OMICS A Journal of Integrative Biology
Volume 14, Number 4, 2010

DOI: 10.1089/omi.2010.0053

ChIP-seq and Functional Analysis of the SOX2 Gene in Colorectal Cancers

Xuefeng Fang,[1,2] Wei Yu,[2] Lisha Li,[2] Jiaofang Shao,[2] Na Zhao,[2] Qiyun Chen,[2] Zhiyun Ye,[3] Sheng-Cai Lin,[3] Shu Zheng,[1] and Biaoyang Lin[2,4,5]

Abstract

SOX2 is an HMG box containing transcription factor that has been implicated in various types of cancer, but its role in colorectal cancers (CRC) has not been studied. Here we show that SOX2 is overexpressed in CRC tissues compared with normal adjacent tissues using immunohistochemical staining and RT-PCR. We also observed an increased SOX2 expression in nucleus of colorectal cancer tissues (46%, 14/30 cases vs. 7%, 2/30 adjacent tissues). Furthermore, knockdown of SOX2 in SW620 colorectal cancer cells decreased their growth rates *in vitro* cell line, and *in vivo* in xenograft models. ChIP-Seq analysis of SOX2 revealed a consensus sequence of wwTGywTT. An integrated expression profiling and ChIP-seq analysis show that SOX2 is involved in the BMP signaling pathway, steroid metabolic process, histone modifications, and many receptor-mediated signaling pathways such as IGF1R and ITPR2 (Inositol 1,4,5-triphosphate receptor, type 2).

Introduction

Colorectal cancer is the second most common malignancy in cancer patients and cause of cancer-related mortality (Wiesner et al., 2003). The SOX (SRY-like HMG box) gene family represents a family of transcriptional factors characterized by the presence of a conserved HMG (high mobility group) box in their genes. Thus far, 20 SOX genes have been identified in humans and mice, and they can be divided into 10 subgroups on the basis of sequence similarity and genomic organization (Schepers et al., 2002). SOX genes bind to the minor groove in DNA to control diverse developmental processes and play critical roles in cell fate determination, differentiation, and proliferation (Schepers et al., 2002). Recently, Takahashi et al. showed that SOX2 is a key transcription factor that can induce pluripotency in both mouse and human somatic cells (Li et al., 2004; Takahashi et al., 2007; Takahashi and Yamanaka, 2006). Importantly, SOX2 is one of the four factors (OCT4, SOX2, NANOG, and LIN28) that can reprogram human somatic cells to pluripotent stem cells that exhibit the essential characteristics of embryonic stem (ES) cells (Yu et al., 2007). Giorgetti et al. (2009) also showed that OCT4 and SOX2 could be used to generate induced pluripotent stem cells from human cord blood.

SOX2 is also overexpressed in several cancers including gastric cancer, breast cancer, pancreatic cancer, pulmonary nonsmall cell carcinomas, lung squamous cell carcinomas, neuroendocrine carcinomas (Gure et al., 2000; Hussenet et al., 2010; Li et al., 2004; Rodriguez-Pinilla et al., 2007; Sanada et al., 2006; Wang et al., 2009a). However, the role of SOX2 in colorectal cancer cells has not been studied. Only one study suggested that levels of SOX2 could be used, together with two other stem cell markers CD133 and OCT4, for prediction of distant recurrence and poor prognosis of rectal cancer patients treated with preoperative CRT (Saigusa et al., 2009). In this study, we have carried out analysis of SOX2 expression in colorectal cancer tissues and found that it is overexpressed in the cancerous tissues compared to normal adjacent tissues. We also carried out *in vitro* and *in vivo* functional analysis of SOX2 by knocking down SOX2 expression SW620 colorectal cancer cells. Furthermore, we characterized the SOX2 response program in colorectal cancer cells through an integrative analysis of expression profiling and ChIP-seq data.

[1]Cancer Institute (Key Laboratory of Cancer Prevention and Intervention, China National Ministry of Education), The Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, Zhejiang, People's Republic of China.
[2]Systems Biology Division, Zhejiang–California International Nanosystems Institute (ZCNI), Zhejiang University, 268 Kaixuan Road, Hangzhou, Zhejiang, People's Republic of China.
[3]Key Laboratory of Ministry of Education for Cell Biology and Tumor Cell Engineering, School of Life Sciences, Xiamen University, Fujian, People's Republic of China.
[4]Swedish Medical Center, Seattle, Washington.
[5]Department of Urology, University of Washington, Seattle, Washington.

Materials and Methods

Cell culture, lentiviral transduction, and transfection

SW620, SW480, HT29, CACO2, RKO, and HCT116 colorectal cancer cell lines were obtained from American Type Culture Collection (Manassas, VA). MISSION shRNA Lentiviral Particles for SOX2 knockdown and MISSION Nontarget shRNA control transduction particles were purchased from Sigma-Aldrich (St. Louis, MO). The SOX2 target sequences for shRNA are: CCGGCAGCTCGCAGACCTACATGAACTCG AGTTCATGTAG GTCTGCGAGCTGTTTTT and CCGGCTG CCGAGAATCCATGTATATCTCGAGATATACATGGATTC TCGGCAGTTTTT. SW620 cell stably expressing siRNAs were generated by transduction with the virus particles followed by selection using puromycin.

Immunoblot and immunohistochemistry (IHC) analysis

The SOX2 (ab59776) antibody (Abcam Inc., Cambridge, MA) was used for both analyses. For immunoblot analysis, some 20 μg of total cellular protein was loaded per lane, separated by 4–12% SDS-polyacrylamide gel electrophoresis, and then transferred to nitrocellulose (Invitrogen, Carlsbad, CA) by electroblotting.

We used the IHC services provided by Superchip, Inc. (Sanford, FL), including antibody optimization, IHC staining, pathological reading, and scoring by experienced pathologists. Primary antibodies for SOX2 were diluted at 1:250 for IHC. Secondary antibody was used at 1:200 dilution. For isotype control antibodies, rabbit IgG (Cat# 0111-01, Southern Biotech, Birmingham, AL) (5 mg/mL) was used at 1:25 dilution. Colorectal cancer tissue array (Lot ID: OD-CT-DgCol03) from Superchip, Inc. (Superchip, Shanghai, China) was used for IHC. Individual tissue samples were arrayed in duplicate cores. The tissue array core diameter is 1.5 mm and the core thickness is 5 μm.

The scoring criteria contain two parameters: percentage of positive cell population, and staining intensities. For percentage of positive cell population, the categories are: 0 = 0%; 1 = 1 to 25%; 2 = 26 to 50%; 3 = 51 to 75%; 4 = 76–100%. The staining intensities were scored as: − = negative staining; + = weak staining intensity; ++ = medium staining intensity; +++ = strong staining intensity.

Wound-healing migration assay

About 1.2 million cells were seeded into six-well plates and incubated at 37°C with 5% CO_2 until confluent. Wounding was introduced to the monolayer of cells by scraping the surface with a sterile pipette tip. The healing process was examined at 72 h later and recorded with a Canon S 80 digital camera with a microscope adapter.

Xenograft studies

Some 2×10^4 Cells were harvested, washed, resuspended in 200 μL phosphate-buffered saline (PBS), and was subcutaneously injected into the flanks of 5-week-old female nude mice. Animal experimental procedures were performed strictly in accordance with the related ethics regulations of our university. Tumor sizes were measured in two dimensions with calipers every week. Tumor volumes (mm^3) were calculated using the following formula: $V = (\text{length}\times\text{width}^2)/2$ (Chen et al., 2008b).

Cell proliferation assays and cell cycle analysis

Cell proliferation was analyzed using the MTT assay kits (Millipore, Billerica, MA) according to the manufacturer's protocol. For cell cycle analysis, cells were harvested and washed with PBS, followed by fixation with 70% ethanol overnight at 4°C. After washing with PBS twice, the cells were resuspended in PBS containing 50 mg/mL propidium iodide and 10 mg/mL RNase A for 30 min at room temperature in the dark. Samples were analyzed for DNA content by a flow cytometer (BD Bioscience, San Jose, CA). The cell-cycle phases were analyzed using CELLQuest software (Becton Dickinson, San Jose, CA) (Tseng et al., 2009).

Apoptosis analysis

Apoptotic cells were quantified using Annexin V-FITC apoptosis detection kit (BD Pharmingen, San Diego, CA). The number of apoptotic cells was analyzed by flow cytometry (Ex = 488 nm; Em = 530 nm) (Cho et al., 2008).

Soft agar colony formation assay

For Soft agar colony formation assay, cells were trypsinized and counted. 10,000 cells were seeded in six-well plates. After 2 weeks of growth, colonies with a diameter greater than 4 mm were counted. Experiments were performed in quadruplicates (Foltz et al., 2006; Lee et al., 2010).

Microarray

Total RNA was extracted from cells expressing SOX2 silencing or nonsilencing shRNA with Trizol (Invitrogen) and followed by further purification step using RNeasy (Qiagen, Valencia, CA). RNA quality was determined with the Agilent Bioanalyzer (Palo Alto, CA). Affymetrix U133 plus2 arrays were used. Microarray hybridization was performed as we described previously (Lin et al., 2009). Genes that showed more than a twofold change were considered differentially expressed.

Chromatin immunoprecipitation (ChIP)-sequencing

About 3×10^6 SW620 cells were used for ChIP according to the manufacturer's instruction (Millipore, Bedford, MA, EZ-Magna ChIP™ A). Antibodies used for Chip included SOX2 (ab59776, Abcam, Cambridge, MA) and normal IgG (SC-2027, Santa Cruz, CA). Briefly, cells were fixed by crosslinking with 1% fresh formaldehyde. The fixed cells were resuspended in lysis buffer. Nuclei were collected and resuspended in nuclei lysis buffer. Samples were sonicated on ice to the length of 200–500 bp. A total of 5μg antibody and 50 μL Dynal protein G beads were incubated for 2 h at 4°C. Sonicated chromatin were incubated with the protein G-antibody complex overnight at 4 °C. Precipitated immunocomplex was treated with proteinase K for 2 h at 65°C, and DNA was purified Qiagen Qiaquick PCR purification kit.

ChIP-seq analysis

ChIP DNA end repairing, adaptor ligation, and amplification were performed as described earlier (Lin et al., 2009). The fragments of about 100 bp (without linkers) were isolated from agarose gel and used for sequencing using the Solexa/Illumina 2 G genetic analyzer. Solexa Pipeline Analysis was

performed as described (Lin et al., 2009). Sequencing tags of 36 nucleotides were mapped to human genome (hg19) using SOAP program (Wang et al., 2009b) allow two mismatches. Sequence reads that map to multiple sites in the human genome were removed. The mapped sequence reads were converted to ELAND format using our own perl script. MACS program were used to determine enriched SOX2 peaks using IgG-ChIP as control. The parameters for MACS were: effective genome size = 2.70e + 09; tag size = 36 (use 5′ sequence 36 bp); model fold = 32; pvalue cutoff = 1.00e-05. The enriched peaks were annotated with the HG19 gene annotation using CisGenome program (Ji et al., 2008).

To compare the SOX2 targets we identified with the *sox2* targets identified in mouse ES cells (Chen et al., 2008a), we used the CisGenome program (Ji et al., 2008) to identify the targets corresponding to the mouse ChIP-Seq data for sox2 using the same parameter of a distance of 50 kb as the cutoff value. We then identified the gene commons to the two lists using the homologene table for human and mouse from NCBI (http://www.ncbi.nlm.nih.gov/homologene).

Validation of differentially expressed genes and ChIP-seq targets by RT-PCR

Randomly selected differentially expressed genes were determined by real-time PCR using the ABI PRISM 7900 HT Sequence Detection System (Applied Biosystems, Carlsbad, CA). The real-time PCR reactions were carried out in a total volume of 20 μL per well containing SYB master mix reagent kit (Applied Biosystems) with primers listed (Supplementary Table 1; http://systemsbiozju.org/data/Supplementary/). For ChIP-seq analysis, the primers were listed in the Supplementary Table 2.

Results

Increased expression of SOX2 in colorectal cancer tissues compared with normal colorectal tissues. Increasing evidence shows that SOX2 is expressed in several types of tumors prompted us to investigate whether SOX2 was also expressed in colorectal cancers. We performed quantitative RT-PCR on a panel of 15 fresh colorectal cancer samples and 11 individual normal colorectal tissues and showed SOX2 mRNA expression were significantly ($p < 0.01$) higher in colorectal cancer tissues compared with normal tissues (Fig. 1A). To analyze its protein expression, we performed immunohistochemical staining in 30 colorectal cancer tissues and 30 normal adjacent tissues. We observed positive immunoreactivities in colorectal cancer with nucleus staining pattern in 46% (14 of 30) of cancer tissues, compared to only 7% in the adjacent mucosa tissue cells. Examples of IHC staining results are shown in Figure 1B. Statistical analysis using Pearson chi-square ($df = 1$, two-sided) indicates that the difference in SOX2 expression between cancer and adjacent tissues is significant ($p < 0.01$).

Western blot analysis of SOX2 in established colorectal cancer cell lines showed that the levels of SOX2 in RKO, SW620, Caco2, and HCT116 lines were considerably higher than those in SW480 and HT29 cell lines (Fig. 1C). SW480 and SW620 were a matched pair of primary and metastatic population of cells from the same patient (Kubens and Zanker, 1998).

Knockdown of SOX2 decreased the growth rate of SW620 cells

We used SW620 cells for the experiments, because SW620 cells have the abundant native expression level of SOX2 protein and have the metastatic property among the six colorectal cancer cell lines that we screened. We established stable SOX2 knockdown clones by lentivirus-mediated ShRNA gene silencing technology (Sigma-Aldrich, St. Louis, MO) and determined the SOX2 mRNA and protein expression in the stable clones using RT-PCR and Western blot analysis (Fig. 2A and B). MTT assay revealed that knockdown of SOX2 significantly decreased cell proliferation of SW620 cells (Fig. 2C). To determine the mechanism of the reduced cell proliferation was due to cell growth arrest or apoptosis, we performed FACS analyses. Knocking-down of SOX2 resulted in an accumulation of cells in the G0/G1 phases and a decrease of cells in S and G2/M phases (Fig. 2D). However, the knockdown of SOX2 expression had no measurable effect on the apoptosis of SW620 cells (Fig. 2E).

SOX2 knockdown decreased cell colony formation on soft agar and in vivo *tumorigenesis potential*

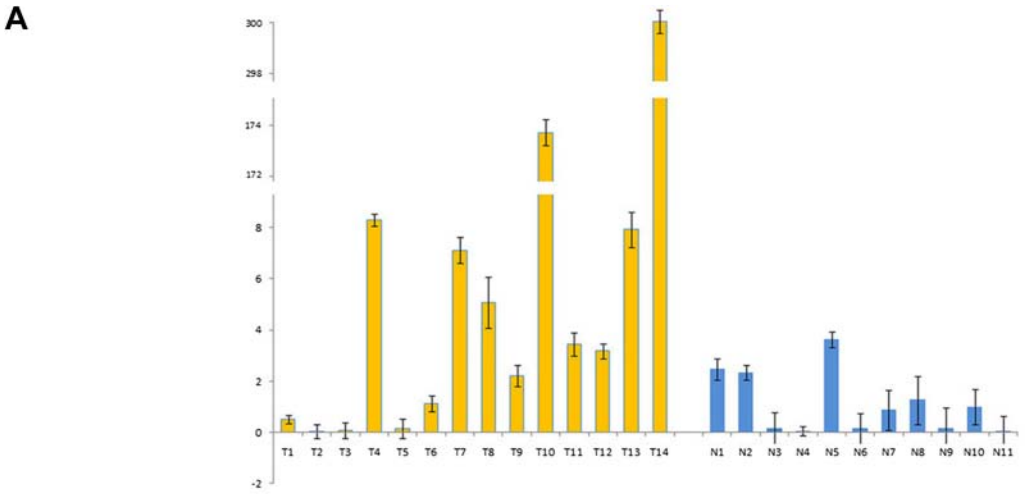
Anchorage-independent growth assays demonstrated that knockdown of SOX2 expression resulted in a reduction in both the number and sizes of colonies formed on the soft agar (Fig. 3A). In order to assess the role of SOX2 on colorectal tumorigenesis *in vivo*, equal numbers of SW620 cells and SOX2 knockdown SW620 cells were implanted onto flanks of 5-week-old female nude mice, and the growth of the implanted tumors was measured at weeks 1, 2, 3, and 4. The results of these experiments indicated that knockdown of SOX2 expression resulted in a dramatic reduction in tumor volume (Fig. 3B) ($p < 0.01$).

Knockdown of SOX2 decreased migratory property of SW620 cells

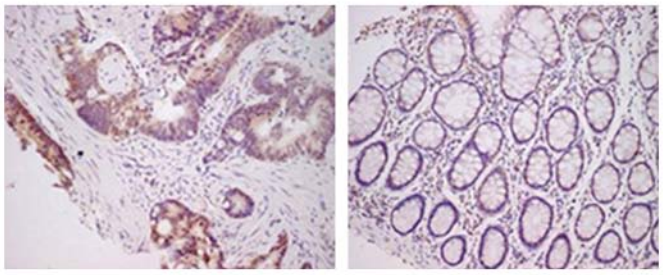
Wound-healing assays have been applied for the analysis of migration of colorectal cancer cells (Dhawan et al., 2005). We also performed a wound-healing assay to examine the effect of SOX2 knockdown on cell migration. Cells were seeded in six-well plates until confluent, and wounds were introduced by scraping a line on the cell lawn using a pipette tip. At the time of the next 72 h, cells migration was monitored. We found that the migration of the parental SW620 cells across the wounded area was faster than SOX2 knockdown cells (Fig. 3C). Interestingly, remarkable cell morphologic alterations were observed in SW620 cells with SOX2 knockdown compared with the parental SW620 cells, with the cells displaying a compact shape and lacked ruffles, protrusions on their surfaces (Fig. 3D).

Identification of SOX2 regulated genes by microarrays

We used Affymetrix array U133 plus 2 to identify genes regulated by SOX2 by comparing expression profiles of parental SW620 cells and SOX2-knockdown SW620 cells. Data were submitted to GEO with the accession number GSE20689. We found that the expression of 1,715 probes (corresponding to 1,205 genes and 124 unannotated probes) was decreased by twofold or more in the SOX2 knockdown SW620 cells comparing with the mock control cells (Supplementary Table 3). In



B



	Cancer tissues	normal tissues
Total number of samples	30	30
Samples with sox2 expression in nucleus	14 (46%)	2(7%)

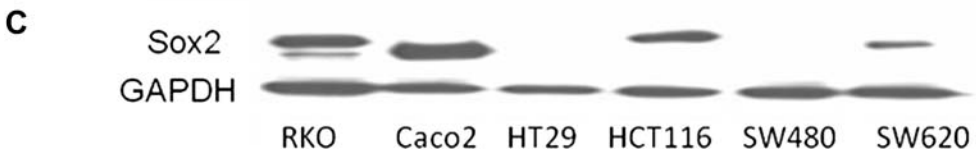


FIG. 1. Upregulation of SOX2 in human colorectal carcinomas. (**A**) A bar chart showing real-time PCR of SOX2 in colon adenocarcinomas tissue (T, tumor tissue) and normal tissue mucosa (N, normal tissue). Y-axis, normalized relative expression levels. (**B**) Immunostaining for SOX2 protein in tissue of carcinomas and adjacent normal tissue mucosa. Top: representative pictures of carcinomas (left) and normal tissue (right). Bottom: a summary of the IHC findings. (**C**) Western blot detection of SOX2 protein (38 kDa) in different colon cancer cell lines.

addition, 2,109 probes (corresponding to 1,381 annotated genes, and 182 unannotated probes) were upregulated in SOX2-knockdown cells compared with the mock control cells.

To further validate the SOX2-responsive gene expression detected by microarrays, we selected six genes (ST14, MTSS1, NGFRAP1, FGFR2, Jun, and CDKN2C) and determined their expression by real-time-PCR. We confirmed the expression of ST14, MTSS1, NGFRAP1, Jun, FGFR2, and CDKN2C in SW620 cells of SOX2 knockdown comparing to that in the controls (Fig. 4A).

Gene Ontology (GO) analysis of downregulated genes by SOX2 knockdown revealed that the GO terms cell surface receptor linked signal transduction (GO: 0007166), BMP signaling pathway (GO:0030509), response to retinoic acid (GOL0032526), response to vitamin A (GO: 0033189), and steroid metabolic process (GO:0008202) were enriched at FDR <0.05 (Table 1). However, there were no significantly (FDR <0.05) enriched GO terms for the genes that were upregulated by SOX2 known down.

We also analyzed the differentially expressed genes using GSEA with the C2:CP gene sets, which is the collection of the pathways from various data bases, http://www.broadinstitute.org/gsea/msigdb/collection_details.jsp). We found the SOX2 upregulated genes (i.e., those under ex-

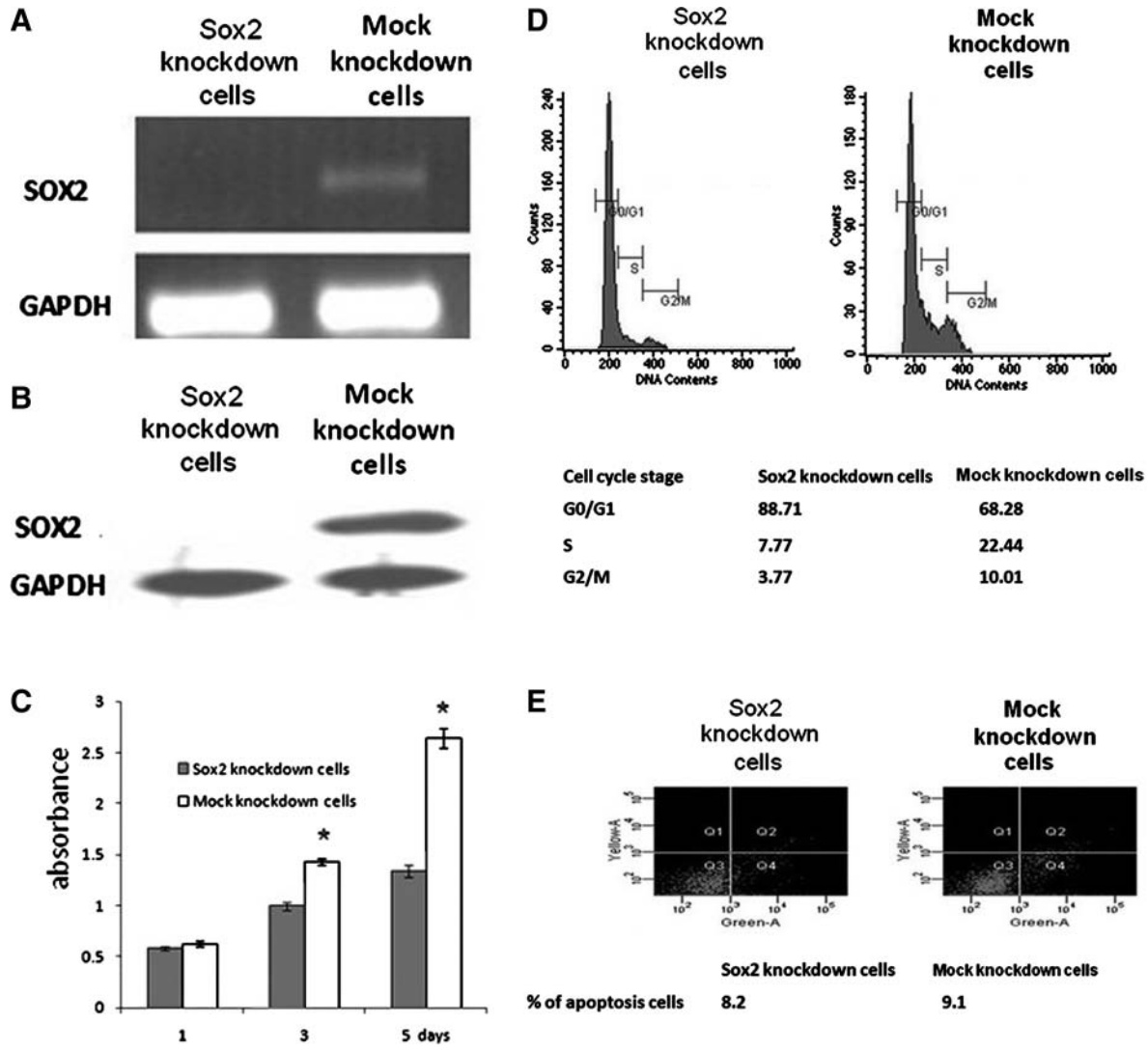


FIG. 2. Knockdown of Sox2 inhibits cell growth in colorectal cells. (**A**) Analysis of SOX2 RNA levels by RT-PCR in SOX2 knockdown SW620 cells and SW620 MOCK knockdown cells. (**B**) Western blot analysis of SOX2 protein expression in SOX2 knockdown SW620 cells and SW620 MOCK knockdown cells. (**C**) Effect of SOX2 knockdown on cell proliferation by MTT assay. The asterisks indicate statistical significance at $p < 0.01$. (**D**) Cell cycle analysis SOX2 knockdown SW620 cells and SW620 MOCK knockdown cells by FACS. A summary table is provided for cells in different stages of cell cycles. (**E**) Evaluation of apoptosis in SOX2 knockdown SW620 cells and SW620 MOCK knockdown cells. The percentage of cells entering apoptosis was determined by FACS analysis using FITC-labeled annexin V kit and shown at the bottom panel.

pressed in SOX2 knockdown cells compared to MOCK knock down cells) were enriched for many pathways including HSA04514 cell adhesion molecules, HSA04670 leukocyte transendothelial migration, HSA01430 cell communication, HSA04530 tight junction, etc. ($p < 0.05$) (Supplementary Table 4). Key genes regulated by SOX2 and their functions are listed in Table 2.

ChIP-seq analysis of SOX2 binding sites in colorectal cancer cells

In order to understand the genome-wide binding patterns of SOX2, we applied ChIP-seq technology, which is a novel approach for identifying transcription factor binding sites genome-wide (Barski et al., 2007; Jothi et al., 2008). We performed replicate SOX2 ChIP and IgG ChIP. After sequencing analysis. We obtained 13,437,404 and 16,165,014 raw reads, respectively, for SOX2_ChIP and IgG_ChIP, and 7,241,344 and 8,291,710 passed the default quality filter of the GA pipeline of Illumina Inc., respectively. We then applied MACS to identify SOX2 enriched binding regions (Zhang et al., 2008). We identified 1,086 enriched SOX2 binding region in the human genome with mfold = 32 and p-value cutoff <1.00E-05 (Supplementary Table 5). We randomly picked 10 genes for which the promoter regions are enriched for the SOX2 IP, and we were able to confirm all 10 genes to be enriched in the SOX2 IP DNAs compared to the IGG-IP DNAs using real-time quantitative PCR (Fig. 4B), suggesting that the false positive rate is negligible in our dataset.

We mapped the SOX2 binding regions to genes using the UCSC human genome annotations (hg19) using a distance of 50 kb as the cutoff value (Supplementary Table 5). We

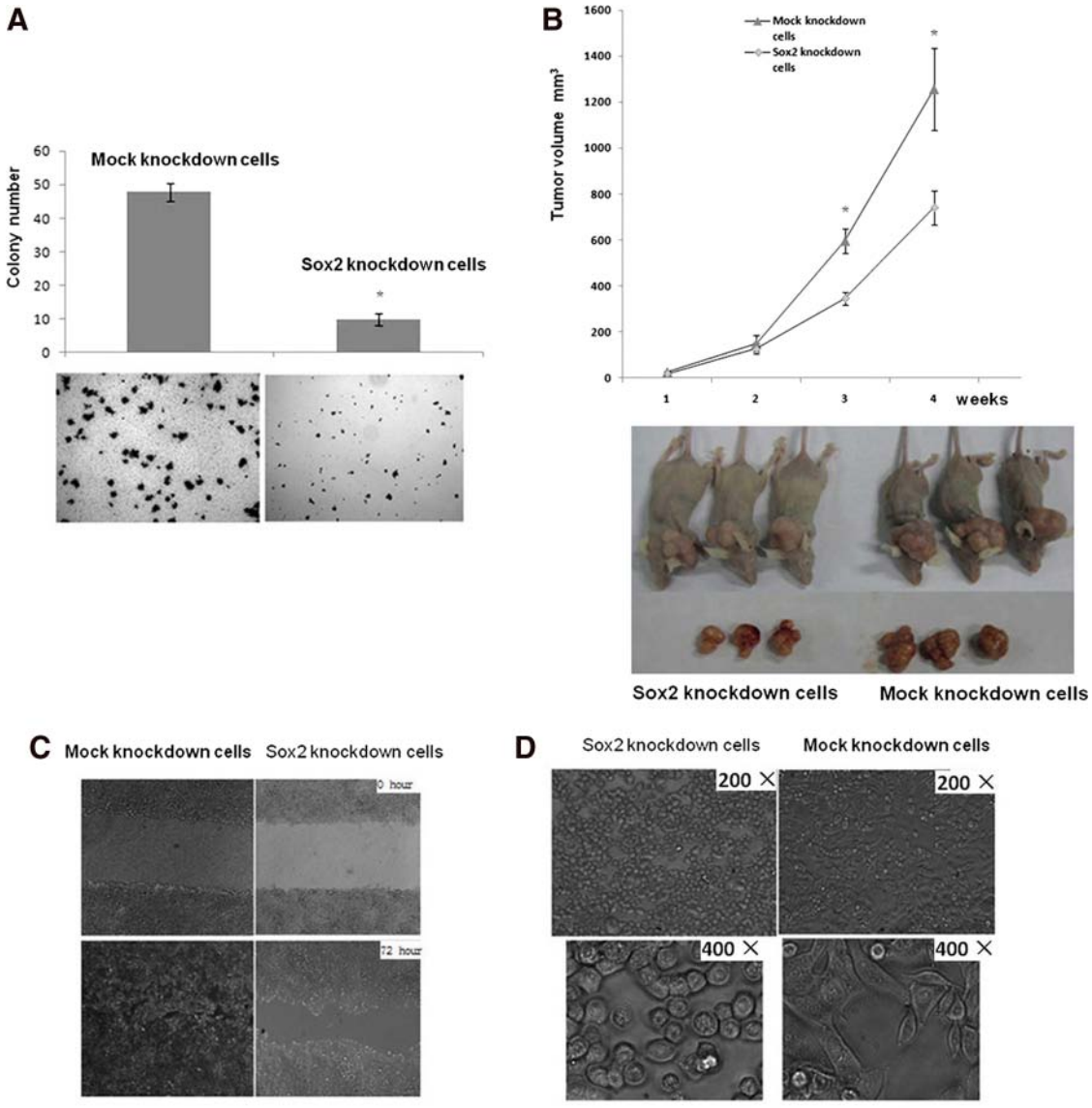


FIG. 3. SOX2 knockdown decreases cell colony formation on soft agar, *in vivo* tumorigenesis potential and migratory property of SW620 cells. (**A**) Sox2 knockdown SW620 cells and SW620 MOCK knockdown cells were analyzed for their ability to form colonies. Data are representatives of three independent experiments. The asterisk indicates statistical significance at $p < 0.01$. (**B**) Growth rates of SOX2 knockdown SW620 cells and SW620 MOCK knockdown cells in *in vivo* mouse model. Volumes of tumors were monitored every week. Y-axis, tumor volumes; X-axis, weeks after inoculation of cells. The asterisk indicates statistical significance at $p < 0.01$. Bottom: representative pictures of tumor samples. (**C**) A wound-healing assay of SOX2 knockdown SW620 cells and SW620 MOCK knockdown cells. Photographs were taken at the time of 0 h and 72 h time points. Representative photos from one of three replicate experiments are shown (100× original magnification). (**D**) Morphological alterations of SW620 cells with SOX2 knockdown. The Sox2 knockdown cells show a compact shape and lacked ruffles, protrusions on their surfaces compared with MOCK knockdown cells. Photos shown at 200× and 400× original magnification.

found that 676 (62.3%) SOX2 binding regions could be mapped to within 50 kb of TSS or TES of genes. Among them, only 136 (20%) mapped to within 10 kb of TSSs and 56 mapped to within 2 kb of TSSs, suggesting that the majority of the SOX2 binding sites are far away from proximal promoter regions.

Using the homologene table for human and mouse from NCBI (http://www.ncbi.nlm.nih.gov/homologene), we compared the SOX2 targets that we identified in CRC cells with the *sox2* targets that were identified in mouse ES cells (Chen et al., 2008a). We identified 122 common targets (Table 3). Many interesting genes were among the common targets including two other SOX family members (SOX5 and SOX6), tumor suppressor or oncogene such as WT1 (Wilms tumor 1) and PIM1 (Pim-1 oncogene), two catennins {CTNNA3 [Catenin (cadherin-associated protein), alpha 3] and CTNNBL1 (Catenin, beta like 1)}, several solute carrier proteins including SLC23A2 [Solute carrier family 23 (nucleobase transporters), member 2], SLC24A3 (Solute carrier family 24, sodium/potassium/calcium exchanger, member 3), SLC35F1 (Solute

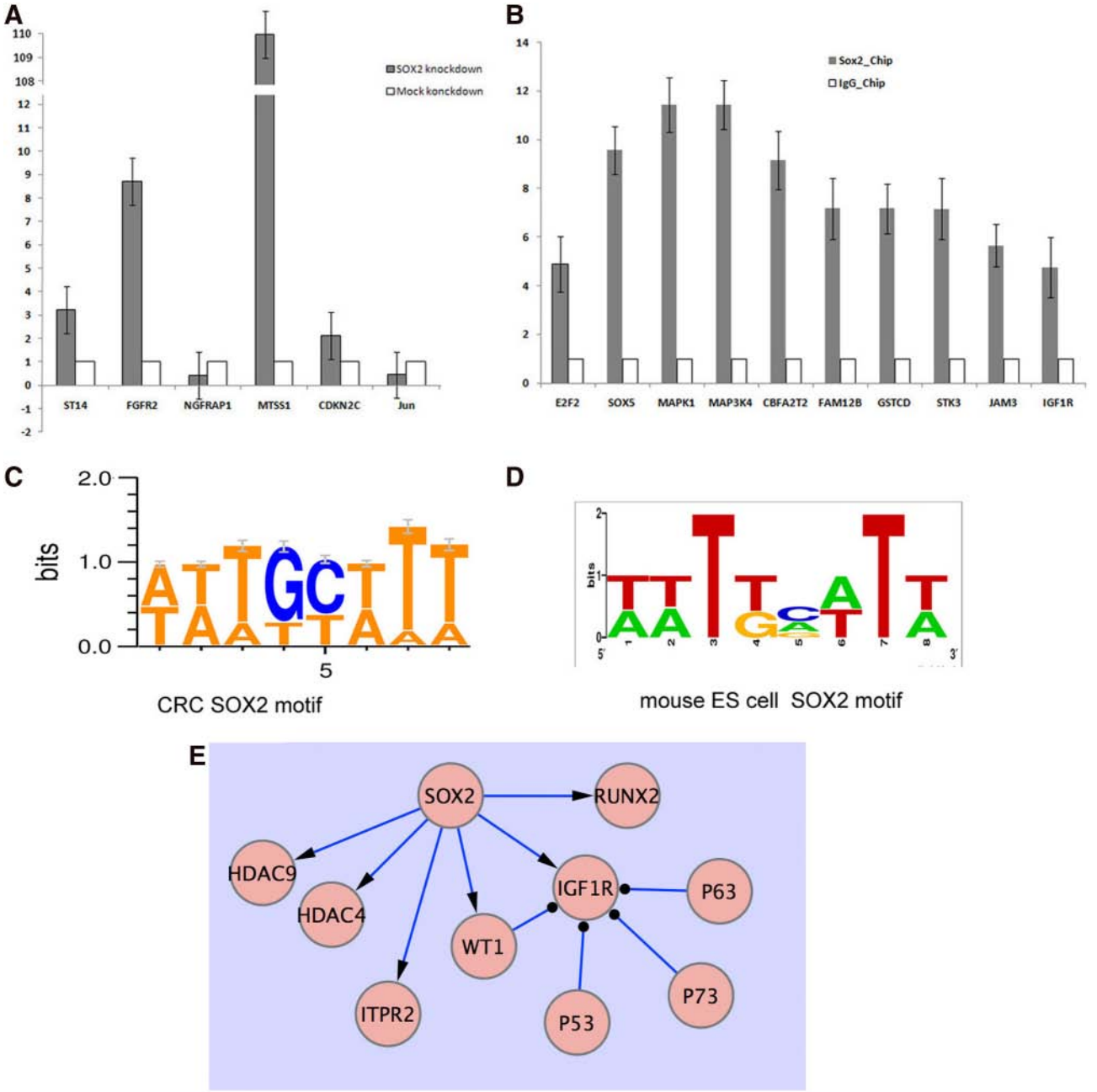


FIG. 4. Validation of differentially expressed genes and ChIP-seq targets by RT-PCR and the consensus sequence of SOX2 binding. (**A**) Expression of *ST14, MTSS1, FGFR2, CDKN2C, NGFRAP1,* and *Jun* was analyzed with real-time PCR in SOX2 knockdown and MOCK knockdown SW620 cells. (**B**) Quantitative real-time PCR for the confirmation of ChIP-seq peaks. Relative amount of PCR products from SOX2–ChIP and IgG–ChIP are shown as bar graphs. Standard deviations are also shown. (**C**) The DNA binding consensus sequence of SOX2 in SW620 cells identified by the MotifSampler program. (**D**) The DNA binding consensus sequence of the mouse ES cells from Chen et al. (2008a) and reanalyzed by the motifSampler program. (**E**) Cytoscape network view of SOX2 and its key interacting proteins. Direction lines showed direct SOX2-DNA binding interaction from our data and the lines ending with cycles indicate negative regulation information derived from literatures. The network was drawn with the Cytoscape program.

carrier family 35, member F1), SLC35F3 (Solute carrier family 35, member F3), and HDAC9 (Histone deacetylase 9). These data demonstrated versatile functions of SOX2.

An integrative analysis of SOX2 ChIP-seq data and expression profiling data revealed that only 94 of the 676 SOX2 binding regions (that could be mapped to within 50 kb of a gene) (about 14%) changed in gene expression (Supplementary Table 5). Among them, 53 SOX2 binding regions showed positive correlation (i.e., SOX2 binding would result in overexpression) and 41 SOX2 binding region showed negative

Table 1. Enriched GO Terms for SOX2-Regulated Genes in CRC

GO category	*Changed genes*	*Enrichment*	*Log10(p)*
Downregulated genes by SOX2 knockdown			
GO:0007165_signal_transduction	1,064	169	1.34
GO:0007154_cell_communication	1,175	182	1.31
GO:0007166_cell_surface_receptor_linked_signal_transduction	384	71	1.56
GO:0008544_epidermis_development	54	17	2.66
GO:0032501_multicellular_organismal_process	989	152	1.30
GO:0030509_BMP_signaling_pathway	19	9	4.00
GO:0021700_developmental_maturation	12	7	4.92
GO:0007398_ectoderm_development	61	18	2.49
GO:0042060_wound_healing	53	16	2.55
GO:0007167_enzyme_linked_receptor_protein_signaling_pathway	147	32	1.84
GO:0007275_multicellular_organismal_development	706	111	1.33
GO:0001501_skeletal_system_development	73	19	2.20
GO:0009611_response_to_wounding	126	28	1.88
GO:0008202_steroid_metabolic_process	52	15	2.44
GO:0032526_response_to_retinoic_acid	11	6	4.60
GO:0033189_response_to_vitamin_A	11	6	4.60
GO:0051240_positive_regulation_of_multicellular_organismal_process	47	14	2.51

correlation (SOX2 binding would result in under expression) (Supplementary Table 5).

Identification of DNA binding consensus and SOX2 collaborating TFs in the SOX2 binding regions

To see whether the human SOX2 binding regions in CRC cells have their own unique and enriched binding motif, we used the MotifSampler program (http://homes.esat.kuleuven.be/~thijs/Work/MotifSampler.html) (Thijs et al., 2002a, 2002b) to identify binding consensus sequences enriched in the SOX2 binding regions that we identified. We found a consensus sequence wwTGywTT (Fig. 4C) with a very high log-likelihood score of 1,535.58. The output matrix for this consensus sequence is shown in Supplementary Document 6, and there are 628 instances of this motif in 1086 SOX2 binding regions (Supplementary Table 5). The consensus logo is shown in Figure 4C. We were curious whether the sequence motif shows any similarities to known binding motifs. Using JASPAR: an open-access database for eukaryotic transcription factor binding profiles (http://jaspar.genereg.net/) (Sandelin et al., 2004), we found that the highest ranked match is to FOX11 (matrix MA0042.1) with a score of 7.6 and a percent score of 95.39. FOX11 belongs to the forkead family transcription factor, winged helix-turn-helix class. SOX2 belongs to the High Mobility Group (HMG) transcription factor, other alpha-helix class (http://jaspar.genereg.net/).

We then wondered whether known TFs could bind to the SOX2 binding regions that we identified and act as SOX2 cooperators for the regulation of gene expression. In order to systematically search for potential bindings of other transcription factors, we used the MotifScanner program (http://homes.esat.kuleuven.be/~thijs/download.html) and scanned all TF motif matrices (PWM databases) using the human transcription factor subset of the Transfac professional 7.0. Matched matrices with likelihood (LR) ratios of 500 or higher were tabulated and frequencies calculated (Supplementary Table 7). We found that three transcription factor OCT1, SOX1, and CEBP seem to colocalize at >50% of the times with SOX2 (Supplementary Table 7).

Discussion

In the present study, we have demonstrated that expression levels of SOX2 mRNA and SOX2 protein were frequently overexpressed in primary colorectal cancer tissues (Fig. 1A and B) ($p < 0.01$). We knocked down the expression of SOX2 SW620 colorectal cancer cell lines, and found that reduced expression of SOX2 correlates with decreased growth rates, colony formation, and migration abilities of SW620 cells (Figs. 2C and 3A–C). In addition, knocking down of SOX2 resulted in an accumulation of cells in the G0/G1 phases and a decrease of cells in S and G2/M phases (Fig. 2D). However, the knockdown of SOX2 expression had no measurable effect on the apoptosis of SW620 cells (Fig. 2E). Using *in vivo* mouse models, we showed that SOX2 knockdown cells had reduced tumor sizes compared with the MOCK known down cells (Fig. 3B). Taken together, these data suggested functional important roles of SOX2 in colorectal carcinogenesis. Expression changes revealed by DNA microarray analysis showed that many pathways involved in cell adhesion, cell communication, cell migration, tight junction, fatty acid metabolism, mitochondria pathway, TGF-beta signaling are enriched in the MOCK transfected cells (Supplementary Table 4). These expression changes may explain the underlying mechanism of different growth and migration properties of SOX2 knock down and MOCK knock down cells.

An examination of the cell morphological changes revealed that the SOX2 knockdown SW620 cells had a compact shape and lacked ruffles, protrusions on their surfaces compared with control cells (Fig. 3D). Furthermore, by IHC staining, we detected dramatic increase of nuclear localization of SOX2 proteins in colorectal cancer cells compared with cells from normal adjacent tissues (Fig. 1B), suggesting a process of nuclear translocation and accumulation of SOX2 during colorectal cancer carcinogenesis. These data suggested that SOX2 have dramatic effects on cell morphology and cellular biology. It would be of great interest to study the mechanism and the role in cell growth control of SOX2 nuclear translocation as well as the underlying mechanisms for its role in induction of morphological changes.

Table 2. Key Genes Regulated by SOX2 and Their Functions

ACVR1	Activin A receptor, type I	BMP signaling pathway
BMP2	Bone morphogenetic protein 2	BMP signaling pathway
BMP4	Bone morphogenetic protein 4	BMP signaling pathway
BMPR2	Bone morphogenetic protein receptor, type II (serine/threonine kinase)	BMP signaling pathway
PCSK6	Proprotein convertase subtilisin/ kexin type 6	BMP signaling pathway
SMAD5	SMAD family member 5	BMP signaling pathway
SMAD6	SMAD family member 6	BMP signaling pathway
SMAD7	SMAD family member 7	BMP signaling pathway
SMURF1	SMAD specific E3 ubiquitin protein ligase 1	BMP signaling pathway
PCSK6	Proprotein convertase subtilisin/ kexin type 6	cell surface receptor linked signal transduction
IGF1R	Insulin-like growth factor 1 receptor	cell surface receptor linked signal transduction
GNA15	Guanine nucleotide binding protein (G protein), alpha 15 (Gq class)	cell surface receptor linked signal transduction
TGFB2	Transforming growth factor, beta 2	cell surface receptor linked signal transduction
RGS3	Regulator of G-protein signaling 3	cell surface receptor linked signal transduction
FGFR3	Fibroblast growth factor receptor 3 (achondroplasia, thanatophoric dwarfism)	cell surface receptor linked signal transduction
IL6ST	Interleukin 6 signal transducer (gp130, oncostatin M receptor)	cell surface receptor linked signal transduction
EREG	Epiregulin	cell surface receptor linked signal transduction
SOCS5	Suppressor of cytokine signaling 5	cell surface receptor linked signal transduction
IRAK1	Interleukin-1 receptor-associated kinase 1	cell surface receptor linked signal transduction
PIK3R1	Phosphoinositide-3-kinase, regulatory subunit 1 (p85 alpha)	cell surface receptor linked signal transduction
EDN1	Endothelin 1	cell surface receptor linked signal transduction
OPN3	Opsin 3 (encephalopsin, panopsin)	cell surface receptor linked signal transduction
PYCARD	PYD and CARD domain containing	cell surface receptor linked signal transduction
SMURF1	SMAD specific E3 ubiquitin protein ligase 1	cell surface receptor linked signal transduction
SIGIRR	Single immunoglobulin and toll-interleukin 1 receptor (TIR) domain	cell surface receptor linked signal transduction
CD24	CD24 molecule	cell surface receptor linked signal transduction
SMAD7	SMAD family member 7	cell surface receptor linked signal transduction
AGRN	Agrin	cell surface receptor linked signal transduction
RHOQ	Ras homolog gene family, member Q	cell surface receptor linked signal transduction
IRS1	Insulin receptor substrate 1	cell surface receptor linked signal transduction
MAML3	Mastermind-like 3 (*Drosophila*)	cell surface receptor linked signal transduction
GNB2	Guanine nucleotide binding protein (G protein), beta polypeptide 2	cell surface receptor linked signal transduction
COL16A1	Collagen, type XVI, alpha 1	cell surface receptor linked signal transduction
MTSS1	Metastasis suppressor 1	cell surface receptor linked signal transduction
TRIB1	Tribbles homolog 1 (*Drosophila*)	cell surface receptor linked signal transduction
PTK2	PTK2 protein tyrosine kinase 2	cell surface receptor linked signal transduction
BMPR2	Bone morphogenetic protein receptor, type II (serine/threonine kinase)	cell surface receptor linked signal transduction
IFITM1	interferon induced transmembrane protein 1	cell surface receptor linked signal transduction
ACVR1	Activin A receptor, type I	cell surface receptor linked signal transduction
CCRL2	Chemokine (C-C motif) receptor-like 2	cell surface receptor linked signal transduction
TACSTD2	Tumor-associated calcium signal transducer 2	cell surface receptor linked signal transduction
PTPRF	Protein tyrosine phosphatase, receptor type, F	cell surface receptor linked signal transduction
SMAD3	SMAD family member 3	cell surface receptor linked signal transduction
DGKQ	Diacylglycerol kinase, theta 110 kDa	cell surface receptor linked signal transduction
SRC	V-src sarcoma (Schmidt-Ruppin A-2) viral oncogene homolog (avian)	cell surface receptor linked signal transduction
BMP4	Bone morphogenetic protein 4	cell surface receptor linked signal transduction
EGFR	Epidermal growth factor receptor (erythroblastic leukemia viral (v-erb-b) oncogene homolog, avian)	cell surface receptor linked signal transduction
SMAD5	SMAD family member 5	cell surface receptor linked signal transduction

(*continued*)

Table 2. (Continued)

DGKH	Diacylglycerol kinase, eta	cell surface receptor linked signal transduction
KLK6	Kallikrein-related peptidase 6	cell surface receptor linked signal transduction
GAB1	GRB2-associated binding protein 1	cell surface receptor linked signal transduction
VIPR1	Vasoactive intestinal peptide receptor 1	cell surface receptor linked signal transduction
BMP2	Bone morphogenetic protein 2	cell surface receptor linked signal transduction
PARD3	Par-3 partitioning defective 3 homolog (*C. elegans*)	cell surface receptor linked signal transduction
CD14	CD14 molecule	cell surface receptor linked signal transduction
ROR1	Receptor tyrosine kinase-like orphan receptor 1	cell surface receptor linked signal transduction
ADAM17	ADAM metallopeptidase domain 17 (tumor necrosis factor, alpha, converting enzyme)	cell surface receptor linked signal transduction
ADAM10	ADAM metallopeptidase domain 10	cell surface receptor linked signal transduction
NUP62	Nucleoporin 62 kDa	cell surface receptor linked signal transduction
LEPR	Leptin receptor	cell surface receptor linked signal transduction
P2RY2	Purinergic receptor P2Y, G-protein coupled, 2	cell surface receptor linked signal transduction
AFAP1L2	actin filament associated protein 1-like 2	cell surface receptor linked signal transduction
AREG	Amphiregulin (schwannoma-derived growth factor)	cell surface receptor linked signal transduction
HPGD	Hydroxyprostaglandin dehydrogenase 15-(NAD)	cell surface receptor linked signal transduction
STAT3	Signal transducer and activator of transcription 3 (acute-phase response factor)	cell surface receptor linked signal transduction
CXCL1	Chemokine (C-X-C motif) ligand 1 (melanoma growth stimulating activity, alpha)	cell surface receptor linked signal transduction
GPR37	G protein-coupled receptor 37 (endothelin receptor type B-like)	cell surface receptor linked signal transduction
CD59	CD59 molecule, complement regulatory protein	cell surface receptor linked signal transduction
LY6E	Lymphocyte antigen 6 complex, locus E	cell surface receptor linked signal transduction
CBLC	Cas-Br-M (murine) ecotropic retroviral transforming sequence c	cell surface receptor linked signal transduction
PPAP2A	Phosphatidic acid phosphatase type 2A	cell surface receptor linked signal transduction
PSENEN	Presenilin enhancer 2 homolog (C. elegans)	cell surface receptor linked signal transduction
THBS1	Thrombospondin 1	cell surface receptor linked signal transduction
DEFB1	Defensin, beta 1	cell surface receptor linked signal transduction
SMAD6	SMAD family member 6	cell surface receptor linked signal transduction
NMUR2	Neuromedin U receptor 2	cell surface receptor linked signal transduction
ADRA2A	Adrenergic, alpha-2A-, receptor	cell surface receptor linked signal transduction
CENTA1	Centaurin, alpha 1	cell surface receptor linked signal transduction
RGS2	Regulator of G-protein signaling 2, 24kDa	cell surface receptor linked signal transduction
ITPR3	Inositol 1,4,5-triphosphate receptor, type 3	cell surface receptor linked signal transduction
RBP4	Retinol binding protein 4, plasma	response to retinoic acid
HSD17B2	Hydroxysteroid (17-beta) dehydrogenase 2	response to retinoic acid
RXRA	Retinoid X receptor, alpha	response to retinoic acid
AQP3	Aquaporin 3 (Gill blood group)	response to retinoic acid
RARA	Retinoic acid receptor, alpha	response to retinoic acid
TRIM16	Tripartite motif-containing 16	response to retinoic acid
RBP4	Retinol binding protein 4, plasma	response to vitamin A
HSD17B2	Hydroxysteroid (17-beta) dehydrogenase 2	response to vitamin A
RXRA	Retinoid X receptor, alpha	response to vitamin A
AQP3	Aquaporin 3 (Gill blood group)	response to vitamin A
RARA	Retinoic acid receptor, alpha	response to vitamin A
TRIM16	Tripartite motif-containing 16	response to vitamin A

As a vital step toward comprehensive understanding the molecular mechanism of SOX2 function, we set to identify SOX2 regulated genes (all potential genes) by comparing SOX2 knockdown colorectal cancer cells with MOCK knockdown cancer cells. We identified 1,205 genes and 124 unannotated genes that are potentially regulated by SOX2 (Supplementary Table 3). GO analysis revealed that SOX2 is involved the many important biological processes (GO terms) including cell surface receptor-linked signal transduction (GO: 0007166), BMP signaling pathway (GO:0030509), response to retinoic acid (GO:0032526), response to vitamin A (GO: 0033189), and steroid metabolic process (GO:0008202) (Table 1). Looking further into the detailed gene lists identified to be regulated by SOX2 in these GO terms, we found that SOX2 regulated 9 out of 19 genes in the BMP signaling pathway (GO: 0030509) including SMAD5, SMAD6, SMAD7, BMP2, and BMP7, ACVR1 (activin Receptor 1), PCSK6, SMURF1, and BMPR2 (Table 2). SMAD proteins were found to be involved in colorectal cancers. For example, SAMD7 was found to induce hepatic metastasis in colorectal cancer (Halder et al., 2008). SOX2 also regulated many genes that cell surface receptor that are signaling molecules including TGFB2, IGF1R, FGFR3, and EGFR (Table 2), suggesting that SOX2 is probably a master gene that control all these receptor mediated signaling pathways. Many of these pathways have been implicated in colorectal cancers. For example, IGF1R (insulin-like growth factor 1 receptor) enhances invasion and induces resistance to apoptosis of colon cancer cells (Sekharam et al., 2003). The EGFR expression was shown to correlate with more aggressive disease and a poorer prognosis of colorectal cancers, and EGFR is considered a critical target for the treatment of colorectal cancers (O'Dwyer and Benson, 2002). Hu et al. (2010) showed that EGFR–SOX2–EGFR forms a feedback loop that positively regulates the self-renewal of neural precursor cells. Our data showed that SOX2 regulated EGFR. However, whether such an EGFR–SOX2–EGFR feedback loop exists in colorectal cancer cells or their stem cells remain to be investigated.

In order to differentiate SOX2 directly regulated genes from SOX2 indirect regulated genes, we applied ChIP-seq, a novel global technology in identifying transcriptional factor binding sites (Barski et al., 2007; Johnson et al., 2007), to the identification of global SOX2 binding sites in colorectal cancer cells. We identified 1,086 significantly enriched SOX2 binding region in the human genome (Supplementary Table 5). Among them, only 136 (20%) mapped to within 10 kb of transcription start sites (TSSs) of genes, and majority mapped to >10 kb from TSSs, suggesting that SOX2 is probably an enhancer binding transcription factors (i.e., activator) rather than a promoter binding transcription factor. Indeed, only 46 (6.8%) SOX2 binding regions could be mapped to within 1 kb of TSSs. An interesting observation is that SOX2 binds to two enhancer regions in SOX5 and to one enhancer region in SOX6 (Supplementary Table 5), suggesting a regulatory mechanism within the SOX family proteins. However, the resulting binding did not change the expression of SOX5 and SOX6, judged by two fold changes from the comparison of SOX2 knockdown to MOCK cells.

Analysis of the consensus sequences bound to SOX2 revealed a consensus sequence of wwTGywTT (W = A, T; Y = C, T), a GC rich dinucleotide sequence core surrounded by three AT-rich sequences on each site. Our consensus motif is similar to the last eight nucleotides (the reverse and complement stand sequence TTWGCATA,) of the 15-nucleotide consensus sequence identified in embryonic stem cells (Chen et al., 2008a). As there are many programs for identifying DNA binding motifs from DNA binding sequence data including AlignACE (Roth et al., 1998), MEME (Bailey et al., 2006), MotifSampler (Thijs et al., 2002a, 2002b), and PoSSuMsearch (Beckstette et al., 2006), and Chen et al. used their custom written program for analyzing their mouse ES SOX2 ChIP-seq data, it would be best to compare the motif identified using the sample motif finding program. We therefore analyzed the mouse SOX2 ChIP-seq data from Chen's article (Chen et al., 2008a) using the MotifSampler program and the same parameters that we used to colorectal cancer SOX2 ChIP-seq data analysis, and we found that the eight nucleotide motif is wwTkmwTw (K = G, T; M = A, C) (Fig. 4C), which is similar to wwTGywTT (Fig. 4D) that we found for the SOX2 in colorectal cancer cells.

We found that 53 and 41 SOX2 binding regions showed positive and negative correlation with expression changes of genes respectively (Supplementary Table 5). These suggest that other transcription factor or coactivators also play a role in the final outcome of the SOX2 binding. Interesting positive correlated genes include WT1 (Wilms tumor 1), ITPR2 (Inositol 1,4,5-triphosphate receptor, type 2), HDAC4 (Histone deacetylase 4), and HDAC9 (Histone deacetylase 9). WT1 gene is overexpressed in colorectal cancers (Oji et al., 2003). ITPR2 plays an important role in intracellular Ca^{2+} signaling and phosphatidylinositol signaling (KEGG pathway HSA04020 and HSA04070). Histone modification has been shown to be an important process for carcinogenesis (Esteller, 2007). HDAC1 are overexpressed in colorectal cancers (Ishihama et al., 2007) and HDAC inhibitors sensitize colorectal cancer cells to chemotherapy and radiotherapy (Chen et al., 2009; Flis et al., 2009). We showed that two histone modification enzymes HDAC4 and HDAC9 are regulated by SOX2. This finding is novel and may provide a link between histone modification and stem cell gene SOX2. However, additional experiments need to be carried out for confirm this observation.

The negative correlated genes include RUNX2 (Runt-related transcription factor 2) and IGF1R (Insulin-like growth factor 1 receptor). Wai et al. (2006) showed that murine *RUNX2* regulated transcription of a metastatic gene, osteopontin, in murine colorectal cancer cells. The insulin-like growth factor I receptor (IGF-I-R) plays a critical role in cell growth, transformation, and apoptosis. In many cancers including colorectal cancers, it is overexpressed and functions as an antiapoptotic agent by enhancing cell survival (Baserga et al., 1997; Donovan and Kummar, 2008; Hakam et al., 1999). The mechanism controlling IGF1R is complex. Multiple proteins including P53, WT1, P63, and P73 were shown to inhibit IGF1R expression (Nahor et al., 2005; Werner et al., 1996). Our data added SOX2 as a negative regulator of IGF1R in colorectal cancer cells. A network view of the above relationships is shown in Figure 4E.

Conclusion

We showed for the first time that SOX2 is overexpressed in CRC tissues compared with normal adjacent tissues, and that knockdown of SOX2 in colorectal cancer cells decreased their growth rates using both *in vitro* cell line and *in vivo* xenograft

Table 3. Common SOX2 Targets in CRC Cells and Mouse ES Cells Identified through the NCBI's Homologene Table

Homo sapiens	*GeneID*	*Mus musculus*	*GeneID*	*Description*
ABCA4	24	Abca4	11304	ATP-binding cassette, subfamily A (ABC1), member 4
ACTR2	10097	Actr2	66713	ARP2 actin-related protein 2 homolog (yeast)
ADCY5	111	Adcy5	224129	Adenylate cyclase 5
ADK	132	Adk	11534	Adenosine kinase
AFF3	3899	Aff3	16764	AF4/FMR2 family, member 3
AKAP7	9465	Akap7	432442	A kinase (PRKA) anchor protein 7
ALDH1L2	160428	Aldh1l2	216188	Aldehyde dehydrogenase 1 family, member L2
ALK	238	Alk	11682	Anaplastic lymphoma kinase (Ki-1)
ANGPT1	284	Angpt1	11600	Angiopoietin 1
ANKRD55	79722	Ankrd55	77318	Ankyrin repeat domain 55
ARHGAP10	79658	Arhgap10	78514	Rho GTPase activating protein 10
ARID2	196528	Arid2	77044	AT rich interactive domain 2 (ARID, RFX-like)
ASCC1	51008	Ascc1	69090	Activating signal cointegrator 1 complex subunit 1
BCMO1	53630	Bcmo1	63857	Beta-carotene 15,15′-monooxygenase 1
BTBD9	114781	Btbd9	224671	BTB (POZ) domain containing 9
C11orf61	79684	BC024479	235184	Chromosome 11 open reading frame 61
CBFA2T2	9139	Cbfa2t2	12396	Core-binding factor, runt domain, alpha subunit 2; translocated to, 2
CDH12	1010	Cdh12	215654	Cadherin 12, type 2 (N-cadherin 2)
CDH4	1002	Cdh4	12561	Cadherin 4, type 1, R-cadherin (retinal)
CDKAL1	54901	Cdkal1	68916	CDK5 regulatory subunit associated protein 1-like 1
CDS1	1040	Cds1	74596	CDP-diacylglycerol synthase (phosphatidate cytidylyltransferase) 1
CHD1L	9557	Chd1l	68058	Chromodomain helicase DNA binding protein 1-like
CHI3L1	1116	Chi3l1	12654	Chitinase 3-like 1 (cartilage glycoprotein-39)
CLYBL	171425	Clybl	69634	Citrate lyase beta like
CNTNAP2	26047	Cntnap2	66797	Contactin associated protein-like 2
CNTNAP4	85445	Cntnap4	170571	Contactin associated protein-like 4
COL18A1	80781	Col18a1	12822	Collagen, type XVIII, alpha 1
COL4A1	1282	Col4a1	12826	Collagen, type IV, alpha 1
CTNNA3	29119	Ctnna3	216033	Catenin (cadherin-associated protein), alpha 3
CTNNBL1	56259	Ctnnbl1	66642	Catenin, beta like 1
DENND1A	57706	Dennd1a	227801	DENN/MADD domain containing 1A
DEPDC6	64798	Depdc6	97998	DEP domain containing 6
DIAPH3	81624	Diap3	56419	Diaphanous homolog 3 (*Drosophila*)
DLGAP1	9229	Dlgap1	224997	Discs, large (*Drosophila*) homolog-associated protein 1
DNAJC15	29103	Dnajc15	66148	DnaJ (Hsp40) homolog, subfamily C, member 15
DOCK4	9732	Dock4	238130	Dedicator of cytokinesis 4
DOK5	55816	Dok5	76829	Docking protein 5
DTNA	1837	Dtna	13527	Dystrobrevin, alpha
DTNB	1838	Dtnb	13528	Dystrobrevin, beta
DTX4	23220	Dtx4	207521	Deltex 4 homolog (*Drosophila*)
ETV1	2115	Etv1	14009	Ets variant gene 1
EXT1	2131	Ext1	14042	Exostoses (multiple) 1
EYA4	2070	Eya4	14051	Eyes absent homolog 4 (*Drosophila*)
FBXL18	80028	Fbxl18	231863	F-box and leucine-rich repeat protein 18
FER	2241	Fert2	14158	Fer (fps/fes related) tyrosine kinase (phosphoprotein NCP94)
FRMD4A	55691	Frmd4a	209630	FERM domain containing 4A
GRHL2	79977	Grhl2	252973	Grainyhead-like 2 (*Drosophila*)
GRPEL2	134266	Grpel2	17714	GrpE-like 2, mitochondrial (*E. coli*)
HDAC9	9734	Hdac9	79221	Histone deacetylase 9
HNF4G	3174	Hnf4g	30942	Hepatocyte nuclear factor 4, gamma
HS2ST1	9653	Hs2st1	23908	Heparan sulfate 2-O-sulfotransferase 1
IGF1R	3480	Igf1r	16001	Insulin-like growth factor 1 receptor

Table 3. (Continued)

Homo sapiens	*GeneID*	*Mus musculus*	*GeneID*	*Description*
IMMP2L	83943	Immp2l	93757	IMP2 inner mitochondrial membrane peptidase-like (*S. cerevisiae*)
ITGA1	3672	Itga1	109700	
JRK	8629	Jrk	16469	Jerky homolog (mouse)
KIAA0182	23199	Gse1	382034	KIAA0182
KIAA1383	54627	4933403G14Rik	74393	KIAA1383
KLF12	11278	Klf12	16597	Kruppel-like factor 12
KRAS	3845	Kras	16653	V-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog
LAMA1	284217	Lama1	16772	Laminin, alpha 1
LAMA4	3910	Lama4	16775	Laminin, alpha 4
LOC647309	647309	Gm606	239789	Hypothetical LOC647309
LRP1B	53353	Lrp1b	94217	Low density lipoprotein-related protein 1B (deleted in tumors)
LSAMP	4045	Lsamp	268890	Limbic system-associated membrane protein
LYST	1130	Lyst	17101	Lysosomal trafficking regulator
MPP6	51678	Mpp6	56524	Membrane protein, palmitoylated 6 (MAGUK p55 subfamily member 6)
MSI2	124540	Msi2	76626	Musashi homolog 2 (*Drosophila*)
MT4	84560	Mt4	17752	Metallothionein 4
MTHFD1L	25902	Mthfd1l	270685	Methylenetetrahydrofolate dehydrogenase (NADP + dependent) 1-like
MYB	4602	Myb	17863	V-myb myeloblastosis viral oncogene homolog (avian)
MYO1E	4643	Myo1e	71602	Myosin IE
NRARP	441478	Nrarp	67122	Notch-regulated ankyrin repeat protein
NRP1	8829	Nrp1	18186	Neuropilin 1
PANX1	24145	Panx1	55991	Pannexin 1
PDE4D	5144	Pde4d	238871	Phosphodiesterase 4D, cAMP-specific (phosphodiesterase E3 dunce homolog, *Drosophila*)
PDIA4	9601	Pdia4	12304	Protein disulfide isomerase family A, member 4
PIK3R3	8503	Pik3r3	18710	Phosphoinositide-3-kinase, regulatory subunit 3 (p55, gamma)
PIM1	5292	Pim1	18712	Pim-1 oncogene
PLCH1	23007	Plch1	269437	Phospholipase C, eta 1
PLEKHA5	54477	Plekha5	109135	Pleckstrin homology domain containing, family A member 5
PRKRIR	5612	Prkrir	72981	Protein-kinase, interferon-inducible double stranded RNA dependent inhibitor, repressor of (P58 repressor)
PTPRK	5796	Ptprk	19272	Protein tyrosine phosphatase, receptor type, K
PTPRT	11122	Ptprt	19281	Protein tyrosine phosphatase, receptor type, T
RAD51L1	5890	Rad51l1	19363	RAD51-like 1 (S. cerevisiae)
RBMS1	5937	Rbms1	56878	RNA binding motif, single stranded interacting protein 1
REPS1	85021	Reps1	19707	RALBP1 associated Eps domain containing 1
RPS6KA2	6196	Rps6ka2	20112	Ribosomal protein S6 kinase, 90kDa, polypeptide 2
RTEL1	51750	Rtel1	269400	
RTTN	25914	Rttn	246102	Rotatin
RUNX1T1	862	Runx1t1	12395	Runt-related transcription factor 1; translocated to, 1 (cyclin D-related)
SCHIP1	29970	Schip1	30953	Schwannomin interacting protein 1
SERINC5	256987	Serinc5	218442	Serine incorporator 5
SETX	23064	Setx	269254	Senataxin
SHC4	399694	Shc4	271849	SHC (Src homology 2 domain containing) family, member 4
SLC23A2	9962	Slc23a2	54338	Solute carrier family 23 (nucleobase transporters), member 2
SLC24A3	57419	Slc24a3	94249	Solute carrier family 24 (sodium/potassium/ calcium exchanger), member 3
SLC35F1	222553	Slc35f1	215085	Solute carrier family 35, member F1

(*continued*)

Table 3. (Continued)

Homo sapiens	*GeneID*	*Mus musculus*	*GeneID*	*Description*
SLC35F3	148641	Slc35f3	210027	Solute carrier family 35, member F3
SNTG1	54212	Sntg1	71096	Syntrophin, gamma 1
SOX5	6660	Sox5	20678	SRY (sex determining region Y)-box 5
SOX6	55553	Sox6	20679	SRY (sex determining region Y)-box 6
SQLE	6713	Sqle	20775	Squalene epoxidase
STARD13	90627	Stard13	243362	StAR-related lipid transfer (START) domain containing 13
STK3	6788	Stk3	56274	Serine/threonine kinase 3 (STE20 homolog, yeast)
STK39	27347	Stk39	53416	Serine threonine kinase 39 (STE20/SPS1 homolog, yeast)
STX8	9482	Stx8	55943	Syntaxin 8
SYN3	8224	Syn3	27204	Synapsin III
TANC1	85461	Tanc1	66860	Tetratricopeptide repeat, ankyrin repeat and coiled-coil containing 1
TEAD1	7003	Tead1	21676	TEA domain family member 1 (SV40 transcriptional enhancer factor)
THSD4	79875	Thsd4	207596	Thrombospondin, type I, domain containing 4
TLE3	7090	Tle3	21887	Transducin-like enhancer of split 3 (E(sp1) homolog, *Drosophila*)
TMEM131	23505	Tmem131	56030	Transmembrane protein 131
TNFSF4	7292	Tnfsf4	22164	Tumor necrosis factor (ligand) superfamily, member 4 (tax-transcriptionally activated glycoprotein 1, 34 kDa)
TOM1L2	146691	Tom1l2	216810	Target of myb1-like 2 (chicken)
TRIM44	54765	Trim44	80985	Tripartite motif-containing 44
TTL	150465	Ttl	69737	Tubulin tyrosine ligase
TTLL11	158135	Ttll11	74410	Tubulin tyrosine ligase-like family, member 11
TTN	7273	Ttn	22138	Titin
UGT2B10	7365	Ugt2b34	100727	UDP glucuronosyltransferase 2 family, polypeptide B10
USP6NL	9712	Usp6nl	98910	USP6 N-terminal like
UTRN	7402	Utrn	22288	Utrophin
WT1	7490	Wt1	22431	Wilms tumor 1

models, suggesting its essential roles in the tumorigenesis of colorectal cancers. Applying next-generation sequencing coupled with Chromatin IP, we identified a consensus sequence wwTGywTT for SOX2 binding to promoter/enhancer regions of its regulated genes in colorectal cancers. An integrated expression profiling and ChIP-seq analysis show that SOX2 is involved in the BMP signaling pathway, steroid metabolic process, histone modifications, and many receptor-mediated signaling pathways such as IGF1R and ITPR2 (Inositol 1,4,5-triphosphate receptor, type 2).

Acknowledgments

This work was supported by grants 2006AA02A303, 2006AA02Z4A2, 2006DFA32950, 2007DFC30360, and 2004CB 518707 from the MOST, China.

Supplementary Materials

The supplementary tables for this manuscript are at the link http://systemsbiozju.org/data/Supplementary/

Author Disclosure Statement

The authors declare that no conflicting financial interests exist.

References

Bailey, T.L., Williams, N., Misleh, C., and Li, W.W. (2006). MEME: discovering and analyzing DNA and protein sequence motifs. Nucleic Acids Res 34, W369–W373.

Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., et al. (2007). High-resolution profiling of histone methylations in the human genome. Cell 129, 823–837.

Baserga, R., Hongo, A., Rubini, M., Prisco, M., and Valentinis, B. (1997). The IGF-I receptor in cell growth, transformation and apoptosis. Biochim Biophys Acta 1332, F105–F126.

Beckstette, M., Homann, R., Giegerich, R., and Kurtz, S. (2006). Fast index based algorithms and software for matching position specific scoring matrices. BMC Bioinformatics 7, 389.

Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V.B., et al. (2008a). Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. Cell 133, 1106–1117.

Chen, X., Wong, P., Radany, E., and Wong, J.Y. (2009). HDAC inhibitor, valproic acid, induces p53-dependent radiosensitization of colon cancer cells. Cancer Biother Radiopharm 24, 689–699.

Chen, Y., Shi, L., Zhang, L., Li, R., Liang, J., Yu, W., et al. (2008b). The molecular mechanism governing the oncogenic potential of SOX2 in breast cancer. J Biol Chem 283, 17969–17978.

Cho, S.J., Kim, J.S., Kim, J.M., Lee, J.Y., Jung, H.C., and Song, I.S. (2008). Simvastatin induces apoptosis in human colon cancer cells and in tumor xenografts, and attenuates colitis-associated colon cancer in mice. Int J Cancer 123, 951–957.

Dhawan, P., Singh, A.B., Deane, N.G., No, Y., Shiou, S.R., Schmidt, C., et al. (2005). Claudin-1 regulates cellular transformation and metastatic behavior in colon cancer. J Clin Invest 115, 1765–1776.

Donovan, E.A., and Kummar, S. (2008). Role of insulin-like growth factor-1R system in colorectal carcinogenesis. Crit Rev Oncol Hematol 66, 91–98.

Esteller, M. (2007). Cancer epigenomics: DNA methylomes and histone-modification maps. Nat Rev Genet 8, 286–298.

Flis, S., Gnyszka, A., and Splawinski, J. (2009). HDAC inhibitors, MS275 and SBHA, enhances cytotoxicity induced by oxaliplatin in the colorectal cancer cell lines. Biochem Biophys Res Commun 387, 336–341.

Foltz, G., Ryu, G.Y., Yoon, J.G., Nelson, T., Fahey, J., Frakes, A., et al. (2006). Genome-wide analysis of epigenetic silencing identifies BEX1 and BEX2 as candidate tumor suppressor genes in malignant glioma. Cancer Res 66, 6665–6674.

Giorgetti, A., Montserrat, N., Aasen, T., Gonzalez, F., Rodriguez-Piza, I., Vassena, R., et al. (2009). Generation of induced pluripotent stem cells from human cord blood using OCT4 and SOX2. Cell Stem Cell 5, 353–357.

Gure, A.O., Stockert, E., Scanlan, M.J., Keresztes, R.S., Jager, D., Altorki, N.K., et al. (2000). Serological identification of embryonic neural proteins as highly immunogenic tumor antigens in small cell lung cancer. Proc Natl Acad Sci USA 97, 4198–4203.

Hakam, A., Yeatman, T.J., Lu, L., Mora, L., Marcet, G., Nicosia, S.V., et al. (1999). Expression of insulin-like growth factor-1 receptor in human colorectal cancer. Hum Pathol 30, 1128–1133.

Halder, S.K., Rachakonda, G., Deane, N.G., and Datta, P.K. (2008). Smad7 induces hepatic metastasis in colorectal cancer. Br J Cancer 99, 957–965.

Hu, Q., Zhang, L., Wen, J., Wang, S., Li, M., Feng, R., et al. The EGF receptor-sox2-EGF receptor feedback loop positively regulates the self-renewal of neural precursor cells. Stem Cells 28, 279–286.

Hussenet, T., Dali, S., Exinger, J., Monga, B., Jost, B., Dembele, D., et al. (2010). SOX2 is an oncogene activated by recurrent 3q26.3 amplifications in human lung squamous cell carcinomas. PLoS One 5, e8960.

Ishihama, K., Yamakawa, M., Semba, S., Takeda, H., Kawata, S., Kimura, S., et al. (2007). Expression of HDAC1 and CBP/p300 in human colorectal carcinomas. J Clin Pathol 60, 1205–1210.

Ji, H., Jiang, H., Ma, W., Johnson, D.S., Myers, R.M., and Wong, W.H. (2008). An integrated software system for analyzing ChIP-chip and ChIP-seq data. Nat Biotechnol 26, 1293–1300.

Johnson, D.S., Mortaza Vi, A., Myers, R.M., and Wold, B. (2007). Genome-wide mapping of in vivo protein–DNA interactions. Science 316, 1497–1502.

Jothi, R., Cuddapah, S., Barski, A., Cui, K., and Zhao, K. (2008). Genome-wide identification of in vivo protein-DNA binding sites from ChIP-Seq data. Nucleic Acids Res 36, 5221–5231.

Kubens, B.S., and Zanker, K.S. (1998). Differences in the migration capacity of primary human colon carcinoma cells (SW480) and their lymph node metastatic derivatives (SW620). Cancer Lett 131, 55–64.

Lee, H.J., Chattopadhyay, S., Yoon, W.H., Bahk, J.Y., Kim, T.H., Kang, H.S., et al. (2010). Overexpression of hepatocyte nuclear factor-3alpha induces apoptosis through the upregulation and accumulation of cytoplasmic p53 in prostate cancer cells. Prostate 70, 353–361.

Li, X.L., Eishi, Y., Bai, Y.Q., Sakai, H., Akiyama, Y., Tani, M., et al. (2004). Expression of the SRY-related HMG box protein SOX2 in human gastric carcinoma. Int J Oncol 24, 257–263.

Lin, B., Wang, J., Hong, X., Yan, X., Hwang, D., Cho, J.H., et al. (2009). Integrated expression profiling and ChIP-seq analyses of the growth inhibition response program of the androgen receptor. PLoS One 4, e6589.

Nahor, I., Abramovitch, S., Engeland, K., and Werner, H. (2005). The p53-family members p63 and p73 inhibit insulin-like growth factor-I receptor gene expression in colon cancer cells. Growth Horm IGF Res 15, 388–396.

O'Dwyer, P.J., and Benson, A.B., 3rd. (2002). Epidermal growth factor receptor-targeted therapy in colorectal cancer. Semin Oncol 29, 10–17.

Oji, Y., Yamamoto, H., Nomura, M., Nakano, Y., Ikeba, A., Nakatsuka, S., et al. (2003). Overexpression of the Wilms' tumor gene WT1 in colorectal adenocarcinoma. Cancer Sci 94, 712–717.

Rodriguez-Pinilla, S.M., Sarrio, D., Moreno-Bueno, G., Rodriguez-Gil, Y., Martinez, M.A., Hernandez, L., et al. (2007). Sox2: a possible driver of the basal-like phenotype in sporadic breast cancer. Mod Pathol 20, 474–481.

Roth, F.P., Hughes, J.D., Estep, P.W., and Church, G.M. (1998). Finding DNA regulatory motifs within unaligned noncoding sequences clustered by whole-genome mRNA quantitation. Nat Biotechnol 16, 939–945.

Saigusa, S., Tanaka, K., Toiyama, Y., Yokoe, T., Okugawa, Y., Ioue, Y., et al. (2009). Correlation of CD133, OCT4, and SOX2 in rectal cancer and their association with distant recurrence after chemoradiotherapy. Ann Surg Oncol 16, 3488–3498.

Sanada, Y., Yoshida, K., Ohara, M., Oeda, M., Konishi, K., and Tsutani, Y. (2006). Histopathologic evaluation of stepwise progression of pancreatic carcinoma with immunohistochemical analysis of gastric epithelial transcription factor SOX2: comparison of expression patterns between invasive components and cancerous or nonneoplastic intraductal components. Pancreas 32, 164–170.

Sandelin, A., Alkema, W., Engstrom, P., Wasserman, W.W., and Lenhard, B. (2004). JASPAR: an open-access database for eukaryotic transcription factor binding profiles. Nucleic Acids Res 32, D91–D94.

Schepers, G.E., Teasdale, R.D., and Koopman, P. (2002). Twenty pairs of sox: extent, homology, and nomenclature of the mouse and human sox transcription factor gene families. Dev Cell 3, 167–170.

Sekharam, M., Zhao, H., Sun, M., Fang, Q., Zhang, Q., Yuan, Z., et al. (2003). Insulin-like growth factor 1 receptor enhances invasion and induces resistance to apoptosis of colon cancer cells through the Akt/Bcl-x(L) pathway. Cancer Res 63, 7708–7716.

Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell 126, 663–676.

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., et al. (2007). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell 131, 861–872.

Thijs, G., Marchal, K., Lescot, M., Rombauts, S., De Moor, B., Rouze, P., et al. (2002a). A Gibbs sampling method to detect overrepresented motifs in the upstream regions of coexpressed genes. J Comput Biol 9, 447–464.

Thijs, G., Moreau, Y., De Smet, F., Mathys, J., Lescot, M., Rombauts, S., et al. (2002b). INCLUSive: integrated clustering, upstream sequence retrieval and motif sampling. Bioinformatics 18, 331–332.

Tseng, Y.S., Lee, J.C., Huang, C.Y., and Liu, H.S. (2009). Aurora-A overexpression enhances cell-aggregation of Ha-ras transformants through the MEK/ERK signaling pathway. BMC Cancer 9, 435.

Wai, P.Y., Mi, Z., Gao, C., Guo, H., Marroquin, C., and Kuo, P.C. (2006). Ets-1 and runx2 regulate transcription of a metastatic gene, osteopontin, in murine colorectal cancer cells. J Biol Chem 281, 18973–18982.

Wang, Q., He, W., Lu, C., Wang, Z., Wang, J., Giercksky, K.E., et al. (2009a). Oct3/4 and Sox2 are significantly associated with an unfavorable clinical outcome in human esophageal squamous cell carcinoma. Anticancer Res 29, 1233–1241.

Wang, W., Zhang, P., and Liu, X. (2009b). Short read DNA fragment anchoring algorithm. BMC Bioinformatics 10(Suppl 1), S17.

Werner, H., Karnieli, E., Rauscher, F.J., and Leroith, D. (1996). Wild-type and mutant p53 differentially regulate transcription of the insulin-like growth factor I receptor gene. Proc Natl Acad Sci USA 93, 8318–8323.

Wiesner, G.L., Daley, D., Lewis, S., Ticknor, C., Platzer, P., Lutterbaugh, J., et al. (2003). A subset of familial colorectal neoplasia kindreds linked to chromosome 9q22.2–31.2. Proc Natl Acad Sci USA 100, 12961–12965.

Yu, J., Vodyanik, M.A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J.L., Tian, S., et al. (2007). Induced pluripotent stem cell lines derived from human somatic cells. Science 318, 1917–1920.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., et al. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol 9, R137.

Address correspondence to:
Dr. Biaoyang Lin
Systems Biology Division
Zhejiang-California International
Nanosystems Institute (ZCNI)
Zhejiang University
268 Kaixuan Road
Hangzhou, Zhejiang, 310029, P.R. China

E-mail: bylin@u.washington.edu

or

Prof. Shu Zheng
The Second Affiliated Hospital
Zhejiang University
Hangzhou, Zhejiang, P.R. China

E-mail: zhengshu@zju.edu.cn