SOX15 Governs Transcription in Human Stratified Epithelia and a Subset of Esophageal Adenocarcinomas

Rita Sulahian,1,2 Justina Chen,1 Zoltan Arany,3 Unmesh Jadhav,1,2 Shouyong Peng,1 Anil K. Rustgi,1 Adam J. Bass,1,2 Amitabh Srivastava,5 Jason L. Hornick,5 and Ramesh A. Shivdasani1,2

1Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, Massachusetts; 2Department of Medicine, Harvard Medical School, Boston, Massachusetts; 3Cardiovascular Institute and 4Division of Gastroenterology, Department of Pathology, Brigham & Women’s Hospital, and Department of Pathology, Harvard Medical School, Harvard University, Boston, Massachusetts

SUMMARY
This study identifies SOX15 as a direct transcriptional regulator of a substantial fraction of cell type-specific genes in stratified epithelial cells. SOX15 expression is attenuated in intestinal metaplasia (Barrett’s esophagus) but is active in many esophageal adenocarcinomas.

BACKGROUND & AIMS: Intestinal metaplasia (Barrett’s esophagus, BE) is the principal risk factor for esophageal adenocarcinoma (EAC). Study of the basis for BE has centered on intestinal factors, but loss of esophageal identity likely also reflects the absence of key squamous-cell factors. As few determinants of stratified epithelial cell-specific gene expression have been characterized, identifying the necessary transcription factors is important.

METHODS: We tested regional expression of mRNAs for all putative DNA-binding proteins in the mouse digestive tract and verified the esophageal-specific factors in human tissues and cell lines. Integration of diverse data defined a human squamous esophageal-specific transcriptome. We used chromatin immunoprecipitation with high-throughput sequencing (ChIP-seq) to locate transcription factor binding sites, computational approaches to profile the transcripts in cancer data sets, and immunohistochemistry to reveal protein expression.

RESULTS: The transcription factor Sex-determining region Y-box 15 (SOX15) is restricted to esophageal and other murine and human stratified epithelia. SOX15 mRNA levels are attenuated in BE, and its deletion in human esophageal cells reduces esophageal transcripts significantly and specifically. SOX15 binding is highly enriched near esophagus-expressed genes, indicating direct transcriptional control. SOX15 and hundreds of genes coexpressed in squamous cells are reactivated in up to 30% of EAC specimens. Genes normally confined to the esophagus or intestine appear in different cells within the same malignant glands.

CONCLUSIONS: These data identify a novel transcriptional regulator of stratified epithelial cells and a subtype of EAC with bi-lineage gene expression. Broad activation of squamous-cell genes may shed light on whether EACs arise in the native stratified epithelium or in ectopic columnar cells. (Cell Mol

BACKGROUND & AIMS: Intestinal metaplasia (Barrett’s esophagus, BE) is a common, chronic condition in which an epithelium containing intestinal goblet and other columnar cells replaces the native stratified squamous mucosa.1 BE results from chronic acid and bile reflux. Over time, the metaplastic tissue may become dysplastic, and it progresses to invasive cancer in three to five cases per 1000 person-years.2 Esophageal adenocarcinoma (EAC) arises principally in the setting of BE, and the incidence of this cancer in the West increased about eightfold between 1970 and 2010, with about 18,000 new U.S. cases and 15,000 deaths expected in 2015 (http://seer.cancer.gov).

Investigation into the mechanisms of BE has centered largely on determinants of intestinal identity,3 particularly the intestine-restricted transcription factors (TFs) Caudal type homeobox 1 (CDX1) and CDX2, which specify the embryonic intestine.4 Forced expression of CDX2 or CDX1 in the mouse stomach induces ectopic intestinal differentiation,5,6 and both factors are implicated in activating intestinal genes in BE.7,8 though forced CDX2 expression in the

| Abbreviations used in this paper: BE, Barrett’s esophagus; CDX1/2, caudal type homeobox 1/2; ChIP, chromatin immunoprecipitation; ChIP-seq, chromatin immunoprecipitation with high-throughput sequencing; EAC, esophageal adenocarcinoma; G-E, gastroesophageal; KRT5, keratin 5, type II; KRT6A, keratin 6A, type II; PAX9, paired box 9; PBS, phosphate-buffered saline; qRT-PCR, quantitative reverse-transcription polymerase chain reaction; shRNA, small hairpin RNA; SIM2, single-minded family bHLH transcription factor 2; SOX2, 15, sex-determining region Y-box 2, -box 15; TCGA, The Cancer Genome Atlas; TF, transcription factor; TP63, tumor protein P63; TRIM29, tripartite motif containing 29. |

© 2015 The Authors. Published by Elsevier Inc. on behalf of the AGA Institute. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
mouse esophagus does not induce BE per se.\textsuperscript{9} Loss of esophagus-specific transcripts and of stratified squamous morphology probably reflects parallel loss of transcriptional determinants of the native epithelium, which are largely unknown. Tumor protein P63 (TP63) regulates differentiation of all stratified epithelia, such as those in the esophagus and skin,\textsuperscript{10,11} acting in part through another transcription factor, basonuclin 1 (BNC1).\textsuperscript{12} Sex-determining region Y-box 15 (SOX15) as such a TF and we show that it directly participate.\textsuperscript{16} Forkhead box A2 (FOXA2) is expressed in embryonic but not in adult esophageal cells.\textsuperscript{17} We sought to identify other tissue-restricted TFs that might control the characteristic stratified epithelium.

Among all putative DNA-binding proteins, we searched first for those with esophagus-restricted expression among digestive epithelia and then for factors with attenuated expression in BE. We identified sex-determining region Y-box 15 (SOX15) as such a TF and we show that it directly controls transcription of a large fraction of human esophagus-expressed genes. SOX15 is absent from most EACs, but up to 30\% of cases retain expression of SOX15 and its target genes, coexpressing representative intestinal and squamous-specific genes within the same tissue. Together, these data identify a novel regulator of stratified epithelial genes and a subtype of EAC with bi-lineage gene expression.

**Materials and Methods**

**Tissue Preparation and Transcription Factor Expression Screen**

We isolated epithelial sheets from the esophagus, gastric corpus-antrum, and duodenum of 1-month old CD1 and C57BL/6 mice. Before peeling the mucosa using fine forceps, the esophagus was treated with 0.1\% collagenase-dispase (cat. no. 11097113001; Roche Applied Science, Indianapolis, IN) in phosphate-buffered saline (PBS) for 15 minutes at 37°C, whereas stomach and duodenum were incubated in 1 mM ethylenediaminetetraacetic acid (EDTA) in PBS at 37°C. To determine the relative transcript levels (Figure 1A–C), we used quantitative reverse-transcription polymerase chain reaction (qRT-PCR) screen. Expression of 1880 TF mRNAs was assessed in epithelial cell isolates from adult CD1 mouse esophagus (red), stomach and intestine (blue), and then for factors with attenuated expression in BE. We identified sex-determining region Y-box 15 (SOX15) as such a TF and we show that it directly controls transcription of a large fraction of human esophagus-expressed genes. SOX15 is absent from most EACs, but up to 30\% of cases retain expression of SOX15 and its target genes, coexpressing representative intestinal and squamous-specific genes within the same tissue. Together, these data identify a novel regulator of stratified epithelial genes and a subtype of EAC with bi-lineage gene expression.

**Figure 1. Differential transcription factor (TF) expression in the normal mouse gut and other tissues.** (A) Distribution of all TFs in wild-type mouse digestive epithelia, as revealed in a quantitative reverse-transcription polymerase chain reaction (qRT-PCR) screen. Expression of 1880 TF mRNAs was assessed in epithelial cell isolates from adult CD1 mouse esophagus (red), stomach and intestine (blue), and then for factors with attenuated expression in BE. We identified sex-determining region Y-box 15 (SOX15) as such a TF and we show that it directly controls transcription of a large fraction of human esophagus-expressed genes. SOX15 is absent from most EACs, but up to 30\% of cases retain expression of SOX15 and its target genes, coexpressing representative intestinal and squamous-specific genes within the same tissue. Together, these data identify a novel regulator of stratified epithelial genes and a subtype of EAC with bi-lineage gene expression.
polymerase chain reaction (qRT-PCR) and a library containing oligonucleotide primers specific to 1880 known and putative TFs. Tissue-specific TFs were identified using the comparative CT method. To further determine the tissue specificity (Figure 1D), other whole organs were harvested from adult C57BL/6 mice.

Cell Lines
We cultured CP-A (KR-42421), CP-B (CP-52731), and CP-C (CP-94251) cells (American Type Culture Collection, Manassas, VA) in MCDB-153 medium (cat. no. M7403; Sigma-Aldrich, St. Louis, MO) supplemented with 0.4 μg/mL hydrocortisone, 20 ng/mL recombinant human epidermal growth factor (cat. no. E9644; Sigma-Aldrich), 8.4 μg/L cholera toxin (cat. no. H0135; Sigma-Aldrich), 20 mg/L adenine (cat. no. A2786; Sigma-Aldrich), 140 μg/mL bovine pituitary extract (cat. no. P1476; Sigma-Aldrich), ITS supplement (cat. no. I1884; Sigma-Aldrich) (final concentrations: 5 μg/mL insulin, 5 μg/mL transferrin, 5 ng/mL sodium selenite), 4 mM glutamine, and 5% fetal bovine serum. EPC2-hTert cells were cultured in Keratinocyte-generated keratinocyte growth supplement. Each cell line was used to create three different cell lines, each of which was subcloned to obtain a homogeneous population of cells.

Figure 2. Differential transcription factor (TF) expression in normal and metaplastic human esophagus. (A) Expression profiles of PAX9, SIM2, SOX15, and TRIM29 in 65 human organs. Data from necropsies analyzed using Oncomine tools showed selective expression in the esophagus and other stratified epithelia such as the oropharyngeal mucosa and skin derivatives. (B) Relative expression of esophagus-active keratin genes in the human Barrett’s esophagus (BE) cell line series (CP-A, CP-B, and CP-C) with increasing dysplasia. Results of quantitative reverse-transcription polymerase chain reaction analysis are represented with respect to transcript levels in the immortalized human esophageal cell line EPC2-hTert. (C) Relative expression of esophagus-specific TF mRNAs in human BE cell lines CP-A, CP-B and CP-C, expressed in relation to levels in EPC2-hTert cells. (D) Fold-enriched expression of esophagus-specific TF mRNAs in fresh human esophageal epithelial biopsy samples relative to areas of BE in the same patients. (E) Expression of esophagus-specific TFs and intestine-specific genes in normal human esophagus and BE resection specimens. Data used were analyzed using Oncomine tools.
SFM medium (GIBCO/Life Technologies, Grand Island, NY) supplemented with bovine pituitary extract and recombinant human epidermal growth factor (GIBCO). Soybean trypsin inhibitor (Sigma-Aldrich) was used to quench trypsin activity during cell passage.

**Gene Analyses**

Figures 2A, D, and E and selected other figures show analyses of relative mRNA expression levels from published studies of 65 adult human tissues, of human esophageal biopsy specimens, and of human normal esophagus, BE,
and EAC samples. The data were reanalyzed with respect to SOX15 using Oncomine tools (Compendia Bioscience, Ann Arbor, MI; www.oncomine.com), considering all samples in each data set. Genes significantly associated with SOX15 were ranked on the basis of correlation values. Enriched Gene Ontology terms were determined using DAVID tools (http://david.abcc.ncifcrf.gov/).

We examined processed RNA-seq data from a Cancer Genome Atlas (TCGA) study on stomach cancer, first isolating 30 GIN tumors arising at the gastroesophageal (G-E) junction or gastric cardia for unsupervised clustering (Supplementary Figure 1). To this group we applied hierarchical clustering (using hclust from the R package; http://cran.r-project.org) on the 1000 most variable transcripts normalized according to expression z-scores, followed by a second hierarchical clustering on the set of 317 genes coexpressed with SOX15. To assess the specificity of SOX15 overexpression in tumors of the gastric cardia, we compared with RNA-seq data from TCGA studies on colon and distal gastric adenocarcinomas.

**Experimental RNA Analyses**

Total RNA was isolated using TRIzol (Invitrogen/Life Technologies, Carlsbad, CA), treated with the RNeasy Mini Kit (Qiagen, Valencia, CA), and DNA was digested using Turbo DNA-Free (Ambion/Life Technologies, Austin, TX). For qRT-PCR analysis (Figure 1A–C, Figure 2B and C, etc. 1 μg of total RNA was reverse-transcribed with Superscript III First Strand Synthesis System (Invitrogen), and cDNA was amplified using SYBRGreen PCR Master Mix (Applied Biosystems, Foster City, CA). RNA-seq libraries (the full data set is deposited in the Gene Expression Omnibus with accession number GSE62909) were prepared from 300 ng of total RNA using TruSeq RNA Sample Preparation kits (Illumina, San Diego, CA), and 75-base pair (bp) single-end sequences were obtained using a NextSeq 500 instrument (Illumina). The reads were aligned to human genome build Hg19 using TopHat v2.0.6 (https://ccb.jhu.edu/software/tophat/index.shtml). Expression levels of transcripts in duplicate samples were calculated as fragments per kb per 106 mapped reads (FPKM) using Cufflinks v2.0.2 (http://cole-trapnell-lab.github.io/cufflinks/), and differential expression was determined using CuffDiff (http://cole-trapnell-lab.github.io/cuffdiff/). Chi-square tests with 1 degree of freedom and two-tailed P values were used to assess significance. Log2 (FPKM + 1) values for control and SOX15-depleted samples were plotted to display differential expression.

**Depletion of SOX15 and Expression of Biotin-Tagged SOX15**

The cells were infected with lentiviruses generated from the pLKO.1 vector (Open Biosystems/GE Dharmacon, Huntsville, AL) carrying either a SOX15-targeting small hairpin RNA (shRNA) (TGCCCTGGCACATTGGCTCTT) or a control, nonspecific shRNA that does not complement any human gene and is not toxic to cultured human cells (CTTAAGGTATAGCCTGCGGG). Human SOX15 cDNA was cloned into the pLJLa vector (cat. no. 24129; Addgene, Cambridge, MA) together with cassettes for the T2A sequence, biotin, and BirA-V5 (gift of Ben Ebert, Brigham & Women’s Hospital, Boston, MA).

**Chromatin Immunoprecipitation and Chromatin Immunoprecipitation With High-Throughput Sequencing**

Cells were cross-linked with 2 mM disuccinimidyl glutarate (catalog no. 20593; Pierce Biotechnology, Rockford, IL) in PBS for 45 minutes, followed by 10 minutes with 1% formaldehyde (Pierce Biotechnology) in PBS at room temperature. Chromatin immunoprecipitation (ChIP) and chromatin immunoprecipitation with high-throughput sequencing (ChIP-seq) were performed as described previously elsewhere using a 30-μL slurry of streptavidin-conjugated magnetic beads (cat. no. 65601; Invitrogen). We used Cistrome tools (www.cistrome.org) to identify and annotate TF binding sites, generate wiggle files and conservation plots, identify enriched sequence motifs and linked genes, and compare data across ChIP-seq libraries. Wiggle traces were projected on the Integrative Genome Viewer (www.broadinstitute.org/igv/).

Functions of genes within 50 kb of SOX15 occupancy were determined using GREAT (http://omictools.com/great-s1664.html). ChIP-seq data are deposited in the Gene Expression Omnibus (GEO) database with accession number GSE62909 (www.ncbi.nlm.nih.gov/geo/).
Immunohistochemistry

We baked 4-μm-thick tissue paraffin sections overnight at 37°C; they then were deparaffinized in xylene, rehydrated, and peroxidase activity was blocked with 1.5% H2O2 in methanol for 10 minutes. Slides were treated with 0.01 M citrate buffer, pH 6.0, in a pressure cooker at 120°C for 30 minutes for antigen retrieval, then transferred to Tris-buffered saline. Sections were first incubated with mouse CDX2 Ab (clone CDX2-88, Biogenex mu392A-uc, 1:200) for 40 minutes, followed by Dako Envision+ Mouse (Dako K4007; Dako, Carpinteria, CA) secondary Ab for 30 minutes, and developed with 3,3′-diaminobenzidine (Dako). Sections were
then incubated with mouse KRT5 Ab (clone XM26, Leica NCL-L-CK, 1:500) for 40 minutes, followed by PowerVision AP mouse (catalog no. PV6110; Leica Biosystems, Buffalo Grove, IL) secondary Ab for 30 minutes, developed with Permanent Red, and counterstained with Mayer’s hematoxylin. To stain resection specimens that carried areas of BE, slides were treated with the same mouse KRT5 Ab, followed by Dako Envision+ Mouse (Dako K4007) secondary Ab for 30 minutes, and developed with 3,3’-diaminobenzidine (Dako).

Results
Identification of Transcription Factors That Are Specific to the Esophageal Epithelium and Attenuated in Barrett’s Esophagus

To identify candidate regulators of esophageal squamous identity, we first examined epithelia isolated from different regions of the mouse alimentary tract—esophagus, glandular stomach, and intestine (duodenum)—with a goal to identify TF mRNAs expressed selectively in the stratified esophageal epithelium (Figure 1A). Among 1880 known and putative DNA-binding proteins, those showing ≥32-fold higher expression in the intestinal mucosa included the known intestinal factors Atoh1, Cdx1, Creb313, Hnf4g, and Isx, underscoring the fidelity of the experimental approach (Figure 1B). Forty factors and 59 TF genes showed considerably higher expression in esophageal cells, compared with the gastric corpus and the intestine, respectively (Figure 1A), and 21 TFs were common to the two esophagus-specific groups (Figure 1B).

To exclude variability among mouse strains and to assess specificity relative to nondigestive organs, we measured expression of these 21 mRNAs in nine diverse tissues from C57BL/6 mice, including the skin. Six TFs gave consistent evidence of high tissue specificity (Figure 1C and D). Sim2 (single-minded family bHLH transcription factor 2) and Pax9 (paired box 9) showed the greatest specificity, followed by Sox15 and Trim29 (tripartite motif containing 29), which showed some expression in murine skin. Additional data from 65 adult human tissues revealed robust expression of each of these four TF mRNAs in the esophagus, with varying levels in other stratified squamous...
tissues, such as the tongue, mouth, pharynx, and skin derivatives (Figure 2A).

To determine whether these TFs may function in the identity of stratified epithelia, we examined expression data from immortalized EPC2-hTert esophageal keratinocytes and found high expression of each factor except ELF5 (data not shown). Next we tested a series of three cell lines: CP-A, which represents nondysplastic BE, and CP-B and CP-C, which represent BE with high-grade dysplasia. This cell line series replicates disease progression with reduced levels of multiple keratin mRNAs (Figure 2B).

We observed a concomitant decline in SOX15 and TRIM29 levels, matching or exceeding that of TP63 mRNA, with little variance in the other factors (Figure 2C). Although these findings do not in isolation give robust information about a relation to mucosal dysplasia per se, they reveal the squamous cell specificity of SOX15 and TRIM29. Furthermore, gene expression data from a collection of human esophageal biopsy specimens showed significantly fewer SOX15 and TRIM29 mRNAs in primary BE compared with adjacent normal esophageal mucosa (Figure 2D).

Finally, we used Oncomine tools to analyze mRNA data from an independent series of 28 frozen human normal esophagus and 15 frozen BE biopsy specimens. Levels of PAX9, SOX15, and TRIM29 were uniformly high in normal esophagus and attenuated in BE specimens (Figure 2E).

Together, these data identify SOX15, PAX9, and TRIM29 as conserved candidate determinants of squamous cell identity. PAX9 levels were similar in CP-A, CP-B and CP-C cells (Figure 2C), and although TRIM29 has a putative DNA-binding domain, its role in transcriptional regulation is poorly defined and uncertain. By contrast, SOX proteins control differentiation of diverse tissues, often in conjunction with other family members, and related factors such as SOX2 and SOX7 are known to regulate aspects of esophageal organogenesis and squamous cell cancer.

We therefore concentrated on human SOX15, which shares 85% homology (100% in the DNA-binding domain) with the mouse protein. SOX15 was previously noted as one among hundreds of genes in various expression profiling studies and we proceeded to investigate its functions.

SOX15 Depletion Affects Genes Specific to Stratified Epithelium

To test whether SOX15 might regulate genes specific to the stratified human esophageal epithelium, we needed to delineate the corresponding transcriptome. To this end, first we considered public data from adult human postmortem tissues (see Figure 2A) and identified 362 genes that express at greater than threefold higher levels (P < .05) in the esophagus than in any of seven diverse tissues, including glandular stomach, from the same collection of postmortem samples (Figure 3A). Second, we identified 300 genes with greater than threefold higher mRNA levels (P < .05) in normal fresh human esophageal biopsy specimens than in adjacent areas of BE or in fresh intestinal biopsies from the same study. Consistent with specific roles in stratified epithelia, both gene sets were highly enriched for functions...
related to ectodermal, epidermal, and keratinocyte differentiation, and they shared 114 genes (Figure 3A). Accordingly, we regard the union set of 548 genes as a good representation of human esophagus-specific transcripts and the intersection set of 114 genes as an especially robust subset.

To determine whether SOX15 regulates any part of this transcriptome, we used lentiviral delivered shRNA to deplete the TF in CP-A cells. These cells express keratin and TF genes specific to stratified epithelia, including SOX15, at levels similar to immortalized EPC2-hTert esophageal epithelial cells (Figure 2B and C), and they tolerate lentiviral infection and drug selection. Because SOX15 depletion retarded CP-A cell growth and survival, we harvested cells 72 hours after infection, when they appeared healthy but the SOX15 mRNA levels were appreciably reduced (Figure 3B).

RNA-seq analysis showed reduced and increased levels of 2950 and 717 transcripts, respectively, compared with cells treated with a nonspecific shRNA (Figure 3C). In agreement with the deficit in cell growth, these genes were enriched for Gene Ontology terms related to the cell cycle (Supplementary Table 1). More importantly, genes reduced in SOX15-depleted cells included 26.4% of the human esophagus-specific “union” transcriptome, compared with 15.34% overlap with multiple sets of 2950 random genes expressed in CP-A cells (Figure 3D; P < .0001). Correspondence was even higher for the esophagus-specific “intersection” transcriptome, where 31.5% of genes were reduced in SOX15-depleted cells compared with 14.68% of random genes (P < .0001). SOX15-dependent genes were highly enriched for functions related to stratified epithelia (Figure 3D). None of the 114 genes in the esophagus-specific “intersection set,” and only 23 genes in the “union set” were increased in SOX15-depleted CP-A cells, and we observed no increase in intestinal genes. Rather, the 717 increased transcripts were enriched for functions such as apoptosis and vesicular transport (Supplementary Table 1). Thus, beyond cell survival or proliferation, a substantial portion of the esophageal transcriptome depends on SOX15.

SOX15 Directly Regulates Esophagus-Specific Genes

Depletion of SOX15 could affect transcript levels as a consequence of its cis-regulatory activity or indirectly. To determine whether SOX15 might regulate dependent genes directly, we used ChIP-seq to map its cisrome. Because available antibodies performed poorly in ChIP assays, we expressed biotin-tagged SOX15 stably in CP-A cells and precipitated chromatin using streptavidin beads. The nearly 5000 high-confidence binding sites we identified by this approach showed high sequence conservation and greatest enrichment for the SOX consensus motif, which was present in >97% of sites, implying direct TF occupancy (Figure 4A). Similar to other tissue-specific TFs, SOX15 occupied few promoters (6.2% of all binding sites) and bound DNA predominantly in intergenic regions and introns (Figure 4A).

GREAT analysis of the nearest flanking genes within 50 kb of SOX15 occupancy revealed enrichment of pathways known to be vital in stratified epithelia, such as epidermal growth factor and Rho/Rac signaling, and in cell survival (Figure 4B). Moreover, 20.9% of genes in the human esophagus-specific transcriptome and 31.5% of genes common to the two esophagus transcript sources showed at least one SOX15-binding site within 50 kb of the transcription start site, compared with about 5% of comparable numbers of random genes (P < .0001, Figure 4C).

Gene Ontology terms related to stratified epithelia were further enriched among SOX15-bound genes (Figure 4C). Most importantly, genes affected by SOX15 depletion in CP-A cells were highly enriched for nearby SOX15 binding, compared with random gene sets of equal size (P < .0001), and SOX15-bound genes reduced in SOX15-depleted cells far outnumbered genes that were increased (Figure 4D and E).

Taken together, these data indicate direct SOX15 regulation of genes specific to the stratified squamous epithelium, with a strong bias toward gene activation. Canonical esophageal genes such as KRT6A (keratin 6A, type II) illustrate SOX15 occupancy at putative cis-regulatory sites and reduced expression in SOX15-depleted cells (Figure 4F).

SOX15 in Human Esophageal Adenocarcinoma

SOX15 is expressed highly in normal human esophagus, but not in the BE cell lines CP-B and CP-C (Figure 2C) or in areas of intestinal metaplasia in vivo (Figure 2D). To our surprise, RNA expression data from a large collection of frozen primary esophagus, BE, and EAC biopsy specimens revealed high SOX15 mRNA levels in up to one-third of EACs (Figure 3A, left; note that all samples in this study, including EAC, were frozen biopsy specimens). Moreover, at least 317 transcripts that are strongly coexpressed with SOX15 in the normal esophagus (r > 0.81; Supplementary Table 2) were also present in the same EAC specimens (Figure 5A, which shows the 100 genes with highest correlation), suggesting broad activation of the squamous cell transcriptional program.

Accordingly, functions related to stratified epithelia were significantly enriched among the genes coexpressed with SOX15 (Figure 5B). The canonical intestinal marker CDX2 and its coexpressed genes (r > 0.81) were expressed in many SOX15+ and also in SOX15− specimens (Figure 5A, right), revealing coexpression of esophageal and intestinal genes in some cases. Moreover, 21.6% of genes coexpressed with SOX15 in this analysis showed SOX15 binding within 50 kb in CP-A cells, compared with ~6% of random genes (P < .0001; Figure 5C, top), which implies that many of these genes are direct transcriptional targets. Indeed, the effects of SOX15 depletion were statistically significantly greater on these genes than on random sets of genes expressed in CP-A cells (P < .017; Figure 5C, bottom). These features collectively suggest direct SOX15 regulation of many esophagus-restricted genes that are silent in BE and reactivated in up to one-third of human EACs.

To exclude the possibility that EACs expressing SOX15 were simply contamined with normal SOX15+ esophageal cells, we studied cases from an independent collection, the Cancer Genome Atlas (TCGA), where nonmalignant cells were meticulously minimized. Cancers of the G-E junction
typically arise in a background of BE and, when associated with chromosomal instability (CIN), usually represent distal EACs. Among 30 cases of CIN+ tumors from the G-E junction or gastric cardia in the TCGA collection of gastric cancers, some samples showed robust levels of SOX15 and of genes coexpressed with SOX15 in normal esophageal epithelium (Supplementary Figure 1A and B). Transcripts specific to the squamous esophageal epithelium were thus again evident in a fraction of EACs. To determine whether this extent of SOX15 expression is specific to EACs, we evaluated other gastrointestinal cancers in the TCGA collection: gastric fundus, body, and antrum, and colorectal tumors. Extreme outliers for SOX15 mRNA expression were present only among tumors of the gastric cardia (Figure 5D).

Finally, we examined 99 separate EACs by immunohistochemistry on resection specimens. Because several antibodies failed to detect SOX15, we used KRT5 (keratin 5, type II) as a proxy for expression of SOX15 and other stratified cell-specific products, noting nearly total concordance of SOX15 and KRT5 mRNA expression in the large aforementioned tissue collection 42 (Figure 6A; \( r = 0.97 \)). We also stained the same samples for the intestinal marker CDX2. Nineteen cases (19%) showed cytoplasmic KRT5 expression within malignant glands, and most of these cases coexpressed nuclear CDX2 (Figure 6B). Levels of KRT5 were variable (Figure 6B) but did not correlate with tumor grade or other pathologic features such as mucin production. Importantly, KRT5 was not expressed in rare pockets of squamous differentiation but rather in bona fide glandular structures. In fact, and of particular note, cytoplasmic KRT5 and nuclear CDX2 were almost always present in different cells within the same glands (see Figures 6C and D for examples from two different cases). Coexpression of esophagus- and intestine-specific genes within individual glands reveals the malignant cells’ potential to express genes from distinct cell lineages.

To corroborate the observation that mRNA levels of stratified cell genes are low in BE but elevated in many EACs (Figures 5A and 6A), we used immunohistochemistry to assess areas of BE that were present in 24 of the 99 resection specimens. KRT5 was uniformly absent from these areas, and although disease heterogeneity is apparent in the clinic, the underlying determinants are unclear. We show here that one-fifth to one-third of EACs simultaneously express products specific to the esophageal squamous epithelium and columnar intestinal cells. It will be important in the future to identify the clinical and genetic correlates of these EACs showing bi-lineage gene expression and to determine whether they reflect a distinctive pathophysiology or harbor unique therapeutic vulnerabilities.

**Discussion**

**Implications for Esophageal Squamous Differentiation and Esophageal Adenocarcinoma**

Insights into transcriptional control of the esophageal squamous epithelium are largely limited to the broad functions of TP63 and SOX2. 10,11 Our identification of SOX15 as a novel, conserved, and likely direct regulator of many human stratified epithelial genes extends our understanding of esophageal differentiation and pathology. The lack of overt esophageal defects in Sox15 mutant mice 31,42 is compatible with the considerable known redundancies among SOX-family TFs. 33

There is much debate whether BE and particularly EAC originate in the native esophageal epithelium through bona fide metaplasia or in ectopic cells that may colonize the esophagus from the gastric cardia, as in mice. 13,44 Clearly, the best way to answer the question is through lineage-tracing studies, which are possible in animals but not in humans. When lineage tracing is not feasible, cell-specific transcript patterns offer clues, and the expression of SOX15 and other squamous epithelial genes may be informative. Consider, for example, the observation that the BE cell lines CP-A/B/C show reduced levels of SOX15, esophageal keratins, TP63, and other esophage-specific TF genes, implying loss of an esophageal program, but some esophagus-restricted TF genes such as SIM2 and TFPAR2 (transcription factor activating enhancer binding protein 2α) are highly expressed in these cells (Figure 2C) and in BE biopsy specimens (Figure 2D and E). Moreover, >300 esophageal genes, including SOX15, are active in up to 30% of human EACs (Figure 3A), with intestinal genes such as CDX2 often coexpressed in the same glands as esophageal genes (Figure 4C and D). These findings in BE and EAC could indicate residual squamous cell-specific transcription or fortuitous ectopic gene activity. If diseased cells are better equipped to express genes from their native transcriptional program than are genes from a heterologous cell lineage, because native genes and their cis-regulatory elements are inherently primed and accessible, then the first possibility may be more likely. Our findings do not of course rule out the alternative model, which will require additional, independent lines of evidence.

EAC is a particularly recalcitrant disease, with poor 5-year survival rates. Surgery and empiric cytotoxic chemotherapy anchor current treatment approaches, 14,45 and although disease heterogeneity is apparent in the clinic, the underlying determinants are unclear. We show here that one-fifth to one-third of EACs simultaneously express products specific to the esophageal squamous epithelium and columnar intestinal cells. It will be important in the future to identify the clinical and genetic correlates of these EACs showing bi-lineage gene expression and to determine whether they reflect a distinctive pathophysiology or harbor unique therapeutic vulnerabilities.

**References**

1. Spechler SJ, Souza RF. Barrett’s esophagus. N Engl J Med 2014;371:836–845.

2. Hvid-Jensen F, Pedersen L, Drewes AM, et al. Incidence of adenocarcinoma among patients with Barrett’s esophagus. N Engl J Med 2011;365:1375–1383.

3. Souza RF, Krishnan K, Spechler SJ. Acid, bile, and CDX: the ABCs of making Barrett’s metaplasia. Am J Physiol Gastrointest Liver Physiol 2008;295:G211–G218.
4. Gao N, White P, Kaestner KH. Establishment of intestinal identity and epithelial-mesenchymal signaling by Cdx2. Dev Cell 2009;16:588–599.

5. Silberg DG, Sullivan J, Kang E, et al. Cdx2 ectopic expression induces gastric intestinal metaplasia in transgenic mice. Gastroenterology 2002;122:689–696.

6. Mutoh H, Sakurai S, Satoh K, et al. Cdx1 induced intestinal metaplasia in the transgenic mouse stomach: comparative study with Cdx2 transgenic mice. Gut 2004;53:1416–1423.

7. Eda A, Osawa H, Satoh K, et al. Aberrant expression of Cdx2 in Barrett’s epithelium and inflammatory esophageal mucosa. J Gastroenterol 2003;38:14–22.

8. Stairs DB, Nakagawa H, Klein-Szanto A, et al. Ectopic Cdx2 expression in murine esophagus models an intermediate stage in the emergence of Barrett’s esophagus. PLoS One 2011;6:e18280.

9. Que J, Okubo T, Goldenring JR, et al. Critical role of p63 in the development of a normal esophageal and tracheobronchial epithelium. Am J Physiol Cell Physiol 2004;287:C171–C181.

10. Daniely Y, Liao G, Dixon D, et al. p63 and p53: a family of transcribers in development, immunity, and disease. J Cell Biochem 2004;91:317–330.

11. Wang X, Ouyang H, Yamamoto Y, et al. Residual epithelialization and epithelial-mesenchymal signaling by Cdx2 in Barrett’s epithelium and inflammatory esophageal mucosa. J Gastroenterol 2003;38:14–22.

12. Harada H, Nakagawa H, Oyama K, et al. Telomerase induces immortalization of human esophageal keratinocytes without p16INK4a inactivation. Mol Cancer Res 2003;1:729–738.

13. Roth RB, Hevezi P, Lee J, et al. Gene expression analyses reveal molecular relationships among 20 regions of the human CNS. Neurogenetics 2006;7:67–80.

14. Kim SM, Park YY, Park ES, et al. Prognostic biomarkers for esophageal adenocarcinoma identified by analysis of tumor transcriptome. PLoS One 2010;5:e15074.

15. Cancer Genome Atlas Research Network. Comprehensive molecular characterization of gastric adenocarcinoma. Nature 2014;513:202–209.

16. Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. Nature 2012;487:330–337.

17. Trappel C, Roberts A, Goff L, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nature Protoc 2012;7:562–578.

18. Kim J, Cantor AB, Orkin SH, et al. Integrative genomics viewer. Nat Biotechnol 2011;29:24–26.

19. McLean CY, Bristor D, Hiller M, et al. GREAT improves functional interpretation of cis-regulatory regions. Nat Biotechnol 2010;28:495–501.

20. Hackett NR, Shaykhiev R, Walters MS, et al. The human airway epithelial basal cell transcriptome. PLoS One 2011;6:e18378.

21. Hatakeyama S. TRIM proteins and cancer. Nat Rev Cancer 2011;11:792–804.

22. Palanca-Wessels MC, Barrett MT, Galipeau PC, et al. Genetic analysis of long-term Barrett’s esophagus epithelial cultures exhibiting cytogenetic and ploidy abnormalities. Gastroenterology 1998;114:295–304.

23. Rhodes DR, Kalyana-Sundaram S, Mahavisno V, et al. Oncomine 3.0: genes, pathways, and networks in a collection of 18,000 cancer gene expression profiles. Neoplasia 2007;9:116–180.

24. Hatakeyama S. TRIM proteins and cancer. Nat Rev Cancer 2011;11:792–804.

25. Kamachi Y, Kondoh H. Sox proteins: regulators of cell fate specification and differentiation. Development 2013;140:4129–4144.

26. Hackett NR, Shaykhiev R, Walters MS, et al. The human airway epithelial basal cell transcriptome. PLoS One 2011;6:e18378.

27. Thu KL, Becker-Santos DD, Radulovich N, et al. SOX15 and other SOX family members are important mediators of tumorigenesis in multiple cancer types. Oncoscience 2014;1:326–335.

28. Hyland PL, Hu N, Rotunno M, et al. Global changes in gene expression of Barrett’s esophagus compared to normal squamous esophagus and gastric cardia tissues. PLoS One 2014;9:e93219.

29. El-Zimaity HM, Graham DY. Cytokeratin subsets for distinguishing Barrett’s esophagus from intestinal metaplasia in the cardia using endoscopic biopsy specimens. Am J Gastroenterol 2001;96:1378–1382.
38. Nurgalieva Z, Lowrey A, El-Serag HB. The use of cytokeratin stain to distinguish Barrett’s esophagus from contiguous tissues: a systematic review. Dig Dis Sci 2007;52:1345–1354.

39. van Baal JW, Bozikas A, Pronk R, et al. Cytokeratin and CDX-2 expression in Barrett’s esophagus. Scand J Gastroenterol 2008;43:132–140.

40. DiMaio MA, Kwok S, Montgomery KD, et al. Immunohistochemical panel for distinguishing esophageal adenocarcinoma from squamous cell carcinoma: a combination of p63, cytokeratin 5/6, MUC5AC, and anterior gradient homolog 2 allows optimal subtyping. Hum Pathol 2012;43:1799–1807.

41. Lee HJ, Göring W, Ochs M, et al. Sox15 is required for skeletal muscle regeneration. Mol Cell Biol 2004;24:8428–8436.

42. Maruyama M, Ichisaka T, Nakagawa M, et al. Differential roles for Sox15 and Sox2 in transcriptional control in mouse embryonic stem cells. J Biol Chem 2005;280:24371–24379.

43. Quante M, Bhagat G, Abrams JA, et al. Bile acid and inflammation activate gastric cardia stem cells in a mouse model of Barrett-like metaplasia. Cancer Cell 2012;21:36–51.

44. Ajani JA, Rodriguez W, Bodoky G, et al. Multicenter phase III comparison of cisplatin/S-1 with cisplatin/infusional fluorouracil in advanced gastric or gastroesophageal adenocarcinoma study: the FLAGS trial. J Clin Oncol 2010;28:1547–1553.

45. Cunningham D, Starling N, Rao S, et al. Capecitabine and oxaliplatin for advanced esophagogastric cancer. N Engl J Med 2008;358:36–46.
Supplementary Figure 1. Genes coexpressed with SOX15 in the Cancer Genome Atlas (TCGA) data set of gastric cancers. (A) Correlations of all mRNAs with SOX15 levels. The 317 genes coexpressed with SOX15 in esophageal epithelium are marked with red lines. (B) Expression of SOX15 coexpressed genes in 30 cases of CIN⁺ (chromosomal instability) adenocarcinomas from the gastroesophageal junction or gastric cardia.
### Supplementary Table 1. Top Gene Ontology (GO) terms for transcripts altered after SOX15 depletion in CPA cells

| Term RNAseq Down                                      | P Value         | Term RNAseq UP                           | P Value         |
|-------------------------------------------------------|-----------------|------------------------------------------|-----------------|
| GO:0000278 ~ mitotic cell cycle                       | 3.26E–20        | GO:0000079 ~ regulation of cyclin-dependent protein kinase activity | 4.40E–06        |
| GO:0051301 ~ cell division                            | 6.39E–20        | GO:0016192 ~ vesicle-mediated transport  | 6.89E–06        |
| GO:0000280 ~ nuclear division                         | 1.39E–19        | GO:0051726 ~ regulation of cell cycle    | 8.53E–06        |
| GO:0007067 ~ mitosis                                  | 1.39E–19        | GO:0009991 ~ response to extracellular stimulus | 3.09E–05        |
| GO:0022403 ~ cell cycle phase                         | 1.48E–19        | GO:0006091 ~ generation of precursor metabolites and energy | 5.73E–05        |
| GO:0000087 ~ M phase of mitotic cell cycle            | 1.52E–19        | GO:0045767 ~ regulation of antiapoptosis  | 7.05E–05        |
| GO:0022402 ~ cell cycle process                        | 4.58E–19        | GO:0031667 ~ response to nutrient levels | 1.71E–04        |
| GO:0007049 ~ cell cycle                               | 2.49E–18        | GO:0007584 ~ response to nutrient        | 2.63E–04        |
| GO:0000279 ~ M phase                                  | 2.96E–18        | GO:0008219 ~ cell death                  | 5.87E–04        |
| GO:0048285 ~ organelle fission                        | 3.08E–18        | GO:0010033 ~ response to organic substance | 6.29E–04        |
| GO:0007059 ~ chromosome segregation                    | 1.93E–10        | GO:0042981 ~ regulation of apoptosis     | 6.32E–04        |
| GO:0008104 ~ protein localization                     | 4.68E–10        | GO:0055114 ~ oxidation reduction         | 6.36E–04        |
| GO:0045184 ~ establishment of protein localization     | 5.39E–10        | GO:0042127 ~ regulation of cell proliferation | 6.77E–04        |
| GO:0008654 ~ phospholipid biosynthetic process         | 6.48E–10        | GO:0016265 ~ death                       | 6.77E–04        |
| GO:0015031 ~ protein transport                        | 2.46E–09        | GO:0043067 ~ regulation of programmed cell death | 7.56E–04        |
| Gene     | Correlation With SOX15 Expression |
|----------|-----------------------------------|
| SOX15    | 1.000                             |
| GRHL3    | 0.981                             |
| FAM46B   | 0.981                             |
| ZNF750   | 0.981                             |
| LYPD3    | 0.981                             |
| CAPNS2   | 0.979                             |
| IL20RB   | 0.978                             |
| BNIPL    | 0.978                             |
| ANXA8L2  | 0.975                             |
| GPR87    | 0.971                             |
| TMPRSS11D| 0.971                             |
| LY6D     | 0.971                             |
| LASS3    | 0.971                             |
| CLCA2    | 0.971                             |
| PKP1     | 0.971                             |
| KRT5     | 0.971                             |
| GSDMC    | 0.970                             |
| TMEM40   | 0.970                             |
| FAM83C   | 0.970                             |
| TP63     | 0.970                             |
| DSC3     | 0.970                             |
| TGM5     | 0.964                             |
| TMPRSS11A| 0.964                             |
| GJB6     | 0.964                             |
| KLK13    | 0.964                             |
| LYNXI    | 0.964                             |
| SPINK5   | 0.961                             |
| ENDOU    | 0.961                             |
| RNF222   | 0.961                             |
| PRSS27   | 0.961                             |
| KRT78    | 0.961                             |
| CRC1T    | 0.961                             |
| SCEL     | 0.961                             |
| A2ML1    | 0.961                             |
| SLURP1   | 0.961                             |
| c9orf169 | 0.961                             |
| CSTA     | 0.961                             |
| MAL      | 0.961                             |
| KRT6C    | 0.961                             |
| AHRGAP6  | 0.961                             |
| DSG3     | 0.961                             |
| TGM1     | 0.961                             |
| SBSN     | 0.961                             |
| SPRR1B   | 0.961                             |
| CLCA4    | 0.961                             |
| CALML3   | 0.961                             |
| RHCG     | 0.961                             |

**Supplementary Table 2. Continued**

| Gene     | Correlation With SOX15 Expression |
|----------|-----------------------------------|
| KRT4     | 0.961                             |
| SERPINB13| 0.961                             |
| GBP6     | 0.961                             |
| NCCRIP1  | 0.961                             |
| TMRSS11B | 0.961                             |
| CNFN     | 0.961                             |
| TGM3     | 0.961                             |
| CRNN     | 0.961                             |
| HSPB8    | 0.961                             |
| SERPINB2 | 0.961                             |
| S100A2   | 0.961                             |
| LGALS7B  | 0.952                             |
| LGALS7   | 0.952                             |
| DUOX1    | 0.949                             |
| DUOXA1   | 0.949                             |
| CSNK1E   | 0.944                             |
| CLIC3    | 0.942                             |
| HOPX     | 0.940                             |
| SERPINB4 | 0.939                             |
| SERPINB3 | 0.939                             |
| ECM1     | 0.937                             |
| Trim29   | 0.937                             |
| SLC39A2  | 0.937                             |
| RAET1G   | 0.937                             |
| AQP3     | 0.937                             |
| TMEM154  | 0.937                             |
| GNA15    | 0.937                             |
| SULT2B1  | 0.937                             |
| ALDH3B2  | 0.937                             |
| EVPL     | 0.937                             |
| GRHL1    | 0.937                             |
| KAZ      | 0.937                             |
| PITX1    | 0.937                             |
| TMEM79   | 0.937                             |
| DENND2C  | 0.937                             |
| VSIG10L  | 0.937                             |
| ZNF185   | 0.937                             |
| PPL      | 0.937                             |
| GJB5     | 0.937                             |
| KRT15    | 0.937                             |
| c10orf99 | 0.931                             |
| EPHX3    | 0.928                             |
| AIF1L    | 0.928                             |
| CRABP2   | 0.928                             |
| PPP1R3C  | 0.917                             |
| ZNF365   | 0.916                             |
| CPA4     | 0.916                             |
| SPINK7   | 0.916                             |
| RNASE7   | 0.916                             |
| LOC643479| 0.916                             |
### Supplementary Table 2. Continued

| Gene     | Correlation With SOX15 Expression |
|----------|-----------------------------------|
| TIAM1    | 0.912                             |
| ARL4D    | 0.912                             |
| LASS4    | 0.912                             |
| IVL      | 0.912                             |
| P2RY1    | 0.912                             |
| DLK2     | 0.912                             |
| ANXA8    | 0.912                             |
| BBOX1    | 0.911                             |
| CYP4F22  | 0.911                             |
| SCNN18   | 0.911                             |
| MUC15    | 0.911                             |
| CWH43    | 0.911                             |
| CALML5   | 0.909                             |
| CST6     | 0.906                             |
| FAM83A   | 0.904                             |
| CDA      | 0.904                             |
| KRT80    | 0.904                             |
| LYPD2    | 0.901                             |
| FGFI1    | 0.899                             |
| PPP2R2C  | 0.899                             |
| TLE3     | 0.881                             |
| DOCK9    | 0.881                             |
| PLD2     | 0.881                             |
| PYGL     | 0.881                             |
| BNIIP3   | 0.881                             |
| TUBB6    | 0.881                             |
| NDUFA4L2 | 0.881                             |
| BDKRB1   | 0.881                             |
| NDRG4    | 0.881                             |
| CBR3     | 0.881                             |
| SLC22A17 | 0.881                             |
| SRPX2    | 0.881                             |
| FRMD6    | 0.881                             |
| MID2     | 0.881                             |
| EFS      | 0.881                             |
| PARD6G   | 0.881                             |
| c3orf54  | 0.881                             |
| RGM4     | 0.881                             |
| RRAGD    | 0.881                             |
| ANKRD35  | 0.881                             |
| TNAIP8L3 | 0.881                             |
| ELOVL4   | 0.881                             |
| CRYAB    | 0.881                             |
| GPC1     | 0.881                             |
| ZNF385A  | 0.881                             |
| WDFY2    | 0.881                             |
| NOD2     | 0.881                             |
| PTPN13   | 0.881                             |
| TFAP2C   | 0.881                             |
| CDK5R1   | 0.881                             |
| VSNL1    | 0.881                             |
| MICALL1  | 0.872                             |
| PAK6     | 0.872                             |
| PVRL1    | 0.872                             |
| PVRL4    | 0.872                             |
| ITPKC    | 0.872                             |
| SPTBN2   | 0.872                             |
| SAMD9    | 0.872                             |
| AIM1L    | 0.872                             |
| PLCD1    | 0.872                             |
| NLRX1    | 0.872                             |
| MPZL2    | 0.872                             |
| PRRG4    | 0.872                             |
| PPP1R13L | 0.872                             |
| URGCP    | 0.872                             |
| XG       | 0.863                             |
| HES2     | 0.863                             |
| c12orf54 | 0.863                             |
| IRX4     | 0.863                             |
| DST      | 0.863                             |
| BNC1     | 0.863                             |
| ARHGEF4  | 0.858                             |
| DMKN     | 0.855                             |
| KLK10    | 0.855                             |
| LTB4R    | 0.855                             |
| CYP2E1   | 0.853                             |
| KCTD1    | 0.852                             |
| ATP13A4  | 0.849                             |
| ATP6V0A4 | 0.849                             |
| RASGRP1  | 0.849                             |
| TRIM6    | 0.849                             |
| SHROOM2  | 0.849                             |
| CLIP4    | 0.849                             |
| KLK12    | 0.840                             |
| S100A8   | 0.840                             |
| S100A9   | 0.840                             |
| KRT6B    | 0.840                             |
| SPRR3    | 0.840                             |
| KRT13    | 0.840                             |
| SPRR1A   | 0.840                             |
| SPRR2A   | 0.840                             |
| NCKAP5   | 0.833                             |
| GPR1     | 0.833                             |
| CDKN2B   | 0.833                             |
| LOC653110| 0.833                             |
| SFTP    | 0.833                             |
| FBXO27   | 0.833                             |
| ZNF433   | 0.824                             |
| CA12     | 0.824                             |
| WNT4     | 0.824                             |
### Supplementary Table 2. Continued

| Gene          | Correlation With SOX15 Expression |
|---------------|-----------------------------------|
| MTSS1         | 0.824                             |
| WDR47         | 0.824                             |
| ELL2          | 0.824                             |
| RORA          | 0.824                             |
| SLC9A9        | 0.824                             |
| SEMA4A        | 0.824                             |
| MAF           | 0.824                             |
| BEX4          | 0.824                             |
| SERPINB8      | 0.824                             |
| ZDHHC21       | 0.824                             |
| ZNF425        | 0.824                             |
| SNX24         | 0.824                             |
| ALDH4A1       | 0.824                             |
| NOTCH2NL      | 0.824                             |
| DUSP22        | 0.824                             |
| MBD2          | 0.824                             |
| C4ORF3        | 0.824                             |
| CNOT1         | 0.824                             |
| CTNBP2NL      | 0.824                             |
| CUL4B         | 0.824                             |
| ADK           | 0.824                             |
| MOXD1         | 0.824                             |
| PDZD2         | 0.824                             |
| CAB33P        | 0.824                             |
| ZNF431        | 0.824                             |
| RASAL2        | 0.824                             |
| TMOD3         | 0.824                             |
| ARHGAP10      | 0.824                             |
| DAPP1         | 0.824                             |
| SLC2A6        | 0.824                             |
| ZNF426        | 0.824                             |
| RIT1          | 0.824                             |
| UBE2H         | 0.824                             |
| SPTLC1        | 0.824                             |
| KAT2B         | 0.824                             |
| SECISBP2L     | 0.824                             |
| KIAA1370      | 0.824                             |
| RNF11         | 0.824                             |
| WDR26         | 0.824                             |
| RANBP9        | 0.824                             |
| ABHD5         | 0.824                             |
| YOD1          | 0.824                             |
| SEPT10        | 0.824                             |
| UBE2G1        | 0.824                             |
| SASH1         | 0.824                             |
| GAB1          | 0.824                             |
| PARD3         | 0.824                             |
| RSC1A1        | 0.824                             |
| TPD52L1       | 0.824                             |
| IL34          | 0.824                             |
**Supplementary Table 2. Continued**

| Gene   | Correlation With SOX15 Expression |
|--------|-----------------------------------|
| CRISP3 | 0.824                             |
| EHD3   | 0.824                             |
| ACYP2  | 0.824                             |
| C2ORF54| 0.824                             |
| KLF8   | 0.824                             |
| KREM1  | 0.824                             |
| TPRG1  | 0.824                             |
| PAX9   | 0.824                             |
| SUSD4  | 0.824                             |
| DAPL1  | 0.824                             |
| ARSF   | 0.816                             |
| ALOX15B| 0.816                             |
| C18ORF26| 0.816                          |
| FMO2   | 0.816                             |
| FAM63A | 0.806                             |
| SH3GL3 | 0.806                             |
| LY6G6C | 0.806                             |
| DSG1   | 0.806                             |
| CLDN17 | 0.806                             |
| KPRP   | 0.806                             |
| IGFL1  | 0.806                             |