Title Page

SOX2 drives bronchial dysplasia in a novel organotypic model of early human squamous lung cancer

Lúcia L Correia[1]

Jo-Anne Johnson[5]

Peter McErlean[2]

Julien Bauer[6]

Hassan Farah[2]

Doris M Rassl[3]

Robert C Rintoul[4]

Tariq Sethi[2]

Paul Lavender[2]

Emma L Rawlins[5]

Trevor D Littlewood[1]

Gerard I Evan[1]

Frank M McCaughan[1,2]

1. Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge CB2 1GA

2. Department of Asthma, Allergy and Lung Biology, 5th Floor. Tower Wing, Guy's Hospital, King's College London SE1 9RT

3. Department of Pathology, Papworth Hospital Foundation NHS Trust, Papworth Everard, Cambridge, CB23 3RE

4. Department of Respiratory Medicine, Papworth Hospital Foundation NHS Trust, Papworth Everard, Cambridge, CB23 3RE

5. Wellcome Trust/CRUK Gurdon Institute, University of Cambridge, Cambridge, CB2 1QN, UK.

6. Cambridge Genomic Services, Department of Pathology, Tennis Court Road, Cambridge CB2 1QP, UK.

Corresponding Author: Frank McCaughan fm319@cam.ac.uk, frank.mccaughan@kcl.ac.uk

Author Contributions:

LC performed most of the experimental work. FM conceived of the study, performed some of the experimental work and wrote the manuscript. JJ performed some experimental work. HF and TS contributed protocol advice. DMR and RCR contributed the tissue microarray. DMR reviewed histology. PM and PL contributed to study design and bioinformatic analyses. ER, GIE and TDL contributed to study conception and design. LC, RCR, JJ, ER, TDL, GIE helped with editing the manuscript.

Sources of Support: This work is supported by the Wellcome Trust. FM is a Wellcome Trust Intermediate Clinical Fellow (WT097143MA). TDL and GIE are supported by Cancer Research UK (C4750/A12077 and C4750/A19013). This work was also supported by the National Institute for Health Research (NIHR) Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust and King's College London. PL is supported by MRC Programme Grant G1100238. RCR and DMR are supported in part by the NIHR Biomedical

Research Centre in Cambridge and the Cambridge Cancer Centre. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health. The authors would like to thank Professor JW Shay and Professor JD Minna, University of Texas for the kind gift of the HBEC3 KT cells; Dr Luca Pellegrinet for assistance with flow cytometry; Nicholas Salisbury for A549 pTRIPZ-RFP cells and the histology core at the Wellcome Trust Medical Research Council Cambridge Stem Cell Institute for histological services. FUW-tetO-lox-hSOX2 and FUW M2rtTA were gifts from Rudolf Jaenisch (Addgene plasmids #s 20324, 20729) and pLVUHshp53-tTR-KRAB and pLVU-tTR-KRAB were gifts from Patrick Aebischer & Didier Trono (Addgene # 11645 and 11654). We thank Cambridge Genomic Services for able assistance with RNA-Seq and bioinformatic analyses.

Running Title: SOX2 and bronchial dysplasia

Descriptor: 9.26

Total Word Count:

Abstract Word Count: 250 words

At a glance commentary:

Scientific knowledge on the subject: There are no *in vitro* models of preinvasive squamous lung cancer that faithfully recapitulate the human disease. Here we show that SOX2 activation in a 3-D culture system and in the appropriate molecular context is sufficient to drive the dysplastic phenotype. This implicates SOX2 deregulation as a key early event in squamous carcinogenesis and provides

an excellent model for studying chemoprevention and the molecular pathogenesis of early disease.

This article has an online data supplement, which is accessible from this issue's table of content online at www.atsjournals.org.

Abstract

Rationale

Improving the early detection and chemoprevention of lung cancer are key to improving outcomes. The pathobiology of early squamous lung cancer is poorly understood. We have shown that amplification of SOX2 is an early and consistent event in the pathogenesis of this disease but its functional oncogenic potential remains uncertain. We tested the impact of deregulated SOX2 expression in a novel organotypic system that recreates the molecular and microenvironmental context in which squamous carcinogenesis occurs.

Objectives

1) To develop an *in vitro* model of bronchial dysplasia that recapitulates key molecular and phenotypic characteristics of the human disease
2) To test the hypothesis that SOX2 deregulation is a key early event in the pathogenesis of bronchial dysplasia
3) To use the model for studies on pathogenesis and chemoprevention

Methods

We engineer the inducible activation of oncogenes in immortalised bronchial epithelial cells. We use 3-dimensional tissue culture to build an organotypic model of bronchial dysplasia.

Measurements and Main Results

We recapitulate human bronchial dysplasia *in vitro*. SOX2 deregulation drives dysplasia, and loss of TP53 is a co-operating genetic event that potentiates the dysplastic phenotype. Deregulated SOX2 alters critical genes implicated in hallmarks of cancer progression. Targeted inhibition of AKT prevents the initiation of the dysplastic phenotype.

Conclusion

In the appropriate genetic and microenvironmental context acute deregulation of SOX2 drives bronchial dysplasia. This confirms it's oncogenic potential in human cells and affords novel insights into the impact of SOX2 deregulation. This model can be used to test therapeutic agents aimed at chemoprevention.

Word count 250 words

Key words: Early lung cancer, SOX2, bronchial dysplasia, squamous lung cancer, organotypic culture

Introduction

Lung cancer is one of the most common causes of cancer-related mortality in the world (1). Squamous lung cancer (SQC) accounts for approximately 30% of all lung cancers. There are a number of challenges to reducing the mortality from SQC and these include improving smoking cessation, early detection and diagnosis, and the development of novel approaches to the chemoprevention and treatment of lung cancer (2-4).

A key aspiration is to understand the molecular pathogenesis of SQC and to use this information to drive developments in early detection, chemoprevention and treatment (2). This would be significantly facilitated by model systems that faithfully recapitulate the human disease. There has recently been progress in the development of murine models of squamous lung cancer and these models will lead to new insights (5-7). However there are advantages to *in vitro* models, including the use of human cells, genetic tractability, throughput, costs and reduction in animal use.

The epithelial cell-microenvironment interaction is critical to the pathogenesis of epithelial malignancies, and to resistance to therapeutics (8). It has been repeatedly demonstrated that cancer cells behave differently and have very different phenotypic outcomes in complex compared to traditional culture systems (9). Advances in tissue engineering have made it possible to use *in vitro* models of the bronchial wall to investigate pathological mechanisms in asthma, COPD and cancer (10, 11). In the current study we have built on these studies to

build a rational model of bronchial dysplasia, the precursor lesion to squamous lung cancer.

The key molecular drivers in the early pathogenesis of squamous lung cancer (SQC) remain obscure. In a series of reports on a small number of patients it has been demonstrated that loss or mutation of p53 is an early event (12, 13) that appears to afford a clonal survival and expansion benefit, even at the earliest histopathological stages. Further, TP53 is disrupted in almost all invasive squamous lung cancers (14).

It has repeatedly been demonstrated that 3q amplification occurs prior to invasion in the pathogenesis of SQC (15). In previous work we have shown that 3q amplification marks the progression between low and high-grade dysplasia and has implications for prognosis (16). This has been confirmed in a larger cohort (17). It is likely that multiple genes on 3q can contribute to squamous carcinogenesis (18). However, previous mapping studies, both in SQC and bronchial dysplasia have identified SOX2 as a likely key target of the 3q amplicon (16, 19, 20).

Therefore the available molecular epidemiological evidence suggest that p53 is disrupted early in the pathogenesis of the disease and that 3q amplification and associated deregulation of SOX2 expression occurs at the epidemiological and histological bottle-neck between low- and high-grade bronchial dysplasia (16, 17, 21).

SOX2 is a single exon nuclear transcription factor. It is pleiotropic, with critical roles in stem cell and developmental biology and it is particularly important in the development of the lung (22, 23). There have been a number of reports regarding the potential for SOX2 overexpression to drive carcinogenesis in murine models. In one report deregulated SOX2 was unable to drive carcinogenesis when its overexpression was restricted to Club cells using the Scgb1a1 promoter (24); in a separate report marked overexpression was sufficient to form phenotypic adenocarcinomas that were positive for p63, a squamous carcinoma marker (25). More recently, the delivery of intratracheal lentivirus containing SOX2 in conjunction with loss of LKB1 was sufficient to drive squamous carcinogenesis, albeit at relatively low penetrance (6).

In vitro experiments using immortalised human cell lines have given varying results. Overexpression of SOX2 was sufficient to transform SV40 large T-antigen immortalised BEAS2B cells (20); virally immortalised small airway epithelial cells (AALE) were transformed by SOX2, but only in conjunction with co-operating genetic lesions (19).

We have modelled the earliest stages of bronchial epithelial carcinogenesis using non-virally immortalised human bronchial epithelial cells with a stable karyotype and an intact p53 signalling response/pathway (26). Using lentiviral constructs we sequentially knockdown TP53 and inducibly deregulate SOX2 expression. We establish short-term organotypic cultures and demonstrate that SOX2 deregulation, on a background of p53 knockdown in a confluent epithelial

monolayer at the air-liquid interface, drives the dysplastic phenotype *in vitro*. We show that SOX2 alters key cell signaling pathways, recapitulating signatures seen in invasive SQC and that it deregulates multiple genes implicated in hallmarks of cancer progression - including inhibition of apoptosis, driving cell proliferation, and cell migration and epithelial mesenchymal transition (EMT). We demonstrate that the model can be used to test chemoprevention strategies, and that a specific inhibitor of Akt that is already in early phase trials prevents the development of the dysplastic phenotype. This work establishes that SOX2, in the appropriate molecular and environmental context is sufficient to drive lung squamous carcinogenesis in human cells and establishes a novel model of human bronchial carcinogenesis with broad utility.

This work has been presented in part in abstract format at a number of scientific meetings (27-29).

Methods

Cell culture.

Human bronchial epithelial cells immortalized with hTERT and CDK4 (hereafter called KT cells) were kindly provided by Professors JW Shay and JD Minna (University of Texas Southwestern Medical Center, Dallas, Texas). The following plasmids were used to generate lentiviral particles using standard protocols – FUW M2rtTA and FUW tetO-lox-hSOX2 (30), pLVUHshp53-tTR-KRAB and pLVU-tTR-KRAB (31). KT cells were sequentially transduced with lentiviral particles to create a suite of cell lines with different genotypes. Cell lines were cultured in Keratinocyte Serum-Free Medium (KSFM) supplemented with human recombinant Epidermal Growth Factor (EGF) and Bovine Pituitary Extract (BPE) (Life Technology, Paisley, UK) and Bronchial Epithelial Basal Medium (BEBM, (Lonza, Basel, Switzerland)). A549 cells with a tetracycline inducible RFP reporter were maintained in DMEM/10% Foetal Bovine Serum (FBS).

Organotypic culture.

Organotypic cultures were established in 12-well transwell plates with a porous membrane insert (pore size 0.4μm) (Corning 3460, Corning, Weisbaden, Germany). On day 1, a mixture of collagen (Corning Collagen I, Rat Tail), 10X DMEM, 10% tetracycline low FBS, 7.5% sodium bicarbonate and $5x10^5$/ml MRC-5 fibroblasts was prepared on ice. This solution was inverted gently in order to obtain a homogenous solution without bubbles. The pH was adjusted to 7.25-7.3

with 1M sodium hydroxide and 600 µl of the mixture poured into each transwell. Plates were incubated at 37°C for 30 min and medium (KSFM plus supplements) was added to both the upper and lower chambers. The fibroblasts were left to contract the collagen for 5 days and HBECs then overlaid on the collagen layer. HBECs were kept covered until they reached confluence and on day 8 an air-liquid interface (ALI) was generated through the removal of medium from the upper chamber. Medium with or without 2-5µg/ml doxycycline supplementation in the lower chamber was changed every 48h and samples fixed on or after day 14 with 10% formaldehyde at 4°C prior to further processing. 4.5µm sections were prepared for immunofluorescence or immunohistochemistry using standard protocols. The following antibodies and dilutions were used: SOX2 (Cat No. MAB2018, R&D Systems-dilution 1:200), Ki67 (Cat No. RM-9106-S1, Thermo Scientific-dilution 1:200), p-AKT (Ser473) (Cat No. 3787, Cell Signaling Technology-dilution 1:50), p-ERK1/2 (Cat No. 4376, Cell Signaling Technology-dilution 1:1000) and Acetylated Tubulin (Cat No. T7451, Sigma-dilution 1:1000). For chemoprevention studies AZD5363 was dissolved in DMSO and added to the lower chamber to a final concentration of 10µM at the same time that doxycycline was added.

Differentiation culture

24-well transwells (Cat No. 353095, Corning) were coated with collagen $1x10^4$ cells were plated onto each transwell and submerged in 200µl of Complete PneumaCult™ expansion medium (Cat No. 05008, Stem Cell Technologies) plus penicillin/streptomycin, with 500µl in the lower chamber. On day 5, an air-liquid

interface was generated by the removal of the medium from the upper chamber. 500µl of PneumaCult™ ALI Maintenance medium plus penicillin/streptomycin was added to the lower chamber (Cat No. 05001, Stem Cell Technologies) and cells were washed with PBS and medium changed every 3 days. 23-day ALI cultures were then fixed with 4% PFA for 10min at room temperature. Membranes were cut from the transwells and cells permeabilised with PBS + 0.3% Triton for 15 min followed by blocking with PBS + 0.1% Triton + 5% NGS + 1% BSA for 45 min. Membranes were then incubated with the following antibodies and dilutions: Keratin 5 (Cat No. PRB-160P, Covance - dilution 1:500), Mucin 5AC (Cat. No. MS-145P0, Thermo Scientific - dilution 1:200), γ-Tubulin (Cat. No. SAB4503045, Sigma - dilution 1:200), Acetylated tubulin (Cat No. T7451, Sigma - dilution 1:1000) and CC10 (Cat No. sc25555, Santa Cruz Biotechnology – dilution 1:500). Nuclei were stained using Hoechst 33258 dye (Sigma), samples were mounted in Fluoromount Aqueous Mounting Medium (Sigma) and, analysed using an AxioImager compound microscope (Carl Zeiss).

Western blotting

Protein extracts were obtained using RIPA buffer. 40 µg of protein was loaded onto 4–12% Tris-Acrylamide gels followed by transference to a PVDF membrane (Millipore). The following antibodies were blotted: SOX2 (Cat No. MAB2018, R&D Systems-dilution 1:1000), p53 (Cat No. sc-126, Santa Cruz Biotechnology-dilution 1:1000), Keratin 5 (Cat No. ab24647, Abcam-dilution 1:1000), p63 (Cat No. ab735, Abcam-dilution 1:1000) and β-actin (Cat No. sc-47778, Santa Cruz Biotechnology-dilution 1:5000). Blotting results were detected by an ECL

chemiluminescence kit (Thermo Scientific) before exposing to X-ray film (Fujifilm).

Cell culture assays and cell cycle analysis

For cell growth assays in standard culture $0.1x10^5$ cells from each cell line were seeded with or without 1µg/ml doxycycline. Cells were trypsinised at specified timepoints and counted using the Countess Automated Cell Counter (ThermoFisher Scientific, Paisley, UK). For cell cycle analysis**,** $0.1x10^6$ cells were plated in a 6 well plate. For 3D cultures, to harvest the cells at the air-liquid interface medium was removed from both lower and upper chamber and the porous membrane of each transwell wrapped in cling film. Collagen discs were washed twice with PBS and treated with trypsin for 12 min at 37˚C. Cells were collected in DMEM/10% FBS and the pellet resuspended in phosphate-buffered saline-5mM EDTA. Cells were fixed and stained with propidium iodide. Samples were analysed on an Accuri C6 (BD Biosciences, UK) and further analysis was performed using FlowJo software (FlowJo LLC, Ashland, Oregon, USA).

Tissue microarray

A tissue microarray was constructed using surgical and endobronchial patient biopsy specimens from Papworth Hospital NHS Trust, Cambridge, UK. Biopsies assessed by a consultant thoracic histopathologist to be high-grade bronchial dysplasia were used for this study (limited to one core per individual patient).

Chromatin Immunoprecipitation-PCR

Bronchial epithelial cells were cultured in the organotypic culture system (OTC) and untreated/treated with 2μg/ml doxycycline for 8 days prior to being trypsinised and harvested. Cells were processed using the Magna ChIP™ G - Chromatin Immunoprecipitation kit as per manufacturer's instructions (Millipore, UK). Chromatin shearing was performed using Covaris S2 series and the following programme: Duty cycle: 2%, Intensity: 3, Cycles per burst: 200, for 12 min. 50 μl of sheared chromatin was incubated with magnetic beads and the following antibodies: SOX2 (Cat No. AF2018, R&D systems- 5μg) and IgG control (Cat No. AB-108-C, R&D systems- 5μg). Samples were validated using the manufacturer's recommended antibody followed by PCR of the eluted DNA (ChIPAb+ RNA Pol II - ChIP Validated Antibody and Primer set, Millipore).

RNA extraction and quantitative PCR (QPCR):

RNA was extracted from cells cultured in 3D, untreated/treated with 2μg/ml doxycycline for 8 days, using the miRNeasy Mini Kit (Qiagen, Manchester, UK). Collagen disks were solubilized with 1ml of QIAzol Lysis Reagent (Qiagen, Manchester, UK) by vortexing and RNA was isolated following manufacturer's instructions. RNA was eluted with RNase-free water and treated with DNase I (Thermo Scientific, Paisley, Scotland). cDNA was prepared with 1μg of total RNA using the High Capacity cDNA Reverse Transcription Kit (Thermo Scientific, Paisley, Scotland). Quantitative PCR was performed using SYBR green reagents and GAPDH as internal control. Thermal cycling conditions were as follows: 95°C for 3min, 40 cycles at 95°C for 15s and 55°C for 30s. The relative amount of mRNA in each sample was quantified using ΔΔCt method.

RNA sequencing

Sample sequencing libraries were prepared from 200ng of total RNA extracted from treated and control organotypic cultures four days after the addition of doxycycline. The TruSeq Stranded mRNA HT sample preparation kit (Illumina, Chesterford UK) was used according to the manufactures instructions. Samples were individually indexed for pooling using a dual index strategy. Libraries were quantified on a Tapestation DNA 1000 Screen tape (Agilent, Cheadle UK) and by qPCR using an NGS Library Quantification Kit (KAPA Biosystems, London UK) on an AriaMx qPCR system (Agilent). Libraries were then normalised, pooled, diluted and denatured for sequencing on the NextSeq 500 (Illumina) according to the manufacturer's instructions. Samples were pooled such that a minimum of 25M unique clusters per sample was achieved. PhiX control library (Illumina) was spiked into the main library pool at 1% v/v for quality control purposes. Sequencing was performed using a high output flow cell with 2x75 cycles of sequencing providing 800M paired end reads from 400M unique clusters.

RNA-seq was performed and analysed as a service by Cambridge Genomic Services (Department of Pathology, University of Cambridge). RNA was sequenced (2x75bp) using the Illumina NextSeq500. The quality control of the reads was evaluated using FastQC (32) and reads were trimmed using Trim Galore (33). Reads were mapped to the reference human genome (Ensembl Homo_sapiens.GRCh38.80 GTF file) using STAR (34). The number of reads that map to a genomic feature was calculated using HTSeq (35) and differential expression analysis was performed using the counted reads and the R package edgeR (v3.8.6) (36). We used the classic edgeR approach to make pairwise

comparisons between treated and control groups. This involves computing an exact test for each gene and correcting for multiple testing. EdgeR reported a p-value for differential expression and a False Discovery Rate (FDR) adjusted for multiple testing. Pathway analysis was undertaken using Ingenuity Pathway Analysis (IPA, Qiagen, USA) restricted to significantly altered genes at a p-value of $<1x10^{-6}$. An expression heatmap was produced using Multiple Experiment Viewer and plotted by scaling the normalised counts per million (CPM) of each gene across biological replicates (37). Sequencing reads are deposited in a public database hosted by the National Centre for Biotechnology Information (SRA Bioproject Number PRJNA344067).

Statistics

Data was analysed using Graphpad Prism using standard settings (California, USA). Details of specific analyses are provided in figure legends.

Results

HBEC-KT cells were sequentially transduced with lentiviral constructs to create a series of cell lines with multiple genotypes including TRE-SOX2 (iSOX2) (Figure 1a). SOX2 expression was readily induced in the presence of doxycycline in those cell lines that expressed both SOX2 and the reverse tetracycline transactivator (M2rtTA). Upregulation of SOX2 was apparent after 24 hours in standard tissue culture. Further genotypes included a stable knockdown of TP53 alone (p53lo) and a compound genotype with inducible SOX2 and TP53 knockdown (iSOX2p53lo) The expression of the key markers associated with squamous differentiation – Keratin 5 (K5) and p63 – was retained after SOX2 induction with doxycycline in cells with iSOX2 alone and iSOX2/p53lo.

We established a three-dimensional organotypic culture system (OTC) in which SOX2 expression could be induced after cells were established in a monolayer at the air-liquid interface (Figure 2a). Using A549 cells transduced with a lentiviral reporter vector, we demonstrated that the addition of doxycycline to the medium in the lower chamber was sufficient to induce expression of a red fluorescent reporter in the bronchial epithelial monolayer (Figure 2b). The successful induction of SOX2 in KTiSOX2/KTiSOXp53lo cells cultured at the air-liquid interface in the OTC was demonstrated (Figure 2c). We then demonstrated that the genetically manipulated KTiSOX2p53lo cells retain the previously reported potential of the parental cells (38) to differentiate into the constituent cells of the adult human airway when cultured at the air-liquid interface on collagen coated transwell inserts and using bronchial differentiation medium. The KTiSOX2p53lo

cells differentiated into ciliated cells, mucin-secreting cells and Club cells as well as basal cells (Figure 2d-h). Therefore despite multiple rounds of genetic manipulation these cells retain the capacity for “normal” differentiation.

The response to SOX2 induction was compared in standard cell population assay conditions and the OTC system using cell cycle analysis. SOX2 induction in standard conditions did not lead to an increase in cell numbers (Figure 3a) and there was no alteration in the cell cycle profile (Figure 3b). However, in the organotypic system induction of SOX2 in the confluent bronchial epithelial monolayer led to a significant increase in cells in S/G2/M (Figure 3b) and in total cell number (Figure 3 c)

Consistent with the cell cycle results, acute deregulation of SOX2 expression led to an early phenotypic difference in the organotypic model; using phase contrast microscopy, as early as 72 hours after the induction of SOX2 expression there were focal “outgrowths” in the confluent monolayer with cells appearing to grow on top of each other. Loss of p53 was also sufficient for modest focal outgrowths to appear in the monolayer. Stable transduction with the equivalent empty lentivirus had no phenotypic impact (Supplementary Figure 1). The combination of SOX2 deregulation and p53 knockdown led to a much more dramatic phenotype, with diffuse outgrowths (Figure 3 d-i).

Histological analysis corroborated these findings (Figure 3j-o). The ‘outgrowths’ were areas in which there was increased thickness of the epithelial layer with apparent loss of contact inhibition and cells protruding into the putative

bronchial lumen. The cells also had a high nuclear to cytoplasmic ratio. These outgrowths were focal for the KTiSOX2 genotype and more diffuse for the iSOX2p53lo genotype. They were consistent with focal areas of dysplasia arising from an epithelial monolayer.

We performed immunofluorescence and immunohistochemistry and demonstrated that the dysplastic foci in doxycycline-treated KTiSOX2 and KTiSOX2p53lo cultures were positive for SOX2 and that there were no SOX2 positive cells in the control (untreated) cultures (Figure 4 a-b, Supplementary Figure 2).

We next assessed the impact of deregulated SOX2 expression on key downstream signaling pathways. In squamous lung cancer there is typically upregulation of phosphoAKT (pAKT) and downregulation of phosphoERK (pERK) - signaling features that distinguish it from adenocarcinoma (6). SOX2 expression was associated with upregulated pAKT in dysplastic foci and a marked reduction in intensity of cytoplasmic pERK staining, recapitulating the signaling patterns reported in more advanced human squamous lung cancer (Figure 4 c-j).

We compared the findings in this organotypic model with a series of high-grade preinvasive lesions from a tissue microarray. Lesions from 15 patients were assessed; 9 of which were from patients with simultaneous invasive squamous lung cancer. SOX2 is expressed in normal human bronchial epithelium and in basal cell hyperplasia but at steady state Ki67 positive cells are relatively rare

and tend to be adjacent to the basement membrane (Figure 5 a, b and Supplementary Figure 3). SOX2 was expressed in high-grade dysplastic lesions from 14 of 15 patients, qualitatively at a higher level than that seen in normal bronchial epithelium or basal cell hyperplasia (Figure 5 c, e, Supplementary Figure 3), and it was associated with an increased proliferative index (Figure 5 d, f). Unfortunately, staining for pAKT was unreliable in these archived specimens that had undergone variable processing. However, in one lesion in which a clear transition from normal epithelium to high-grade dysplasia was captured pAKT expression was restricted to a few cells abutting the basement membrane in the pseudostratified epithelium, whereas in the dysplastic lesion there was significant upregulation of SOX2, Ki67 and pAKT throughout the epithelium (Figure 5 g-i).

SOX2 is a pleiotropic nuclear transcription factor that exerts variable phenotypic effects in different cell types and cell culture systems (39-41). Deregulated SOX2 has been shown to directly regulate CyclinD1 and BCL2 in breast cancer and melanoma respectively (39, 40). We performed ChIP-PCR on KTiSOX2p53lo cells at the air-liquid interface in the presence and absence of doxycycline. We demonstrated that CYCLIND1 and BCL2 promoters were specifically bound by SOX2 (Figure 6 a) and that both target oncogenes were upregulated (Figure 6 b-c). A broader survey of key cell cycle regulatory proteins showed that many were upregulated (Figure 6 c) and that CDKN1A, a key negative regulator of the cell cycle (p21), was downregulated.

The results are consistent with SOX2 acting as a mitogenic oncogene and targeting, in this novel organotypic system, key downstream effectors

anticipated from other experimental systems. SOX2 is a pleiotropic nuclear transcription factor that controls the expression of large gene sets that vary depending on the cell type and experimental system (42, 43). We therefore undertook RNA-Seq analysis to perform an unbiased transcriptomic analysis of the broader impact and potential downstream effectors of deregulated SOX2 in this OTC model of bronchial dysplasia. Over two thousand genes were significantly deregulated by SOX2 induction ($p<0.01$) and there was clear segregation between the SOX2-induced and non-induced experiments (Supplementary Figure 4, SRA Bioproject Number PRJNA344067). Further analysis was undertaken using Ingenuity Pathway Analysis with a cut-off of significance of $p<1x10^{-6.}$ (Qiagen). Importantly, and consistent with the observed phenotype in the OTC system, the most significantly altered cellular functions associated with deregulated SOX2 expression were "Cellular movement" and "Cellular growth and proliferation" (Figure 7 a). The 197 genes annotated as "cell migration" show clear segregation between SOX2-deregulated and control experiments and include many genes previously implicated in carcinogenesis and known direct targets of SOX2 such as ETV4 and DKK1 (44, 45) (Figure 7 b and Supplementary Table 1).

Epithelial mesenchymal transition (EMT) is a hallmark of cancer implicated in the critical early stages of the disease including cell migration. We used QPCR to corroborate RNA-Seq results and analyse an extended panel of canonical mediators of EMT (Figure 7 c). The RNA-Seq data also implicated SOX2 in the upregulation of emerging drivers of EMT and cell migration phenotypes – including AGR2, CEMIP (KIAA1199) and SERPINI1 (46-48)- also confirmed using QPCR (Figure 7 d). Of these, SERPINI1 is most frequently altered in early lung

squamous cell carcinoma, often in conjunction with SOX2 (Figure 7 e). We confirmed that SERPINI1 upregulation was accompanied by protein overexpression (Figure 7 f). We then optimised primer sets around the SERPINI1 transcriptional start site to demonstrate that it was a direct transcriptional target of SOX2 (Figure 7 g). We then corroborated this with clinical samples and demonstrated that SERPINI1 is expressed in high-grade SOX2 positive bronchial dysplastic lesions but not in basal cell hyperplasia (Figure 7 h). We have previously demonstrated that TP53 mutated, SOX2-amplified clonal populations expand, migrate and colonise the respiratory epithelium. The current data comprehensively links SOX2 mechanistically to this clinically important preinvasive migratory phenotype (49),.

Finally, building on the immunohistochemistry (Figure 5) and the published literature linking activation of the PI3K/AKT pathway and bronchial preneoplasia (50-53), we assessed the impact of a specific inhibitor of Protein Kinase-B (Akt) in preventing the development of SOX2-driven bronchial dysplasia in this organotypic model system. AZD5363 is a potent and specific inhibitor of AKT1-3 that is already in clinical trials (54). AZD5363 had a clear impact on reducing the dysplastic phenotype (Figure 8 a-d). No dysplastic outgrowths were seen on phase contrast microscopy in the treated cultures and this was confirmed by histology. However the epithelial cells in contact with the collagen basement membrane and at the air-liquid interface were apparently healthy, entering cell cycle and expressing SOX2 (Figure 8 e-f), indicating that there was not general severe cellular toxicity as a result of the AZD5363 treatment.

Discussion

In this work we provide direct evidence that deregulation of a putative oncogene, SOX2 - in the appropriate molecular and cellular context; and with the appropriate microenvironmental cues (collagen, fibroblasts, air-liquid interface) – is sufficient to drive bronchial dysplasia in a short-term three-dimensional culture system.

The cellular, molecular and environmental context in which an oncogene becomes deregulated is critical to its impact. The cells used in this study are immortalised human bronchial epithelial cells (HBECs) that have been well characterised. They have an almost normal karyotype, intact p53 signalling and express markers, K5 and p63, that are associated with airway basal cells - the putative cells of origin of squamous lung cancer (SQC) (26). They have previously been used in experiments examining the role of mutated KRAS with or without the expression of MYC in carcinogenesis (11, 55). KRAS mutations are very common in lung adenocarcinoma but very rare in SQC. In this work we model the key events reported in bronchial dysplasia/early squamous carcinogenesis.

When cultured using specific medium and at the air-liquid interface the HBECs can differentiate into the constituent cells of the bronchial epithelium including secretory (Club) and ciliated cells (38). If cultured in matrigel they have the capacity to form branching structures and express differentiation markers

consistent with distal airway differentiation (38). In this study we show that, notwithstanding multiple rounds of defined genetic manipulation to create the complex genotype we sought, they retain their capacity to differentiate into constituent airway cell types. In contrast, a commonly used SV40 large T antigen immortalized bronchial epithelial cell line - BEAS2B – reportedly does not differentiate at the air-liquid interface (56).

The limited available molecular epidemiological data from human samples suggests that p53 is disrupted early in the pathogenesis of SQC and that 3q amplification defines the transition between low and high-grade dysplasia (12, 16, 57, 58). The HBECs in this study are immortalised by the introduction of two specific genes, CDK4 and hTERT. CDK4 overexpression overcomes the cell cycle inhibition mediated by the p16INK4A-RB tumour suppressor pathway and hTERT (telomerase) facilitates clonal expansion through crisis (59). There are limited data on CDKN2A (p16INK4A locus at 9p21) in preinvasive disease but Wistuba et al reported loss of heterozygosity at 9p21 in dysplasia and carcinoma-in-situ (58), and the CDKN2A locus is disrupted in up to 75% of SQCs (14). Telomerase overexpression increases with dysplastic progression in immunohistochemical studies (60). Therefore, the deregulated overexpression of SOX2 in the current model is in a rational cellular and molecular context that reflects the human disease.

Many groups have previously demonstrated the importance of the physical and cellular microenvironment in the phenotypic response to oncogenic stimuli. The data we present are consistent with these studies and emphasise the significant

benefit of using organotypic as opposed to traditional cell culture. The system adopted here reflects that used by others studying the pathobiology of bronchial epithelial cells. We show that it is possible to genetically manipulate the epithelial cells and induce deregulated expression of SOX2 when an intact respiratory epithelial monolayer at the air-liquid interface is already established.

SOX2 has previously been implicated as an oncogene in the pathogenesis of SQC. In a mouse model of SQC almost 50% of mice develop cancer 9 months after inhalation of a virus that contains SOX2 and deletes LKB1. The majority of the cancers that developed were of the squamous subtype (6). SOX2 was also sufficient to transform SV40 large T antigen transformed BEAS2B cells (20) but was insufficient to transform similarly immortalised small airway epithelial cells (AALE) alone (19). SV40 large T antigen has multiple impacts on key tumour suppressor pathways pathways including p53 and RB; and these cell lines have not been characterized karyotypically. Therefore, it is difficult to measure the impact of SOX2 in these cellular contexts.

Here we provide novel data regarding the impact of SOX2 in an organotypic human system with a well-characterised cell line that has been non-virally immortalised and has an almost normal karyotype. The data demonstrate that SOX2, in the appropriate molecular and physical environmental context, drives dysplastic change. Moreover its impact when deregulated is relatively acute – outgrowths from the monolayer are readily visible at 72 hours. Further, the combination of SOX2 deregulation and p53 knockdown leads to an impressive dysplastic phenotype. This is entirely consistent with what has been previously

demonstrated in studies on human specimens: p53 loss/disruption is likely to be a very early event in the pathogenesis of this disease and SOX2 deregulation is superimposed on it at the key epidemiological and molecular bottlenecks between low and high-grade dysplasia (12, 13, 16). We observed in this model similar signaling alterations to those reported in the invasive disease and previously reported in preinvasive disease (51, 52). We then used a unique tissue microarray of human specimens to demonstrate that the increased proliferative index seen in high-grade dysplastic lesion co-exists with an increase in SOX2 expression.

SOX2 is a pleiotropic nuclear transcription factor that exerts variable phenotypic effects in different cell types and cell culture systems. In driving dysplastic change in this model system we show that it directly regulates the expression of key proteins implicated in carcinogenesis – CyclinD1 and Bcl2, a negative regulator of apoptosis. We further show that overexpression of SOX2 upregulates multiple kinases involved in cell cycle progression and downregulates p21, a potent cyclin-dependent kinase inhibitor. Taken together with the presented Ki67 and cell cycle data this shows that SOX2 is driving cell proliferation, consistent with the results seen in the TMA staining of human specimens. Further, we demonstrated that the phenotypic changes associated with deregulated SOX2 are accompanied by a broad transcriptional programme encompassing multiple hallmarks of cancer - cell proliferation, inhibition of apoptosis and cell migration, including canonical and emerging genes associated with EMT and metastasis. Serine protease inhibitors have diverse functional roles but recent data strongly associates SERPINI1 with EMT and lung cancer

metastatic cerebral invasion. This is the first study to show that SOX2 is a direct transcriptional regulator of *SERPINI1* and to demonstrate that it is overexpressed at a preinvasive stage in clinical specimens. SOX2 orchestrates a complex genetic programme resulting in the dysplastic phenotype. Although nuclear transcription factors are notoriously challenging drug targets, emergent medicinal chemistry strategies, and these data, strongly support the direct therapeutic targeting of SOX2.

EMT is regarded as a key early event in the pathogenesis of epithelial cancers. We, and others, have shown previously that cell migration is also a key early feature of squamous lung cancer; and that SOX2 amplified/p53 disrupted clonal populations *in vivo*, as reproduced in the OTC, have the capacity to migrate and colonise the bronchial epithelium (12, 13). Taken together, these data suggest that SOX2-driven cell migration is a mechanism underpinning the so-called field-change effect.

No model system can recapitulate the *in vivo* disease perfectly. In previous work we have shown that the amplification of SOX2 is likely to be an incremental process (13) whereas in this case SOX2 is turned on over a very short time period. It is not possible to correlate the SOX2 expression levels in the OTC to human preinvasive specimens. However, we have shown that SOX2 is amplified and significantly overexpressed in preinvasive lesions (16). Further, there are clear data from the TCGA studies that show that the expression of SOX2 is grossly elevated in SQC compared to other tumour types (14) (Supplementary Figure 5). The *in vivo* complexity in terms of immunological response and the

microenvironment cannot yet be recapitulated *in vitro*. Finally, doxycycline binds collagen and data using a fluorescent reporter suggests that it continues to be present at levels that maintain transgene expression for at least 3 weeks after withdrawal from the medium (not shown). Therefore withdrawal experiments to establish the need for continued SOX2 expression in the maintenance of these outgrowths were not possible.

The confirmation of the impact of SOX2 overexpression as an early driver of bronchial carcinogenesis and the ability to phenocopy the human disease in short-term culture are important steps forward that should help to meet a key challenge - the development of agents to use for primary and secondary chemoprevention (61). We have begun to address this challenge - we demonstrated that AZD 5363, a targeted inhibitor of Akt already in early phase clinical trials completely abrogated the development of the dysplastic phenotype. This is consistent with prior studies suggesting that the PI3K-Akt axis may be an appropriate target for intervention in bronchial preneoplasia including a landmark clinical trial that showed that this pathway could be targeted in a chemoprevention strategy (50-53). The current study mimics using AZD5363 to prevent the emergence of bronchial dysplasia. It has a side effect profile that would be likely to preclude its use in primary chemoprevention of at-risk groups but there may be patients in whom a trial of this or a similar compound may be appropriate in the secondary chemoprevention setting.

This platform will be suitable for adaptation to moderate throughput screening of potential chemoprevention compounds and further screens for potential

complementary genetic lesions that may drive the progression of this devastating disease.

Figure Legends

Figure 1

KT cells were sequentially transduced with lentiviral constructs to create a series of cell lines with different genotypes (a). A mean of 49% of cells were positive for SOX2 on induction with doxycycline (b). Western blot with a range of the cell lines used with different genotypes (c). The expression of SOX2 was tightly controlled and only induced in the presence of M2rtTA and doxycycline. Cells were positive for Keratin 5 and p63 in the presence and absence of doxycycline. (1b bar=100μm).

Figure 2

Cell lines were established in short-term culture on top of a disc consisting of Type I collagen and pulmonary fibroblasts (a). Supplementation of medium in the lower chamber with doxycycline induces the expression of red fluorescence in A549 cells transduced with an inducible RFP reporter gene that are in established confluent culture at the air-liquid interface (ALI) (b). Transduced HBECs were established in the organotypic system described in (a); doxycycline supplementation in the lower chamber led to induction of SOX2 expression in the monolayer (c). Genetically manipulated KTiSOX2p53lo cells retain the capacity to differentiate into constituent cells of the human airway epithelium (d-h). ALI transwell cultures using differentiation medium at day 23 show a mixed

population of cells with clear differentiation into ciliated cells (d,e) with acetylated tubulin representing mature cilia in red; (f) mucin-secreting (goblet) cells (MUC5AC-green, γ-tubulin/cilia basal bodies–red); (g) club cells (CC10 – green; acetylated tubulin-red); (h) and basal cells (KRT5 – green, acetylated tubulin-red). (2b bar=400μm; 2d bar=50μm).

Figure 3

A cell population growth assay was performed in the presence and absence of doxycycline (a). Induction of SOX2 expression did not lead to an increase in the number of cells over 120 hours. Induction of SOX2 expression did not alter cell cycle distribution in standard culture conditions but led to a significant increase in the proportion of cells in S-G2/M in the organotypic culture (b, paired student t-test, 2-tailed) and in total epithelial cell number (c, paired student t-test, 2-tailed). Data presented in (a) represents means (+/- s.d.) of three separate experiments in triplicate. Control cell lines (KTM2) were treated with doxycycline with no significant alteration in morphology on phase contrast or histology (d, e). Loss of p53 was led to modest focal outgrowths in the monolayer (h). Induction of SOX2 led to an obvious phenotype with focal (KTiSOX2, f, g) and more diffuse changes (KTiSOX2p53lo, h,i) in the epithelial monolayer. Histological analysis was consistent with the changes seen using phase contrast microscopy (j-o,). There were diffuse outgrowths demonstrated in the KTiSOX2p53lo upon induction of SOX2 expression in the confluent monolayer. The histological features were typical of high-grade preinvasive lesions – loss of contact inhibition, a thickened epithelial layer with dysplasia, absence of

maturation and a high nuclear to cytoplasmic ratio. (3d-f: bar=400um; 3g-i: bar=1000um; 3j-o: bar=100um).

Figure 4

Activation of SOX2 leads to focal dysplastic changes and alterations in key cell signalling cascades. Dysplastic lesions in doxycycline-treated cultures were strongly positive for SOX2 with no SOX2 staining in the untreated cultures (a, b). Serial sections from dysplastic lesions show a downregulation in pERK and an upregulation in pAKT in those cells that are SOX2 positive and in cycle (Ki67 positive) (c-j). (bar=100μm).

Figure 5

Representative images from a tissue microarray with high-grade preinvasive bronchial lesions. Basal cell hyperplasia has weakly positive staining for SOX2 and rare Ki67 +ve cells (a, b). High-grade dysplastic lesions exhibit much more marked SOX2 and Ki67 staining that extends throughout the epithelial layer (c-f). pAKT staining proved challenging on archived TMA specimens; however there was a single sample in which the border between normal ciliated pseudostratified epithelium and high-grade dysplasia was available that provided a robust internal control for the antibody performance. There was a

clear upregulation in pAKT staining in those dysplastic cells in which SOX2 and Ki67 were upregulated. (bar=100μm).

Figure 6

Deregulated SOX2 targets key cancer-related genes in the organotypic culture system. ChIP-PCR was performed using input cross-linked DNA and the same DNA incubated with both a SOX2 specific antibody and an IgG control. There was enrichment for the BCL2 and CCND1 (CyclinD1) promoters in the SOX2 pulldown treated with doxycycline but not the untreated (No Dox) nor the IgG control confirming that SOX2 binds both gene promoters and regulates their transcription in this system (a). QPCR experiments confirmed that BCL2 and CCND1 were both upregulated transcriptionally on induction of SOX2 (b, c). The expression of a series of key cell cycle genes was analysed. There was upregulation of key cyclin-dependent kinases (CCNA2, CCNB1, CCNB2, CCNE2); and downregulation of CKN1A (p21), a critical inhibitor of cell cycle progression (c). These alterations were seen on treatment with doxycycline compared to the untreated control. The following kinases were also tested but were not significantly altered: CCNA1, CCNE1, CCND2, CCND3. The results shown are representative of three independent experiments and data is presented as the mean (n = 3) ± standard deviation. For the QPCR experiments ΔΔCt values were log (base 10) transformed and results for each marker was compared to GAPDH. Statistical analysis: a student's t-test (unpaired) was performed to test significance: ns p>0/05, * p≤ 0.05, ** p≤ 0.01, *** p ≤ 0.001 and ****p≤ 0.0001.

Figure 7

The most significantly differentially expressed genes ($p<1x10^{-6}$, n=832) were analysed using Ingenuity Pathway Analysis. The most significantly altered molecular and cellular functions were cellular movement and cellular proliferation (a). A heatmap of 197 genes associated with cell migration illustrating up and downregulation of genes with induction of SOX2 is presented (b). The results show 3 paired replicates with (italics) and without (normal font) SOX2 induction. A number of genes are highlighted including reported direct targets of SOX2 - DLL3 and ETV4; and those directly implicated in EMT (CDH2, CEMIP). Further details are in Supplementary Table 1. QPCR of canonical mediators of EMT was performed to corroborate RNA-Seq results and confirmed their upregulation by SOX2 (c). In addition a number of the most significantly altered genes in differential gene expression analyses – CEMIP, SERPINI1 and AGR2 (all $p<1x10^{-13}$) are recently implicated in EMT in the literature but not annotated in the "cellular movement" group. QPCR experiments confirmed significant upregulation in each case (d). Interrogation of the TCGA database showed that SERPINI1 was the most commonly altered of these in early squamous lung cancer and commonly altered with SOX2 (e). The corresponding protein SerpinI1 was also overexpressed on SOX2 induction in the OTC (f) and ChIP-PCR confirmed enrichment for SOX2 binding near the SERPINI1 transcriptional start site (g). Further sections from the TMA were stained for SerpinI1 and showed clear cytoplasmic staining with in high-grade dysplasia compared to normal epithelium. This is particularly reinforced by the section

containing the transition from relatively normal pseudostratified epithelium to disorganised high-grade dysplasia.

Figure 8

Representative phase contrast (a-b) and H&E stains (c-d) of doxycycline treated (a,c) and doxycycline and AZD5363 treated (b, d) organotypic cultures. All test and control cultures were treated with DMSO. As before induction of SOX2 led to a diffuse dysplastic phenotype. Treatment with a pan AKT inhibitor (AZD5363, 10μM) abrogated the development of the dysplastic lesions (b, d). Cells in the treated culture were alive and continued to express SOX2 (e) and Ki67 (f). Images are representative of 3 independent experiments. (7a,b: bar=1000μm; c-f: bar=100μm).

References

1. CRUK. Cancer Mortality Statistics. 2014 [cited 2014 July 2014]. Available from: http://www.cancerresearchuk.org/cancer-info/cancerstats/mortality/.
2. Keith RL, Miller YE. Lung cancer chemoprevention: current status and future prospects. *Nature reviews Clinical oncology* 2013; 10: 334-343.
3. Spira A, Beane JE, Shah V, Steiling K, Liu G, Schembri F, Gilman S, Dumas YM, Calner P, Sebastiani P, Sridhar S, Beamis J, Lamb C, Anderson T, Gerry N, Keane J, Lenburg ME, Brody JS. Airway epithelial gene expression in the diagnostic evaluation of smokers with suspect lung cancer. *Nat Med* 2007; 13: 361-366.
4. Gesthalter YB, Vick J, Steiling K, Spira A. Translating the transcriptome into tools for the early detection and prevention of lung cancer. *Thorax* 2015; 70: 476-481.
5. Xiao Z, Jiang Q, Willette-Brown J, Xi S, Zhu F, Burkett S, Back T, Song NY, Datla M, Sun Z, Goldszmid R, Lin F, Cohoon T, Pike K, Wu X, Schrump DS, Wong KK, Young HA, Trinchieri G, Wiltrout RH, Hu Y. The pivotal role of IKKalpha in the development of spontaneous lung squamous cell carcinomas. *Cancer Cell* 2013; 23: 527-540.
6. Mukhopadhyay A, Berrett KC, Kc U, Clair PM, Pop SM, Carr SR, Witt BL, Oliver TG. Sox2 cooperates with Lkb1 loss in a mouse model of squamous cell lung cancer. *Cell reports* 2014; 8: 40-49.
7. Xu C, Fillmore CM, Koyama S, Wu H, Zhao Y, Chen Z, Herter-Sprie GS, Akbay EA, Tchaicha JH, Altabef A, Reibel JB, Walton Z, Ji H, Watanabe H, Janne PA, Castrillon DH, Rustgi AK, Bass AJ, Freeman GJ, Padera RF, Dranoff G, Hammerman PS, Kim CF, Wong KK. Loss of Lkb1 and Pten leads to lung squamous cell carcinoma with elevated PD-L1 expression. *Cancer Cell* 2014; 25: 590-604.
8. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell* 2011; 144: 646-674.
9. Slorach EM, Chou J, Werb Z. Zeppo1 is a novel metastasis promoter that represses E-cadherin expression and regulates p120-catenin isoform expression and localization. *Genes Dev* 2011; 25: 471-484.
10. Choe MM, Tomei AA, Swartz MA. Physiological 3D tissue model of the airway wall and mucosa. *Nat Protoc* 2006; 1: 357-362.
11. Sato M, Vaughan MB, Girard L, Peyton M, Lee W, Shames DS, Ramirez RD, Sunaga N, Gazdar AF, Shay JW, Minna JD. Multiple oncogenic changes (K-RAS(V12), p53 knockdown, mutant EGFRs, p16 bypass, telomerase) are not sufficient to confer a full malignant phenotype on human bronchial epithelial cells. *Cancer Res* 2006; 66: 2116-2128.
12. Franklin WA, Gazdar AF, Haney J, Wistuba, II, La Rosa FG, Kennedy T, Ritchey DM, Miller YE. Widely dispersed p53 mutation in respiratory epithelium.

A novel mechanism for field carcinogenesis. *J Clin Invest* 1997; 100: 2133-2137.
13. McCaughan F, Pipinikas CP, Janes SM, George PJ, Rabbitts PH, Dear PH. Genomic evidence of pre-invasive clonal expansion, dispersal and progression in bronchial dysplasia. *J Pathol* 2011; 224: 153-159.
14. Network TCGA. Comprehensive genomic characterizationof squamous cell lung cancers. *Nature* 2012; 489: 519-525.
15. Qian J, Massion PP. Role of chromosome 3q amplification in lung cancer. *J Thorac Oncol* 2008; 3: 212-215.
16. McCaughan F, Pole JC, Bankier AT, Konfortov BA, Carroll B, Falzon M, Rabbitts TH, George PJ, Dear PH, Rabbitts PH. Progressive 3q amplification consistently targets SOX2 in preinvasive squamous lung cancer. *Am J Respir Crit Care Med* 2010; 182: 83-91.
17. van Boerdonk RA, Daniels JM, Snijders PJ, Grunberg K, Thunnissen E, van de Wiel MA, Ylstra B, Postmus PE, Meijer CJ, Meijer GA, Smit EF, Sutedja TG, Heideman DA. DNA copy number aberrations in endobronchial lesions: a validated predictor for cancer. *Thorax* 2014; 69: 451-457.
18. Qian J, Hassanein M, Hoeksema MD, Harris BK, Zou Y, Chen H, Lu P, Eisenberg R, Wang J, Espinosa A, Ji X, Harris FT, Rahman SM, Massion PP. The RNA binding protein FXR1 is a new driver in the 3q26-29 amplicon and predicts poor prognosis in human cancers. *Proc Natl Acad Sci U S A* 2015; 112: 3469-3474.
19. Bass AJ, Watanabe H, Mermel CH, Yu S, Perner S, Verhaak RG, Kim SY, Wardwell L, Tamayo P, Gat-Viks I, Ramos AH, Woo MS, Weir BA, Getz G, Beroukhim R, O'Kelly M, Dutt A, Rozenblatt-Rosen O, Dziunycz P, Komisarof J, Chirieac LR, Lafargue CJ, Scheble V, Wilbertz T, Ma C, Rao S, Nakagawa H, Stairs DB, Lin L, Giordano TJ, Wagner P, Minna JD, Gazdar AF, Zhu CQ, Brose MS, Cecconello I, Jr UR, Marie SK, Dahl O, Shivdasani RA, Tsao MS, Rubin MA, Wong KK, Regev A, Hahn WC, Beer DG, Rustgi AK, Meyerson M. SOX2 is an amplified lineage-survival oncogene in lung and esophageal squamous cell carcinomas. *Nat Genet* 2009; 41: 1238-1242.
20. Hussenet T, Dali S, Exinger J, Monga B, Jost B, Dembele D, Martinet N, Thibault C, Huelsken J, Brambilla E, du Manoir S. SOX2 Is an Oncogene Activated by Recurrent 3q26.3 Amplifications in Human Lung Squamous Cell Carcinomas. *PLoS One* 2010; 5: e8960.
21. Massion PP, Zou Y, Uner H, Kiatsimkul P, Wolf HJ, Baron AE, Byers T, Jonsson S, Lam S, Hirsch FR, Miller YE, Franklin WA, Varella-Garcia M. Recurrent genomic gains in preinvasive lesions as a biomarker of risk for lung cancer. *PLoS One* 2009; 4: e5611.
22. Que J, Luo X, Schwartz RJ, Hogan BL. Multiple roles for Sox2 in the developing and adult mouse trachea. *Development* 2009; 136: 1899-1907.
23. Gontan C, de Munck A, Vermeij M, Grosveld F, Tibboel D, Rottier R. Sox2 is important for two crucial processes in lung development: branching morphogenesis and epithelial cell differentiation. *Dev Biol* 2008; 317: 296-309.
24. Tompkins DH, Besnard V, Lange AW, Keiser AR, Wert SE, Bruno MD, Whitsett JA. Sox2 Activates Cell Proliferation and Differentiation in the Respiratory Epithelium. *Am J Respir Cell Mol Biol* 2010.

25. Lu Y, Futtner C, Rock JR, Xu X, Whitworth W, Hogan BL, Onaitis MW. Evidence that SOX2 overexpression is oncogenic in the lung. *PLoS One* 2010; 5: e11022.
26. Ramirez RD, Sheridan S, Girard L, Sato M, Kim Y, Pollack J, Peyton M, Zou Y, Kurie JM, Dimaio JM, Milchgrub S, Smith AL, Souza RF, Gilbey L, Zhang X, Gandia K, Vaughan MB, Wright WE, Gazdar AF, Shay JW, Minna JD. Immortalization of human bronchial epithelial cells in the absence of viral oncoproteins. *Cancer Res* 2004; 64: 9027-9034.
27. Correia L, Farah H, Rassl D, Rintoul R, Sethi T, Littlewood T, Evan G, McCaughan F. SOX2 initiates carcinogenesis in a novel organotypic model of bronchial dysplasia. British Thoracic Society. London; 2015.
28. Correia L, Littlewood T, Evan G, Mccaughan F. A rational organotypic model of human bronchial dysplasia. British Thoracic Oncology Group. Dublin; 2016.
29. Correia LL, Littlewood TD, Evan G, McCaughan F. Deregulated SOX2 drives dysplasia in a novel 3D organotypic model of bronchial dysplasia. AACR-IASCLC. San Diego; 2016.
30. Soldner F, Hockemeyer D, Beard C, Gao Q, Bell GW, Cook EG, Hargus G, Blak A, Cooper O, Mitalipova M, Isacson O, Jaenisch R. Parkinson's disease patient-derived induced pluripotent stem cells free of viral reprogramming factors. *Cell* 2009; 136: 964-977.
31. Szulc J, Wiznerowicz M, Sauvain MO, Trono D, Aebischer P. A versatile tool for conditional gene expression and knockdown. *Nat Methods* 2006; 3: 109-116.
32. Andrews S. FastQC. 2016 [cited 2016 11 October 2016]. Available from: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.
33. Kreuger F. Trim Galore! 2016 [cited 2016 11 October 2016]. Available from: http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/.
34. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 2013; 29: 15-21.
35. Anders S, Pyl PT, Huber W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* 2015; 31: 166-169.
36. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 2010; 26: 139-140.
37. Saeed AI, Sharov V, White J, Li J, Liang W, Bhagabati N, Braisted J, Klapa M, Currier T, Thiagarajan M, Sturn A, Snuffin M, Rezantsev A, Popov D, Ryltsov A, Kostukovich E, Borisovsky I, Liu Z, Vinsavich A, Trush V, Quackenbush J. TM4: a free, open-source system for microarray data management and analysis. *Biotechniques* 2003; 34: 374-378.
38. Delgado O, Kaisani AA, Spinola M, Xie XJ, Batten KG, Minna JD, Wright WE, Shay JW. Multipotent capacity of immortalized human bronchial epithelial cells. *PLoS One* 2011; 6: e22023.
39. Santini R, Pietrobono S, Pandolfi S, Montagnani V, D'Amico M, Penachioni JY, Vinci MC, Borgognoni L, Stecca B. SOX2 regulates self-renewal and tumorigenicity of human melanoma-initiating cells. *Oncogene* 2014; 33: 4697-4708.

40. Chen Y, Shi L, Zhang L, Li R, Liang J, Yu W, Sun L, Yang X, Wang Y, Zhang Y, Shang Y. The molecular mechanism governing the oncogenic potential of SOX2 in breast cancer. *J Biol Chem* 2008; 283: 17969-17978.
41. Hagey DW, Muhr J. Sox2 acts in a dose-dependent fashion to regulate proliferation of cortical progenitors. *Cell reports* 2014; 9: 1908-1920.
42. Fang X, Yoon JG, Li L, Yu W, Shao J, Hua D, Zheng S, Hood L, Goodlett DR, Foltz G, Lin B. The SOX2 response program in glioblastoma multiforme: an integrated ChIP-seq, expression microarray, and microRNA analysis. *BMC Genomics* 2011; 12: 11.
43. Sharov AA, Masui S, Sharova LV, Piao Y, Aiba K, Matoba R, Xin L, Niwa H, Ko MS. Identification of Pou5f1, Sox2, and Nanog downstream target genes with statistical confidence by applying a novel algorithm to time course microarray and genome-wide chromatin immunoprecipitation data. *BMC Genomics* 2008; 9: 269.
44. Watanabe H, Ma Q, Peng S, Adelmant G, Swain D, Song W, Fox C, Francis JM, Pedamallu CS, DeLuca DS, Brooks AN, Wang S, Que J, Rustgi AK, Wong KK, Ligon KL, Liu XS, Marto JA, Meyerson M, Bass AJ. SOX2 and p63 colocalize at genetic loci in squamous cell carcinomas. *J Clin Invest* 2014; 124: 1636-1645.
45. Malladi S, Macalinao DG, Jin X, He L, Basnet H, Zou Y, de Stanchina E, Massague J. Metastatic Latency and Immune Evasion through Autocrine Inhibition of WNT. *Cell* 2016; 165: 45-60.
46. Ma SR, Wang WM, Huang CF, Zhang WF, Sun ZJ. Anterior gradient protein 2 expression in high grade head and neck squamous cell carcinoma correlated with cancer stem cell and epithelial mesenchymal transition. *Oncotarget* 2015; 6: 8807-8821.
47. Shostak K, Zhang X, Hubert P, Goktuna SI, Jiang Z, Klevernic I, Hildebrand J, Roncarati P, Hennuy B, Ladang A, Somja J, Gothot A, Close P, Delvenne P, Chariot A. NF-kappaB-induced KIAA1199 promotes survival through EGFR signalling. *Nat Commun* 2014; 5: 5232.
48. Matsuda Y, Miura K, Yamane J, Shima H, Fujibuchi W, Ishida K, Fujishima F, Ohnuma S, Sasaki H, Nagao M, Tanaka N, Satoh K, Naitoh T, Unno M. SERPINI1 regulates epithelial-mesenchymal transition in an orthotopic implantation model of colorectal cancer. *Cancer Sci* 2016; 107: 619-628.
49. Singh A, Settleman J. EMT, cancer stem cells and drug resistance: an emerging axis of evil in the war on cancer. *Oncogene* 2010; 29: 4741-4751.
50. Gustafson AM, Soldi R, Anderlind C, Scholand MB, Qian J, Zhang X, Cooper K, Walker D, McWilliams A, Liu G, Szabo E, Brody J, Massion PP, Lenburg ME, Lam S, Bild AH, Spira A. Airway PI3K pathway activation is an early and reversible event in lung cancer development. *Sci Transl Med* 2010; 2: 26ra25.
51. Balsara BR, Pei J, Mitsuuchi Y, Page R, Klein-Szanto A, Wang H, Unger M, Testa JR. Frequent activation of AKT in non-small cell lung carcinomas and preneoplastic bronchial lesions. *Carcinogenesis* 2004; 25: 2053-2059.
52. Massion PP, Taflan PM, Shyr Y, Rahman SM, Yildiz P, Shakthour B, Edgerton ME, Ninan M, Andersen JJ, Gonzalez AL. Early involvement of the phosphatidylinositol 3-kinase/Akt pathway in lung cancer progression. *Am J Respir Crit Care Med* 2004; 170: 1088-1094.

53. Han W, Gills JJ, Memmott RM, Lam S, Dennis PA. The chemopreventive agent myoinositol inhibits Akt and extracellular signal-regulated kinase in bronchial lesions from heavy smokers. *Cancer Prev Res (Phila)* 2009; 2: 370-376.
54. Targeting Mutant AKT in Cancer. *Cancer Discov* 2016; 6: OF7.
55. Sato M, Larsen JE, Lee W, Sun H, Shames DS, Dalvi MP, Ramirez RD, Tang H, DiMaio JM, Gao B, Xie Y, Wistuba, II, Gazdar AF, Shay JW, Minna JD. Human lung epithelial cells progressed to malignancy through specific oncogenic manipulations. *Mol Cancer Res* 2013; 11: 638-650.
56. Stewart CE, Torr EE, Mohd Jamili NH, Bosquillon C, Sayers I. Evaluation of differentiated human bronchial epithelial cell culture systems for asthma research. *J Allergy (Cairo)* 2012; 2012: 943982.
57. Chung GT, Sundaresan V, Hasleton P, Rudd R, Taylor R, Rabbitts PH. Sequential molecular genetic changes in lung cancer development. *Oncogene* 1995; 11: 2591-2598.
58. Wistuba, II, Behrens C, Milchgrub S, Bryant D, Hung J, Minna JD, Gazdar AF. Sequential molecular abnormalities are involved in the multistage development of squamous cell lung carcinoma. *Oncogene* 1999; 18: 643-650.
59. Shay JW, Wright WE. Role of telomeres and telomerase in cancer. *Semin Cancer Biol* 2011; 21: 349-353.
60. Lantuejoul S, Raynaud C, Salameire D, Gazzeri S, Moro-Sibilot D, Soria JC, Brambilla C, Brambilla E. Telomere maintenance and DNA damage responses during lung carcinogenesis. *Clin Cancer Res* 2010; 16: 2979-2988.
61. Lee J, Ryu SH, Kang SM, Chung WC, Gold KA, Kim ES, Hittelman WN, Ki Hong W, Koo JS. Prevention of bronchial hyperplasia by EGFR pathway inhibitors in an organotypic culture model. *Cancer Prev Res (Phila)* 2011; 4: 1306-1315.