Novel Transcriptional Targets of the SRY-HMG Box Transcription Factor SOX4 Link Its Expression to the Development of Small Cell Lung Cancer

Sandra D. Castillo1, Ander Matheu2, Niccolo Mariani1, Julian Carretero3, Fernando Lopez-Rios4, Robin Lovell-Badge2, and Montse Sanchez-Cespedes1

Abstract

The HMG box transcription factor SOX4 involved in neuronal development is amplified and overexpressed in a subset of lung cancers, suggesting that it may be a driver oncogene. In this study, we sought to develop this hypothesis including by defining targets of SOX4 that may mediate its involvement in lung cancer. Ablating SOX4 expression in SOX4-amplified lung cancer cells revealed a gene expression signature that included genes involved in neuronal development such as PCDH8, MYB, RBPI, and TEAD2. Direct recruitment of SOX4 to gene promoters was associated with their upregulation upon ectopic overexpression of SOX4. We confirmed upregulation of the SOX4 expression signature in a panel of primary lung tumors, validating their specific response by a comparison using embryonic fibroblasts from Sox4-deficient mice. Interestingly, we found that small cell lung cancer (SCLC), a subtype of lung cancer with neuroendocrine characteristics, was generally characterized by high levels of SOX2, SOX4, and SOX11 along with the SOX4-specific gene expression signature identified. Taken together, our findings identify a functional role for SOX4 genes in SCLC, particularly for SOX4 and several novel targets defined in this study.

Cancer Res; 72(1); 176–86. ©2011 AACR.

Introduction

As with other types of cancer, lung cancer is undergoing a therapeutic revolution characterized by the identification of novel driver oncogenes and the generation of drugs that inhibit their activity in a very specific manner. The success of erlotinib/gefitinib in epidermal growth factor receptor (EGFR)-mutant tumors and crizotinib in tumors carrying a translocation of the ALK oncogene are some of the current paradigms (1). Because few tumors carry alterations of these genes, effort is required to identify additional targetable oncogenes. We previously carried out a wide-ranging DNA copy number analysis of lung cancer cell lines and identified an amplicon on chromosome 6p22, containing the SOX4 gene, which was expressed at very high levels and was the best candidate for an oncogene in the amplicon (2).

The SOX4 gene belongs to the SOX family, which is divided into 8 groups. A to H, according to protein identity (3). Sox4, Sox11, and Sox12 form the Sox-C group, sharing a high degree of identity in the high-mobility group domain, and in a group-specific transactivation domain (4). Sox4 is predominantly expressed during embryonic development in the heart, central nervous system, lung and thymus (5–7). SOX4 protein levels are increased in several types of carcinoma (8–11), and knocking down SOX4 induces apoptosis and growth suppression in cancer cells (12–14). Providing further evidence of the oncogenic potential of SOX4, we reported that its overexpression in NIH3T3 cells increases the number of foci induced by the mutant RHOA-Q63L (2). In addition, several independent studies have shown that Sox4 (also known as ecotropic viral integration site 16, Evi16) is a frequent target of retroviral insertional mutagenesis, leading to neoplastic transformation in murine hematopoietic cells (15–18). In the particular case of lung cancer, we previously reported the presence of a high level of SOX4 amplification in a subset of lung primary tumors and cancer cell lines (2). The relevance of SOX4 to lung cancer development has also been observed by others. A meta-analysis examining the transcriptional profiles of human tumors found SOX4 to be one of 64 genes uniquely upregulated in cancer thereby making it part of a general gene expression signature of cancer (19). Lung cancer was among the tumors with the greatest levels of SOX4 expression. In addition, Sox4 was among the set of genes overexpressed in the lungs of c-myc transgenic

Authors' Affiliations: 1Genes and Cancer Group, Cancer Epigenetics and Biology Program (PEBC), Bellvitge Biomedical Research Institute-IDIBELL, Barcelona, Spain; 2Division of Stem Cell Biology and Developmental Genetics, MRC National Institute for Medical Research, London, United Kingdom; 3Department of Physiology, Faculty of Medicine and Odontology, University of Valencia; and 4Hospital Universitario Madrid-Sanchinarro, Laboratorio Dianas Terapeuticas, Madrid, Spain

Note: Supplementary data for this article are available at Cancer Research Online (http://cancerres.aacrjournals.org/).

Array data deposited in the Gene Expression Omnibus (GEO) under accession number GSE31612.

Corresponding Author: Montse Sanchez-Cespedes, Cancer Epigenetics and Biology Program (PEBC), Bellvitge Biomedical Research Institute-IDIBELL 08908, Hospital de Llobregat, Barcelona 08907, Spain. Phone: 34-932-607132; Fax: 34-932-607219; E-mail: mcscespedes@idibell.cat

doi: 10.1158/0008-5472.CAN-11-3506

©2011 American Association for Cancer Research.
mice, and c-myc was also found to be one of the transcription factors with overrepresented Sox4-binding sites among the set of overexpressed genes (20).

SOX4 is involved in neural development and the maintenance of some stem cell types (21, 22). Recent studies have reported SOX4 to be a direct TGF-beta target that activates SOX2 transcription while retaining the stem cell properties of glioma-initiating cells (21). In addition, in normal hair follicles, Sox4 is expressed in the developing hair germ (23).

In spite of this, and although SOX4 was one of the first members of the SOX family to be isolated and characterized, including the demonstration that it has separable DNA-binding and transactivation domains, our present knowledge of the genes controlled by SOX4 activity is scarce. This paper reports on the role of SOX4 in human lung carcinogenesis, focusing especially on identifying novel targets of SOX4 transcriptional activity.

Materials and Methods

Cancer cell lines and primary tumors

Cancer cell lines were obtained from the American Type Culture Collection and grown under recommended conditions. DNA and RNA from additional lung cancer cell lines used for real-time quantitative reverse transcriptase (RT) PCR were kindly provided by Luis M. Montuenga and Ruben Pio of the Centro de Investigación Medica Aplicada (CIMA), University of Navarra, Spain, and Jun Yokota, National Cancer Center Research Institute, Tokyo, Japan. Fresh frozen lung primary tumors were provided by the CNIO Tumour Bank Network, CNIO, Spain, and were selected as previously described (24).

Expression plasmids and reporters

We stably transduced the H522 cell line with a Tet repressor (TetR) expression construct, pCMB1b-Tr5, kindly provided by M.V. de Wetering (Hubrecht Institute, Utrecht, the Netherlands) using hygromycin selection. We examined TetR activity by transfecting the cells with the pcDNA4/TO/Luc construct that expressed the firefly luciferase gene under the control of a Tet operator and then measured the luciferase activity with the Dual Luciferase Assay Kit (Promega). The pRL-Tk plasmid (Promega) that constitutively codified for the Renilla luciferase was used to measure the transfection efficiency. Clones that stably expressed the TetR were transfected with pSUPERIOR-SOX4/S1 construct, kindly provided by C.A. Moskaluk (University of Virginia, Charlottesville, VA), which expressed a short hairpin against the sequence 5'-AAGACGACCTGCTC-GACCTGA-3' of the SOX4 coding region under the control of a Tet operator and selected using geneticin. These oligonucleotides specifically inhibited the expression of SOX4 but not of other SOX family genes (13). Selected clones were analyzed by SOX4 expression by quantitative RT-PCR and Western blot analysis before and after doxycycline induction. For transient transfection, full-length human wt SOX4 and the mutant SOX4 S395X carrying an HA tag were cloned as previously described (2).

Gene expression microarrays and real-time quantitative PCRs

The mRNA was extracted using conventional methods, and 1 µg of it was amplified from each sample and used for gene expression microarray analysis. Universal Human Reference RNA (P/N 74000: Stratagene), was used as reference for hybridization and analysis. For labeling we used the commercial Two-Color Microarray-Based Gene Expression Analysis Kit (version 5.5). MMLV-RT retrotranscription of sample from a T7 promoter primer was followed by a T7 RNA polymerase-catalyzed in vitro transcription reaction in the presence of either Cy3-CTP or Cy5-CTP fluorophores. Cy3 labeling was used for the reference sample. Hybridization was carried out on the Whole Human Genome Microarray (4 × 44 K), scanned with a G2505B DNA microarray scanner and quantified using Agilent Feature Extraction Software (version 9.5: Agilent). Fluorescence intensity from each array element was subtracted from the local background and data normalized as previously described (24). We further selected transcripts that fulfilled the following criteria: (i) repressed at least 2 times in the H522Tr-shSOX4-1 at 48 and 96 hours after induction of shSOX4 relative to the level at 0 hour and (ii) no changes in gene expression between the parental H522 and the H522Tr-shSOX4-1. The mRNA and genomic DNA were measured by real-time quantitative PCR. DNase-treated RNA was reverse-transcribed. The cDNA and genomic DNA were amplified using SYBR green real-time PCR. Reactions were carried out in triplicate. As controls we used the human GAPDH, B-Actin, and TATA box–binding protein (TBP) to correct for inter-individual/tumor variations. The primer sequences used are included in Supplementary Table S2.

Antibodies, Western blot analysis, and immunostaining

For Western blot analysis, cells were scraped from the dishes into the lysis buffer. A total of 25 µg of total protein was separated by SDS-PAGE and blotted with rabbit anti-SOX4 (A574) 1:5,000 (CS-129-100, Diagenode), mouse anti-SOX2 1:2,500 (R&D Systems); rabbit anti-NSE 1:1,000 (Abcam), or mouse anti-GAPDH. The secondary goat anti-mouse-IgG:HRP (horseradish peroxidase) and goat anti-rabbit:HRP antibodies (DAKO) were added at 1:5,000. We carried out immunohistochemical analysis of SOX9 (1:750 dilution; Chemicon) in the Autostainer Low FLEX (DAKO), using previously described protocols (25). Sections were counterstained with hematoxylin and evaluated by 2 independent researchers.

Chromatin immunoprecipitation assays

Cells were fixed in 1% formaldehyde for 10 minutes at 37°C. Cross-linking was quenched by adding 125 mmol/L glycine. Cells were then washed with cold PBS, harvested and resuspended in SDS lysis buffer containing a protease inhibitor cocktail. Chromatin was sheared by sonication (average length 0.25–1 Kb) and incubated with 60 µl protein A/G agarose/salmon sperm DNA (50% slurry; Millipore) with gentle agitation for 30 minutes. The supernatant was then immunoprecipitated with anti-SOX4 antibody 1:500 or its matched nonimmune crude serum 1:500 (IgG: Diagenode) at 4°C overnight.
Protein A/G agarose (60 μL of a 50% slurry) was then added and incubated for 1 hour. Pellets were washed and protein-DNA cross-links were reversed by overnight incubation at 65°C with proteinase K. DNA was purified following a conventional phenol–chloroform protocol and eluted in 50 μL water. At least 2 independent Chromatin immunoprecipitation (ChIP) experiments were carried out. Real-time quantitative PCR was carried out using SYBR Green Master Mix on an ABI Prism 7900 (Applied Biosystems). The primer sequences used are included in Supplementary Table S2.

Mouse embryonic fibroblast processing and RNA extraction from mouse tissues

Mice were housed in the pathogen-free barrier area of the National Institute for Medical Research, London, UK. Tissues from young (1–2 months) and old (1.5–2 years) C57BL6 mice were homogenized at high speed and total RNA extracted with TRIzol. Mouse embryo fibroblasts (MEF) were isolated and cultured as previously described (26) from Sox4+/+, Sox4+/−, and Sox4−/− E13.5 mouse embryos. Mice were genotyped with primers included in Supplementary Table S2.

Statistical and bioinformatic analysis

Statistical analyses were carried out with SPSS software. Differences between groups were analyzed by the unpaired t test or Fishers exact test, as appropriate. Differences in gene expression between lung tumors with and without high levels of SOX4 were determined by the Mann–Whitney U test. Correlations between SOX genes were estimated by Spearman rank correlation. Values of P < 0.05 were considered statistically significant. For the Gene set enrichment analysis (GSEA), raw data from the Gene Expression Omnibus (GEO) database repository (27) were normalized using the Robust Multiarray Average (RMA) algorithm available in Bioconductor’s Affy package. We used the LIMMA package to obtain LIMMA-moderated t statistics to build a ranked list of expression. GSEA was applied to this ranked list. After testing for normality of the data with the Kolmogorov–Smirnoff test, our gene sets were considered significantly enriched between classes under comparison for values of FDR less than 0.25, a well-established cutoff for the identification of biologically relevant gene sets (28).

Results

Differential gene expression upon SOX4 depletion in lung cancer cells

The NCI-H522 lung cancer cell line (henceforth referred to as H522) is known to feature SOX4 gene amplification (>20 copies/nuclei) and SOX4 overexpression (2). To identify transcriptional targets regulated by SOX4 we generated H522-derived cells that downregulate SOX4 in an inducible manner.

![Figure 1. Changes in gene expression after depletion of SOX4 in the H522Tr-shSOX4-1 cells. A, Western blot analysis of SOX4 of the H522 parental, H522Tr-shSOX4-1–derived cells with (+) or without (−) doxycycline (dox; 2.5 ng/mL) after 48 hours. B, representative examples of real-time quantitative RT-PCR of the indicated transcripts relative to GAPDH. The values were also relative to the uninduced control for the H522Tr-shSOX4-1 (dox−) and to the parental H1299 (pCEFL-mock) for the H1299 pCEFL-SOX4-wt and H1299 pCEFL-SOX4-S395X. Bars, mean ± SD from triplicates, from at least 2 independent experiments. C, Western blot analysis of SOX4 and SOX2 of the H1299, transfected with pCEFL-mock, pCEFL-SOX4, and pCEFL-SOX4-S395X. GAPDH is also shown for protein loading comparison. D, Western blot analysis of SOX4 and SOX2 of the H522Tr-shSOX4-1. GAPDH is also shown for protein loading comparison. GAPDH, glyceraldehyde-3-phosphate dehydrogenase.](cancerres.aacrjournals.org)
with a shSOX4 that was reported to specifically deplete SOX4 expression (13). Several clones were found to downregulate SOX4 in an inducible-dependent manner. A stable clone, H522Tr-shSOX4-1, which downregulates SOX4 expression by 90%, 48 hours after the addition of doxycycline, was chosen for further analysis (Fig. 1A).

To determine the gene expression profile characteristic of SOX4-depleted expression we compared the global gene expression of the parental H522 cells with that of H522Tr-shSOX4-1 cells 0, 48, and 96 hours after inducing shSOX4 with doxycycline. We found 123 genes with a more than 2-fold change of expression (85 downregulated and 38 upregulated) in the H522Tr-shSOX4-1 cells after shSOX4-inducible expression relative to the H522 parental and to the H522Tr-shSOX4-1 cells before adding doxycycline (at 0 hour; Supplementary Table S1). We then measured the expression of these transcripts by quantitative RT-PCR in the H522Tr-shSOX4-1 cells to validate the microarray observations. Half of the 50 downregulated genes chosen for validation were confirmed and selected for further analysis. The TEAD2 and Tubb3 genes were previously identified as SOX4 transcriptional targets in neural progenitors (29–32). All the transcripts, including the TEAD2 and Tubb3 controls, showed an approximately 50% drop in expression after interfering SOX4 expression (Fig. 1B), validating the robustness of our microarray assay. It is of note that many of the transcripts found to be downregulated after inducing shSOX4 (e.g., EGR3, MLT11, NBEA, PCDHBs, RBP1, TMEFF2, TUBB3, and VASH2) are highly expressed in neural-derived tissues, including brain, spinal cord, and retina (GSE7905 from GEO database).

Next, we aimed to determine whether these observations could be extended to other lung cancer cell lines. To this end, we transfected the H1299 cells, which have low levels of endogenous SOX4, with a construct carrying wild-type SOX4 or a mutant form (S395X). The latter is devoid of transactivation activity because of a mutation that eliminates the serine-rich C-terminal domain of SOX4 (Fig. 1C; ref. 2). As shown in Fig. 1B, ectopic expression of the wild type, but not the mutant SOX4 form, leads to a significant, more than 2-fold upregulation of most targets. In particular, EGR3 and RBP-1 were strongly upregulated (>16 times) relative to the control and to the SOX4-mutant.

It has previously been described that SOX4 regulates the expression of SOX2 in glioma-initiating cells (21). Although SOX2 was not downregulated in our microarray analysis after depletion of SOX4, we specifically tested for SOX2 levels in a Western blot analysis. No change in protein expression levels of SOX2 was observed after ectopic expression of SOX4 in the H1299 cells or after abrogation of SOX4 in the H522Tr-shSOX4-1 cells (Fig. 1C and D).

Analysis of SOX4 recruitment to the gene promoters

We developed a quantitative ChIP assay to investigate which of the genes downregulated after SOX4 depletion were direct transcriptional targets of SOX4. The SOX4 HMG-box binds preferentially to the 4-bp DNA-motif 5′-ACAA-3′ sequence (33) and we searched for putative SOX4-binding sequences in the promoters of these genes. We used the MatInspector program (Genomatix; ref. 34) and designed primers flanking the regions enriched in these sequences (Supplementary Fig. S1). To determine the specificity of SOX4 occupancy in the selected regions, we also screened distant regions (>5,000 bp from the SOX4-binding sequences) that were considered negative controls. We immunoprecipitated the H522Tr-shSOX4-1 chromatin with the rabbit polyclonal anti-SOX4 antibody and its matched crude serum as IgG control (Fig. 2A) followed by quantitative PCR. We confirmed the recruitment of SOX4 to the known targets TEAD2 and TUBB3 (Fig. 2B; refs. 29, 32). We also determined that the MYB and VASH2 promoters were significantly enriched (>4-fold) in the anti-SOX4 immunoprecipitated chromatin from the cells prior to doxycycline induction, compared with cells after depletion of SOX4 expression and IgG immunoprecipitates (Fig. 2C). In all cases tested, the enrichment was specific to the indicated SOX4-binding regions, and was negative in the control regions. This indicates that SOX4 binds to the MYB and VASH2 promoters. The protocadherin genes, PCDHA, PCDHB, and PCDHG, are tandemly arranged in clusters on the same chromosome. To determine the possible direct involvement of SOX4 in the transactivation of the different PCDH genes we first screened for binding of SOX4 in regions upstream of individual genes but found no SOX4 recruitment. While conducting the screening, the PCDHB cluster was found to be transcriptionally regulated by a control region downstream of the PCDHB cluster (35). Therefore, we searched and found SOX4-binding sequences within this control region. Next, we tested and confirmed SOX4 recruitment to this regulatory region (Fig. 2C), although it remains to be understood how the binding of SOX4 to this region affects the transcriptional activation of the PCDHB cluster. Thus, we found evidence that the PCDHB clusters, as well as the MYB and VASH2 genes, are direct SOX4 transcriptional targets. Other genes that are strongly upregulated by SOX4 were not found to recruit SOX4 to their promoters (EGR3, GPC2, and RBP1), at least in the regions that we selected for the analysis.

Expression of SOX4 targets is increased in lung tumors with SOX4 gene overexpression and is decreased in Sox4−/− MEF

We extended our analysis to lung primary tumors selected on the basis of high (SOX4-H) and low (SOX4-L) levels of SOX4 gene expression. We compared the differences in the expression of the transcripts between the 2 groups. As depicted in Fig. 3A, most of the transcripts, especially CCNG2, DLG3, VASH2, PCDH9, and -11, were expressed at significantly higher levels in the SOX4-H group than in the SOX4-L group. We also used GSEA to compare our SOX4-specific gene expression signature, composed of the genes most significantly downregulated in H522Tr-shSOX4-1 with the data set containing the gene expression profile of the SOX4-H and SOX4-L lung primary tumors (GEO Series accession number GSE35869; ref. 36). The GSEA showed that the set of downregulated genes after inducing the expression of the shSOX4 were upregulated in the SOX4-H tumors (Fig. 3B). Collectively, these observations are consistent with these being transcriptional targets of SOX4 and having a role in lung carcinogenesis.
Sox4 is key to embryogenesis and, while Sox4+/− embryos are viable, Sox4−/− embryos die by day 14 of embryonic development (37). Taking advantage of the availability of MEFs from Sox4+/−/− and Sox4+/+ embryos, we characterized the expression levels of the newly identified SOX4 targets in this material (Fig. 3C). Compared with mouse lung tissues, some of the targets were expressed at very low levels in the Sox4+/+ MEFs (i.e., Dlg3, Dock4, Egr3, Mta2, Myb, Rbp1, and most Pcdhb transcripts) and were then excluded from the analysis. Our results showed that the Gpc2, Rnf122, Tead2, Tubb3, and Vash2 transcripts exhibited significant downregulation in the Sox4−/− MEFs, relative to their wild-type counterparts (Fig. 3D).

**Ageing is associated with decreased expression of Sox4 and its targets in lung tissue**

Ageing tissues experience changes in the expression levels of some tumor suppressors and oncogenes that have been associated with a diminished potential for self-renewal of multipotent progenitors. In particular, an increase in the levels of p16/INK4a has been reported in several tissues including the neural system and in beta cells from the pancreas of ageing mice (38, 39). This background, coupled with the reported connection of Sox4 with the maintenance of specific stem cell properties (23), prompted us to test whether the levels of Sox4 in the lung tissue vary as mice age. We used real-time quantitative RT-PCR to compare the gene expression levels of Sox4 in different tissues (lung, spleen, liver, kidney, and heart) of young (1–2 month) and old (1.5–2 years) C57BL6 mice, from 4 and 5 different individuals, respectively. As a positive control we included p16/INK4a. We found that Sox4 levels were significantly reduced (>70%) in the lungs of old mice compared with young individuals (Fig. 4A). We also tested the gene expression levels in many of the newly identified targets of Sox4. Again we discarded some of the transcripts because they were expressed at very low levels in normal lung (e.g., Pcdhb2, Pcdhb5, Pcdhb7, Tubb3, and Vash2). Among those depicting measurable levels of gene expression, 10 (Ccng2, Dlg3, Dock4, Gcp2, Mta2, Nbea, Rbp1, Rnf122, Tead2, and Tmeff1) exhibited significantly higher levels in the lungs of young mice than in their older counterparts (Fig. 4B).

**Analysis and characterization of gene expression levels among the SOX family of genes in human lung cancer**

We previously reported that SOX4 expression levels were significantly higher in small cell lung cancer (SCLC) than in non–small cell lung cancer (NSCLC) histopathology, indicating differences in their underlying biology (2). A functional, but nonoverlapping relationship has been reported between SOX4...
and other members of the SOX family of transcription factors (6, 29, 31, 32). Thus, we searched for a possible preferential expression of different SOX-family members (i.e., SOX2, SOX4, SOX9, SOX11, and SOX12) in the distinct and most common lung cancer histopathologies. First, we tested a panel of lung cancer cell lines, using quantitative RT-PCR, and found that the expression of SOX2, SOX4, and SOX11 transcripts was significantly higher in SCLC than in NSCLC (Fig. 5A–B and Supplementary Fig. S2). This occurred in the absence of gene amplification (data not shown). The differences were particularly striking in the case of SOX11, in which 9 of 18 SCLCs had very high levels of SOX11 (p < 0.005; Fig. 5A). For unknown reasons, commercially available lung cancer cell lines are under-represented in the squamous cell carcinoma (SCC) type, which affects the comparative assessment of differences between the 2 main NSCLC histopathologies: adenocarcinomas (AC) and SCCs. To overcome this hurdle we also tested in primary lung tumors: 40 SCCs and 30 ACs. We found that levels of SOX2 and SOX9 were significantly higher in SCCs, whereas high levels of SOX4 predominated in the AC type (Supplementary Fig. S3A). It is important to note that SOX2 is located on chromosome 3q, a region that is frequently amplified in SCC and also includes the PIK3CA oncogene (24, 40). In our samples, overexpression of SOX2 was concordant with the presence of gene amplification, as we previously observed (24). In the case of SOX9, taking advantage of available antibodies, we carried out immunostaining and confirmed the high protein levels of SOX9 and its association with the SCC histopathology (Fig. 5C). Although more commonly overexpressed in SCC, strong SOX9 immunostaining was also observed in some ACs. No differences in the levels of SOX12 were observed among the different types of lung cancer.
Finally, we determined the correlations between the levels of the different SOX family genes. In lung cancer cell lines, there were significant direct correlations between all pairs of SOX transcripts, except SOX9 (Supplementary Fig. S3B). The most significant direct correlation involved SOX4 and SOX2 expression levels, which occurred in both NSCLC and SCLC cell lines and suggests a functional correlation. The exception was in the lung primaries of the SCC type, which probably reflects the fact that SOX2 overexpression in this type of lung cancer is mainly because of gene amplification.

The gene expression signature of SOX4 relates to that of SCLC

Taking into account the following observations: (i) the high level of expression of several SOX-family genes in the SCLC type; (ii) the neural-derived origin of SCLC, and (iii) the neural-tissue specificity of many of the SOX4 targets (Supplementary Table S1), we decided to examine whether there was a relationship between the gene expression signature of SOX4 and that of SCLC. First, we compared the expression levels of the various SOX4 targets identified here with a data set containing the gene expression profile of a panel of 68 cell lines (48 and 20 of the NSCLC and SCLC types, respectively) extracted from the GEO database (GSE4824). As depicted in Fig. 6A, most of the SOX4 targets exhibited higher expression levels in the SCLC than in the NSCLC type. Two of the lung AC cell lines, H2009 and H2887, had a profile for the SOX4 targets with similarities to that of SCLC. One of these cell lines, H2009, had simultaneous mutations at KRAS and RB (http://www.sanger.ac.uk/). While the former mutation is characteristic of lung ACs, the latter is almost exclusively found in SCLCs. This is puzzling as it suggests that this cell line has a neuroendocrine origin.

Next, we used the same data set to obtain the gene expression profile of the SCLC cell lines (Fig. 6B) and to conduct GSEA. Our results show significant similarities between the genes downregulated after inducing the expression of shSOX4 in the H522Tr-shSOX4-1 and the SCLC gene expression signature (Fig. 6B). Taken together these observations suggest that SOX4 is a key element required for the development of SCLC.

Discussion

We report the changes in global gene expression after depletion of SOX4 in a lung cancer cell line that has SOX4 overexpression because of gene amplification. Taking this approach, we identified novel and high confidence direct and indirect targets of SOX4 activity. We validated our findings in various models, attesting to the robustness of our approach. ChIP of SOX4 indicated that some of these are direct targets, activated by the position of SOX4 in their promoters, whereas others are indirectly activated by SOX4. Apart from TEAD2 and TUBB3, previously described as SOX4 direct targets in neural progenitors (29–32), we identified novel genes that recruit SOX4 to their promoters, including the MYB and VASH2 genes. Some human protocadherin (PCDH) genes are clustered together on chromosome 5 and classified into 3 subfamilies, PCDHA, PCDHB, and PCDHG. Of special interest are the PCDHB genes, many of which are downregulated after SOX4 depletion. The PCDH proteins are located, at least in part, at synapses and suggest a functional correlation. The exception was in the lung primaries of the SCC type, which probably reflects the fact that SOX2 overexpression in this type of lung cancer is mainly because of gene amplification.

The gene expression signature of SOX4 relates to that of SCLC

Taking into account the following observations: (i) the high level of expression of several SOX-family genes in the SCLC type; (ii) the neural-derived origin of SCLC, and (iii) the neural-tissue specificity of many of the SOX4 targets (Supplementary Table S1), we decided to examine whether there was a relationship between the gene expression signature of SOX4 and that of SCLC. First, we compared the expression levels of the various SOX4 targets identified here with a data set containing the gene expression profile of a panel of 68 cell lines (48 and 20 of the NSCLC and SCLC types, respectively) extracted from the GEO database (GSE4824). As depicted in Fig. 6A, most of the SOX4 targets exhibited higher expression levels in the SCLC than in the NSCLC type. Two of the lung AC cell lines, H2009 and H2887, had a profile for the SOX4 targets with similarities to that of SCLC. One of these cell lines, H2009, had simultaneous mutations at KRAS and RB (http://www.sanger.ac.uk/). While the former mutation is characteristic of lung ACs, the latter is almost exclusively found in SCLCs. This is puzzling as it suggests that this cell line has a neuroendocrine origin.

Next, we used the same data set to obtain the gene expression profile of the SCLC cell lines (Fig. 6B) and to conduct GSEA. Our results show significant similarities between the genes downregulated after inducing the expression of shSOX4 in the H522Tr-shSOX4-1 and the SCLC gene expression signature (Fig. 6B). Taken together these observations suggest that SOX4 is a key element required for the development of SCLC.

Discussion

We report the changes in global gene expression after depletion of SOX4 in a lung cancer cell line that has SOX4 overexpression because of gene amplification. Taking this approach, we identified novel and high confidence direct and indirect targets of SOX4 activity. We validated our findings in various models, attesting to the robustness of our approach. ChIP of SOX4 indicated that some of these are direct targets, activated by the position of SOX4 in their promoters, whereas others are indirectly activated by SOX4. Apart from TEAD2 and TUBB3, previously described as SOX4 direct targets in neural progenitors (29–32), we identified novel genes that recruit SOX4 to their promoters, including the MYB and VASH2 genes. Some human protocadherin (PCDH) genes are clustered together on chromosome 5 and classified into 3 subfamilies, PCDHA, PCDHB, and PCDHG. Of special interest are the PCDHB genes, many of which are downregulated after SOX4 depletion. The PCDH proteins are located, at least in part, at synapses and suggest a functional correlation. The exception was in the lung primaries of the SCC type, which probably reflects the fact that SOX2 overexpression in this type of lung cancer is mainly because of gene amplification.
neural formation. Each clustered PCDH isof orm has its own promoter but it has recently been reported that genes of the Pcdh-β family are controlled by a region located downstream of the Pcdh-γ cluster. This region activates the expression of the Pcdh-β gene cluster in cis, and its deletion dramatically decreases their expression levels (35). We found that this control region, containing many SOX4-binding sites, strongly recruits SOX4, implying that it is an important regulator of the PCDHB gene family. Although the exact mechanisms that underlie this regulation remain to be elucidated, the formation of a looped architecture is a possible explanation that has been reported for other transcription factors. Apart from the of a looped architecture is a possible explanation that has been reported for other transcription factors.

Figure 5. Characterization of gene expression in SOX family members among the different lung cancer histopathologies. A, mRNA levels of SOX11 and SOX2 genes assessed by real-time quantitative RT-PCR, relative to the internal control GAPDH, in lung cancer cell lines of different histopathologies and in normal lung. LCC, large cell carcinoma; AL, normal lung. Bars, means ± SD from triplicates. B, Western blot analysis of SOX4, SOX2, and from the neural marker NSE (neuron-specific enolase) of the indicated lung cancer cell lines. GAPDH is also shown for protein loading comparison. The histopathologic subtypes are also indicated. GAPDH, glyceraldehyde-3-phosphate dehydrogenase. C, representative examples of SOX9-positive immunostaining (original magnification, 200×), in 2 different lung primary tumors. Cells from normal tissue that do not express detectable SOX9 protein are included in the images and serve as negative controls. (P = 0.01; Fisher exact test).

Although it was reported that SOX2 constitutes a direct target of SOX4 in glioma cells (21), we could not confirm this in lung cancer-derived cells.

 Sox4+/− mice are viable but Sox4 homozygous null mutants are embryonic lethal because of cardiac defects and other abnormalities (37). This indicates an involvement of Sox4 in embryonic development, which is consistent with the observed fluctuations of Sox4 levels during embryogenesis; these are significantly reduced in later stages (42). In addition to embryogenesis, we have shown that levels of Sox4 and of most Sox4 targets are reduced in the lungs of older mice compared with those of younger ones, thereby implying a role for Sox4 in lung ageing. In parallel, the levels of the p16Ink4a tumor suppressor were significantly increased in aged lungs. This agrees with the previously reported increase in the expression of p16Ink4a in the neural system and in beta cells from the pancreas of ageing mice (38, 39) and suggests that genes involved in cancer development also have a role during ageing, possibly related to the diminished self-renewal potential of multipotent progenitors. In this regard, a crucial role for Sox4 in cell fate decisions and maintenance of stem cell properties has been described (21, 23).

A variety of Sox family genes are expressed in the developing lung, including Sox2, Sox4, Sox9, and Sox11, and become repressed as embryonic development progresses (32, 42, 43), indicating that this family of genes actively control lung embryogenesis. On the other hand, lung cancer comprises several histopathologic lineages that arise from various genetically distinct cell types (45). In this study, we found strong differences in the pattern of expression of Sox-related genes (SOX2, SOX4, SOX9, and SOX11) among the lung cancer types. In NSCLC, increased levels of SOX4.
preferentially occurred in ACs whereas SOX2 and SOX9 were more commonly upregulated in the SCC type. It is important to note that SOX2 is located on chromosome 3q26 and is amplified and overexpressed in lung SCCs (24, 40). Moreover, shRNA targeting SOX2 decreased colony formation in SOX2-amplified cells lines (40), supporting the causative effect of SOX2 amplification in lung cancer development. However, because the 3q26 amplicon includes another bona fide oncogene, PIK3CA (24), it is not possible to determine which is the target for gene amplification in each case. Interestingly, SOX2, SOX4, and SOX11 exhibited very strong expression levels in most SCLCs, implying specific roles for the SOX family in this lung cancer type. This occurs in the absence of gene amplification and indicates that there are alternative mechanisms for SOX-gene upregulation in this lung tumor type. Sox4 and Sox2 are known to be TGF-β target genes (21) and it cannot be ruled out that the high levels of SOX4 and SOX2 in SCLC in the absence of an associated gene alteration may be because of an active TGF-β pathway. While the high expression levels of SOX4 in SCLC had been described by us and others (2, 44), the strong and unique association between high levels of SOX11 and the SCLC type is a new finding. Here, we also report that the SOX4-associated expression profile has significant similarities with the SCLC
Novel SOX4 Targets and Small Cell Lung Cancer Development

gene expression signature. Collectively, these observations strongly implicate SOX4 and other SOX family genes in the development of the SCLC type. The SOX4 and SOX11 proteins are members of the Sox-C group, are molecularly very similar (4) and both are critical to the induction of the expression of neuronal traits and the maintenance of neural stem cells (21, 29). Taken into account this background and the fact that many of the SOX4 targets are related to neural tissues, the high levels of SOX4 and SOX11 in SCLC may also indicate specific cells of origin. In fact, it has been suspected for a long time that the SCLC type arises from neuroendocrine cells, commonly found in clusters known as neuroendocrine bodies. Recent observations, using mouse models for targeted Trp53 and Rb1 inactivation in distinct cell types of the adult lung, support the neuroendocrine origin of SCLC (45). It is likely that along with their role in the stemness of the neuroendocrine cells of the lung, SOX4 and possibly SOX11, are required for the development of SCLC.

In conclusion, through the abrogation of SOX4 expression in lung cancer cells carrying SOX4 amplification and overexpression, we have identified novel transcriptional targets for SOX4. We show that these are upregulated in lung tumors with SOX4 overexpression, indicating involvement in lung cancer development. We have also found that SOX4 and other SOX family genes are strongly upregulated in the SCLC type. Given the neural-related nature of many of the SOX4 targets identified here and the similarity between the gene expression signatures of SCLC and ACC, we propose that SOX4 is involved in the development of lung tumors with neuroendocrine characteristics, especially of the SCLC type.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Authors’ Contributions

S.J. Castillo, A. Matheu, R. Lovell-Badge, and M. Sanchez-Cespedes conceived and designed the experiments. S.J. Castillo, A. Matheu, and N. Mariani carried out the experiments. F. Lopez-Bios contributed to the histopathologic and immunohistochemical analysis and evaluation of the data. J. Carretero did the bioinformatic analysis. S.J. Castillo and M. Sanchez-Cespedes analyzed the data and wrote the manuscript.

Acknowledgments

The authors acknowledge the technical assistance of Albert Coll (PEBC) and of Orlando Dominguez at the Genomics Unit (CNIO).

Grant Support

S.J. Castillo was supported for fellowships from the CAM, AGAUR, and EMBO short-term fellowship–ASTF–2009 and A. Matheu by a postdoctoral fellowship from the HFSP. Work of R. Lovell-Badge is funded by UK-MRC (U117512772) and of M. Sanchez-Cespedes by MICINN SAF2008-02691, RTIIC (RD06/0026/0062), and EU–Framework Programme (HEALTH-F2-2010-258677).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received October 25, 2011; revised October 31, 2011; accepted November 2, 2011; published OnlineFirst November 14, 2011.

References


Downloaded from cancerres.aacrjournals.org on April 15, 2017. © 2012 American Association for Cancer Research.
Novel Transcriptional Targets of the SRY-HMG Box Transcription Factor SOX4 Link Its Expression to the Development of Small Cell Lung Cancer

Sandra D. Castillo, Ander Matheu, Niccolo Mariani, et al.


Updated version
Access the most recent version of this article at:
doi:10.1158/0008-5472.CAN-11-3506

Supplementary Material
Access the most recent supplemental material at:
http://cancerres.aacrjournals.org/content/suppl/2011/11/14/0008-5472.CAN-11-3506.DC1

Cited articles
This article cites 45 articles, 19 of which you can access for free at:
http://cancerres.aacrjournals.org/content/72/1/176.full.html#ref-list-1

Citing articles
This article has been cited by 6 HighWire-hosted articles. Access the articles at:
/content/72/1/176.full.html#related-urls

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.