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. 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 ($p = 2 \times 10\text{-}5$, n=311), Cervical = Cervical squamous cell Carcinoma and endocervical adenocarcinoma ($p = 1 \times 10\text{-}6$, n=191), Head & neck = Head & neck squamous cell carcinoma ($p = 5 \times 10\text{-}9$, n=504), Esophagus = Esophageal carcinoma ($p= 7 \times 10\text{-}12$, n=184), Bladder = Bladder urothelial carcinoma ($p = 1 \times 10\text{-}14$, n=126), Stomach = Stomach adenocarcinoma ($p = 2 \times 10\text{-}43$, n=393), DLBC = Diffuse Large B-cell Lymphoma ($p = 8 \times 10\text{-}19$, n=48), Breast = Breast invasive carcinoma ($p < 1 \times 10\text{-}89$, n=963), Lung ADC = Lung adenocarcinoma ($p = 6 \times 10\text{-}47$, n=230), Uterine = Uterine corpus endometrial carcinoma ($p = 2 \times 10\text{-}51$, n=242), GBM = Glioblastoma multiforme ($p = 6 \times 10\text{-}60$, n=273), Melanoma = Skin cutaneous melanoma ($p = 4 \times 10\text{-}62$, n=287), Sarcoma ($p = 1 \times 10\text{-}59$, n=243), Pancreas = Pancreatic adenocarcinoma ($p = 2 \times 10\text{-}49$, n=149), Colorectal = Colorectal adenocarcinoma ($p = 7 \times 10\text{-}59$, n=220), ccRCC = Kidney renal clear cell carcinoma ($p = 1 \times 10\text{-}76$, n=448), Liver = Liver hepatocellular carcinoma ($p = 3 \times 10\text{-}72$, n=366), Prostate = Prostate adenocarcinoma ($p = 3 \times 10\text{-}85$, n=492), AML = Acute Myeloid Leukemia ($p = 2 \times 10\text{-}89$, 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 x 10-8 (gains), p = 0.03 (losses), n=48], Prostate = Prostate adenocarcinoma [p = 7 x 10-46 (gains), p = 0.001 (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). [a]Statistical significance was calculated using the Mann-Whitney U Test (Wilcoxon Rank-Sum Test) and applying a Bonferonni 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,

Table S5. **Comparative analyses of the SOX2 BioID and previous AP-MS studies.** The SOX2 AP-MS datasets were from Watanabe *et al*, *J Clin Invest*, 2014; Mallanna *et al*, *Stem Cells*, 2010; Gao *et al*, *J Biol Chem*, 2012; Ding *et al*, *Cell Stem Cell*, 2015; Myers *et al*, *eLife*, 2016; Engelen *et al*, *Nat Genet*, 2011; Cox et al, *PLoS One*, 2013; Fang *et al*, *Proteomics*, 2011. The BioID data were from our study. For Mallanna *et al*, Gao *et al*, and Cox *et al*, high confidence SOX2-interactors were obtained from Tables S2A, S2B, S3A; S1-S3; and S1-S3, respectively, from those studies.

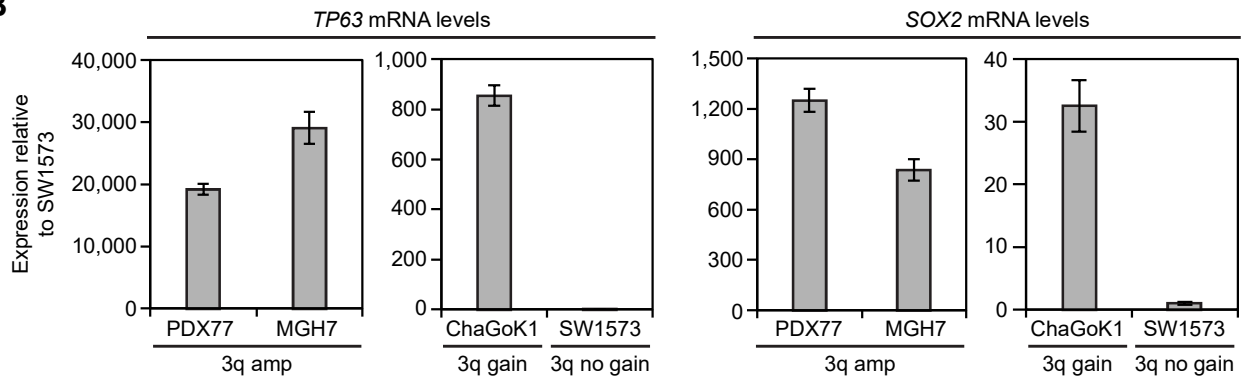
using the Hallmark (H), Curated (C2), and Oncogenic signature (C6) gene sets. The top 100 enriched gene sets having a false discovery rate (FDR q-value) ≤ 0.05 are shown.

Fig. S1

A

Cell line	Cancer type	3q gain/loss	*EP300* gain/loss/ mutation	*KRAS* mutation
MGH7	Lung squamous	High copy amp	NA	NA
ChaGoK1	Lung bronchogenic	Low copy gain	Diploid/wildtype	Wildtype
SW1573	Lung alveolar	Het loss	Diploid/wildtype	G12C

B



C

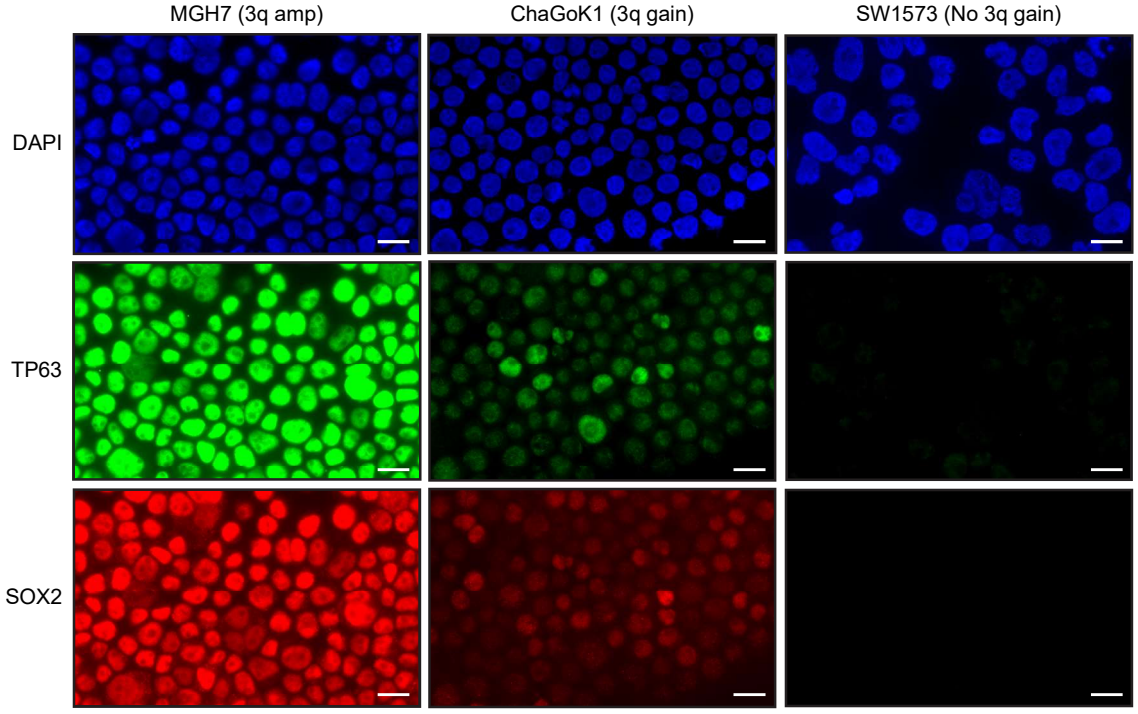


Fig. S2

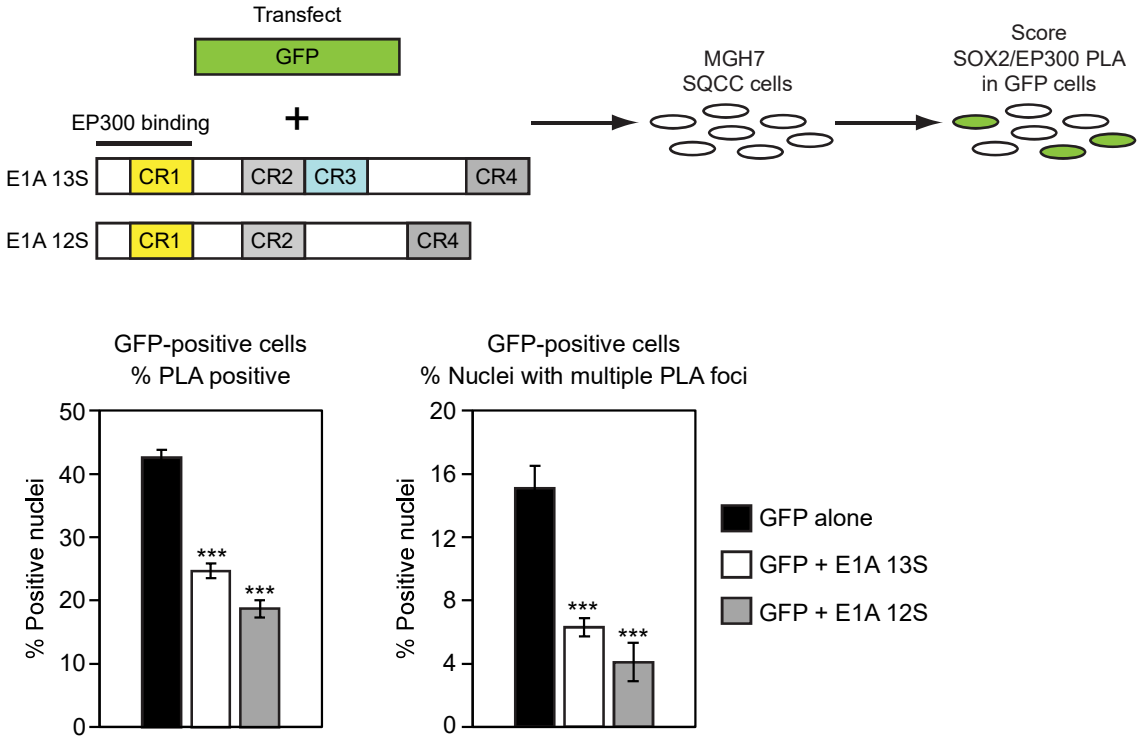

Transfect
GFP
EP300 binding
+
E1A 13S
CR1
CR2
CR3
CR4
E1A 12S
CR1
CR2
CR4
MGH7
SQCC cells
Score
SOX2/EP300 PLA
in GFP cells
GFP-positive cells
% PLA positive
% Positive nuclei
50
40
30
20
10
0
***
***
GFP-positive cells
% Nuclei with multiple PLA foci
% Positive nuclei
20
16
12
8
4
0
***
***
GFP alone
GFP + E1A 13S
GFP + E1A 12S

Fig. S3

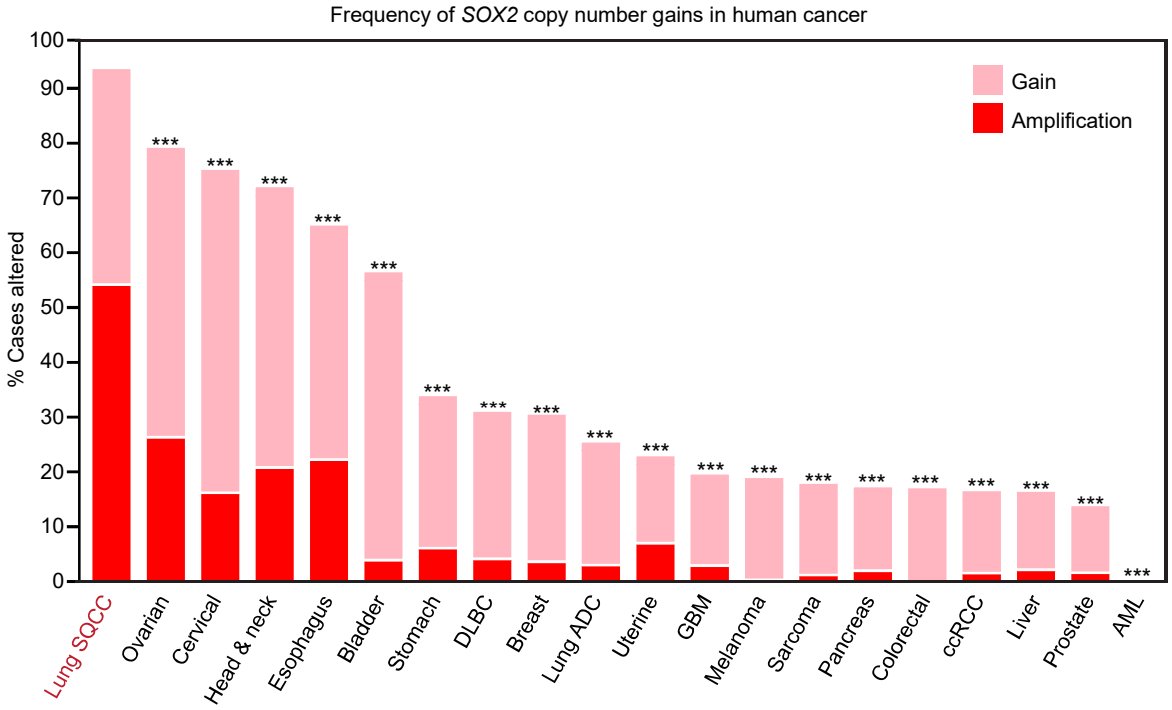

Frequency of SOX2 copy number gains in human cancer
% Cases altered
100
90
80
70
60
50
40
30
20
10
0
Gain
Amplification
***
Lung SQCC
Ovarian
Cervical
Head & neck
Esophagus
Bladder
Stomach
DLBC
Breast
Lung ADC
Uterine
GBM
Melanoma
Sarcoma
Pancreas
Colorectal
ccRCC
Liver
Prostate
AML

Fig. S4

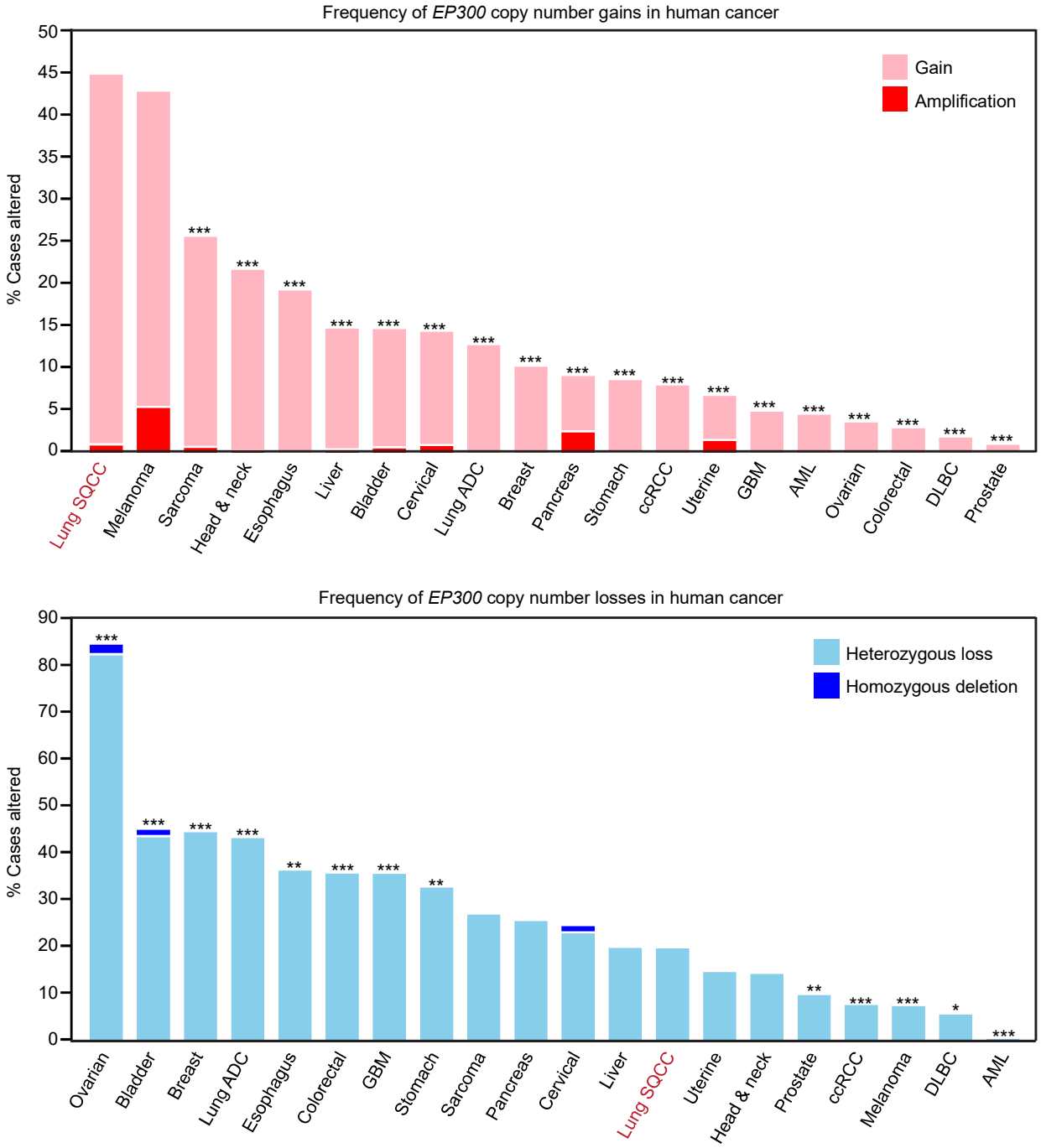
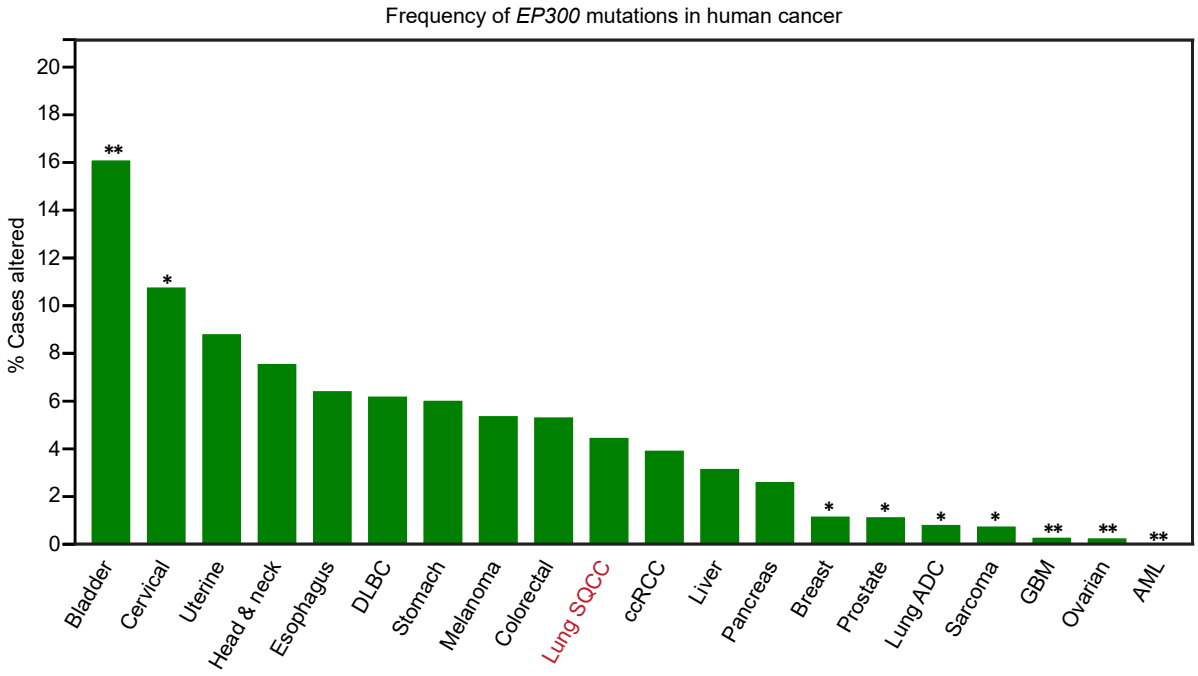

Frequency of *EP300* copy number gains in human cancer
% Cases altered
Gain
Amplification
Lung SQCC
Melanoma
Sarcoma
Head & neck
Esophagus
Liver
Bladder
Cervical
Lung ADC
Breast
Pancreas
Stomach
ccRCC
Uterine
GBM
AML
Ovarian
Colorectal
DLBC
Prostate
Frequency of *EP300* copy number losses in human cancer
% Cases altered
Heterozygous loss
Homozygous deletion
Ovarian
Bladder
Breast
Lung ADC
Esophagus
Colorectal
GBM
Stomach
Sarcoma
Pancreas
Cervical
Liver
Lung SQCC
Uterine
Head & neck
Prostate
ccRCC
Melanoma
DLBC
AML

Fig. S5


Frequency of *EP300* mutations in human cancer
% Cases altered
0
2
4
6
8
10
12
14
16
18
20
**
*
*
*
*
**
**
**
Bladder
Cervical
Uterine
Head & neck
Esophagus
DLBC
Stomach
Melanoma
Colorectal
Lung SQCC
ccRCC
Liver
Pancreas
Breast
Prostate
Lung ADC
Sarcoma
GBM
Ovarian
AML