SOX2 gene amplification and protein overexpression are associated with better outcome in squamous cell lung cancer

Theresia Wilbertz[1], Patrick Wagner[2,3], Karen Petersen[1], Ann-Cathrin Stiedl[1], Veit J Scheble[1], Sebastian Maier[1], Markus Reischl[4], Ralf Mikut[4], Nasser K Altorki[2], Holger Moch[5], Falko Fend[1], Annette Staebler[1], Adam J Bass[6,7,8], Matthew Meyerson[6,7,8,9], Mark A Rubin[3], Alex Soltermann[5], Claudia Lengerke[10] and Sven Perner[1]

[1]Institute of Pathology, Comprehensive Cancer Center, University Hospital Tuebingen, Tuebingen, Germany; [2]Cardiothoracic Surgery, Weill Cornell Medical Center, New York, NY, USA; [3]Department of Pathology and Laboratory Medicine, Weill Cornell Medical Center, New York, NY, USA; [4]Institute for Applied Computer Science, Karlsruhe Institute of Technology, Karlsruhe, Germany; [5]Institute of Surgical Pathology, University Hospital Zurich, Zurich, Switzerland; [6]Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA, USA; [7]Broad Institute, Cambridge, MA, USA; [8]Center for Cancer Genome Discovery, Dana-Farber Cancer Institute, Boston, MA, USA; [9]Department of Pathology, Harvard Medical School, Boston, MA, USA and [10]Division of Hematology and Oncology, University of Tuebingen Medical Center II, Tuebingen, Germany

The transcription factor SOX2 (3q26.3–q27) is a key regulator of foregut development and an embryonic stem cell factor cooperating during induction of pluripotency in terminally differentiated somatic cells. Recently, we found SOX2 to be amplified in a subset of squamous cell lung and esophageal cancers. The aim of this study was to explore the prognostic role of SOX2 in a large series of squamous cell carcinomas and adenocarcinomas of the lung. A total of 891 samples from two independent population-based cohorts were assessed by fluorescence in situ hybridization and immunohistochemistry. Furthermore, we assessed for associations between SOX2 amplification/upregulation and clinicopathological features. Similar results were found in the two cohorts. Within squamous cell carcinoma cases, 8% high-level as well as 68 and 65% low-level SOX2 amplifications occurred in the two cohorts, respectively. In adenocarcinomas, no high-level amplification was found and low-level amplification occurred in 6% of the two cohorts. Within squamous cell carcinomas of one cohort, SOX2 amplification was associated with lower tumor grade, while higher levels of SOX2 expression were related to younger age, smaller tumor size, and lower probability of angiolymphatic invasion and metastasis. High SOX2 expression levels proved to be a marker for prolonged overall survival among patients with squamous cell carcinomas. In conclusion, SOX2 amplification and upregulation are frequent events in squamous cell carcinomas of the lung and are associated with indicators of favorable prognosis.
Modern Pathology (2011) 24, 944–953; doi:10.1038/modpathol.2011.49; published online 1 April 2011

Keywords: amplification; expression; lung cancer; SOX2; squamous cell carcinoma

Based on current knowledge, several cancers arise and progress through an accumulation of complex genetic and molecular alterations. Classical genetic events, including increases in gene copy number, are frequently found in human cancer genomes and have proven to be important as diagnostic markers as well as potential targets for directed anticancer therapies.[1,2] Such molecular findings enable a better understanding of tumor biology and may contribute to the identification of distinct clinically relevant tumor subclasses.

Several oncogenes have been proven to be frequently amplified in lung cancer. For example,

in lung adenocarcinomas, *NK2 homeobox 1* (*NKX2-1*) (also known as *thyroid transcription factor 1* (*TTF1* or *TITF1*)) is amplified as a lineage-specific onco-gene in a significant subset of tumors[3–5] and recent data suggest that NKX2-1 upregulation is associated with better outcome in patients with lung adeno-carcinomas.[6,7] Lung squamous cell carcinomas are known to commonly harbor an amplification of the genomic region 3q.[8] Recently, we and others identified the transcription factor *SOX2* (SRY (sex determining region Y)-box 2) as the most promising candidate disease-related gene within this amplicon.[9,10]

SOX2 is a member of the SRY-related HMG-box (SOX) family of transcription factors[11] and has been recently studied for its ability to reprogram adult cells into induced pluripotent stem cells when co-expressed with other embryonic stem cell factors like Nanog and Oct3/4.[12] SOX2 is expressed during early embryogenesis and has an important role in embryonic and extra-embryonic cell types.[13] In detail, SOX2 is believed to inhibit neurogenesis and to preserve cellular proliferative potential during development of the central nervous system.[14] Furthermore, SOX2 is a key regulator of foregut development. High SOX2 expression is found in all endoderm cells of the undivided foregut with protein levels being higher in the dorsal foregut (the future esophagus) and lower in the ventral foregut (the future trachea).[15,16] During bronchogenesis in the developing lung, SOX2 is precisely regulated and forced overexpression of SOX2 leads to a block of airway branching.[15] In the adult, SOX2 expression is found in the basal cell layer of the esophagus and in bronchial epithelial cells of the lung, whereas it is absent in alveolar cells.[17]

In lung cancer, *SOX2* amplification and expression can be found in a large subset of squamous cell carcinomas.[9,10,17,18] Though at much lower frequencies, expression of SOX2 has also been reported to occur in adenocarcinomas of the lung.[17–19] *In vitro* studies suggest that SOX2 upregulation induces cell proliferation and anchorage-independent growth in squamous cell lung cancer cell lines.[9,10] Furthermore, *SOX2* amplification and expression have been detected in preinvasive lesions of the lung, such as squamous dysplasia and carcinoma *in situ*.[17,20] SOX2 may have similar oncogenic effects in other organs of foregut and non-foregut origin, since amplification and expression also occur in squa-mous cell carcinomas of the oral cavity, as well as in esophageal and gastric tumors.[9,21–23] Recently, SOX2 expression has been detected in early-stage breast carcinoma.[24] These data suggest that expression of SOX2 drives oncogenesis and therefore overexpres-sion in tumors may be associated with a more aggressive tumor phenotype and poorer clinical outcome. However, in lung cancers, contradictory results have been reported so far: while one smaller study suggested that higher SOX2 expression is associated with better outcome in patients suffering

from lung squamous cell carcinomas,[25,26] data from a third group present SOX2 expression to be a marker of poor prognosis in stage I lung adenocarcinomas.[19]

Together, these findings support the notion that SOX2 upregulation has a major impact on lung cancer initiation and progression. The prognostic value of SOX2 expression levels may vary depend-ing on histologic tumor type and additional studies performed on larger cohorts of patients are needed to consistently define these effects. In the present study, we assessed the prognostic impact of *SOX2* amplification and expression levels in lung adeno-carcinomas and squamous cell carcinomas by analyzing two large independent cohorts for their *SOX2* status and performing correlations with relevant clinicopathological data.

Materials and methods

Cohort Characterization and Clinicopathological Data Collection

A total of 891 patients with lung squamous cell carcinomas and adenocarcinomas from two inde-pendent cohorts (Zurich, Switzerland and New York, NY, USA) were assessed by fluorescence *in situ* hybridization (FISH) and immunohistochem-istry to study *SOX2* gene amplification and SOX2 protein expression status. The Zurich cohort (*n*=512) is a population-based series of 250 adeno-carcinomas and 262 squamous cell carcinomas. Patients from this cohort underwent surgery with curative intent between January 1993 and December 2002. Out of these 512 patients, 472 were treatment-naive and 40 received neoadjuvant radiation or chemotherapy. Surgical lung specimens were pro-cessed according to the guidelines of the Swiss Society of Pathology. Overall survival time was obtained from the Cancer Registry of the Kanton Zurich and was defined as the interval between the date of surgery and the date of death. The end date of follow-up was 31 July 2007, and the median duration of follow-up was 45 (range 1–169) months.

The New York cohort represents patients treated in a single institution (*n* = 379) and comprises 325 adeno-carcinomas and 54 squamous cell carcinomas. Patients underwent resection between May 1992 and October 2007. Clinicopathological data, including follow-up data, were obtained via review of the patients' pathology reports and electronic medical record. A prospective database of patient outcome was used to obtain overall survival time. The mean duration of follow-up was 32 (range 0–152) months, with the end date of follow-up being 10 December 2009.

Tissue Microarray Construction

Tissue microarray recipient blocks were constructed using a custom-made, semiautomatic tissue arrayer (Beecher Instruments, Sun Prairie, WI, USA), as

previously described. Two to three representative 0.6 mm cores of viable tissue from each tumor were included.[6,27] Additional control cores of normal lung tissue were incorporated. The recipient blocks were cut into 4 μm thick sections, which were then mounted onto superfrost slides.

SOX2 Amplification FISH Assay

To assess for *SOX2* amplification status at the chromosomal level on tissue microarray slides, we applied the same two-color interphase FISH assay as described by Bass *et al*.[9] Briefly, a probe spanning the locus 3q26.33 (BAC clone CTD-2348H10) was applied to detect *SOX2* copy number status and was compared with a reference probe hybridizing to 3p22.3–3p22.2 (BAC clone RP11-286G5). Both clones were purchased from Invitrogen, Carlsbad, CA, USA. The target probe was labeled with biotin and detected with a streptavidin-conjugated red fluorochrome (SAV-Alexa Fluor-594, Invitrogen). The reference probe was labeled with digoxigenin and detected via an anti-digoxigenin-conjugated green fluorochrome (FITC, Roche, Basel, Switzerland).

Assessment of the *SOX2* amplification status was performed semiquantitatively by comparing the number of red signals (*SOX2* target region) to the number of corresponding green signals (reference region).

A non-amplified nucleus showed one red target signal for every corresponding green reference signal, with a red/green ratio of 1:1 (Figure 1a). In most cases, cancer cells were present in a diploid state (two red signals and two green signals), with rare cases revealing triploidy or tetraploidy (three red and three green signals or four red and four green signals, respectively). Cores with additional 2–9 red target signals exceeding the number of green signals in at least 30% of tumor cells were defined as low-level amplification (Figure 1b). Cells with

additional ≥ 10 red signals displaying a cluster-like formation were defined as high-level amplification (Figure 1c). All cutoff values were defined *a priori* of the assessment and were the same as applied in our previous study.[9]

All tissue microarray slides were analyzed under a 63 × oil immersion objective using a fluorescence microscope (Zeiss, Jena, Germany) equipped with appropriate filters, a charge-coupled device camera and the FISH imaging and capturing software Metafer 4 (Metasystems, Altlussheim, Germany). FISH evaluation was independently performed by two experienced observers (TW and AS). At least 100 nuclei per case were assessed. Cases were included into the analysis if there was at least one core assessable. In all, 42 adenocarcinomas and 52 squamous cell carcinomas from the Zurich cohort and 34 adenocarcinomas and 5 squamous cell carcinomas from the New York cohort were not assessable for their *SOX2* amplification status due to insufficient target tissue or hybridization. In the rare case of heterogeneity of *SOX2* amplification status within a single core, only the predominant region was taken into account.

SOX2 Immunohistochemistry and Evaluation

Immunohistochemistry was conducted with the Ventana Discovery automated immunostaining system (Ventana Medical Systems, Tucson, AZ, USA), using Ventana reagents. Sections mounted on superfrost slides were deparaffinized in inorganic buffer, then pretreated at pH 8.4 in an EDTA-based buffer. A primary antibody (polyclonal goat anti-human SOX2 antibody, AF2018, R&D systems, dilution 1:40, heat-induced epitope retrieval) was applied to probe SOX2 protein. Dilution was performed with a Ventana diluent. The antibody was detected using a biotinylated detection kit based on diaminobenzidine and horseradish peroxidase (DABMap-kit, Ventana). Slides were counterstained with
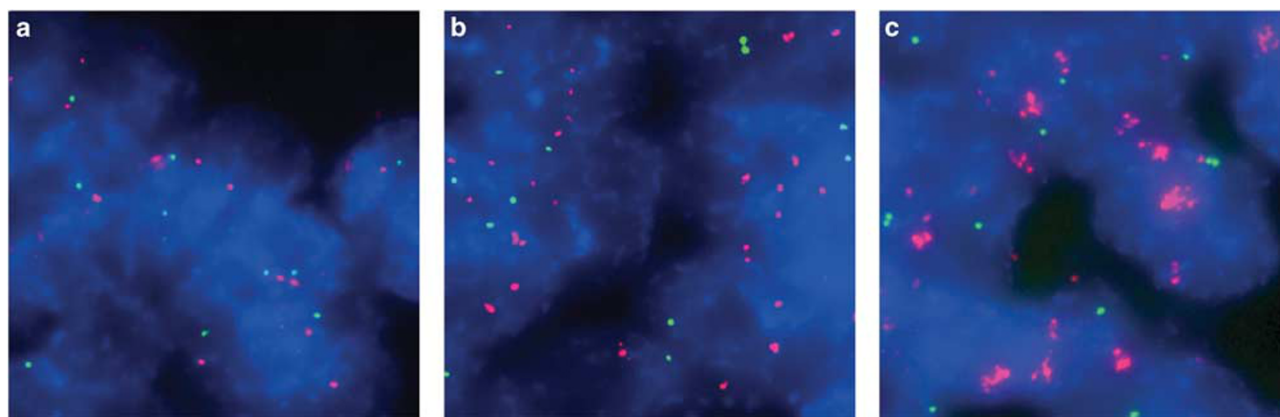

Figure 1 *SOX2* amplification in lung squamous cell carcinomas assessed by fluorescence *in situ* hybridization. (**a**) Nucleus without *SOX2* amplification as found in normal alveolar cells and non-amplified tumor cells. (**b**) Nucleus with low-level *SOX2* amplification. (**c**) Nucleus with cluster-like high-level *SOX2* amplification.

SOX2 in lung cancer
T Wilbertz *et al*
947

hematoxylin and Blueing Reagent (Ventana). Subsequently, sections were washed, dehydrated in an ascending alcohol series, and covered using Cytoseal.

Stained tissue microarray slides were digitalized using the Zeiss MIRAX DESK scanner. A semiautomated quantitative image analysis software (Definiens Architect XD 1.2, Definiens, Munich, Germany) was applied to obtain a continuous spectrum of average nuclear brown staining intensity in arbitrary units (maximum range of readout 0–2) for the tumor areas of each core. Cases were included into the analysis if at least one core was assessable. Because of insufficient target tissue or artefact, 10 adenocarcinomas and 7 squamous cell carcinomas from the Zurich cohort and 10 adenocarcinomas and 1 squamous cell carcinoma from the New York cohort were not assessable for their SOX2 expression status.

Statistics

For statistical analysis of the Zurich cohort, only treatment-naive patients were included. Nominal variables (gender, smoking status, angiolymphatic invasion, pleural infiltration) were compared among subgroups using χ^2 or Fisher's exact test, as appropriate. Statistical assessment of two ordinal variables (TNM status, stage, grade), one ordinal and one continuous variable or comparisons between two continuous variables (age, packyears, tumor size, expression level) were performed using Kendall's τ-b correlation test. Comparisons between dichotomous and continuous variables were accomplished using Mann–Whitney U-test.

For survival analysis, patients that had received neoadjuvant therapy and all patients with overall survival <1 month post-surgery were excluded. Overall survival curves were generated using the Kaplan–Meier method and compared using a log-rank test. A multivariate Cox proportional hazards regression model of independent prognostic factors for overall survival was performed. For all tests, statistical significance was defined as $P < 0.05$. Statistics was performed using SPSS software (SPSS, Chicago, IL, USA).

This study was approved by the institutional ethical review boards of the University Hospital Zurich, the New York-Presbyterian Hospital/Weill Cornell Medical Center, and the University Hospital Tuebingen.

Results

SOX2 Gene Amplification and Protein Expression Status

We found high-level SOX2 amplification (Figure 1c) in 8% of Zurich (16/210) and New York (4/49) squamous cell carcinomas, whereas no adenocarcinomas from either cohort exhibited this alteration. Low-level amplification (Figure 1b) was found in the majority of squamous cell carcinomas: Zurich 68% (143/210), New York 65% (32/49), respectively. In contrast, only 6% of Zurich (13/208) and New York (16/291) adenocarcinomas harbored a low-level amplification of SOX2. In all, 24% (51/210) of Zurich squamous cell carcinomas and 27% (13/49) of New York squamous cell carcinomas did not show any alteration concerning SOX2 copy number status, whereas the majority of adenocarcinoma samples (94% (195/208) from Zurich and 95% (275/291) from New York) were negative for SOX2 copy number alteration (Figure 2). Similarly, normal lung tissue did not exhibit amplification of SOX2. Taken together, SOX2 amplification was much more frequently detected in squamous cell carcinomas than adenocarcinomas and comparable results were attained in the two independent cohorts.
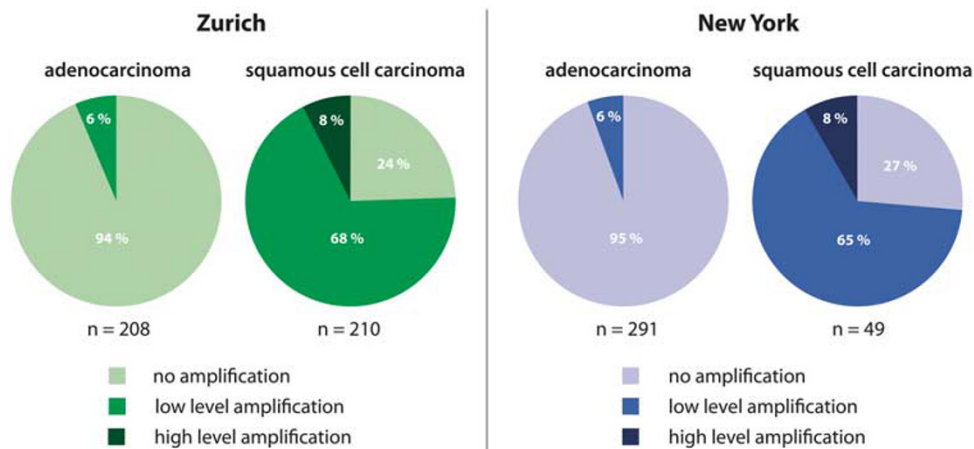

Figure 2 Distribution of SOX2 amplification status. The two independent cohorts reveal an almost identical amplification pattern. High-level SOX2 amplification was seen only in squamous cell carcinomas (Zurich: 8% (16/210), New York: 8% (4/49)). Low-level amplification was seen in the majority of squamous cell carcinomas (Zurich: 68% (143/210), New York: 65% (32/49)), and in a small minority of adenocarcinomas (Zurich: 6% (13/208), New York: 6% (16/291)) (difference to 100% due to rounding).
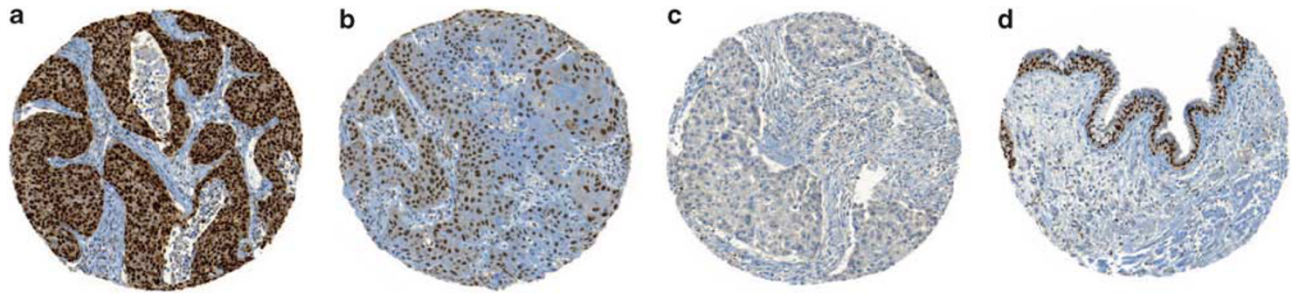
Figure 3 SOX2 expression in lung squamous cell carcinomas assessed by immunohistochemistry. (**a**) SOX2 expression in a representative *SOX2*-amplified lung squamous cell carcinoma: strong nuclear expression in the majority of tumor cells. (**b**) SOX2 expression as found in both amplified and some non-amplified lung squamous cell carcinomas with homogenous moderate nuclear expression within the tumor cell population. (**c**) SOX2 expression in a non-amplified lung squamous cell carcinoma: The majority of tumor cells do not display nuclear SOX2 expression. (**d**) SOX2 protein expression in normal bronchus epithelium. The majority of normal bronchus epithelial cells show moderate nuclear SOX2 expression.
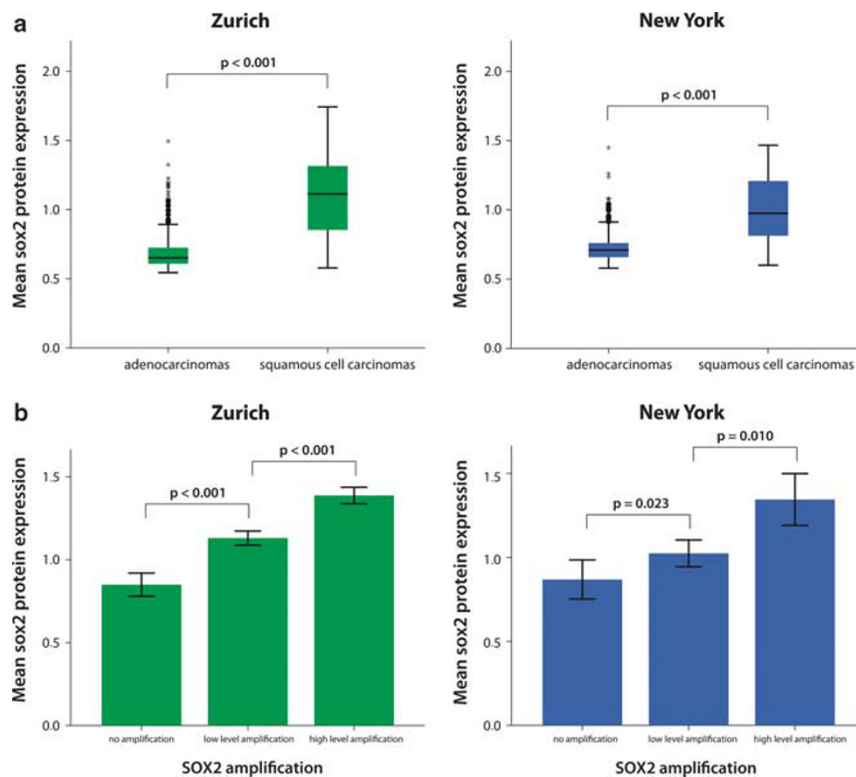

Figure 4 (**a**) SOX2 expression by immunohistochemistry. SOX2 expression is significantly higher in lung squamous cell carcinoma tissue than in adenocarcinomas of the lung (Mann–Whitney *U*-test). Expression levels are consistently low in adenocarcinoma specimens, whereas squamous cell carcinoma samples exhibit a wider distribution of values. (**b**) Bar graphs displaying mean SOX2 expression stratified by *SOX2* amplification status. *SOX2* amplification is associated with significantly increased SOX2 expression in lung squamous cell carcinomas (Mann–Whitney *U*-test).

Concerning SOX2 protein expression, similar expression patterns were found in the two cohorts: a broad range of SOX2 expression could be detected in squamous cell carcinoma samples of both cohorts with mean brown staining values ranging from 0.58 to 1.74 (arbitrary units). Adenocarcinoma specimens showed much less SOX2 expression, with 95% of cases exhibiting mean values between 0.54 and 0.94, and only 22 of 555 cases showing expression values higher than 1. Consistent with the data presented above on *SOX2* amplification status, the overall mean SOX2 expression was significantly higher in squamous cell carcinomas than in adenocarcinomas in both cohorts ($P < 0.001$) (Figures 3 and 4a). Squamous cell carcinoma cases with higher *SOX2* amplification status exhibited significantly higher levels of SOX2 protein expression (Figure 4b). Interestingly, this relationship was not significant in adenocarcinoma samples of either cohort (data not shown).

Correlation of *SOX2* Amplification/Expression Status with Clinicopathological Data

In order to assess the prognostic value of *SOX2*-amplified and SOX2-overexpressing lung squamous cell carcinomas and adenocarcinomas, we correlated relevant clinicopathological patient data with *SOX2* amplification and expression status.

In the Zurich squamous cell carcinoma cohort, tumors harboring higher levels of *SOX2* amplification were significantly associated with lower tumor grade ($P = 0.014$). As for SOX2 expression, smaller tumors and tumors with lower pT stage (according to TNM classification) exhibited significantly higher levels of SOX2 protein expression ($P = 0.001$, $P = 0.037$). Increased SOX2 expression was associated with lower UICC/WHO stage ($P = 0.028$) and with cases lacking angiolymphatic invasion and metastasis ($P = 0.001$, $P = 0.034$). Interestingly, younger age at diagnosis was associated with elevated SOX2 expression ($P = 0.003$) (Table 1).

As the New York cohort comprised only 49 valid squamous cell carcinoma samples, only four of which showed high-level *SOX2* amplification, this cohort was underpowered to perform formal statistical analysis. However, all four high-level amplified cases were staged pT1 and pT2, had a lower mean tumor size compared with cases with low-level or no *SOX2* amplification, were either moderately or well differentiated and proved to be negative for angiolymphatic invasion and metastasis.

As for adenocarcinomas, we found low-level *SOX2* amplification to be associated with lower probability of distant metastasis ($P = 0.003$) in the Zurich cohort. Higher levels of SOX2 expression proved to be associated with higher tumor grade. In the New York cohort low-level *SOX2* amplification was associated with angiolymphatic invasion ($P = 0.017$) and less differentiated tumors ($P = 0.011$). In addition, increased levels of SOX2 expression were associated with higher probability of metastasis ($P = 0.022$) and more smoking pack-years ($P = 0.027$).

Survival Analysis

To determine whether *SOX2* amplification or SOX2 expression is a prognostic factor in squamous cell carcinomas, we performed a Kaplan–Meier survival analysis. The different levels of *SOX2* amplification and expression (with the expression scale being divided into quartiles) were compared using a log-rank test.

Comparing high-level SOX2 expression (ie expression levels within the highest quartile) with lower levels of SOX2 expression (1.–3. quartile), high-level SOX2 expression was significantly associated with better survival ($P = 0.036$) (Figure 5b).

Among squamous cell carcinoma samples of the Zurich cohort, high-level amplified cases showed an association with better outcome as compared with all other cases, although this finding did not reach statistical significance ($P = 0.183$) (Figure 5a). When patients with squamous cell carcinomas from the New York cohort were included, a strong trend toward better survival was seen in patients harboring high-level *SOX2* amplification ($P = 0.078$). There was no significant difference in survival between low-level amplified cases and cases without *SOX2* copy number alteration.

In a Cox regression model considering all parameters with significant impact on patient survival (pT, pN, pM, stage, grade, angiolymphatic invasion, pleural infiltration), neither *SOX2* amplification status nor SOX2 expression level proved to be an independent predictor for survival in patients with squamous cell carcinomas.

Among adenocarcinomas, SOX2 expression level was not associated with patient survival in either cohort. However, even in spite of limited case numbers, low-level amplification proved to be a highly significant predictor for worse patient survival in the New York cohort ($P = 0.009$).

Discussion

SOX2 is a member of the SOX (SRY-related HMG-box) family of transcription factors.[11] Besides its role during neurogenesis and embryonic foregut development, SOX2 can aid the reprogramming of adult cells into induced pluripotent stem cells, indicating a major potency for induction of cellular fate changes.[12,14,28] More recently, studies have identified *SOX2* as the driving gene of the well-known 3q amplicon in squamous cell carcinomas of the lung and the esophagus, suggesting *SOX2* as a novel candidate oncogene[9,10] in lung carcinogenesis.

In the current study, we investigated the prognostic significance of *SOX2* amplification and overexpression in squamous cell carcinomas and adenocarcinomas of the lung.

SOX2 in Lung Squamous Cell Carcinomas

Recently, we and others have characterized the *SOX2* amplification status in lung squamous cell carcinomas with reported frequencies varying from 20 to 60%.[9,10,17] These variations in frequency are most likely due to methodological discrepancies applied by different laboratories (eg the molecular method applied for detection of *SOX2* copy number status and the threshold set to distinguish between different levels of amplification). Furthermore, differences between the cohorts and tumor heterogeneity have to be considered.

In our current study, we assessed two large independent cohorts from different continents (Zurich, Switzerland and New York, NY, USA) in a semiquantitative manner and found 67% low-level and 8% high-level amplifications with notably uniform distribution in the two cohorts. To control

Table 1 *SOX2* amplification and expression status and clinicopathological parameters for patients with lung squamous cell carcinomas (Zurich cohort)

Squamous cell cancer	Total (Σ = 196)	SOX2 amplification status				Total (Σ = 239)	SOX2 expression status	
		None (Σ = 45)	Low level (Σ = 137)	High level (Σ = 14)	P		Mean expression	P
Gender					0.457[a]			0.354[b]
Male	159	39	108	12		192	1.09 ± 0.28	
Female	37	6	29	2		47	1.06 ± 0.27	
Age	65.65 ± 9.44	66.31 ± 11.72	65.79 ± 8.19	62.21 ± 12.54	0.115[c]	239	—	0.003[c]
Smoking status					0.408[a]			0.456[b]
Non-smoker	7	3	4	0		8	1.01 ± 0.29	
Smoker	154	32	110	12		186	1.09 ± 0.28	
Missing	35					45		
Packyears	55.61 ± 30.02	53.13 ± 32.32	58.42 ± 29.31	36.50 ± 24.17	0.700[c]	186	—	0.721[c]
Tumor size (cm)	4.21 ± 2.13	4.47 ± 1.91	4.11 ± 2.20	4.33 ± 2.16	0.242[c]	239	—	0.001[c]
pT					0.220[c]			0.037[c]
pT1	38	7	29	2		45	1.19 ± 0.26	
pT2	104	23	71	10		131	1.06 ± 0.27	
pT3	39	11	26	2		45	1.12 ± 0.32	
pT4	15	4	11	0		18	0.93 ± 0.22	
pN					0.200[c]			0.804[c]
pN0	98	23	71	4		122	1.09 ± 0.29	
pN1	72	18	49	5		87	1.07 ± 0.26	
pN2	21	3	14	4		25	1.14 ± 0.29	
pN3	5	1	3	1		5	0.85 ± 0.39	
pM					0.153[c]			0.034[c]
pM0	184	39	132	13		226	1.10 ± 0.28	
pM1	12	6	5	1		13	0.87 ± 0.31	
Stage (UICC/WHO)					0.318[c]			0.028[c]
IA	26	5	19	2		31	1.21 ± 0.26	
IB	45	8	36	1		60	1.04 ± 0.27	
IIA	10	1	9	0		12	1.18 ± 0.25	
IIB	55	14	36	5		67	1.12 ± 0.26	
IIIA	36	7	24	5		41	1.12 ± 0.31	
IIIB	14	4	10	0		17	0.93 ± 0.23	
IV	10	6	3	1		11	0.77 ± 0.22	
Grade					0.014[c]			0.128[c]
Grade 1	3	1	2	0		3	1.02 ± 0.30	
Grade 2	107	18	78	11		132	1.12 ± 0.26	
Grade 3	86	26	57	3		104	1.04 ± 0.30	
Angiolymphatic invasion					0.592[a]			0.001[b]
Negative	112	23	80	9		136	1.14 ± 0.28	
Positive	84	22	57	5		103	1.01 ± 0.27	
Pleural infiltration					0.445[a]			0.534[b]
Negative	129	27	94	8		159	1.09 ± 0.28	
Positive	67	18	43	6		80	1.07 ± 0.28	

[a]χ^2 test (or Fisher's exact test if expected frequency was < 5 in > 20%).

[b]Mann–Whitney *U*-test.

[c]Kendall's τ-b correlation.

the problem of tumor heterogeneity in our study, we assessed the *SOX2* copy number status in samples from at least two physically distant parts of the primary tumor.

A FISH probe precisely covering the *SOX2* locus and encompassing only the genes *SOX2* and *SOX2OT* (*SOX2 overlapping transcript*) was applied to specifically focus on *SOX2* within the 3q amplicon. However, the amplification locus of 3q comprises additional genes, such as *DCUN1D1* and *PIK3CA* that previously have been proposed to be target oncogenes of the 3q amplicon.[8] Recent studies have provided strong evidence that *SOX2* is the primary amplification target within the common 3q amplicon and functional studies using shRNA against SOX2 showed an impact on tumor biology, thus making *SOX2* the most promising candidate 3q oncogene.[9,10,28]

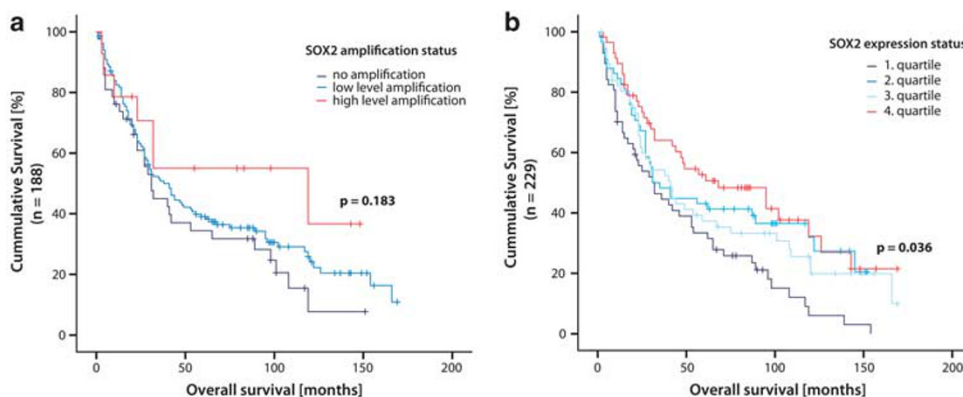We found *SOX2* amplification to be associated with higher SOX2 expression in lung squamous cell

Figure 5 Kaplan–Meier estimates for different *SOX2* amplification and expression levels. (**a**) Cases with high-level *SOX2* amplification show a trend toward longer overall survival in patients with squamous cell carcinomas (median survival 119 *vs* 36 months, *P* = 0.183). (**b**) Higher SOX2 expression levels significantly correlate with longer overall survival (median survival 68 *vs* 35 months, *P* = 0.036).

carcinomas, likely reflecting a gene dosage effect. While this association was highly significant, there were a few cases with high SOX2 expression levels despite the absence of *SOX2* amplification, indicating that in some instances mechanisms other than *SOX2* copy number may explain SOX2 protein overexpression. However, based on our results it would appear that *SOX2* amplification is the most important driver of SOX2 expression in lung squamous cell carcinomas.

It has been proposed that embryonic stem cell genes such as *SOX2* may mark the tumor stem cell population within cancers (eg breast cancer[29,30]). Such tumor stem cells are believed to persist in tumors as a distinct subpopulation and to cause relapse and metastasis by giving rise to new tumors.[31] Notably, we found *SOX2* amplification and expression to be homogenous in virtually all cases, both within one region as well as between the different regions of a specific case.

Role of *SOX2* in Carcinogenesis of Lung Squamous Cell Carcinoma

In univariate statistical analyses, our data clearly demonstrate an association between elevated SOX2 expression and indicators of better patient outcome, most importantly prolonged overall survival. Furthermore, increased levels of *SOX2* amplification indicate a better histological differentiation grade and a trend to improved patient survival. These results support our previous findings, indicating that patients with lung squamous cell carcinomas harboring an embryonic stem cell-like gene expression signature, including SOX2 expression, exhibit significant better survival than patients with tumors lacking this signature.[9] Consistently, Lu *et al*[25] were able to show in a cohort of early stage lung squamous cell carcinomas that patients with SOX2 expression above the median show prolonged overall survival.

The underlying molecular mechanisms accounting for SOX2 being associated with favorable prognosis in lung squamous cell carcinomas still remain to be elucidated. Consistent with the theory of lineage survival oncogenes,[5] SOX2 overexpression might recapitulate transcription networks active in normal squamous precursor cells and thus counteract the chaos of malignant dedifferentiation. Alternatively, SOX2 overexpression might occur early during lung squamous cell carcinoma carcinogenesis and might be lost during disease progression, due to genetic inactivation. Further studies will be needed to clarify these functional aspects of SOX2.

The discovery of *SOX2* as the driving gene of the common 3q amplicon suggests *SOX2* to be a potential and frequently occurring oncogene in squamous cell carcinomas. Functional studies underscore the oncogenic role of *SOX2*. In squamous cell carcinoma cell lines harboring *SOX2* amplification, suppression of SOX2 had an anti-proliferative effect.[9] Furthermore, cell lines overexpressing SOX2 exhibited increased migratory activity and enhanced colony formation.[10] In preinvasive lesions of the lung, SOX2 expression has been reported to occur in normal bronchial epithelium, alveolar bronchiolization, squamous dysplasia, as well as carcinoma *in situ*.[17] Furthermore, McCaughan *et al*[20] reported *SOX2* amplification in none of a series of low-grade bronchial lesions, but in all high-grade lesions, suggesting upregulation during preinvasive disease progression. Consistently, Lu *et al*[25] showed that conditional homozygous SOX2 overexpression in clara cells induces bronchial epithelial hyperplasia with 50% of cases showing a progression to lung cancer in mice.

Taken together, these results strongly indicate that *SOX2* might be oncogenic during lung carcinogenesis. However, our results indicate a strong association between SOX2 upregulation and clinicohistopathological parameters indicative of better patient outcome, including lower tumor grade.

Based on these data, we propose a model in which *SOX2* harbors oncogenic potential and has a role during tumorigenesis. At the same time, we believe that those tumors arising from an upregulation of SOX2 exhibit a clear squamous cell differentiation and thus are associated with better prognostic features, similar to *NKX2-1* in lung adenocarcinomas.[3,4] In conclusion, we propose that *SOX2* most likely is a lineage survival oncogene in lung squamous cell carcinomas.

SOX2 in Squamous Cell Carcinomas from Other Organ Sites

In agreement with previous reports[9,10,17] the findings of our study suggest that among lung cancers, *SOX2* amplification and expression is highly specific for squamous cell histology. In support of this, SOX2 proved to be significantly co-expressed with typical markers of squamous cell carcinoma differentiation, such as p63.[9,15,25] Besides the lung, *SOX2* has been found to be amplified and expressed in squamous cell carcinomas originating from other organ sites, predominantly derived from the embryonic foregut.[9,21,22] Recently, we found *SOX2* to be amplified and expressed in squamous cell carcinomas originating from non-foregut tissues, such as the skin, the cervix, and the penis,[23] indicating that *SOX2* might be a general marker for squamous cell carcinoma differentiation regardless the tissue of origin. Squamous carcinogenesis from diverse body sites may thus share similar underlying mechanisms.

SOX2 in Lung Adenocarcinomas

Previous findings revealed that SOX2 is expressed in bronchial epithelial cells of the lung, whereas it is absent in alveolar cells.[17,22] Likewise, adenocarcinoma precursor lesions, such as atypical adenomatous hyperplasia, proved to be negative for SOX2 expression.[17] In concordance, we could not detect any high-level *SOX2* amplification in lung adenocarcinomas and only low-level *SOX2* amplification in about 6% of cases. Protein expression levels were significantly lower in lung adenocarcinomas compared with squamous cell carcinomas. Neither the correlation analysis with clinicopathological patient data, nor the survival analysis comparing different levels of SOX2 expression revealed any consistent results to predict patient outcome in lung adenocarcinomas. Notably, we found low-level amplification of *SOX2* to be associated with shorter patient survival in the New York cohort, but despite otherwise homogenous results in the two cohorts we were not able to confirm this particular finding in the Zurich cohort. Interestingly, Sholl *et al*[19] found SOX2 expression to be a prognostic factor of worse patient outcome in stage I lung adenocarcinomas.

Taken together, these results might suggest a prognostic role for *SOX2* in lung adenocarcinomas.

However, given the overall low frequency of *SOX2* amplification and overexpression, the significance of this finding needs further evaluation. Most importantly, the finding of *SOX2* amplification/upregulation being frequent in lung squamous cell carcinomas, but rare in lung adenocarcinomas might reflect a fundamental molecular difference in carcinogenesis between these tumor entities.

Conclusion

Amplification and overexpression of *SOX2* are common in lung squamous cell carcinomas, but not adenocarcinomas, of the lung. *SOX2* amplification is associated with higher SOX2 expression levels, suggesting a gene dosage effect. Both *SOX2* amplification and expression are associated with favorable clinicopathological features in squamous cell carcinomas, including longer overall survival.

Acknowledgement

Disclosure/conflict of interest

The authors declare no conflict of interest.

References

1 Frohling S, Dohner H. Chromosomal abnormalities in cancer. N Engl J Med 2008;359:722–734.

2 Albertson DG. Gene amplification in cancer. Trends Genet 2006;22:447–455.

3 Weir BA, Woo MS, Getz G, *et al.* Characterizing the cancer genome in lung adenocarcinoma. Nature 2007;450:893–898.

4 Kwei KA, Kim YH, Girard L, *et al.* Genomic profiling identifies TITF1 as a lineage-specific oncogene amplified in lung cancer. Oncogene 2008;27:3635–3640.

5 Garraway LA, Sellers WR. Lineage dependency and lineage-survival oncogenes in human cancer. Nat Rev Cancer 2006;6:593–602.

6 Perner S, Wagner PL, Soltermann A, *et al.* TTF1 expression in non-small cell lung carcinoma: association with TTF1 gene amplification and improved survival. J Pathol 2009;217:65–72.

7 Barletta JA, Perner S, Iafrate AJ, *et al.* Clinical significance of TTF-1 protein expression and TTF-1 gene amplification in lung adenocarcinoma. J Cell Mol Med 2009;13:1977–1986.

8 Qian J, Massion PP. Role of chromosome 3q amplification in lung cancer. J Thorac Oncol 2008;3:212–215.

9 Bass AJ, Watanabe H, Mermel CH, *et al.* SOX2 is an amplified lineage-survival oncogene in lung and esophageal squamous cell carcinomas. Nat Genet 2009;41:1238–1242.

10 Hussenet T, Dali S, Exinger J, *et al.* SOX2 is an oncogene activated by recurrent 3q26.3 amplifications in human lung squamous cell carcinomas. PLoS One 2010;5:e8960.

11 Kamachi Y, Uchikawa M, Kondoh H. Pairing SOX off: with partners in the regulation of embryonic development. Trends Genet 2000;16:182–187.

12 Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell 2006;126:663–676.

13 Avilion AA, Nicolis SK, Pevny LH, *et al.* Multipotent cell lineages in early mouse development depend on SOX2 function. Genes Dev 2003;17:126–140.

14 Chew LJ, Gallo V. The Yin and Yang of Sox proteins: activation and repression in development and disease. J Neurosci Res 2009;87:3277–3287.

15 Gontan C, de Munck A, Vermeij M, *et al.* Sox2 is important for two crucial processes in lung development: branching morphogenesis and epithelial cell differentiation. Dev Biol 2008;317:296–309.

16 Que J, Okubo T, Goldenring JR, *et al.* Multiple dose-dependent roles for Sox2 in the patterning and differentiation of anterior foregut endoderm. Development 2007;134:2521–2531.

17 Yuan P, Kadara H, Behrens C, *et al.* Sex determining region Y-Box 2 (SOX2) is a potential cell-lineage gene highly expressed in the pathogenesis of squamous cell carcinomas of the lung. PLoS One 2010;5:e9112.

18 Sholl LM, Long KB, Hornick JL. Sox2 expression in pulmonary non-small cell and neuroendocrine carcinomas. Appl Immunohistochem Mol Morphol 2010;18:55–61.

19 Sholl LM, Barletta JA, Yeap BY, *et al.* Sox2 protein expression is an independent poor prognostic indicator in stage I lung adenocarcinoma. Am J Surg Pathol 2010;34:1193–1198.

20 McCaughan F, Pole JC, Bankier AT, *et al.* Progressive 3q amplification consistently targets SOX2 in preinvasive squamous lung cancer. Am J Respir Crit Care Med 2010;182:83–91.

21 Freier K, Knoepfle K, Flechtenmacher C, *et al.* Recurrent copy number gain of transcription factor SOX2 and corresponding high protein expression in oral squamous cell carcinoma. Genes Chromosomes Cancer 2010;49:9–16.

22 Long KB, Hornick JL. SOX2 is highly expressed in squamous cell carcinomas of the gastrointestinal tract. Hum Pathol 2009;40:1768–1773.

23 Maier S, Wilbertz T, Braun M, *et al.* SOX2 amplification is a common event in squamous cell carcinomas of different organ sites. Hum Pathol 2011 (e-pub ahead of print).

24 Lengerke C, Fehm T, Kurth R, *et al.* Expression of the embryonic stem cell marker SOX2 in early-stage breast carcinoma. BMC Cancer 2011;11.

25 Lu Y, Futtner C, Rock JR, *et al.* Evidence that SOX2 overexpression is oncogenic in the lung. PLoS One 2010;5:e11022.

26 Zhang X, Yu H, Yang Y, *et al.* SOX2 in gastric carcinoma, but not Hath1, is related to patients' clinicopathological features and prognosis. J Gastrointest Surg 2010;14:1220–1226.

27 Soltermann A, Tischler V, Arbogast S, *et al.* Prognostic significance of epithelial-mesenchymal and mesenchymal-epithelial transition protein expression in non-small cell lung cancer. Clin Cancer Res 2008;14:7430–7437.

28 Hussenet T, du Manoir S. SOX2 in squamous cell carcinoma: amplifying a pleiotropic oncogene along carcinogenesis. Cell Cycle 2010;9.

29 Ben-Porath I, Thomson M, Carey V, *et al.* An embryonic stem cell-like gene expression signature in poorly differentiated aggressive human tumors. Nat Genet 2008;40:499–507.

30 Jeter C, Badeaux M, Choy G, *et al.* Functional evidence that the self-renewal gene NANOG regulates human tumor development. Stem Cells 2009;27:993–1005.

31 Rosen J, Jordan C. The increasing complexity of the cancer stem cell paradigm. Science 2009;324:1670–1673.