al. 2013). Nkx2.1 not only marks the early lung endoderm progenitors within the foregut endoderm, it also plays an essential role in the morphological development of the airway structure and differentiation of the various epithelial lineages within the lung (Minoo et al. 1995, 1997). Other critical transcription factors, such as Gata6 and Foxa1/2 (forkhead box a1/2), also play key roles in promoting foregut and respiratory endoderm morphogenesis and differentiation (Yang et al. 2002; Wan et al. 2004a,b, 2005; Zhang et al. 2008). However, how these factors themselves are regulated during foregut and lung development, at the level of either expression or function, remains poorly understood. While small RNAs such as miRNAs have been shown to play an important role in development of the respiratory system (Lu et al. 2007; Ventura et al. 2008; Tian et al. 2011), little is known about the role of lncRNAs in this process. Several lncRNAs are expressed in the lung, including MALAT1, which has been shown to play an important role in lung cancer progression (Schmidt et al. 2011; Xu et al. 2011; Lai et al. 2012). Although loss of MALAT1 does not affect lung development (Zhang et al. 2012a), chromosomal deletions encompassing other lncRNAs can cause lethal lung development disorders, suggesting that they may regulate lung development (Stankiewicz et al. 2009; Barnett et al. 2012; Parris et al. 2013; Szafranski et al. 2013). Loss of the lncRNA Fendrr has also been shown to result in respiratory defects and perinatal lethality (Sauvageau et al. 2013). However, there is little data on the specific expression patterns or function of lncRNAs in most tissues, including the foregut endoderm and the lung.

In this study, we used RNA sequencing (RNA-seq) combined with other methods to identify lncRNAs in the developing and adult lung. Our conservative pipeline provides a systematic and comprehensive bridge from transcriptome-wide analysis of lung lncRNAs to the interrogation of individual lncRNA biological function. We show that a subset of lncRNAs in the lung are located near transcription factors, including Nkx2.1, Gata6, Foxa2, and Foxf1, that play important roles in foregut and lung development. In addition to this subset, we identified several known and novel intergenic lncRNAs that are not associated with PCGs, and many of these have unique expression patterns, suggesting an important functional role in lung development and homeostasis. We assessed biological function for two lung lncRNAs and show that they exhibit distinct and important roles in regulating endoderm gene expression, including developmental and signaling pathways. Importantly, we show that one of these lncRNAs, LL18, plays an important role in lung development by acting upstream of the critical transcription factor Nkx2.1 and downstream from Wnt/ $\beta$ -catenin signaling to regulate lung endoderm gene expression and morphogenesis.

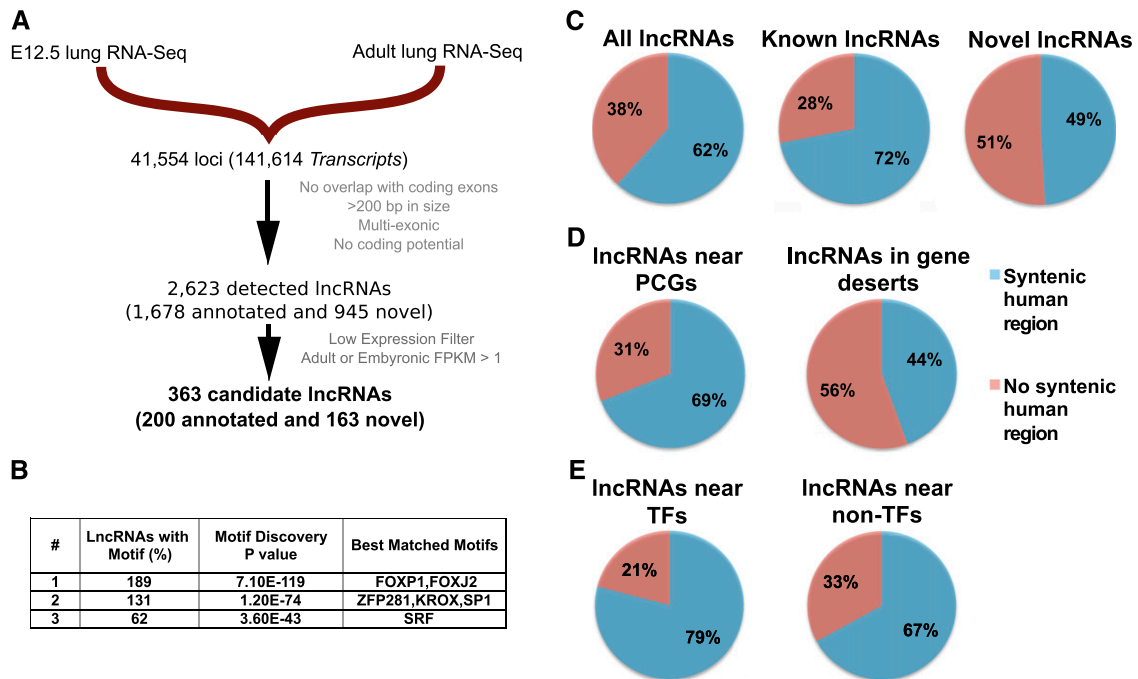
## Results

### *Identification of lncRNAs in the developing and postnatal lung*

To identify and begin to characterize the lncRNA transcriptome in the lung, we performed RNA-seq on embryonic day 12.5 (E12.5) and adult mouse lung polyadenylated RNA. Paired-end reads were aligned to the mouse reference NCBI 37 release using GSNAP (Wu and Nacu 2010). Cufflinks was used to assemble 141,614 unique transcripts from E12.5 and adult lung RNA-seq. The assembled transcripts were compared and annotated using Cuffmerge against the Ensembl (version 67) gene annotations. Transcripts that mapped to known lncRNAs were retained, and transcripts mapping to all other known genes were removed. Transcripts defined by Cuffmerge as unknown that were also multiexonic and >200 nt were considered putative lncRNAs. These putative lncRNAs were then filtered for potential coding ability using phylogenetic codon substitution frequency (PhyloCSF) and by scanning all potential ORFs per transcript against the protein domains in Pfam. Moreover, we only included lncRNAs that did not overlap extensively with exons of coding genes and were not located entirely within the intron of a coding gene. These analyses resulted in the identification of 363 lncRNAs, 200 of which were previously annotated, and 163 of which are novel (Fig. 1A; Supplemental Table 1). Using DNase hypersensitivity sites 5' of each lncRNA locus as a guide (Rosenbloom et al. 2013), we found that DNA-binding motifs for Fox, SP1/Krox, and serum response factor (SRF) factors were enriched in lung lncRNA regulatory regions (Fig. 1B). The high enrichment of Fox-binding motifs upstream of lung lncRNA loci suggests regulation by forkhead transcription factors, which are known to regulate foregut development as well as lung endoderm and mesoderm development (Wan et al. 2004b; Shu et al. 2007; Zhou et al. 2008; Chokas et al. 2010; Yu et al. 2010; Li et al. 2012; Parris et al. 2013).

While nucleotide conservation of lung lncRNAs across placental mammals was low (29%), as has been reported previously for lncRNAs in general (Supplemental Fig. 1; Derrien et al. 2012), many of these lncRNAs have syntenic transcripts in the human genome, indicating that they are conserved based on genomic location rather than nucleotide sequence (Fig. 1C). Our data also show that previously annotated lncRNAs that we uncovered in our analysis are more likely to have syntenic human transcripts than our novel subset of lncRNAs (Fig. 1C). Lung lncRNAs located near PCGs are more syntenic than lncRNAs located in gene deserts (Fig. 1D). Furthermore, lung lncRNAs located near transcription factor loci more often have syntenic human transcripts than lncRNAs found near nontranscription factor PCGs (Fig. 1E). This syntenic conservation of lncRNAs near transcription factor loci suggests a potential spatial relationship between these two gene subsets.

To further assess the potential spatial correlation between lncRNAs and transcription factor loci, we measured the genomic distance between lncRNAs and transcription



**Figure 1.** Identification of the mammalian lung lncRNA transcriptome. RNA-seq was performed on E12.5 and adult lungs to identify 141,614 transcripts generated from 41,554 loci. (A) These were filtered for size, coding potential, and expression to generate a catalog of 363 lncRNAs expressed in the mouse lung. (B) Motif enrichment analysis highlights the three most enriched transcription factor DNA-binding sites in the 5' regulatory regions of lung lncRNAs. (C) The majority of lncRNAs in this study have a human syntenic transcript, with previously annotated lncRNAs being more likely than novel lncRNAs to have a human syntenic transcript. (D) lncRNAs found within 20 kb of PCGs are more likely to have a human syntenic transcript than lncRNAs with no PCG gene nearby. (E) lncRNAs with PCG genes that were transcription factors are more likely to have syntenic human transcripts than those with nontranscription factor PCGs.

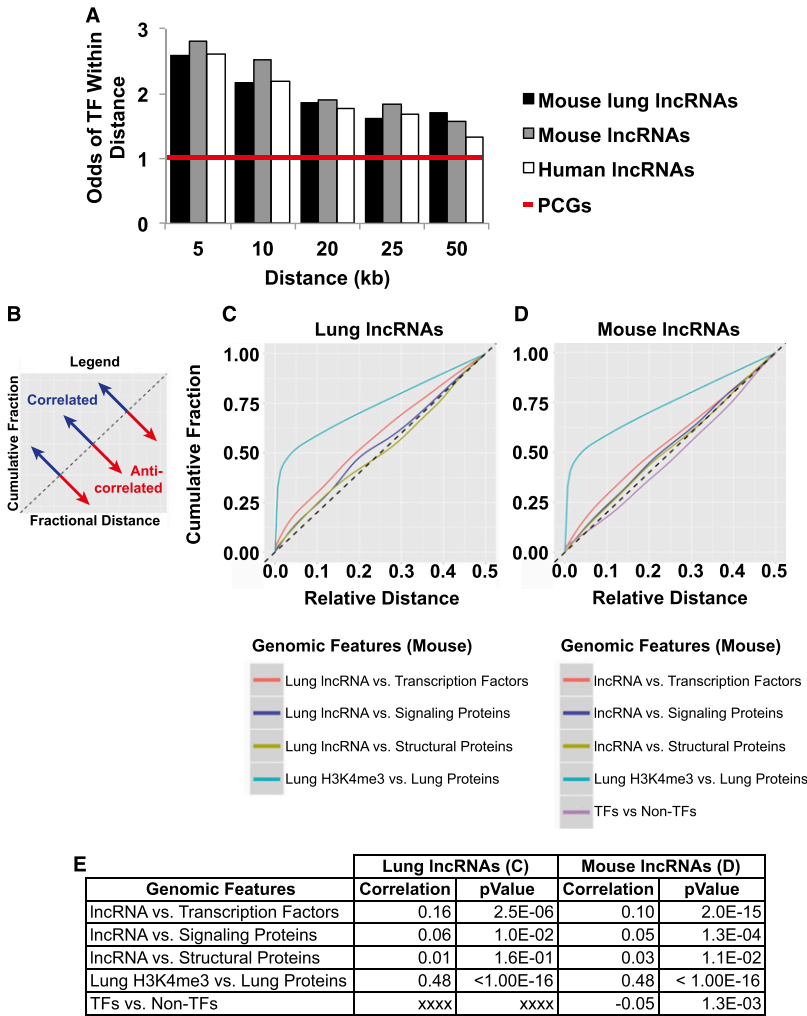
factor loci. Using our mouse lung lncRNA data set, we found that the lung lncRNAs in our database were more likely to be genomically located near transcription factors than other nontranscription factor PCGs (Fig. 2A). This enrichment was observed with all lncRNAs annotated in both the mouse and human genomes and increases as the distance from lncRNA loci decreases (Fig. 2A). Together, these observations support a conserved spatial correlation between lncRNA loci and transcription factor loci.

To validate this spatial correlation between lncRNAs and transcription factors, we used a recently published methodology that allows a direct comparison between different classes of PCGs and their spatial relationship with genomic loci (Fig. 2B; Favorov et al. 2012). This method also indicates a statistically positive spatial correlation between lung lncRNAs and transcription factors as well as those annotated in the mouse Ensembl database (Fig. 2C–E). Importantly, lncRNAs were spatially correlated with transcription factors more than other classes of PCGs, such as signaling proteins and structural proteins (Fig. 2C–E). While other classes of PCGs may also be positionally correlated with lncRNAs, our data suggest that lncRNAs are particularly enriched near transcription factors.

#### Characterization of the genomic organization of a subset of lung lncRNAs

From the initial RNA-seq-based screen of the lung transcriptome, we chose seven lung lncRNAs for in-depth

characterization. Due to the ability of some lncRNAs to regulate the expression of neighboring PCGs (Orom et al. 2010b; Wang et al. 2011; Berghoff et al. 2013), four lung lncRNAs (LL18, LL33, LL35, and LL66) were selected based on the close genomic proximity to transcription factors known to play a critical role in lung development. LL18, which we named NINCI (for Nkx2.1-associated noncoding intergenic RNA), is 2.5 kb downstream from Nkx2.1, a transcription factor that is a global regulator of anterior foregut and lung development and the earliest known marker of lung endoderm progenitor fate (Fig. 3A; Kimura et al. 1996). LL35 is 2.5 kb downstream from Foxa2, a forkhead transcription factor that is essential for endoderm and lung development (Fig. 3B; Wan et al. 2004a, 2005). Both NINCI and LL35 are transcribed in the sense orientation relative to their protein-coding neighbors. PCR using multiple primer pairs, 5' RACE, and the presence of unique RNA polymerase II (Pol II) and H3K4me3 marks indicates that NINCI and LL35 transcripts are driven from unique loci that are not linked to their PCGs (Fig. 3A,B; data not shown). In contrast, LL33 and LL66 are transcribed in an antisense orientation to their protein-coding neighbors. LL33 has multiple start sites that are upstream of or overlapping with the first exon of Gata6, an important endoderm transcription factor that is essential for lung epithelial differentiation and regeneration (Supplemental Fig. 3; Yang et al. 2002; Zhang et al. 2008). LL66 is located 1.4 kb upstream of Foxf1,



**Figure 2.** IncRNAs are spatially correlated with transcription factors. (A) The odds of a lncRNA being close to a transcription factor locus are statistically greater in lung lncRNAs identified in this study and from the Ensembl mouse and human annotated databases than in PCGs in general. (B) Spatial relationships were determined using the GenomTriCorr program, which computes relative distances between genomic features, which were then plotted and compared with a uniform random distribution (dashed line). The distributions *above* the line show a positive spatial correlation, while those *below* the line show an anti-correlation in spatial position. The distance between lung H3K4me3 marks and lung PCGs provides a positive control for spatial correlation between two genomic regions. (C) Lung lncRNAs identified in this study show a greater spatial correlation to transcription factors than other functional classes of PCGs, such as signaling or structural proteins. (D) This spatial correlation is also true for all lncRNAs annotated in the mouse Ensembl database. (E) Table of correlation coefficients and *P*-values from comparisons of the distance distributions for gene groups versus a uniform distribution.

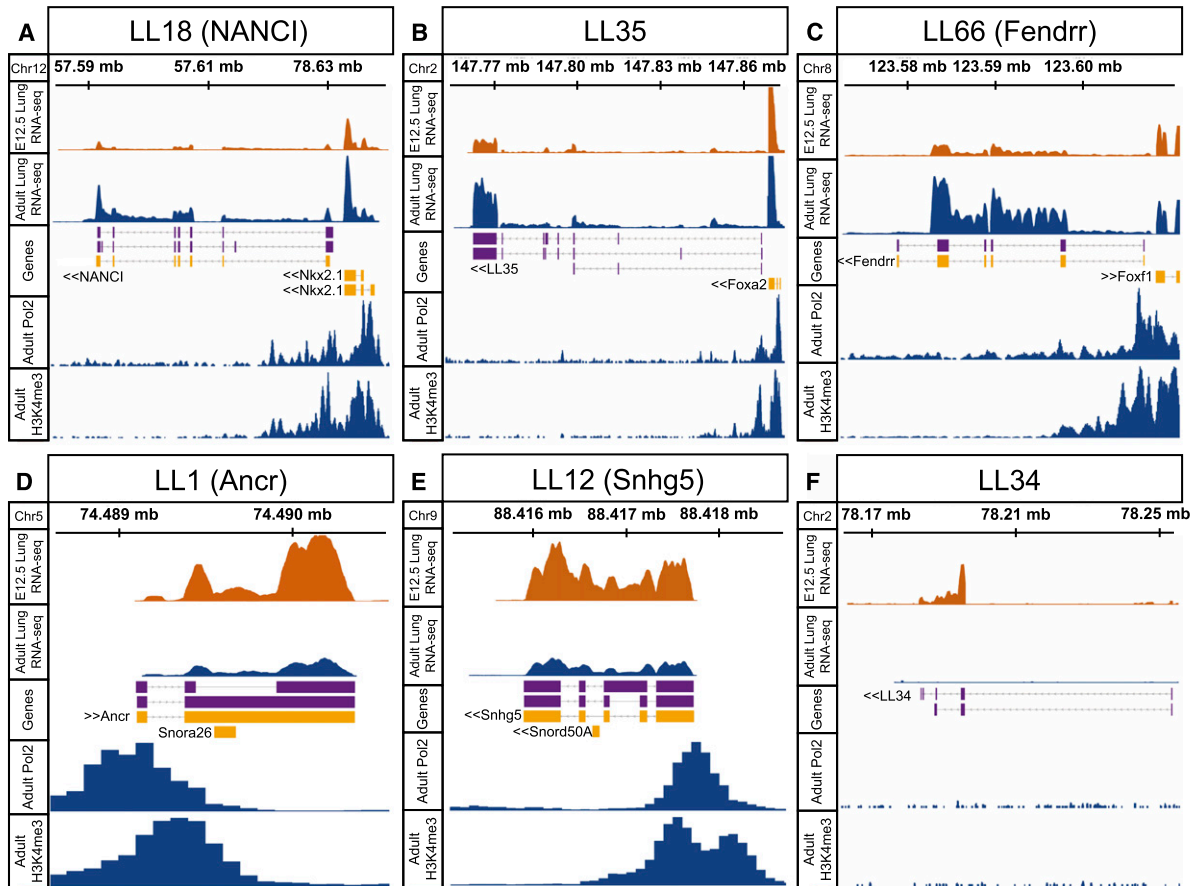
a forkhead transcription factor that plays an important role in the development of the pulmonary mesenchyme and vasculature and has been implicated in the pathogenesis of the lethal disorder alveolar capillary dysplasia (Fig. 3C; Stankiewicz et al. 2009). LL66, also known as Fendrl, was recently shown to be important in lateral plate mesoderm development, similar to its neighboring PCG, Foxf1 (Grote et al. 2013; Sauvageau et al. 2013). Quantitative PCR (qPCR) analysis confirmed the differences in E12.5 and adult mouse lung expression observed in the RNA-seq analysis for Nanci and LL35 (Supplemental Fig. 2A–D). However, RNA-seq revealed increased expression of LL66 in the adult lung relative to the E12.5 lung, whereas qPCR showed little change (Supplemental Fig. 2E,F).

Three other lncRNAs (LL1, LL12, and LL34), all of which are located in relative gene deserts, were selected due to their high level of expression in the E12.5 lung relative to the adult lung. LL1 and LL12 are previously identified lncRNAs and are known snoRNA host genes (Fig. 3D,E; Tanaka et al. 2000; Nakamura et al. 2008; Kretz et al. 2012). LL1 is the mouse ortholog of the human lncRNA anti-differentiation noncoding RNA (ANCR), which is up-regulated in multiple types of cancer and

inhibits osteoblast and keratinocyte differentiation (Kretz et al. 2012; Du et al. 2013; Zhu and Xu 2013). LL12 is annotated as snoRNA host gene 5 (Snhg5) in both the mouse and human genome and has been implicated in translocation induced B-cell lymphoma and recurrent prostate cancer (Du et al. 2013; Tanaka et al. 2000). In contrast, LL34 is a novel lncRNA that is expressed in the E12.5 lung but is nearly undetectable in the adult lung (Fig. 3F). qPCR analysis confirms that expression of all three of these lncRNAs is dramatically down-regulated in the adult lung compared with the E12.5 lung (Supplemental Fig. 2G,H). We also confirmed that these lncRNAs do not encode for a detectable protein using *in vitro* transcription/translation assays (Supplemental Fig. 4).

*lncRNAs associated with transcription factors exhibit spatial and temporal patterns of expression highly similar to their protein-coding neighbors*

While the genome-wide expression of some lncRNAs has been reported, little data are available on the spatial and temporal pattern of lncRNA expression during tissue development. Therefore, we performed extensive *in situ*



**Figure 3.** Genomic structure of lncRNAs expressed during lung development and homeostasis. Gene plots for LL18/NANCI (A), LL35 (B), LL66/Fendrr (C), LL1/Ancr (D), LL12 (E), and LL34 (F). The gene plots contain RNA-seq data from both E12.5 and adult lungs combined with Pol II and H3K4me3 ChIP-seq (chromatin immunoprecipitation [ChIP] combined with deep sequencing) data are from ENCODE (Rosenbloom et al. 2013). Transcripts were assembled from RNA-seq data using Cufflinks (purple) and compared with gene annotations in Ensembl (yellow).

hybridization (ISH) and qPCR analysis at different times of lung development and in different tissues to characterize our subset of lung lncRNAs. NANCI expression is observed in the developing lung, thyroid, and ventral forebrain in a pattern nearly identical to Nkx2.1 (Fig. 4A; Supplemental Fig. 5A–D). Within the developing lung, both genes are expressed throughout the developing epithelium, with lower levels found in the large airway epithelium (Fig. 4A). Expression of NANCI persists in the adult, with notable expression observed in large airway epithelium (Fig. 4A). The human ortholog of NANCI, which is annotated as surfactant-associated transcript A3 (SFTA3), is also a lncRNA and is expressed in a pattern similar to that of human NKX2.1 in both the airway and alveolar epithelium (Supplemental Fig. 6). This conserved expression pattern may be explained by conserved binding motifs for known regulators of foregut and lung development found in the NANCI promoter locus, including Smad, Forkhead, and Gata (Supplemental Fig. 7). Despite low sequence conservation between mouse and human NANCI (26%), their identical expression patterns suggest that there may be an evolutionarily conserved function. LL35 exhibits strong expression in the lung

epithelium, floor plate, gut endoderm, and liver in a pattern similar to that of Foxa2 (Fig. 4B; Supplemental Fig. 5E–H). LL66 is expressed primarily in the mesenchyme of the developing lung and posterior gut, similar to Foxf1 (Fig. 4C; Supplemental Fig. 5I–L). The expression of these lncRNAs was confirmed using qPCR and cDNA from E12.5 and adult lungs (Supplemental Fig. 5M,N).

The expression of LL33 is more restricted than Gata6 itself. Within the developing lung, Gata6 is robustly expressed in the developing epithelium and vascular smooth muscle (VSM), while LL33 expression can be detected in the VSM but is expressed at low levels in the developing epithelium (Supplemental Fig. 3). Outside of the lung, LL33 expression is observed primarily in the heart (Supplemental Fig. 3). During early heart development, LL33 was detectable in the atrio-ventricular canal region and the developing heart valves (Supplemental Fig. 3). In contrast, Gata6 is expressed throughout most of the heart, including the myocardium (Supplemental Fig. 3; Morrisey et al. 1996). By E16.5, LL33 expression expands into the myocardium such that, by E18.5, its expression is similar to that of Gata6 in the heart (Supplemental Fig. 3).

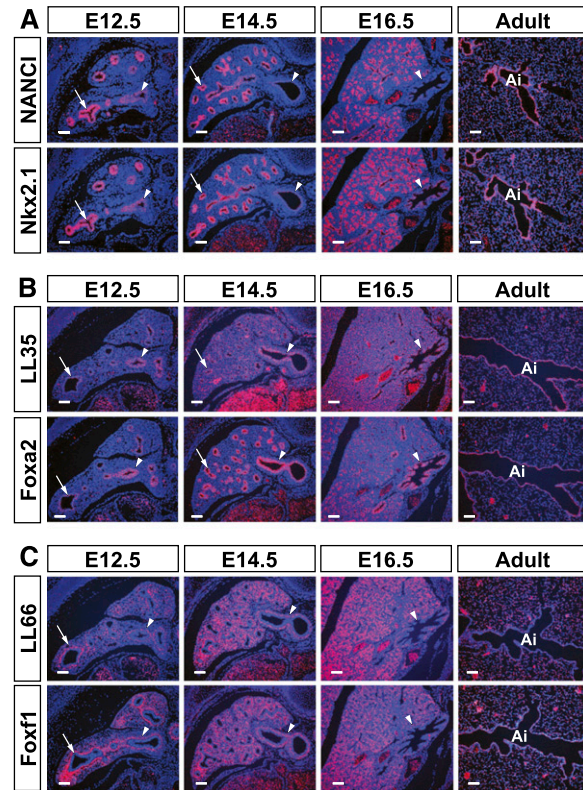
### Expression patterns of lung lncRNAs located in gene deserts

We next looked at expression patterns of those lung lncRNAs not associated with a transcription factor or PCG. At E12.5 and E14.5, LL1/ANCR is expressed in the distal airway epithelium and at lower levels in the lung mesenchyme (Fig. 5A, arrows; Supplemental Fig. 5O). Expression in the proximal airway epithelium is not observed (Fig. 5A, arrowheads). During development, LL1 expression diminishes until it is no longer detectable in the adult lung by ISH. Of note, the developing epidermis expresses high levels of LL1, as has been reported for human ANCR (Supplemental Fig. 5P; Kretz et al. 2012). LL12 expression is observed in the embryonic lung epithelium and persists at low levels in the adult lung (Fig. 5B; Supplemental Fig. 5Q). LL12 is also expressed robustly in the embryonic liver, where expression remains high in the adult (Supplemental Fig. 5Q,R). LL34 is expressed in the developing lung endoderm at E12.5 and E14.5, after which it is rapidly down-regulated such that its expression is not observed in the adult lung (Fig. 5C). Expression of LL34 is absent in tissues other than the lung at E12.5 by qPCR (Supplemental Fig. 5S). At E10.5, LL34 expression is observed throughout the anterior-to-posterior axis of the gut endoderm (Supplemental Fig. 5T–W). In the adult, LL34 expression is only detectable at low levels in the intestine by qPCR (Supplemental Fig. 5S).

### Loss of function for select lung lncRNAs reveals important roles in endoderm development

To define the function of a select subset of lncRNAs in lung epithelial cells, we performed shRNA-mediated knockdown of NANCI and LL34 in the mouse lung epithelial cell line MLE12. These lncRNAs represent two major categories of transcripts identified by our analysis, including those located near transcription factor loci (NANCI) and novel lncRNAs expressed in the developing but not the adult lung (LL34). Of note, both lncRNAs are expressed in MLE12 cells, and the selected shRNAs exhibited between 80%–95% knockdown efficiency upon optimization of viral transduction (Supplemental Fig. 8; data not shown). Changes in the transcriptome of MLE12 cells upon loss of lncRNA expression were assessed using microarray assays, which demonstrated that each lncRNA regulated unique cellular pathways in lung epithelial cells. Loss of NANCI expression resulted in a significant change in lung epithelial gene expression, with >1000 genes up-regulated or down-regulated compared with scrambled shRNA expression (Fig. 6A,B; Supplemental Table 2). Less dramatic effects were observed with knockdown of LL34, although loss of this lncRNA resulted in the altered expression of several hundred genes (Fig. 6A,B). Interestingly, there was little overlap between the genes altered by NANCI knockdown in comparison with LL34 (Fig. 6A,B).

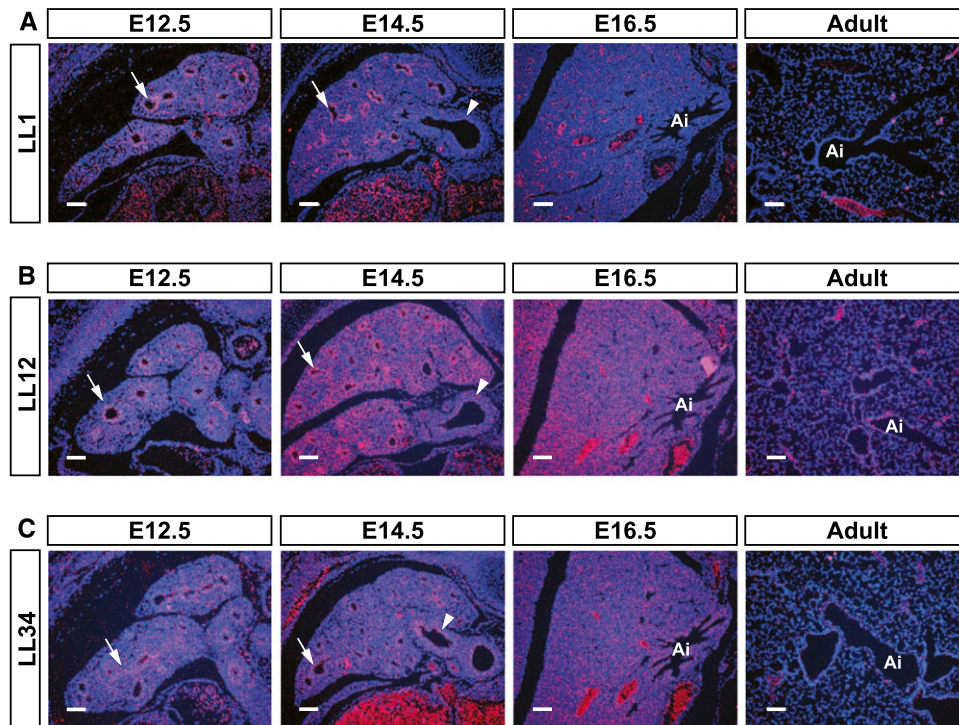
A large number of lung epithelial development genes were down-regulated by loss of NANCI expression such that respiratory and lung development were top categories of gene expression changes upon NANCI knockdown (Fig. 6C,D; Supplemental Tables 4, 5). Importantly,



**Figure 4.** Temporal and spatial expression of lncRNAs located near transcription factor loci during lung development. ISH was used to determine the expression patterns of NANCI and Nkx2.1 (A), LL35 and Foxa2 (B), and LL66 and Foxf1 (C), during lung development. Note the expression of Nkx2.1 and NANCI in the distal lung epithelium (A, arrows) and at lower levels in the developing proximal airways (A, arrowheads). Note the expression of Foxa2 and LL35 in the distal lung epithelium (B, arrows) and at higher levels in the developing proximal airways (B, arrowheads). Expression of Foxf1 and LL66 is observed in the mesenchyme surrounding both the distal lung epithelium (C, arrows) and proximal airways (C, arrowheads). (Ai) Airways. Bars, 100  $\mu$ m.

Nkx2.1 was down-regulated along with many targets of Nkx2.1, such as Sftpc, Abca3, and Aqp5 (Fig. 6E). Foxp1, Foxp2, Sox9, and other genes not known to be directly regulated by Nkx2.1 were also down-regulated after loss of NANCI expression by both microarray and qPCR analysis (Fig. 6F; Supplemental Table 2). Knockdown of NANCI in the human lung adenocarcinoma cell line H441 likewise resulted in decreased expression of NKX2.1 and AQP5 (Supplemental Fig. 9). These data suggest that in both mice and humans, NANCI regulates expression of the essential lung transcription factor Nkx2.1 and, in turn, Nkx2.1 target genes.

LL34 was found to be essential for the expression of several critical lung developmental genes, including Bmp2, Foxa1, Gata6, Fgfr2, Fgfr3, Fgf18, Jag1, and Tgfb3, as noted by microarray and qPCR analysis (Fig. 6G,J; Supplemental Table 3). Tube development was one of the top categories of genes altered by knockdown of LL34 (Fig. 6H; Supplemental Tables 6, 7). Interestingly, loss of LL34 resulted in decreased expression of several important components of



**Figure 5.** Temporal and spatial expression of lung lncRNAs located in relative gene deserts. ISH was used to determine the expression of LL1/ANCR (A), LL12 (B), and LL34 (C) during lung development. (A–C) Expression of all three lncRNAs is noted in the distal lung epithelium at E12.5 and E14.5 (arrows). At E14.5, expression of LL1/ANCR and LL12 is low in proximal airway epithelium of the lung, while LL34 expression is observed in this region (arrowheads). (A,C) By E16.5, expression of LL1 and LL34 is significantly decreased in the lung. (B) Only LL12 exhibits notable expression at E16.5 and in the adult. (Ai) Airways. Bars, 100 $\mu$ m.

the retinoic acid (RA) pathway, including *Aldh1a1* and *Rarg* (Fig. 6I–K; Supplemental Table 3). Decreased expression of these genes as well as other components of the RA pathway, including *Aldh1a2*, *Aldh1a3*, and *Cyp2b1*, was confirmed by qPCR (Fig. 6I). Many of the developmental regulators whose expression was decreased upon loss of LL34 are known targets of RA signaling (Fig. 6J). Of the 403 known RA targets expressed in MLE12 cells, 58 demonstrated altered levels of expression following knockdown of LL34 (Supplemental Table 3; Balmer and Blomhoff 2002). These data, along with the restricted expression of LL34 in early foregut endoderm, suggest that LL34 may modulate RA signaling in the early foregut and lung endoderm.

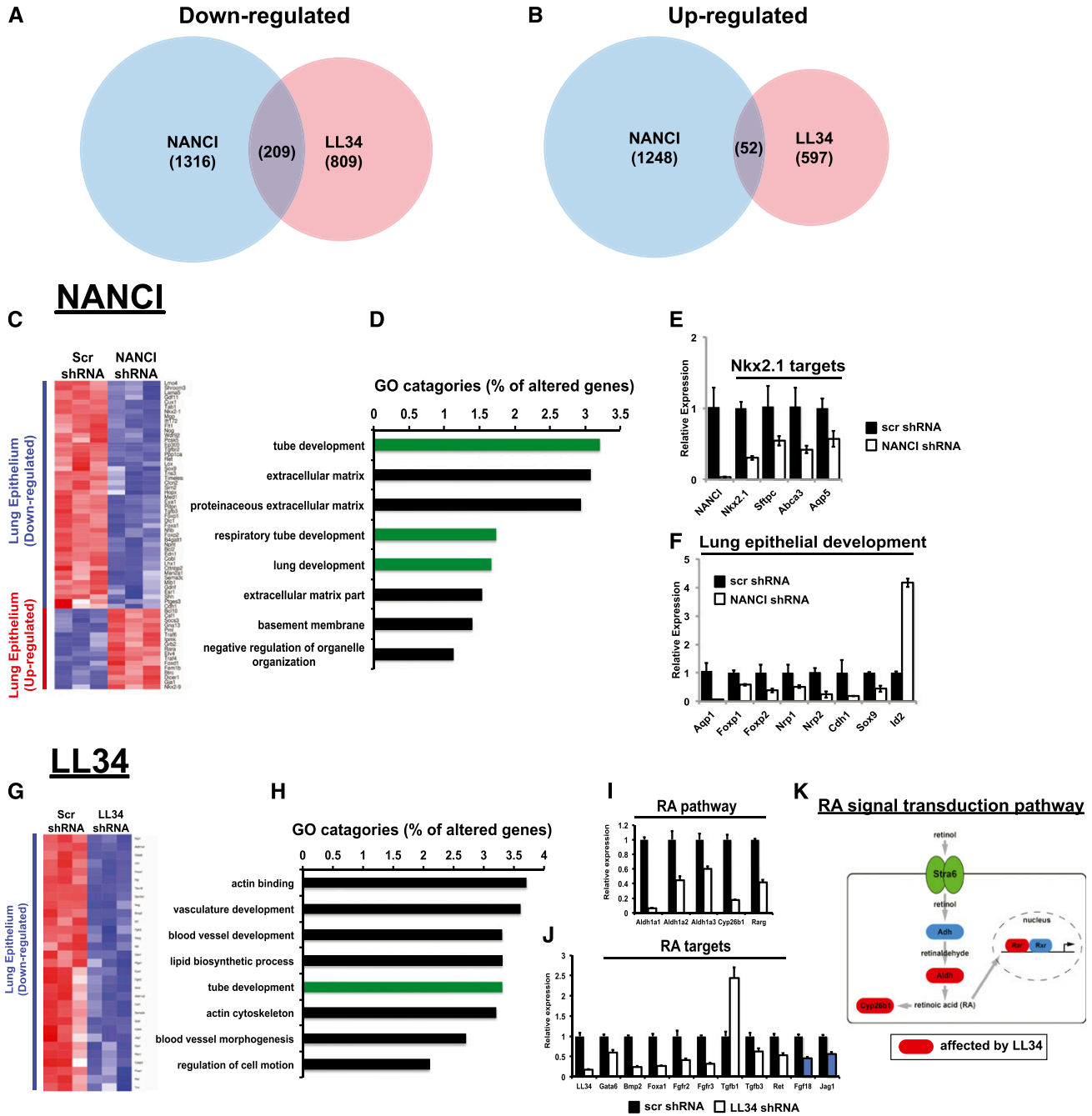
*NANCI acts upstream of Nkx2.1 and downstream from Wnt signaling to regulate lung epithelial gene expression*

The loss-of-function data for NANCI indicate its importance for Nkx2.1 expression and function during lung endoderm development. Given the central role that Nkx2.1 plays in anterior foregut and lung endoderm development, we chose to further explore the role of NANCI and determine how its regulation of Nkx2.1 affects lung epithelial differentiation. To determine how much of NANCI's effects on gene expression are mediated in part or in full by Nkx2.1, we attempted to rescue

the knockdown of NANCI and, in turn, loss of Nkx2.1 expression with ectopic overexpression of Nkx2.1. These experiments showed that increased Nkx2.1 expression rescued the expression of its target genes, *Abca3*, *Sftpc*, and *Aqp5*, and partially rescued *Aqp1* expression (Fig. 7A–C). Moreover, ectopic expression of Nkx2.1 led to increased endogenous Nkx2.1 expression (Fig. 7A). Such autoregulation of Nkx2.1 has been previously suggested by the finding of conserved Nkx2.1-binding sites in its promoter (Shaw-White et al. 1999; Das et al. 2011). Interestingly, expression of NANCI itself was unaffected by overexpression of Nkx2.1, indicating that Nkx2.1 does not regulate NANCI expression (Fig. 7A), further demonstrating distinct differences in regulation of these two loci. Overexpression of Nkx2.1 did not rescue expression of *Nrp1*, *Nrp2*, *Cdh1*, *Foxp1*, and *Foxp2*, suggesting that NANCI acts independently of Nkx2.1 to regulate the expression of these important developmental regulators (Supplemental Fig. 10). We next assessed whether NANCI could promote Nkx2.1 expression using overexpression of the full-length NANCI transcript in MLE12 cells. Overexpression of NANCI led to increased Nkx2.1 expression, suggesting that NANCI is capable of promoting Nkx2.1 expression in lung epithelium (Fig. 7D).

Previous data have demonstrated that Wnt signaling is an essential inducer of Nkx2.1<sup>+</sup> lung endoderm progenitors during anterior foregut development and in vitro differentiation of pluripotent stem cells into lung





**Figure 6.** shRNA-mediated loss-of-function experiments reveal distinct roles for NANCI and LL34 in lung epithelial gene transcription. (A,B) Venn diagram of microarray analysis of NANCI and LL34 shRNA knockdowns showing the number of both down-regulated and up-regulated genes. (C) Heat map of lung epithelial genes altered by knockdown of NANCI. (D) DAVID/GO (gene ontology) analysis shows that genes related to tube, respiratory, and lung development (green bars) are top categories altered by NANCI knockdown. (E) qPCR shows that Nkx2.1 and several of its targets are decreased by knockdown of NANCI expression. (F) qPCR also shows that expression of additional genes important for lung development were also down-regulated by knockdown of NANCI. (G) Heat map for lung developmental and RA pathway genes altered by knockdown of LL34. (H) DAVID/GO analysis shows that genes related to tube development (green bar) are a top category altered by LL34 knockdown. Expression of RA pathway components (I) and lung developmental genes known as RA targets (J) was verified by qPCR. (J) Note that Fgf18 and Jag1 are not RA targets (blue bars). (K) RA pathway diagram showing genes affected by decreased LL34 expression in red. All heat maps represent genes in the indicated category selected by GO analysis. All changes in gene expression shown by qPCR are statistically significant ( $P < 0.05$ ).

epithelium (Goss et al. 2009; Harris-Johnson et al. 2009; Longmire et al. 2012; Mou et al. 2012). Our data showing that ectopic Nkx2.1 expression in the context of loss of NANCI expression results in restoration of the expression of several Nkx2.1 target genes as well as the ability of ectopic NANCI to increase Nkx2.1 expression suggest that NANCI acts upstream of Nkx2.1 to regulate lung epithelial gene expression. To further explore how NANCI acts in the hierarchy of lung epithelial gene transcription, we assessed NANCI expression in *in vivo* models of loss of Nkx2.1 and Wnt signaling in the early lung endoderm. Nkx2.1<sup>-/-</sup> lungs are dysplastic and exhibit decreased expression of several known target genes, including Sftpc, Scgb1a1, and Aqp5 (Minoo et al. 1999). While the Nkx2.1 transcript was absent in Nkx2.1<sup>-/-</sup> lung epithelium as expected, NANCI expression was readily observed and maintained in a proximal-to-distal expression gradient as in normal embryonic lungs (Fig. 7E–H). In contrast, loss of Wnt/ $\beta$ -catenin in the anterior foregut endoderm of Shh<sup>cre</sup>;Cttnb1<sup>flox/flox</sup> mutants results in a loss of both Nkx2.1 and NANCI expression (Fig. 7I–L). Together, these data imply a model in which NANCI acts upstream of Nkx2.1 and downstream from Wnt/ $\beta$ -catenin signaling to regulate a suite of critical lung developmental genes (Fig. 7M).

Consistent with the observation that NANCI promotes Nkx2.1 expression, NANCI interacts strongly with Wdr5/trithorax transcriptional activating complex but not Ezh2/polycomb-repressive complex, similar to other activating lncRNAs (Supplemental Fig. 11; Zhao et al. 2008; Khalil et al. 2009; Wang et al. 2011; Ulitsky and Bartel 2013). In contrast, LL12 interacts strongly with Ezh2/polycomb, as has been previously shown (Zhao et al. 2010). Moreover, loss of NANCI reduces H3K4me3 marks imposed by the trithorax complex at the Nkx2.1 locus (Fig. 7N,O). Moreover, these data suggest that NANCI recruits the trithorax complex to activate Nkx2.1 expression, making it one of the earliest markers for lung endoderm specification (Fig. 7P).

*Knockdown of NANCI in the developing mouse lung results in disruption of epithelial differentiation and loss of Nkx2.1 target gene expression*

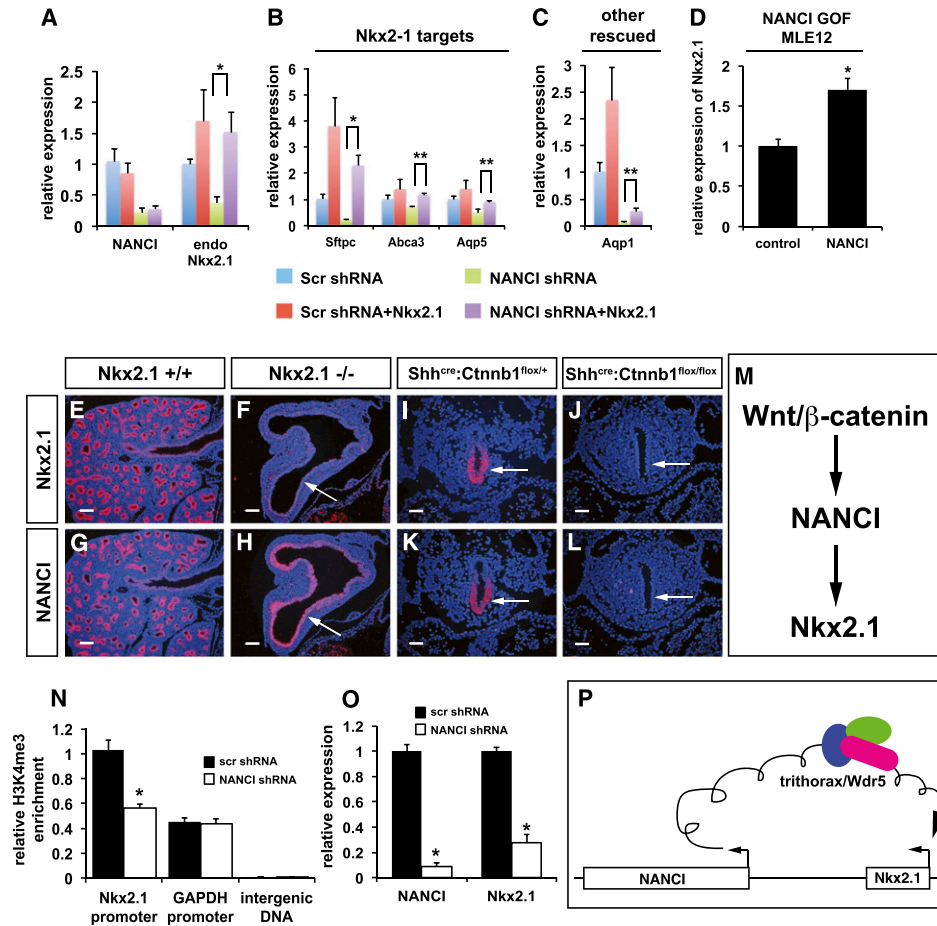
Many lncRNAs, especially ones located close to PCGs, are transcribed from genomic regions that overlap potential enhancer elements (Orom et al. 2010a). Therefore, to further characterize the role of the NANCI noncoding transcript in lung epithelial development without affecting the NANCI genomic locus, we performed *in vivo* knockdown of NANCI using the human SFTPC promoter, which is active throughout the developing lung epithelium. We used the previously demonstrated ability of shRNAs to be processed from miRNA precursor transcripts to generate a SFTPC-driven knockdown construct (Fig. 8A; Wang et al. 2007). Five F0 SFTPC-NANCI<sup>shRNA</sup> transgenic mice were generated, with an average NANCI knockdown of 43% at E17.5 (Fig. 8F; Supplemental Fig. 12A). SFTPC-NANCI<sup>shRNA</sup> transgenic lungs had reduced sacculle formation, as noted by H&E

staining and increased mesenchymal thickness between airways (Fig. 8B,D). Many of the genes down-regulated after NANCI knockdown in MLE12 cells were similarly affected in SFTPC-NANCI<sup>shRNA</sup> transgenic lungs, including Nkx2.1 and its targets, such as Sftpc, Abca3, and Aqp5 (Supplemental Fig. 12A). Likewise, Scgb1a1, an Nkx2.1 target gene not expressed in MLE12 cells, was down-regulated in SFTPC-NANCI<sup>shRNA</sup> transgenic lungs (Supplemental Fig. 12A). Decreased expression of both Aqp1 and Cdh1 was also observed, similar to the NANCI knockdown in MLE12 cells (Supplemental Fig. 12B). However, levels for Sox2, Sox9, and Id2 were unchanged, suggesting that lung endoderm progenitors were not drastically perturbed (Supplemental Fig. 12B). Immunostaining for Cdh1 shows that the distal airway epithelium of SFTPC-NANCI<sup>shRNA</sup> transgenic lungs retained a columnar morphology as compared with control littermates, indicating defects in distal sacculation/alveolarization (Supplemental Fig. 12E). Sftpc<sup>+</sup> alveolar epithelial type 2 cells were more focally located in SFTPC-NANCI<sup>shRNA</sup> transgenic lungs rather than dispersed as observed in transgene-negative littermates (Fig. 8G).

To determine whether the *in vivo* knockdown of NANCI mimicked decreased Nkx2.1 expression, Nkx2.1<sup>+/-</sup> lungs were examined at E17.5. Nkx2.1<sup>+/-</sup> lungs had an ~40% decrease in Nkx2.1 expression (Supplemental Fig. 12C). As in SFTPC-NANCI<sup>shRNA</sup> transgenic lungs, the Nkx2.1<sup>+/-</sup> mutant lungs had reduced sacculle formation and increased mesenchymal thickness, as noted by H&E staining (Fig. 8C,D). These mutants also had similar reductions in Sftpc, Abca3, Aqp5, Scgb1a1, and Aqp1, as noted in SFTPC-NANCI<sup>shRNA</sup> transgenic lungs (Supplemental Fig. 12C,D). The number of Sftpc<sup>+</sup> distal sacculles was also decreased in a manner similar to the SFTPC-NANCI<sup>shRNA</sup> transgenic lungs (Fig. 8H). However, the level of Cdh1 expression was not affected in Nkx2.1<sup>+/-</sup> lungs (Supplemental Fig. 12D). In contrast, while there were changes in the morphology of the distal airway epithelium of Nkx2.1<sup>+/-</sup> lungs as noted by Cdh1 immunohistochemistry, these changes were not as dramatic as noted in SFTPC-NANCI<sup>shRNA</sup> transgenic lungs (Supplemental Fig. 12F). Together, these data suggest that NANCI regulates lung epithelial development *in vivo* in large part through its regulation of the critical transcription factor Nkx2.1.

## Discussion

While different methodologies have resulted in the cataloging of thousands of lncRNAs, it has remained difficult to assess which portion of these transcripts plays a functional role in the cells and tissues where they are expressed. Using conservative criteria to focus on lncRNAs that may play a functional role in lung development, we identified 363 lncRNAs and show that they are spatially correlated with transcription factors important for development of the anterior foregut and lung. Importantly, we identified and characterized NANCI, a lncRNA located near the important lung endoderm regulator Nkx2.1, and show that NANCI acts upstream of Nkx2.1

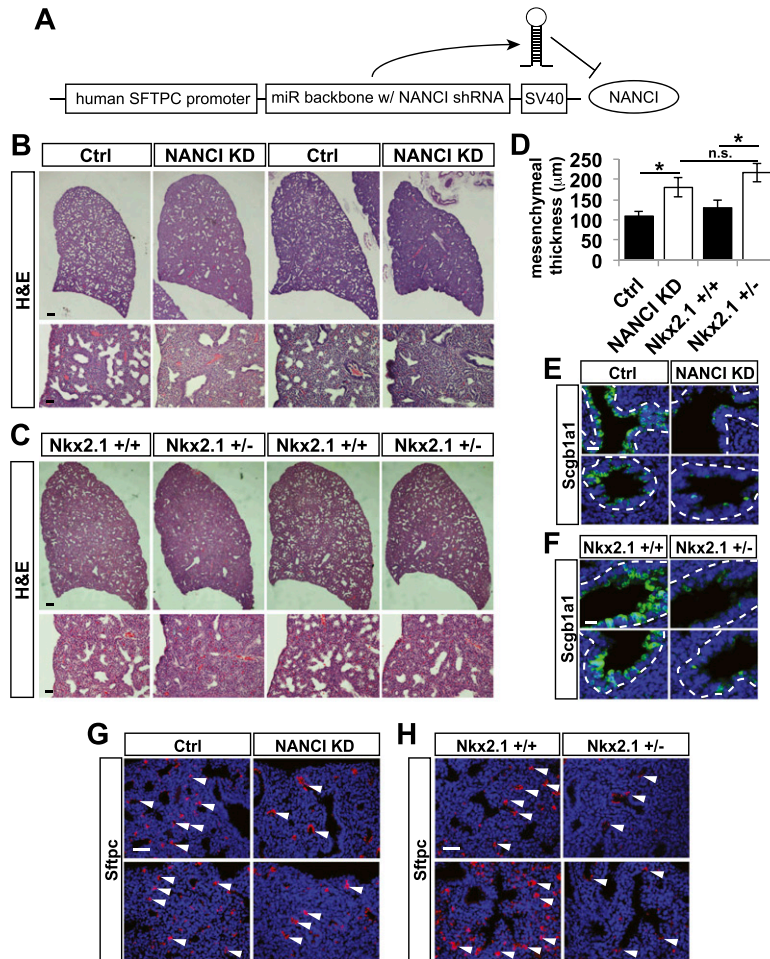


**Figure 7.** Nanci acts upstream of Nkx2.1 and downstream from Wnt/ $\beta$ -catenin to regulate lung gene expression. (A) Ectopic Nkx2.1 expression rescues endogenous Nkx2.1 expression but not Nanci expression in the presence of Nanci shRNA knockdown. (B) Ectopic Nkx2.1 expression rescues endogenous Sftpc, Abca3, and Aqp5 expression in the presence of Nanci shRNA knockdown. (C) Expression of Aqp1 is partially rescued by exogenous Nkx2.1 expression in the presence of Nanci knockdown. (D) Overexpression of Nanci in MLE12 increases Nkx2.1 expression. Nkx2.1 (E,F) and Nanci (G,H) expression in control lungs (E,G) or Nkx2.1<sup>-/-</sup> lungs (F,H) at E14.5. Note the loss of Nkx2.1 expression but persistent Nanci expression in Nkx2.1<sup>-/-</sup> lungs. Nkx2.1 (I,J) and Nanci (K,L) expression in the control (I,K) or Shh<sup>cre</sup>:Ctnnb1<sup>fllox/fllox</sup> mutant (J,L) anterior foregut at E9.5. (M) Together, these data indicate that Nanci acts downstream from Wnt/ $\beta$ -catenin but upstream of Nkx2.1 to regulate lung gene expression. (N) ChIP assay for H3K4me3 marks at the Nkx2.1 promoter, the GAPDH promoter, and an unrelated intergenic region. (O) The level of knockdown achieved for Nanci and the level of decreased Nkx2.1 expression using scrambled and Nanci-specific shRNAs used in the H3K4me3 ChIP assays. (P) Model of how Nanci potentially recruits trithorax/Wdr5 complexes to the Nkx2.1 locus to promote Nkx2.1 gene expression. Bars, 100  $\mu$ m. (\*)  $P < 0.01$ ; (\*\*)  $P < 0.05$ .

and downstream from Wnt/ $\beta$ -catenin to regulate lung epithelial gene expression. In contrast, LL34 regulates RA signaling, which is critical for foregut endoderm development. Taken together, our data provide a novel catalog of lncRNAs in the foregut and lung and show that several of these play unique and essential roles in foregut and lung epithelial gene expression and development.

Our data show that lncRNA loci are spatially correlated with transcription factor loci. These transcription factor-associated lncRNAs are more conserved at the syntenic level with the human genome than other lncRNAs. Moreover, in several cases, they have expression patterns remarkably similar to their neighboring transcription factors, suggesting that these lncRNAs are closely associated with the expression and function of

their neighboring transcription factors. We identified Nanci as a critically important lncRNA located near the lineage regulator Nkx2.1. Nanci's expression is identical to Nkx2.1, which correlates with its functional importance for regulating Nkx2.1 expression. The correlation between NKX2.1 and Nanci is also found in the human lung even though Nanci is poorly conserved at the nucleotide level. Surprisingly, Nanci expression persists in the absence of Nkx2.1 expression, indicating that Nkx2.1 is not required for Nanci expression. This is supported by the lack of conserved Nkx2.1 DNA-binding motifs within the Nanci promoter. This finding was unexpected, as the Nkx2.1 expression pattern is unique and has been previously recognized as the earliest marker of anterior foregut endoderm committed to the



**Figure 8.** In vivo knockdown of NANCI results in defective lung epithelial morphogenesis similar to Nkx2.1 haploinsufficiency. (A) The shRNA used in the in vitro studies was cloned into the miR-34 pre-miRNA backbone as previously described (Wang et al. 2007) and expressed in the developing lung epithelium using the human SFTPC promoter. (B) H&E staining shows a thickened mesenchyme and narrower airways in SFTPC-NANCI<sup>shRNA</sup> transgenic lungs versus nontransgenic controls. (D) Calculation of mesenchymal thickness confirms reduced airway sacculcation in SFTPC-NANCI<sup>shRNA</sup> transgenic lungs. (C,D) This same mesenchymal thickening and reduced airway sacculcation was also seen in Nkx2.1<sup>+/-</sup> lungs. SFTPC-NANCI<sup>shRNA</sup> transgenic lungs and Nkx2.1<sup>+/-</sup> lungs also exhibit decreased Scgb1a1 expression in the proximal airways (E,F, dotted lines) as well as decreased expression and improper organization of Sftpc<sup>+</sup> cells in the distal airways (G,H, arrowheads). Two different examples from different transgenic animals or mutants are shown in E–H to demonstrate consistency of phenotype. Bars: B,C top panels, 500 μm; B,C (bottom panels), G,H, 100 μm; E,F, 50 μm. All changes in gene expression shown by qPCR are statistically significant ([\*]  $P < 0.05$ ) except for the difference in mesenchymal thickness between SFTPC-NANCI<sup>shRNA</sup> transgenic lungs and Nkx2.1<sup>+/-</sup> lungs. (n.s.) Nonsignificant.

lung and thyroid fate (Longmire et al. 2012; Mou et al. 2012). Our data showing that NANCI acts upstream of Nkx2.1 indicate that NANCI is an earlier marker than Nkx2.1 of anterior foregut lung fate. Our data show that both NANCI and Nkx2.1 expression is lost upon inactivation of Wnt/ $\beta$ -catenin signaling in the anterior foregut endoderm, placing NANCI downstream from this pathway in lung endoderm development. Interestingly, a patient was recently identified with a chromosomal deletion that encompassed most of NANCI but left the NKX2.1 locus intact (Barnett et al. 2012). This patient had symptoms characteristic of NKX2.1 haploinsufficiency, including respiratory distress, suggesting that NANCI has a similar role of regulating NKX2.1 in the human lung. Our data show that knockdown of NANCI in human cells results in decreased NKX2.1 expression, further supporting a conserved role for NANCI in human development. Together, these findings indicate that NANCI acts between Wnt/ $\beta$ -catenin signaling and Nkx2.1 to regulate a broad array of lung endoderm-specific genes in part through regulation of Nkx2.1 itself.

Not all lncRNAs located near transcription factors that we identified were expressed in a pattern identical to their neighboring PCG. LL33, which is located next to the Gata6 locus, exhibited a divergent expression pattern in

both the developing lung and heart. Where Gata6 is expressed at high levels in the developing lung epithelium, LL33 is expressed at low levels in this tissue. Moreover, LL33 expression is not observed at significant levels in the developing myocardium of the heart until E18.5, whereas Gata6 is expressed in this tissue from the earliest stages of cardiac development. Such distinct differences may indicate tissue-specific regulatory roles for these lncRNAs during development.

The lncRNA LL34 is specifically expressed throughout the anterior-to-posterior axis of the early gut endoderm, and its expression continues in the early lung endoderm. However, its expression is extinguished by late gestation, making it the one of the most embryonic-specific lung lncRNAs identified in our study. This expression pattern correlates with data showing that loss of LL34 leads to decreased expression of several essential components of the RA signaling pathway as well as RA target genes. RA signaling plays a critical role in both early foregut patterning and lung development. Loss of RA signaling leads to lung agenesis, while additional studies suggest a role for RA signaling in other regions of gut endoderm, including in the pancreas (Martin et al. 2005; Wang et al. 2006; Chen et al. 2007, 2010; Bayha et al. 2009). Given LL34's restricted expression in the gut endoderm in early development and

the important role for RA signaling in endoderm development, this lncRNA is likely a key regulator of foregut development through modulation of RA signaling.

The lncRNAs in our data set that are in close proximity with important transcription factors could be considered enhancer RNAs (eRNAs), which are thought to regulate their neighboring PCGs through various mechanisms, including recruiting chromatin remodeling complexes or acting as genomic enhancers of the neighboring PCGs (Orom et al. 2010a). Our data showing that NANCI binds to TrxG/Wdr5 complexes suggest that this transcript may serve this type of function in regulating Nkx2.1 gene expression during foregut and lung development. Interestingly, other lncRNAs in our database, such as LL12, bind to both TrxG/Wdr5 and PRC2/Ezh2 at high levels, suggesting that they may play both activating and repressive roles in regulating gene transcription. Of note, LL12 was previously found to be associated with Ezh2 in a genome-wide survey of polycomb-associated lncRNAs (Zhao et al. 2010). Further experiments disrupting these interactions specifically will be required to fully assess their importance in regulating NANCI function.

Our database of lung lncRNAs reveals important features of this class of transcripts and also identifies several important players in both lung and foregut development. We characterized the tissue-specific expression patterns of several of these lncRNAs, revealing that they are often expressed in a pattern exquisitely similar to their neighboring PCGs. These data bolster and extend an emerging concept that lncRNAs are an important regulator of cell lineage differentiation during tissue development and suggest that a subset acts on neighboring transcription factors to regulate their expression and function.

## Materials and methods

### Library preparation

Total RNA was extracted from E12.5 and adult mouse lungs using the RNAEasy kit (Qiagen) and used in the preparation of sequencing libraries using the Illumina TruSeq IRNA sample preparation kit. Paired-end reads were sequenced on a HiSeq 2000 at a depth of four samples per lane to achieve an average 60 million reads per library. Two libraries per time point were analyzed.

### Alignment and transcript assembly

Paired-end reads were aligned to the mouse reference genome (build NCBI37/mm9) using the read mapper GSNAP version 2012-07-20 (Wu and Nacu 2010). GSNAP was executed in the mode to detect both novel and known splice sites. Known splice sites were obtained from Ensembl build 67. Cufflinks was then used to assemble the transcriptome for each of the four individual samples (Trapnell et al. 2013). These individual assemblies were then merged and annotated using Ensembl 67 into a single collection of transcripts using Cuffmerge. Expression quantification and differential expression analysis were performed on the transcript using Cuffdiff. Expression values in the form of fragments per kilobase of exon per million reads mapped (FPKM) were imported into a custom database for further analysis.

### lncRNA identification pipeline

Potential lncRNAs were discovered using the same pipeline described by Cabili et al. (2011). Briefly, transcripts that were multiexonic, >200 nt, and called with a class code of "u" in Cuffmerge were retained as potential novel lncRNAs. To remove transcripts that might be novel PCGs, PhyloCSF was used to assess the protein-coding potential of each transcript by examining the overrepresentation of evolutionary signatures characteristic of alignments of conserved coding regions, such as the high frequencies of synonymous codon substitutions and conservative amino acid substitution (Lin et al. 2011). Transcripts with a PhyloCSF score >100, which roughly indicated a false positive rate of 10% when comparing known noncoding transcripts, were excluded for further analysis. Transcripts were also excluded if they contained a known protein-coding domain. Pfam scan (<ftp://ftp.sanger.ac.uk/pub/databases/Pfam>) was used to search all potential ORFs against Pfam (release version 26, PfamA and PfamB). GenBank accession numbers are KF856953 for mouse LL18/NANCI, KF856952 for human LL18/NANCI, and KF876689 for LL34.

### Motif enrichment

DNase hypersensitivity sites (ENCODE) were used to predict protein-binding regions within lncRNA promoters 2.5 kb upstream of and 300 base pairs (bp) downstream from lncRNA transcription start sites. DNA sequences overlapping with lncRNA promoter regions and DNase hypersensitivity sites were created using BEDTools (Quinlan and Hall 2010). Novel motifs were discovered using the tool MEME (Bailey and Elkan 1994) and set to find the five most overrepresented motifs in the input sequences. Discovered motifs were then matched to known motifs in the TRANSFAC database (Matys et al. 2006) using the tool TOMTOM (Gupta et al. 2007).

### Neighbor gene analysis and spatial correlation

Genes from Ensembl were annotated as transcription factors using data from the animal transcription factor database (<http://www.bioguo.org/AnimalTFDB/index.php>) (Zhang et al. 2012b) A 2×2 table using distance windows of 5, 10, 20, 25, and 50 kb of counts was constructed with lncRNAs near transcription factor loci, lincRNAs near nontranscription factor PCG, PCGs near transcription loci, and PCGs near nontranscription factor loci. This count table was used in a single-sided Fisher's exact test to determine whether the proportion of lncRNAs near transcription factors was different from that of PCGs with near transcription factors.

The R package GenometriCorr was used to examine the spatial correlation of lncRNAs with PCGs (Favorov et al. 2012). Gene groups were formed using the animalTFDB database for transcription factors, signal proteins defined by gene ontology (GO) (GO:0038023), and structural proteins defined by GO (GO:0005198). GenometriCorr implements various statistical approaches to assess whether the positions of two sets of intervals are associated in genomic space. The 363 lncRNAs from the mouse lung and lncRNAs in Ensembl were compared against groups of PCGs defined as transcription factors, signaling proteins, or structural proteins using the GenometriCorr relative distance test. For a positive correlation reference the adult mouse lung H3K4me3 sites, a promoter mark was compared against PCGs expressed in the mouse lung determined by using our RNA-seq data (FPKM >5) The distributions' relative distances between lncRNA and the categories of PCGs were plotted using R and

tested for a significant difference from the uniform null distribution using a Kolmogorov-Smirnov test.

#### Human syntenic transcripts

Mouse lung lncRNA loci were mapped to syntenic human (GRCh37/hg19) regions using the LiftOver tool available on the University of California at Santa Cruz (UCSC) Genome Browser Web site. Human Ensembl version 75 was used to determine which genes were located in syntenic regions.

#### 5' and 3' RACE and full-length lncRNA cloning

5' RACE was performed with the TaKaRa 5' RACE core kit (Clontech) per the manufacturer's instructions using the PCR primers listed in Supplemental Table 8. The final reaction products were cloned and sequenced using the TOPO TA cloning kit and pCR4 vector (Life Technologies). Primer 1 (Supplemental Table 8) was used for first strand synthesis from total RNA extracted from MLE12 cells with Superscript II per the manufacturer's protocol. 3' RACE products were subsequently amplified with nested PCR primers (see Supplemental Table 8). These products were also subsequently cloned into the TOPO vector pCR4. Based on the results of 5' and 3' RACE, PCR primers were designed to amplify full-length transcripts for each lncRNA of interest. LA Taq (Clontech) was used for full-length transcript amplification. Each product was cloned into TOPO pCR4, sequenced, and then subcloned into the plasmid backbones as described in the Materials and Methods.

#### In vitro transcription/translation assay

The Promega TnT quick-coupled transcription/translation system and transcend nonradioactive translation detection system was used to in vitro transcribe and translate each full-length lncRNA from the T7 promoter of a pcDNA3.1<sup>+</sup> plasmid per the manufacturer's instructions. Following completion of the transcription/translation reaction, 1  $\mu$ L of the reaction product was added to 15  $\mu$ L of SDS sample buffer, heated for 2 min to 100°C to denature the proteins, and then loaded onto a 4%–12% 1-mm 15-W NuPage SDS-polyacrylamide gel (Life Technologies). Protein products labeled with the biotinylated Transcend tRNA were detected with streptavidin antibody and Western Blue reagent per the manufacturer's instructions. The luciferase T7 control DNA plasmid supplied with the TnT quick-coupled transcription/translation system was used as a positive control.

#### Animals

The generation and genotyping of the Nkx2.1<sup>-/-</sup>, Shh<sup>cre</sup>, and Ctnnb1<sup>fllox/fllox</sup> mice have been previously described (Brault et al. 2001; Harfe et al. 2004; Taniguchi et al. 2011, 2013). All animal work was performed under the approval and guidance of the University of Pennsylvania Institutional Animal Care and Use Committee.

#### Quantitative real-time PCR

Total RNA was isolated from the indicated tissues or cell lines using Trizol (Invitrogen) per the manufacturer's protocol. cDNA was synthesized from total RNA by using the SuperScript strand synthesis system (Invitrogen). Quantitative real-time PCR was performed using the SYBR Green system (Applied Biosystems) with primers listed in Supplemental Table 8. GAPDH expression values were used to control for RNA quality and quantity. Five biologic replicates were used for all cell line experiments, and

biologic triplicates were used for in vivo experiments unless otherwise noted. Data shown are the mean  $\pm$  SEM.

#### Histology

Tissues were fixed in 4% paraformaldehyde, embedded in paraffin wax, and sectioned at 5- $\mu$ m intervals. Immunohistochemistry was performed using the following antibodies: mouse anti-phospho-histone3 (1:200; Cell Signaling Technology), goat anti-CC10 (1:20; Santa Cruz Biotechnology), rabbit anti-SP-C (1:500; Chemicon), rabbit anti-Nkx2.1 (1:50; Santa Cruz Biotechnology), mouse anti-Cdkn1a (1:100; Santa Cruz Biotechnology), rabbit anti-Sox9 (1:100; Santa Cruz Biotechnology), rabbit anti-Rbl2/p130 (1:50; Abcam), rabbit anti-Par3 (1:100; Upstate Biotechnology), rabbit anti-Sox2 (1:500; Seven Hills Bioreagents), and rabbit anti-Cdh1 (1:100; Cell Signaling). Slides were mounted with Vectashield mounting medium containing DAPI (Vector Laboratories). ISH was performed as described (Zhang et al. 2008). Probes were amplified from each lncRNA using the primers listed in Supplemental Table 8.

#### Human lung sample collection

Neonatal lung samples were collected through an institutional review board (IRB)-approved registry biorepository, "Molecular and Clinical Analysis of Syndromic and Isolated Anomalies of the Foregut," at The Children's Hospital of Philadelphia. Fetal lung samples were collected through an IRB protocol at Temple University (Philadelphia, PA).

#### Cell culture methods

HEK293T, MLE12, and H441 cell lines obtained from American Type Culture Collection were used for all experiments. HEK293T cells were cultured in DMEM (Life Technologies) plus 10% FBS plus 1% penicillin/streptomycin at 37°C and 5% CO<sub>2</sub>. Antibiotic-free conditions were used during transfection. MLE12 cells were cultured in the HITES medium (Ham's F12 [Mediatech], 0.005 mg/mL insulin, 0.01 mg/mL transferrin, 0.01 mg/mL transferrin, 30 nM sodium selenite, 10 nM hydrocortisone, 10 nM  $\beta$ -estradiol, 10 mM HEPES, 2 mM L-glutamine) plus 2% FBS plus 1% penicillin/streptomycin. H441 cells were cultured in RPMI-1640 (Mediatech) plus 10% FBS plus 1% penicillin/streptomycin at 37°C and 5% CO<sub>2</sub>.

#### Lentiviral shRNA knockdown

Suitable target regions for each lncRNA were chosen based on the guidelines outlined by Birmingham et al. (2007). Five distinct regions of each target were chosen for shRNA design. Regions containing known SNPs or repeat elements were specifically avoided. A minimum of five shRNAs were designed against these regions using the RNAi Consortium's publically available design tool (<http://www.broadinstitute.org/rnai/public/seq/search>). Each target sequence was independently verified to be unique through a separate BLAST search. shRNAs were screened for efficiency; the shRNA construct with maximum knockdown was selected for subsequent analyses. Sequences for shRNAs and target sites can be located in Supplemental Table 9.

For knockdown in MLE12 cells, shRNA oligos were synthesized by Integrated DNA Technologies, phosphorylated, annealed to their complementary partner, and then cloned into the pLKO.1-GFP vector. Lentivirus was produced by transfecting HEK293T cells at 50%–70% confluency with the viral packaging plasmids psPAX2 and pMD2.G and the appropriate pLKO.1-shRNA plasmid with FuGENE6 (Promega). A total of 4  $\mu$ g of

DNA (2  $\mu\text{g}$  of pLKO.1-shRNA, 1.5  $\mu\text{g}$  of psPAX2, 500 ng of pMD2.G) plus 12  $\mu\text{L}$  of FuGENE 6 was used to transfect each well of a six-well plate for screening assays. A total of 10  $\mu\text{g}$  of DNA (5  $\mu\text{g}$  of pLKO.1-shRNA, 3.75  $\mu\text{g}$  of psPAX2, 1.25  $\mu\text{g}$  of pMD2.G) plus 30  $\mu\text{L}$  of FuGENE 6 was used to transfect 10-cm plates for subsequent experiments. Fresh medium was added 12 h after transfection. Virus was collected 48 h following transfection by passing the virus-containing medium through a 0.45- $\mu\text{m}$  strainer and then concentrated with Amicon 100k Ultracentrifugation filters (Millipore) by centrifugation at 2800g for 15 min. The viral concentrate was reconstituted in an equivalent volume of MLE12 medium (composition described above) and then applied to MLE12 cells with fresh medium in a 1:4 ratio with 8  $\mu\text{g}$  of polybrene per milliliter of medium. Fresh medium was added 12 h after addition of virus. MLE12 cells were cultured for 65 h after infection before collecting and isolating total RNA. A similar protocol was used to infect H441 cells, except the H441 cells were exposed to virus for 18 h and cultured in H441 medium (composition described above).

#### *Overexpression of NANCI and Nkx2.1 rescue of NANCI knockdown*

NANCI was cloned into pLenti 7.3/V5-DEST using the gateway cloning system. Lentivirus was produced in HEK293T cells and used to infect MLE12 cells at 50%–70% confluency, as described above. Infected MLE12 cells were grown for 65 h before collecting and isolating total RNA. Full-length Nkx2.1 was cloned into the pCMV-Tag2B vector. MLE12 cells were infected with NANCI shRNA lentivirus as described above and allowed to grow for 24 h. MLE12 cells were then transfected with a pCMV-Tag2B-Nkx2.1 vector using the Lonza 4D-Nucleofector system and allowed to grow for another 48 h before isolating RNA.

#### *Microarray analysis*

RNA was isolated from MLE12 cells following infection with either a scrambled control shRNA or lncRNA target shRNA in triplicate. Biotinylated cRNA probe libraries were generated from these RNA samples and used with Affymetrix Mouse Gene 2.0ST arrays. Microarray data were analyzed using the Oligo package available at the Bioconductor Web site (<http://www.bioconductor.org>). The raw data were background-corrected by the robust multichip average (RMA) method and then normalized by an invariant set method. Genes with 80% of samples with an expression signal above the negative control probes were considered detectable or present. Differential gene expression performed on the detected gene set between the control and mutant mice was analyzed by the Limma package available at the Bioconductor Web site. *P*-values obtained from the multiple comparison tests were corrected by false discovery rates. Heat map displays were created using R and the R library heatmap2. See Supplemental Tables 2 and 3 for the full list of genes that were significantly altered with a 1.4-fold change in expression. GO enrichment analysis using separate gene sets for up-regulated and down-regulated genes for each assay was compared against a background using the detected genes using the DAVID Web site (Huang et al. 2009) using separate gene sets (see Supplemental Tables 4–7). The Gene Expression Omnibus accession number for the microarray data is GSE52389.

#### *NANCI promoter motif analysis*

The promoter region of NANCI was searched for transcription factor motifs (TRANSFAC) of transcription factors with known

roles in foregut endoderm development using the search tool FIMO (Grant et al. 2011). A phastCons score of >0.75 in the 30-way placental mammal UCSC was used to classify a motif as being evolutionarily conserved.

#### *RNA coimmunoprecipitation*

RNA coimmunoprecipitation was carried out as previously described (Moran et al. 2012). MLE12 cells were grown to 90% confluency on 12 10-cm plates and then harvested following trypsinization. The cells were washed twice with sterile PBS and then cross-linked with 0.3% formaldehyde in PBS for 10 min at room temperature. The reaction was quenched with glycine at a final concentration of 0.125 M. Cells were then pelleted, washed with PBS twice, and resuspended in 1.2 mL of RIPA buffer supplemented with protease inhibitors (Complete protease inhibitor tablet [Roche] and RNase inhibitors (62.5  $\mu\text{L}$  of RNaseOUT plus 125  $\mu\text{L}$  of SuperaseIN [Life Technologies])). The cell suspension was incubated for 30 min at 37°C and, during this period, vortexed for 30 sec at 5-min intervals. The cells were then homogenized with a Dounce homogenizer for 1 min. The cell lysate was centrifuged to pellet the debris, and the supernatant was transferred to a clean tube. A 200- $\mu\text{L}$  aliquot was set aside as the input sample, and 500  $\mu\text{L}$  was used for each pull-down reaction (500  $\mu\text{L}$  for the IgG reaction; 500  $\mu\text{L}$  for the target antibody reaction). Dynabeads Protein G (Life Technologies) were used for immunoprecipitation following the manufacturer's instructions. The beads were resuspended by vortexing for 30 sec, and then 100- $\mu\text{L}$  aliquots were transferred to clean tubes and resuspended in RIPA buffer supplemented with protease and RNase inhibitors. Two micrograms of target antibody (either anti-Ezh2 [Millipore] or anti-Wdr5 [Bethyl Laboratories, Inc.]) was added to the first tube, and 2  $\mu\text{g}$  of polyclonal IgG antibody was added to the second tube. The antibody bead mixture was incubated for 2 h at 4°C with shaking. The beads were then washed with 500  $\mu\text{L}$  of RIPA buffer, and resuspended in 500  $\mu\text{L}$  of MLE12 cell lysate. The antibody-bead-cell lysate mixture was incubated for 2 h at 4°C with shaking. Next, the beads were washed with high-salt RIPA buffer (1 M NaCl, 1% NP-40, 0.5% sodium deoxycholate, 0.1% SDS, 50 mM Tris-HCl, 1 mM EDTA). Ten micrograms of proteinase K was added to each reaction mixture and incubated for 30 min at 42°C, followed by incubation for 4 h at 65°C to reverse the formaldehyde cross-links and free the bound RNA. RNA was then isolated from this sample as well as the input using Trizol in a standard fashion per the manufacturer's protocol. cDNA was synthesized as described above with random hexamers, and cDNA content was quantified by qPCR (see Supplemental Table 8 for qPCR primer sequences). U1 rRNA was used as an internal control, Hmbs was used as a negative control, and LL12/Snhg5 was used as a positive control for the EZH2 pull-down (Zhao et al. 2010; Grote et al. 2013).

#### *Generation of NANCI shRNA knockdown transgenics*

The 3.7-kb human SFTPC promoter transgenic expression vector has been described previously (Yang et al. 2002; Li et al. 2012). The NANCI shRNA construct used above was used to generate an shRNA-miRNA cassette construct, as described previously (Wang et al. 2007). This was cloned downstream from the human SFTPC promoter and the SV40 polyadenylation sequence, excised from the resulting plasmid, purified, and injected into FVBN fertilized oocytes as previously described (Yang et al. 2002; Li et al. 2012). Embryos were collected at E18.5 and genotyped by standard PCR (see Supplemental Table 8). Whole lungs were

isolated; the left lung was fixed for histology, and the right lung was used to isolate total RNA.

#### Mesenchymal thickness calculations

H&E staining was performed on SFTPC-NANCI<sup>shRNA</sup> transgenic lungs, Nkx2.1<sup>+/-</sup> lungs, and littermate control lungs for both genotypes. For each lung section, two pictures were taken at 200× near the top and middle of the lung section. These pictures were overlaid with a uniform grid. Two horizontal lines and two vertical lines were chosen at random from each grid. Along a uniform length of each line, the mesenchymal length and number of intersecting airways were measured; and these values were then summed across all four lines. The total mesenchymal length was then divided by the total number of intersecting airways to obtain the average mesenchymal thickness for that picture. The mesenchymal thicknesses of the two pictures were then averaged to get the mesenchymal thickness for a lung section. The number of lung sections analyzed was five SFTPC-NANCI<sup>shRNA</sup> transgenic lungs and control littermates, six Nkx2.1<sup>+/-</sup> lungs, or four Nkx2.1<sup>+/+</sup> lungs.

#### Acknowledgments

Studies in the Morrisey laboratory were supported by funding from the National Institutes of Health (NIH; HL100405 and HL110942). D.T.S. received funding from NIH T32 GM008638 and a Thrasher Foundation Early Career Award. M.J.H received funding from NIH T32-HD007516.

#### References

- Bailey TL, Elkan C. 1994. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol* **2**: 28–36.
- Balmer JE, Blomhoff R. 2002. Gene expression regulation by retinoic acid. *J Lipid Res* **43**: 1773–1808.
- Barnett CP, Mencil JJ, Gecz J, Waters W, Kirwin SM, Vnette KM, Uppill M, Nicholl J. 2012. Choroathetosis, congenital hypothyroidism and neonatal respiratory distress syndrome with intact NKX2-1. *Am J Med Genet A* **158A**: 3168–3173.
- Bayha E, Jorgensen MC, Serup P, Grapin-Botton A. 2009. Retinoic acid signaling organizes endodermal organ specification along the entire antero-posterior axis. *PLoS ONE* **4**: e5845.
- Berghoff EG, Clark MF, Chen S, Cajigas I, Leib DE, Kohtz JD. 2013. Evf2 (Dlx6as) lncRNA regulates ultraconserved enhancer methylation and the differential transcriptional control of adjacent genes. *Development* **140**: 4407–4416.
- Birmingham A, Anderson E, Sullivan K, Reynolds A, Boese Q, Leake D, Karpilow J, Khvorova A. 2007. A protocol for designing siRNAs with high functionality and specificity. *Nat Protoc* **2**: 2068–2078.
- Braut V, Moore R, Kutsch S, Ishibashi M, Rowitch DH, McMahon AP, Sommer L, Boussadia O, Kemler R. 2001. Inactivation of the  $\beta$ -catenin gene by Wnt1-Cre-mediated deletion results in dramatic brain malformation and failure of craniofacial development. *Development* **128**: 1253–1264.
- Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-Vega B, Regev A, Rinn JL. 2011. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev* **25**: 1915–1927.
- Chen F, Desai TJ, Qian J, Niederreither K, Lü J, Cardoso WV. 2007. Inhibition of Tgf $\beta$  signaling by endogenous retinoic acid is essential for primary lung bud induction. *Development* **134**: 2969–2979.
- Chen F, Cao Y, Qian J, Shao F, Niederreither K, Cardoso WV. 2010. A retinoic acid-dependent network in the foregut controls formation of the mouse lung primordium. *J Clin Invest* **120**: 2040–2048.
- Chokas AL, Trivedi CM, Lu MM, Tucker PW, Li S, Epstein JA, Morrisey EE. 2010. Foxp1/2/4–NuRD interactions regulate gene expression and epithelial injury response in the lung via regulation of interleukin-6. *J Biol Chem* **285**: 13304–13313.
- Clark MB, Johnston RL, Inostroza-Ponta M, Fox AH, Fortini E, Moscato P, Dinger ME, Mattick JS. 2012. Genome-wide analysis of long noncoding RNA stability. *Genome Res* **22**: 885–898.
- Das A, Acharya S, Gottipati KR, McKnight JB, Chandru H, Alcorn JL, Boggaram V. 2011. Thyroid transcription factor-1 (TTF-1) gene: identification of ZBP-89, Sp1, and TTF-1 sites in the promoter and regulation by TNF- $\alpha$  in lung epithelial cells. *Am J Physiol Lung Cell Mol Physiol* **301**: L427–L440.
- Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, Guernec G, Martin D, Merkel A, Knowles DG, et al. 2012. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res* **22**: 1775–1789.
- Dieci G, Preti M, Montanini B. 2009. Eukaryotic snoRNAs: a paradigm for gene expression flexibility. *Genomics* **94**: 83–88.
- Du Z, Fei T, Verhaak RG, Su Z, Zhang Y, Brown M, Chen Y, Liu XS. 2013. Integrative genomic analyses reveal clinically relevant long noncoding RNAs in human cancer. *Nat Struct Mol Biol* **20**: 908–913.
- Favorov A, Mularoni L, Cope LM, Medvedeva Y, Mironov AA, Makeev VJ, Wheelan SJ. 2012. Exploring massive, genome scale datasets with the GenometriCorr package. *PLoS Comput Biol* **8**: e1002529.
- Goss AM, Tian Y, Tsukiyama T, Cohen ED, Zhou D, Lu MM, Yamaguchi TP, Morrisey EE. 2009. Wnt2/2b and  $\beta$ -catenin signaling are necessary and sufficient to specify lung progenitors in the foregut. *Dev Cell* **17**: 290–298.
- Grant CE, Bailey TL, Noble WS. 2011. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**: 1017–1018.
- Grote P, Wittler L, Hendrix D, Koch F, Wahrisch S, Beisaw A, Macura K, Blass G, Kellis M, Werber M, et al. 2013. The tissue-specific lncRNA *Fendrr* is an essential regulator of heart and body wall development in the mouse. *Dev Cell* **24**: 206–214.
- Gupta S, Stamatoyannopoulos JA, Bailey TL, Noble WS. 2007. Quantifying similarity between motifs. *Genome Biol* **8**: R24.
- Harfe BD, Scherz PJ, Nissim S, Tian H, McMahon AP, Tabin CJ. 2004. Evidence for an expansion-based temporal Shh gradient in specifying vertebrate digit identities. *Cell* **118**: 517–528.
- Harris-Johnson KS, Domyan ET, Vezina CM, Sun X. 2009.  $\beta$ -Catenin promotes respiratory progenitor identity in mouse foregut. *Proc Natl Acad Sci* **106**: 16287–16292.
- Huang DW, Sherman BT, Lempicki RA. 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**: 44–57.
- Khalil AM, Guttman M, Huarte M, Garber M, Raj A, Rivea Morales D, Thomas K, Presser A, Bernstein BE, van Oudenaarden A, et al. 2009. Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc Natl Acad Sci* **106**: 11667–11672.
- Kimura S, Hara Y, Pineau T, Fernandez-Salguero P, Fox CH, Ward JM, Gonzalez FJ. 1996. The T/ebp null mouse: thyroid-specific enhancer-binding protein is essential for the organogenesis of the thyroid, lung, ventral forebrain, and pituitary. *Genes Dev* **10**: 60–69.



- Kretz M, Webster DE, Flockhart RJ, Lee CS, Zehnder A, Lopez-Pajares V, Qu K, Zheng GX, Chow J, Kim GE, et al. 2012. Suppression of progenitor differentiation requires the long noncoding RNA ANCR. *Genes Dev* **26**: 338–343.
- Lai MC, Yang Z, Zhou L, Zhu QQ, Xie HY, Zhang F, Wu LM, Chen LM, Zheng SS. 2012. Long non-coding RNA MALAT-1 overexpression predicts tumor recurrence of hepatocellular carcinoma after liver transplantation. *Med Oncol* **29**: 1810–1816.
- Li S, Wang Y, Zhang Y, Lu MM, DeMayo FJ, Dekker JD, Tucker PW, Morrisey EE. 2012. Foxp1/4 control epithelial cell fate during lung development and regeneration through regulation of anterior gradient 2. *Development* **139**: 2500–2509.
- Lin MF, Jungreis I, Kellis M. 2011. PhyloCSF: a comparative genomics method to distinguish protein coding and non-coding regions. *Bioinformatics* **27**: i275–i282.
- Longmire TA, Ikonomou L, Hawkins F, Christodoulou C, Cao Y, Jean JC, Kwok LW, Mou H, Rajagopal J, Shen SS, et al. 2012. Efficient derivation of purified lung and thyroid progenitors from embryonic stem cells. *Cell Stem Cell* **10**: 398–411.
- Lu Y, Thomson JM, Wong HY, Hammond SM, Hogan BL. 2007. Transgenic over-expression of the microRNA miR-17-92 cluster promotes proliferation and inhibits differentiation of lung epithelial progenitor cells. *Dev Biol* **310**: 442–453.
- Luteijn MJ, Ketting RF. 2013. PIWI-interacting RNAs: from generation to transgenerational epigenetics. *Nat Rev Genet* **14**: 523–534.
- Martin M, Gallego-Llamas J, Ribes V, Keding M, Niederreither K, Chambon P, Dollé P, Gradwohl G. 2005. Dorsal pancreas agenesis in retinoic acid-deficient Raldh2 mutant mice. *Dev Biol* **284**: 399–411.
- Matys V, Kel-Margoulis OV, Fricke E, Liebich I, Land S, Barre-Dirrie A, Reuter I, Chekmenov D, Krull M, Hornischer K, et al. 2006. TRANSFAC and its module TRANSCOMP: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res* **34**: D108–D110.
- Minoo P, Hamdan H, Bu D, Warburton D, Stepanik P, deLemos R. 1995. TTF-1 regulates lung epithelial morphogenesis. *Dev Biol* **172**: 694–698.
- Minoo P, Li C, Liu HB, Hamdan H, deLemos R. 1997. TTF-1 is an epithelial morphoregulatory transcriptional factor. *Chest* **111**: 135S–137S.
- Minoo P, Su G, Drum H, Bringas P, Kimura S. 1999. Defects in tracheoesophageal and lung morphogenesis in Nkx2.1<sup>-/-</sup> mouse embryos. *Dev Biol* **209**: 60–71.
- Moran VA, Niland CN, Khalil AM. 2012. Co-immunoprecipitation of long noncoding RNAs. *Methods Mol Biol* **925**: 219–228.
- Morrisey EE, Ip HS, Lu MM, Parmacek MS. 1996. GATA-6: a zinc finger transcription factor that is expressed in multiple cell lineages derived from lateral mesoderm. *Dev Biol* **177**: 309–322.
- Mou H, Zhao R, Sherwood R, Ahfeldt T, Lapey A, Wain J, Sicilian L, Izvolsky K, Musunuru K, Cowan C, et al. 2012. Generation of multipotent lung and airway progenitors from mouse ESCs and patient-specific cystic fibrosis iPSCs. *Cell Stem Cell* **10**: 385–397.
- Nakamura Y, Takahashi N, Kakegawa E, Yoshida K, Ito Y, Kayano H, Niitsu N, Jinnai I, Bessho M. 2008. The GAS5 (growth arrest-specific transcript 5) gene fuses to BCL6 as a result of t(1;3)(q25;q27) in a patient with B-cell lymphoma. *Cancer Genet Cytogenet* **182**: 144–149.
- Orom UA, Derrien T, Beringer M, Gumireddy K, Gardini A, Bussotti G, Lai F, Zytnicki M, Notredame C, Huang Q, et al. 2010a. Long noncoding RNAs with enhancer-like function in human cells. *Cell* **143**: 46–58.
- Orom UA, Derrien T, Guigo R, Shiekhattar R. 2010b. Long noncoding RNAs as enhancers of gene expression. *Cold Spring Harb Symp Quant Biol* **75**: 325–331.
- Parris T, Nik AM, Kotecha S, Langston C, Helou K, Platt C, Carlsson P. 2013. Inversion upstream of FOXF1 in a case of lethal alveolar capillary dysplasia with misalignment of pulmonary veins. *Am J Med Genet A* **161A**: 764–770.
- Peng T, Tian Y, Boogerd CJ, Lu MM, Kadzik RS, Stewart KM, Evans SM, Morrisey EE. 2013. Coordination of heart and lung co-development by a multipotent cardiopulmonary progenitor. *Nature* **500**: 589–592.
- Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841–842.
- Rosenbloom KR, Sloan CA, Malladi VS, Dreszer TR, Learned K, Kirkup VM, Wong MC, Maddren M, Fang R, Heitner SG, et al. 2013. ENCODE data in the UCSC Genome Browser: year 5 update. *Nucleic Acids Res* **41**: D56–D63.
- Sauvageau M, Goff LA, Lodato S, Bonev B, Groff AF, Gerhardinger C, Sanchez-Gomez DB, Hanculeyman E, Li E, Spence M, et al. 2013. Multiple knockout mouse models reveal lincRNAs are required for life and brain development. *eLife* **2**: e01749.
- Schmidt LH, Spieker T, Koschmieder S, Schaffers S, Humberg J, Jungen D, Bulk E, Hascher A, Wittmer D, Marra A, et al. 2011. The long noncoding MALAT-1 RNA indicates a poor prognosis in non-small cell lung cancer and induces migration and tumor growth. *J Thorac Oncol* **6**: 1984–1992.
- Shaw-White JR, Bruno MD, Whitsett JA. 1999. GATA-6 activates transcription of thyroid transcription factor-1. *J Biol Chem* **274**: 2658–2664.
- Shu W, Lu MM, Zhang Y, Tucker PW, Zhou D, Morrisey EE. 2007. Foxp2 and Foxp1 cooperatively regulate lung and esophagus development. *Development* **134**: 1991–2000.
- Stankiewicz P, Sen P, Bhatt SS, Storer M, Xia Z, Bejjani BA, Ou Z, Wiszniewska J, Driscoll DJ, Maisenbacher MK, et al. 2009. Genomic and genic deletions of the FOX gene cluster on 16q24.1 and inactivating mutations of FOXF1 cause alveolar capillary dysplasia and other malformations. *Am J Hum Genet* **84**: 780–791.
- Szafranski P, Dharmadhikari AV, Brosens E, Gurha P, Kolodziejska KE, Zhishuo O, Dittwald P, Majewski T, Mohan KN, Chen B, et al. 2013. Small noncoding differentially methylated copy-number variants, including lincRNA genes, cause a lethal lung developmental disorder. *Genome Res* **23**: 23–33.
- Tanaka R, Satoh H, Moriyama M, Satoh K, Morishita Y, Yoshida S, Watanabe T, Nakamura Y, Mori S. 2000. Intronic U50 small-nucleolar-RNA (snRNA) host gene of no protein-coding potential is mapped at the chromosome breakpoint t(3;6)(q27;q15) of human B-cell lymphoma. *Genes Cells* **5**: 277–287.
- Taniguchi H, He M, Wu P, Kim S, Paik R, Sugino K, Kvitsiani D, Fu Y, Lu J, Lin Y, et al. 2011. A resource of Cre driver lines for genetic targeting of GABAergic neurons in cerebral cortex. *Neuron* **71**: 995–1013.
- Taniguchi H, Lu J, Huang ZJ. 2013. The spatial and temporal origin of chandelier cells in mouse neocortex. *Science* **339**: 70–74.
- Tian Y, Zhang Y, Hurd L, Hannehalli S, Liu F, Lu MM, Morrisey EE. 2011. Regulation of lung endoderm progenitor cell behavior by miR302/367. *Development* **138**: 1235–1245.
- Tilgner H, Knowles DG, Johnson R, Davis CA, Chakraborty S, Djebali S, Curado J, Snyder M, Gingeras TR, Guigo R. 2012. Deep sequencing of subcellular RNA fractions shows splicing

- to be predominantly co-transcriptional in the human genome but inefficient for lncRNAs. *Genome Res* **22**: 1616–1625.
- Trapnell C, Hendrickson DG, Sauvageau M, Goff L, Rinn JL, Pachter L. 2013. Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat Biotechnol* **31**: 46–53.
- Ulitsky I, Bartel DP. 2013. lincRNAs: genomics, evolution, and mechanisms. *Cell* **154**: 26–46.
- Ventura A, Young AG, Winslow MM, Lintault L, Meissner A, Erkeland SJ, Newman J, Bronson RT, Crowley D, Stone JR, et al. 2008. Targeted deletion reveals essential and overlapping functions of the miR-17 through 92 family of miRNA clusters. *Cell* **132**: 875–886.
- Wan H, Kaestner KH, Ang SL, Ikegami M, Finkelman FD, Stahlman MT, Fulkerson PC, Rothenberg ME, Whitsett JA. 2004a. Foxa2 regulates alveolarization and goblet cell hyperplasia. *Development* **131**: 953–964.
- Wan H, Xu Y, Ikegami M, Stahlman MT, Kaestner KH, Ang SL, Whitsett JA. 2004b. Foxa2 is required for transition to air breathing at birth. *Proc Natl Acad Sci* **101**: 14449–14454.
- Wan H, Dingle S, Xu Y, Besnard V, Kaestner KH, Ang SL, Wert S, Stahlman MT, Whitsett JA. 2005. Compensatory roles of Foxa1 and Foxa2 during lung morphogenesis. *J Biol Chem* **280**: 13809–13816.
- Wang Z, Dolle P, Cardoso WV, Niederreither K. 2006. Retinoic acid regulates morphogenesis and patterning of posterior foregut derivatives. *Dev Biol* **297**: 433–445.
- Wang J, Theunissen TW, Orkin SH. 2007. Site-directed, virus-free, and inducible RNAi in embryonic stem cells. *Proc Natl Acad Sci* **104**: 20850–20855.
- Wang KC, Yang YW, Liu B, Sanyal A, Corces-Zimmerman R, Chen Y, Lajoie BR, Protacio A, Flynn RA, Gupta RA, et al. 2011. A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature* **472**: 120–124.
- Wapinski O, Chang HY. 2011. Long noncoding RNAs and human disease. *Trends Cell Biol* **21**: 354–361.
- Wu TD, Nacu S. 2010. Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics* **26**: 873–881.
- Xu C, Yang M, Tian J, Wang X, Li Z. 2011. MALAT-1: a long non-coding RNA and its important 3' end functional motif in colorectal cancer metastasis. *Int J Oncol* **39**: 169–175.
- Yang H, Lu MM, Zhang L, Whitsett JA, Morrisey EE. 2002. GATA6 regulates differentiation of distal lung epithelium. *Development* **129**: 2233–2246.
- Yates LA, Norbury CJ, Gilbert RJ. 2013. The long and short of microRNA. *Cell* **153**: 516–519.
- Yu S, Shao L, Kilbride H, Zwick DL. 2010. Haploinsufficiencies of FOXF1 and FOXC2 genes associated with lethal alveolar capillary dysplasia and congenital heart disease. *Am J Med Genet A* **152A**: 1257–1262.
- Zhang Y, Goss AM, Cohen ED, Kadzik R, Lepore JJ, Muthukumaraswamy K, Yang J, DeMayo FJ, Whitsett JA, Parmacek MS, et al. 2008. A Gata6–Wnt pathway required for epithelial stem cell development and airway regeneration. *Nat Genet* **40**: 862–870.
- Zhang B, Arun G, Mao YS, Lazar Z, Hung G, Bhattacharjee G, Xiao X, Booth CJ, Wu J, Zhang C, et al. 2012a. The lncRNA Malat1 is dispensable for mouse development but its transcription plays a cis-regulatory role in the adult. *Cell Reports* **2**: 111–123.
- Zhang HM, Chen H, Liu W, Liu H, Gong J, Wang H, Guo AY. 2012b. AnimalTFDB: a comprehensive animal transcription factor database. *Nucleic Acids Res* **40**: D144–D149.
- Zhao J, Sun BK, Erwin JA, Song JJ, Lee JT. 2008. Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science* **322**: 750–756.
- Zhao J, Ohsumi TK, Kung JT, Ogawa Y, Grau DJ, Sarma K, Song JJ, Kingston RE, Borowsky M, Lee JT. 2010. Genome-wide identification of polycomb-associated RNAs by RIP-seq. *Mol Cell* **40**: 939–953.
- Zhou B, Zhong Q, Minoo P, Li C, Ann DK, Frenkel B, Morrisey EE, Crandall ED, Borok Z. 2008. Foxp2 inhibits Nkx2.1-mediated transcription of SP-C via interactions with the Nkx2.1 homeodomain. *Am J Respir Cell Mol Biol* **38**: 750–758.
- Zhu L, Xu PC. 2013. Downregulated lncRNA-ANCR promotes osteoblast differentiation by targeting EZH2 and regulating Runx2 expression. *Biochem Biophys Res Commun* **432**: 612–617.