SUPPLEMENTAL FIGURE AND TABLE LEGENDS

Fig. S1. **Characteristics of lung cancer cell lines used in this study.** (A) Summary of origins and a subset of genomic alterations in the cell lines. For ChaGoK1 and SW1573, genomic alteration information was obtained from the Cancer Cell Line Encyclopedia (CCLE) and downloaded from [www.cbioportal.org](http://www.cbioportal.org). MGH7 3q copy number information was from Luk et al, *Cancer Genet Cytogenet*, 2001. NA = not available. (B) Analysis of TP63 and SOX2 mRNA expression in the cell lines. Gene expression was quantified by qRT-PCR and normalized to TBP levels. Means ± SEM from three technical replicates are shown, and are plotted relative to expression in SW1573 cells, which was assigned a value of 1.0. (C) Analysis of TP63 and SOX2 protein expression in the cell lines. Cells were subjected to immunofluorescence staining with α-TP63 and α-SOX2 antibodies. Images were captured using the same exposure time. Scale bar is 20 μm.

Fig. S2. **Inhibition of the SOX2-EP300 interaction in SQCC cells by adenoviral E1A oncoproteins.** MGH7 SQCC cells were transiently transfected with either a GFP expression vector alone or the GFP vector and an expression vector for the adenoviral E1A 13S or 12S isoform, both of which bind to EP300. After 3 days, cells were fixed and subjected to the proximity ligation assay (PLA). Cells were stained with both α-SOX2 and α-EP300 primary antibodies. Only GFP-expressing cells were scored for presence of nuclear foci. Mean percentages of foci-positive nuclei and nuclei with multiple foci ± SEM are shown. Data were obtained by scoring 5 fields comprising 53-79 nuclei per field. Significance was calculated using a 2-tailed t test relative to transfection of GFP vector alone. ***p = 5.2 x 10^-6 and 1.2 x
10-6 for 13S and 12S, respectively (left panel, foci-positive nuclei) and 0.0003 for both 13S and 12S (right panel, nuclei with multiple foci).

Fig. S3. **Frequency of SOX2 copy number gains in human cancers.** Provisional TCGA data for the different human cancers are graphed, and include both low and high copy number gains (amplification). Significance of differences in frequencies of SOX2 copy number gains between individual cancers and lung SQCC (squamous cell carcinoma, n=177) was calculated by a two-tailed Fisher’s exact test. Ovarian = Ovarian serous cystadenocarcinoma (p = 2 x 10-5, n=311), Cervical = Cervical squamous cell Carcinoma and endocervical adenocarcinoma (p = 1 x 10-6, n=191), Head & neck = Head & neck squamous cell carcinoma (p = 5 x 10-9, n=504), Esophagus = Esophageal carcinoma (p= 7 x 10-12, n=184), Bladder = Bladder urothelial carcinoma (p = 1 x 10-14, n=126), Stomach = Stomach adenocarcinoma (p = 2 x 10-43, n=393), DLBC = Diffuse Large B-cell Lymphoma (p = 8 x 10-19, n=48), Breast = Breast invasive carcinoma (p < 1 x 10-89, n=963), Lung ADC = Lung adenocarcinoma (p = 6 x 10-47, n=230), Uterine = Uterine corpus endometrial carcinoma (p = 2 x 10-51, n=242), GBM = Glioblastoma multiforme (p = 6 x 10-60, n=273), Melanoma = Skin cutaneous melanoma (p = 4 x 10-62, n=287), Sarcoma (p = 1 x 10-59, n=243), Pancreas = Pancreatic adenocarcinoma (p = 2 x 10-49, n=149), Colorectal = Colorectal adenocarcinoma (p = 7 x 10-59, n=220), ccRCC = Kidney renal clear cell carcinoma (p = 1 x 10-76, n=448), Liver = Liver hepatocellular carcinoma (p = 3 x 10-72, n=366), Prostate = Prostate adenocarcinoma (p = 3 x 10-85, n=492), AML = Acute Myeloid Leukemia (p = 2 x 10-89, n=188).
Fig. S4. **Frequency of EP300 copy number variations in human cancers.** Provisional TCGA data for the different human cancers are graphed, and include low and high copy number gains (amplification), as well as heterozygous and homozygous losses. Significance of differences in frequencies of EP300 copy number variations between individual cancers and lung SQCC (squamous cell carcinoma, n=177) was calculated by a two-tailed Fisher’s exact test. Melanoma = Skin cutaneous melanoma [p = 0.7 (gains), p =0.0002 (losses), n=287], Sarcoma [p = 5 x 10-5 (gains), p=0.1 (losses), n=243], Head & neck = Head & neck squamous cell carcinoma [p = 8 x 10-9 (gains), p = 0.1 (losses), n=504], Esophagus = Esophageal carcinoma [p = 2 x 10-7 (gains), p = 0.0007 (losses), n=184], Liver = Liver hepatocellular carcinoma [p = 1 x 10-13 (gains), p =1.0 (losses), n=366], Bladder = Bladder urothelial carcinoma [p = 3 x 10-8 (gains), p = 3 x 10-6 (losses), n=126], Cervical = Cervical squamous cell Carcinoma and endocervical adenocarcinoma [p = 9 x 10-11 (gains), p = 0.3 (losses), n=191], Lung ADC = Lung adenocarcinoma [p = 4 x 10-13 (gains), p = 6 x 10-7 (losses), n=230], Breast = Breast invasive carcinoma [p < 1 x 10-46 (gains), p < 1 x 10-47 (losses), n=963], Pancreas = Pancreatic adenocarcinoma [p = 8 x 10-13 (gains), p = 0.2 (losses), n=149], Stomach = Stomach adenocarcinoma [p = 6 x 10-22 (gains), p = 0.001 (losses), n=393], ccRCC = Kidney renal clear cell carcinoma [p = 4 x 10-24 (gains), p = 4 x 10-5 (losses), n=448], Uterine = Uterine corpus endometrial carcinoma [p = 2 x 10-20 (gains), p = 0.2 (losses), n=242], GBM = Glioblastoma multiforme [p = 9 x 10-25 (gains), p = 0.0003 (losses), n=273], AML = Acute Myeloid Leukemia [p = 7 x 10-21 (gains), p = 1 x 10-10 (losses), n=188], Ovarian = Ovarian serous cystadenocarcinoma [p = 4 x 10-29 (gains), p = 3 x 10-47 (losses), n=311], Colorectal = Colorectal adenocarcinoma [p = 1 x 10-25 (gains), p = 0.0004 (losses), n=220], DLBC = Diffuse
Large B-cell Lymphoma \[p = 3 \times 10^{-8} \text{ (gains)}, p = 0.03 \text{ (losses), n=48}\], Prostate = Prostate adenocarcinoma \[p = 7 \times 10^{-46} \text{ (gains)}, p = 0.001 \text{ (losses), n=492}\].

Fig. S5. **Frequency of EP300 mutations in human cancers.** Provisional TCGA data for the different human cancers are graphed. Significance of differences in frequencies of EP300 mutations between individual cancers and lung SQCC (squamous cell carcinoma, n=177) was calculated by a two-tailed Fisher’s exact test. Bladder = Bladder urothelial carcinoma (p = 0.0006, n=126), Cervical = Cervical squamous cell Carcinoma and endocervical adenocarcinoma (p = 0.03, n=191), Uterine = Uterine corpus endometrial carcinoma (p = 0.09, n=242), Head & neck = Head & neck squamous cell carcinoma (p = 0.2, n=504), Esophagus = Esophageal carcinoma (p = 0.5, n=184), DLBC = Diffuse Large B-cell Lymphoma (p = 0.7, n=48), Stomach = Stomach adenocarcinoma (p = 0.6, n=393), Melanoma = Skin cutaneous melanoma (p = 0.3, n=287), Colorectal = Colorectal adenocarcinoma (p = 0.8, n=220), ccRCC = Kidney renal clear cell carcinoma (p = 0.8, n=448), Liver = Liver hepatocellular carcinoma (p = 0.6, n=366), Pancreas = Pancreatic adenocarcinoma (p = 0.6, n=149), Breast = Breast invasive carcinoma (p = 0.009, n=963), Prostate = Prostate adenocarcinoma (p = 0.01, n=492), Lung ADC = Lung adenocarcinoma (p = 0.02, n=230), Sarcoma (p = 0.02, n=243), GBM = Glioblastoma multiforme (p = 0.003, n=273), Ovarian = Ovarian serous cystadenocarcinoma (p = 0.002, n=311), AML = Acute Myeloid Leukemia (p = 0.003, n=188).
Table S1. **List of all peptides detected in the BioID analysis.**

Table S2. **List of all proteins identified from more than one unique peptide in the BioID analysis.**

Table S3. **Identification of Gene Ontology (GO) annotations that are significantly enriched in the SOX2 BioID dataset.** Gene Ontology annotations were restricted to “biological processes” and were determined using PANTHER (www.pantherdb.org). Statistical significance was calculated using the Mann-Whitney U Test (Wilcoxon Rank-Sum Test) and applying a Bonferroni correction. Results with a p-value ≤ 0.05 are shown.

Table S4. **Identification of gene sets enriched in the SOX2 BioID dataset.** Gene set enrichment analysis (GSEA) was performed at http://software.broadinstitute.org/gsea/index.jsp.

using the Hallmark (H), Curated (C2), and Oncogenic signature (C6) gene sets. The top 100 enriched gene sets having a false discovery rate (FDR q-value) ≤ 0.05 are shown.
<table>
<thead>
<tr>
<th>Cell line</th>
<th>Cancer type</th>
<th>3q gain/loss</th>
<th>EP300 gain/loss/mutation</th>
<th>KRAS mutation</th>
</tr>
</thead>
<tbody>
<tr>
<td>MGH7</td>
<td>Lung squamous</td>
<td>High copy amp</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>ChaGoK1</td>
<td>Lung bronchogenic</td>
<td>Low copy gain</td>
<td>Diploid/wildtype</td>
<td>Wildtype</td>
</tr>
<tr>
<td>SW1573</td>
<td>Lung alveolar</td>
<td>Het loss</td>
<td>Diploid/wildtype</td>
<td>G12C</td>
</tr>
</tbody>
</table>

**Fig. S1**

**A**

<table>
<thead>
<tr>
<th>Cell line</th>
<th>Cancer type</th>
<th>3q gain/loss</th>
<th>EP300 gain/loss/mutation</th>
<th>KRAS mutation</th>
</tr>
</thead>
<tbody>
<tr>
<td>MGH7</td>
<td>Lung squamous</td>
<td>High copy amp</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>ChaGoK1</td>
<td>Lung bronchogenic</td>
<td>Low copy gain</td>
<td>Diploid/wildtype</td>
<td>Wildtype</td>
</tr>
<tr>
<td>SW1573</td>
<td>Lung alveolar</td>
<td>Het loss</td>
<td>Diploid/wildtype</td>
<td>G12C</td>
</tr>
</tbody>
</table>

**B**

**TP63 mRNA levels**

- PDX77 (3q amp)
- MGH7 (3q amp)
- ChaGoK1 (3q gain)
- SW1573 (3q no gain)

**SOX2 mRNA levels**

- PDX77 (3q amp)
- MGH7 (3q amp)
- ChaGoK1 (3q gain)
- SW1573 (3q no gain)

**C**

- MGH7 (3q amp)
- ChaGoK1 (3q gain)
- SW1573 (No 3q gain)

**DAPI**

- MGH7 (3q amp)
- ChaGoK1 (3q gain)
- SW1573 (No 3q gain)

**TP63**

- MGH7 (3q amp)
- ChaGoK1 (3q gain)
- SW1573 (No 3q gain)

**SOX2**

- MGH7 (3q amp)
- ChaGoK1 (3q gain)
- SW1573 (No 3q gain)
Fig. S2

Transfect

EP300 binding

E1A 13S
CR1 CR2 CR3 CR4

E1A 12S
CR1 CR2 CR4

GFP-positive cells
% PLA positive

GFP-positive cells
% Nuclei with multiple PLA foci

Score
SOX2/EP300 PLA
in GFP cells

MGH7
SQCC cells

***
***

GFP alone
GFP + E1A 13S
GFP + E1A 12S
Frequency of SOX2 copy number gains in human cancer

Fig. S3
Frequency of EP300 copy number losses in human cancer

Frequency of EP300 copy number gains in human cancer

Fig. S4