# BCL11A interacts with SOX2 to control the expression of epigenetic regulators in lung squamous cell carcinoma

**Kyren A. Lazarus[1,7], Fazal Hadi[1,7], Elisabetta Zambon[1,7], Karsten Bach[1,7], Maria-Francesca Santolla[1,3], Julie K. Watson[6], Lucia L. Correia[2], Madhumita Das[8], Rosemary Ugur[1,7], Sara Pensa[1,7], Lukas Becker[1], Lia S. Campos, Graham Ladds[1], Pentao Liu[4], Gerard Evan[2], Frank McCaughan[2], John Le Quesne[8,9,10], Joo-Hyeon Lee[6], Dinis Calado[5] and Walid T. Khaled[1,7]**

**1. Department of Pharmacology, University of Cambridge, Cambridge, UK**

**2. Department of Biochemistry, University of Cambridge, Cambridge, UK**

**3. Department of Pharmacy, Health and Nutritional Sciences, University of Calabria, Rende, Italy**

**4. Wellcome Trust Sanger Institute, Cambridge, UK**

**5. The Francis Crick Institute, London, UK**

**6. WT-MRC Stem Cell Institute, University of Cambridge, UK**

**7. Cambridge Cancer Centre, Cambridge, UK**

**8. MRC Toxicology Unit, Lancaster Road, Leicester**

**9. Cancer Research Centre, University of Leicester**

**10. University Hospitals Leicester NHS trust**

**Correspondence to: Walid T. Khaled (wtk22@cam.ac.uk)**

**Abstract**

Patients diagnosed with lung squamous cell carcinoma (LUSC) have limited targeted therapeutic options. We report here the identification and characterisation of the transcriptional regulator, *BCL11A,* as a LUSC oncogene. Analysis of cancer genomics datasets revealed *BCL11A* to be upregulated in LUSC but not in lung adenocarcinoma (LUAD). We demonstrated that knockdown of *BCL11A* in LUSC cell lines abolished xenograft tumour growth and its overexpression *in vivo* led to lung airway hyperplasia and the development of reserve cell hyperplastic lesions. In addition, deletion of *Bcl11a* in the tracheal basal cells abolished the development of tracheosphere organoids while its overexpression led to solid tracheospheres with a squamous phenotype. At the molecular level we found *BCL11A* to be a target of SOX2 and we show that it is required for the oncogenic role of SOX2 in LUSC. Furthermore, we showed that *BCL11A* and SOX2 interact at the protein level and that they regulate the expression of several transcription factors, including *SETD8*, *SKIL* and *TBX2*. We demonstrate that shRNA-mediated or pharmacological-mediated inhibition of SETD8 selectively affects *in vitro* and xenograft growth of LUSC cells in comparison to LUAD. Collectively, our study indicates that BCL11A is integral to LUSC pathology and that the disruption of the BCL11A-SOX2 transcriptional program could provide a future framework for the development of targeted therapeutics.

## Main

Lung cancer accounts for the highest rate of cancer related diagnosis and mortality worldwide[1]. Broadly, there are two major types of lung cancer; small cell lung cancer (SCLC) accounting for 15% of cases and non-small cell lung cancer (NSCLC) accounting for 85% of cases[1]. NSCLC patients have a poor outcome in the clinic with only 15% of patients surviving five years or more[2]. At present NSCLC is defined histo-pathologically in the clinic into four broad categories: lung adenocarcinoma (LUAD), lung squamous cell carcinoma (LUSC), large cell carcinoma and undifferentiated non-small cell lung cancer. LUAD and LUSC are the most common types of NSCLC (90% of cases). LUSC accounts for more than 400,000 deaths worldwide each year[3] and unlike LUAD there are limited targeted therapies available for LUSC. Platinum-based chemotherapy remains the first-line treatment for LUSC and although the recent FDA approval of Necitumumab in combination with platinum-based chemotherapy for metastatic LUSC has shown positive signs, a great deal of work still needs to be done in this field[4].

At the cellular level, LUAD tends to originate from the secretory epithelial cells in the lung while LUSC usually originates from basal cells in the main and central airways[2]. At the molecular level LUAD is known to harbour mutations in epidermal growth factor receptor (EGFR), V-Ki-Ras2 Kirsten Rat Sarcoma Viral Oncogene Homolog (KRAS) and Anaplastic Lymphoma Receptor Tyrosine Kinase (ALK), which are also well modelled and studied both *in vitro* and *in vivo* (for review see [5,6]). On the other hand, LUSC is less studied but it is known that amplifications of SRY (Sex Determining Region Y)-Box 2 (SOX2) tend to be present in 70-80% of patients[7-10].

Recently, a detailed picture of the molecular differences between LUAD and LUSC has been made available through 'The Cancer Genome Atlas' (TCGA)[11,12]. To identify key drivers responsible for the differences between LUAD and LUSC we reanalysed the gene expression data from TCGA and focused on transcriptional regulators in the genome. As reported previously *SOX2* was the most amplified gene in LUSC and its expression level was also significantly higher in LUSC vs. LUAD (Figure 1a and Supplementary Fig. 1a). The second most amplified locus in LUSC patients revealed by TCGA analysis contains the transcription factors *BCL11A* and *REL*[11,12]. *BCL11A* has been shown to be an oncogene in B-cell lymphoma and triple negative breast cancer[13-16].

We found that *BCL11A* expression levels were also significantly higher in LUSC vs. LUAD (Figure 1a and Supplementary Figure 1a). Moreover, the expression of both *BCL11A* and *SOX2* was significantly higher in LUSC but not in LUAD tumour samples compared to patient matched normal samples (Figure 1b-c and Supplementary Figure 1b-c) supporting a driver role for these transcription factors in LUSC pathology. In contrast, *REL* expression was unchanged between LUSC and LUAD (Figure 1a-c and Supplementary Figure 1a-c) suggesting that *BCL11A* amplification is a key driving event in LUSC. This observation is supported by the recent report from the TRACERx (TRAcking Cancer Evolution through therapy (Rx)) study demonstrating the amplification of *BCL11A* as an early event in LUSC[17]. Furthermore, BCL11A IHC staining on LUAD (n=99) and LUSC (n=120) TMAs revealed little or no staining in 99% of LUAD patients. In contrast, 25% of LUSC patients had moderate staining and 24% of LUSC patients had strong staining, which is in agreement with previous IHC staining of NSCLC tumours[18] (Figure 1d). This confirms our transcriptomic analyses indicating the specificity of BCL11A in LUSC patients.

To determine if high levels of *BCL11A* expression are oncogenic in LUSC, we performed shRNA mediated knockdown (KD) of *BCL11A* using two independent shRNAs in two LUSC cell lines, LK2 and H520 (Supplementary Fig. 2a and b). We first tested the clonogenic capacity of control or *BCL11A-KD* cells by seeding them into matrigel for 3D colony formation assays. We found that *BCL11A-KD* cells had a significant reduction in colony formation capacity (Supplementary Fig. 2c and d). We then injected control or *BCL11A*-KD cells into immune compromised mice to test for their tumour formation capacity and found a significant reduction in tumour burden from *BCL11A-KD* cells compared to control cells (Fig 1e and f). In addition, we found the squamous markers *KRT5* and *TP63* levels were significantly reduced in *BCL11A*-KD LUSC cells (Supplementary Figure 2e-h). However, we found no change in the expression of *SOX2* in *BCL11A*-KD LUSC cells suggesting that BCL11A activity is downstream of SOX2 (Supplementary Figure 2i and j). Moreover, to test if the role of BCL11A is context dependant we knocked down *BCL11A* in a LUAD cell line H1792 and found no change in 3D colony growth indicating specificity at the cellular level (Supplementary Figure 2k-l).

To explore the role of BCL11A in lung biology, we utilised a novel Cre-inducible mouse model that allows for the overexpression of *BCL11A*. Essentially, *BCL11A* was inserted into the *Rosa26* locus with a LoxP-Stop-LoxP (*LSL*) cassette upstream, under the control of a CAGG promoter, thus preventing the expression of *BCL11A* unless the *LSL* is excised by Cre recombinase. To test the effect of *BCL11A* overexpression on lung morphology, we allowed the *BCL11A*-overexpression ($BCL11A^{ovx}$) mice to inhale Adenovirus-Cre (Figure 2a). Eight months after infection, we analysed the lungs and found signs of airway

hyperplasia (Figure 2b and c) accompanied by aggregates of small hyperchromatic cells with irregular nuclei that represent reserve cell hyperplasia (Figure 2b, arrows and inset) which are precursors to squamous metaplasia[19]. IHC analysis of the lungs from *BCL11A*$^{ovx}$ also indicated an increase in positivity for the proliferative marker Ki-67 (Supplementary Figure 3a) and Sox2 indicating a transition to squamous differentiation (Supplementary Figure 3b). However, we found little difference in Cc10, Krt5 and Trp63 staining at this stage (Supplementary Figure 3a and b).

To further investigate the role of *BCL11A* in LUSC, we employed a 3D organoid culture system for basal cells (BCs) from the trachea, as they have been suggested to be the cell of origin for LUSC[20-22]. BCs from human and mouse lungs have higher expression of *BCL11A* when compared to the other epithelial compartments indicating its importance in lung biology (Supplementary Figure 4a and b)[22]. Therefore, we crossed the *BCL11A*$^{ovx}$ mice to *R26-CreERT2* mice, which allowed us to induce Cre recombination by the administration of tamoxifen. In addition, we also used a *Bcl11a* conditional knockout mouse under the control of *R26-CreERT2* (called *Bcl11a*$^{CKO}$ from this point) to elucidate the importance of BCL11A in tracheosphere organoid formation (Figure 2a).

We found that in contrast to the hollow organoids normally formed by *BCL11A*$^{OVX}$ or *Bcl11a*$^{CKO}$ organoids treated with ethanol *in vitro*, BCs from *BCL11A*$^{ovx}$ mice treated with tamoxifen *in vitro* formed solid organoid structures with no hollow lumen suggesting hyper-proliferation and loss of organisation (Figure 2d). However, BCs from *Bcl11a*$^{cko}$ mice failed to form any organoid structures suggesting that *Bcl11a* is necessary for organoid formation (Figure 2d). Quantitative analysis revealed a significant decrease in organoid numbers from *Bcl11a*$^{cko}$ BCs but no significant difference in *BCL11A*$^{ovx}$ BCs

(Supplementary Figure 4c and d). Tamoxifen or ethanol treated organoids from WT mice showed no difference in organoid morphology indicating that the differences are attributed to changes in *BCL11A* expression (Supplementary Figure 4e)

H&E staining confirmed the hollowness of the organoids from the control mice which was in stark contrast to the filled organoids from the *BCL11A*$^{ovx}$ mice (Figure 2e and Supplementary Figure 5). IF staining revealed that the *BCL11A*$^{ovx}$ organoids were also positive for Ki-67 (Supplementary Figure 5), Sox2, Krt5, Trp63 and negative for the luminal marker Krt8 (Figure 2e and Supplementary Figure 6) indicating that BCL11A maintains a squamous phenotype.

Given the importance of SOX2 in driving LUSC[8,9] we next investigated if *BCL11A* is regulated by SOX2. To achieve this, we knocked down *SOX2* (*SOX2-KD*) in two LUSC cell lines using two independent shRNAs (Figure 3a and b and Supplementary Figure 7a and b). Similar to *BCL11A* we found that *SOX2-KD* cells had a significantly reduced colony and tumour formation abilities (Supplementary Figure 7c-f). At the molecular level we found a significant reduction in the expression levels of *BCL11A* and similar to *BCL11A-KD*, reduction in squamous markers *KRT5* and *TP63* in the *SOX2-KD* cells (Figure 3c, d and Supplementary Figure 7g-j).

To investigate if BCL11A is required for SOX2 mediated oncogenesis we introduced a doxycycline inducible *BCL11A* overexpression vector into *SOX2-KD* cell lines and found that *BCL11A* overexpression partially rescues the colony and tumour formation abilities of *SOX2-KD* cells (Figure 3e-h, and Supplementary Figure 8a-c). These results suggest

169 that BCL11A is partially responsible for SOX2's mediated transcriptional changes in
170 LUSC cells.
171
172 To understand how BCL11A and SOX2 contribute to the LUSC transcriptional program
173 we performed BCL11A and SOX2 ChIP-Seq analysis on LK2 cells in the presence or
174 absence of BCL11A (*BCL11A-KD*) (Figure 3i). We identified 49,567 peaks for SOX2
175 and 4,294 peaks for BCL11A (Figure 3j and Supplementary Tables 1 and 2). Out of the
176 4,294 BCL11A peaks identified, 3,946 were not present in *BCL11A-KD* cells validating
177 the true nature of BCL11A binding at these regions of the genome.
178
179 We then compared regions bound by both BCL11A and SOX2 in LK2 control cells and
180 identified 1,114 peaks suggesting that a relationship between these two transcription
181 factors (Figure 3i and j). Subsequently, we identified the nearest genes to the common
182 peaks (Supplementary Table 3) and performed Gene Ontology (GO) analysis to identify
183 if these common peaks are enriched for specific biological process (Figure 3j). The top
184 hits from the GO analysis revealed enrichment for transcriptional and epigenetic
185 regulators including *SETD8, SKIL,* and *TBX2* (Figure 3k and l). These three factors have
186 been reported to have roles in NSCLC. *SKIL* and *TBX2* have been reported to be
187 upregulated in NSCLC[23,24]. SETD8 on the other hand has been indirectly linked to
188 NSCLC as a target of mir-382 in NSCLC[25]. We also found that SOX2 binds the *BCL11A*
189 locus at multiple sites suggesting a strong direct regulation at the transcriptional level
190 further supporting the data in Figure 3c and d (Figure 3l). BCL11A and SOX2 peaks on
191 *SETD8*[26,27], *SKIL*[28], *TBX2*[29] and *BCL11A* were validated and confirmed by ChIP-qPCR
192 (Supplementary Figure 9). The overlap of the ChIP-Seq peaks suggests a direct

interaction between BCL11A and SOX2 proteins, which was confirmed in co-immunoprecipitation experiments on LK2 and H520 (Supplementary Figure 10a, b).

To further asses the importance of these three factors we first analysed the expression of *SETD8, SKIL* and *TBX2* in multiple NSCLC cell lines and found that *SETD8* and *SKIL* expression was significantly higher in LUSC cell lines (n=5) compared to LUAD cell lines (n=6) (Figure 4a and b) which was in correlation with the expression levels of *BCL11A* and *SOX2* (Figure 4d and e). In contrast, *TBX2* levels were not different in LUSC cell lines vs LUAD cell lines (Figure 4c). To understand if BCL11A regulates these genes, we sorted basal stem cells ($Epcam^{+ve}/GSI\beta4^{+ve}$)[30,31] and $Epcam^{+ve}$ cells from trachea excised from tamoxifen treated $BCL11A^{ovx}$ or WT mice (Figure 4f, g and k). We found that *Setd8* and *Skil*, but not *Tbx2* and *Sox2* are upregulated in both basal stem cell and epithelial cell compartment from the trachea of $BCL11A^{ovx}$ mice (Figure 4g-k and l-p). These results provide the first evidence that BCL11A regulates *Setd8* and *Skil* independent of SOX2 driven mechanisms.

We then investigated the expression of *SETD8*, *SKIL* and *TBX2* in *BCL11A-KD* or *SOX2-KD* cells and found a modest reduction in the expression of these genes in *BCL11A-KD*, which was more pronounced in *SOX2-KD* cells (Supplementary Figure 10c-n). We reasoned that this could be due to SOX2 redundantly regulating *Setd8* in *BCL11A-KD* cells.

Collectively, our results thus far suggest that disrupting the BCL11A-SOX2 transcriptional program could be selectively detrimental to LUSC cells. To test this hypothesis we employed a dox-inducible shRNA system [32], to knockdown *SETD8*, *SKIL*

and *TBX2* in two LUSC cell lines and one LUAD cell line (H520 Supplementary Figure 11a, b, and c; LK2 Supplementary Figure 11d, e, and f; H1792 Supplementary Figure 11g, h, and i). We found that SETD8-KD had a pronounced effect on colony formation selectively in LUSC cells but not in LUAD cells (Figure 5a, b and c). SKIL-KD, also reduced colony formation but affecting both LUSC and LUAD cells suggesting its importance in NSCLC (Figure 5e, f, and g). In contrast, TBX2-KD had no effect on either LUSC or LUAD cells suggesting that it does not play a key role in the pathology of these cells (Figure 5i, j, and k).

Given the selectivity of the SETD8-KD on LUSC vs LUAD, we decided to focus on this gene for further studies. SETD8 is a member of the SET domain containing family and is known to catalyse the monomethylation of histone H4 Lys20[26,27] which is involved in recruiting signalling proteins or chromatin modifications[33]. In agreement with the qPCR data from sorted *BCL11A*$^{ovx}$ cells (Figure 4) we found that SETD8 protein levels were upregulated in airways and hyperplastic lesions of *BCL11A*$^{ovx}$ mice (Supplementary Figure 12a, b). Furthermore, analysis of the TCGA datasets revealed that *SETD8* expression correlates with *BCL11A* and *SOX2* expression in LUSC patients (Supplementary figure 12c and d).

To test the impact of SETD8 on xenograft tumour growth we injected cells with either Scram or *SETD8* shRNA vectors into immune compromised mice. Once tumours reached 0.25cm$^2$ in size, we supplemented the mouse diet with doxycycline to induce the *SETD8* knockdown and measured tumour growth periodically. This setup would also mimic a therapeutic intervention in patients presenting with LUSC tumours. We found that initial tumour growth rates were comparable between Scram and shRNA cells across all three

cell lines (Figure 5d, h and l and Supplementary Figure 13a-c) However, upon the addition of doxycycline tumours from  tumours formed by the *SETD8-KD* cells slowed down significantly (Figure 5d, h, l). At the end of the experiment, tumours formed by the *SETD8-KD* were approximately 50% smaller in size compared to Scram cells (Figure 5d, h and l and Supplementary Figure 13a-c). Interestingly, this effect was only found in the LUSC cell lines and not the LUAD which, is in agreement with the colony assays results.

To expand our analysis further we tested the effect of a SETD8 inhibitor, NSC663284[34], on 11 NSCLC cell lines (5 LUSC and 6 LUAD) *in vitro*. We treated all the NSCLC cell lines with a range of NSC663284 concentrations (full details of setup in materials and methods) for 72 hours and measured cell viability. Remarkably, we found that LUSC cells had significantly lower $IC_{50}$ (average 0.30 μM) compared to LUAD cells (average 7.65 μM) (Figure 6a and b). To understand if SETD8 inhibition would add a clinical benefit to patients, we tested the effect of combining NSC663284 and cisplatin. First, we found that cisplatin treatment for 24hrs alone affected LUSC and LUAD in a similar way with both cell types demonstrating a similar $IC_{50}$ (LUAD = 64.12 μM and LUSC = 31.67 μM) (Figure 6c and d). However, if we pre-treat NSCLC cell lines with NSC663284 for 48hrs and then combine cisplatin with NSC663284 for a further 24hrs we found that LUSC ($IC_{50}$ = 4.66 μM) cells are more sensitive to cisplatin than LUAD cells ($IC_{50}$ = 32.21 μM) (Figure 6e and f). It is important to note that NSC663284 has also been reported as an inhibitor of Cdc25[35-37]. However, our shRNA data which is SETD8 specific corroborate the NSC663284 data and strongly supports the case for exploring SETD8 as a novel target for LUSC treatment alone or in combination with other chemotherapeutics.

## Discussion

Currently there is an unmet clinical need for patients with LUSC. Understanding and developing therapies for LUSC lags behind LUAD largely due to the association between tobacco smoking and LUSC incidences. Indeed, cancer genomics studies such as TCGA and TRACERx show that generally LUSC have a mutational signature associated with smoking[12,17]. However, even with the decrease in smoking rates LUSC incidences are still high[38], reinforcing the calls for better molecular understanding and the development of new therapies to tackle the disease. In this study we have demonstrated that the transcription regulator, *BCL11A* is upregulated in LUSC vs LUAD and that it is a direct target of *SOX2*. BCL11A was initially discovered as an oncogene in B-Cell lymphomas[39] and we have recently reported that it is as an oncogene in TNBC[40]. We show that *BCL11A* upregulation leads to early stages of tumour development in the mouse and is critical for tumour maintenance even in the presence of SOX2. Our data suggests that disrupting the BCL11A-SOX2 transcriptional program might provide a selective therapeutic window for LUSC patients. This was demonstrated by the sensitivity of LUSC cell lines to the inhibition of *SETD8*, a gene we identified to be regulated by both BCL11A and SOX2 specifically in LUSC. SETD8 is a member of the SET domain containing family and is known to catalyse the monomethylation of histone H4 Lys20[26,27] which is involved in recruiting signalling proteins or chromatin modifications[33]. In addition, SETD8 has been shown to play a role in maintaining skin differentiation[41] and is dysregulated in multiple cancer types[42-44].We found that *Setd8* was also upregulated in the mouse model of *BCL11A* overexpression in the lung. It will be important in the future to investigate the downstream targets of Setd8 and their role in LUSC development. In conclusion, our results describe an oncogenic role for *BCL11A* in LUSC and provide a potential future framework for tackling the unmet clinical need for LUSC patients.

## Figure Legends

### Figure 1. *BCL11A* is a lung squamous cell carcinoma (LUSC) oncogene

**(a)** Volcano plots of The Cancer Genome Atlas (TCGA) RNAseq data[11,12] indicating that *BCL11A* and *SOX2* are highly expressed in LUSC compared to lung adenocarcinoma (LUAD). The plots show that *REL* is not differentially expressed in LUSC vs LUAD patients. The x-axis represents log2 expression fold-change (FC) in LUSC patients vs LUAD patients and the y-axis represents –log10(pValue). The vertical dashed lines represent FC = 1.0 and the horizontal dashed line represents p value = 0.01. **(b)** Volcano plots indicating that *BCL11A* and *SOX2* are differentially expressed in LUSC patients vs matched normal samples. The plot indicates that *REL* is not differentially expressed in LUSC vs matched normal samples **(c)** Volcano plots indicating that neither BCL11A, SOX2 nor *REL* are differentially expressed in LUAD patients vs matched normal. **(d)** Images and scoring of BCL11A IHC staining on LUAD and LUSC tumours (see Methods for scoring). Graph depicting reduction in tumour size observed when shRNA1 or shRNA2 against *BCL11A* transfected LK2 **(e)** and H520 **(f)** cells are injected subcutaneously into mice compared to control. Five mice per cell line were monitored for 25 days after which tumours were removed and measured. On the right are images showing actual tumours measured. Data presented as mean ± s.d. One way ANOVA with post Dunnett test performed, * indicates $p<0.05$ and ** $p<0.005$ and *** indicates $p<0.001$.

**Figure 2. *BCL11A* overexpression leads to thickening of the airways and abnormal organoid formation**

**(a)** Schematic representing strategy to explore the role of *BCL11A in vivo* and *ex vivo*. Left Panel: Adenovirus-Cre was nasally administered to $BCL11A^{ovx}$ mice and the lungs were analysed after eight months. Right panel: for the tracheosphere organoid model, basal cells from the trachea of either $BCL11A^{ovx}$ or $BCL11A^{cko}$ mice were FACS sorted, embedded in matrigel and analysed after 15 days. Three independent mice were used for each experiment. **(b)** Images of airways from control and $BCL11A^{ovx}$. Arrows indicate small hyperchromatic cells with irregular nuclei. **(c)** Quantificaiton of airway epithelial layer hyperplasia from two control and $BCL11A^{ovx}$ mice. **(d)** Bright field images of organoids from $Bcl11a^{cko}$ and $BCL11A^{ovx}$ mice treated with vehicle or tamoxifen. **(e)** Sectioned organoids from $BCL11A^{ovx}$ mice stained with haematoxylin and eosin, BCL11A, Sox2, Krt5, Trp63 and Krt8. Arrows indicate positive staining. Scale bar indicates 50μm.

**Figure 3. BCL11A and SOX2 occupy independent and common loci in the genome of LUSC cells**

**(a and b)** Western blot showing SOX2 and BCL11A expression in *SOX2-KD* in LK2 **(a)** and H520 **(b)** cells transfected with control (scramble), shRNA1 or shRNA2 vectors. **(c and d)** *BCL11A* expression in *SOX2*-KD *LK2* **(c)** and *H520* **(d)** cells. **(e)** Western blot showing BCL11A rescue in *SOX2*-KD cells. Doxycycline (Dox) inducible *BCL11A* overexpression vector was transfected into control and *SOX2* shRNA1 LK2 cells and Dox treatment was performed for 48 hours. **(f)** Graph depicting 3D matrigel assay in control, *SOX2-KD* and *BCL11A* rescue cells indicating a partial rescue in *SOX2-KD*, *BCL11A* overexpressing cells. **(g)** Graph indicating partial rescue of tumour size from *BCL11A*

overexpressing *SOX2-KD* cells injected subcutaneously. **(h)** Images of actual tumours measured. Four mice per cell line were monitored for 15 days after which tumours were removed and measured. Data presented as mean ± s.d. (n=4). One way Anova test performed, * indicates $p<0.05$ and ** $p<0.005$ and *** indicates $p<0.001$. **(i)** LK2 cell line either transfected with control or shRNA1 vectors were used for BCL11A and SOX2 ChIP-Seq. Heatmaps showing BCL11A only, SOX2 only or common peaks in BCL11A or SOX2 IP in control and *BCL11A*-KD cells. **(j)** Venn diagram indicating the overlap of BCL11A and SOX2 target genes in LK2 cells. BCL11A target genes were derived by BCL11A IP in LK2 control cells. SOX2 target genes were derived from SOX2 IP in LK2 control cells. Image below show top five biological GO terms from GO analysis performed using DAVID. **(k)** IGV genome browser views for *SETD8*, *SKIL*, *TBX2* and *BCL11A*.

**Figure 4. BCL11A regulates both *SETD8* and *SKIL* gene expression**

*SETD8* **(a),** *SKIL* **(b)**, *TBX2* **(c)**, *BCL11A* **(d)**, *SOX2* **(e)** gene expression in NSCLC cell lines. **(f)** Sorting strategy for RNA extraction from WT or $BCL11A^{OVX}$ mice treated with tamoxifen. SETD8 **(g),** *SKIL* **(h)**, *TBX2* **(i)**, *BCL11A* **(j)**, *SOX2* **(k)** gene expression in basal stem cells. *SETD8* **(l),** *SKIL* **(m)**, *TBX2* **(n)**, *BCL11A* **(o)**, *SOX2* **(p)** gene expression in epithelial $Epcam^{+ve}$ cells. Data presented as mean ± s.d. (LUSC n=5 and LUAD n=6; WT n=3 and $BCL11A^{OVX}$ n=3). Student's t-test performed, * indicates $p<0.05$ and ** $p<0.005$ and *** indicates $p<0.001$.

**Figure 5. SETD8-KD has a selective effect on LUSC cell line oncogenic capacity.**

**(a-c)** Comparison of colony numbers in 3D matrigel assay from control, shRNA1 or shRNA2 against *SETD8* **(a)**, *SKIL* **(b)** and *TBX2* **(c)** in H520 cells. **(e-g)** Comparison of

colony numbers in 3D matrigel assay from control, shRNA1 or shRNA2 against *SETD8* **(e)**, *SKIL* **(f)** and *TBX2* **(g)** in LK2 cells. Data presented as mean ± s.d. (n=3). **(i-k)** Comparison of colony numbers in 3D matrigel assay from control, shRNA1 or shRNA2 against *SETD8* **(i),** *SKIL* **(j)** and *TBX2* **(k)** in H1792 cells. Data presented as mean ± s.d. (n=3). **(d, h, l)** Tumour kinetics of control or SETD8-KD cells injected into immune compromised mice. Arrow indicates administration of doxycycline. Data presented as mean ± s.d. One way ANOVA with post Dunnett test performed, * indicates $p<0.05$ and ** $p<0.005$ and *** indicates $p<0.001$.

**Figure 6. SETD8 inhibition preferentially sensitises LUSC cell lines to chemotherapy**

**(a)** Dose-response curves were derived by treating NSCLC cell lines with SETD8 inhibitor NSC663284. 1000 cells were seeded and allowed to recover for 24 hours. The inhibitor was then added at increasing concentrations to LUSC (red) and LUAD (blue) cells and Cell Titre (see Methods) assay was performed after 72 hours. **(b)** $IC_{50}$ values were derived from the dose-response assay indicating LUSC cells are significantly more responsive to SETD8 inhibition than LUAD cells. **(c)** Dose-response curves were derived by treating NSCLC cell lines with cisplatin as above. **(d)** $IC_{50}$ values were derived from the dose-response assay indicating cisplatin effects LUSC and LUAD cells equally. **(e)** Dose-response curves derived from treating NSCLC cell lines with cisplatin and NSC663284 IC50 concentration for each cell line as above. **(f)** $IC_{50}$ values were derived from the dose-response assay indicating SETD8 inhibition preferentially enhances cisplatin efficacy in LUSC cells. Data presented as mean ± s.d. (LUSC n=5 and LUAD n=6). Student's t-test performed, * indicates $p<0.05$ and ** $p<0.005$ and *** indicates $p<0.001$.

**Supplementary Figure 1. BCL11A, SOX2, and REL expression in Lung TCGA dataset**

**(a)** Violin plot showing the distribution of *BCL11A*, *SOX2* and *REL* expression in LUSC vs LUAD patients. FPKM, fragments per kilobase per million mapped reads. **(b)** Violin plot showing the distribution of *BCL11A*, *SOX2* and *REL* expression in LUSC tumour vs normal matched patients. **(c)** Violin plot showing the distribution of *BCL11A*, *SOX2* and *REL* expression in LUAD tumour vs normal matched patients.

**Supplementary Figure 2. BCL11A KD reduces squamous markers in LUSC cells**

**(a and b)** qPCR and western blot shows *BCL11A* reduction in LK2 (**a)** and H520 **(b)** cells transfected with shRNA1 and shRNA2 vectors. **(c and d)** Comparison of colony numbers in 3D matrigel assay from control, shRNA1 or shRNA2 in LK2 **(c)** and **(d)** H520 cells. Data presented as mean ± s.d. (n=3). **(e and f)** *KRT5* expression is reduced in LK2 **(e)** and H520 (f) *BCL11A-KD* cells. **(g and h)** *TP63* expression is reduced in LK2 **(g)** and H520 (h) *BCL11A*-KD cells. **(i and j)** *SOX2* expression is unchanged in LK2 **(i)** and H520 **(j)** *BCL11A-KD* cells. **(k)** qPCR and western blot shows BCL11A reduction in H1792 cells transfected with shRNA1 and shRNA2 vectors. **(l)** Comparison of colony numbers in 3D matrigel assay from control, shRNA1 or shRNA2 in H1792 cells. Data presented as mean ± s.d. One way ANOVA with post Dunnett test performed, * indicates $p<0.05$ and ** $p<0.005$ and *** indicates $p<0.001$.

**Supplementary Figure 3. Airways from the *BCLL1A$^{ovx}$* mice demonstrate proliferative preneoplastic lesions**

**(a)** Immunohistochemistry for BCL11A, Ki67 and Cc10 expression in control vs BCL11A$^{ovx}$ airways. BCL11A panel, arrows indicating positive staining especially in preneoplastic lesions. Ki67 panel, arrows indicating positive staining. **(b)** Sox2, Trp63 and Krt5 expression in control vs *BCL11A$^{ovx}$* airways. Scale bar indicates 50 μm.

**Supplementary Figure 4. Organoids from *BCL11A$^{ovx}$* basal cells exhibit increase in squamous markers**

**(a)** RNA-Seq data analysed from Weeden et al [22] indicating that in human samples FACS fraction labelled as proximal BCs has approximately 400 fold increase in human *BCL11A* levels in comparison to other epithelial fractions. **(b)** RNA-Seq data from the same dataset indicating that mouse *Bcl11a* is approximately 500-1000 fold higher in basal fractions compared to other epithelial subtypes. **(c, d and e)** Quantification of organoids in vehicle vs tamoxifen treated **(c)** BCL11A$^{ovx}$, **(d)** *BCL11A$^{cko}$* and **(e)** WT organoids. Data presented as mean ± s.d. Paired student t-test performed, * indicates $p<0.05$ and ** $p<0.005$ and *** indicates $p<0.001$.

**Supplementary Figure 5. Organoids from the BCL11A$^{ovx}$ basal cells display an abnormal proliferative phenotype**

Vehicle and tamoxifen treated BCL11A$^{ovx}$ organoids stained with H&E, GFP (which is also expressed if the LSL is efficiently excised), Bcl11a and Ki67. Nuclei stained by DAPI illustrated in blue. Arrows indicate positive staining. Scale bar indicates 50μm.

**Supplementary Figure 6. BCL11A overexpression inhibits organoid luminal differentiation**

Vehicle and tamoxifen treated BCL11A$^{ovx}$ organoids stained with Sox2, Krt5, Trp63 and Krt8. Nuclei stained by DAPI illustrated in blue. Arrows indicate positive staining. Scale bar indicates 50 μm.

**Supplementary Figure 7. SOX2 KD affects LUSC cell lines tumourigenicity**

**(a and b)** Graph showing *SOX2* expression in *SOX2-KD* LK2 (a) and H520 (b) cells. **(c and d)** Graph showing decrease in colony numbers in *SOX2-KD* LK2 **(c)** and H520 **(d)** cells. **(e and f)** Graph depicting reduction in tumour size observed when *SOX2* shRNA1 or shRNA2 transfected LK2 **(e)** and H520 **(f)** cells are injected subcutaneously into mice compared with control. Five (LK2) and four (H520) mice per cell line were monitored for 15 days after which tumours were removed and measured. On the right are images showing actual tumours measured. Data presented as mean ± s.d. One way ANNOVA with post Dunnett test performed, * indicates $p<0.05$ and ** $p<0.005$ and *** indicates $p<0.001$. **(g and h)** *KRT5* expression is reduced in LK2 **(g)** and H520 **(h)** *BCL11A*-KD cells. **(i and j)** *P63* expression is reduced in LK2 **(i)** and H520 **(j)** *BCL11A*-KD cells.

**Supplementary Figure 8. BCL11A is required for SOX2 mediated LUSC phenotype**

**(a and b)** Graph showing *BCL11A* **(a)** and *SOX2* **(b)** expression in BCL11A rescue in *SOX2*-KD cells. Dox inducible *BCL11A* overexpression vector was transfected into control and *SOX2* shRNA1 LK2 cells and Dox treatment was performed for 48 hours. **(c)** Images from 3D matrigel experiment showing reduction in colony numbers after *SOX2-KD* and partial rescue after *BCL11A* overexpression.

**Supplementary Figure 9. BCL11A and SOX2 bind to common loci on the genome**

**(a-h)** ChIP-qPCR showing enrichment for the indicated primers on *SETD8, SKIL, TBX2* and *BCL11A* regions after Chip pull down using anti-BCL11A or anti-SOX2 antibodies in LK2 and H520 cell lines. **(i-l)** Schematic of amplicon locations for ChIP-qPCR validations performed in this study. Arrows represent location of primers used. Error bars represent mean ± s.e.m.; n = 3 (technical replicates).

**Supplementary Figure 10. BCL11A and SOX2 co-regulate transcriptional regulators**

**(a and b)** Co-Immunoprecipitation of endogenous BCL11A and SOX2 proteins in H520 **(a)** and LK2 **(b)** cell lines. **(c-f)** Graph depicting *SETD8* gene expression in LK2, H520 *BCL11A-KD* and LK2, H520 *SOX2 KD* cells. **(g-j)** *SKIL* gene expression in *BCL11A*-KD or *SOX*-KD LK2 and H520 cell lines. **(k-n)** *TBX2* gene expression in *BCL11A*-KD or *SOX2*-KD LK2 and H520 cell lines. Data presented as mean ± s.d. (n=3). One way ANOVA with post Dunnett test performed, * indicates $p<0.05$ and ** $p<0.005$ and *** indicates $p<0.001$.

**Supplementary Figure 11. Generation of *SETD8*, *SKIL* and *TBX2* KD cells**

**(a-c)** SETD8 gene expression in SETD8-KD **(a)** , SKIL gene expression in SKIL-KD **(b)**, TBX2 gene expression in TBX2-KD **(c)** H520 cells. **(d-f)** SETD8 gene expression in SETD8-KD **(d)** , SKIL gene expression in SKIL-KD **(e)**, TBX2 gene expression in TBX2-KD **(f)** LK2 cells. **(g-i)** SETD8 gene expression in SETD8-KD **(g)** , SKIL gene expression in SKIL-KD **(h)**, TBX2 gene expression in TBX2-KD **(i)** H1792 cells. Data presented as mean ± s.d. (n=3). One way ANOVA with post Dunnett test performed, * indicates $p<0.05$ and ** $p<0.005$ and *** indicates $p<0.001$.

**Supplementary Figure 12. *SETD8* correlates with *BCL11A* and *SOX2* in LUSC patients**

**(a and b)** SETD8 immunofluorescence in control or BCL11AOVX mouse airways. Scale bar indicates 50 μm. **(c)** Scatter plot showing *SETD8* and *BCL11A* expression in TCGA patient tumour samples. **(d)** Scatter plot showing *SETD8* and *SOX2* expression in TCGA patient tumour samples.

**Supplementary Figure 13. SETD8-KD confers a survival disadvantage to xenograft tumours**

Tumour kinetics of individual mice injected with either control or SETD8-KD vectors in H520 (a), LK2 (b) or H1792 (c) cells.

**Supplementary Figure 14. Uncropped western blot images**

**Supplementary Figure 15. Uncropped western blot images**

**Supplementary Table 1: Nearest downstream genes to BCL11A only peaks**

**Supplementary Table 2: Nearest downstream genes to SOX2 only peaks**

**Supplementary Table 3: Nearest downstream genes to overlap between BCL11A only and SOX2 only peaks**

**Supplementary Table 4: List of primers used for genotyping**

**Supplementary Table 5: List of qPCR primers used in this study**

**Supplementary Table 6: List of CHIP qPCR primers used in this study**

## Materials and Methods

### Mouse models

All mice used in this study were maintained at the Sanger Institute or the University of Cambridge. Housing and breeding of mice and experimental procedures were performed according to the UK 1986 Animals Scientific Procedure Act and local institute ethics committee regulations. The *BCL11A*$^{ovx}$ allele was generated following a strategy previously described[45]. Briefly, the *ROSA26* allele was targeted with a construct containing human *BCL11A* cDNA preceded by a loxP flanked STOP cassette and marked *eGFP* under the control of an internal ribosomal entry site (IRES) downstream of the inserted cDNA and transgene transcription is controlled by the *CAGG* promoter. The generation of the *Bcl11a*$^{cko}$ mice was described previously[16]. All mice were 8–12 weeks of age at the time of experiments, and at least 3 mice per cohort were used in each experiment. The primers used for genotyping are listed in Supplementary Table 4.

### Mouse Tracheal Isolation, BASC isolation and tracheoshpere culture

Tracheae were incubated in 50 U/ml dispase (Sigma) for 45 minutes at 37$^{o}$C. 10 ml PBS was injected through each trachea using a 25G 5/8" needle to flush out sheets of epithelial cells. Cells were incubated in 0.25% trypsin for 5 minutes at 37$^{o}$C.

For isolating basal stem cells, mice were first injected intraperitoneally with tamoxifen (Sigma) dissolved in corn oil. Trachea was isolated as described above and cells were stained in PBS+10%FBS with 1:500 anti-EpCAM BV421 (Biolegend) and GSIβ4-Biotin. Basal cells were considered Epcam$^{+ve}$/GSIB4$^{+ve}$ [30,31]. Cells were sorted directly into RLT lysis buffer for mRNA isolation.

For tracheosphere culture, cells were stained in PBS + 10%FBS + 1:100 anti-EpCAM PE-Cy7 (Biolegend) + 1:5000 DAPI on ice for 20 minutes. Live EpCAM$^{+ve}$ cells were isolated using a MoFlo sorter. 2500 cells were plated in 100µl 1:1 mix of Mouse Tracheal Epithelial Cell (MTEC)/Plus media which is DMEM-Ham's F-12 (1:1 vol/vol), 15 mM HEPES, 3.6 mM sodium bicarbonate, 4 mMl-glutamine, 100 U/ml penicillin, 100 µg/ml streptomycin, and 0.25 µg/ml fungizone; supplemented with 10 µg/ml insulin, 5 µg/ml transferrin, 0.1 µg/ml cholera toxin, 25 ng/ml epidermal growth factor (Becton-Dickinson, Bedford, MA), 30 µg/ml bovine pituitary extract, 5% FBS, and freshly added 0.01 µM retinoic acid[46] and growth factor reduced-Matrigel (Corning) per 24-well insert, in duplicate with 500 nM 4-hydroxy tamoxifen (Sigma) or ethanol (vehicle) for each mouse, and cultured for 15 days.

**Cell lines**

LK2, NCI-H520 and LUDLU1 cells were generously gifted to us by Dr Frank McCaughan. NCI-H157, LUDLU1, SW900, NCI-H1792, NCI-H522, NCI-HCC78, A549, NCI-H1563 and NCI-H1975 were obtained from the Sanger Cancer Project. All these cells were all maintained in RPMI 1640 (Gibco), 10% FCS and 1% Penicillin/streptomycin in a 37$^{o}$C incubator with 5% $CO_2$. All cell lines were routinely tested for mycoplasma.

**ShRNA mediated knockdown**

*BCL11A* shRNA sequences were obtained from TRC consortium (TRCN0000033449 and TRCN0000033453) and cloned into a *piggyBac* transposon vector (PB-H1-shRNA-GFP) as describe previously[40]. Sox2 shRNA sequences (TRCN0000355694 and TRCN0000257314) were also cloned as above. H520, H1792 and LK2 cells were

transfected with 4.0 μg of respective vector using Lipofectamine 3000 or Lipofectamine LTX (Invitrogen). Cells were treated with G418 (400 μg/mL) (Gibco) for 5 days after which $GFP^{+ve}$ cell were sorted and cultured.

**Doxycycline inducible shRNA mediated knockdown**

SETD8 shRNA (TRCN0000359304 and TRCN0000359373), SKIL shRNA (TRCN0000431894 and TRCN0000424201); and TBX2 shRNA (TRCN0000232147 and TRCN0000232146) were cloned into pLKO-Tet-On vector [32]. Lentiviruses were generated by co-transfecting HEK293T cells with 3μg of shRNA-encoding plasmid and 1μg each of pMD2.G, pMDLg/pRRE and pRSVRev plasmids using Lipofectamine LTX. Growth media was exchanged after 5hrs and lentivirus-containing supernatant was harvested after 48hrs. The supernatant was filtered using a 0.45μm filter cartridge and then Lenti-X concentrator was added. The solution was incubated O/N at 4oC and then centrifuged at 1,500xg for 45 mins. The pellet was then resuspended in $1/20^{th}$ of the original volume using RPMI media. Cells were then infected with the virus for 48hrs and selected in 1μg/ml puromycin. Cells infected with virus were grown in RPMI supplemented with 10% Tet-spproved FBS (Clontech) and 1μg/ml puromycin. shRNA expression was induced by culturing cells in the presence of 1μg/ml doxycycline for 72hrs.

**Transfection and 3D colony assays**

The control or the *BCL11A* overexpression piggybac vectors were delivered into NSCLC cells using Lipofectamine LTX (Invitrogen). Transfected cells were maintained for 48 hours and then cultured in puromycin (1μg/ml). To induce BCL11A expression in LK2 *SOX2-KD* cells, doxycycline (Clonetech) was used at a final concentration of (1 μg/ml).

3D colony assays were performed by suspending 500 cells in matrigel (BD Biosciences) and seeding this cell-matrigel suspension onto a 6-well plate. The plate was then incubated for 15 minutes in 37ºC/5% $CO_2$ to allow hardening of suspension. Growth media was added to the well and changed every 2-3 days for 20 days. All experiments were performed in triplicates.

**Preparation of RNA**

RNA was extracted using the RNeasy mini kit (Qiagen). Cell cultures in T25 flasks were first washed with cold PBS, and 350 µl of RLT was added. Cells were scraped, passed through a 20G syringe five times and RNA was extracted using the RNeasy mini kit (Qiagen) according to manufacturer instructions. DNA was degraded by adding 20U Rnase-free DnaseI (Roche) for 30 minutes at room temperature. DnaseI treatment was performed on columns.

**Preparation of cDNA and qRT-PCR**

1 µg of total RNA was diluted to a final volume of 11 µl. 2 µl of random primers (Promega) were added after which the mixture was incubated at 65ºC for 5 mins. A master mix containing Transcriptor Reverse Transcriptase (Roche), Reverse Transcriptase buffer, 2 mM dNTP mix and RNasin Ribonuclease Inhibitors (Promega). This mixture was incubated at 25 ºC for 10 minutes, then 42ºC for 40 minutes and finally 70ºC for 10 minutes. The resulting cDNA was then diluted 1:2.5 in $H_2O$ for subsequent use. qPCR was performed using a Step-One Plus Real-Time PCR System (Thermofisher Scientific). Either Taqman (ThermoFisher Scientific) probes with GoTaq Real Time qPCR Master Mix (Promega) or primers (Sigma) with PowerUp SYBR Green Master Mix (ThermoFisher Scientific) were used. All probe and primer details can be found in

Supplementary Table 5 and 6. The enrichment was normalised with control mRNA levels of GAPDH and relative mRNA levels were calculated using the ΔΔCt method comparing to control group.

**Western Blot**

Cells were lysed using RIPA (Cell signalling) and protease inhibitors (Roche) as per manufacturer instructions. Total protein was measured using the bicinchoninic acid (BCA) method (Pierce Biotechnology). In total, 50 mg cell lysate was separated using 7.5% SDS-PAGE gels and transferred to PVDF membranes by electro-blotting. Membranes were blocked in 5% (w/v) milk in Tris-buffered saline containing 0.05% Tween-20 (TBST). Blots were then incubated at 4$^{o}$C overnight with primary antibodies as indicated, washed in TBST and subsequently probed with secondary antibodies for 1h at room temperature. ECL solution was then added to the membrane and analysed. Antibodies used were, anti-BCL11A (ab191401, Abcam, 1:3000), anti-SOX2 (ab97959, Abcam, 1:2000) and anti-TUBULIN (ab7291, Abcam, 1:10000).

**Co-Immunoprecipitation**

Cells were lysed using RIPA (Cell signalling) and protease inhibitors (Roche) as per the manufacturer's instructions. Total protein was measured using the BCA method as above (Pierce Biotechnology). Briefly, 500µg cell lysates were pre-cleared for 3h at 4°C to remove nonspecific binding. Then, the pre-cleared lysates were incubated with anti-BCL11A (Bethyl, A300-382A) and SOX2 (R&D Systems, AF2018) or control IgG at 4°C overnight. Next day 50µl of Dynabeads Protein G (Thermo Fisher Scientific) were added to each sample. After 3h, the complex was washed three times with RIPA buffer, and then analysed by Western Blot performed as described above.

643

644 **Histology, Immunohistochemistry and Immunofluorescence**

645 Cultured organoid were fixed with 4% paraformaldehyde in PBS for 4h at room
646 temperature. After rinsing with PBS, fixed colonies were immobilised with Histogel
647 (Thermo Scientific) for paraffin embedding. 5μm sections of lung tissues or embedded
648 colonies were stained with haematoxylin and eosin (H&E) and immunostained with
649 antibodies for BCL11A (IHC - ab191402, Abcam,1:1000), SOX2 (IHC – ab97959,
650 1:1000; IF – 14-9811-82, eBioscience, 1:200), Ki67 (IHC - MA5-14520, Thermo
651 Scientific, 1:1000; IF - ab16667, Abcam, 1:300), GFP (IF - ab13970 Abcam, 1:1000),
652 Keratin 8 (IF - TROMA-I, DSHB, 1:100), Keratin 5 (IHC – ab52635, Abcam, 1:1000; IF
653 - 905501, Biolegend, 1:1000), P63 (IHC – ab735, Abcam, 1:200; IF – ab735, Abcam,
654 1:200), CCP10 (IHC – sc25555, Santa Cruz Biotechnology, 1:500), SETD8 (IF – ab3798,
655 Abcam 1:100). IHC secondary staining involved an HRP-conjugated donkey anti-rabbit
656 or donkey anti-mouse secondary (1:250, Thermo Scientific) and were detected using
657 DAB reagent (Thermo Scientific). IF secondary staining involved goat anti-chicken 488,
658 goat anti-rabbit 647, goat anti-rat 647, goat anti-rabbit 488 and goat anti-mouse 647
659 (1:2000, Invitrogen). Nuclear stain was detected using Haematoxylin (IHC) or ProLong
660 Gold Antifade Mountant DAPI (Thermofisher, P36941) (IF). IHC images were acquired
661 using a Zeiss Axioplan 2 microscope and IF images were acquired using a Leica TCS
662 SP5 confocal microscope and analysed on Image J.

663

664 **Xenograft tumour assays**

665 One million H520 and LK2 cells were suspended in 25% matrigel (BD Biosciences) and
666 HBSS. This mixture was subcutaneously injected in 6-12 week old NSG mice. To induce
667 BCL11A overexpression in *SOX2-KD* xenografts or to induce shRNA expression, mice

were fed doxycycline pellets (Envigo, TD.01306, 625mg/kg). Tumours were measured blindly by animal technicians who did not know what was injected into the specific mouse. Mice were culled as specified in figure legends and resulting tumours were analysed.

**ChIP-Seq and ChIP-qPCR**

ChIP-Seq experiments were performed as described[47]. Antibodies used were BCL11A (Bethyl, A300-382A) and SOX2 (R&D Systems, AF2018). Briefly 2 x 15cm plates per cell line were formaldehyde crosslinked, nuclear fraction was isolated and chromatin sonicated using Bioruptor Pico (Diagenode). IP was performed using 100µl of Dynabeads Protein G (Thermo Fisher Scientific) and 10µg of antibody. The samples were then reverse crosslinked and DNA was eluted using Qiagen MinElute column. Sample was then processed either for sequencing or qPCR. Primers used for qPCR are listed in Supplementary Table 6.

**Drug assays**

SETD8 inhibitor NSC663284 (Cayman Chemical Company, 13303) was suspended using DMSO in 10mM stock concentration. Cisplatin (LKT Laboratories, C3374) was suspended in 154mM NaCl at a 3mM stock concentration. Cells were cultured as above and seeded at 1000 cells per 96-well plate and left to recover for 24h. The edges of the 96-well plate were avoided to ensure accuracy in measurement. For NSC663284, an initial dilution of 1:100 from stock was performed in RPMI media for the first concentration of 10-4M. Half log dilutions were performed in RPMI media reaching 10-6M and after which full log dilutions were performed reaching 10-10M. For cisplatin, initial dilution in RPMI media were made to achieve 100µM and 75µM. The 100µM

693 solution was used to make the following solutions 50μM, 25μM, 12.5μM, 6.75μM,
694 3.75μM, 1.5μM and 0.5μM. Cells were treated with vehicle for 48h after which the above
695 doses of cisplatin were added. For cisplatin + NSC663284 experiment the NSC663284
696 IC50 concentration for each cell line was calculated and added initially for 48hrs and then
697 added again along with the cisplatin dilution series as above for 24hrs. Data analysis for
698 drug inhibitor assays was performed in GraphPad Prism 7.02 (San Diego, CA). Data were
699 fitted to obtain concentration-response curves using the three-parameter logistic equation
700 (for pIC50 values). Emax was constrained to 100% while the basal (Emin) parameter was
701 contrained to 20%. Statistical differences were analyzed using one-way ANOVA or
702 Student's t test as appropriate with post hoc Dunnett's multiple comparisons, and $p < 0.05$
703 was considered significant.

704

705 **BCL11A IHC on patient tumours**

706 TMAs contained LUAD (n=99) and LUSC (n=120) cases of archival primary pulmonary
707 tumours collected under East Midlands NRES REC approved project (ref. 14/EM/1159).
708 1mm (n=3) cores are present per case which were initially scored (average nuclear
709 staining intensity) as 0=neg, 1=weak, 2=moderate, 3=strong. A median score was
710 calculated for each case and re classified as 0=neg, 1=moderate, 2=strong. All tissues and
711 data are anonymised to the research team. IHC was performed using BCL11A antibody
712 (ab19487, 1:200) with CC1 antigen retrieval using a Ventana discovery xt. Digital images
713 of stained TMAs were scanned with a Nanozoomer RX instrument and scored on-screen
714 by MD.

715

716 **TCGA gene expression analysis**

The TCGA data was accessed from the recount2 database[48] containing gene level count data from RNA-seq and clinical data from primary tumor samples from patients diagnosed with LUAD and LUSC, respectively. EdgeR[49] was used to test for differential expression of transcription factors (as defined by Tfcheckpoint.org). For this, compositional differences between samples were normalized using the trimmed median of M-values method and a gene-specific dispersion was estimated for each gene. A negative binomial generalized log-linear model was fit to each gene with the covariates "plate+disease_type" (for the LUAD versus LUSC comparison) or "patient+disease_type" (for the tumor versus normal comparison). A log-likelihood ratio test was conducted to test whether the coefficient of the disease_type variable is non-zero, followed by Benjamini-Hochberg adjustment of P values to account for multiple testing. Plate A277 and A278 from the LUAD dataset were removed prior to analysis as they showed a clear separation along the first component in a principal component analysis from all other LUAD samples.

**ChIP-Seq analysis**

ChIP-Seq libraries were sequenced on illumina Hiseq2000 platform at the Wellcome Sanger Institute. Each library was divided into two and sequenced on different lanes. Reads were subsequently run through a pipeline at the sequencing facility to remove adapter sequences and align to the reference genome among others. Alignment was done using mem algorithm in BWA (version 0.7.15) and human_g1k_v37 was used as the reference genome. Aligned and processed reads were received as compressed CRAM files. Samtools (version 1.3.1) was used to decompress the CRAM files and filter uniquely mapped reads in proper pairs. Next, reads from the two runs were combined into a single BAM file using 'merge' function in samtools. Bedtools intersect was then used to remove

reads falling into blacklisted genomic regions or unplaced genomic contigs of the GRCh37 assembly before marking and removing duplicate reads using MarkDuplicates function in Picard tools. Next, DownsampleSam in picard tools was used to sample ~105 million reads from each BAM file. Significantly enriched genomic regions relative to input DNA were identified using MACS2 (version 2.1.1.20160309) with p-value cutoff of 1.00e-05. Heatmaps generation: Mapped read counts were calculated in a 10bp window and normalised as RPKM (Reads Per Kilobase per Million mapped reads) using bamCoverage module from deeptools (version 2.5.1)[50]. This coverage file was used to compute score matrix ±1kb around peak summits using computeMatrix reference-point module (from deeptools version 2.5.1)[50]. Heatmaps of binding profiles around peak summits were then generated using plotHeatmap module in deeptools (version 2.5.1)[50]. Number of overlapping peaks between BCL11A and SOX2 and nearest downstream genes to peaks were determined using ODS and NDG utilities respectively in PeakAnnotator (version 1.4). For annotating nearest downstream genes, Homo sapiens GRCh37 (release 64) from ensembl was used.

**Data and code availability**

All data and source code will be available on request.

**Author contribution**

K.A.L. designed and performed the majority of the experiments and analysed most of the data. F.H. analysed the ChIP-seq data. E.Z. performed the ChIP-qPCR experiments. K.B. analysed the TCGA data. S.P. performed some cell line work. R.U assisted with the BCL11A rescue experiment. M.F.S performed Co-IP experiment. L.B. measured the airway thickness. L.S.C. characterised the lung pathology of adenovirus mice. G.L

designed and analysed the drug assays. J.K and J.H.L performed the tracheoshpere organoid experiment. L.L.C and F.M performed the mouse tissue IHC. M.D and J.LQ performed and analysed the BCL11A IHC on patient tumours. P.L. assisted with the NGS sequencing and provided the *Bcl11a*$^{cko}$ mice. Adenovirus Cre administration was performed under G.E. supervision. D.C. generated the *BCL11A*$^{ovx}$ mice. W.T.K conceptualised and supervised the study. K.A.L and W.T.K wrote the manuscript.

**Acknowledgements**

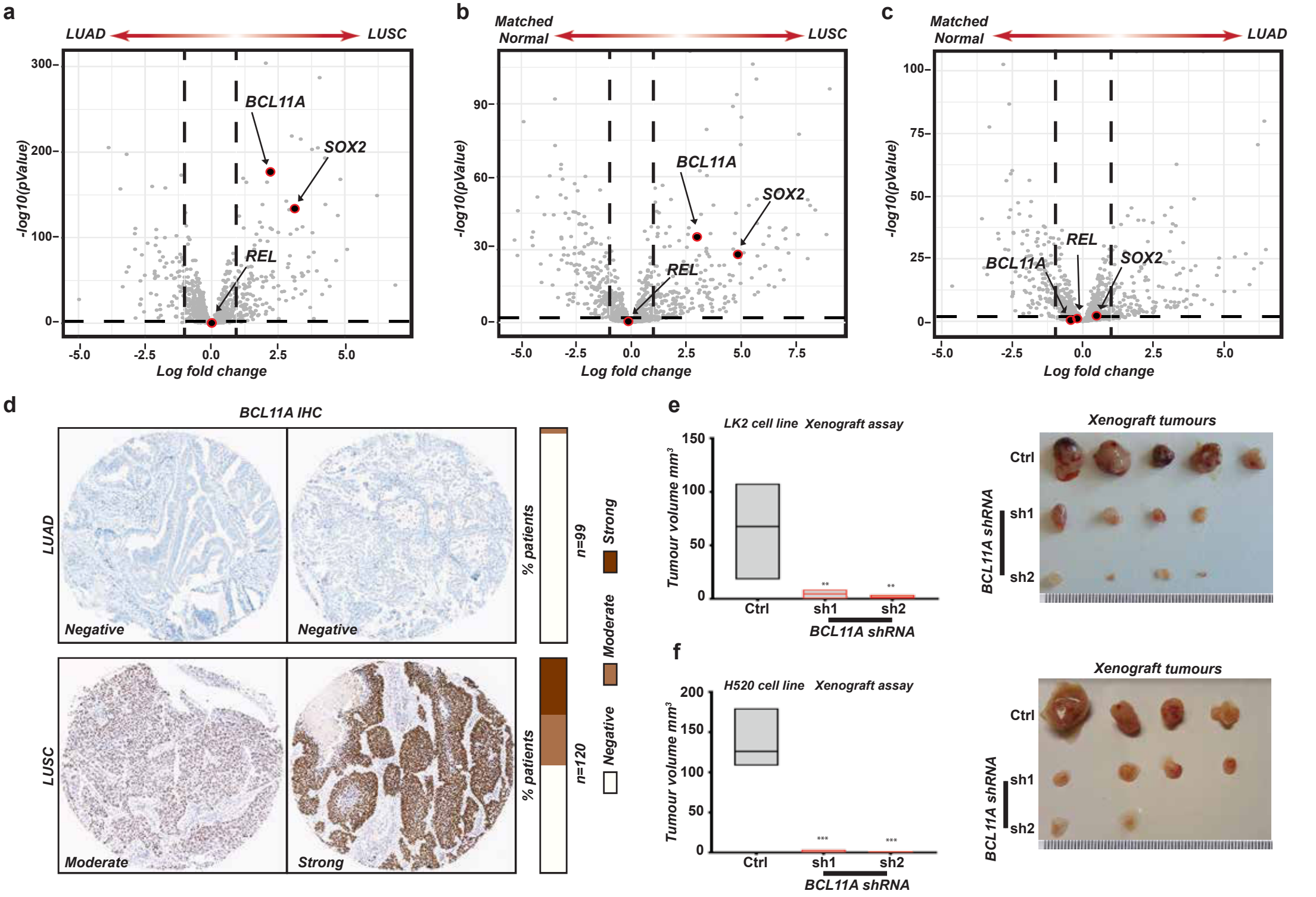
We would like to thank the staff at Sanger Institute, Research Service Facility (RSF) for their assistance. We would like to thank Dr. Catherine Wilson and Dr. Deborah Burkhart (Department of Biochemistry, Cambridge) for her help with the Adenovirus experiment. We would like to thank Dr. Emma Rawlins for helpful discussions and comments. We would like to thank Agnetta Lazarus and members of the WTK Lab for comments on the manuscript. G.L. was supported by the BBSRC (Grant BB/M00015X/2). W.T.K is funded by a CRUK career establishment award (C47525/A17348), University of Cambridge, The Isaac Newton Trust Grant (16.38c) and Magdalene College, Cambridge.

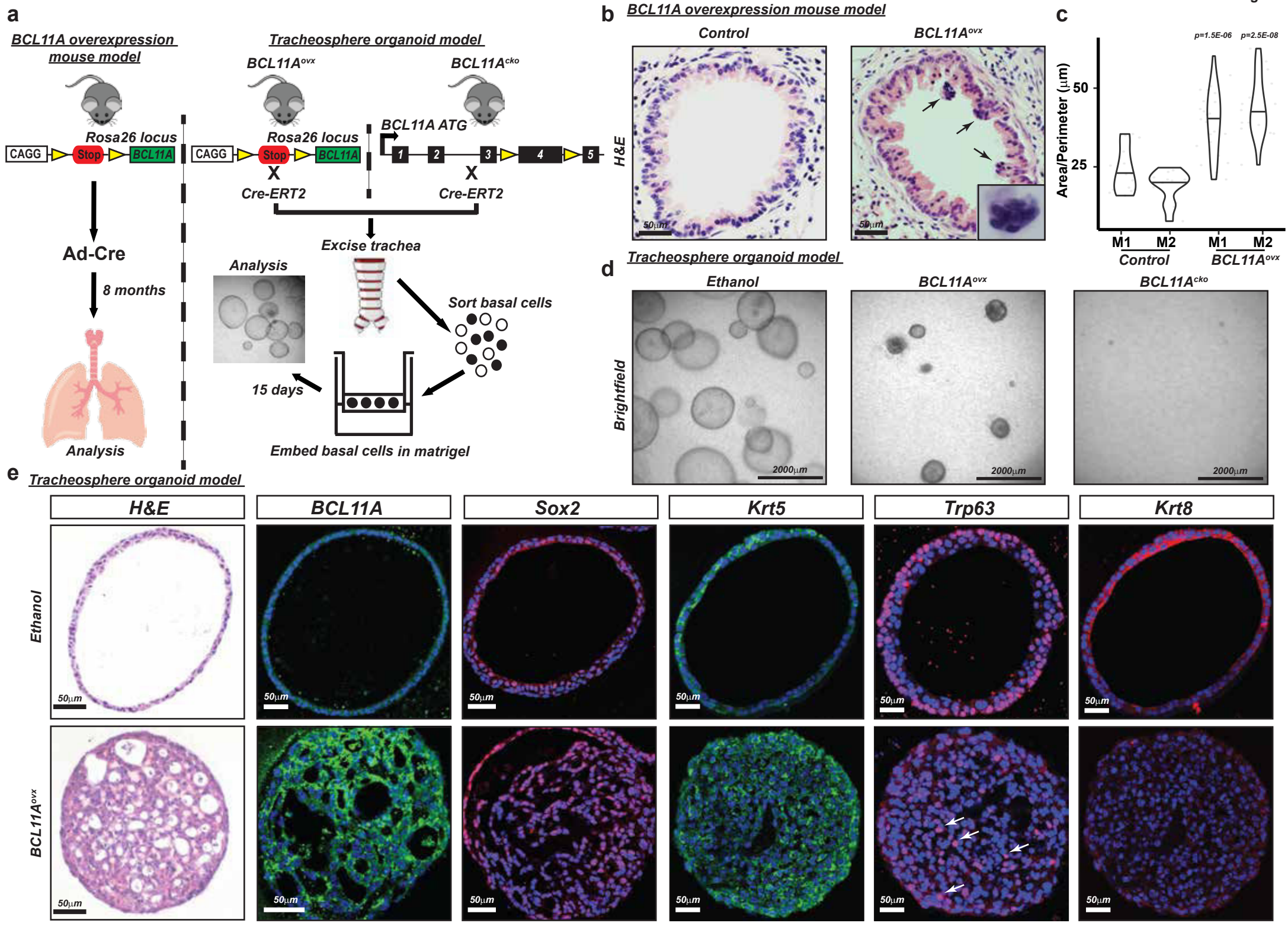
## References

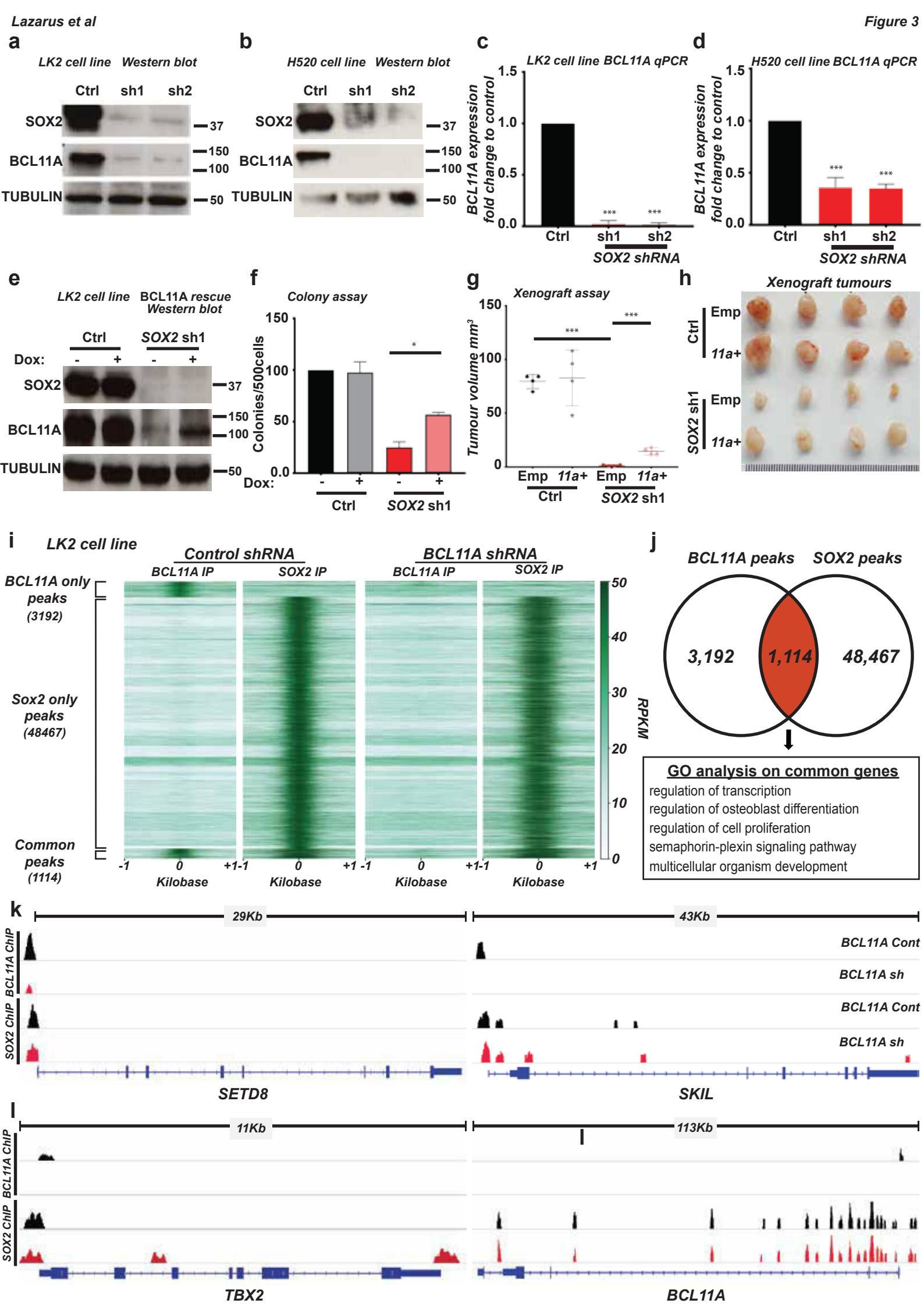

1. Siegel, R.L., Miller, K.D. & Jemal, A. Cancer statistics, 2017. *CA: A Cancer Journal for Clinicians* **67**, 7-30 (2017).
2. Chen, Z., Fillmore, C.M., Hammerman, P.S., Kim, C.F. & Wong, K.K. Non-small-cell lung cancers: a heterogeneous set of diseases. *Nature reviews. Cancer* **14**, 535-546 (2014).
3. Jemal, A.*, et al.* Global cancer statistics. *CA Cancer Journal for Clinicians* **61**, 69-90 (2011).
4. Pillai, R.N. & Ramalingam, S.S. Necitumumab: a new therapeutic option for squamous cell lung cancer? *Translational Lung Cancer Research* **3**, 382-383 (2014).
5. Kwon, M.-c. & Berns, A. Mouse models for lung cancer. *Molecular Oncology* **7**, 165-177 (2013).
6. The Cancer Genome Atlas Research, N. Comprehensive molecular profiling of lung adenocarcinoma. *Nature* **511**, 543-550 (2014).
7. McCaughan, F.*, et al.* Progressive 3q Amplification Consistently Targets SOX2 in Preinvasive Squamous Lung Cancer. *American Journal of Respiratory and Critical Care Medicine* **182**, 83-91 (2010).
8. Bass, A.J.*, et al.* SOX2 is an amplified lineage-survival oncogene in lung and esophageal squamous cell carcinomas. *Nat Genet* **41**, 1238-1242 (2009).
9. Hussenet, T.*, et al.* SOX2 Is an Oncogene Activated by Recurrent 3q26.3 Amplifications in Human Lung Squamous Cell Carcinomas. *PLOS ONE* **5**, e8960 (2010).
10. Lu, Y.*, et al.* Evidence That SOX2 Overexpression Is Oncogenic in the Lung. *PLOS ONE* **5**, e11022 (2010).
11. Network, T.C.G.A.R. Comprehensive genomic characterization of squamous cell lung cancers. *Nature* **489**, 519-525 (2012).
12. Campbell, J.D.*, et al.* Distinct patterns of somatic genome alterations in lung adenocarcinomas and squamous cell carcinomas. *Nat Genet* **48**, 607-616 (2016).
13. Khaled, W.T.*, et al.* BCL11A is a triple-negative breast cancer gene with critical functions in stem and progenitor cells. *Nat Commun* **6**(2015).
14. Tao, H.*, et al.* BCL11A expression in acute myeloid leukemia. *Leukemia Research* **41**, 71-75 (2016).
15. Weniger, M.A.*, et al.* Gains of the proto-oncogene BCL11A and nuclear accumulation of BCL11AXL protein are frequent in primary mediastinal B-cell lymphoma. *Leukemia* **20**, 1880-1882 (2006).
16. Yu, Y.*, et al.* Bcl11a is essential for lymphoid development and negatively regulates p53. *The Journal of Experimental Medicine* **209**, 2467 (2012).
17. Jamal-Hanjani, M.*, et al.* Tracking the Evolution of Non–Small-Cell Lung Cancer. *New England Journal of Medicine* **376**, 2109-2121 (2017).
18. Zhang, N.*, et al.* The BCL11A-XL expression predicts relapse in squamous cell carcinoma and large cell carcinoma. *Journal of Thoracic Disease* **7**, 1630-1636 (2015).
19. Idowu, M.O. & Powers, C.N. Lung cancer cytology: potential pitfalls and mimics - a review. *International Journal of Clinical and Experimental Pathology* **3**, 367-385 (2010).
20. Rock, J.R.*, et al.* Basal cells as stem cells of the mouse trachea and human airway epithelium. *Proceedings of the National Academy of Sciences* **106**, 12771-12775 (2009).
21. Ferone, G.*, et al.* SOX2 Is the Determining Oncogenic Switch in Promoting Lung Squamous Cell Carcinoma from Different Cells of Origin. *Cancer Cell* **30**, 519-532 (2016).
22. Weeden, C.E.*, et al.* Lung Basal Stem Cells Rapidly Repair DNA Damage Using the Error-Prone Nonhomologous End-Joining Pathway. *PLOS Biology* **15**, e2000731 (2017).
23. Hagerstrand, D.*, et al.* Systematic interrogation of 3q26 identifies TLOC1 and SKIL as cancer drivers. *Cancer discovery* **3**, 1044-1057 (2013).

24. Zhang, Z. & Guo, Y. *High TBX2 expression predicts poor prognosis in non-small cell lung cancer*, (2014).
25. Chen, T.*, et al.* miR-382 inhibits tumor progression by targeting SETD8 in non-small cell lung cancer. *Biomedicine & Pharmacotherapy* **86**, 248-253 (2017).
26. Nishioka, K.*, et al.* PR-Set7 Is a Nucleosome-Specific Methyltransferase that Modifies Lysine 20 of Histone H4 and Is Associated with Silent Chromatin. *Molecular Cell* **9**, 1201-1213 (2002).
27. Wu, S. & Rice, J.C. A new regulator of the cell cycle. *Cell Cycle* **10**, 68-72 (2011).
28. Tecalco-Cruz, A.C.*, et al.* Transforming Growth Factor-β/SMAD Target Gene SKIL Is Negatively Regulated by the Transcriptional Cofactor Complex SNON-SMAD4. *Journal of Biological Chemistry* **287**, 26764-26776 (2012).
29. Prince, S., Carreira, S., Vance, K.W., Abrahams, A. & Goding, C.R. Tbx2 Directly Represses the Expression of the p21 (WAF1) Cyclin-Dependent Kinase Inhibitor. *Cancer Research* **64**, 1669 (2004).
30. Lynch, T.J.*, et al.* Submucosal Gland Myoepithelial Cells Are Reserve Stem Cells That Can Regenerate Mouse Tracheal Epithelium. *Cell Stem Cell* **22**, 653-667.e655 (2018).
31. Hong, K.U., Reynolds, S.D., Watkins, S., Fuchs, E. & Stripp, B.R. Basal Cells Are a Multipotent Progenitor Capable of Renewing the Bronchial Epithelium. *The American Journal of Pathology* **164**, 577-588 (2004).
32. Wiederschain, D.*, et al.* Single-vector inducible lentiviral RNAi system for oncology target validation. *Cell Cycle* **8**, 498-504 (2009).
33. Lu, X.*, et al.* The effect of H3K79 dimethylation and H4K20 trimethylation on nucleosome and chromatin structure. *Nat Struct Mol Biol* **15**, 1122-1124 (2008).
34. Blum, G.*, et al.* Small-Molecule Inhibitors of SETD8 with Cellular Activity. *ACS Chemical Biology* **9**, 2471-2478 (2014).
35. Pu, L., Amoscato, A.A., Bier, M.E. & Lazo, J.S. Dual G1 and G2 Phase Inhibition by a Novel, Selective Cdc25 Inhibitor 7-Chloro-6-(2-morpholin-4-ylethylamino)- quinoline-5,8-dione. *Journal of Biological Chemistry* **277**, 46877-46885 (2002).
36. Brisson, M.*, et al.* Redox Regulation of Cdc25B by Cell-Active Quinolinediones. *Molecular Pharmacology* **68**, 1810 (2005).
37. Lazo, J.S.*, et al.* Discovery and Biological Evaluation of a New Family of Potent Inhibitors of the Dual Specificity Protein Phosphatase Cdc25. *Journal of Medicinal Chemistry* **44**, 4042-4049 (2001).
38. Gandara, D.R., Hammerman, P.S., Sos, M.L., Lara, P.N. & Hirsch, F.R. Squamous Cell Lung Cancer: From Tumor Genomics to Cancer Therapeutics. *Clinical Cancer Research* **21**, 2236 (2015).
39. Nakamura, T.*, et al.* Evi9 Encodes a Novel Zinc Finger Protein That Physically Interacts with BCL6, a Known Human B-Cell Proto-Oncogene Product. *Molecular and Cellular Biology* **20**, 3178-3186 (2000).
40. Khaled, W.T.*, et al.* BCL11A is a triple-negative breast cancer gene with critical functions in stem and progenitor cells. **6**, 5987 (2015).
41. Driskell, I.*, et al.* The histone methyltransferase Setd8 acts in concert with c-Myc and is required to maintain skin. *The EMBO Journal* **31**, 616-629 (2012).
42. Takawa, M.*, et al.* Histone Lysine Methyltransferase SETD8 Promotes Carcinogenesis by Deregulating PCNA Expression. *Cancer Research* **72**, 3217 (2012).
43. Shi, X.*, et al.* Modulation of p53 Function by SET8-Mediated Methylation at Lysine 382. *Molecular Cell* **27**, 636-646 (2007).
44. Yang, F.*, et al.* SET8 promotes epithelial–mesenchymal transition and confers TWIST dual transcriptional activities. *The EMBO Journal* **31**, 110 (2011).
45. Sasaki, Y.*, et al.* Canonical NF-κB Activity, Dispensable for B Cell Development, Replaces BAFF-Receptor Signals and Promotes B Cell Proliferation upon Activation. *Immunity* **24**, 729-739 (2006).

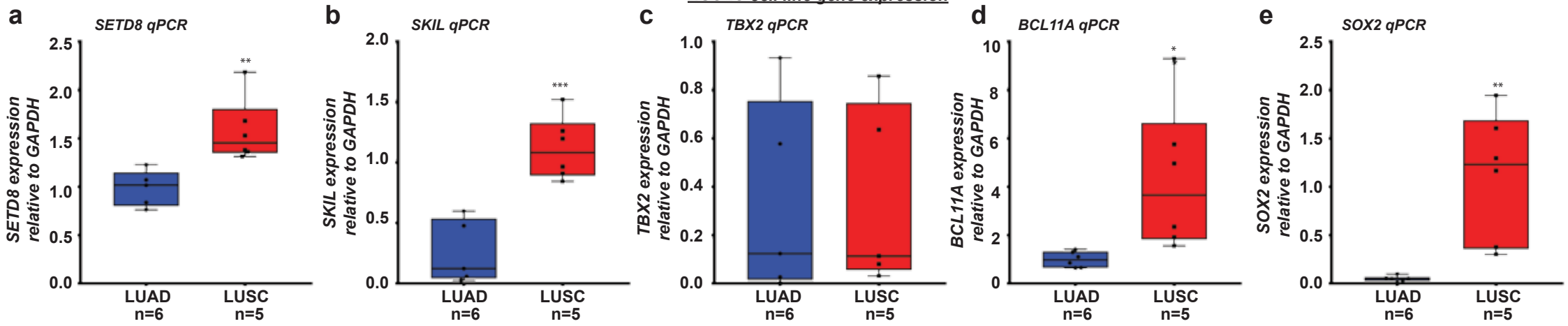
46. You, Y., Richer, E.J., Huang, T. & Brody, S.L. Growth and differentiation of mouse tracheal epithelial cells: selection of a proliferative population. *American Journal of Physiology - Lung Cellular and Molecular Physiology* **283**, L1315 (2002).
47. Schmidt, D.*, et al.* ChIP-seq: Using high-throughput sequencing to discover protein–DNA interactions. *Methods* **48**, 240-248 (2009).
48. Collado-Torres, L.*, et al.* Reproducible RNA-seq analysis using recount2. *Nat Biotech* **35**, 319-321 (2017).
49. Robinson, M.D., McCarthy, D.J. & Smyth, G.K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-140 (2010).
50. Ramírez, F., Dündar, F., Diehl, S., Grüning, B.A. & Manke, T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Research* **42**, W187-W191 (2014).


a
LUAD
LUSC
BCL11A
SOX2
REL
-log10(pValue)
Log fold change
b
Matched Normal
LUSC
BCL11A
SOX2
REL
-log10(pValue)
Log fold change
c
Matched Normal
LUAD
REL
BCL11A
SOX2
-log10(pValue)
Log fold change
d
BCL11A IHC
LUAD
Negative
Negative
LUSC
Moderate
Strong
% patients
n=99
n=120
Strong
Moderate
Negative
e
LK2 cell line Xenograft assay
Tumour volume mm3
Ctrl
sh1
sh2
BCL11A shRNA
Xenograft tumours
BCL11A shRNA
f
H520 cell line Xenograft assay
Tumour volume mm3
Ctrl
sh1
sh2
BCL11A shRNA
Xenograft tumours
BCL11A shRNA


a
BCL11A overexpression mouse model
Tracheosphere organoid model
BCL11A^ovx
BCL11A^cko
Rosa26 locus
CAGG
Stop
BCL11A
BCL11A ATG
1
2
3
4
5
X
Cre-ERT2
Ad-Cre
8 months
Analysis
Excise trachea
Sort basal cells
15 days
Embed basal cells in matrigel
b
BCL11A overexpression mouse model
Control
BCL11A^ovx
H&E
50μm
c
p=1.5E-06
p=2.5E-08
Area/Perimeter (μm)
50
25
M1
M2
M1
M2
Control
BCL11A^ovx
d
Tracheosphere organoid model
Ethanol
BCL11A^ovx
BCL11A^cko
Brightfield
2000μm
e
Tracheosphere organoid model
H&E
BCL11A
Sox2
Krt5
Trp63
Krt8
Ethanol
BCL11A^ovx
50μm

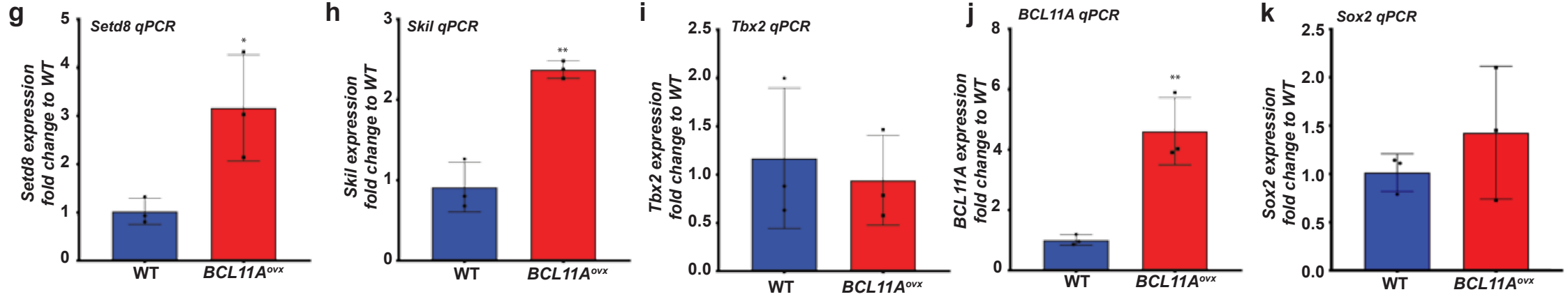

a
LK2 cell line Western blot
Ctrl sh1 sh2
SOX2
BCL11A
TUBULIN
37
150
100
50
b
H520 cell line Western blot
Ctrl sh1 sh2
SOX2
BCL11A
TUBULIN
37
150
100
50
c
LK2 cell line BCL11A qPCR
BCL11A expression fold change to control
1.5
1.0
0.5
0.0
***
***
Ctrl sh1 sh2
SOX2 shRNA
d
H520 cell line BCL11A qPCR
BCL11A expression fold change to control
1.5
1.0
0.5
0.0
***
***
Ctrl sh1 sh2
SOX2 shRNA
e
LK2 cell line BCL11A rescue Western blot
Ctrl SOX2 sh1
Dox: - + - +
SOX2
BCL11A
TUBULIN
37
150
100
50
f
Colony assay
Colonies/500cells
150
100
50
0.0
*
Dox: - + - +
Ctrl SOX2 sh1
g
Xenograft assay
Tumour volume mm3
150
100
50
0
***
***
Emp 11a+ Emp 11a+
Ctrl SOX2 sh1
h
Xenograft tumours
Ctrl
Emp
11a+
SOX2 sh1
Emp
11a+
i
LK2 cell line
Control shRNA
BCL11A shRNA
BCL11A IP
SOX2 IP
BCL11A IP
SOX2 IP
BCL11A only peaks (3192)
Sox2 only peaks (48467)
Common peaks (1114)
50
40
30
20
10
0
RPKM
-1 0 +1
Kilobase
j
BCL11A peaks
SOX2 peaks
3,192
1,114
48,467
GO analysis on common genes
regulation of transcription
regulation of osteoblast differentiation
regulation of cell proliferation
semaphorin-plexin signaling pathway
multicellular organism development
k
29Kb
43Kb
BCL11A ChIP
SOX2 ChIP
BCL11A Cont
BCL11A sh
BCL11A Cont
BCL11A sh
SETD8
SKIL
l
11Kb
113Kb
BCL11A ChIP
SOX2 ChIP
TBX2
BCL11A

NSCLC cell line gene expression

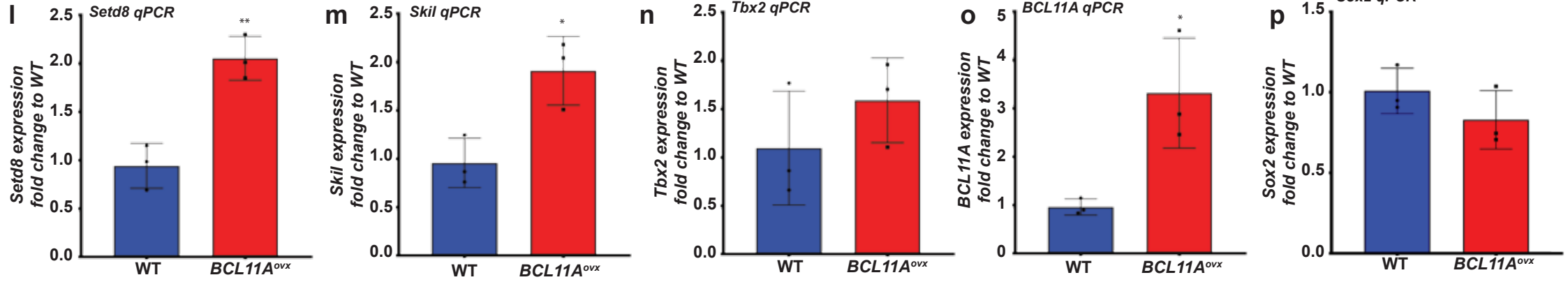


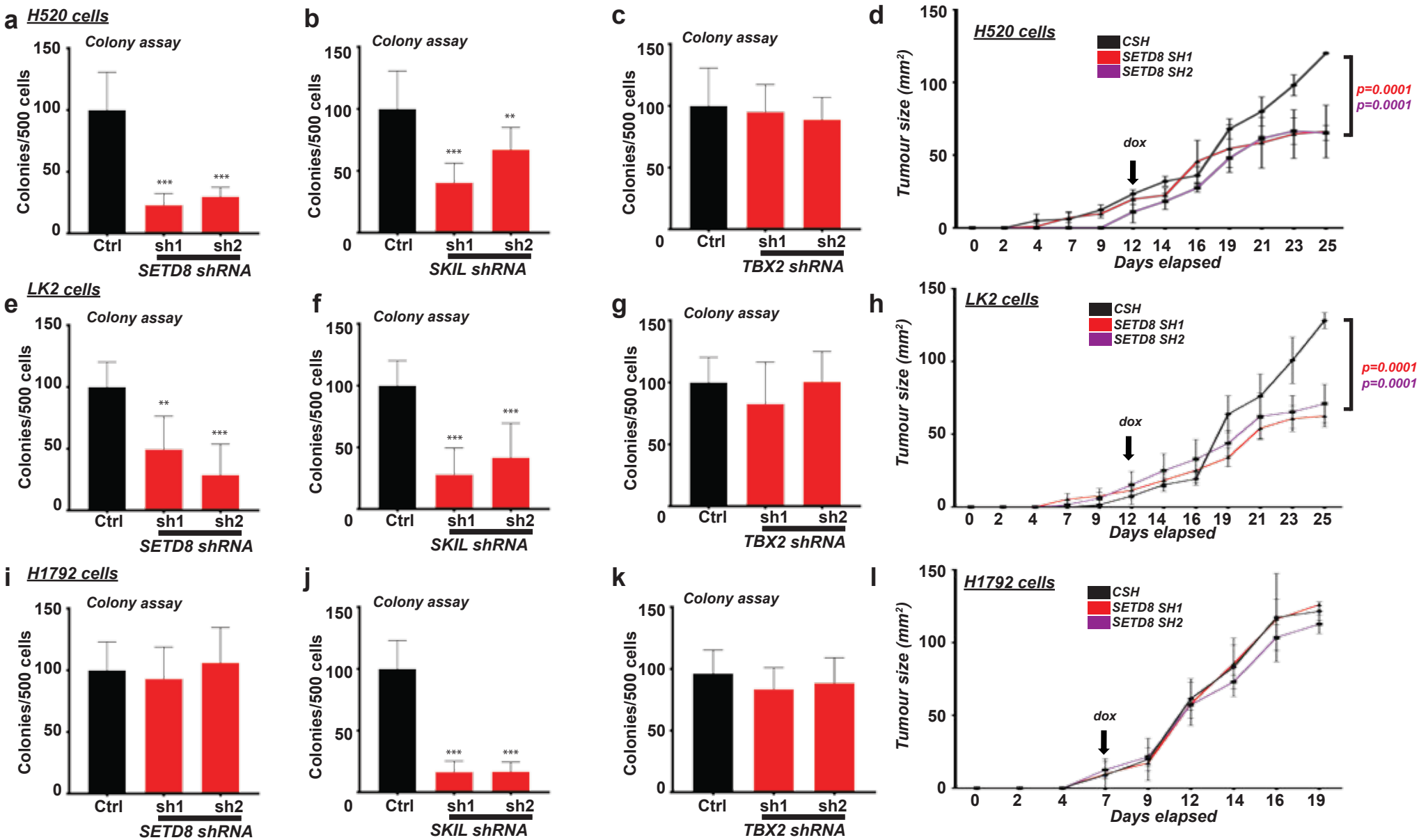


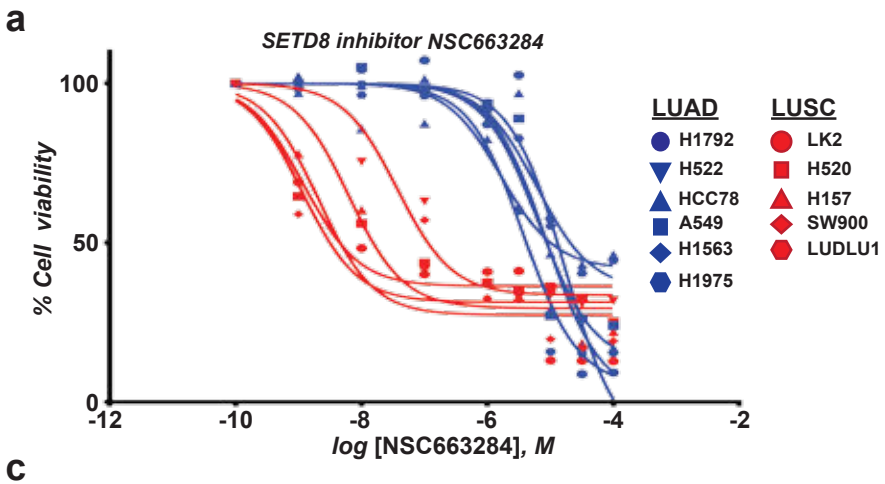
Sort Basal stem cells Epcam+/GSIB4+

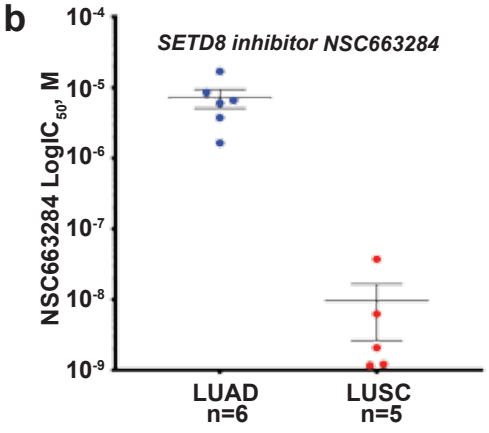


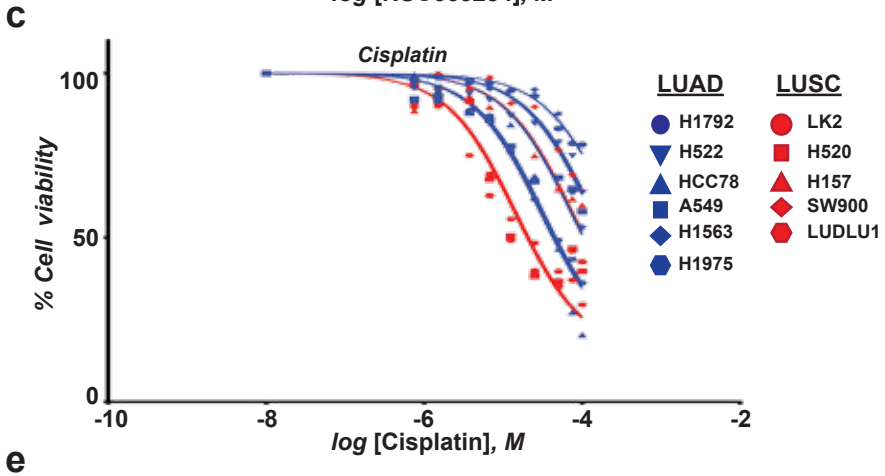

Sort Epcam+ cells

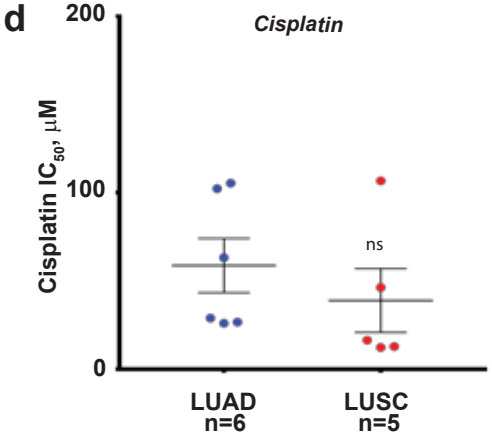


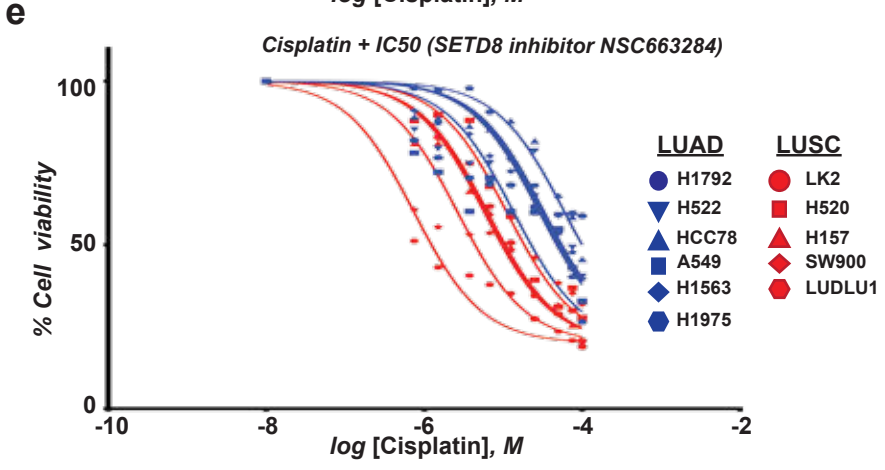

a
H520 cells
Colony assay
Colonies/500 cells
150
100
50
0
***
***
Ctrl
sh1
sh2
SETD8 shRNA
b
Colony assay
Colonies/500 cells
***
**
Ctrl
sh1
sh2
SKIL shRNA
c
Colony assay
Colonies/500 cells
Ctrl
sh1
sh2
TBX2 shRNA
d
H520 cells
CSH
SETD8 SH1
SETD8 SH2
Tumour size (mm²)
dox
0 2 4 7 9 12 14 16 19 21 23 25
Days elapsed
p=0.0001
p=0.0001
e
LK2 cells
Colony assay
Colonies/500 cells
**
***
Ctrl
sh1
sh2
SETD8 shRNA
f
Colony assay
Colonies/500 cells
***
***
Ctrl
sh1
sh2
SKIL shRNA
g
Colony assay
Colonies/500 cells
Ctrl
sh1
sh2
TBX2 shRNA
h
LK2 cells
CSH
SETD8 SH1
SETD8 SH2
Tumour size (mm²)
dox
0 2 4 7 9 12 14 16 19 21 23 25
Days elapsed
p=0.0001
p=0.0001
i
H1792 cells
Colony assay
Colonies/500 cells
Ctrl
sh1
sh2
SETD8 shRNA
j
Colony assay
Colonies/500 cells
***
***
Ctrl
sh1
sh2
SKIL shRNA
k
Colony assay
Colonies/500 cells
Ctrl
sh1
sh2
TBX2 shRNA
l
H1792 cells
CSH
SETD8 SH1
SETD8 SH2
Tumour size (mm²)
dox
0 2 4 7 9 12 14 16 19
Days elapsed

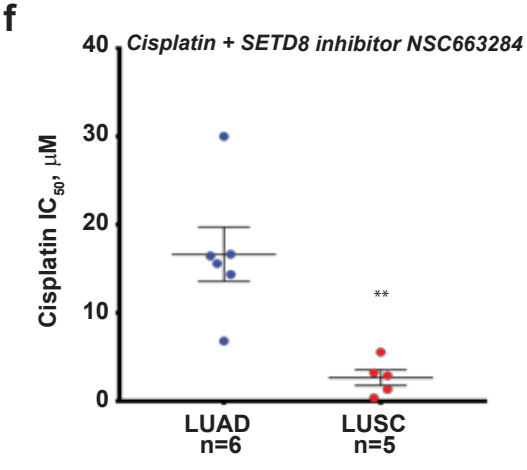

a
SETD8 inhibitor NSC663284
% Cell viability
log [NSC663284], M
LUAD
H1792
H522
HCC78
A549
H1563
H1975
LUSC
LK2
H520
H157
SW900
LUDLU1

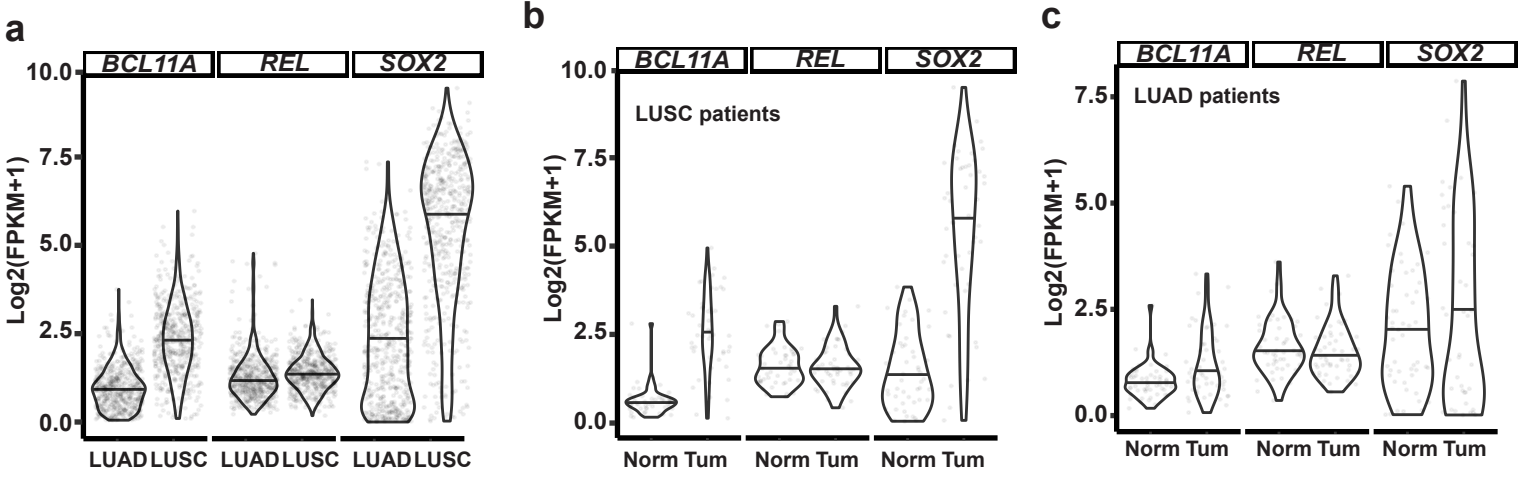

b
SETD8 inhibitor NSC663284
NSC663284 LogIC50, M
LUAD n=6
LUSC n=5


c
Cisplatin
% Cell viability
log [Cisplatin], M
LUAD
H1792
H522
HCC78
A549
H1563
H1975
LUSC
LK2
H520
H157
SW900
LUDLU1

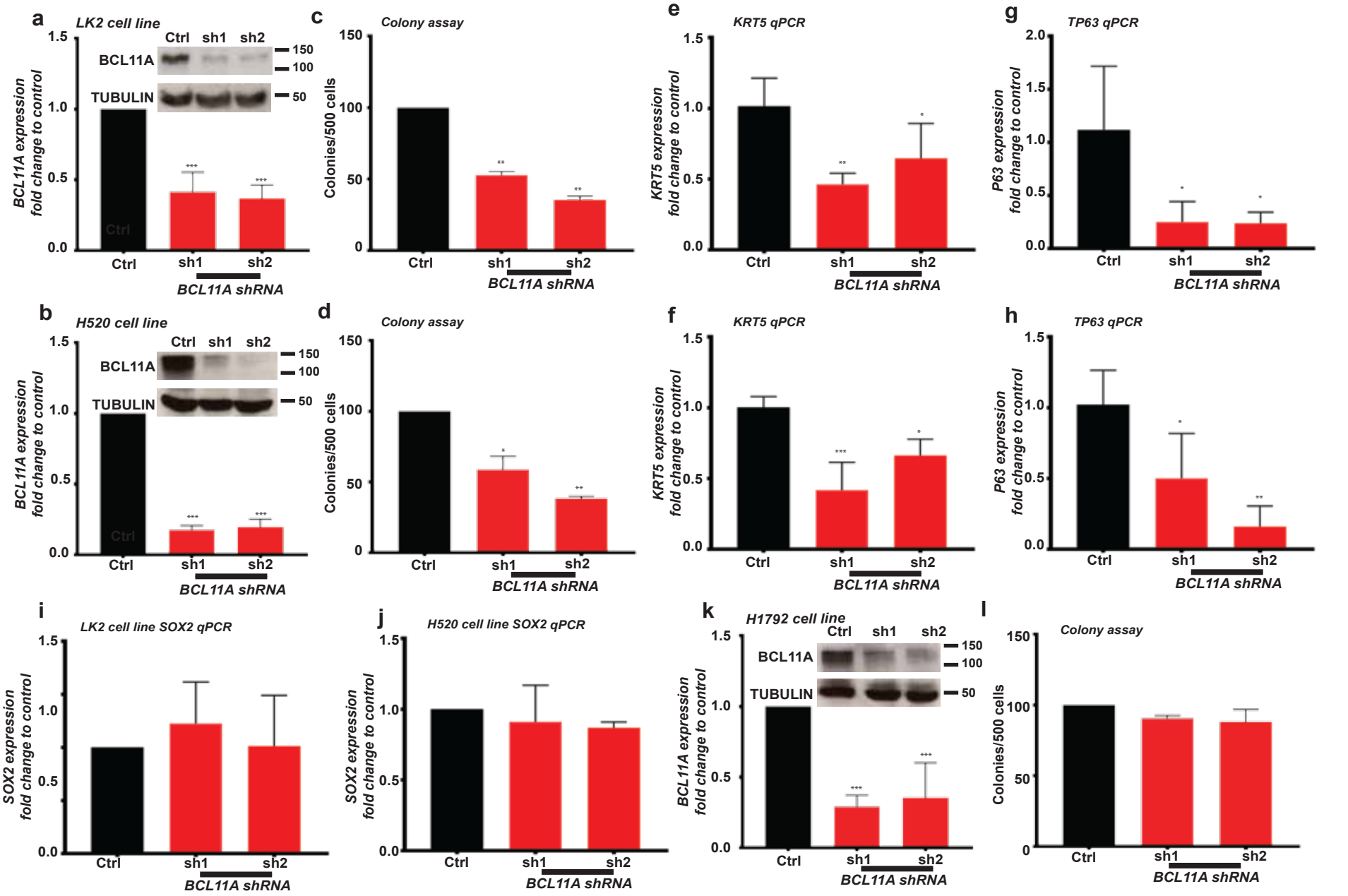

d
Cisplatin
Cisplatin IC50, µM
ns
LUAD n=6
LUSC n=5


e
Cisplatin + IC50 (SETD8 inhibitor NSC663284)
% Cell viability
log [Cisplatin], M
LUAD
H1792
H522
HCC78
A549
H1563
H1975
LUSC
LK2
H520
H157
SW900
LUDLU1

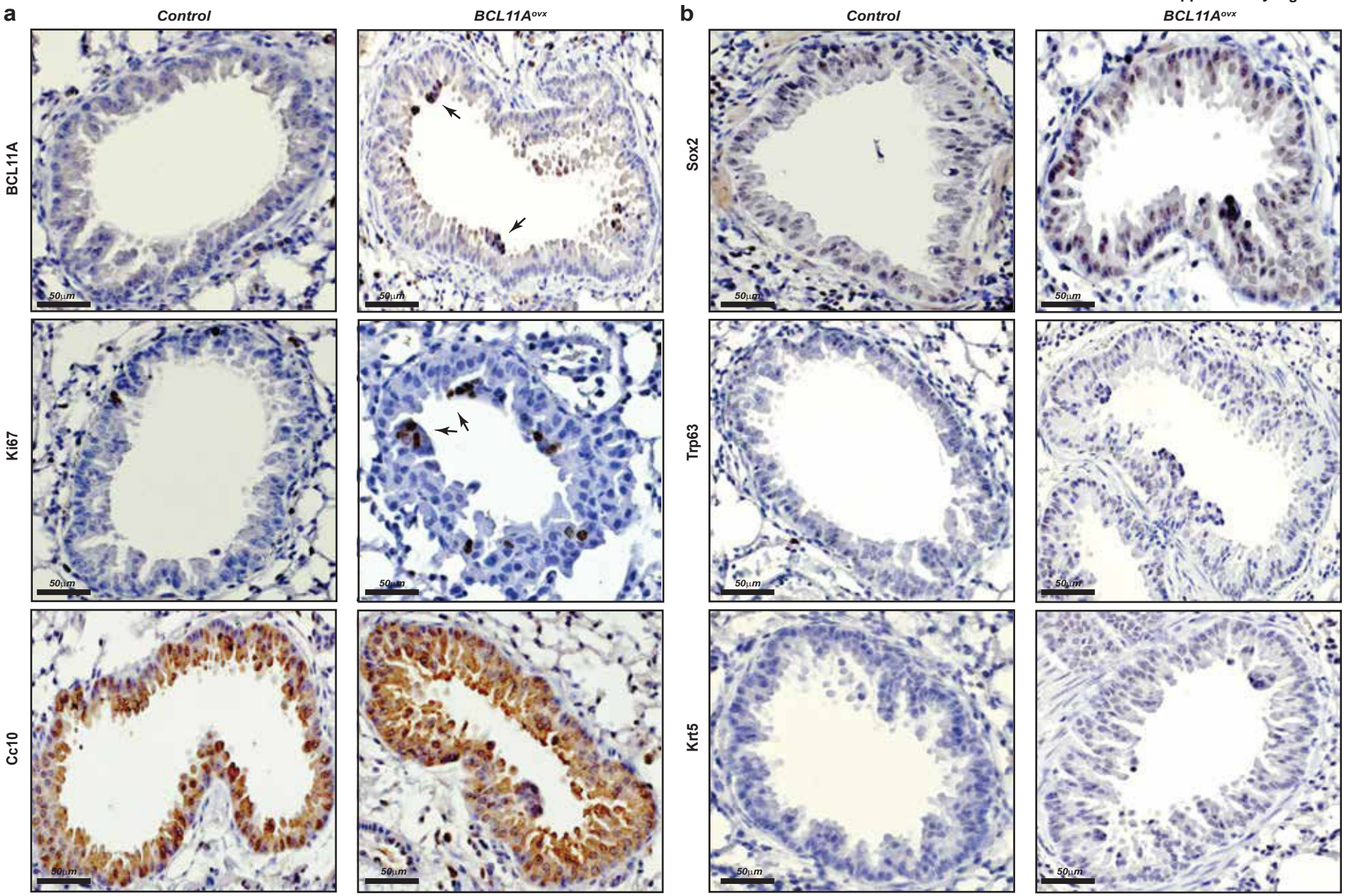

f
Cisplatin + SETD8 inhibitor NSC663284
Cisplatin IC50, µM
**
LUAD n=6
LUSC n=5



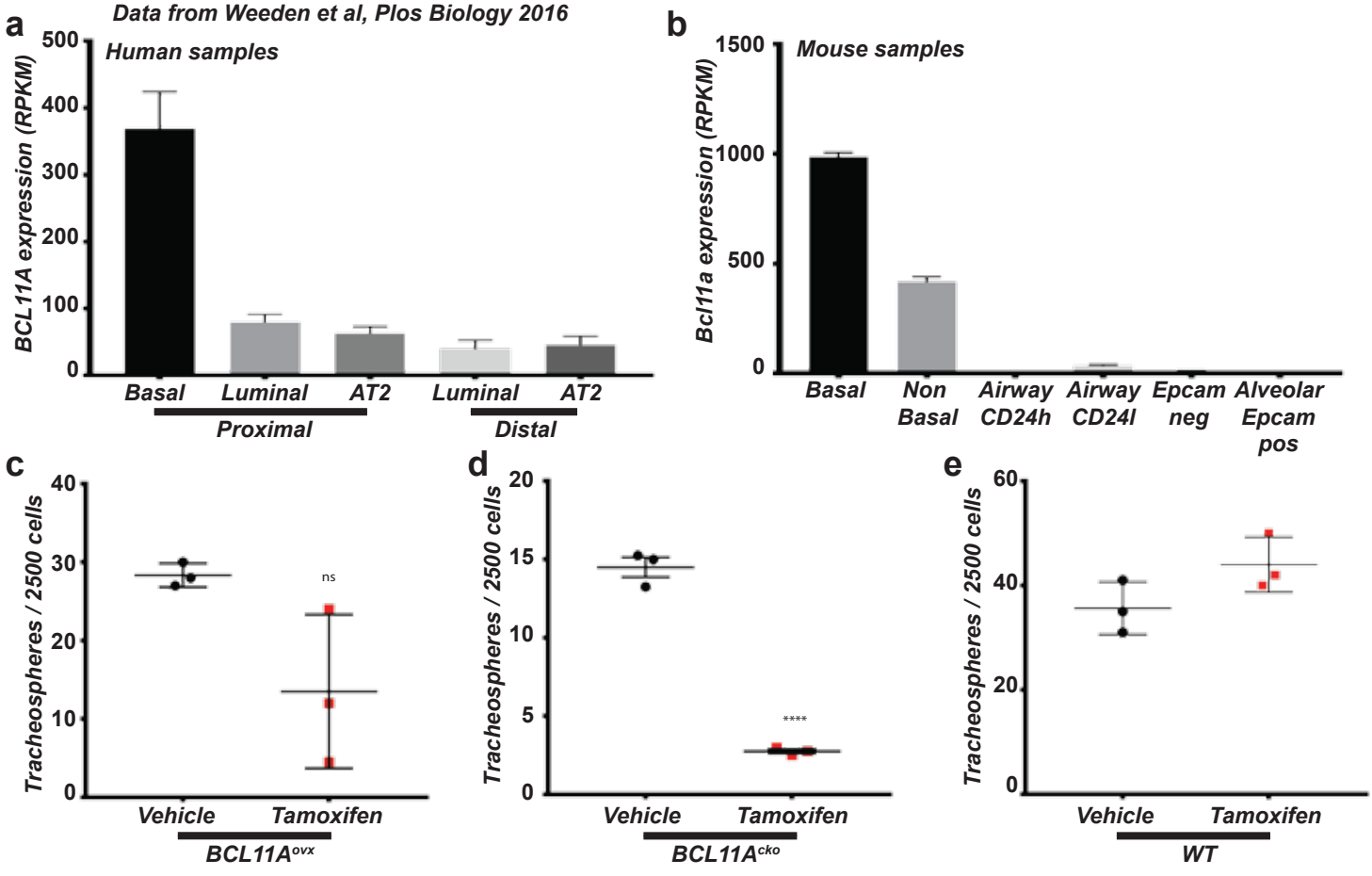

**Supplementary Figure 1. BCL11A, SOX2, and REL expression in Lung TCGA dataset**
**(a) Violin plot showing BCL11A, SOX2 and REL expression in LUSC vs LUAD. FPKM, fragments per kilobase per million mapped reads. (b) Violin plot showing BCL11A, SOX2 and REL expression in LUSC tumour vs normal matched patients. (c) Violin plot showing BCL11A, SOX2 and REL expression in LUAD tumour vs normal matched patients.**



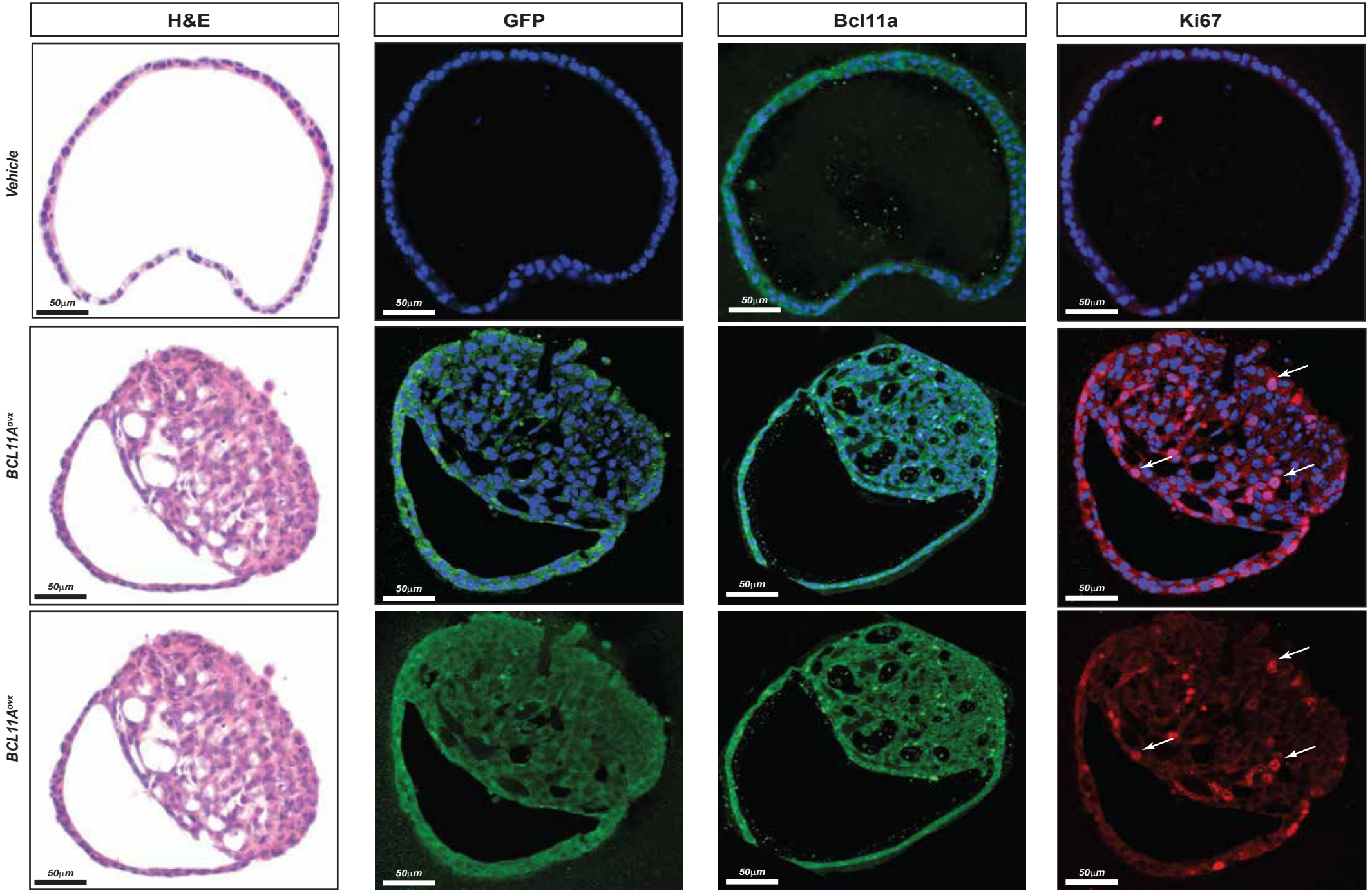

*Supplementary Figure 2. BCL11A KD reduces squamous markers in LUSC cells*
*(a and b) qPCR and western blot shows BCL11A reduction in LK2 (a) and H520 (b) cells transfected with shRNA1 and shRNA2 vectors. (c and d) Comparison of colony numbers in 3D matrigel assay from control, shRNA1 or shRNA2 in LK2 (c) and (d) H520 cells. Data presented as mean ± s.d. (n=3). (e and f) KRT5 expression is reduced in LK2 (e) and H520 (f) BCL11A-KD cells. (g and h) TP63 expression is reduced in LK2 (g) and H520 (h) BCL11A-KD cells. (i and j) SOX2 expression is unchanged in LK2 (i) and H520 (j) BCL11A-KD cells. (k) qPCR and western blot shows BCL11A reduction in H1792 cells transfected with shRNA1 and shRNA2 vectors. (l) Comparison of colony numbers in 3D matrigel assay from control, shRNA1 or shRNA2 in H1792 cells. Data presented as mean ± s.d. One way ANOVA with post Dunnett test performed, * indicates p<0.05 and ** p<0.005 and *** indicates p<0.001.*



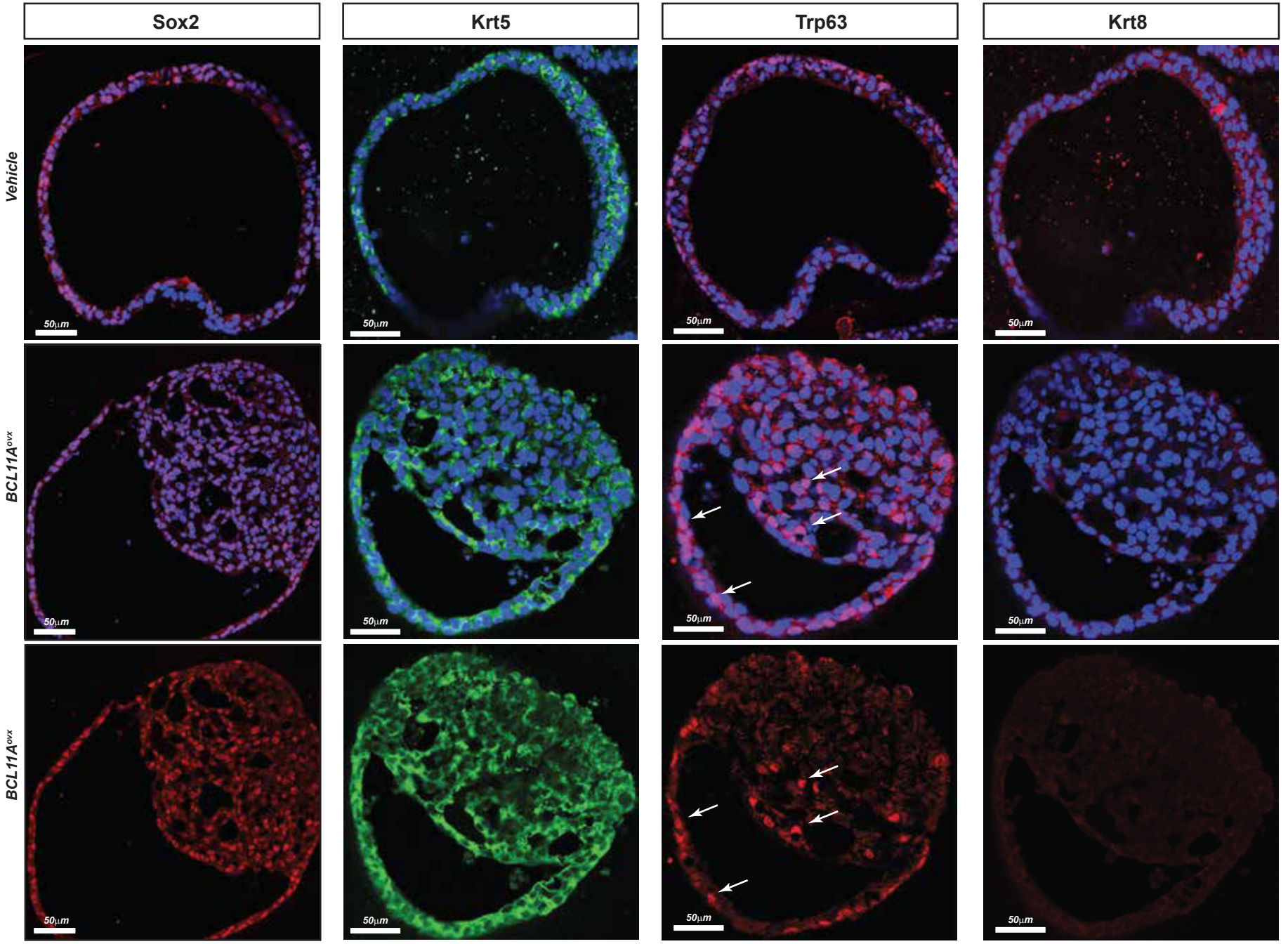

*Supplementary Figure 3. Airways from the BCLL1Aovx mice demonstrate proliferative preneoplastic lesions*
*(a) Immunohistochemistry for BCL11A, Ki67 and Cc10 expression in control vs BCL11Aovx airways. BCL11A panel, arrows indicating positive staining especially in preneoplastic lesions. Ki67 panel, arrows indicating positive staining. (b) Sox2, Trp63 and Krt5 expression in control vs BCL11Aovx airways. Scale bar indicates 50 μm.*



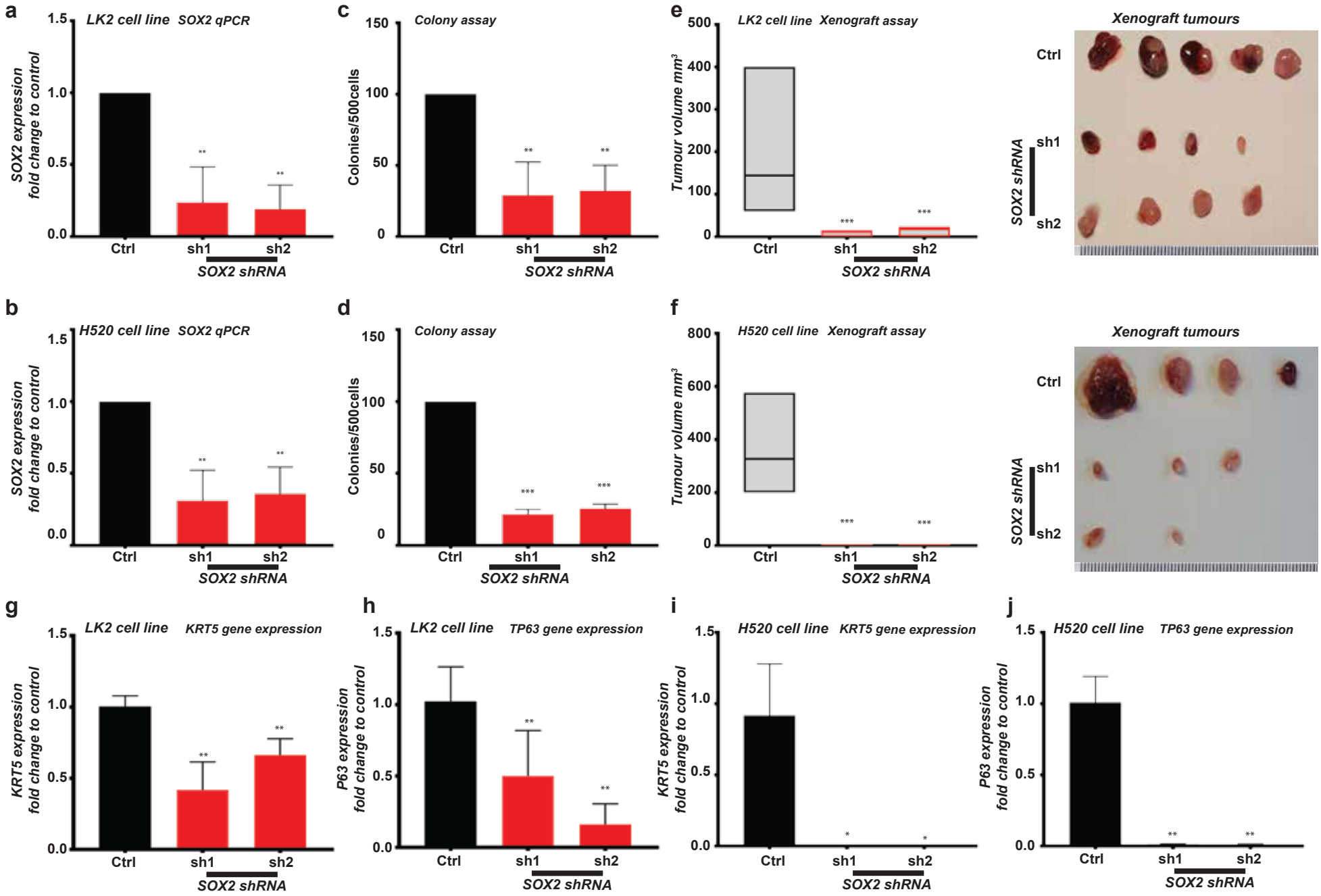

*Supplementary Figure 4. Organoids from BCL11Aovx basal cells exhibit increase in squamous markers*
*(a) RNA-Seq data analysed from Weeden et al 22 indicating that in human samples FACS fraction labelled as proximal BCs has approximately 400 fold increase in human BCL11A levels in comparison to other epithelial fractions. (b) RNA-Seq data from the same dataset indicating that mouse Bcl11a is approximately 500-1000 fold higher in basal fractions compared to other epithelial subtypes. (c, d and e) Quantification of organoids in vehicle vs tamoxifen treated (c) BCL11Aovx, (d) BCL11Acko and (e) WT organoids. Data presented as mean ± s.d. Paired student t-test performed, * indicates p<0.05 and ** p<0.005 and *** indicates p<0.001.*



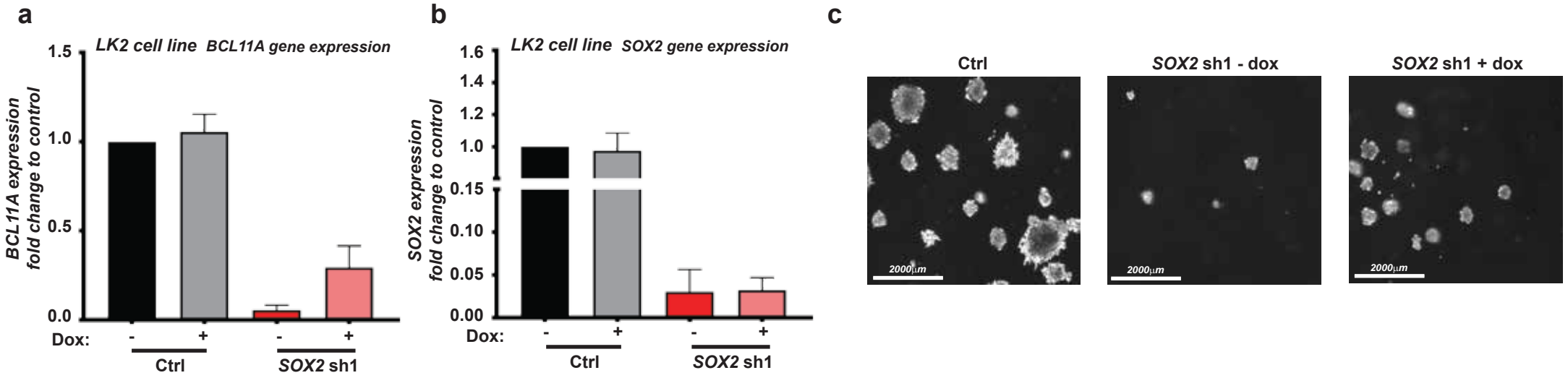

*Supplementary Figure 5. Organoids from the BCL11Aovx basal cells display an abnormal proliferative phenotype*
*Vehicle and tamoxifen treated BCL11Aovx organoids stained with H&E, GFP (which is also expressed if the LSL is efficiently excised), Bcl11a and Ki67. Nuclei stained by DAPI illustrated in blue. Arrows indicate positive staining. Scale bar indicates 50 μm.*



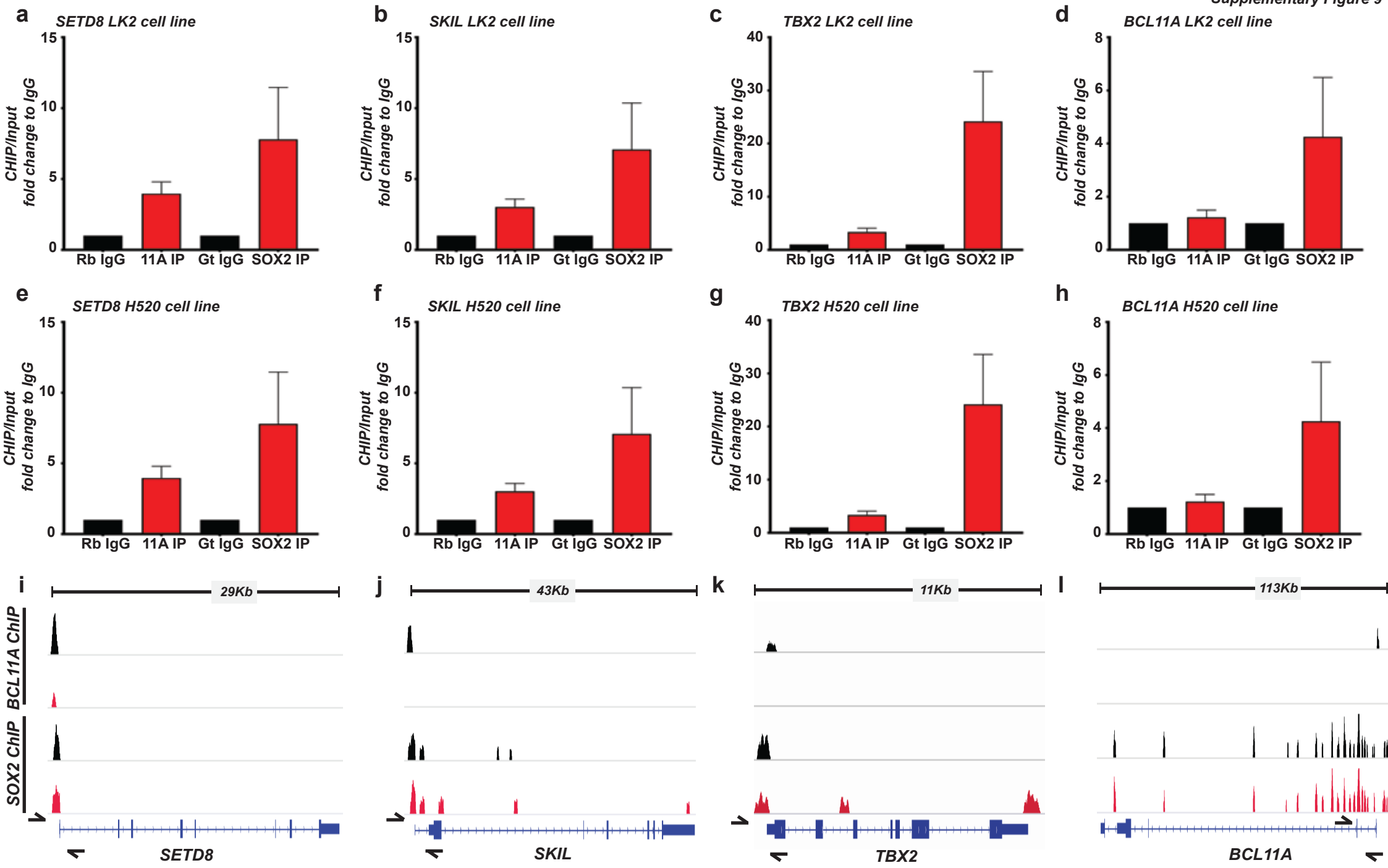

***Supplementary Figure 6. BCL11A overexpression inhibits organoid luminal differentiation***
***Vehicle and tamoxifen treated BCL11Aovx organoids stained with Sox2, Krt5, Trp63 and Krt8. Nuclei stained by DAPI illustrated in blue. Arrows indicate positive staining. Scale bar indicates 50 μm.***



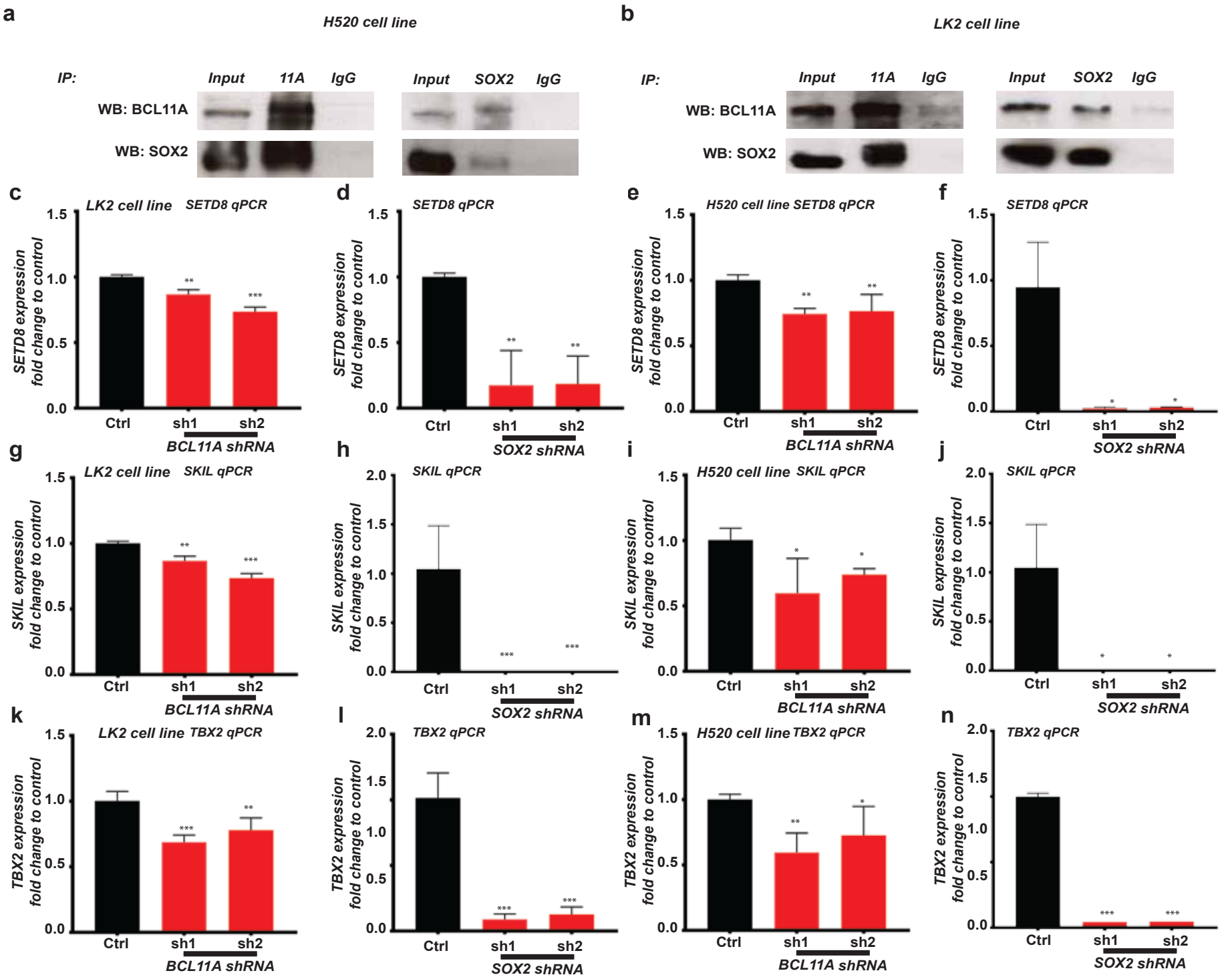

*Supplementary Figure 7. SOX2 KD affects LUSC cell lines tumourigenicity*
*(a and b) Graph showing SOX2 expression in SOX2-KD LK2 (a) and H520 (b) cells. (c and d) Graph showing decrease in colony numbers in SOX2-KD LK2 (c) and H520 (d) cells. (e and f) Graph depicting reduction in tumour size observed when SOX2 shRNA1 or shRNA2 transfected LK2 (e) and H520 (f) cells are injected subcutaneously into mice compared with control. Five (LK2) and four (H520) mice per cell line were monitored for 15 days after which tumours were removed and measured. On the right are images showing actual tumours measured. Data presented as mean ± s.d. One way ANNOVA with post Dunnett test performed, * indicates p<0.05 and ** p<0.005 and *** indicates p<0.001. (g and h) KRT5 expression is reduced in LK2 (g) and H520 (h) BCL11A-KD cells. (i and j) P63 expression is reduced in LK2 (i) and H520 (j) BCL11A-KD cells.*



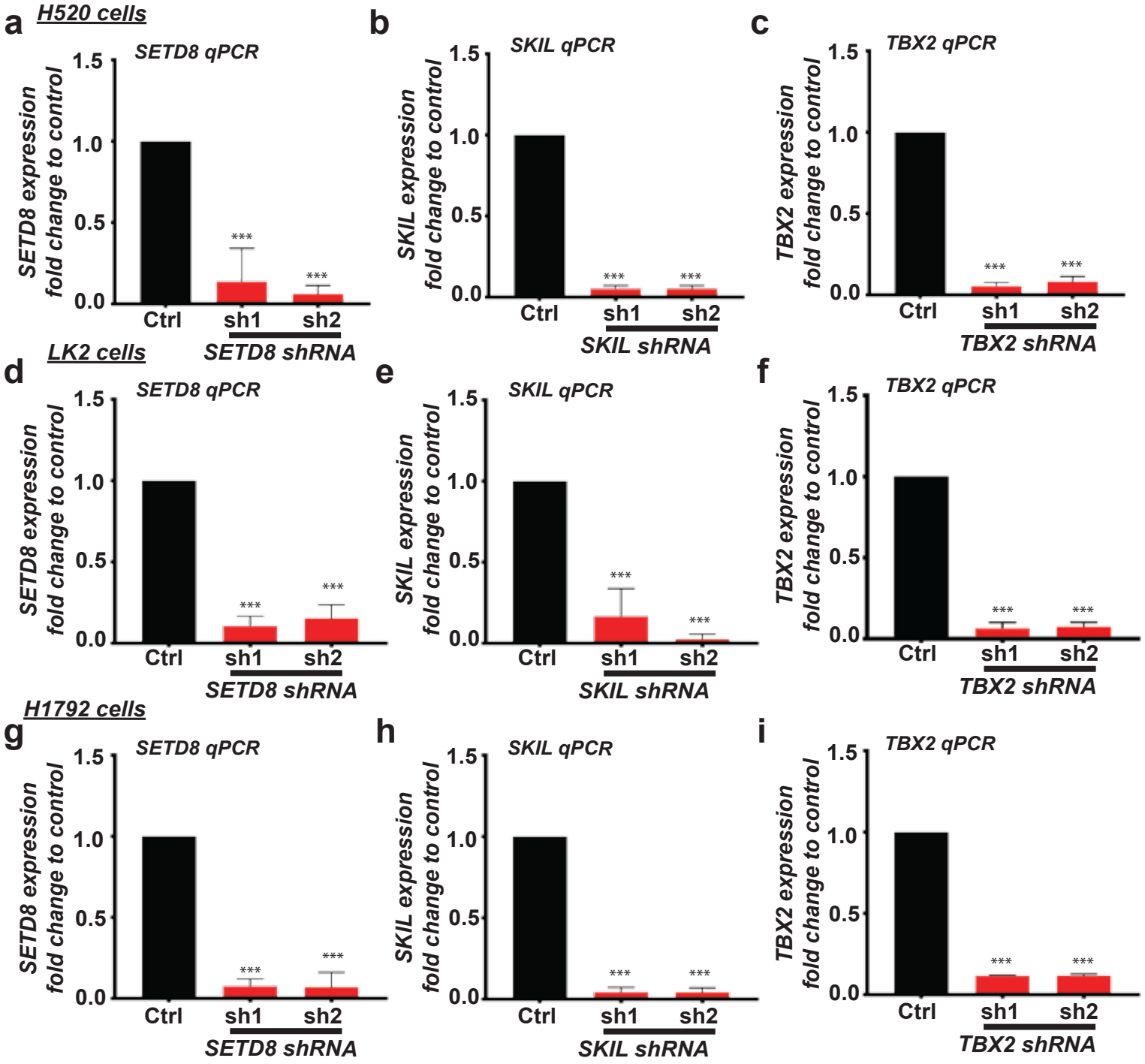

*Supplementary Figure 8. BCL11A is required for SOX2 mediated LUSC phenotype*
*(a and b) Graph showing BCL11A (a) and SOX2 (b) expression in BCL11A rescue in SOX2-KD cells. Dox inducible BCL11A overexpression vector was transfected into control and SOX2 shRNA1 LK2 cells and Dox treatment was performed for 48 hours. (c) Images from 3D matrigel experiment showing reduction in colony numbers after SOX2-KD and partial rescue after BCL11A overexpression.*



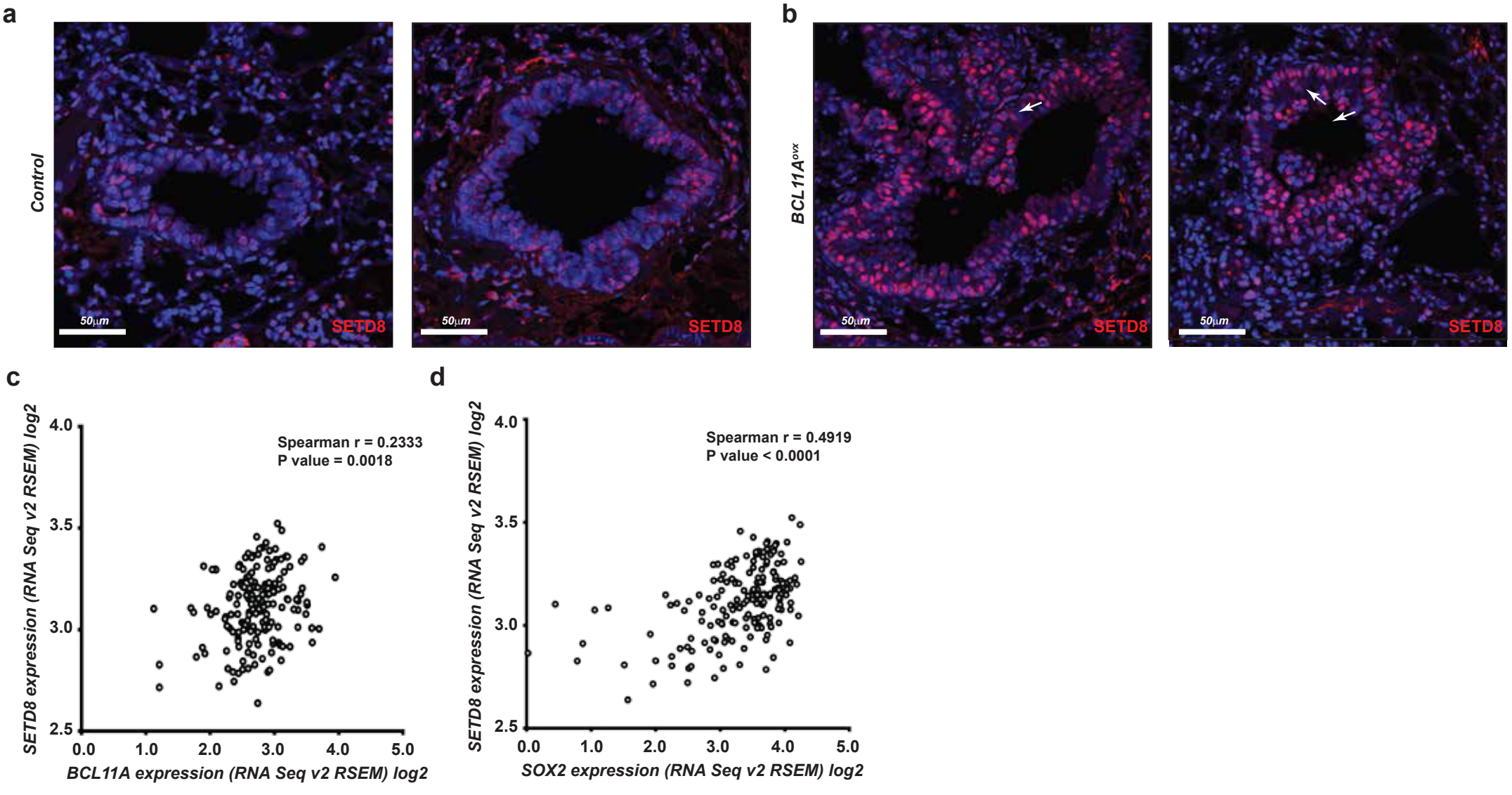

*Supplementary Figure 9. BCL11A and SOX2 bind to common loci on the genome*
*(a-h) ChIP-qPCR showing enrichment for the indicated primers on SETD8, SKIL, TBX2 and BCL11A regions after Chip pull down using anti-BCL11A or anti-SOX2 antibodies in LK2 and H520 cell lines. (i-l) Schematic of amplicon locations for ChIP-qPCR validations performed in this study. Arrows represent location of primers used. Error bars represent mean ± s.e.m.; n = 3 (technical replicates).*



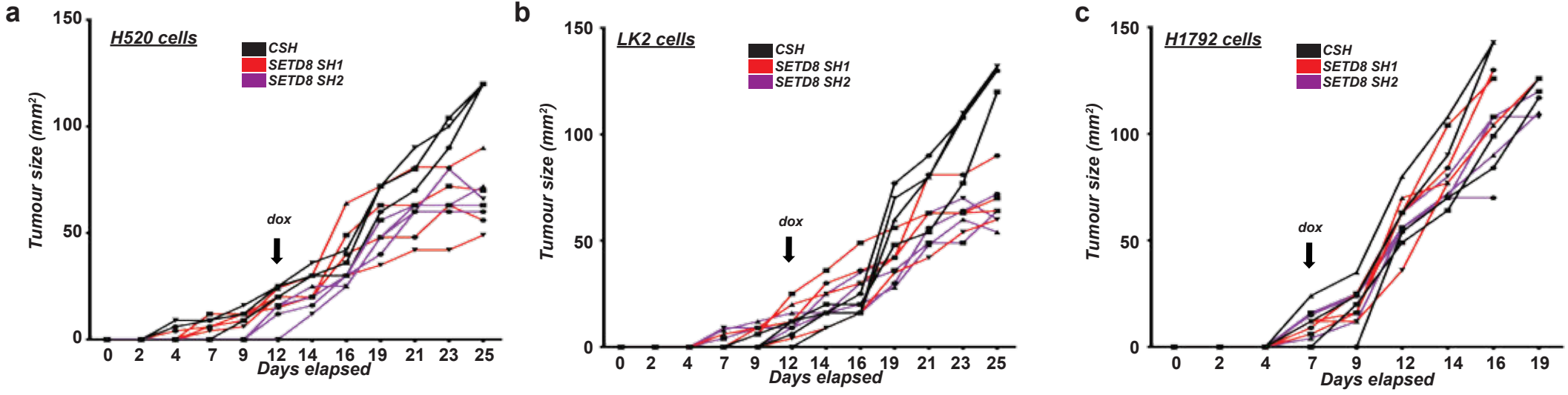

*Supplementary Figure 10. BCL11A and SOX2 co-regulate transcriptional regulators*
*(a and b) Co-Immunoprecipitation of endogenous BCL11A and SOX2 proteins in H520 (a) and LK2 (b) cell lines. (c-f) Graph depicting SETD8 gene expression in LK2, H520 BCL11A-KD and LK2, H520 SOX2 KD cells. (g-j) SKIL gene expression in BCL11A-KD or SOX-KD LK2 and H520 cell lines. (k-n) TBX2 gene expression in BCL11A-KD or SOX2-KD LK2 and H520 cell lines. Data presented as mean ± s.d. (n=3). One way ANOVA with post Dunnett test performed, * indicates p<0.05 and ** p<0.005 and *** indicates p<0.001.*



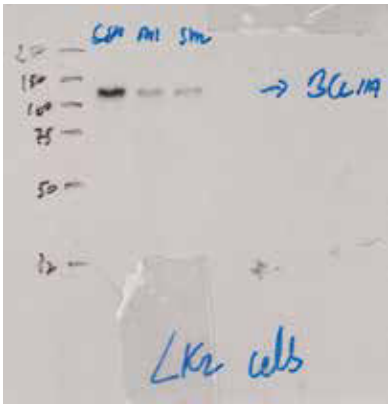

*Supplementary Figure 11. Generation of SETD8, SKIL and TBX2 KD cells*
*(a-c) SETD8 gene expression in SETD8-KD (a) , SKIL gene expression in SKIL-KD (b), TBX2 gene expression in TBX2-KD (c) H520 cells.*
*(d-f) SETD8 gene expression in SETD8-KD (d) , SKIL gene expression in SKIL-KD (e), TBX2 gene expression in TBX2-KD (f) LK2 cells.*
*(g-i) SETD8 gene expression in SETD8-KD (g) , SKIL gene expression in SKIL-KD (h), TBX2 gene expression in TBX2-KD (i) H1792 cells.*
*Data presented as mean ± s.d. (n=3). One way ANOVA with post Dunnett test performed, * indicates p<0.05 and ** p<0.005 and *** indicates p<0.001.*

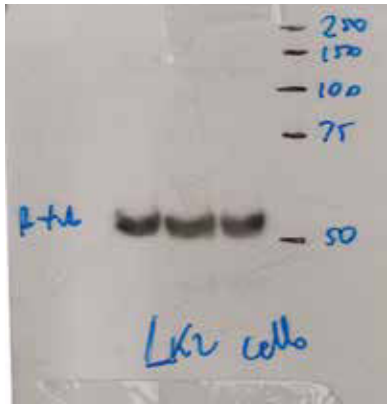



*Supplementary Figure 12. SETD8 correlates with BCL11A and SOX2 in LUSC patients*
*(a and b) SETD8 immunofluorescence in control or BCL11AOVX mouse airways. Scale bar indicates 50 μm. (c) Scatter plot showing SETD8 and BCL11A expression in TCGA patient tumour samples. (d) Scatter plot showing SETD8 and SOX2 expression in TCGA patient tumour samples.*

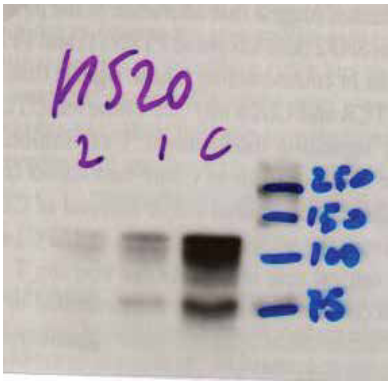



*Supplementary Figure 13. SETD8-KD confers a survival disadvantage to xenograft tumours*
*Tumour kinetics of individual mice injected with either control or SETD8-KD vectors in H520 (a), LK2 (b) or H1792 (c) cells.*



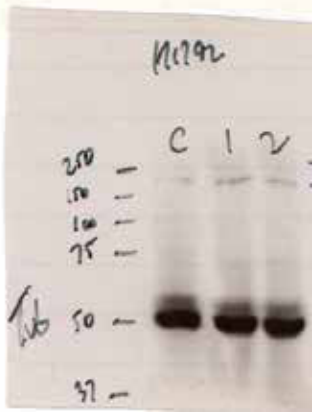




*Western blot used for Supplementary Figure 2a*





*Western blot used for Supplementary Figure 2b*


*Western blot used for Supplementary Figure 2c*

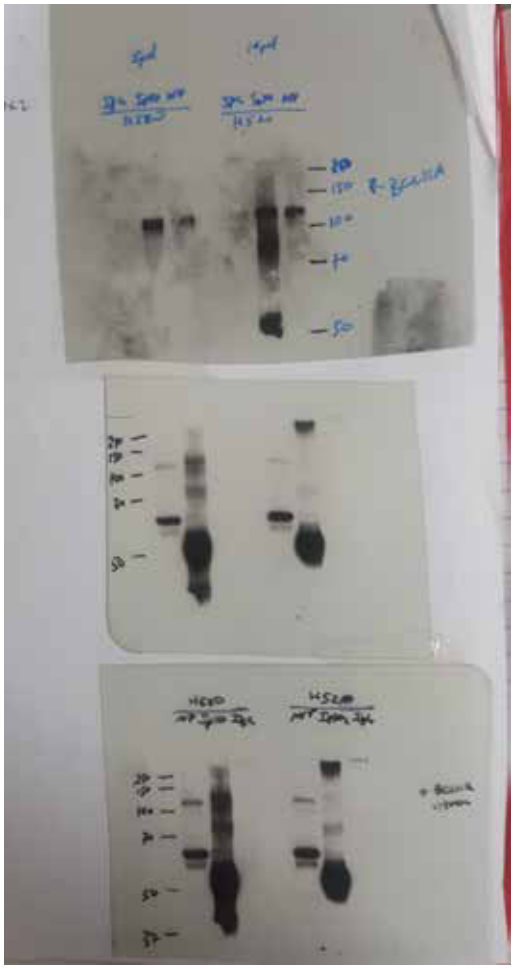

*Western blot used for Supplementary Figure 3a*


*Western blot used for Supplementary Figure 3b*

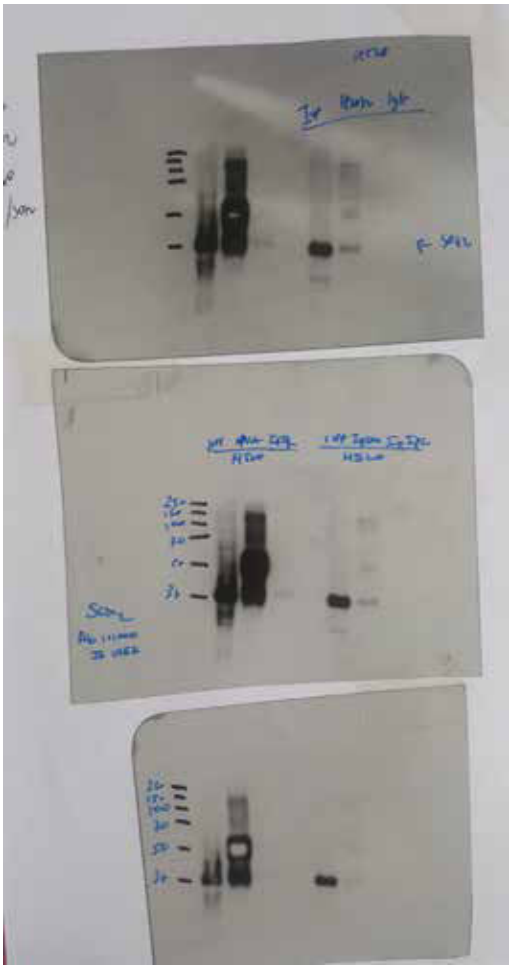
***Supplementary Figure 14. Uncropped western blot images***

***Western blot used for Figure 3e***

***Western blot used for Supplementary Figure 10a***



***Western blot used for Supplementary Figure 10a***



***Western blot used for Supplementary Figure 10b***

***Western blot used for Supplementary Figure 10b***

***Supplementary Figure 15. Uncropped western blot images***