Gene expression

Computational prediction of novel components of lung transcriptional networks

M. Juanita Martinez, Andrew D. Smith[1], Bilan Li, Michael Q. Zhang[1] and Kevin S. Harrod[*]

Lovelace Respiratory Research Institute, 2425 Ridgecrest Dr SE, Albuquerque, NM 87108 and [1]Cold Spring Harbor Laboratory, One Bungtown Road, Cold Spring Harbor, NY 11724, USA

ABSTRACT

Motivation: Little is known regarding the transcriptional mechanisms involved in forming and maintaining epithelial cell lineages of the mammalian respiratory tract.

Results: Herein, a motif discovery approach was used to identify novel transcriptional regulators in the lung using genes previously found to be regulated by Foxa2 or Wnt signaling pathways. A human–mouse comparison of both novel and known motifs was also performed. Some of the factors and families identified here were previously shown to be involved epithelial cell differentiation (ETS family, HES-1 and MEIS-1), and ciliogenesis (RFX family), but have never been characterized in lung epithelia. Other unidentified over-represented motifs suggest the existence of novel mammalian lung transcription factors. Of the fraction of motifs examined we describe 25 transcription factor family predictions for lung. Fifteen novel factors were shown here to be expressed in mouse lung, and/or human bronchial or distal lung epithelial tissues or lung epithelial cell lineages.

Availability: DME: http://rulai.cshl.edu/dme. MATCOMPARE: http://rulai.cshl.edu/MatCompare. MOTIFCLASS is available from the authors.

Contact: kharrod@lrri.org

Supplementary information: http://www.lrri.org/Martinez2006motifs Supplement/ and *Bioinformatics* Online.

1 INTRODUCTION

The lung is a complex organ consisting of a network of branching tubules culminating into a vascularized gas exchange tissue. Highly specialized, differentiated epithelial cells of the lung originate from common progenitors during development to provide unique functions related to host defense, gas exchange and ion transport (Perl *et al*., 2002a). This large number of specialized functions suggests that there are unique mechanisms of gene expression in lung epithelia during and following differentiation. Indeed, a transcriptional module of cooperating factors from the transcription factor families Nkx, Fox and C/EBP is hypothesized to be essential for regulating lung-specific surfactant protein and secretoglobin protein gene expression (Whitsett, 2002).

Recent advances have allowed for the computational discovery of putative *cis*-regulatory elements that are over-represented in promoters of co-regulated genes. Computational classification of both known and putative *cis*-regulatory elements has been previously performed to identify tissue-specific transcriptional mechanisms in tissue selective gene sets (Frech *et al*., 1998; Nelander *et al*., 2003), as well as genome wide (Nelander *et al*., 2005; Smith *et al*., 2006), using a variety of methods. The Discriminating Matrix Enumerator (DME) algorithm (Smith *et al*., 2005) is useful for identifying motifs that are over-represented in promoters of co-regulated genes (foreground) relative to other promoters of the genome (background). For a relatively large and complex organ, such as the lung, which contains an estimated 40–60 different cell types, obtaining collections of co-regulated genes within specific epithelial cell types has been challenging. Recently, motif discovery analyses utilizing co-regulated genes from whole lung tissue has produced very few novel lung factors (Nelander *et al*., 2005).

Here, we used lung epithelial gene expression coinciding with two experimentally validated gene regulatory pathways in the lung epithelia, Foxa2 and Wnt, to discover motifs and predict novel transcription factors of the lung epithelia. The motif discovery algorithm DME and the programs MATCOMPARE (Schones *et al*., 2005) and MOTIFCLASS (Smith *et al*., 2006) were used to identify and analyze both novel and known transcription factor binding motifs in promoters of lung gene subsets known to be responsive to Foxa2 transactivation during lung embryogenesis or Wnt-mediated lung cell differentiation (Okubo and Hogan 2004; Wan *et al*., 2004). Furthermore, experimental validation of positive expression of newly elucidated transcription factors indicated a high degree of success in predicting regulators of lung transcriptional networks. These findings provide a framework for future identification of novel transcriptional regulators in unique biological states of the lung, such as development or disease.

2 METHODS

2.1 Selection of promoter sequences

Proximal promoters of 24 genes down-regulated in mouse Foxa2$^{-/-}$ fetal lung and 17 specialized lung cell-marker genes down-regulated after constitutive activation of the Wnt pathway in mouse lung were obtained from previously published work (Okubo and Hogan 2004; Wan *et al*., 2004). The promoters were acquired from mouse genome build 34 (v.6, March 2005) and human genome build 35 (v.17, May 2004) through UCSC genome browser (Karolchik *et al*., 2003). For each co-regulated gene, the proximal promoter was defined as 1000 bases upstream and 200 bases downstream of

---

[*]To whom correspondence should be addressed.

the 5′ end of a known Refseq cDNA (Pruitt *et al*., 2005). Two hundred bases downstream of the 5′ end were included to accommodate for the possible existence of additional *cis*-regulatory elements. These mouse (m) promoters were compiled into two sets of sequences and named $mFox_{fg}$, and $mWnt_{fg}$, which were used as the foreground sequence sets for discovering over-represented motifs relative to the genome (background).

Human homologues of the mouse co-regulated genes were determined using sequence alignments from the UCSC genome browser as well as homologue annotation from both GeneLynx (www.genelynx.org) and Ensembl (www.ensembl.org) databases. The genomic sequence upstream of the human homologue to the mouse gene was assumed to be the human homologous promoter. The human homologous promoters of $mFox_{fg}$ and $mWnt_{fg}$ were compiled into the foreground sequence sets named $hFox_{fg}$ and $hWnt_{fg}$. Two mouse genes of the $mFox_{fg}$, Aox3 and Retnla, had no definitive human homologues. Similarly for $mWnt_{fg}$, no definitive homologues were found for Chi3l3 or Retnla, and were not included. The promoter for human LYZ was considered the homologue for both mouse Lys-s and Lyz and, therefore, was included twice in the human foregrounds to represent both genes (for sequences, see Supplementary material).

2.2 Discovery of novel motifs

The algorithms DME v. 1.44 (Smith *et al*., 2005) and DME–B (Smith *et al*., 2006) were used to discover over-represented motifs of the foregrounds described above relative to background sequence sets containing 1000 promoters randomly selected from the remaining genome for each organism. Under-represented motifs of the mouse promoters were also discovered (for detailed methods, see Supplementary material).

The most over-represented motifs of a foreground relative to its background promoter set were those with the lowest reported relative error-rate (a.k.a. classification error) (Smith *et al*., 2006), as calculated by MOTIFCLASS v. 1.20. The relative error-rates of all novel motifs were calculated and sorted, and the novel motifs with the 10 lowest relative error-rates (10 most over-represented) were reported. The relative error-rates for these motifs were significantly lower $P < 0.01$ on these promoter sets relative to random promoters.

MOTIFCLASS first scores the motif at every position in each promoter (Stormo, 2000; Wasserman and Sandelin, 2004). The highest scoring site for each promoter was determined and the score assigned to that promoter. The relative error-rate is the average of the false-positive and false-negative rates and accounts for the relative size difference of the foreground and background sequence sets for a fixed threshold score. The threshold is optimized to give the lowest possible relative error-rate for the motif, and the lowest relative error-rate is reported. The novel motifs were then compared to the TRANSFAC Professional v. 9.3 matrix library using MATCOMPARE v. 1.10 (Schones *et al*., 2005).

2.3 Cross-species comparison of motifs

Novel motifs with a relative error-rate of 0.300 or less were mutually compared between mice and human for each Fox_{fg} and Wnt_{fg} set using MATCOMPARE v. 1.10 (see Supplementary material). To compare over-representation of known motifs between mouse and human promoters, the relative error-rates of all matrices of TRANSFAC v. 9.3 were calculated with respect to mouse and human, Fox_{fg} and Wnt_{fg}, and then sorted. The factor names, not accession numbers, associated with each motif were first used to compare the known over-represented motifs between the mouse and human to accommodate for multiple matrices that exist for the same factor in TRANSFAC. Next, the sequences of compared motifs were examined and those motifs corresponding to multiple members of the same family having similar sequence were retained (e.g. GATA-3 compared to GATA-4). Motifs with relative error-rates less than 0.38 for Fox_{fg} and 0.34 for Wnt_{fg} were selected to visualize less than 10 pairs. Motifs beyond these cutoffs are available in the Supplementary material. For each novel and known motif shown here, the statistical significance of the relative error-rate was determined by calculating a distribution of 1000 relative error-rates based on foregrounds and backgrounds randomly sampled from the genome.

2.4 RT–PCR analysis of lung cell lines and tissue

Human lung carcinoma cell line A549 (American Type Cell Culture [ATCC]# CCL-185) and human lung papillary adenocarcinoma cell line NCI-H441 (ATCC# HTB-174) were grown in accordance with ATCC guidelines. Adult mouse whole lung (BALB/C and FVB/N) tissues were dissected from the apical and cardiac lobes. The tissue was snap-frozen in RNAlater (Ambion, Inc., Austin, TX) and stored at −70°C overnight. Mouse C22 cells, a generous gift from Daphne Demello, St. Louis University, MO, were grown at 33°C in permissive media until confluent as described (Demello *et al*., 2002). The cells were then passaged in nonpermissive media (without IFN-γ) and grown at 39°C for 4 days, with media changed every other day. MLE-15 cells were a generous gift from Jeffrey Whitsett, Children's Hospital, Cincinnati, OH.

Total RNA of all mouse and human tissue and cell lines was prepared using Tri-Reagent® (Molecular Research Center Inc., Cincinnati, OH) according to the manufacturer's protocol. Primers were designed to selectively amplify cDNA and not genomic DNA. Reverse transcription on 2 μg total RNA was performed with MMLV reverse transcriptase (Invitrogen, Carlsbad, CA) in the presence of RNase inhibitor (Roche Diagnostics, Indianapolis, IN) at 37°C for 50 min, then at 70°C for 15 min. PCR primers were designed with PrimerQuest (http://scitools.idtdna.com/Primerquest/). PCR conditions for all human genes and cDNAs (except A549) were performed using *Taq* polymerase (Promega Corp., Madison, WI) in standard reaction conditions as follows: 94°C for 2 min, and 29 cycles of (55°C for 30 s, 68°C for 30 s or 1 min, 94°C for 30 s), and a final extension at 72°C for 5 min. All mouse experiments and human A549 experiments were performed with GoTaq (Promega) according to manufacturer's protocol. cDNA synthesis was monitored with PCR amplification of β-actin. Negative controls were included in which no transcriptase or no template were added to the reaction mixture. Some of the human gene amplified fragments were sequenced to confirm their identity.

3 RESULTS

3.1 Co-regulated promoters in the lung epithelia

The genes co-regulated by Foxa2 or Wnt signaling pathways in mouse lung were obtained from Wan *et al*., (2004) and Okubo and Hogan (2004), respectively (See Supplementary material). These gene sets contain many previously characterized epithelial cell-marker genes. The proximal promoters of these genes ($mFox_{fg}$ and $mWnt_{fg}$) and their human homologues ($hFox_{fg}$ and $hWnt_{fg}$, respectively) were used to identify over-represented motifs discovered *de novo* or from known *cis*-regulatory elements (see Methods). Under-represented motifs were also identified by DME. Furthermore, a comparison with a second motif discovery algorithm, BioProspector (Liu *et al*., 2001) was used to determine similarities in findings (Supplementary material).

3.2 Novel motif discovery

To search for novel *cis*-regulatory elements in co-expressed lung genes, DME was used to identify 8–12 base motifs that were over-represented among each of the mouse and human foreground promoters relative to a background of 1000 randomly selected promoters for that genome. Under-represented motifs were discovered by switching the background and foreground promoter sets. Subsequent analysis of these motifs was performed using MATCOMPARE and MOTIFCLASS (See Methods and Supplementary material).

Table 1. Top 10 DME novel motifs from mouse Fox$_{fg}$

Motif ID	Logo	Number of sites fg	Number of sites bg[a]	R.E.[b]	Comparison to TRANSFAC (MATCOMPARE) Matrix[c]	Comparison to TRANSFAC (MATCOMPARE) Alignment[d]
E-9-70		17	2	0.212	M00280:RFX1	HGHAGSAAC | |:| ||| NNVNNRHNHVGNAACNN
E-11-42		20	5	0.228	M00340:ETS2	KRSAVRARRCA ::||..|.|:: DVSAGGAARYVNBH
B-10-1		20	4	0.228	M00743:CETS168	ADGCAGRRRG |:|...| CVGGAWGY
E-10-1		26	7	0.231	M00340:ETS2	ARDCAGRRRG :.:||....: DVSAGGAARYVNBH
E-11-2		18	4	0.238	M00450:ZIC3	ARKCAGGRRRR ||.: NGGGDNNNN
E-10-15		28	8	0.248	M01004:HELIOSA	AGRAAHDARK :|.|:|. NNNWGGAWHAN
E-10-33		26	6	0.252	M00971:ETS	YTSCYDCWTK |:.|.. NYWTCCBN
E-10-13		22	4	0.259	M00009:TTK69[e]	YSMTCMTGSY ..||.||| GSTCCTGS
B-10-7		13	2	0.263	M00238:BARBIE	RAAGCYAVRG .||| : NNNNAAAGNNNNNDN
E-10-34		20	5	0.263	M00971:ETS	WYCCTVWTBT |.||: NYWTCCBN

A colour version of this table is available as supplementary data.

[a]Corrected based on relative size of background sequence.

[b]Significantly lower ($P < 0.001$) for this foreground set relative to random promoters.

[c]Based on TRANSFAC v. 9.3; Vertebrate matrices only unless noted.

[d]Top sequence = novel motif; lower sequence = TRANSFAC motif; (|) = exact match; (:) = strong similarity; (.) = weak similarity; (x) = no match; ( ) = no comparison.

[e]*Drosophila Tramtrack-69*.

Over-representation of a motif was determined by its relative error-rate (see Methods). Ten motifs shown to be the most over-represented in the mFox$_{fg}$ and mWnt$_{fg}$ by having the lowest relative error-rate are presented in Tables 1 and 2, respectively. Under-represented motifs (over-represented in background promoters) on the mouse promoters are shown in the Supplementary material. Approximately 60 over-represented motifs in mFox$_{fg}$ and approximately 170 over-represented motifs in mWnt$_{fg}$ foreground promoters were found to be within a relative error-rate of 0.300, whereas far fewer motifs were found to be under-represented (see Supplementary material). None of the motifs of mFox$_{fg}$ and six of the mWnt$_{fg}$ motifs had a relative error-rate below 0.200.

To assess whether the novel motifs resembled any known *cis*-regulatory elements, all novel motifs were compared to all position weight matrices of the TRANSFAC database (Matys *et al*., 2006) using the program MATCOMPARE. The DME motif with the lowest relative error-rate from mFox$_{fg}$ (E-9-70) resembled the binding site matrix for the vertebrate RFX family of transcription factors while the second and third motifs both resembled binding site matrices for ETS family members

(Table 1). A motif identical to E-9-70 was also discovered by Bio-Prospector (Supplementary material). From mWnt$_{fg}$, the top three motifs resembled binding matrices for MEIS-1, Dof-1 (found only in plants), and the IRF family, respectively (Table 2). Interestingly, a motif resembling LEF-1, a member of the Wnt signaling pathway, was among the top 10. Ten under-represented motifs with the lowest relative error-rate matched factors that are involved in cell cycle, neuron and lymphoid tissue gene expression (see Supplementary material).

3.3 Cross-species comparison of over-represented motifs

Cis-regulatory elements important for critical lung functions, such as host defense or respiration, are expected to be evolutionarily conserved from rodents to humans. Motifs discovered *de novo* from the Fox$_{fg}$ dataset with a relative error-rate less than 0.3 were computationally compared between mouse and human using the MATCOMPARE algorithm (see Methods). Of the 57 mouse and 79 human motifs from Fox$_{fg}$ within a relative error-rate less than 0.3, there were 23 mouse motifs that mutually matched

Table 2. Top 10 DME novel motifs from mouse Wnt$_{fg}$

Motif ID	Logo	Number of Sites fg	bg[a]	R.E.[b]	Comparison to TRANSFAC (MATCOMPARE) Matrix[c]	Alignment[d]
E-11-21		15	2	0.181	M00419:MEIS1	MAKDRRCASAW . . . | | | NNNTGACASNNN
E-10-6		23	5	0.186	M00352:DOF1[e]	SWRAAKYYMT | . | | . NNNWAAAGNNN
E-11-22		22	4	0.191	M00972:IRF	ADVMAGAARHS | . . | | | . RDAANNGAAAN
B-10-9		11	1	0.193	M01022:LEF1	AGDMMCARAG : . | | . | | NHWCAAAGNN
E-9-39		20	6	0.197	M00262:STAF	CMSAAWRMA | . : : | . . . : NNNDCCCRBVATSCMHYDBNNN
E-8-10		18	6	0.198	M00292:FREAC4	ARACASWA | . | | | : : SHWRDGTAAACANHVN
B-10-1		12	2	0.204	M01004:HELIOSA	AGGRMMVAGA : | | . . HNHHGGGANNN
E-10-19		21	5	0.206	M00261:OLF	YMRGWGASYY . : : . | : : : NNVDDBTCCCHNRRGDVBNBVN
E-9-64		18	3	0.208	M00935:NFAT	GSYAGVAAM : | . | | . NHGGAAANNN
E-10-4		24	5	0.210	M00261:ETS	MHYTCCTBTK . . | | | : NYWTCCBN

A colour version of this table is available as supplementary data.

[a]Corrected based on relative size of background sequence.

[b]Significantly lower ($P < 0.001$) for this foreground set relative to random promoters.

[c]Based on TRANSFAC v. 9.3; Vertebrate matrices only unless noted.

[d]Top sequence = novel motif; lower sequence = TRANSFAC motif; (|) = exact match; (:) = strong similarity; (.) = weak similarity; (x) = no match; ( ) = no comparison.

[e]*Zea mays*.

human motifs. Nine out of the 10 top mouse Fox$_{fg}$ motifs had a mutual match to human motifs (Table 3). The matched human motifs were then compared to TRANSFAC v. 9.3 using MATCOM-PARE, and those results were compared to the TRANSFAC matches of the mouse motifs listed in Table 1. Three mouse Fox$_{fg}$ motifs resembled the same TRANSFAC matrices as their human counterpart. These motif pairs resembled matrices for the vertebrate ETS family and insect *Ttk*-69 (Table 3). A similar analysis was performed on novel motifs of Wnt$_{fg}$ (Table 4). Of the 186 mouse and 241 human motifs from Wnt$_{fg}$, 84 mouse motifs mutually matched human motifs. Four of the top 10 regulated mouse Wnt$_{fg}$ motifs matched seven human motifs. The cross-species comparison identified novel motifs resembling binding matrices for *Ttk*-69 and ETS and NFAT families over-represented in both mouse and human on lung-expressed promoters regulated by Foxa2 and/or Wnt signaling. Interestingly motifs resembling these same matrices (*Ttk*-69, ETS and NFAT) were also discovered by BioProspector (see Supplementary material).

Motifs associated with known *cis*-regulatory elements were also analyzed for their over-representation among lung promoters in both mouse and human promoters. The relative error-rates of all motifs contained in the TRANSFAC v. 9.3 databases were calculated with respect to mouse and human Fox$_{fg}$ and Wnt$_{fg}$,

and then the motifs were sorted by relative error-rate. The most over-represented motifs of mouse and human promoters were then compared and similarities identified. Relative error-rate limits of 0.38 and 0.34 were applied to motifs of Fox$_{fg}$ and Wnt$_{fg}$, respectively, to display 10 or fewer known motifs comparisons (Table 5). Within these relative error-rate limits, four TRANSFAC motifs were significantly over-represented in both mouse and human Fox$_{fg}$ (HES-1, TFE, *Ttk*-69 and AP-1). TRANSFAC motifs representing a total of three additional transcription factors (SRF, ER and CDP), and four factor families (SNAIL, C/EBP, GATA and STAT) were over-represented in Wnt$_{fg}$ of both mouse and human. For each known motif, the calculated relative error-rate with respect to the lung-expressed promoters was significantly different ($P < 0.05$) from a relative error-rate calculated from randomly selected promoters.

3.4 Computationally predicted transcription factors are expressed in lung epithelium

A number of the motifs identified herein corresponded to a variety of transcription factors never before characterized in the lung. To validate the predicted importance of these factors in mammalian lung, the corresponding transcription factor genes were monitored

Table 3. Mutual comparison of Fox$_{fg}$ novel motifs from mouse to human

Mouse motif		Similar human motif			Motif comparison		TRANSFAC[e]
Motif ID	Logo	Motif ID[a]	Logo	R.E.[b]	Alignment[c]	Div.[d]	
E-9-70	*(logo)*	E-9-12-rc	*(logo)*	0.299	GTTSCTDCD :|:.|:.||. STKYCWDCT	1.8	—
E-11-42	*(logo)*	E-11-22-rc	*(logo)*	0.261	TGYYTYBTSYM |:.|||.|..: MRVWTYCTGWH	2.9	M00771: ETS
E-10-1	*(logo)*	E-10-20	*(logo)*	0.277	ARDCAGRRRG ||.:|:...x ARMSARGAHM	2.6	M00971: ETS
E-11-2	*(logo)*	E-11-7	*(logo)*	0.250	ARKCAGGRRRR :|.:|::.... RRWMASRMASK	2.9	—
E-10-13	*(logo)*	E-10-5-rc	*(logo)*	0.284	YSMTCMTGSY x..:.:.|.. GGSWMCBGCM	2.9	M00009:[f] TTK69
B-10-7	*(logo)*	E-10-37	*(logo)*	0.253	CYBTRGCTTY :.:.|..:.||. YBSTGKYTTS	1.9	—
E-10-34	*(logo)*	E-10-41	*(logo)*	0.290	WYCCTVWTBT ..||x.:.: TKCCCCHDKW	2.8	—

A colour version of this table is available as supplementary data.
[a]rc = reverse complement of the motif.
[b]significantly lower ($P < 0.01$) for this foreground set relative to random promoters.
[c]Top line = mouse; bottom line = human; (|) = exact match; (:) = strong similarity; (.) = weak similarity; (x) = no match; ( ) = no data.
[d]Kullback-Leibler divergence score. Perfect match if div. = 0.
[e]Both human and mouse motifs matched the same TRANSFAC motif.
[f]*Drosophila Tramtrack*-69.

using nonquantitative RT–PCR analysis of total RNA preparations of mouse whole lung, human dissected lung tissue, and mouse or human cell lines representing distinct epithelial cell lineages. The transcription factors tested corresponded to the novel motifs with the three lowest relative error-rates. Motifs that were similar between mFox$_{fg}$ and mWnt$_{fg}$ (HELIOSA) or between the mouse and human (HES-1, SNAIL, NFAT and ETS) were also considered. In light of known examples where multiple members of a transcription factor family bind to very similar *cis*-regulatory elements (Merika and Orkin, 1993; Sementchenko and Watson, 2000), multiple members of three selected families (ETS, RFX and SNAIL) were investigated. There are no known vertebrate homologues for plant Dof-1. RT–PCR detection of transcription factors matching the under-represented identified is presented in Supplementary Figure S1.

Consistent with the predicted importance of over-represented motifs of Fox$_{fg}$, multiple members of the RFX and ETS families were found to be expressed in both mouse and human lung tissue and cell lines (see Fig. 1). RFX-1, -3 and -5 mRNAs were not detectable in parenchyma (alveoli) but all were detectable in A549 cells, and bronchial epithelial cells. RFX-4 was only found in human cell lines suggesting that RFX-4 is not expressed, expressed at low levels, or expressed in very few cells of adult lung. RFX-2, RFXAP and RFXANK transcripts were detectable in both airway and alveolar type tissues and cells, indicating ubiquitous expression throughout the lung epithelium. Transcripts

of all known members of the RFX family, except RFX-4, as well as two ETS members tested (ETS-1 and ETS-2) were detectable in whole lungs of all mouse strains tested (BALB/C and FVB/N) as well as in all mouse cell lines tested (C22 and MLE-15), suggesting that the roles for these ETS family members on promoters of genes regulated by Foxa2 in lung epithelia is conserved across phylogeny.

MEIS-1 and IRF-1 binding matrices were similar to the most over-represented novel motifs of mWnt$_{fg}$-E-11-21 and -E-11-22 (see Table 2). Both MEIS-1 and IRF-1 were expressed in all human and mouse lung tissues tested. HELIOS, whose binding matrix was similar to motifs over-represented in both Fox$_{fg}$ and Wnt$_{fg}$, was also expressed in both mouse and human lung.

A similar investigation of expression patterns was performed on transcription factors corresponding to known motifs (HES-1 and SNAIL) that were over-represented in both mouse and human promoters as shown in Table 5. SNAIL-1 was not detectable in airway and barely detectable in bronchial epithelial cells, but it was present in alveolar tissue, H441 and A549 cells. SNAIL-2 was detectable in all human tissues tested. SNAIL-3, was not detected in A549 cells. HES-1 was detectable only in H441 and bronchial epithelial cell lines, but it was detectable in all mouse cells and tissues tested.

Taken together, these data show that the transcription factors corresponding to the most highly over-represented novel motifs

Table 4. Mutual comparison of Wnt$_{fg}$ novel motifs from mouse and human

Mouse motif		Similar human motif			Motif comparison		
Motif ID	Logo	Motif ID[a]	Logo	R.E.[b]	Alignment[c]	Div.[d]	TRANSFAC[e]
E-11-22		E-11-7		0.275	ADVMAGAARHS |...||::..x ARMCAGRRVMW	1.9	—
		E-11-15-rc		0.126	RGWMARMMACH .:.||...:.. ADVMAGAARHS	2.5	—
		E-11-38-rc		0.279	SDYTTCTKBHT ..:||||..x MMHWTCTKTTM	2.7	—
B-10-1		E-10-41		0.221	AGGRMMVAGA ::::....:|: WKRGRRARGR	2.7	—
E-9-64		E-9-95		0.243	KTTBCTRSC |:.:|:|.: KYTYCWRGS	1.4	M00935: NFAT
E-10-4		E-10-31-rc		0.300	MHYTCCTBTK .|.:::|.|| CHTBSYTCTK	1.4	M00658: PU1
		E-10-17-rc		0.229	MHYTCCTBTK ...::||.:. ASTWYCTKHY	2.2	M00771: ETS

A colour version of this table is available as supplementary data.

[a]rc = reverse complement of the motif.

[b]Significantly lower ($P < 0.01$) for this foreground set relative to random promoters.

[c]Top line = mouse; bottom line = human; (|) = exact match; (:) = strong similarity; (.) = weak similarity; (x) = no match; ( ) = no data.

[d]Kullback–Leibler divergence score. Perfect match if div. = 0.

[e]Both human and mouse motifs matched the same TRANSFAC motif.

Table 5. Known motifs over-represented in mouse and human Fox$_{fg}$ and Wnt$_{fg}$

	Mouse				Human			
	Acc.	Name	R.E.[a]	*P* value[b]	Acc.	Name	R.E.	*P* value
Fox$_{fg}$	M01009	HES1	0.345	$P<0.01$	M01009	HES1	0.353	$P<0.05$
	M01029	TFE	0.325	$P<0.01$	M01029	TFE	0.364	$P<0.05$
	M00009	*Ttk-69*[c]	0.363	$P<0.01$	M00009	*Ttk-69*	0.370	$P<0.05$
	M00199	AP-1	0.377	$P<0.05$	M00174	AP-1	0.371	$P<0.05$
Wnt$_{fg}$	M00044	*Snail*[d]	0.336	$P<0.05$	M00060	[d]*Snail*	0.296	$P<0.01$
	M00632	GATA-4	0.279	$P<0.001$	M00077	GATA-3	0.301	$P<0.01$
	M00922	SRF	0.281	$P<0.01$	M00186	SRF	0.301	$P<0.01$
	M00810	SRF	0.300	$P<0.01$	M00215	SRF	0.314	$P<0.05$
	M00191	ER-α	0.329	$P<0.05$	M00191	ER-α	0.305	$P<0.01$
					M00959	ER-α/β	0.336	$P<0.05$
	M00102	CDP	0.304	$P<0.01$	M00106	CDP-HD	0.326	$P<0.05$
	M00106	CDP-HD[e]	0.323	$P<0.05$				
	M00621	C/EBP-δ	0.316	$P<0.05$	M00109	C/EBP-β	0.331	$P<0.05$
	M00460	STAT5A	0.335	$P<0.05$	M00494	STAT6	0.331	$P<0.05$

[a]Relative error-rate was calculated using MOTIFCLASS (see Methods).

[b]Significance based on 1000 calculations of relative error-rate where fg and bg were randomly shuffled.

[c]*Drosophila Tramtrack-69*.

[d]*Drosophila Snail1*.
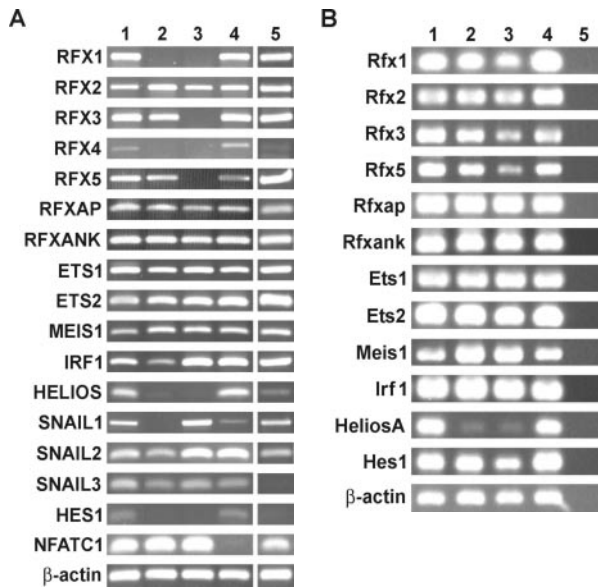
[e]CR3 domain–Homeodomain fusion protein.

Fig. 1. Expressed mRNA transcripts of candidate transcription factors were detected in human (**A**) and mouse (**B**) lung associated tissues or *in vitro* cell cultures. (A) Nonquantitative RT–PCR analysis of total RNA preparations of H441 cells (1), distal airway (2), lung parenchyma (3), bronchial epithelial cells (4) and A549 cells (5). (B) Nonquantitative RT–PCR analysis of total RNA preparations of C22 cells (1), BALB/C lung (2), FVB/N-WT lung (3) and MLE-15 cells (4). Negative control: no cDNA (5). β-actin was used a positive control for reverse transcription efficiency for both human [(A), shown for NFATC1] and mouse (B).

and known motifs identified by the computational methods described herein are expressed in both the mouse and human lung epithelium and may exhibit localized patterns within various regions of the lung.

4 DISCUSSION

The complexity of the human lung with regard to its number of cell types, and numerous physiological functions suggests unique regulatory pathways of gene expression. Herein, we have made computational predictions of transcriptional regulators of lung epithelial gene networks. This work resulted in the identification of 15 lung-expressed factors with uncharacterized lung functions.

DME is one of many motif discovery algorithms currently available but was chosen for its ability to utilize a background of real promoters (random promoters from the genome) rather than randomized sequence. It is unique in that it uses a refinement step to optimize the motifs and has been tested in tissue-specific promoters of higher organisms (Smith *et al.*, 2005). Additionally, the information content of the discovered motif can be specified. DME output files are compatible with MATCOMPARE which was modified for this study to compare multiple sets motifs. Using MATCOMPARE we were able to perform a cross-species comparison of mouse to human promoters with consideration to two independently derived types of promoter motifs, those discovered *de novo* by DME and those previously known and compiled into the TRANSFAC database. The comparison of data between different motif discovery algorithms is difficult due to differences in their motif search

parameters and scoring criteria. Regardless, both DME and BioProspector detected motifs matching to RFX-1, MEIS-1, RFX, *Ttk*-1 and NFAT binding sites.

Genes known to be co-regulated and specific to distinct differentiated epithelial cell lineages would serve as an ideal source of co-regulated promoters; however, such datasets have not been widely published. Consideration should be given to the fact that within the respiratory tract there are approximately 40–60 different cell types of ectodermal, endodermal and mesodermal origin. To overcome this complexity, technological advances have made it possible to investigate molecular pathways specifically in the lung epithelium (Perl *et al.*, 2002b). Herein, we used two gene array analyses specific to the lung epithelium as sources for co-expressed genes, one with targeted disruption of Foxa2 (HNF3-β) and the other with targeted constitutive activation of Wnt signaling (Okubo and Hogan, 2004; Wan *et al.*, 2004; Wan *et al.*, 2005). Many of the genes contained within these lists were already known to be involved in lung-specific physiological processes and expressed in pulmonary epithelial cell lineages. The goal of this work was to predict factors that function within these networks but not to address the mechanism by which Foxa2 or β-catenin-Lef-1 regulates these promoter sets.

Motif discovery *de novo* using the mouse Fox$_{fg}$ and Wnt$_{fg}$ gene promoter subsets identified members of the vertebrate ETS family as candidate regulators in both sets of promoters (see Tables 1 and 2). The additional finding that ETS binding motifs were also over-represented in the human homologue promoters further suggests the importance of this transcription factor family in regulating crucial lung functions (see Tables 3 and 4). Previously published work of ETS family members indicated expression in stratified epithelial cells, including the lung epithelium (Kola *et al.*, 1993; Kathuria *et al.*, 2004). Other reports have found ETS family members, the epithelial-specific enhancers (ESE), are expressed in airway epithelium (Oettgen *et al.*, 1997), although their overall function, and their role in maintenance of epithelial differentiation, is currently unknown. Work from this lab has shown that ESE-1, -2 and -3 influence expression of the host defense genes, lysozyme (Lys), chitinase (Ch3l1) and secretoglobin (Scgb1a1) (W. Lei and K. Harrod, unpublished data). These host defense genes are also regulated by Foxa2 and the Wnt pathway and are contained in Fox$_{fg}$ and Wnt$_{fg}$. It is possible that the ESE binding sites are contributing to over-representation of the ETS motif on these promoters.

Other over-represented novel motifs from Fox$_{fg}$ or Wnt$_{fg}$ resembled binding sites for the RFX factors, LEF-1 and MEIS-1, consistent with previous reports that these factors are expressed in lung tissue (Reith *et al.*, 1994; Su *et al.*, 2002; Okubo and Hogan, 2004; Su *et al.*, 2004; Dintilhac *et al.*, 2005; Steel *et al.*, 2005). Finding an over-representation of Lef-1 sites of promoters' down-regulated by Lef-1-β-catenin suggests the repression of these lung markers may involve HDAC1 (Billin *et al.*, 2000). Over-representation of motifs resembling the FREAC-4 motif was not surprising since some FREAC proteins are expressed in lung (Pierrou *et al.*, 1994; Hellqvist *et al.*, 1996). Using the occurrences of these novel motifs we can experimentally investigate these putative binding sites on promoters of these foreground sets. Members of the RFX family have been implicated in the formation of cilia in *Caenorhabditis elegans* (Daf19), *Drosophila* (*RFX*) and mouse (Rfx-3) (Perkins *et al.*, 1986; Dubruille *et al.*, 2002; Bonnafe *et al.*, 2004) but their roles in lung epithelia have not been

elucidated. The function of members of the RFX family in lung epithelial cell subsets, especially ciliated cells will require further investigation. The over-representation of the motifs for *Drosophila Ttk*-69 and plant Dof-1 may indicate the existence of a mammalian structural counterpart to the DNA-binding domains of these factors.

From the novel and known motif analysis, the Wnt_{fg} gene subset consistently identified motifs corresponding to factors that have been associated with inflammatory signaling cascades (IRF, NFAT, STAT, C/EBP) (Look *et al*., 1995; Xanthoudakis *et al*., 1996; Burgess-Beusse and Darlington, 1998; Sampath *et al*., 1999). Both the Fox_{fg} and Wnt_{fg} subsets used for the computational approaches described herein include an abundance of genes important in host defense, such as surfactant proteins, secretoglobins, lysozyme and chitinase. Our data are consistent with the hypothesis that lung host defense is coupled with transcriptional regulatory control during lung development, cell differentiation and epithelial maintenance (Whitsett, 2002). Our discovery of putative novel transcriptional regulators of lung host defense genes may lead to further the understanding of mechanisms by which infection and inflammation lead to pulmonary disease.

Expression of the factors HES-1, MEIS-1 and members of the RFX, ETS and SNAIL families were shown in mouse and/or human lung tissues herein. Our results are consistent with prior evidence of these factors having roles in epithelial cell function (Dubruille *et al*., 2002; Bomgardner *et al*., 2003; Bonnafe *et al*., 2004; Kathuria *et al*., 2004; Nakamura *et al*., 2004; Parent *et al*., 2004; Bachelder *et al*., 2005; Dintilhac *et al*., 2005; van Tuyl *et al*., 2005). The expression of IRF-1 and NFAT in human lung epithelia is consistent with prior studies of these factors in mouse lung epithelial cells (Zhao *et al*., 2003; Dave *et al*., 2004). The findings presented here represent the initial observations for these transcription factors in human lung epithelia.

Although not tested here, the over-representation of the sites of some of these factors (TFE and SRF) is consistent with enriched expression in lung or airway tissues as reported in the GNF Atlas microarray dataset (Reith *et al*., 1994; Su *et al*., 2002; Karolchik *et al*., 2003; Su *et al*., 2004) or from prior studies of lung epithelial cells (AP-1, ER, CDP and C/EBP) (Ellis *et al*., 2001; Reddy and Mossman, 2002; Vuong *et al*., 2002; Cassel and Nord, 2003; Patrone *et al*., 2003). Our findings also indicated expression for transcription factors Helios, SNAIL-1 and -3 in lung cells and tissues, while the GNF Atlas indicated these were expressed but not enriched in the lung.

Many motifs that were shown to be conserved using the known motif analysis were not represented by DME motifs (e.g. SNAIL and HES-1) and vice versa (e.g. RFX and MEIS-1). In most cases, this may be due to DME motifs being shorter in length as compared with the binding sites contained in the TRANSFAC database, consequently making proper alignments difficult. A second possibility is that the DME input parameters excluded certain *cis*-regulatory elements. Additionally, some factors predicted by the novel motif analysis were not predicted by the known motifs. This could occur if the known motifs do not accurately represent *cis*-regulatory elements on lung-expressed promoters in mouse and/or human. Co-factors, cellular biochemistry, promoter organization or protein modifications may all play a role in changing the DNA-binding specificity of a factor (Naar *et al*., 2001; Werner *et al*., 2003).

These findings suggest reliance on a single computational tool may be limiting.

Our results identified highly over-represented motifs for GATA-3 in mouse Wnt_{fg} and GATA-4 in human Wnt_{fg}. This result may represent multiple *cis*-regulatory elements of the entire GATA family functioning in both organisms. GATA-4 is hypothesized to play a role in foregut endoderm organ development (Kuo *et al*., 1997), and GATA-6 expression has been identified to be important in lung-specific gene expression and development (Bruno *et al*., 2000; Yang *et al*., 2002; Liu *et al*., 2003). Similarly, motifs representing different members of the C/EBP and STAT families were also identified in the two species. Although, factors within these families have very similar binding specificities (Akira *et al*., 1990; Merika and Orkin, 1993; Ehret *et al*., 2001), our results may demonstrate a weakness in the specificity of the TRANSFAC models for these families and more examples may exist.

Experimental validation of our computational approach is important in light of the vast numbers of motifs identified and their corresponding transcription factors. Ultimate validation would involve testing each factor's activity on each promoter via discovered occurrences to confirm that over-represented sites are functional, a task requiring additional mapping and high-throughput experiments. As a first step toward this goal, we have shown positive expression of these factors in many healthy lung tissue and cell samples for both human and mouse; however, our work does not try to identify their functional targets. Under-represented motifs matched with factors having functions not specifically associated with lung epithelia, and of three factors tested, two were not expressed in epithelial cell lines. Computational validation identified seven motifs with position preference relative to the start of transcription (see Supplementary material). Out of hundreds of motifs examined, the finding of only a few motifs with position preference is consistent with previous findings (FitzGerald *et al*., 2004). Overall, our results are consistent with published data describing function or expression for these predicted transcription factors.

We performed a computational analysis on two groups of co-expressed lung promoters to identify over-represented promoter motifs. We identify motifs associated with a total of 25 transcription factor families, of which 17 factors were experimentally found to be expressed in the lung. Our work presents the largest number of computationally predicted lung *cis*-regulatory elements presented to date. Future work will focus on further characterization of the predicted *cis*-regulatory elements and corresponding binding factors and their functionality in lung cell subsets.

ACKNOWLEDGEMENTS

REFERENCES

Akira,S. *et al*. (1990) A nuclear factor for IL-6 expression (NF-IL6) is a member of a C/EBP family. *EMBO J.*, **9**, 1897–1906.
Bachelder,R.E. *et al*. (2005) Glycogen synthase kinase-3 is an endogenous inhibitor of Snail transcription: implications for the epithelial-mesenchymal transition. *J. Cell. Biol.*, **168**, 29–33.

Billin,A.N. *et al.* (2000) Beta-catenin-histone deacetylase interactions regulate the transition of lef1 from a transcriptional repressor to an activator. *Mol. Cell. Biol.*, **20**, 6882–6890.

Bomgardner,D. *et al.* (2003) 5' Hox genes and Meis 1, a hox-DNA binding cofactor, are expressed in the adult mouse epididymis. *Biol. Reprod.*, **68**, 644–650.

Bonnafe,E. *et al.* (2004) The transcription factor RFX3 directs nodal cilium development and left-right asymmetry specification. *Mol. Cell. Biol.*, **24**, 4417–4427.

Bruno,M.D. *et al.* (2000) GATA-6 activates transcription of surfactant protein A. *J. Biol. Chem.*, **275**, 1043–1009.

Burgess-Beusse,B.L. and Darlington,G.J. (1998) C/EBPalpha is critical for the neonatal acute-phase response to inflammation. *Mol. Cell. Biol.*, **18**, 7269–7277.

Cassel,T.N. and Nord,M. (2003) C/EBP transcription factors in the lung epithelium. *Am. J. Physiol. Lung Cell. Mol. Physiol.*, **285**, L773–L781.

Dave,V. *et al.* (2004) Nuclear factor of activated T cells regulates transcription of the surfactant protein D gene (Sftpd) via direct interaction with thyroid transcription factor-1 in lung epithelial cells. *J. Biol. Chem.*, **279**, 34578–34588.

Demello,D.E. *et al.* (2002) Generation and characterization of a conditionally immortalized lung clara cell line from the h-2kb-tsa58 transgenic mouse. *In Vitro Cell. Dev. Biol. Anim.*, **38**, 154–164.

Dintilhac,A. *et al.* (2005) PBX1 intracellular localization is independent of meis1 in epithelial cells of the developing female genital tract. *Int. J. Dev. Biol.*, **49**, 851–858.

Dubruille,R. *et al.* (2002) *Drosophila* regulatory factor X is necessary for ciliated sensory neuron differentiation. *Development*, **129**, 5487–5498.

Ehret,G.B. *et al.* (2001) DNA binding specificity of different STAT proteins. Comparison of *in vitro* specificity with natural target sites. *J. Biol. Chem.*, **276**, 6675–6688.

Ellis,T. *et al.* (2001) The transcriptional repressor CDP (Cutl1) is essential for epithelial cell differentiation of the lung and the hair follicle. *Genes Dev.*, **15**, 2307–2319.

FitzGerald,P.C. *et al.* (2004) Clustering of DNA sequences in human promoters. *Genome Res.*, **14**, 1562–1574.

Frech,K. *et al.* (1998) Muscle actin genes: a first step towards computational classification of tissue specific promoters. *In Silico Biol.*, **1**, 29–38.

Hellqvist,M. *et al.* (1996) Differential activation of lung-specific genes by two forkhead proteins, FREAC-1 and FREAC-2. *J. Biol. Chem.*, **271**, 4482–4490.

Karolchik,D. *et al.* (2003) The UCSC genome browser database. *Nucleic Acids Res.*, **31**, 51–54.

Kathuria,H. *et al.* (2004) Transcription of the caveolin-1 gene is differentially regulated in lung type I epithelial and endothelial cell lines: a role for ETS proteins in epithelial cell expression. *J. Biol. Chem.*, **279**, 30028–30036.

Kola,I. *et al.* (1993) The Ets1 transcription factor is widely expressed during murine embryo development and is associated with mesodermal cells involved in morphogenetic processes such as organ formation. *Proc. Natl Acad. Sci. USA*, **90**, 7588–7592.

Kuo,C.T. *et al.* (1997) GATA4 transcription factor is required for ventral morphogenesis and heart tube formation. *Genes Dev.*, **11**, 1048–1060.

Liu,C. *et al.* (2003) Inhibition of alveolarization and altered pulmonary mechanics in mice expressing GATA-6. *Am. J. Physiol. Lung Cell. Mol. Physiol.*, **285**, L1246–L1254.

Liu,X. *et al.* (2001) Bioprospector: discovering conserved DNA motifs in upstream regulatory regions of co-expressed genes. *Pac. Symp. Biocomput.*, **6**, 127–138.

Look,D.C. *et al.* (1995) Stat1 depends on transcriptional synergy with Sp1. *J. Biol. Chem.*, **270**, 30264–30267.

Matys,V. *et al.* (2006) TRANSFAC and its module transcompel: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res.*, **34**, D108–D110.

Merika,M. and Orkin,S.H. (1993) DNA-binding specificity of GATA family transcription factors. *Mol. Cell. Biol.*, **13**, 3999–4010.

Naar,A.M. *et al.* (2001) Transcriptional coactivator complexes. *Annu. Rev. Biochem.*, **70**, 475–501.

Nakamura,Y. *et al.* (2004) Ets-1 regulates TNF-alpha-induced matrix metalloproteinase-9 and tenascin expression in primary bronchial fibroblasts. *J. Immunol.*, **172**, 1945–1952.

Nelander,S. *et al.* (2005) Predictive screening for regulators of conserved functional gene modules (gene batteries) in mammals. *BMC Genomics*, **6**, 68.

Nelander,S. *et al.* (2003) Prediction of cell type-specific gene modules: identification and initial characterization of a core set of smooth muscle-specific genes. *Genome Res.*, **13**, 1838–1854.

Oettgen,P. *et al.* (1997) Isolation and characterization of a novel epithelium-specific transcription factor, ESE-1, a member of the *ets* family. *Mol. Cell. Biol.*, **17**, 4419–4433.

Okubo,T. and Hogan,B.L. (2004) Hyperactive Wnt signaling changes the developmental potential of embryonic endoderm. *J. Biol.*, **3**, 11.1–11.17.

Parent,A.E. *et al.* (2004) The developmental transcription factor slug is widely expressed in tissues of adult mice. *J. Histochem. Cytochem.*, **52**, 959–965.

Patrone,C. *et al.* (2003) Regulation of postnatal lung development and homeostasis by estrogen receptor-beta. *Mol. Cell. Biol.*, **23**, 8542–8552.

Perkins,L.A. (1986) Mutant sensory cilia in the nematode *Caenorhabditis elegans*. *Dev. Biol.*, **117**, 456–487.

Perl,A.K. *et al.* (2002a) Conditional gene expression in the respiratory epithelium of the mouse. *Transgenic Res.*, **11**, 21–29.

Perl,A.K. *et al.* (2002b) Early restriction of peripheral and proximal cell lineages during formation of the lung. *Proc. Natl Acad. Sci. USA*, **99**, 10482–10487.

Pierrou,S. *et al.* (1994) Cloning and characterization of seven human forkhead proteins: binding site specificity and DNA bending. *EMBO J.*, **13**, 5002–5012.

Pruitt,K.D. *et al.* (2005) Ncbi reference sequence (refseq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.*, **33**, D501–D504.

Reddy,S.P. and Mossman,B.T. (2002) Role and regulation of activator protein-1 in toxicant-induced responses of the lung. *Am. J. Physiol. Lung Cell. Mol. Physiol.*, **283**, L1161–L1178.

Reith,W. *et al.* (1994) RFX1, a transactivator of hepatitis B virus enhancer I, belongs to a novel family of homodimeric and heterodimeric DNA-binding proteins. *Mol. Cell. Biol.*, **14**, 1230–1244.

Sampath,D. *et al.* (1999) Constitutive activation of an epithelial signal transducer and activator of transcription (STAT) pathway in asthma. *J. Clin. Invest.*, **103**, 1353–1361.

Schones,D.E. *et al.* (2005) Similarity of position frequency matrices for transcription factor binding sites. *Bioinformatics*, **21**, 307–313.

Sementchenko,V.I. and Watson,D.K. (2000) Ets target genes: past, present and future. *Oncogene*, **19**, 6533–6548.

Smith,A.D. *et al.* (2005) Identifying tissue-selective transcription factor binding sites in vertebrate promoters. *Proc. Natl Acad. Sci. USA*, **102**, 1560–1565.

Smith,A.S. *et al.* (2006) DNA motifs in human and mouse proximal promoters predict tissue specific expression. *Proc. Natl Acad. Sci. USA*, **103**, 6275–6280.

Steel,M.D. *et al.* (2005) Beta-catenin/T-cell factor-mediated transcription is modulated by cell density in human bronchial epithelial cells. *Int. J. Biochem. Cell. Biol.*, **37**, 1281–1295.

Stormo,G.D. (2000) DNA binding sites: representation and discovery. *Bioinformatics*, **16**, 16–23.

Su,A.I. *et al.* (2002) Large-scale analysis of the human and mouse transcriptomes. *Proc. Natl Acad. Sci. USA*, **99**, 4465–4470.

Su,A.I. *et al.* (2004) A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc. Natl Acad. Sci. USA*, **101**, 6062–6067.

van Tuyl,M. *et al.* (2005) Overexpression of lunatic fringe does not affect epithelial cell differentiation in the developing mouse lung. *Am. J. Physiol. Lung Cell. Mol. Physiol.*, **288**, L672–L682.

Vuong,H. *et al.* (2002) JNK1 and AP-1 regulate PMA-inducible squamous differentiation marker expression in clara-like H441 cells. *Am. J. Physiol. Lung Cell. Mol. Physiol.*, **282**, L215–L225.

Wan,H. *et al.* (2005) Compensatory roles of Foxa1 and Foxa2 during lung morphogenesis. *J. Biol. Chem.*, **280**, 13809–13816.

Wan,H. *et al.* (2004) Foxa2 is required for transition to air breathing at birth. *Proc. Natl Acad. Sci. USA*, **101**, 14449–14454.

Wasserman,W.W. and Sandelin,A. (2004) Applied bioinformatics for the identification of regulatory elements. *Nat. Rev. Genet.*, **5**, 276–287.

Werner,T. *et al.* (2003) Computer modeling of promoter organization as a tool to study transcriptional coregulation. *FASEB*, **17**, 1228–1237.

Whitsett,J.A. (2002) Intrinsic and innate defenses in the lung: intersection of pathways regulating lung morphogenesis, host defense, and repair. *J. Clin. Invest*, **109**, 565–569.

Xanthoudakis,S. *et al.* (1996) An enhanced immune response in mice lacking the transcription factor NFAT1. *Science*, **272**, 892–895.

Yang,H. *et al.* (2002) GATA6 regulates differentiation of distal lung epithelium. *Development*, **129**, 2233–2246.

Zhao,Z. *et al.* (2003) IFN regulatory factor-1 is required for the up-regulation of the CD40-NF-kappa B activator 1 axis during airway inflammation. *J. Immunol.*, **170**, 5674–5680.