Genome-Wide Promoter Analysis of the SOX4 Transcriptional Network in Prostate Cancer Cells

Christopher D. Scharer,1,2 Colleen D. McCabe,2 Mohamed Ali-Seyed,2 Michael F. Berger,4,7 Martha L. Bulyk,4,6,7 and Carlos S. Moreno3,3

1Program in Genetics and Molecular Biology, 2Department of Pathology and Laboratory Medicine, and 3Winship Cancer Institute, Emory University School of Medicine, Atlanta, Georgia; 4Division of Genetics, Department of Medicine and 5Department of Pathology, Brigham and Women’s Hospital and Harvard Medical School; 6Harvard/MIT Division of Health Sciences and Technology, Harvard Medical School, Boston, Massachusetts; and 7Committee on Higher Degrees in Biophysics, Harvard University, Cambridge, Massachusetts

Abstract

SOX4 is a critical developmental transcription factor in vertebrates and is required for precise differentiation and proliferation in multiple tissues. In addition, SOX4 is overexpressed in many human malignancies, but the exact role of SOX4 in cancer progression is not well understood. Here, we have identified the direct transcriptional targets of SOX4 using a combination of genome-wide localization chromatin immunoprecipitation–chip analysis and transient overexpression followed by expression profiling in a prostate cancer model cell line. We have also used protein-binding microarrays to derive a novel SOX4-specific position-weight matrix and determined that SOX4 binding sites are enriched in SOX4-bound promoter regions. Direct transcriptional targets of SOX4 include several key cellular regulators, such as EGFR, HSP70, Tenascin C, Frizzled-5, Patched-1, and Delta-like 1. We also show that SOX4 targets 23 transcription factors, such as MLL, FOXA1, ZNF281, and NKK3-1. In addition, SOX4 directly regulates expression of three components of the RNA-induced silencing complex, namely Dicer, Argonaute 1, and RNA Helicase A. These data provide new insights into how SOX4 affects developmental signaling pathways and how these changes may influence cancer progression via regulation of gene networks involved in microRNA processing, transcriptional regulation, the TGFβ, Wnt, Hedgehog, and Notch pathways, growth factor signaling, and tumor metastasis. [Cancer Res 2009;69(2):709–17]

Introduction

The sex determining region Y-box 4 (SOX4) gene is a developmental transcription factor important for progenitor cell development and Wnt signaling (1, 2). SOX4 is a 47-kDa protein that is encoded by a single exon and contains a conserved high-mobility group DNA-binding domain (DBD) related to the TCF/LEF family of transcription factors that mediate transcriptional responses to Wnt signals. SOX4 directly interacts with β-catenin, but its precise role in the Wnt pathway is unknown (2). In adult mice, SOX4 is expressed in the gonads, thymus, T-lymphocyte and pro-B-lymphocyte lineages, and to a lesser extent in the lungs, lymph nodes, and heart (1). Embryonic knockout of SOX4 is lethal around day E14 due to cardiac failure, and these mice also showed impaired lymphocyte development (3). Tissue-specific knockout of SOX4 in the pancreas results in failure of normal development of pancreatic islets (4). SOX4 heterozygous mice have impaired bone development (5), whereas prolonged expression of SOX4 inhibits correct neuronal differentiation (6). These studies suggest a critical role for SOX4 in cell fate decisions and differentiation.

Whereas SOX2 is critical for maintenance of stem cells (7), SOX4 may specify transit-amplifying progenitor cells that are the immediate daughters of adult stem cells and have been proposed to be the population that gives rise to cancer stem cells. In humans, SOX4 is expressed in the developing breast and osteoblasts and up-regulated in response to progestins (8). SOX4 is up-regulated at the mRNA and protein level in prostate cancer cell lines and patient samples, and this up-regulation is correlated with Gleason score or tumor grade (9). In addition, SOX4 is overexpressed in many other types of human cancers, including leukemias, melanomas, glioblastomas, medulloblastomas (10), and cancers of the bladder (11) and lung (12). A meta-analysis examining the transcriptional profiles of human cancers found SOX4 to be 1 of 64 genes up-regulated as a general cancer signature (12), suggesting that SOX4 has a role in many malignancies. Furthermore, SOX4 cooperates with Evi1 in mouse models of myeloid leukemogenesis (13). Recently, we showed that SOX4 can induce anchorage-independent growth in prostate cancer cells (9). Consistent with the concept that SOX4 is an oncogene, three independent studies searching for oncogenes have found SOX4 to be one of the most common retroviral integration sites, resulting in increased mRNA (14–16).

Despite these findings, the role that SOX4 plays in carcinogenesis remains poorly defined. Whereas the transactivational properties of SOX4 have been characterized (17), genuine transcriptional targets remain elusive. To date, three studies have used expression profiling of cells after either small interfering RNA (siRNA) knockdown or overexpression of SOX4 to identify candidate downstream target genes (9, 11, 18). Very recently, 31 SOX4 target genes were confirmed by chromatin immunoprecipitation (ChIP) in a hepatocellular carcinoma cell line (19). Although interesting, this study was limited by the fact that it focused on a specific tumor stage transition and did not use a genome-wide localization approach.

Here, we have performed a genome-wide localization analysis using a ChIP-chip approach to identify those genes that have SOX4 bound at their proximal promoters in human prostate cancer cells. We have identified 282 genes that are high-confidence direct SOX4 targets, including many genes involved in microRNA (miRNA) processing, transcriptional regulation, developmental pathways, growth factor signaling, and tumor metastasis. We have also used...
unique protein-binding DNA microarrays (PBM; refs. 20–22) to query the binding of recombinant SOX4 to every possible 8-mer. The PBM-derived SOX4 DNA binding data will further facilitate computational analyses of genomic SOX4 binding sites. These data provide new insights into how SOX4 affects key growth factor and developmental pathways and how these changes may influence cancer progression.

Materials and Methods

Cell culture and stable cell line construction. All cell lines were cultured, as described by American Type Culture Collection except LNCaP cells, which were cultured with T-Medium (Invitrogen). HA-tagged SOX4 was cloned into the pHRL-UBQ-IRES-eYFP-ΔU3 lentiviral vector (gift from Dr. Hihn Ly, Emory University), and stable cells were isolated, as previously described (23).

ChIP. Two 90% confluent P150s of both LNCaP-YFP and LNCaP-YFP/HA-SOX4 or RWPE-1-YFP and RWPE-1-YFP/HA-SOX4 cells were formaldehyde fixed and sonicated, and ChIP assay was performed, as described previously (23). Anti-HA 12CA5 or mouse IgG was used to immunoprecipitate protein-DNA complexes overnight at 4°C and collected using Dyna M280 sheep anti-mouse IgG beads for 2 h. Dyna beads were washed, protein-DNA complexes were eluted, and DNA was purified, as described previously (24). A detailed description of the ChIP-chip protocol can be found in Supplementary Methods. Anti-HA 12CA5, anti–Flag-M2 (Sigma-Aldrich), or mouse IgG was used to immunoprecipitate protein-DNA complexes overnight at 4°C. All PCR primers used in ChIP-PCR can be found in Supplementary Table S7.

ChIP-chip analysis. To determine the direct SOX4 target genes on a global scale, we performed ChIP assays in triplicate from the LNCaP cell line stably expressing SOX4 and in duplicate from a control cell line that expressed YFP alone. Immunoprecipitated and input DNA were subjected to whole genome amplification, Cy3/Cy5 fluorescent labeling, and hybridization to the NimbleGen 25K human promoter array set. Input and immunoprecipitated DNA isolated from LNCaP-YFP and LNCaP-YFP/HA-SOX4 cells were amplified using linker-mediated PCR as described previously (25). Amplified DNA was labeled and hybridized in triplicate by NimbleGen Systems, Inc., to their human 25K promoter array. This set consists of two microarrays that tile 4 kb of upstream promoter sequence and 750 bp of downstream intronic sequence on average, with a total genomic coverage of 110 Mb. Raw hybridization data were Z-score normalized, and ratios of immunoprecipitation to input DNA were determined for each sample. ChIPOTle software was used to determine enriched peaks using a 500-bp sliding window every 50 bp, as previously described (23). NimbleGen microarray data are available from the GEO database accession number GEO11915.

Luciferase assays. PCR fragments representing the binding sites in the EGFR, ERBB2, and TLE1 genes were cloned in front of the pGL3-promoter luciferase construct (Promega). Primers sequences used can be found in Supplementary Table S7. LNCaP cells were transfected with 100 ng of TK-Basic construct, 500 ng of pGL3-promoter vector alone and with cloned inserts, and 500 ng of either a SOX4 or vector expression construct. Dual luciferase assays were performed 48 h posttransfection, according to the manufacturer’s guidelines (Promega). All assays were performed in triplicate on separate days.

Quantitative real-time PCR. LNCaP cells were plated in six-well culture dishes and grown to 90% confluency before transfection with 1 μg of SOX4 plasmid or vector control using Lipofectamine 2000 (Invitrogen). At 24 h posttransfection, total RNA was harvested using the RNeasy kit (Qiagen), and reverse transcription was performed using Superscript III reverse transcriptase (Invitrogen). Quantitative real-time PCR (qPCR) was performed using SYBR Green I (Invitrogen) on a Bio-Rad iCycler using 18s or y-actin as a control, and data were analyzed using the ΔΔCt method (26). All primers used in this study are listed in Supplementary Table S7.

Microarray analysis. Total RNA was isolated from three independent experiments of either vector control or SOX4-transfected LNCaP cells, as described (23). Dual luciferase assays were performed 48 h posttransfection, according to the manufacturer’s guidelines (Promega). All assays were performed in triplicate on separate days.

Affymetrix U133A GeneChip microarray analysis of SOX4 overexpression and knockdown in LNCaP prostate cancer cells. Overexpression of SOX4 leads to increased EGFR expression, whereas siRNA knockdown of SOX4 results in decreased EGFR expression. B, schematic showing the location of the SOX4 binding site in the first intron of the EGFR (top) and ERBB2 (bottom) genes. Arrows denote location of the SOX4 binding site. C, ChIP assay of FLAG-SOX4 bound to the introns of EGFR, ERBB2, and TLE1. PSMA is shown as a negative control. SOX4 bound DNA is specifically amplified in the FLAG immunoprecipitation lane from FLAG-SOX4 expressing cells (lane 3) and not control cells (lane 5) or with a nonspecific antibody (lanes 2 and 4).

Figure 1. A, Affymetrix U133A GeneChip microarray analysis of SOX4 overexpression and knockdown in LNCaP prostate cancer cells. Overexpression of SOX4 leads to increased EGFR expression, whereas siRNA knockdown of SOX4 results in decreased EGFR expression. B, schematic showing the location of the SOX4 binding site in the first intron of the EGFR (top) and ERBB2 (bottom) genes. Arrows denote location of the SOX4 binding site. C, ChIP assay of FLAG-SOX4 bound to the introns of EGFR, ERBB2, and TLE1. PSMA is shown as a negative control. SOX4 bound DNA is specifically amplified in the FLAG immunoprecipitation lane from FLAG-SOX4 expressing cells (lane 3) and not control cells (lane 5) or with a nonspecific antibody (lanes 2 and 4). D, luciferase reporter assays with SOX4 binding sites showing activation in the presence of SOX4 compared with empty vector. *, P < 0.01 by Student’s t test; bars, SD (n = 3 independent biological replicates performed on separate days).
described above. Each transfection was performed in triplicate, and each sample was hybridized in duplicate, creating six data points for each condition. Total RNA was submitted to the Winship Cancer Institute DNA Microarray Core facility. All samples showed RNA integrity of 8.3 or greater using an Agilent 2100 Bioanalyzer. RNA was hybridized to the Illumina Human6 v2 Expression Beadchip that query roughly 47,000 transcripts with 48,701 probes, and after normalization, significantly changed probes were calculated using significance analysis of microarrays (SAM) software (27). Settings for SAM were two-class unpaired (versus vector control) imputation engine (10 nearest neighbor), permutations (500), RNG seed (1234567), Delta (1.316), fold change (1.5), and false discovery rate (0.749%). Microarray data are available in the GEO database accession number GEO11915.

Results
SOX4 transcriptionally activates EGFR. Using expression profiling to determine the genes whose mRNA levels change when SOX4 is either overexpressed or eliminated using siRNA (9), we identified EGFR as a candidate SOX4 transcriptional target (Fig. 1A). Analysis of the promoter and first intron of EGFR and other family members with CONFAC software (28) revealed the presence of potential SOX4 binding sites within the first intron of EGFR and ERBB2 (Fig. 1B). CONFAC functions by identifying the conserved sequences in the 3-kb proximal promoter region and first intron of human-mouse orthologue gene pairs and then identifying transcription factor binding sites (TFBS), defined by position weight matrices from the MATCH software (29), which are conserved between the two species (28).

Whereas limited commercial antibodies exist for SOX4 and show activity in immunoblots, in our hands, none of them have been equal loading, immunoblots were also probed with a mouse monoclonal antibody to protein phosphatase 2A (PP2A) catalytic subunit (BD Biosciences).

Immunoblotting. Cells were lysed in lysis buffer [0.137 mol/L NaCl, 0.02 mol/L TRIS (pH 8.0), 10% glycerol, and 1% NP40], and 50 μg total lysate were separated by SDS-PAGE electrophoresis transferred to nitrocellulose for immunoblotting. Immunoblots were probed with polyclonal rabbit SOX4 antisera described previously (9) and DICER (Santa Cruz). To control for equal loading, immunoblots were also probed with a mouse monoclonal antibody to protein phosphatase 2A (PP2A) catalytic subunit (BD Biosciences).

...
useful in a ChIP assay. Therefore, we used epitope-tagged SOX4, as described in other SOX4 ChIP studies (9, 19). Although the FLAG epitope tag was not tested directly for activity, a glutathione S-transferase (GST)-SOX4 construct showed binding to a known SOX4 motif and not a control motif (Supplementary Fig. S2B), validating that the epitope tag does not interfere with SOX4 binding. To determine if SOX4 directly bound the EGFR and ERBB2 enhancers, we performed ChIP analysis on RWPE-1 prostate cancer cells stably infected with FLAG-SOX4 or a control lentiviral vector. DNA representing the predicted SOX4 sites was specifically amplified from the FLAG-SOX4 cell line and not from the control cell line, indicating that SOX4 binds to intronic sequence of EGFR and ERBB2 (Fig. 1C). EGFR is expressed in RWPE-1 cells, but not in LNCaP cells, and SOX4 did not bind to these sequences in LNCaP cells (data not shown).

To characterize the transcriptional effect of SOX4 levels on the regions bound by SOX4 in ChIP assays, the amplified ChIP fragments were cloned in front of a minimal promoter luciferase reporter plasmid and tested in transient transfections in LNCaP cells. Compared with a vector control, SOX4 significantly increased transcription of the EGFR fragment 3-fold and the TLE1-positive control fragment roughly 4-fold. Although not found significant, ERBB2 was activated 1.5-fold compared with the vector control (Fig. 1D). Consistent with microarray data, SOX4 transcriptionally activates the EGFR enhancer.

**Genome-wide localization analysis.** To determine the direct SOX4 target genes on a global scale, we performed ChIP assays in triplicate from the LNCaP HA-SOX4 stable cell line and in duplicate from the control LNCaP-YFP cell line. Peaks ($P < 0.001$) that overlapped in at least two of the three data sets and were not

![Figure 3.](image-url)
present in the LNCaP-YFP cell line were called significant (Fig. 2A). Based on these variables, we classified 3,600 significant, overlapping peaks as SOX4 target sequences. Because some transcription start sites (TSS) are quite close to each other (<3 kb), it was not always possible to assign a unique gene to every peak. In addition, many genes had multiple peaks in their promoters, and thus, we mapped the 3,600 peaks to 3,470 different genes (Supplementary Table S1).

To verify the set of 3,600 SOX4 peaks, 28 candidate SOX4 target sites representing a range of P values in promoters of genes of biological interest were chosen, primers were designed around the peaks and enrichment was verified by conventional ChIP. Ten of these 28 candidates were analyzed by ChIP qPCR and 18 by ChIP-PCR. Overall, 24 of 28 (86%) of the candidate targets were confirmed, validating our data set. All 10 of the peaks chosen to validate by qPCR were reproducibly enriched over the YF7 control in both the LNCaP-HA-SOX4 cell line and the RWPE-1 cell line (Fig. 2B). Of the target sites validated by conventional PCR, 14 of 18 were confirmed in both the LNCaP and RWPE-1 cell lines, whereas a mock, control PCR was negative (Fig. 2C and D; data not shown). The only exception was ANKRD15, which was enriched only in the LNCaP cell line and not in the RWPE-1 line.

**Target gene expression analysis.** To determine whether SOX4 binding affects transcription of the 3,470 genes that have SOX4 bound at their promoters, we performed whole genome expression analysis on LNCaP cells after transfection with SOX4 or a control vector. To increase the likelihood of identifying direct SOX4 targets, total RNA was isolated at a relatively early time point (24 hours posttransfection) and hybridized to Illumina Human 6-v2 whole genome arrays. A total of 1,766 genes were changed at least 1.5-fold in both the LNCaP-HA-SOX4 cell line and the RWPE-1 cell line (Fig. 3B); of those 1,766 genes, 244 were also direct SOX4 targets by ChIP-chip analysis (Fig. 3; Supplementary Table S2). Seven of these genes were confirmed by qPCR (Fig. 3B).

Our previous expression profiling of LNCaP cells after SOX4 siRNA knockdown (9) identified 465 downstream targets, and we confirmed that SOX4 regulates the expression of DICER, DLL1, and HES2 in LNCaP cells by qPCR (Fig. 3B). We further confirmed SOX4 regulation of DICER at the protein level (Fig. 3C). Out of those 465 candidate targets, 47 genes overlapped with the 3,470 ChIP-chip targets, increasing the number of direct SOX4 targets to 282 genes (Fig. 3A; Supplementary Table S3). We classified these 282 genes bound by SOX4 in ChIP-chip and significantly changed by expression profiling as high confidence direct SOX4 target genes. Nine genes (PHK4CA, DHX9, BTN3A3, CDK2, MVK, ADAM10, RYK, ISG20, and DBI) overlapped in all three data sets. The transcription factor SON and purine biosynthetic enzyme GART, two genes on chromosome 21 that are transcribed in opposite directions and regulated by a bidirectional promoter, were affected in opposite ways. SON was activated by SOX4 1.8-fold, as detected by SOX4 overexpression, whereas GART was increased almost 3-fold as determined by SOX4 siRNA knockdown, suggesting that SOX4 regulates the directionality of this promoter.

We next analyzed the P values of the peaks in our ChIP-chip data set, comparing the P values of the genes that were altered by transient overexpression of SOX4 with those that were not (Supplementary Fig. S2). We found no difference in the distributions of the ChIP-chip P values for those genes that were changed in expression profiling experiments and those that were not. Thus, based on our ChIP-chip validation experiments and the similar P-value distributions, we conclude that SOX4 is genuinely bound at the promoters of the 3,188 genes that did not change but that SOX4 by itself is not limiting or sufficient to generate changes in transcription without corresponding changes in the cellular context, such as activation of cofactors or signaling pathways.

**Novel SOX4 position weight matrix.** To facilitate computational analyses of SOX4 DNA binding sites, we sought to determine the DNA binding preferences of SOX4 using universal PBMs (20). This universal PBM array allows recombinant SOX4 protein to interact with and bind every possible 8-mer, thus allowing in vitro binding site specificities to be calculated.

We generated an NH2 terminal, GST-SOX4-DBD fusion protein, expressed and purified it from E. coli, and tested for activity (Supplementary Fig. S3). The GST-SOX4-DBD was incubated with the protein binding microarray and a novel position weight matrix (PWM; RWYAAWRV) was calculated from the PBM data (Supplementary Table S4) using the Seed-and-Wobble algorithm (Fig. 3D; ref. 20). Three groups have previously reported similar binding site sequences for SOX4: AACAAG (30), AAACAT (31), and WWCATAWG (19). Our PWM confirms the SOX4 core binding sequence of the previously known binding sites but there are some differences in the specificity at the 1st and 7th positions and we find a bias toward A, C, and G at the 8th position. These differences could be due to the fact that earlier reports used no more than 31 sequences to develop the binding motif, whereas our study queried every possible 8-mer.

**SOX4 peaks contain SOX4 binding sites.** Using our newly derived PWM, we applied CONFAC software (28) to analyze the enriched sequences for the presence of SOX4 binding sites. We analyzed the sequences of the peaks in the promoters of our 282 high confidence genes against 10 sets of control promoter sequences to see if SOX4 sites were enriched in our target gene set. Control promoter peaks of equal size to SOX4 peaks were chosen randomly from sequences covered by the NimbleGen array, and each control set contained equal total sequence coverage as our 282 high confidence peaks. With stringent criteria (core similarity, >0.85; matrix similarity, >0.75), we find that 60% of the peaks contain SOX4 binding sites. SOX4 sites were significantly enriched relative to 10 sets of random promoter sequence by Mann-Whitney U test using Benjamini correction for multiple hypothesis testing (q < 0.0019).

To further characterize the SOX4 binding sites, we searched the entire set of 3,600 SOX4 peaks and 10 equal sets of random promoter sequence for the presence of PBM-bound k-mers (here, ungaapped 8-mers). The specificity of PBM k-mers can be quantified by the enrichment score (ES), which ranges from −0.5 to 0.5 (32). We analyzed the enrichment of PBMs k-mers with 0.45 > ES > 0.40 (moderate) and ES > 0.45 (stringent). Whereas both SOX4-bound peaks and random promoter sequence contained moderate and stringent k-mers, SOX4 peaks contained significantly more stringent (P = 0.0002) and moderate (P = 1.08 × 10−5) k-mers by two-tailed Mann-Whitney test (Supplementary Fig. S4).

To investigate interaction with protein partners that may increase SOX4 affinity for poor matching sites in vivo, we searched for enrichment of cooccurring TFBS in the SOX4 peaks. We applied CONFAC software to search the sequences for the presence of co-occurring transcription factor binding sites within the same peak (Table 1). Using the same criteria as above, we determined that the E2F family had the most frequently co-occurring motif (similar to TTTCGCGC, q = 1.78 × 10−11). Interestingly, ingenuity pathway analysis (IPA) identified cell cycle as a functionally enriched process in the 3,470 SOX4 target genes (P = 0.00916), suggesting...
that part of SOX4’s function is to control the expression of genes involved in cell cycle progression.

CONFAC analysis identified other significant TFBS motifs enriched in the SOX4 peaks (Table 1), including those for transcription factors in the TGFβ, Wnt, and NF-κB pathways. SOX4 modulates Wnt signaling via interaction with β-catennin and the TCF4 transcription factor (2), suggesting a possible role for SOX4 in transcriptionally modulating Wnt signals. We confirmed the recent report that SOX4 cooperates with constitutively active β-catennin to activate TOP-Flash luciferase reporters (2) and found that SOX4 synergistically induces activation of these constructs, further highlighting a role for SOX4 in the Wnt pathway (Supplementary Fig. S5).

**SOX4 target genes.** To determine the biological processes and functions of the SOX4 targets, we performed a gene ontology analysis using DAVID software (33) on the 282 high confidence SOX4 targets. Among the SOX4 targets were 23 transcription factors (Table 2), and DAVID analysis determined that the top annotations were transcription (P = 3.7 x 10^-18), transmembrane (P = 5.59 x 10^-10), and protein phosphorylation/dephosphorylation (P = 3.5 x 10^-18/6.6 x 10^-7). These findings are paralleled by expression profiling of SOX4 overexpression in HU609 bladder carcinoma cells where top annotated functions were signal transduction and protein phosphorylation (11).

Commercial IPA software identified biological pathways and functions that are enriched in our 282 high confidence targets, 1,766 significant genes identified by SAM analysis, and the 3,470 functions that are enriched in our 282 high confidence targets, we performed a gene ontology analysis using DAVID software (9) on the 282 high confidence SOX4 target genes (Supplementary Table S2). We confirmed these data by qPCR (Fig. 3B) and Western blot for DICER (Fig. 3C).

Gene set enrichment analysis (GSEA; ref. 36) and GSEA leading edge analysis (37) of these gene sets identified TGFβ--induced SMAD3 direct target genes (Supplementary Table S3) as enriched in SOX4 target genes. SOX4 is up-regulated by TGFβ-1 treatment (4, 38), and we found SMAD4 sites are significantly enriched in the SOX4 ChIP-chip peaks (Table 1), suggesting that SOX4 affects key developmental and growth factor signaling pathways in prostate cancer cells at both the transmembrane signaling and transcriptional levels.

**Table 1.** Benjamini corrected q values for co-occurring transcription factor binding sites

<table>
<thead>
<tr>
<th>Transcription factor</th>
<th>Family</th>
<th>Benjamini corrected q value</th>
</tr>
</thead>
<tbody>
<tr>
<td>E2F4</td>
<td>E2F</td>
<td>1.78E−11</td>
</tr>
<tr>
<td>E2F1</td>
<td>E2F</td>
<td>3.06E−11</td>
</tr>
<tr>
<td>PAX5</td>
<td>Paired box</td>
<td>2.07E−10</td>
</tr>
<tr>
<td>WHN</td>
<td>Forkhead</td>
<td>2.94E−10</td>
</tr>
<tr>
<td>SMAD3</td>
<td>SMAD</td>
<td>1.82E−09</td>
</tr>
<tr>
<td>SMAD4</td>
<td>SMAD</td>
<td>3.33E−09</td>
</tr>
<tr>
<td>MYC</td>
<td>MYC</td>
<td>6.25E−09</td>
</tr>
<tr>
<td>NFKAPPA2B</td>
<td>NF-κB</td>
<td>2.95E−08</td>
</tr>
<tr>
<td>LEF1/TCF1</td>
<td>LEF</td>
<td>1.12E−06</td>
</tr>
</tbody>
</table>

**Table 2.** DAVID analysis identified 23 transcription factors present in our high confidence SOX4 target genes

<table>
<thead>
<tr>
<th>Entrez ID</th>
<th>Symbol</th>
<th>Microarray fold change</th>
</tr>
</thead>
<tbody>
<tr>
<td>196528</td>
<td>ARID2</td>
<td>1.99</td>
</tr>
<tr>
<td>2008</td>
<td>ELF3</td>
<td>2.65</td>
</tr>
<tr>
<td>3169</td>
<td>FOXA1</td>
<td>2.47</td>
</tr>
<tr>
<td>2976</td>
<td>GTF3C2</td>
<td>3.12</td>
</tr>
<tr>
<td>64412</td>
<td>GZF1</td>
<td>2.42</td>
</tr>
<tr>
<td>84585</td>
<td>LCOR</td>
<td>2.41</td>
</tr>
<tr>
<td>4173</td>
<td>MCM4</td>
<td>1.55</td>
</tr>
<tr>
<td>58508</td>
<td>MLL3</td>
<td>2.06</td>
</tr>
<tr>
<td>10933</td>
<td>MORF4L1</td>
<td>2.07</td>
</tr>
<tr>
<td>8031</td>
<td>NCOA4</td>
<td>2.64</td>
</tr>
<tr>
<td>4784</td>
<td>NFIX</td>
<td>2.83</td>
</tr>
<tr>
<td>4824</td>
<td>NKX3-1</td>
<td>4.53</td>
</tr>
<tr>
<td>7779</td>
<td>PRDM2</td>
<td>2.48</td>
</tr>
<tr>
<td>5933</td>
<td>RBL1</td>
<td>1.80</td>
</tr>
<tr>
<td>55309</td>
<td>SNF7</td>
<td>2.32</td>
</tr>
<tr>
<td>6722</td>
<td>SRF</td>
<td>2.03</td>
</tr>
<tr>
<td>54816</td>
<td>SUH4</td>
<td>1.93</td>
</tr>
<tr>
<td>9412</td>
<td>SURB7</td>
<td>2.24</td>
</tr>
<tr>
<td>9338</td>
<td>TCEAL1</td>
<td>1.57</td>
</tr>
<tr>
<td>7718</td>
<td>ZNF165</td>
<td>1.53</td>
</tr>
<tr>
<td>7738</td>
<td>ZNF184</td>
<td>1.66</td>
</tr>
<tr>
<td>23528</td>
<td>ZNF281</td>
<td>1.71</td>
</tr>
<tr>
<td>30834</td>
<td>ZNRD1</td>
<td>1.63</td>
</tr>
</tbody>
</table>

NOTE: Gene ontology term: transcription, DNA dependent (P = 3.7 x 10^-18).

9 http://www.ingenuity.com
to determine a novel PWM specific for SOX4 and show that our ChIP-chip predicted peaks are significantly enriched for SOX4 binding sites. These data provide several new insights into the roles that SOX4 plays in the cell.

**SOX4 direct target genes.** Although only 10% of the significant differentially expressed genes overlapped with the ChIP-chip data, this is likely a conservative estimate because the NimbleGen 25k promoter array only queries proximal promoter sequences and not more than 1 kb downstream of the TSS. We found that SOX4 binds EGFR and ERBB2 in the first intron over 20 kb downstream of the TSS (Fig. 1D), and unsurprisingly, we did not detect EGFR or ERBB2 in our ChIP-chip experiment. Thus, more of the 1,900 genes that responded to changes in SOX4 mRNA levels (but were not detected by ChIP-chip) could still be direct targets. Excellent candidates would be the 40 genes that responded to SOX4 on both microarray platforms, such as the IL6 receptor, SOX12, and NME1 (Supplementary Table S6). Whereas 3,600 is a fairly large number of SOX4 bound regions, some background can be expected. Nevertheless, we were able to validate 24 of 28 (86%) candidate binding sites chosen, adding confidence to our data set. In fact, an even higher number of over 4,200 genomic binding sites had been previously observed for c-Myc in ChIP–positron emission tomography whole genome studies (39). Whole genome tiling arrays or ChIP-seq could provide additional binding sites that may show more overlap with the Illumina expression data set.

Conversely, many of the bound genes may not respond to changes in SOX4 mRNA levels alone but to multiprotein activator complexes of which SOX4 is only one component. Furthermore, the stability of SOX4 bound to a promoter could be greater than unbound SOX4, limiting the effects observed by siRNA knockdown. In different cell types or cellular contexts, SOX4 may activate a different subset of these genes. Of the 31 SOX4 target genes reported by Liao and colleagues (19), only six are represented in our NimbleGen data set and three found to be changed in our Illumina expression profiling data set. The small overlap could be due to the fact that those genes were identified in hepatocellular carcinomas, whereas we have examined prostate cancer cells. Interestingly, DKK was one of the six genes that overlapped in both data sets, further implicating SOX4 in the Wnt pathway. Because SOX4 is known to interact with β-catenin and other coactivators, it may be poised at many of these promoters to enable responses to developmental signals from the Wnt or TGFβ pathway.

**Receptor and signaling regulation.** Our data suggest that SOX4 regulates cellular differentiation through a variety of transcription factors and receptors. SOX4 is up-regulated in response to numerous external ligands ranging from TGFβ (38),...
and BMP-6 (40) to parathyroid hormone and progesterone (8). Previous work has shown that SOX4 directly signals from IL-5R (41), and here, we have shown that SOX4 directly regulates EGFR (Fig. 1). Membrane receptors in the SOX4 transcriptional network also include Frizzled family members FZD3, FZD5, FZD8; the Hedgehog receptor PTCH-1; the Notch ligand DLL1; TRAIL decoy receptor TNFRSF10D; and other growth factor receptors, such as FGFRL1 and IGF2R. DAVID analysis also revealed protein phosphorylation/dephosphorylation ($P = 3.5 \times 10^{-18}/6.6 \times 10^{-7}$) and transcription ($P = 3.7 \times 10^{-18}$) are enriched annotations, identifying 23 transcription factors that are direct targets of SOX4. This evidence suggests that SOX4 regulates signaling events both at the external input level and the internal output or transcription level. This regulation could be direct, as with IL-5R, or through the transcriptional targets SOX4 activates.

**Transcription factors and SOX4.** Here, we have reported DNA binding specificity data for SOX4, which will improve computational analyses for SOX4-specific binding sites. Our data confirm the known SOX family core-binding motif and add new specificity at the 1st, 7th, and 8th positions. Whereas crystal structure evidence from SOX2 has shown the importance of the core-binding motif, it is possible that the specificity for SOX4 is enhanced outside of the core motif at the extra positions. A limitation of these data is that we did not assess how other DNA binding proteins influence the sequences to which SOX4 can bind. The enrichment of SMAD4 sites is particularly interesting in light of the GSEA results, which suggest that SOX4 regulates many TGFβ target genes, including Tenascin C. Thus, we hypothesize that SOX4 may physically interact with SMAD4 in response to TGFβ signals. Experiments to test this hypothesis are under way. Nevertheless, evidence points to a role for SOX4 in modulating other transcriptional programs via hierarchical regulation of 23 downstream transcription factors.

**SOX4 and cancer.** Based on the target genes we identified, SOX4 seems to influence cancer progression in several ways. First, it plays a key role in the activation of and response to developmental pathways, such as Wnt, Notch, Hedgehog, and TGFβ. Second, SOX4 inhibits differentiation via repression of transcription factors, such as NKKX1, and activation of MLL and MLL3, two histone H3 K4 methyltransferases that induce activation of HOX gene expression (42). MLL methyltransferase complexes also facilitate E2F activation of S-phase promoters, facilitating cell cycle progression. Activation of MLL also suggests a mechanism for the role of SOX4 in myeloid leukemogenesis, because MLL is a critical oncogene that is often translocated or amplified in this disease (43). Thirdly, SOX4 targets growth factor receptors, such as EGFR, FGFRL1, and IGF2R, enhancing proliferative signals in tumors and potentially activating the PI3K-AKT pathway. Mice heterozygous for NKKX1 and PTEN in the prostate develop prostate adenocarcinomas and metastases to the lymph node (44). Thus, our data suggest that SOX4 may promote prostate cancer progression directly through NKKX3.1 repression and indirectly through PI3K-AKT activation. Finally, SOX4 seems to promote metastasis via up-regulation of tenascin C. Recently, both SOX4 and tenascin C were shown to enhance metastasis of breast cancer cells to the lung (45), as has the TGFβ pathway, which activates their expression (46). Other metastasis-associated SOX4 target genes include integrin α7, and Rac1. Rac1 was recently shown to control nuclear localization of β-catenin in response to Wnt signals (47).

**SOX4 regulates components of the RISC complex and small RNA pathway.** miRNAs are small noncoding RNA species that regulate the translation and stability of mRNA messages for hundreds of downstream target genes via partial complementarity to short sequences in the 3′ untranslated regions of mRNAs. The RISC, which is composed of AGO1 or AGO2, TRBP, and Dicer processes miRNAs from precursors (pre-miRNA) to their mature form, cleaves target mRNAs, and participates in translational inhibition. RNA Helicase A (RHA/DHX9) interacts with the RISC complex and participates in loading of small RNAs into the RISC complex (48). We observed that three components of the RISC complex, DICER, AGO1, and RHA/DHX9, are high-confidence direct targets of SOX4 (Supplementary Table S3), and we confirmed these data by qPCR (Fig. 3B). Dicer has been independently observed to be overexpressed in prostate cancers (49).

In addition, we observed that Toll-like receptor 3 (TLR3), which binds to double-stranded RNAs, induces gene silencing, and can induce apoptosis (50), was induced 2.8-fold upon overexpression of SOX4. This induction may be indirect because TLR3 was not detected by ChIP-chip, but we cannot exclude the possibility that SOX4 may directly regulate TLR3 from a distal or intronic enhancer.

Our observation that SOX4 targets three genes important in small RNA processing is of particular interest in light of the role of SOX4 in development and cancer progression. miRNAs have been implicated in numerous physiologic processes from development to oncogenesis. miRNAs can also act as suppressors of breast cancer metastasis via targeting of tenascin C and SOX4 (45) and as promoters of breast cancer metastasis (51). The finding that SOX4 can affect expression of multiple components of the RISC complex also provides insight into why long-term loss of SOX4 induces widespread apoptosis (9, 18). In summary, these data shed light on the mechanisms and pathways through which SOX4 may exert its effects during development and cancer progression. Further studies are necessary to elucidate the precise role of SOX4 in the functioning of these pathways.

**Disclosure of Potential Conflicts of Interest**

No potential conflicts of interest were disclosed.

**Acknowledgments**

Received 9/3/2008; revised 10/8/2008; accepted 11/3/2008.

**Grant support:** National Cancer Institute R01 CA106026 (C.S. Moreno), NIH/NIHGNI grant R01 HG003985 (M.F. Berger and M.L. Bulyk), DOD CDMPR Prostate Cancer Predoctoral Training Fellowship PC060145 (C.D. Scharer), and Postdoctoral Training Fellowship PC060114 (C.D. McCabe).

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.

We thank Dr. Maja Orlandic-Kodani at Winship Cancer Institute Microray Core Facility for performing Illumina microarray labeling and hybridization, Robert Karafa at the Emory University FlowCore for cell sorting, Dr. Hinh Ly for IRES-eYFP lentiviral vector, and Dr. Anita Corbett for pGEX-4T-1 plasmid.

References


Cancer Res 2009; 69: (2). January 15, 2009 716 www.aacrjournals.org

Downloaded from cancerres.aacrjournals.org on April 13, 2017. © 2009 American Association for Cancer Research.
Genome-Wide Promoter Analysis of the SOX4 Transcriptional Network in Prostate Cancer Cells

Christopher D. Scharer, Colleen D. McCabe, Mohamed Ali-Seyed, et al.