



Integrated cistromic and expression analysis of amplified *NKX2-1* in lung adenocarcinoma identifies *LMO3* as a functional transcriptional target

Hideo Watanabe, Joshua M. Francis, Michele S. Woo, et al.

Genes Dev. 2013 27: 197-210 originally published online January 15, 2013
Access the most recent version at doi:[10.1101/gad.203208.112](https://doi.org/10.1101/gad.203208.112)

Supplemental Material <http://genesdev.cshlp.org/content/suppl/2013/01/09/gad.203208.112.DC1.html>

References This article cites 71 articles, 32 of which can be accessed free at:
<http://genesdev.cshlp.org/content/27/2/197.full.html#ref-list-1>

Open Access Freely available online through the *Genes & Development* Open Access option.

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

To subscribe to *Genes & Development* go to:
<http://genesdev.cshlp.org/subscriptions>

Integrated cistromic and expression analysis of amplified *NKX2-1* in lung adenocarcinoma identifies *LMO3* as a functional transcriptional target

Hideo Watanabe,^{1,2,3} Joshua M. Francis,^{1,2,3} Michele S. Woo,^{1,2} Banafsheh Etemad,¹ Wenchu Lin,^{1,2} Daniel F. Fries,^{1,3} Shouyong Peng,^{1,3} Eric L. Snyder,⁴ Purushothama Rao Tata,⁵ Francesca Izzo,^{1,6} Anna C. Schinzel,^{1,6} Jeonghee Cho,^{1,2} Peter S. Hammerman,^{1,2,3} Roel G. Verhaak,^{1,2,3} William C. Hahn,^{1,2,3,6} Jayaraj Rajagopal,⁵ Tyler Jacks,^{4,7} and Matthew Meyerson^{1,2,3,8,9}

¹Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, Massachusetts 02215, USA; ²Center for Cancer Genome Discovery, Dana-Farber Cancer Institute, Boston, Massachusetts 02215, USA; ³Cancer Program, Broad Institute of Harvard and Massachusetts Institute of Technology, Cambridge, Massachusetts 02142, USA; ⁴David H. Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA; ⁵Center for Regenerative Medicine, Massachusetts General Hospital, Harvard Stem Cell Institute, Boston, Massachusetts 02114, USA; ⁶RNAi Facility, Dana-Farber Cancer Institute, Boston, Massachusetts 02215, USA; ⁷Howard Hughes Medical Institute, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA; ⁸Department of Pathology, Harvard Medical School, Boston, Massachusetts 02115, USA

The *NKX2-1* transcription factor, a regulator of normal lung development, is the most significantly amplified gene in human lung adenocarcinoma. To study the transcriptional impact of *NKX2-1* amplification, we generated an expression signature associated with *NKX2-1* amplification in human lung adenocarcinoma and analyzed DNA-binding sites of *NKX2-1* by genome-wide chromatin immunoprecipitation. Integration of these expression and cistromic analyses identified *LMO3*, itself encoding a transcription regulator, as a candidate direct transcriptional target of *NKX2-1*. Further cistromic and overexpression analyses indicated that *NKX2-1* can cooperate with the forkhead box transcription factor *FOXA1* to regulate *LMO3* gene expression. RNAi analysis of *NKX2-1*-amplified cells compared with nonamplified cells demonstrated that *LMO3* mediates cell survival downstream from *NKX2-1*. Our findings provide new insight into the transcriptional regulatory network of *NKX2-1* and suggest that *LMO3* is a transcriptional signal transducer in *NKX2-1*-amplified lung adenocarcinomas.

[*Keywords:* lung cancer; lineage-specific oncogene; transcription factor; cistromic analysis; carcinogenesis]

Supplemental material is available for this article.

Received August 20, 2012; revised version accepted December 20, 2012.

Genomic alterations of transcription factors are known to play key roles in the development of many human cancers. For example, *TP53* (the gene encoding p53) is the most frequently mutated gene in all cancers (Levine and Oren 2009), *MYC* is frequently amplified in many types of cancers (including breast carcinomas and small-cell lung carcinomas) and is translocated in Burkitt's lymphomas (Dalla-Favera et al. 1982; Taub et al. 1982), and *ETS* family members are frequently translocated in prostate cancer (Tomlins et al. 2005), Ewing's sarcoma (Delattre et al. 1992), and leukemias (Peeters et al. 1997). For some tran-

scription factors, genomic alterations are only associated with particular types of cancers: For example, *AR* amplification is linked to mechanisms of resistance in recurrent prostate cancers (Visakorpi et al. 1995), *PAX5* deletion is linked to acute lymphocytic leukemia (Mullighan et al. 2007), and *RUNX1* translocation is linked to acute myelogenous leukemia (Miyoshi et al. 1991). In addition, there has been emerging evidence that a lineage-restricted genomic amplification of developmental transcription factors occurs frequently in solid tumors, as exemplified by *MITF* in melanomas and *SOX2* in lung and esophageal squamous cell carcinomas (Garraway et al. 2005; Bass et al. 2009).

NKX2-1 is the most significantly focally amplified gene in lung adenocarcinomas, with amplification detected in ~12% of cases (Kendall et al. 2007; Tanaka et al. 2007; Weir et al. 2007; Kwei et al. 2008). *NKX2-1*, also referred

⁹Corresponding author

E-mail matthew_meyerson@dfci.harvard.edu

Article published online ahead of print. Article and publication date are online at <http://www.genesdev.org/cgi/doi/10.1101/gad.203208.112>. Freely available online through the *Genes & Development* Open Access option.

to as TTF-1 (for thyroid transcription factor 1), is well known as a molecular marker for lung adenocarcinoma and is particularly useful in clinical diagnosis of metastatic carcinomas, where its identification supports the tumor originating in the lung (Bejarano et al. 1996; Holzinger et al. 1996). *Nkx2-1* is required for the development of the trachea, brain, and thyroid in early murine embryonic development and for peripheral lung-branching morphogenesis later in development (Costa et al. 2001; Maeda et al. 2007). Mice lacking *Nkx2-1* die at birth of respiratory failure with hypoplastic lungs that stem from an undivided foregut (Yuan et al. 2000). *NKX2-1* may belong to the class of lineage survival oncogenes, which are ordinarily required for the differentiation and survival of particular cell lineages and later become subject to focal amplification in cancers within their own lineage (Garraway and Sellers 2006). While the specific cell of origin that gives rise to lung adenocarcinomas has yet to be precisely characterized, *NKX2-1* is required for the survival of lung adenocarcinoma cells with amplification of *NKX2-1* (Kendall et al. 2007; Tanaka et al. 2007; Weir et al. 2007; Kwei et al. 2008).

The role of *NKX2-1* in cancer pathogenesis is complex and remains poorly understood. Activating translocations of *NKX2-1* have been reported in ~3% of acute pre-T-cell lymphoblastic leukemias (T-ALL) (Homminga et al. 2011), suggesting that the oncogenic function of *NKX2-1* may not be restricted to the lung. In addition, like *NOTCH1* (Stransky et al. 2011) and *MEN1* (Yokoyama et al. 2005), it appears that *NKX2-1* can play both an oncogenic and a tumor-suppressive role in different settings. While *NKX2-1* amplification is found in human lung adenocarcinoma, loss of mouse *Nkx2-1* promotes metastasis in a *Kras*-driven mouse model of lung adenocarcinoma, and human tumors with low *NKX2-1* expression have generally worse prognoses (Winslow et al. 2011). More recently, a study showed evidence that *Nkx2-1* haploinsufficiency increased *Kras*-induced tumor initiation and progression (Maeda et al. 2012), consistent with previous results (Winslow et al. 2011), but decreased *EGFR*-induced lung adenocarcinomas, indicating context-dependent roles of *Nkx2-1* in lung carcinogenesis (Maeda et al. 2012). These results are consistent with the remarkable genomic heterogeneity of human lung adenocarcinoma (Weir et al. 2007; Ding et al. 2008; Govindan et al. 2012; Imielinski et al. 2012). While the *NKX2-1* locus is the most commonly amplified region in lung adenocarcinoma and RNAi experiments confirm *NKX2-1* as the functional target of this amplification (Kendall et al. 2007; Tanaka et al. 2007; Weir et al. 2007; Kwei et al. 2008), lung adenocarcinomas without *NKX2-1* amplification and/or expression plausibly harbor other genomic alterations that play complementary roles to *NKX2-1*. The observation that tumors with low *NKX2-1* expression (Barletta et al. 2009; Winslow et al. 2011) and with *NKX2-1* amplification (Barletta et al. 2009) are both associated with poor prognosis may not imply any mechanistic relationship to *NKX2-1* itself, as these likely represent the result of different heterogeneous features of the tumors. *NKX2-1* has recently been reported to activate expression of the *ROR1* gene in lung

adenocarcinoma (Yamaguchi et al. 2012); however, the transcriptional consequences of *NKX2-1* amplification in lung adenocarcinoma and the mechanism underlying its oncogenic activity in this disease have not been established.

In the normal lung, *NKX2-1* induces a subset of gene expression changes involved in the differentiation of alveolar type II cells. Among the directly induced genes reported are *LAMP3*, *CEACAM6*, and *CIT* (Kolla et al. 2007), and an *NKX2-1* overexpression signature in BEAS-2B bronchoepithelial cells includes focal adhesion and oxidative phosphorylation pathways (Hsu et al. 2009). Promoter regions directly bound by *Nkx2-1* in developing lungs have been also reported, which include the promoters of *E2f3*, *Ccnb1*, and *Ccnb2* genes (Tagne et al. 2012). Mechanistically, transcriptional activity of *Nkx2-1* has been shown to be facilitated by interaction with several cellular proteins, including nuclear hormone receptors such as the retinoic acid receptor (RAR), zinc finger transcription factors such as Gata-6, and coactivators such as Src (Maeda et al. 2007).

Here, using integrated cistromic and gene expression analysis, we show that *NKX2-1* amplification is associated with overexpression of the *LMO3* gene, a member of the LMO family of oncogenes that are translocated in T-ALL (Boehm et al. 1988b; McGuire et al. 1989), the same disorder in which *NKX2-1* translocation has been observed (Homminga et al. 2011). Furthermore, we show that *NKX2-1* interacts physically with the *LMO3* locus and activates its expression, which preferentially drives survival of cells with *NKX2-1* amplification.

Results

Expression signature of NKX2-1-amplified lung adenocarcinomas

To identify the genes regulated by *NKX2-1* and provide additional insight into the cellular origin of *NKX2-1*-amplified lung adenocarcinomas, we derived an *NKX2-1* expression signature in lung adenocarcinomas that harbor amplification of *NKX2-1*. We performed an unpaired two-class statistical comparison using SAM (significance analysis of microarray) (Tusher et al. 2001) on array-based expression profiling data from 470 primary lung adenocarcinoma tumors (Director's Challenge Consortium for the Molecular Classification of Lung Adenocarcinoma et al. 2008) and 42 non-small-cell lung carcinoma cell lines with matched copy number data (Sos et al. 2009) to identify genes whose expression is most significantly correlated with *NKX2-1* expression in primary tumors and with *NKX2-1* amplification in cell lines (Fig. 1A,B; Supplemental Table S1).

To discern individual genes whose expression is significantly correlated with *NKX2-1* overexpression, we intersected the most significantly differentially expressed genes by SAM analysis in data sets generated from primary tumors and cell lines. Since the statistical power for detection is very different in the two data sets (Fig. 1A,B), we analyzed the 1000 most differentially expressed genes

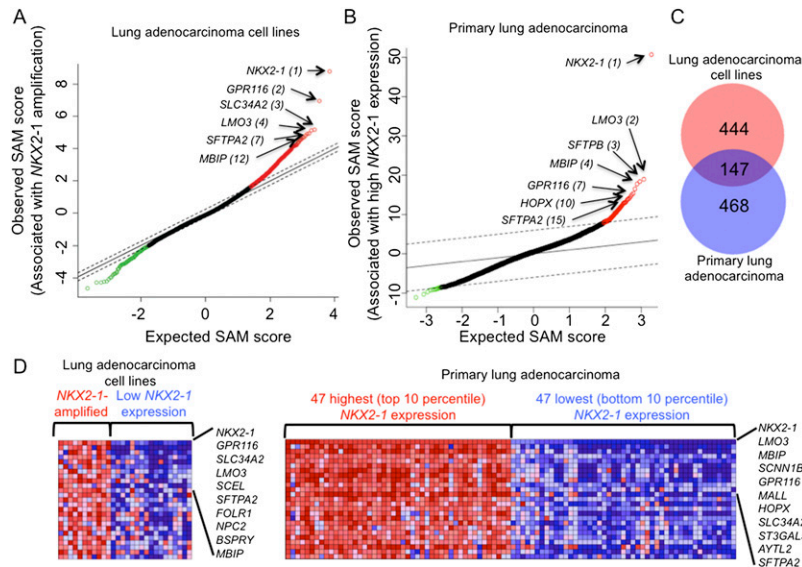


Figure 1. Genes associated with *NKX2-1* amplification and overexpression in human lung adenocarcinoma. (A) Genes (12,328) are rank-ordered by differential expression score (SAM score) and plotted against the expected SAM score in 11 lung adenocarcinoma cell lines with high-level focal *NKX2-1* amplification compared with 17 lung adenocarcinoma cell lines with the lowest *NKX2-1* expression in a panel of 42 lung adenocarcinoma cell lines with matched copy number data from a 250K SNP array. The red circle indicates positively correlated genes with a SAM score that deviates from the expected distribution of the SAM score at a Delta slope of 0.26. The green circle indicates negatively correlated genes with score deviates from the expected distribution of the SAM score at a Delta slope of 0.26. (B) Genes (13,291) are rank-ordered by a differential expression SAM score and plotted against the expected SAM score in 47 (top 10th percentile) primary human lung adenocarcinomas with the highest expression of *NKX2-1* compared with 47 (bottom 10th percentile) tumors with the lowest *NKX2-1* expression from expression profiles of 470 primary lung adenocarcinomas. Genes are plotted against the expected distribution of the SAM score. The red circle indicates positively correlated genes with a SAM score that deviates from the expected distribution of the SAM score at a Delta slope of 3.9. The green circle indicates negatively correlated genes with a score that deviates from the expected distribution of the SAM score at a Delta slope of 3.9. There were more genes whose expression is positively correlated with *NKX2-1* overexpression than negatively correlated. The signature tails, or overall departure from no gene effects (from 0 to 1) in the data set, for positive is 0.77 and for negative is 0.20. (C) Venn diagram depicting the overlap between positively correlated genes of the 1000 most differentially expressed genes associated with *NKX2-1* amplification in lung adenocarcinoma cell lines and positively correlated genes of the 1000 most differentially expressed genes associated with the top 10th percentile highest *NKX2-1* expression in primary lung adenocarcinomas (hypergeometric *P*-value of significant overlap $<1 \times 10^{-12}$). (D, left) Of the 147 overlapped genes, the expression level of the most differentially expressed 25 genes in 11 *NKX2-1*-amplified cell lines and 17 cell lines with the lowest *NKX2-1* expression is shown in a heat map. Cell lines are ordered from left to right by copy number at the *NKX2-1* locus. (Right) Of the 147 overlapped genes, the expression level of the most differentially expressed 25 genes in primary lung adenocarcinomas with the top and bottom 10th percentile *NKX2-1* expression is shown in a heat map. Samples are ordered from left to right by *NKX2-1* expression level. The complete list of the genes is available in Supplemental Table 1.

tumors with the lowest *NKX2-1* expression from expression profiles of 470 primary lung adenocarcinomas. Genes are plotted against the expected distribution of the SAM score. The red circle indicates positively correlated genes with a SAM score that deviates from the expected distribution of the SAM score at a Delta slope of 3.9. The green circle indicates negatively correlated genes with a score that deviates from the expected distribution of the SAM score at a Delta slope of 3.9. There were more genes whose expression is positively correlated with *NKX2-1* overexpression than negatively correlated. The signature tails, or overall departure from no gene effects (from 0 to 1) in the data set, for positive is 0.77 and for negative is 0.20. (C) Venn diagram depicting the overlap between positively correlated genes of the 1000 most differentially expressed genes associated with *NKX2-1* amplification in lung adenocarcinoma cell lines and positively correlated genes of the 1000 most differentially expressed genes associated with the top 10th percentile highest *NKX2-1* expression in primary lung adenocarcinomas (hypergeometric *P*-value of significant overlap $<1 \times 10^{-12}$). (D, left) Of the 147 overlapped genes, the expression level of the most differentially expressed 25 genes in 11 *NKX2-1*-amplified cell lines and 17 cell lines with the lowest *NKX2-1* expression is shown in a heat map. Cell lines are ordered from left to right by copy number at the *NKX2-1* locus. (Right) Of the 147 overlapped genes, the expression level of the most differentially expressed 25 genes in primary lung adenocarcinomas with the top and bottom 10th percentile *NKX2-1* expression is shown in a heat map. Samples are ordered from left to right by *NKX2-1* expression level. The complete list of the genes is available in Supplemental Table 1.

with (observed SAM score – expected score) > 3.9 for primary tumors and > 0.26 for cell lines. As expected, there were significant overlaps; 147 overexpressed genes ($P < 1 \times 10^{-12}$) and 38 underexpressed genes ($P = 1.4 \times 10^{-9}$) were shared between the data sets (Fig. 1C; Supplemental Fig. S1A,B; Supplemental Table S1). The genes whose expression was positively correlated with *NKX2-1* amplification included three categories: genes that reside on chromosome 14q13 and are often coamplified with *NKX2-1*, which include *MBIP* and *NFKBIA*; genes that are known to be positively regulated by *NKX2-1*, including *SFTPA2*, *HOP* (Yin et al. 2006), and *CEACAM6* (Kolla et al. 2007); and newly identified coexpressed genes, including *LMO3*, *GPR116*, and *SLC34A2*. These *NKX2-1* targets are overrepresented in a gene expression signature (cluster 4) of lung adenocarcinomas previously identified with the same primary tumor data set ($P < 1 \times 10^{-16}$) (Fig. 1D; Supplemental Fig. S1C; Director's Challenge Consortium for the Molecular Classification of Lung Adenocarcinoma et al. 2008). This finding is also consistent with an analysis of data from expression profiling arrays of 49 lung adenocarcinomas and 40 lung squamous cell carcinomas that identified *LMO3* (also known as *DAT1*) as the second most differentially expressed

gene after *NKX2-1* in lung adenocarcinomas as compared with squamous cell lung carcinomas (Hofmann et al. 2006).

Genome-wide analysis of the association of *NKX2-1* with chromatin in human lung adenocarcinoma cells

To investigate the DNA-binding sites of *NKX2-1* across the genome, we performed chromatin immunoprecipitation (ChIP) from lysates of the NCI-H3122 cell line, the lung cancer cell line with the highest degree of amplification and the highest expression of *NKX2-1* (Supplemental Fig. S2A,B), followed by massively parallel sequencing (ChIP-seq). We obtained 7,173,491, and 6,118,041 alignable reads for the nonimmunoprecipitated input and the anti-*NKX2-1* immunoprecipitated sample, respectively. To avoid overrepresentation artifacts of amplified regions caused by aneuploidy and copy number alterations, we normalized the number of mappable reads within genome windows of 100 kb.

After the correction, significantly enriched *NKX2-1*-occupied peaks were analyzed by the MACS (model-based analysis of ChIP-seq) algorithm (Zhang et al. 2008). Overall, we identified 7469 significantly enriched regions ($P < 1 \times$

Watanabe et al.

10^{-5}) across the genome (Supplemental Table S2). Analysis of aggregated NKX2-1-binding regions showed sequence conservation across multiple vertebrate species, with the highest average conservation score at the center of the binding sites, representing likely functional occupancies of NKX2-1 (Supplemental Fig. S3A). NKX2-1-binding sites are more enriched in the proximity of transcription start sites (TSSs) and promoter regions (within 1 kb from the TSS) (9.2%; $P = 2.5 \times 10^{-286}$) than expected by chance (Fig. 2A,B). We identified NKX2-1-binding sites in the promoter regions of known transcriptional target genes of NKX2-1, such as *SFTPB*, *CCNB1*, and *ROR1* (Supplemental Fig. S3B; Bohinski et al. 1994; Tagne et al. 2012; Yamaguchi et al. 2012). Independent immunoprecipitation confirmed enrichment by quantitative PCR (qPCR) at all 15 arbitrarily selected enrichment sites (Supplemental Fig. S3C). We further confirmed the specific enrichment of NKX2-1 at five loci by performing ChIP using two additional antibodies against NKX2-1 as well as controls using blocking peptides (Supplemental Fig. S3D).

To investigate whether genome-wide *cis* elements of NKX2-1 are shared across lung adenocarcinoma cell lines with *NKX2-1* amplification, we performed ChIP-seq analyses on two additional lung adenocarcinoma cell lines, NCI-H1819 and NCI-H2087, both of which harbor *NKX2-1* amplification (Supplemental Fig. S2A,B). We identified NKX2-1-bound regions in NCI-H1819 and NCI-H2087 from >16 million aligned reads and compared the enriched sites between these cell lines and NCI-H3122. After normalizing the identified significant ($P < 1 \times 10^{-6}$) peak numbers to the cell line with the fewest peaks (7469 in NCI-H3122), we directly intersected the identified NKX2-1-binding regions of these cell lines (Supplemental Table S2). While there was a significant overlap among these cell lines as expected, only 559 (~7.5% of each

sample) loci were common in all three cell lines and 3171 (~17% of all sites) in any two samples, and there were many unique binding loci in each cell line (Fig. 2C).

Significant overrepresentation of the genes with adjacent NKX2-1-binding sites whose expression is correlated with NKX2-1 amplification and/or overexpression

To determine the functional consequences of NKX2-1 activity and DNA binding in lung adenocarcinomas, we sought to integrate NKX2-1-binding regions with our gene expression data set. For this analysis, we identified 1158 highly significant common NKX2-1-binding site peaks ($P < 1 \times 10^{-60}$) by combining enriched peaks in the three cell lines described above, which also displayed enrichment signals in individual cell lines (Fig. 3A; Supplemental Table S2). To generate the gene set from NKX2-1-binding sites, we then annotated the common 1158 peaks to the adjacent gene and generated a list of genes that have NKX2-1-binding sites within 30 kb of the TSS (Supplemental Table S3). Gene set enrichment analysis (GSEA) (Subramanian et al. 2005) revealed significant overrepresentation of those adjacent genes in primary tumors with high *NKX2-1* expression and in cell lines with a high *NKX2-1* copy number, suggesting that aberrant overexpression of NKX2-1 primarily activates the expression of transcriptional target genes whose expression is not *trans*-activated by NKX2-1 at physiological expression levels (Fig. 3B). In addition, while the ChIP signal was most enriched upstream of TSSs of the genes positively correlated with *NKX2-1* amplification, there was also signal enrichment upstream of TSSs of the genes negatively correlated with *NKX2-1* amplification, indicating that NKX2-1 might directly repress expression of a set of anti-correlated genes (Fig. 2A).

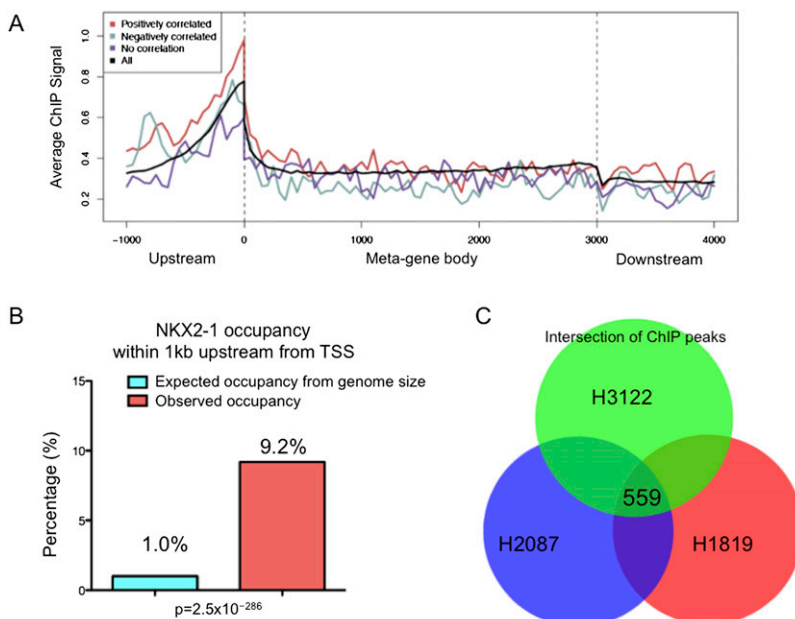


Figure 2. NKX2-1 binds preferentially to promoter regions of the genes whose expression is correlated with *NKX2-1* amplification. (A) The average NKX2-1 ChIP enrichment signal on the gene structure of all genes (gene bodies were normalized to 3 kb) was plotted in black. The signal is highest upstream of the TSS. The NKX2-1 ChIP signal is also plotted for genes whose expression is positively correlated with *NKX2-1* amplification (red) and negatively correlated (cyan) and not correlated and low in NCI-H3122 cells (purple). NKX2-1 occupancy at TSSs is highest in genes whose expression is positively correlated. (B) NKX2-1 occupancy within 1 kb upstream of TSSs is significantly higher than expected by random chance. Of all enriched peaks, 9.2% are within 1 kb upstream of TSSs of any RefSeq genes that comprise 1.0% of the effective genome size. (C) The Venn diagram depicts overlaps of 7469 each of the most significantly enriched genomic intervals by NKX2-1 ChIP for three lung adenocarcinoma cell lines that harbor *NKX2-1* amplification (NCI-H1819, NCI-H2087, and NCI-H3122).

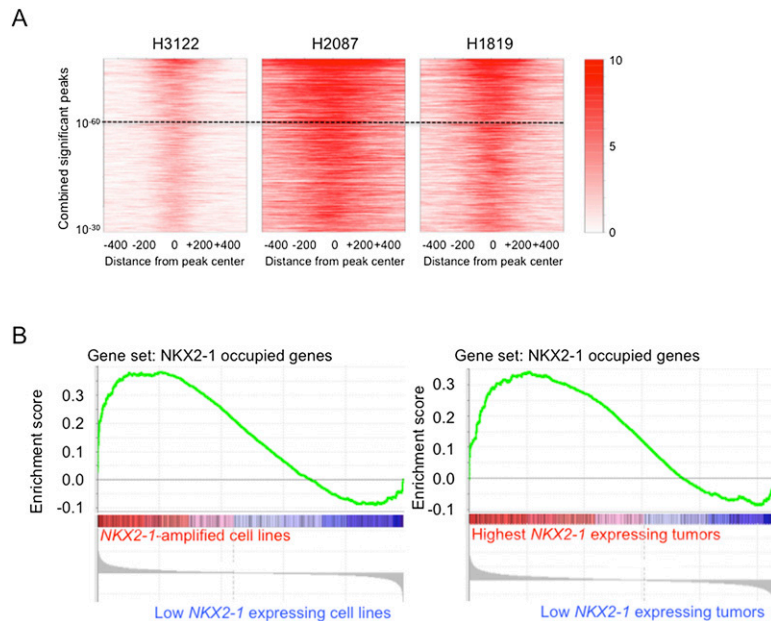


Figure 3. NKX2-1-occupied genes are significantly overrepresented in the *NKX2-1*-amplified lung adenocarcinomas. (A) ChIP signal within 0.5 kb upstream of and downstream from the center of the NKX2-1-bound peaks of 3358 common peaks that are highly significant ($P < 1 \times 10^{-30}$) and 1158 peaks below the P -value of 1×10^{-60} in a combined analysis for each cell line. Peaks are rank-ordered by combined P -value. The scale is shown on the right bar from 0 to 10 tag counts per 10-bp resolution. (B, left) The gene set with the genes that have the common NKX2-1 occupancy (1158 peaks with $P < 10^{-60}$) within 30 kb upstream of or downstream from their TSSs is significantly enriched in cell lines with high-level *NKX2-1* amplification compared with the ones with the lowest *NKX2-1* expression. (Right) The same gene set is also significantly enriched in primary lung adenocarcinomas with the highest *NKX2-1* expression compared with the lowest.

Motif-based analysis of sequences from the *NKX2-1*-binding regions revealed AP-1, Forkhead, and nuclear hormone receptor-binding motifs

Next, we sought to identify consensus NKX2-1-binding region motifs within the central 100 base pairs (bp) of NKX2-1-binding regions with the highest statistical value for enrichment using the MEME-ChIP algorithm (Machanick and Bailey 2011). We used sequences from 1158 common binding regions (combined $P < 1 \times 10^{-60}$) and identified the motif sequence *c/gTg/tGAGa/tGg/c* as the most significant (Fig. 4A). Analysis of sequences from each individual cell line identified the almost identical

motif sequence (data not shown). The identified motif is highly similar to the *Nkx3-2* consensus motif (Kim et al. 2003) and *zeste* consensus motifs in *Drosophila* (Benson and Pirrotta 1988) and has similarity to the *in vitro* motif for NKX2-1 (Berger et al. 2008) and *vnd*, the homologous gene to NKX2-1 in *Drosophila* (Wang et al. 2002).

Many transcription factors function in heterodimeric complexes with other transcription factors or coactivators to synergistically regulate gene expression. To predict potential co-occupancies of NKX2-1 with other DNA-binding proteins, we searched for sequences from the databases of known consensus motifs overrepresented in the proximity of NKX2-1-binding sites. We identified

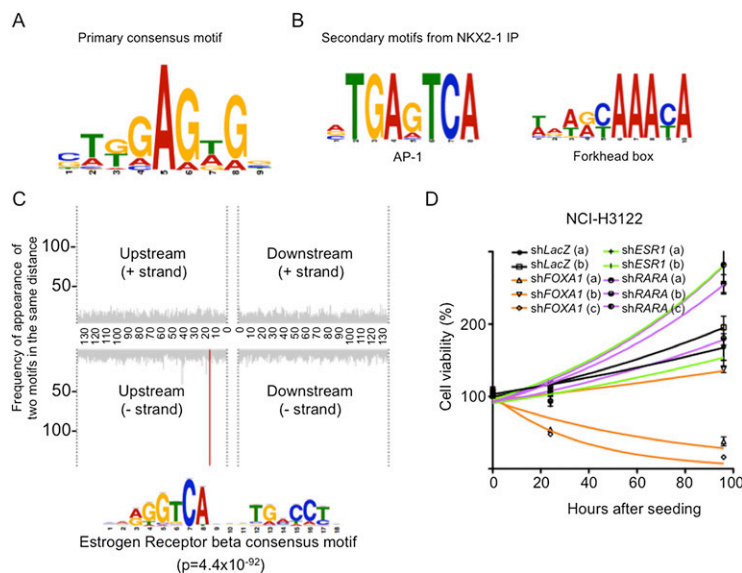


Figure 4. Motif analysis of NKX2-1-bound sequences identified AP-1, Forkhead, and nuclear hormone receptor-binding motifs. (A) De novo primary motif found in the most significant NKX2-1-bound sequences of the combined data shows similarity to the known binding motif of *Nkx3-2* and *zeste*. (B) Other enriched motifs found in common NKX2-1-occupied regions in three *NKX2-1*-amplified cell lines (NCI-H1819, NCI-H2087, and NCI-H3122). The motifs were identified to be similar to known consensus binding motifs of AP-1 and the FOX family, respectively. (C) The known motif most significantly enriched at the fixed distance from the identified primary NKX2-1-binding motif was identified as the ER β motif ($P = 4.4 \times 10^{-92}$). Of 7538 regions that contain sequences of both the identified NKX2-1-binding motif and the ER β consensus binding motif from NKX2-1-binding regions in any of three *NKX2-1*-amplified cell lines, 148 regions were found to have both motifs at a 15-bp distance. (D) Cell growth curve of an *NKX2-1*-amplified cell line, NCI-H3122, after suppression of *FOXA1*, *ESR1*, *RARA*, or *LacZ* control by shRNAs. Suppression of *FOXA1* led to reduced cell viability, whereas that of *ESR1* or *RARA* did not.

significant enrichment of sequences corresponding to AP-1 and Forkhead box (FOX) consensus binding motifs (Fig. 4B). In addition, to more specifically identify DNA motifs immediately adjacent to the primary NKX2-1-binding motifs at fixed spacing intervals, which would suggest a likely hetero-oligomeric interaction between two transcription factors and the DNA, we searched for enrichment of known DNA motifs at a given distance from the primary NKX2-1 motif. The most significantly enriched motif was identified to be the ESR2 (estrogen receptor β [ER β]) found in both common ($P = 1.1 \times 10^{-22}$; 23 of 371 sites with both motifs) and all ($P = 4.4 \times 10^{-92}$; 148 of 7538 sites with both motifs) NKX2-1-binding sequences in the three cell lines at a 15-bp distance from the primary NKX2-1 motif (Fig. 4C). The ER β motif shared commonality between other hormone receptors such as ER α and RARs. Indeed, 425 sites were found to overlap between NKX2-1-occupied regions (3358 sites, combined P -value $< 1 \times 10^{-30}$) (Fig. 3A) and published ER α -occupied regions in the MCF-7 breast cancer cell line (14,505 sites) (Supplemental Fig. S4A; Ross-Innes et al. 2010). Similarly, we attempted to compare the occupancies by NKX2-1 and RAR using RAR α ChIP-seq data (Hua et al. 2009) from the MCF-7 breast adenocarcinoma cell line and found 392 (~8%) regions commonly occupied by RAR α in MCF-7 cells (4870 sites) and by NKX2-1 in lung adenocarcinoma cell lines (Supplemental Fig. S4A) as compared with 98 sites common in 3648 GATA3-bound sites in the MCF7 cell line from the same study (Hua et al. 2009). Analysis with a statistical metric to calculate significant overlap of peaks over genome background (Chikina and Troyanskaya 2012) showed a higher number of peaks that significantly overlapped with RAR α - and ER α -binding peaks as compared with GATA3-binding peaks (Supplemental Fig. S4A).

Further inspection of the expression data set revealed that among 14 AP-1 complex-associated genes, 31 FOX genes, and eight RA/estrogen-related receptor genes, the genes whose expression is most significantly correlated with NKX2-1 overexpression include *FOSB*, *FOS*, *JUN*, *FOXA2*, *FOXA1*, and *FOXJ1* (Supplemental Fig. S4B). To determine the requirement of these potential cooperating DNA-binding proteins with NKX2-1, we examined cell viability after suppressing either of three candidate genes—*FOXA1*, *ESR1*, or *RARA*—in an NKX2-1-amplified cell line, NCI-H3122. *FOXA1* was found to be required for cell survival in this cell line, whereas *ESR1* and *RARA* were dispensable for cell survival (Fig. 4D).

The LMO3 oncogene is a transcriptional target of NKX2-1 in lung adenocarcinomas

To identify the most likely individual target genes of NKX2-1 that transduce its function in lung adenocarcinoma, we analyzed significantly correlated genes in two expression data sets (Ding et al. 2008; Sos et al. 2009) in the context of genes with nearby NKX2-1-binding sites (–3 kb to –1 kb of the TSS) in NCI-H3122 cells and found 14 genes at the intersection of all three lists (Supplemental Fig. S5A). The genes include *SFTPB*, a well-known

target gene of NKX2-1; *MBIP*, the nearest gene to *NKX2-1* that is frequently coamplified; and *LMO3*, a paralog of the *LMO1* and *LMO2* genes, which are well known as translocated oncogenes in T-ALL (Boehm et al. 1988a,b). The *LMO3* gene has also recently been reported to be translocated in T-ALL (Simonis et al. 2009) and implicated in promoting tumorigenesis of neuroblastoma (Aoyama et al. 2005). Moreover, a *Drosophila* homolog of NKX2-1, *vnd*, most preferentially binds to the *Beadex* or *dLMO* gene, the homolog of the LMO family in *Drosophila* (Wang et al. 2005). Indeed, expression of *LMO3* was the most significantly correlated of any gene with expression of *NKX2-1* in the analyzed lung adenocarcinoma tumor data set (Fig. 1B; Supplemental Table S1).

Given that multiple potential NKX2-1 target genes exhibited both NKX2-1 occupancy in NCI-H3122 cells and coexpression with *NKX2-1*, we wished to determine whether any of these novel *NKX2-1* target genes are essential in cells with amplification of the *NKX2-1* locus. To this end, we used a small-scale arrayed shRNA screening approach to determine relative essentiality of 15 genes—14 putative NKX2-1 target genes and *NKX2-1* itself—in four lung adenocarcinoma cell lines (NCI-H3122, NCI-H1819, NCI-H2087, and HCC1833) with amplification of *NKX2-1* and four lung adenocarcinoma cell lines without amplification (NCI-H23, NCI-H1437, HCC461, and Calu3) (Weir et al. 2007). We performed statistical analysis of the shRNA data set using the RIGER (RNAi gene enrichment ranking) algorithm, designed to evaluate the differential impact of multiple shRNA gene sets on different cell lines (Luo et al. 2008). Among the 15 genes tested, weighted sum analysis showed that suppression of the *NKX2-1* gene had the largest differential anti-proliferative effects on *NKX2-1*-amplified lung adenocarcinoma cell lines compared with control lung adenocarcinoma cell lines without such amplification, followed by *LMO3* with the second largest effect (Fig. 5A). These observations indicate that *LMO3* is an essential mediator downstream from *NKX2-1* in driving proliferation and survival of *NKX2-1*-amplified cell lines.

As previous studies have shown that suppression of *NKX2-1* in lung adenocarcinoma cells with amplification of *NKX2-1* results in increased apoptosis (Tanaka et al. 2007; Kwei et al. 2008), we assessed the effect by *LMO3* suppression in NCI-H3122 cells. Suppression of *LMO3* via shRNA similarly increased apoptosis, whereas it did not affect cell proliferation in cell cycle analysis (Supplemental Fig. S5B). However, ectopic expression of *LMO3* in *NKX2-1*-suppressed NCI-H3122 cells failed to overcome decreased cell viability (data not shown). Thus, it is likely that *LMO3* is not the only downstream effector of amplified *NKX2-1* and there may be multiple direct and indirect effectors, as indicated by our expression and cistromic analyses.

To determine whether *LMO3* is more abundantly expressed in lung adenocarcinomas compared with other types of tumor, we analyzed a publicly available expression data set from multiple tumor types (International Genomics Consortium, <https://expo.intgen.org/geo/listPublicGeoTransactions.do>). The *LMO3* gene is

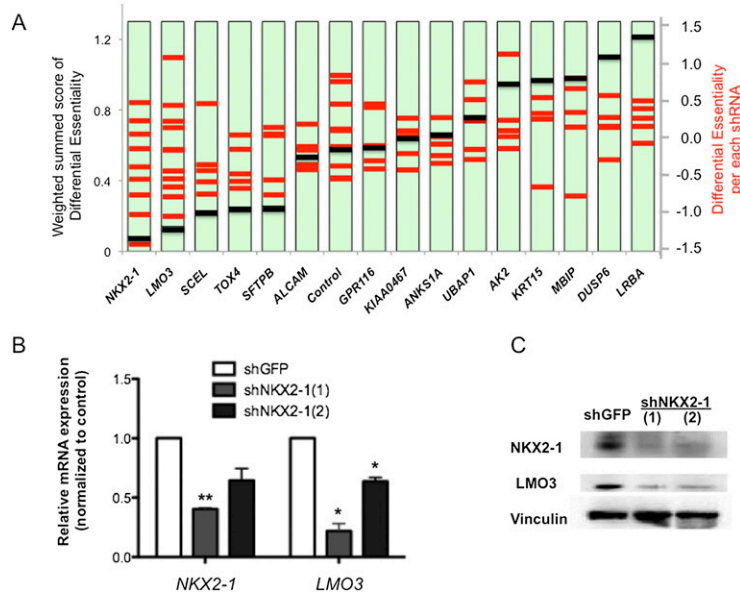


Figure 5. NKX2-1 and *LMO3* are the most differentially essential for cell survival in *NKX2-1*-amplified cell lines as compared with cell lines without amplification, and *NKX2-1* is required for expression of *LMO3*. (A) Anti-proliferative effects were tested for five to 10 shRNAs per gene on four cell lines with *NKX2-1* amplification (H1819, H2087, H3122, and HCC1833) and four cell lines without amplification or expression of *NKX2-1* (H23, H1437, HCC461, and Calu-3). Individual differential essentiality scores for each shRNA (Z-score of differential effects for four cell lines vs. four cell lines) scaled on the right Y-axis are shown in red bars. The composite differential essentiality scores (weighted sum) for each gene tested scaled on the left Y-axis are shown as black bars. (B) mRNA expression of *NKX2-1* and *LMO3* measured by RT-qPCR in NCI-H2009 cells after introduction of two shRNA against *NKX2-1* and one shRNA against *GFP* as control. *LMO3* expression is reduced relative to the suppression of *NKX2-1* expression. (*) $P < 0.05$; (**) $P < 0.01$. (C) Western blot for H2009 cell lysates with anti-NKX2-1, anti-*LMO3*, and anti-vinculin (loading control) antibodies after introduction of two shRNAs against *NKX2-1* and one shRNA against *GFP* as control.

most highly expressed in lung cancer and in thyroid cancer, where *NKX2-1* is also highly expressed (Supplemental Fig. S6A; Fabbro et al. 1994; Zhang et al. 2006). We confirmed that *LMO3* expression is much higher in lung adenocarcinomas compared with lung squamous cell carcinomas (Hofmann et al. 2006) with analysis of additional microarray data (Supplemental Fig. S6B; Bhattacharjee et al. 2001). This result is consistent with our previous finding that *NKX2-1* is primarily amplified in lung adenocarcinoma but not in squamous cell lung carcinomas (Bass et al. 2009). In situ hybridization of a probe against the *Lmo3* gene in normal adult mouse lungs showed that it is highly expressed in bronchiolar epithelium, where *Nkx2-1* is known to be highly expressed (Supplemental Fig. S6C), but is not expressed in the proximal airway, where *Nkx2-1* levels are normally low. However, *Lmo3* was not found to be expressed in alveolar type II cells (Supplemental Fig. S6C), whereas *NKX2-1* is known to be expressed in both bronchiolar epithelium and alveolar type II cells (Yatabe et al. 2002). These results potentially reflect a similarity in cell lineage between lung adenocarcinomas with *NKX2-1* amplification and the bronchiolar epithelium, although both Clara cell (*SCGB1A1* and *CYP2B6*) and alveolar type II cell (*SFTPC*) markers are highly expressed in lung adenocarcinomas that overexpress *NKX2-1* (Supplemental Table S1).

To evaluate the requirement of *NKX2-1* for *LMO3* expression in the cell lines that express both genes, we suppressed the expression of *NKX2-1* using two previously validated shRNAs (Weir et al. 2007). *LMO3* expression was significantly reduced after *NKX2-1* suppression on both the mRNA and protein levels (Fig. 5B,C). To exclude the possibility that reduced expression is a general consequence of decreased transcription by reduced viability of cells with *NKX2-1* suppression, we investigated the degree of decrease of *LMO3* expression by *NKX2-1* compared with the effects on other transcripts

by performing expression profiling of NCI-H2009 cell lines stably expressing two shRNAs against *NKX2-1* or shRNA against *GFP*. *LMO3* was among the 50 most differentially expressed genes between NCI-H2009 cells with shGFP and shNKX2-1, confirming that *LMO3* is a transcriptional target of *NKX2-1* in a cell line with amplification of *NKX2-1* (Supplemental Fig. S6D).

Expression of LMO3 is directly regulated by NKX2-1 in lung adenocarcinomas

To verify the enrichment of *NKX2-1* binding and characterize its precise binding location on the *LMO3* gene locus, we designed several primer sets surrounding the *LMO3* TSS, performed *NKX2-1* ChIP, and analyzed ChIP enrichment by qPCR using the NCI-H3122 cell line. Corresponding to the peak center of enrichment with ChIP-seq data, the highest enrichment of *NKX2-1* was observed with a primer set spanning 693–845 bp downstream from the TSS of the *LMO3* gene, where the sequence is conserved across several species (Supplemental Fig. S7A,B).

Next, we generated reporter constructs to assess the ability of *NKX2-1* binding to the *LMO3* locus to activate expression of the *LMO3* gene. The DNA sequence downstream (103–1149 bp) from the TSS of the *LMO3* gene was cloned upstream of the luciferase gene, and luciferase activity was measured after transfection of 293T cells. We observed induction of luciferase activity when the reporter plasmid was cotransfected with the *NKX2-1* expression vector, although it was modest possibly due to the lack of expression of cofactors in the cellular context of the renal embryonic cell line (Supplemental Fig. S7C).

ChIP after formaldehyde cross-linking enriches DNA loci occupied by the immunoprecipitated protein either directly or indirectly as a multiprotein complex. To determine whether the interaction between *NKX2-1* and

the *LMO3* locus is direct, we tested the binding of in vitro translated NKX2-1 protein with two independent DNA sequences (+773–898 and +823–962) from the *LMO3* locus using an electrophoretic mobility shift assay (EMSA) (Fig. 6A; Supplemental Fig. S7B). Electrophoretic mobility of both *LMO3* sequences was shifted when NKX2-1 protein was added and further shifted (supershift) by addition of anti-NKX2-1 antibody, suggesting direct binding of NKX2-1 to the *LMO3* gene locus (Fig. 6A).

Regulation of LMO3 expression by NKX2-1 is cellular context-dependent and may require cooperative binding with FOXA1

We further evaluated the ability of NKX2-1 to induce expression of the *LMO3* gene in six different cell lines with and without endogenous expression of *NKX2-1*. Ectopic expression of NKX2-1 was able to augment the expression of the *LMO3* gene in the three cell lines that harbor genomic amplification and expression of *NKX2-1*. We did not observe a similar induction in three other cell lines—A549, AALE tracheobronchial epithelial cells, and 293T cells, all of which do not express endogenous *NKX2-1*—whereas *SFTPB* expression was induced in all cell lines except for a nonlung epithelial cell line, 293T (Fig. 6B). In addition, in two lung adenocarcinoma cell lines—HCC78 and H2347—that express *NKX2-1* at high levels without *NKX2-1* amplification (Supplemental Fig. S2A), ectopic expression of *NKX2-1* failed to induce *LMO3* expression (Supplemental Fig. S8A). This suggests that NKX2-1 requires a specific cellular context such as expression of cofactors or a permissive chromatin structure to enable NKX2-1 transactivation of the *LMO3* gene.

While we were able to validate enrichment of the *LMO3* locus as well as the *SFTPB* promoter locus from ChIP-seq data in two additional cell lines—NCI-H1819 and NCI-H2087—that harbor amplification of *NKX2-1* (Fig. 6C), the enrichment of these loci was not observed in the A549 lung adenocarcinoma cell line that lacks endogenous NKX2-1 expression, confirming the specificity of antibody enriching this region (Fig. 6C; Supplemental Fig. S2B). When we ectopically overexpressed NKX2-1 in the A549 cell line, we observed slower cellular growth compared with the control GFP-expressing A549 cell line, consistent with a tumor-suppressive role of NKX2-1 in A549, a mucinous lung adenocarcinoma cell line (Supplemental Fig. S8B; Lieber et al. 1976; Maeda et al. 2012). In A549 cells with ectopic *NKX2-1* expression, NKX2-1 localized to the *SFTPB* promoter region, whereas we could not detect NKX2-1 occupancy at the *LMO3* locus, again suggesting that in this cellular context, additional cofactors are required for NKX2-1 to localize to this *LMO3* site (Fig. 6C).

Since the FOX motif was one of the enriched motif sequences adjacent to NKX2-1-binding regions and the *FOXA1* gene is coamplified with *NKX2-1* at chromosomal segment 14q13 in 80%–90% of *NKX2-1*-amplified cases (Supplemental Fig. S8C) and is known to play an essential role in lung development (Wan et al. 2005), we tested whether FOXA1 can co-occupy the *LMO3* locus by ChIP analysis using an antibody against FOXA1. We found that FOXA1 bound to the *LMO3* locus upstream of the NKX2-1-binding site (Fig. 7A). By independent ChIP-qPCR, we observed that the highest enrichment of FOXA1 binding was with a primer set spanning 385–510 bp downstream from the TSS of the *LMO3* gene (Supplemental Fig. S7A). We also found that four of the *NKX2-1*-

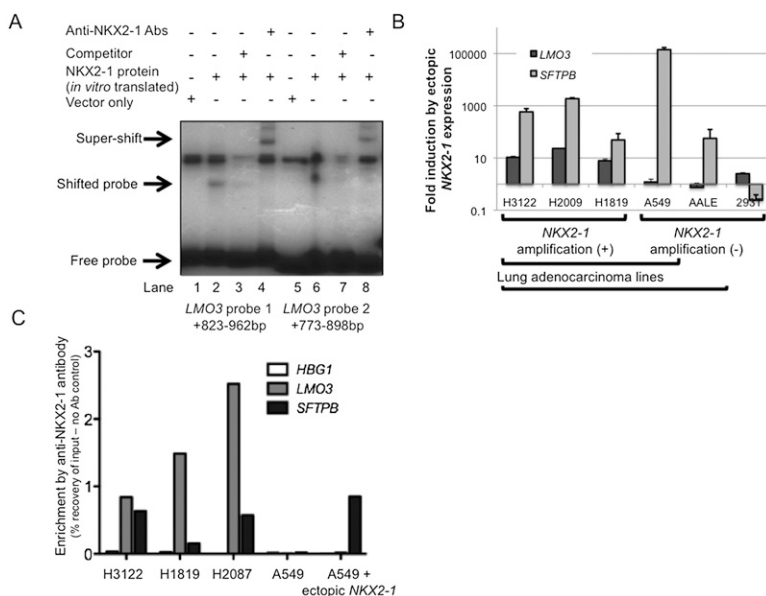


Figure 6. NKX2-1 binds to ~0.8 kb downstream from the TSS of the *LMO3* gene and *trans*-activates *LMO3* expression in cells with *NKX2-1* amplification but not in cells without *NKX2-1* amplification. (A) EMSA with in vitro translated NKX2-1 protein for two ³²P-labeled probes designed from the *LMO3* gene locus. Electrophoretic mobility was shifted when incubated with NKX2-1 protein (lanes 2,6) and further shifted (supershift) by addition of anti-NKX2-1 antibody (lanes 4,8). (Lanes 3,7) Mobility shift was diminished with incubation of nonradioactive self-competitor probes. (B) Ectopic expression of *NKX2-1* induced ~10-fold expression of *LMO3* in three cell lines (H3122, H2009, and H1819) that harbor *NKX2-1* amplification. *LMO3* expression was not induced to a similar level or reduced after ectopic *NKX2-1* expression in three cell lines without *NKX2-1* amplification (A549, AALE, and 293T cells). (C) ChIP enrichment (percent recovery of input) by anti-NKX2-1 antibody measured by qPCR at 0.7 kb downstream from *LMO3* TSS and the *SFTPB* promoter for three cell lines (H1819, H2087, and H3122) with *NKX2-1* amplification, A549 without *NKX2-1* amplification, and A549 expressing ectopic *NKX2-1*.

HBG1 serves as a negative control. NKX2-1 localizes to both *LMO3* and *SFTPB* loci in *NKX2-1*-amplified cell lines but was not able to localize to the *LMO3* locus in A549, whereas NKX2-1 was still able to occupy the *SFTPB* promoter region.

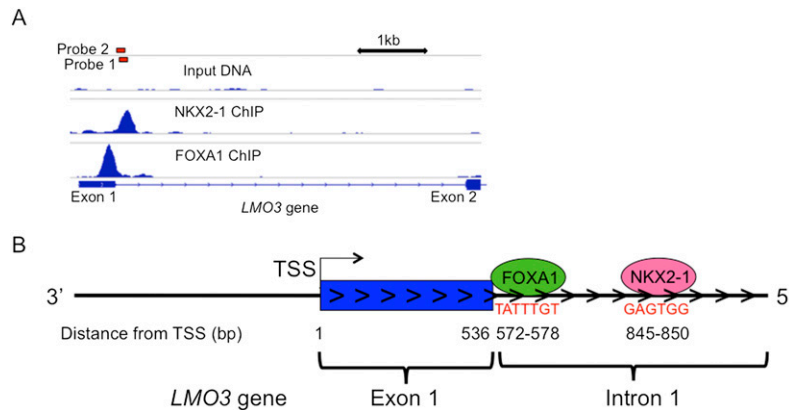


Figure 7. NKX2-1 and FOXA1 colocalize at the genomic regions, including *LMO3*. (A) A view of the *LMO3* gene locus of shifted read counts from ChIP-seq by either the anti-NKX2-1 or anti-FOXA1 antibody as well as read counts from the no-ChIP input control. The location of sequences taken for two probes designed for gel shift assay is shown as red bars. (B) The model depicting cooperative occupancy of NKX2-1 and FOXA1 on the locus downstream from the TSS of the *LMO3* gene to *trans*-activate expression of *LMO3*.

amplified cell lines that we examined have much higher endogenous expression of FOXA1 protein compared with A549 or 293T cells (Supplemental Fig. S2B). Furthermore, ectopic expression of *FOXA1* in A549 cells potentiated the ability of forced expression of *NKX2-1* to induce *LMO3* expression (Supplemental Fig. S8D). Suppression of FOXA1 in NCI-H3122, which resulted in decreased cell survival, led to reduced *LMO3* expression (Supplemental Fig. S8E). Finally, genome-wide analysis of FOXA1 occupancy by ChIP-seq in the NCI-H3122 cell line revealed highly significant overlap with NKX2-1 occupancy (47% of NKX2-1-bound sites) (Supplemental Fig. S8F), suggesting that multiple NKX2-1 target genes could be coregulated by FOXA1 in the *NKX2-1*-amplified cells and that coregulation of not only *LMO3* but other target genes by NKX2-1 and FOXA1 might be a critical mechanism for cell survival of *NKX2-1*-amplified lung adenocarcinomas.

Discussion

NKX2-1 is the most commonly amplified gene in lung adenocarcinoma but plays a complex role in the regulation of lung adenocarcinogenesis. On the one hand, *NKX2-1* is required for the growth of lung adenocarcinoma cells bearing its amplification (Kendall et al. 2007; Tanaka et al. 2007; Weir et al. 2007; Kwei et al. 2008), suggesting that it is a dominant oncogene; this concept is corroborated by the activating translocations of *NKX2-1* seen in acute T-cell lymphoblastic leukemias (Homminga et al. 2011). On the other hand, loss of *NKX2-1* expression promotes metastasis in mouse models, is associated with tumor suppression of *Kras*-driven murine mucinous lung adenocarcinomas, and is associated with an inferior outcome in human lung adenocarcinomas (Winslow et al. 2011; Maeda et al. 2012).

Understanding the downstream signaling from *NKX2-1* can help to identify targets of its transcriptional activity and shed light on its complex roles in lung adenocarcinoma. Here, we used an integrated analysis of data from ChIP, gene expression correlative analysis, and RNAi to nominate the *LMO3* transcriptional regulatory gene as a functional target of *NKX2-1*.

LMO3 is a paralog of *LMO1/2* oncogenes that are recurrently translocated and, as a consequence, overexpressed in T-ALL (Boehm et al. 1988a,b). The *LMO3* gene itself is translocated in at least one T-ALL case (Simonis et al. 2009). Although the mechanism of leukemogenesis by *LMO* gene rearrangements is still unknown, various lines of evidence suggest that the *LMO* genes play a role in differentiation, such as a recent study suggesting that *Lmo2* might reactivate the self-renewing capacity of T-cells in preleukemic thymocytes (McCormack et al. 2010). A developmental/differentiation role for *LMO* genes is consistent with the role of NKX2-1 as a master regulator of lung alveolar cell lineage commitment (Boggarum 2009). Mice lacking *Nkx2-1* fail to develop normal branching morphogenesis of the lung (Yuan et al. 2000). It is conceivable that the developmental roles of *NKX2-1* and the *LMO* genes are linked and that these links may extend to their respective roles in tumorigenesis.

The activity of NKX2-1 to bind to the *LMO3* locus was restricted to a subset of lung adenocarcinoma cell lines; in the limited set of tested cells, only the lines that were *NKX2-1*-amplified showed *LMO3* induction by *NKX2-1*. This result suggested that an essential cofactor might also be required for *NKX2-1* to *trans*-activate *LMO3*. Our data suggest that FOXA1 may be such a cofactor. FOXA1 coexpression with NKX2-1 induces *LMO3* expression, and FOXA1 binds to the *LMO3* locus.

FOX proteins are a family of transcription factors that play roles in regulating genes involved in cell growth, proliferation, and differentiation. While there are as many as 46 genes encoding FOX proteins that have been described in humans, a few have been well characterized and known for consensus binding motifs. Of those, FOXA1/2, FOXJ1, and FOXP1/2 have been implicated in lung morphogenesis. FOXA1/2 are essential in early lung development (Costa et al. 2001), FOXJ1 is essential for differentiation of ciliated epithelial cells in the lung (Costa et al. 2001), and FOXP1/2 has been shown to be important for lung alveolization (Shu et al. 2007). While multiple forkhead proteins may cooperate with NKX2-1 independently, FOXA1 is of particular interest because it is located at chromosomal segment 14q13.3, where *NKX2-1* resides,

and is often coamplified in lung adenocarcinoma. FoxA1 has been shown to interact with Nkx2-1 to regulate the expression of lung-specific genes such as *Sftpb* and *Scgb1a1* (Minoo et al. 2007). We showed that NKX2-1 and FOXA1 bind to many common genomic loci in NCI-H3122 cells, including the *LMO3* gene locus (Fig. 7A; Supplemental Fig. S8F).

Our data provide a new insight into the transcriptional regulatory network of NKX2-1 in *NKX2-1*-amplified lung adenocarcinoma and have nominated *LMO3* as a key transducer of NKX2-1, with FOXA1 as a major cooperative factor with NKX2-1. While aberrant activation of transcriptional regulation for lineage commitment has proven to be important in hematopoietic malignancy, we have just begun to understand dysregulated transcriptional networks in lung adenocarcinomas. Further studies of the NKX2-1 transcriptional network and cooperation between NKX2-1, FOXA1, and *LMO3* as transcriptional regulators in lung adenocarcinoma may elucidate important tumor vulnerabilities.

Materials and methods

Cell culture

NCI-H2347, NCI-H1819, NCI-H2087, NCI-H3122, NCI-H2009, HCC78, HCC1833, and A549 cells were provided by Dr. John Minna (University of Texas Southwestern Medical Center). Cells were maintained in RPMI 1640 growth medium (Mediatech) supplemented with 10% FBS and 1 mM penicillin/streptomycin (Mediatech). Immortalized tracheobronchial epithelial cells (AALE cells) were prepared as described previously (Lundberg et al. 2002) and maintained in SAGM growth medium (Lonza). Stable *NKX2-1* ORF or *GFP*-expressing NCI-H2347, NCI-H3122, NCI-H2009, HCC78, HCC1833, A549, 293T, and AALE cells were established with use of a lentiviral vector (pLEX-Blast) with a standard viral transfer protocol. Stable *FOXA1* ORF or *GFP*-expressing A549 with ectopic *NKX2-1* expression were established with use of a lentiviral vector (pLEX-puro) with a standard viral transfer protocol.

ChIP

About 20 million cells per cell line were cross-linked with 1% formaldehyde for 10 min at 37°C, washed in 5 mg/mL BSA in PBS and then in just cold PBS, resuspended in lysis buffer (50 mM Tris-HCl at pH 8.1, 10 mM EDTA, 1% SDS, 1× complete protease inhibitors [Roche]), and sonicated with the Covaris E210 sonicator to obtain chromatin fragment lengths of 200–1500 bp judged by the Bioanalytical DNA High-Sensitivity kit (Agilent). Fragmented chromatin was diluted in immunoprecipitation buffer (20 mM Tris-HCl at pH 8.1, 150 mM NaCl, 2 mM EDTA, 1% Triton X-100) and incubated overnight at 4°C with protein G magnetic beads (Dynabeads, Invitrogen) that had been pre-incubated with anti-NKX2-1 (Bethyl Laboratories, BL4000 and BL3998; or Santa Cruz Biotechnology, H-190), anti-FOXA1 (Abcam, ab23738), or control rabbit IgG (Bethyl Laboratories) antibodies. Immunoprecipitates were washed six times with wash buffer (50 mM HEPES at pH 7.6, 0.5 M LiCl, 1 mM EDTA, 0.7% Na deoxycholate, 1% NP-40) and twice with TE buffer. DNA was recovered in 1% SDS plus 0.1 M NaHCO₃ over a period of 6 h at 65°C, column-purified with QiaQuick columns (Qiagen), and quantified with use of a PicoGreen assay (Invitrogen). DNA was used as the template for qPCR or for library construction of Illumina sequencing.

All qPCRs were performed in duplicate with Power PCR SYBR Green Master mix (Applied Biosystems) on a 7300 Real-Time PCR system (Applied Biosystems). For quantification of DNA recovery, values for input were adjusted to 100%, and percentages were calculated with use of a fraction of the values from immunoprecipitated DNAs against values for the corresponding 100% input DNAs. The primers used are listed in Supplemental Table S4.

For ChIP-seq, DNA that had been immunoprecipitated with the appropriate antibody and input DNA without immunoprecipitation were treated with RNase A and Proteinase K, followed by sonication to enrich DNA fragment lengths between 100 and 300 bp. Ten nanograms to 50 ng of DNA was used for the library construction. DNA libraries for Illumina cluster generation were prepared according to the manufacturer's protocol. Sequencing was done with Illumina GAIIx or HiSeq 2000.

ChIP-seq peak finding analysis and motif analysis

Sequenced reads were mapped to reference genome build 37, human (hg18), using the Bowtie aligner (Langmead et al. 2009). Alignable reads were subsequently normalized for copy number variation to average DNA copies on each cell line used. Specifically, we counted a number of reads aligned to the genomic loci in each bin size of 100,000 bases. Then, we calculated the average number of the alignable reads per bin and randomly omitted reads in order to adjust to the average copy for each cell line.

Binding sites for NKX2-1 or FOXA1 were detected with the use of MACS as previously described (Zhang et al. 2008), with a *P*-value cutoff of 10⁻⁵ or 10⁻⁶ and with default values for other parameters. Data are publicly available at Gene Expression Omnibus (GEO; accession no. GSE39998).

Identified NKX2-1-binding peaks were then used for analysis for average profiling of genomic features with use of a module of the CEAS (*cis*-regulatory element annotation system) algorithm (Shin et al. 2009). Metagene average profiling was calculated by default setting (3-kb range of normalized gene structures) for either all genes, 260 genes that were positively correlated with NKX2-1 overexpression, 144 genes that were negatively correlated, or 202 genes that were not correlated and expressed at a low level in NCI-H3122 cells.

An equal number (7469) of peaks with the highest statistical significance for enrichment was used to determine the intersection of binding regions from different cell lines. Then, numbers of genomic intervals shared by two or more samples were counted and represented in a Venn diagram. For comparison of NKX2-1 and FOXA1 binding, additional NKX2-1 ChIP-seq was performed on NCI-H3122 cell lysates, and immunoprecipitation and sequencing were done simultaneously under the same condition with FOXA1 ChIP-seq on Illumina HiSeq 2000. These data were used for all visualizations of ChIP-seq data on the genome viewer (Fig. 6C; Supplemental Figs. S4B, S8B).

Significantly enriched peaks in the combined data of three cell lines were also analyzed by first detecting all potential enriched sites with a very low threshold level (nominal *P* < 10⁻²) by MACS for the data from each cell line and intersected common genomic intervals in three data and calculating each combined statistical value using a modified Stouffer's method. Enriched sequence motifs in 1158 binding peaks with the highest statistical significance (*P* < 10⁻⁶⁰) were further determined with the use of the MEME ChIP module in the MEME suite (Machanic and Bailey 2011). Known motifs at the fixed distance from the primary NKX2-1 motif identified from the MEME algorithm were searched using the SpaMo (Whittington et al. 2011) module in the MEME suite as described. For spacing analysis, 1049 intersected NKX2-1-binding peaks of three cell lines or 35,791 NKX2-1-binding peaks in any of three cell lines were used.

Expression data analyses and GSEA

For identification of those genes whose expression is correlated with *NKX2-1*, we used a gene-centric CDF file to generate expression data from existing raw expression files of 470 primary lung adenocarcinomas samples (Director's Challenge Consortium for the Molecular Classification of Lung Adenocarcinoma et al. 2008) and 42 lung adenocarcinoma cell lines with matched copy number data from Affymetrix 250K SNP arrays (Sos et al. 2009). We applied RMA (robust multichip average) and quantile normalization as well as the matchprobes package in the Bioconductor framework to create one single data set. To identify genes linked to *NKX2-1* from the primary lung adenocarcinoma data set, we used the 47 (top 10th percentile) samples that had the highest *NKX2-1* expression and the 47 (bottom 10th percentile) samples with the lowest *NKX2-1* expression, and from the cell line data set, the 11 cell lines with highest *NKX2-1*-amplification and the 17 cell lines without *NKX2-1* amplification or expression were used. To identify correlated genes, we calculated differential expression with use of the samr (SAM in R) package in Bioconductor.

To determine the enrichment of NKX2-1-binding sites at the loci of correlated genes, we first annotated all binding sites with regard to their relative distance from neighboring RefSeq genes. Genes that had NKX2-1-binding sites within 30 kb upstream of or downstream from their TSSs were used as sets for GSEA (Subramanian et al. 2005) on the same expression data described above.

RNAi

shRNA sequences that target *NKX2-1*, *LMO3*, *FOXA1*, *ESR1*, *RARA*, *GFP*, and *LacZ* were cloned into either the pLKO.1 lentiviral vector or the doxycycline-inducible Tet-On-pLKO.1 vector, obtained from the RNAi Consortium (TRC) at the Dana-Farber Cancer Institute. The virus was produced according to TRC's protocol (<http://www.broadinstitute.org/rnai/trc>). Briefly, the pLKO.1-shRNA plasmid was transfected into 293T cells, along with the packaging vectors pCMV-dR8.91 and VSV-G, and viral particles were collected in RPMI-1640 medium at 24 and/or 48 h. Virus titers were assayed in NIH-3T3 mouse fibroblasts to determine the minimum amount of virus needed to produce ~100% infection. Target cells were seeded on 10-cm plates on the day prior to the infection and were subsequently incubated for 4–6 h with diluted virus in 8 µg/mL polybrene. Twenty-four hours post-infection, fresh medium containing 1–2 µg/mL puromycin was added, and cells were selected in puromycin for 72–96 h until controls of noninfected cells were no longer visibly viable. All target sequences for shRNA are listed in Supplemental Table S5.

Arrayed RNAi screen

Lentiviral vectors containing shRNA sequences were obtained from TRC. For two genes that we sought to evaluate (*NKX2-1* and *LMO3*), the 10 top-scoring constructs were used for the RNAi screen. For the remaining 13 genes to be tested in the RNAi screen (*AK2*, *ALCAM*, *ANKS1A*, *DUSP6*, *GPR116*, *KIAA0467*, *KRT15*, *LRBA*, *MBIP*, *SCEL*, *SFTPB*, *TOX4*, and *UBAP1*), five shRNAs per gene in the TRC collection were used and analyzed. For the controls, two shRNAs each against *GFP*, *RFP*, and *LacZ* and one shRNA against *Luciferase* as well as one empty control from pLKO.1 and two from pLKO.3-pgw were included. The screens were performed at the DFCI RNAi Screening Facility. For screening cell lines with this lentivirus, cells were plated in 384-well microtiter plates and, on the following day, infected

with 1–2 µL of shRNA lentivirus in the presence of 8 µg/mL polybrene. All screens were performed in quadruplicate, with two replicates in the presence and two replicates in the absence of puromycin, which was added 24 h post-infection. Six days post-infection, wells were assayed using CellTiter-Glo (Promega) according to the manufacturer's instructions. Raw luminescence scores for the replicate wells with puromycin for a given shRNA construct in each cell line were normalized against readings in that cell line for controls. The results of the RNAi screen were analyzed with the RIGER method. RIGER compares the effects of each shRNA construct on the four *NKX2-1*-amplified lung adenocarcinoma lines with the construct's effects in the four control cell lines to determine an enrichment score for each construct. Lower enrichment scores signify a greater decrease in proliferation in the *NKX2-1*-amplified cell lines with the shRNA constructs for a given gene. The enrichment scores were normalized against an enrichment score that would be generated by random permutation of an shRNA set of the same size to generate a normalized enrichment score for each gene based on the composite scores for the constructs targeting that gene compared with other genes in the screen. In the generation of the normalized score for each gene, the algorithm assumes that not all constructs will be effective at reducing target expression. Comparison of the actual data to this permutation allows calculation of nominal *P*-values and the false discovery rate (FDR).

EMSA

The DNA fragment that lies upstream of the *LMO3* gene was PCR-amplified from human genomic DNA; a total of 250 ng of PCR products was end-labeled with ³²P-dATP by T4 polynucleotide kinase and purified with G-50 columns (GE Healthcare). For DNA-binding reactions, *NKX2-1* ORF was cloned into the pcDNA3 vector. The cloned vector and a control empty pcDNA3 vector were transcribed with T7 polymerase and translated in vitro with use of the TNT Quick-Coupled Transcription/Translation system (Promega) as per the manufacturer's instructions. Prior to DNA-binding reactions, 2 µL of the in vitro translated proteins was incubated in binding buffer (150 mM KCl, 10 mM Tris-HCl at pH 8.0, 0.5 mM EDTA, 0.2 mM DTT, 0.1% Triton X-100, 12.5% glycerol) for 10 min at room temperature with 1 µg of poly(dI-dC) and, when indicated, 1 µg of anti-NKX2-1 antibody (BL4000; Bethyl Laboratories) for supershift assay or 200 ng of nonradiolabeled self-DNAs for competition. The reaction was incubated for an additional 20 min at room temperature after adding the labeled DNA probes (1000 cpm) and electrophoresed on a 6% nondenaturing polyacrylamide gel.

Cell viability assay

Cells are selected for 72 h with puromycin or blasticidin S 24 h after infection and seeded in 96-well assay plates in replicates. Each plate at the indicated time was incubated with CellTiter-Glo reagents (Promega), and luminescence intensity was measured with use of a SpectraMax plate reader (Molecular Device). Luminescence values were normalized to the measurement at day 1.

RT-qPCR

RNA was extracted from cells using the RNeasy minikit (Qiagen) according to the manufacturer's instructions. Reverse transcription was performed with use of the QuantiTECT cDNA synthesis kit (Qiagen). All qPCRs were performed in duplicate with Power PCR SYBR Green Master Mix (Applied Biosystems) on a 7300 Real-Time PCR system (Applied Biosystems). For

Watanabe et al.

quantification of mRNA expression in each sample, values were normalized to expression values for *GAPDH*. The primers used are listed in Supplemental Table S4.

Western blots

Cell lysates were prepared by incubating cells in lysis buffer (150 mM NaCl, 50 mM Tris-HCl at pH 8.0, 1 mM EDTA, 1% Triton X-100, 0.25% NP-40, protease inhibitors) for 20 min at 4°C. After centrifugation to remove insoluble debris, lysates were immunoblotted with use of an anti-NKX2-1 antibody (Bethyl Laboratories, BL-4000), anti-LMO3 antibody (a gift from Dr. Nakagawara, Chiba Cancer Center, Chiba, Japan), anti-FOXA1 antibody (Abcam, ab23738), or an anti-vinculin antibody (Sigma).

Statistical analyses

Statistical analyses were performed with the use of Prism software. Statistical significance was determined using a Student's *t*-test comparison for unpaired data and was indicated as follows: $P < 0.05$ (*) and $P < 0.01$ (**).

Acknowledgments

We thank C.Z. Zhang, Q. Ma, B. Weir, A. Kostic, T. Westerling, H. He, J.K. Lee, N. Hendrick, W. Song, A. Bass, S. Kobayashi, and D. Tenen for technical advice and discussions. H.W. was supported by the NIH/NCI T32 Institutional Training Program fellowship (5T32CA009361-28). This work was supported by National Cancer Institute grants 5R01CA109038 and 5P20CA90578 (to M.M.). M.M. is a paid consultant for and equity holder in Foundation Medicine, a genomics-based oncology diagnostics company. M.M. and W.C.H. are paid consultants for Novartis Pharmaceuticals. Both companies have no connection or rights to the work described in this study.

References

Aoyama M, Ozaki T, Inuzuka H, Tomotsune D, Hirato J, Okamoto Y, Tokita H, Ohira M, Nakagawara A. 2005. LMO3 interacts with neuronal transcription factor, HEN2, and acts as an oncogene in neuroblastoma. *Cancer Res* **65**: 4587–4597.

Barletta JA, Perner S, Iafrate AJ, Yeap BY, Weir BA, Johnson LA, Johnson BE, Watanabe H, Rubin MA, Travis WD, et al. 2009. Clinical significance of TTF-1 protein expression and TTF-1 gene amplification in lung adenocarcinoma. *J Cell Mol Med* **13**: 1977–1986.

Bass A, Watanabe H, Mermel C, Yu S, Perner S, Verhaak R, Kim S, Wardwell L, Tamayo P, Gat-Viks I, et al. 2009. SOX2 is an amplified lineage-survival oncogene in lung and esophageal squamous cell carcinomas. *Nat Genet* **41**: 1238–1280.

Bejarano PA, Baughman RP, Biddinger PW, Miller MA, Fenoglio-Preiser C, al-Kafaji B, Di Lauro R, Whitsett JA. 1996. Surfactant proteins and thyroid transcription factor-1 in pulmonary and breast carcinomas. *Mod Pathol* **9**: 445–452.

Benson M, Pirrotta V. 1988. The *Drosophila* zeste protein binds cooperatively to sites in many gene regulatory regions: Implications for transvection and gene regulation. *EMBO J* **7**: 3907–3915.

Berger MF, Badis G, Gehrke AR, Talukder S, Philippakis AA, Peña-Castillo L, Alleyne TM, Mnaimneh S, Botvinnik OB, Chan ET, et al. 2008. Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences. *Cell* **133**: 1266–1276.

Bhattacharjee A, Richards WG, Staunton J, Li C, Monti S, Vasa P, Ladd C, Beheshti J, Bueno R, Gillette M, et al. 2001. Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. *Proc Natl Acad Sci* **98**: 13790–13795.

Boehm T, Baer R, Lavenir I, Forster A, Waters JJ, Nacheva E, Rabbitts TH. 1988a. The mechanism of chromosomal translocation t(11;14) involving the T-cell receptor C δ locus on human chromosome 14q11 and a transcribed region of chromosome 11p15. *EMBO J* **7**: 385–394.

Boehm T, Buluwela L, Williams D, White L, Rabbitts TH. 1988b. A cluster of chromosome 11p13 translocations found via distinct D-D and D-D-J rearrangements of the human T cell receptor δ chain gene. *EMBO J* **7**: 2011–2017.

Bogaram V. 2009. Thyroid transcription factor-1 (TTF-1/Nkx2.1/TTF1) gene regulation in the lung. *Clin Sci (Lond)* **116**: 27–35.

Bohinski RJ, Di Lauro R, Whitsett JA. 1994. The lung-specific surfactant protein B gene promoter is a target for thyroid transcription factor 1 and hepatocyte nuclear factor 3, indicating common factors for organ-specific gene expression along the foregut axis. *Mol Cell Biol* **14**: 5671–5681.

Chikina MD, Troyanskaya OG. 2012. An effective statistical evaluation of ChIPseq dataset similarity. *Bioinformatics* **28**: 607–613.

Costa RH, Kalinichenko VV, Lim L. 2001. Transcription factors in mouse lung development and function. *Am J Physiol Lung Cell Mol Physiol* **280**: L823–L838.

Dalla-Favera R, Bregni M, Erikson J, Patterson D, Gallo RC, Croce CM. 1982. Human c-myc onc gene is located on the region of chromosome 8 that is translocated in Burkitt lymphoma cells. *Proc Natl Acad Sci* **79**: 7824–7827.

Delattre O, Zucman J, Plougastel B, Desmaziere C, Melot T, Peter M, Kovar H, Joubert I, de Jong P, Rouleau G. 1992. Gene fusion with an ETS DNA-binding domain caused by chromosome translocation in human tumours. *Nature* **359**: 162–165.

Ding L, Getz G, Wheeler D, Mardis E, McLellan M, Cibulskis K, Sougnez C, Greulich H, Muzny D, Morgan M, et al. 2008. Somatic mutations affect key pathways in lung adenocarcinoma. *Nature* **455**: 1069–1144.

Director's Challenge Consortium for the Molecular Classification of Lung Adenocarcinoma, Shedden K, Taylor JMG, Enkemann SA, Tsao M-S, Yeatman TJ, Gerald WL, Eschrich S, Jurisica I, Giordano TJ, et al. 2008. Gene expression-based survival prediction in lung adenocarcinoma: A multi-site, blinded validation study. *Nat Med* **14**: 822–827.

Fabbro D, Di Loreto C, Beltrami CA, Belfiore A, Di Lauro R, Damante G. 1994. Expression of thyroid-specific transcription factors TTF-1 and PAX-8 in human thyroid neoplasms. *Cancer Res* **54**: 4744–4749.

Garraway LA, Sellers WR. 2006. Lineage dependency and lineage-survival oncogenes in human cancer. *Nat Rev Cancer* **6**: 593–602.

Garraway LA, Widlund HR, Rubin MA, Getz G, Berger AJ, Ramaswamy S, Beroukhi R, Milner DA, Granter SR, Du J, et al. 2005. Integrative genomic analyses identify MITF as a lineage survival oncogene amplified in malignant melanoma. *Nature* **436**: 117–122.

Govindan R, Ding L, Griffith M, Subramanian J, Dees ND, Kanchi KL, Maher CA, Fulton R, Fulton L, Wallis J, et al. 2012. Genomic landscape of non-small cell lung cancer in smokers and never-smokers. *Cell* **150**: 1121–1134.

Hofmann H-S, Bartling B, Simm A, Murray R, Aziz N, Hansen G, Silber R-E, Burdach S. 2006. Identification and classification of differentially expressed genes in non-small cell lung

- cancer by expression profiling on a global human 59,620-element oligonucleotide array. *Oncol Rep* **16**: 587–595.
- Holzinger A, Dingle S, Bejarano PA, Miller MA, Weaver TE, DiLauro R, Whitsett JA. 1996. Monoclonal antibody to thyroid transcription factor-1: Production, characterization, and usefulness in tumor diagnosis. *Hybridoma* **15**: 49–53.
- Homminga I, Pieters R, Langerak A, de Rooij J, Stubbs A, Verstegen M, Vuerhard M, Buijs-Gladdines J, Kooi C, Klous P, et al. 2011. Integrated transcript and genome analyses reveal NKX2-1 and MEF2C as potential oncogenes in T cell acute lymphoblastic leukemia. *Cancer Cell* **19**: 484–581.
- Hsu DS, Acharya CR, Balakumaran BS, Riedel RF, Kim MK, Stevenson M, Tuchman S, Mukherjee S, Barry W, Dressman HK, et al. 2009. Characterizing the developmental pathways TTF-1, NKX2-8, and PAX9 in lung cancer. *Proc Natl Acad Sci* **106**: 5312–5317.
- Hua S, Kittler R, White KP. 2009. Genomic antagonism between retinoic acid and estrogen signaling in breast cancer. *Cell* **137**: 1259–1271.
- Imielinski M, Berger AH, Hammerman PS, Hernandez B, Pugh TJ, Hodis E, Cho J, Suh J, Capelletti M, Sivachenko A, et al. 2012. Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. *Cell* **150**: 1107–1120.
- Kendall J, Liu Q, Bakleh A, Krasnitz A, Nguyen K, Lakshmi B, Gerald W, Powers S, Mu D. 2007. Oncogenic cooperation and coamplification of developmental transcription factor genes in lung cancer. *Proc Natl Acad Sci* **104**: 16663–16671.
- Kim D-W, Kempf H, Chen RE, Lassar AB. 2003. Characterization of Nkx3.2 DNA binding specificity and its requirement for somitic chondrogenesis. *J Biol Chem* **278**: 27532–27539.
- Kolla V, Gonzales LW, Gonzales J, Wang P, Angampalli S, Feinstein SI, Ballard PL. 2007. Thyroid transcription factor in differentiating type II cells: Regulation, isoforms, and target genes. *Am J Respir Cell Mol Biol* **36**: 213–225.
- Kwei KA, Kim YH, Girard L, Kao J, Pacyna-Gengelbach M, Salari K, Lee J, Choi Y-L, Sato M, Wang P, et al. 2008. Genomic profiling identifies TITF1 as a lineage-specific oncogene amplified in lung cancer. *Oncogene* **27**: 3635–3640.
- Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**: R25.
- Levine AJ, Oren M. 2009. The first 30 years of p53: Growing ever more complex. *Nat Rev Cancer* **9**: 749–758.
- Lieber M, Smith B, Szakal A, Nelson-Rees W, Todaro G. 1976. A continuous tumor-cell line from a human lung carcinoma with properties of type II alveolar epithelial cells. *Int J Cancer* **17**: 62–70.
- Lundberg AS, Randell SH, Stewart SA, Elenbaas B, Hartwell KA, Brooks MW, Fleming MD, Olsen JC, Miller SW, Weinberg RA, et al. 2002. Immortalization and transformation of primary human airway epithelial cells by gene transfer. *Oncogene* **21**: 4577–4586.
- Luo B, Cheung HW, Subramanian A, Sharifnia T, Okamoto M, Yang X, Hinkle G, Boehm JS, Beroukhim R, Weir BA, et al. 2008. Highly parallel identification of essential genes in cancer cells. *Proc Natl Acad Sci* **105**: 20380–20385.
- Machanic P, Bailey TL. 2011. MEME-ChIP: Motif analysis of large DNA datasets. *Bioinformatics* **27**: 1696–1697.
- Maeda Y, Davé V, Whitsett JA. 2007. Transcriptional control of lung morphogenesis. *Physiol Rev* **87**: 219–244.
- Maeda Y, Tsuchiya T, Hao H, Tompkins DH, Xu Y, Mucenski ML, Du L, Keiser AR, Fukazawa T, Naomoto Y, et al. 2012. KrasG12D and Nkx2-1 haploinsufficiency induce mucinous adenocarcinoma of the lung. *J Clin Invest* **122**: 4388–4400.
- McCormack MP, Young LF, Vasudevan S, de Graaf CA, Codrington R, Rabbitts TH, Jane SM, Curtis DJ. 2010. The *Lmo2* oncogene initiates leukemia in mice by inducing thymocyte self-renewal. *Science* **327**: 879–883.
- McGuire EA, Hockett RD, Pollock KM, Bartholdi MF, O'Brien SJ, Korsmeyer SJ. 1989. The t(11;14)(p15;q11) in a T-cell acute lymphoblastic leukemia cell line activates multiple transcripts, including Ttg-1, a gene encoding a potential zinc finger protein. *Mol Cell Biol* **9**: 2124–2132.
- Minoo P, Hu L, Xing Y, Zhu NL, Chen H, Li M, Borok Z, Li C. 2007. Physical and functional interactions between homeodomain NKX2.1 and winged helix/forkhead FOXA1 in lung epithelial cells. *Mol Cell Biol* **27**: 2155–2165.
- Miyoshi H, Shimizu K, Kozu T, Maseki N, Kaneko Y, Ohki M. 1991. t(8;21) breakpoints on chromosome 21 in acute myeloid leukemia are clustered within a limited region of a single gene, AML1. *Proc Natl Acad Sci* **88**: 10431–10434.
- Mullighan CG, Goorha S, Radtke I, Miller CB, Coustan-Smith E, Dalton JD, Girtman K, Mathew S, Ma J, Pounds SB, et al. 2007. Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. *Nature* **446**: 758–764.
- Peeters P, Raynaud SD, Cools J, Wlodarska I, Grosgeorge J, Philip P, Monpoux F, Van Rompaey L, Baens M, Van den Berghe H, et al. 1997. Fusion of TEL, the ETS-variant gene 6 (ETV6), to the receptor-associated kinase JAK2 as a result of t(9;12) in a lymphoid and t(9;15;12) in a myeloid leukemia. *Blood* **90**: 2535–2540.
- Ross-Innes CS, Stark R, Holmes KA, Schmidt D, Spyrou C, Russell R, Massie CE, Vowler SL, Eldridge M, Carroll JS. 2010. Cooperative interaction between retinoic acid receptor- α and estrogen receptor in breast cancer. *Genes Dev* **24**: 171–182.
- Shin H, Liu T, Manrai AK, Liu XS. 2009. CEAS: Cis-regulatory element annotation system. *Bioinformatics* **25**: 2605–2606.
- Shu W, Lu MM, Zhang Y, Tucker PW, Zhou D, Morrissy EE. 2007. Foxp2 and Foxp1 cooperatively regulate lung and esophagus development. *Development* **134**: 1991–2000.
- Simonis M, Klous P, Homminga I, Galjaard R-J, Rijkers E-J, Grosveld F, Meijerink JPP, de Laat W. 2009. High-resolution identification of balanced and complex chromosomal rearrangements by 4C technology. *Nat Methods* **6**: 837–842.
- Sos ML, Michel K, Zander T, Weiss J, Frommolt P, Peifer M, Li D, Ullrich R, Koker M, Fischer F, et al. 2009. Predicting drug susceptibility of non-small cell lung cancers based on genetic lesions. *J Clin Invest* **119**: 1727–1740.
- Stransky N, Egloff AM, Tward AD, Kostic AD, Cibulskis K, Sivachenko A, Kryukov GV, Lawrence MS, Sougnez C, McKenna A, et al. 2011. The mutational landscape of head and neck squamous cell carcinoma. *Science* **333**: 1157–1160.
- Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, et al. 2005. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci* **102**: 15545–15550.
- Tagne J-B, Gupta S, Gower AC, Shen SS, Varma S, Lakshminarayanan M, Cao Y, Spira A, Volkert TL, Ramirez MI. 2012. Genome-wide analyses of Nkx2-1 binding to transcriptional target genes uncover novel regulatory patterns conserved in lung development and tumors. *PLoS ONE* **7**: e29907.
- Tanaka H, Yanagisawa K, Shinjo K, Taguchi A, Maeno K, Tomida S, Shimada Y, Osada H, Kosaka T, Matsubara H, et al. 2007. Lineage-specific dependency of lung adenocarcinomas on the lung development regulator TTF-1. *Cancer Res* **67**: 6007–6011.
- Taub R, Kirsch I, Morton C, Lenoir G, Swan D, Tronick S, Aaronson S, Leder P. 1982. Translocation of the c-myc gene into the immunoglobulin heavy chain locus in human Burkitt lymphoma and murine plasmacytoma cells. *Proc Natl Acad Sci* **79**: 7837–7841.

Watanabe et al.

- Tomlins SA, Rhodes DR, Perner S, Dhanasekaran SM, Mehra R, Sun X-W, Varambally S, Cao X, Tchinda J, Kuefer R, et al. 2005. Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science* **310**: 644–648.
- Tusher VG, Tibshirani R, Chu G. 2001. Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci* **98**: 5116–5121.
- Visakorpi T, Hyytinen E, Koivisto P, Tanner M, Keinänen R, Palmberg C, Palotie A, Tammela T, Isola J, Kallioniemi OP. 1995. In vivo amplification of the androgen receptor gene and progression of human prostate cancer. *Nat Genet* **9**: 401–406.
- Wan H, Dingle S, Xu Y, Besnard V, Kaestner KH, Ang S-L, Wert S, Stahlman MT, Whitsett JA. 2005. Compensatory roles of Foxa1 and Foxa2 during lung morphogenesis. *J Biol Chem* **280**: 13809–13816.
- Wang L-H, Chmelik R, Nirenberg M. 2002. Sequence-specific DNA binding by the vnd/NK-2 homeodomain of *Drosophila*. *Proc Natl Acad Sci* **99**: 12721–12726.
- Wang L-H, Chmelik R, Tang D, Nirenberg M. 2005. Identification and analysis of vnd/NK-2 homeodomain binding sites in genomic DNA. *Proc Natl Acad Sci* **102**: 7097–7102.
- Weir BA, Woo MS, Getz G, Perner S, Ding L, Beroukhi R, Lin WM, Province MA, Kraja A, Johnson LA, et al. 2007. Characterizing the cancer genome in lung adenocarcinoma. *Nature* **450**: 893–898.
- Whittington T, Frith MC, Johnson J, Bailey TL. 2011. Inferring transcription factor complexes from ChIP-seq data. *Nucleic Acids Res* **39**: e98.
- Winslow MM, Dayton TL, Verhaak RGW, Kim-Kiselak C, Snyder EL, Feldser DM, Hubbard DD, DuPage MJ, Whittaker CA, Hoersch S, et al. 2011. Suppression of lung adenocarcinoma progression by Nkx2-1. *Nature* **473**: 101–104.
- Yamaguchi T, Yanagisawa K, Sugiyama R, Hosono Y, Shimada Y, Arima C, Kato S, Tomida S, Suzuki M, Osada H, et al. 2012. NKX2-1/TTF1/TTF-1-induced ROR1 is required to sustain EGFR survival signaling in lung adenocarcinoma. *Cancer Cell* **21**: 348–361.
- Yatabe Y, Mitsudomi T, Takahashi T. 2002. TTF-1 expression in pulmonary adenocarcinomas. *Am J Surg Pathol* **26**: 767–773.
- Yin Z, Gonzales L, Kolla V, Rath N, Zhang Y, Lu MM, Kimura S, Ballard PL, Beers MF, Epstein JA, et al. 2006. Hop functions downstream of Nkx2.1 and GATA6 to mediate HDAC-dependent negative regulation of pulmonary gene expression. *Am J Physiol Lung Cell Mol Physiol* **291**: L191–L199.
- Yokoyama A, Somervaille TCP, Smith KS, Rozenblatt-Rosen O, Watanabe H, Cleary ML. 2005. The menin tumor suppressor protein is an essential oncogenic cofactor for MLL-associated leukemogenesis. *Cell* **123**: 207–218.
- Yuan B, Li C, Kimura S, Engelhardt RT, Smith BR, Minoo P. 2000. Inhibition of distal lung morphogenesis in Nkx2.1^{-/-} embryos. *Dev Dyn* **217**: 180–190.
- Zhang P, Zuo H, Nakamura Y, Nakamura M, Wakasa T, Kakudo K. 2006. Immunohistochemical analysis of thyroid-specific transcription factors in thyroid tumors. *Pathol Int* **56**: 240–245.
- Zhang Y, Liu T, Meyer C, Eeckhoutte J. 2008. Model-based analysis of ChIP-seq (MACS). *Genome Biol* **9**: R137.