European Journal of Cardio-Thoracic Surgery 41 (2012) e173–e181
doi:10.1093/ejcts/ezs170

ORIGINAL ARTICLE

High expression of octamer-binding transcription factor 4A, prominin-1 and aldehyde dehydrogenase strongly indicates involvement in the initiation of lung adenocarcinoma resulting in shorter disease-free intervals†

Lourdes Cortes-Dericks[a,†], Domenico Galetta[b,†], Lorenzo Spaggiari[b], Ralph Alexander Schmid[a,*] and Golnaz Karoubi[a]

[a] Department of Clinical Research, Division of General Thoracic Surgery, University Hospital Bern, Bern, Switzerland
[b] Division of Thoracic Surgery, European Institute of Oncology, Milan, Italy

* Corresponding author. Division of General Thoracic Surgery, University Hospital Bern, Bern CH-3010, Switzerland. Tel: +41-31-6322330; fax: +41-31-6322327; e-mail: ralph.schmid@insel.ch (R.A. Schmid).

Received 1 September 2011; received in revised form 21 December 2011; accepted 9 January 2012

Abstract

OBJECTIVES: The increasing relevance of the cancer stem cell (CSC) hypothesis and the impact of CSC-associated markers in the carcinogenesis of solid tumours may provide potential prognostic implications in lung cancer. We propose that a collective genetic analysis of established CSC-related markers will generate data to better define the role of putative CSCs in lung adenocarcinoma (LAC).

METHODS: Sixty-four paired tumour and non-tumour biopsies from LAC patients were included in this study. Using the quantitative reverse transcriptase–polymerase chain reaction, we assessed the expression profiles of established CSC-related biomarkers: octamer-binding transcription factor 4 (OCT4A), CD133, aldehyde dehydrogenase (ALDH), BMI-1, ATP-binding cassette subfamily G, member 2 (ABCG2), SRY (sex-determining region Y)-box 2 (SOX2) and uPAR, and evaluated their relation to clinicopathological parameters and disease prognosis.

RESULTS: All of the above-mentioned CSC-related markers were detectable in both tumour and corresponding normal tissues. Importantly, expression levels of OCT4A, CD133, BMI-1, SOX2 and uPAR were significantly higher (OCT4A, $P = 0.0003$; CD133, $P = 0.002$; BMI-1, $P = 0.04$; SOX2, $P = 0.0003$; uPAR, $P = 0.03$) in the tumour compared with those in the non-tumour tissues. By contrast, the quantities of ACBG2 and ALDH were markedly reduced (ACBG2, $P = 0.0006$; ALDH, $P = 0.007$) in the tumour relative to those in the normal biopsies. Using multivariate analysis, elevated ALDH and CD133 revealed significant associations in tumour stage (ALDH, $P = 0.03$; CD133, $P = 0.007$) and differentiation (ALDH, $P = 0.03$; CD133, $P = 0.018$). We observed that ALDH and OCT4A were associated with nodal status (ALDH, $P = 0.05$; OCT4A, $P = 0.03$) having lower mRNA levels in tumours with lymph node metastasis, N+, compared with that in N0. High OCT4A levels were significantly correlated with tumour size of <3 cm, decrease in tumours >3 cm ($P = 0.03$). Kaplan–Meier correlation analyses, showed that OCT4A and CD133 were correlated to short disease-free intervals (OCT4A, $P = 0.047$; CD133, $P = 0.033$) over a period of 29 months.

CONCLUSIONS: Our study reveals that CSC-associated markers: OCT4A, CD133 and ALDH are involved in the initial phase of carcinogenesis of LAC, and can be used as predictors of early stage LAC and poor disease-free intervals. In addition, this work validates the relevance of the CSC hypothesis in LAC.

Keywords: Lung cancer • Adenocarcinoma • Cancer stem cells • Prominin-1 (CD133) • Aldehyde dehydrogenase (ALDH) • Octamer-4 (OCT4A)

BASIC SCIENCE

INTRODUCTION

Lung cancer is one of the major causes of cancer-related death worldwide. Adenocarcinoma, the most common form of non-small cell lung cancer (NSCLC), is the most prevalent, representing 40% of the NSCLC cases, with an increasing incidence and <15% of patients with an overall 5-year survival [1]. Hence, further exploration of the factors surrounding the oncogenesis of this tumour may reveal novel approaches to improve the efficacy of treatment and assure a better clinical outcome.

The cancer stem cell (CSC) model is a conceivable explanation for the oncogenesis of lung adenocarcinoma (LAC). The hallmarks of CSCs to self-renew and undergo asymmetric divisions

[†] Presented at the 25th Annual Meeting of the European Association for Cardio-Thoracic Surgery, Lisbon, Portugal, 1–5 October 2011.
[†] Both authors contributed equally.

producing differentiated progeny enable CSCs to initiate as well as promote cancer progression. An increasing number of studies have underscored the relevance of the expression profiles of CSC-associated markers and/or embryonic stem cell genes in relation to cancer initiation, progression and clinical prognosis, thus presenting these as prognostic tools in predicting patient survival. Established CSC indicators in lung cancer include aldehyde dehydrogenase (ALDH), prominin-1 (CD133), ATP-binding cassette subfamily G, member 2 (ABCG2), octamer-binding transcription factor 4 (OCT4A) and SRY (sex-determining region Y)-box 2 (SOX2). Urokinase plasminogen activator receptor (uPAR) and polycomb ring finger oncogene (BMI-1) have also been linked to lung tumorigenesis.

ALDH is a detoxifying enzyme, sought to protect organisms against various aldehydes that would otherwise cause damage and has been shown to play an oncogenic role in mammary, prostate, ovarian and lung cancers [2]. OCT4A, a transcription factor essential for maintaining pluripotency in embryonic stem cells, has been proposed to have an essential role in the tumorigenesis of solid tumours including those of bladder, gastric and prostate [3]. Moreover, the presence of OCT4A, Nanog and slug have been found in high-grade LAC tumours and marked a poor prognosis for patients [4]. CD133 is a five-transmembrane glycoprotein used to identify both normal and CSCs in different tissues. Increased expression of CD133 has proved to be a useful factor for prediction of correlation to prognosis, recurrence and chemosensitivity in colorectal cancer [5]. Additionally, it has been associated with lower survival resulting in poor prognosis in gastric carcinoma, pancreatic cancer and glioma [6]. Its implications in tumorigenicity and chemoresistance in lung cancer have been confirmed by Bertolini *et al*. [7].

A number of studies have demonstrated the expression of and associations among BMI-1, uPAR, SOX2, ABCG2 in the initiation, maintenance and prognosis of specific neoplastic tissues. BMI-1 is a component of the polycomb repressive complex 1, a key epigenetic factor, which also controls the cell cycle and self-renewal of tissue stem cells. Overexpression of BMI-1 is found to correlate with poor prognosis in non-keratinizing types of nasopharyngeal carcinoma, breast or hepatocellular carcinoma, as well as in the nervous system [8]. SOX2 is a transcription factor essential for early mammalian development and for the sustainment of embryonic and adult stem cells. Overexpression of SOX2 has been observed in different types of solid tumours including the lung [9], thereby manifesting an oncogenic role. uPAR is the cellular receptor for urokinase-type plasminogen activator, which activates plasminogen into plasmin, and has been a target in cancer research due to a role in both the progression of malignant cancers and as a prognostic tool for poor patient survival [10]. ABCG2, expressed in both normal and cancer cells, plays a pivotal role in side population cells and efflux of xenobiotics and drugs [11]. In the lung, a dual positive expression of CD133 and ABCG2 has been shown to be an independent predictor of postoperative recurrence for patients with stage 1 NSCLC [11].

Considering the influence of CSC-associated markers in evaluating disease outcome and patient survival, we reasoned that assessing the mRNA expression of these markers collectively would generate data to support a plausible role of CSC-associated biomarkers in the tumorigenesis of LAC. In this study, we determined the mRNA levels of ALDH, CD133, ABCG2, OCT4A, SOX2, BMI-1 and uPAR on resected paired normal and tumour LAC tissues. Their expression profiles were correlated to the TNM staging and clinicopathological features of this disease. Our results reveal that only the presence of ALDH, CD133 and OCT4A have a significant clinical impact on the initiation rather than progression and/or maintenance of this neoplasm. These markers may serve as possible predictive factors in early stage LAC manifesting significantly shorter disease-free intervals. To the best of our knowledge, this is the first study to examine LAC tumour and non-tumour tissues with respect to CSC-marker expression profiles and their relevance to the clinicopathological parameters of this disease.

MATERIALS AND METHODS

Sample collection

Normal and matched malignant biopsies of LAC were obtained from 64 consenting patients undergoing tumour resection from January 2008 to January 2011 at the European Institute of Oncology in Milan, Italy. Diagnosis was based on clinical and histological parameters. LAC tumour staging (pTNM) was classified according to the current International Staging System. All patients included in the study were subjected to primary surgery. Patient age ranged from 34 to 86; 53% were male and 47% female. None of the patients had undergone chemotherapy or radiotherapy prior to surgery. Paired tumour and non-tumour lung tissues were obtained by surgical resection. Tumour tissue was resected from the periphery of the tumour, whereas normal specimens were sampled at least 6 cm away from the tumour and were confirmed as non-malignant by pathological examination. All biopsies ($\sim 5 \times 5$ mm^2) were immediately stored in RNAlater solution (Qiagen; Hombrechtikon, Switzerland) and stored at −20° C. The study was approved by both the medical ethical committees of University Hospital Berne (Berne, Switzerland) and the European Institute of Oncology (Milan, Italy).

RNA extraction and quantitative real-time reverse transcriptase–polymerase chain reaction

Tissue biopsies, preserved in RNAlater (Qiagen) were processed by total RNA extraction (RNeasy Kit, Qiagen) according to the manufacturer's instructions. Complementary DNA (cDNA) was synthesized using the High Capacity cDNA Reverse Transcription kit (Applied Biosystems; Rotkreuz, Switzerland) and treated with RNase-Free DNase (Qiagen). The mRNA transcript levels of the housekeeping gene *β2*-Microglobulin, *β2M*, and target genes *OCT4A*, *CD133*, *BMI-1*, *SOX2*, *uPAR*, *ALDH1A1 (ALDH)* and *ABCG2* were evaluated with commercially available TaqMan® 'Assay on Demand' primer/probes (*β2M*, Hs00984230_m1; OCT4A, Hs03005111_g1; *CD133*, Hs01009254_m1; *BMI-1*, Hs00180411_m1; *SOX2*, Hs01053049_s1; *uPAR*, Hs00182181_m1; *ALDH1A1*, Hs00964880_m1 and *ABCG2*, Hs01055362_m1) (Applied Biosystems; Rotkreuz). Twenty-five nanograms of resulting cDNAs were subjected to quantitative reverse transcriptase–polymerase chain reaction (qRT–PCR) in a 12.5 μl final reaction volume and analysed in triplicate. Gene expression was detected using ABI 7900 sequence detection system. All target gene C_t values in each parameter were normalized by the reference gene, *β2M*. The baseline and the threshold for C_t calculation were set automatically with the ABI Prism SDS 2.1 software. The qRT–PCR data represent the relative quantity of the target gene mRNA

(target gene mRNA/*β2M* mRNA ratio) in comparison with the human embryonic carcinoma cell line NTERA-2, which was used as a calibrator and set at 1.

Statistical analysis

To assess the statistical significance of differences observed in the mRNA expression levels in tumour specimens and normal counterparts, the Wilcoxon rank-sum, two-tailed test was performed. Correlation between mRNA expression and the clinicopathological parameters of tumour stage, and tumour grade were analysed by an analysis of variance, specifically the Kruskal–Wallis test, followed by the Dunn's *post hoc* method. Patient age was dichotomized to 65 years of age or under and above 65 years. Similarly, nodal status was dichotomized to nodal involvement and no nodal involvement. For associations between the continuous variable (gene expression) and dichotomous variables (patient sex, age and nodal status), the Mann–Whitney *U*, non-parametric, two-tailed test was performed. Spearman's rank correlation coefficients (*r*) were calculated to estimate the correlation between expression levels of different genes studied. Kaplan–Meier survival analysis and the log-rank significance test were utilized to analyse univariate distributions for disease-free interval. Furthermore, a multivariate analysis was done by using Cox proportional hazards regression to determine the prognostic effect of OCT4A, CD133 and ALDH expression on disease-free interval. Hazard ratios and their corresponding 95% confidence intervals were computed to provide quantitative information about the relevance of results of the statistical analysis. *P*-values <0.05 were defined as statistically significant (*). All statistical data were obtained using the GraphPad Prism version 5.03 (GraphPad Software Inc., La Jolla, USA).

RESULTS

Patient demographics

Sixty-four patients with LAC were included in the study. The patients consisted of 34 men and 30 women. The median age was 62, 38 (59%) aged over 65 years. The histologic classification of the tumour was based on the World Health Organization

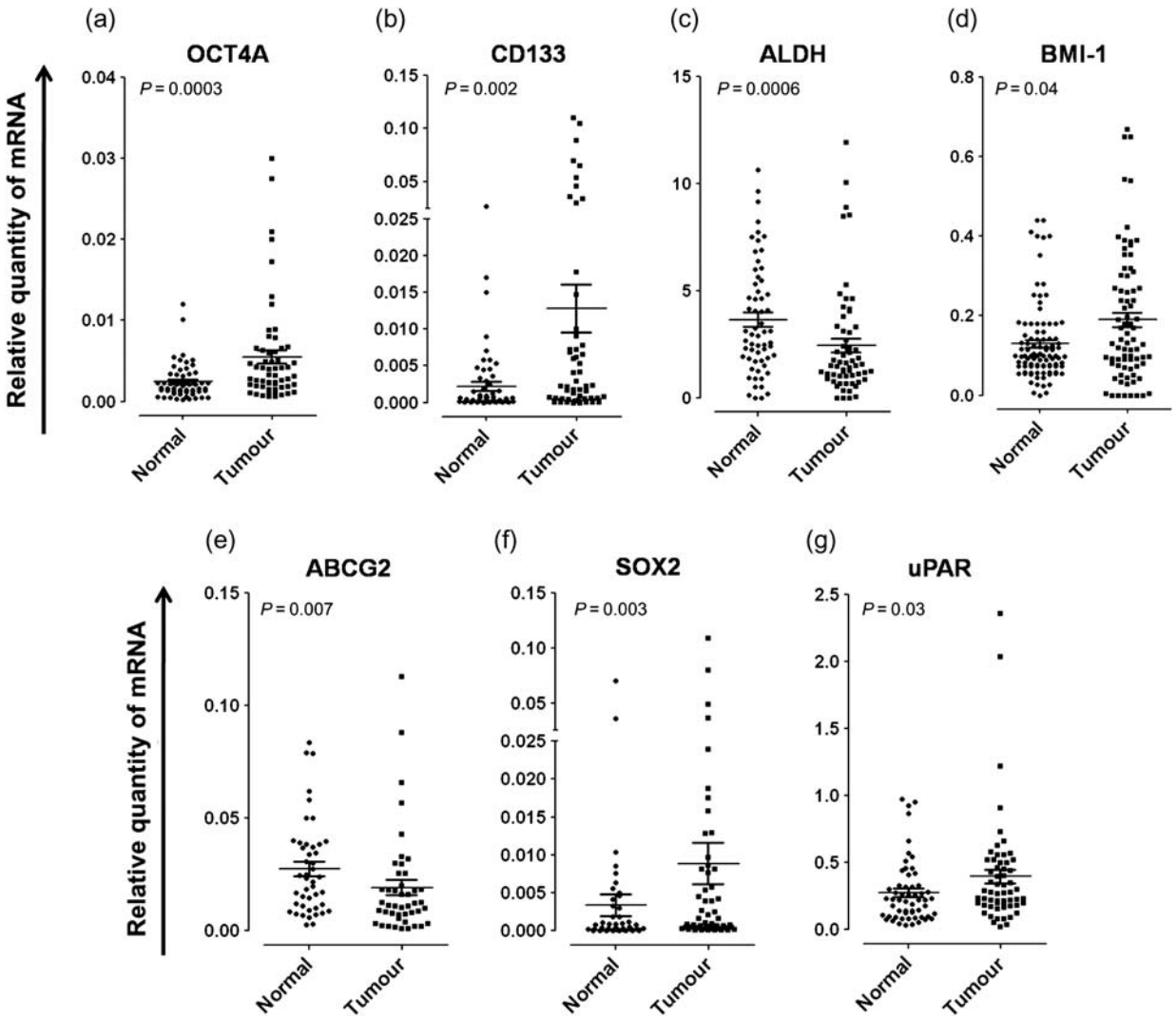

Figure 1: qRT–PCR mRNA analyses showing the relative mRNA levels in paired tumour and non-cancerous lung tissue. (**a**) OCT4A; (**b**) CD133; (**c**) ALDH; (**d**) BMI-1; (**e**) ABCG2; (**f**) SOX2; (**g**) uPAR. Messenger RNA levels were measured using the $\Delta\Delta C_t$ method, relative to the calibrator, NTERA-2, an embryonic carcionoma cell line.

criteria. Thirty-eight (59%) had stage I disease, 14 patients (22%) stage II disease and 12 patients (19%) had stage III disease. Of the 64 patients, 23 (36%) had tumours with nodal metastasis and the majority, 43 (67%) had a tumour size of <3 cm. The grade of the tumours was equally distributed with 18 (28%) G1, 24 (37.5%) G2 and 22 (34%) G3. Sixty-eight per cent of the cases had mixed subtype histology. Acinar pattern was the predominant subtype (34%), followed by papillary (23%), solid (22%) and bronchioloalveolar (20%). Of the cases that contained a single histologic subtype, 12 (60%) were acinar, 6 (30%) were bronchioloalveolar and 5 (25%) were solid.

Cancer stem cell-associated mRNA expression levels in paired normal and tumour biopsies

One of the main objectives of our study was to determine if there was a difference in the frequency of CSC-associated mRNA levels between paired tumour and non-tumour tissues. qRT–PCR analyses showed that CSC-associated biomarkers OCT4A, CD133, BMI-1, SOX2, uPAR, ALDH and ABCG2 were detectable in the majority of the paired tumour and normal LAC tissues. Sixty-one of 64 pairs (95%) and 60/64 (94%) showed OCT4A and SOX2 expression with considerably higher OCT4A ($P = 0.0003$) and SOX2 ($P = 0.0003$) levels in the malignant tissues (Fig. 1a and f). CD133 (Fig. 1b), BMI-1 (Fig. 1d), uPAR (Fig. 1g) signals were found in 63/64 (98%), 62/64 (97%) and 64/64 (100%) of the sample pairs, respectively, likewise, with significantly higher levels in the tumour tissues (CD133, $P = 0.002$; BMI-1, $P = 0.04$; uPAR, $P = 0.03$). ALDH and ABCG2 expression were also detectable in 60/64 (94%) and 55/64 (86%) of the pairs (Fig. 1c and e) but in contrast to the above-mentioned genes, the mRNA signals in the tumour tissues were noticeably lower for both markers (ALDH, $P = 0.0006$; ABCG2, $P = 0.007$). As cigarette smoke has been correlated to increased expression of ALDH, we evaluated whether the high expression of this maker in the normal tissues could be partly attributed to smoking. A comparative expression profile of the cohorts of smokers and non-smokers in normal tissues, illustrated a significantly higher ALDH expression in the smoker group (Fig. 2a). Removal of the smoking cohort resulted in significantly elevated levels of ALDH mRNA in tumour tissue biopsies compared with corresponding normal tissues (Fig. 2b).

Associations between cancer stem cell-associated genes and clinicopathological parameters

To explore the relevance of the CSC-associated genes to clinicopathological parameters, correlation studies were performed. Associations between the clinicopathological features of the 64 patients with LAC and putative CSC biomarkers are shown in Table 1. There were no obvious interdependencies between any of the genes evaluated: OCT4A, CD133, BMI-1, SOX2, uPAR, ALDH and ABCG2 and age or gender of the patients. In addition, no substantial correlation of SOX2, uPAR and ABCG2, to tumour stage, nodal status, tumour size, grade and histological subtype was observed, except for BMI-1, which showed an association with tumour size.

Evaluation by the Kruskal–Wallis test, an analysis of variance, revealed marked associations of ALDH and CD133 with both tumour stage as well as differentiation (Fig. 3a, b, d and e). Elevated ALDH and CD133 mRNA signals were observed in the early phases of tumour stage (ALDH, $P = 0.02$; CD133, $P = 0.007$) and differentiation (ALDH, $P = 0.03$; CD133, $P = 0.02$) which declined upon progression of these parameters. Importantly, higher mRNA levels of these genes were found in poorly differentiated tumours (CD133: G1 vs. G3, $P = 0.02$; ALDH: G1 vs. G3, $P = 0.03$) as well as in stage I tumours (CD133: stage I vs. stage III, $P = 0.006$; ALDH: stage I vs. stage III, $P = 0.08$). We also observed that ALDH and OCT4A were significantly related to nodal status (Fig. 2c and g) with lower mRNA levels in patients with lymph node metastasis, N+ compared with that in N0. Increased OCT4A occurrence correlated to tumour size of <3 cm, which was markedly lower in tumours of >3 cm (Fig. 3f).

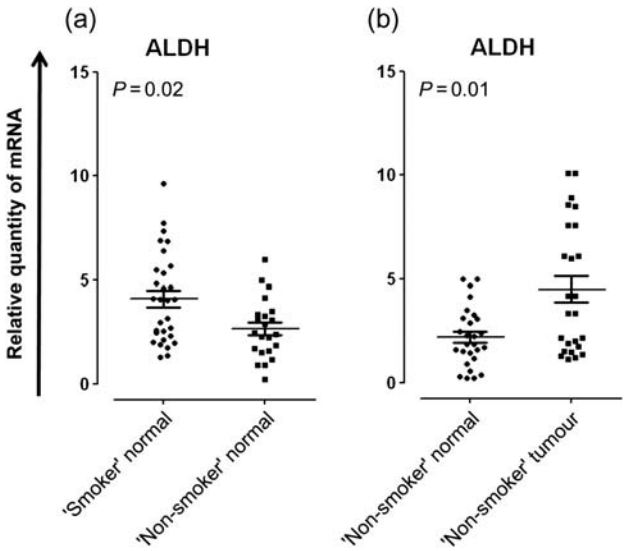

Figure 2: qRT–PCR mRNA analyses showing the relative ALDH mRNA levels in (**a**) normal lung tissue from 'smokers' in comparison with 'non-smokers' and (**b**) paired tumour and non-cancerous lung tissue of the 'non-smoking' cohort. Messenger RNA levels were measured using the $\Delta\Delta C_t$ method, relative to the calibrator, NTERA-2, an embryonic carcionoma cell line.

Given the previously reported evidence of interdependency of CSC markers, we explored this relationship between OCT4A, CD133 and ALDH. Spearman correlation coefficients were calculated to determine a potential correlation of OCT4A, CD133 and ALDH (Fig. 4). The results revealed that the degree of expression levels of OCT4A were positively related with those of CD133 ($r = 0.4$; $P = 0.002$). No correlation was observed between OCT4A and ALDH or CD133 and ALDH.

Cancer stem cell-associated gene expression in relation to disease-free progression and survival

The median disease-free interval and median survival time for the patients were 13 months (range, 1–29 months) and 16.5 months (range, 7–29 months), respectively. A cut-off value was defined by taking the median value for the gene transcript levels of OCT4A, CD133 and ALDH. Values higher than the median were considered as high expression; and values below the median were regarded as low expression. Univariate analysis (log-rank test) of the effect prognostic parameters on disease-free interval depicted that OCT4A ($P = 0.047$) and CD133 ($P = 0.033$) expression levels were independent prognostic factors of disease-free interval (Table 2). Patients with low expression of OCT4A and CD133 had markedly longer disease-free intervals

Table 1: Clinicopathological variables and the corresponding expression of CSC-associated genes

Variables	Total	OCT4A		*P*-value	CD133		*P*-value	BMI-1		*P*-value	SOX2		*P*-value	uPAR		*P*-value	ALDH		*P*-value	ABCG2		*P*-value
		−(*n*)	+(*n*)		−(*n*)	+(*n*)		−(*n*)	+(*n*)		−(*n*)	+(*n*)		−(*n*)	+(*n*)		−(*n*)	+(*n*)		−(*n*)	+(*n*)	
Total	64	3	61		1	63		2	62		4	60		0	64		1	60		9	55	
Age																						
>65	38	2	36	0.72	1	37	0.81	1	37	0.76	3	35	0.68	0	38	0.82	3	35	0.97	4	34	0.88
<65	26	1	25		0	26		1	25		1	25		0	26		1	25		5	21	
Gender																						
Male	34	2	32	0.91	1	33	0.65	2	32	0.83	3	31	0.73	0	34	0.71	2	32	0.59	5	29	0.92
Female	30	1	29		0	30		0	30		1	29		0	30		2	28		4	26	
Smoking																						
Yes	33	2	31	0.65	1	32	0.58	1	32	0.87	1	32	0.79	0	33	0.28	2	31	0.33	6	28	0.16
No	31	1	30		0	31		1	30		3	28		0	31		2	29		3	27	
Pathological stage																						
I	38	1	37	0.95	0	38	0.007	1	37	0.96	4	34	0.49	0	38	0.26	2	36	0.03	5	33	0.64
II	14	1	13		0	14		1	13		0	14		0	14		0	14		1	13	
III	12	0	12		1	11		0	12		0	12		0	12		2	10		3	9	
Nodal status																						
N0	23	1	22	0.25	0	23	0.37	1	22	0.26	3	20	0.98	0	23	0.08	2	21	0.047	5	18	0.85
N+	41	2	39		1	40		1	40		1	40		0	41		2	39		4	37	
Tumour size																						
≤3 cm	43	1	42	0.027	1	42	0.28	1	42	0.002	4	39	0.14	0	43	0.53	0	43	0.21	3	40	0.81
>3 cm	21	2	19		0	21		1	20		0	21		0	21		3	18		6	15	
Tumour grade																						
G1	18	0	18	0.47	0	18	0.02	0	18	0.51	2	16	0.48	0	18	0.42	1	17	0.03	2	18	0.56
G2	24	2	22		0	24		2	22		0	24		0	24		1	23		3	22	
G3	22	1	21		1	21		0	22		2	20		0	22		2	20		4	22	
Histological subtype																						
Solid	14	1	13	0.007	0	14	0.02	0	14	0.20	1	13	0.08	0	14	0.77	1	13	0.009	3	11	0.06
Acinar	22	2	21		0	22		1	21		2	20		0	22		2	20		4	18	
Papillary	15	0	15		1	15		1	14		1	14		0	15		1	14		2	13	
Bronchioloalveolar	13	0	13		0	13		0	13		0	13		0	13		0	13		0	13	

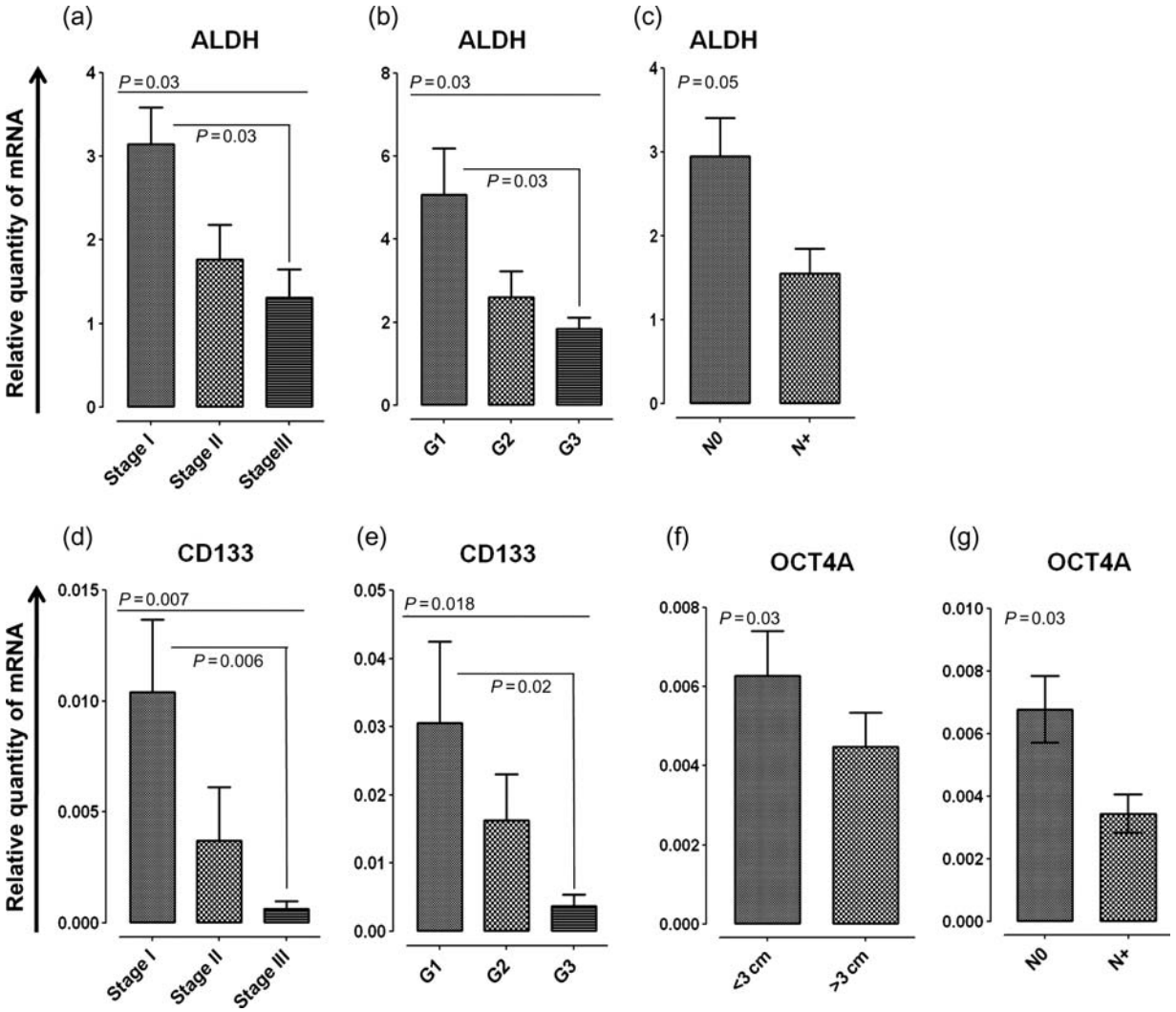

Figure 3: Frequency of mRNA levels of relevant CSC-associated genes based on the pTNM staging. ALDH correlation to stage (**a**) nodal status (**b**) and tumour grade (**c**); association of CD133 with stage (**d**) and tumour differentiation (**e**); and OCT4A correlation to tumour size (**f**) and lymph node metastasis (**g**).

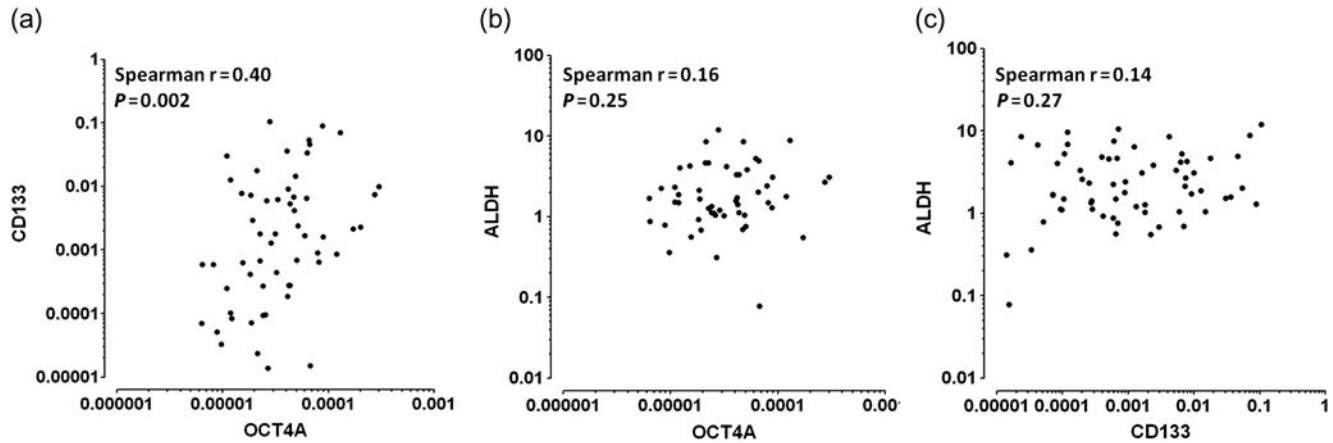

Figure 4: Correlation of CD133, OCT4A and ALDH expression in LAC tumour biopsies. (**a**) A significant association between CD133 and OCT4A (Spearman $r = 0.4$, $P = 0.002$). No correlation was observed between ALDH and OCT4A (**b**), and between ALDH and CD133 (**c**).

(Fig. 5a and b). Although the Kaplan–Meier disease-free interval curve for ALDH shows a trend for a better patient disease outcome in patients with lower ALDH mRNA signals, this did not reach a significant level (Fig. 5c).

DISCUSSION

In the present study, we determined the pattern and expression levels of CSC-associated biomarkers: OCT4A, CD133, ABCG2,

SOX2, BMI-1, uPAR and ALDH in paired tumour and non-tumour LAC biopsies. Further, we assessed their associations with the different clinicopathological parameters, which may serve as a basis for use as prognostic indicators in the neoplasm. Our results illustrated significantly increased transcript levels of OCT4A, SOX2, CD133, BMI-1 and uPAR in tumour tissues compared with those in the corresponding non-tumour biospsies. In contrast, ABCG2 and ALDH were significantly reduced in the tumour relative to those in the corresponding normal biopsies. Correlation data affirmed that only OCT4A, CD133 and ALDH exert significant influence on the clinicopathological parameters and disease-free intervals of LAC patients. Additionally, we observed a strong interdependency in the expression between OCT4A and CD133.

Current studies have reported a strong presence of gene expression, positivity in immunostaining or protein serum levels of OCT4A, SOX2, CD133, uPAR, ALDH, ABCG2 and BMI-1, in LAC tumour specimens [4, 9, 12, 13]. Data emerging from our study indicate not only the presence of these markers in tumours but, importantly, higher expression profiles of OCT4A, CD133, BMI-1, SOX2 and uPAR in the tumours compared with their matched normal tissue counterparts. Up-regulation of these CSC-associated biomarkers strongly suggest the existence of CSCs in LAC. Although not yet evident, one current belief of lung tumorigenesis proposes that CSCs originate from the transformation of lung progenitor cells or lung epithelial cells by mutagens. Once mutated, these CSCs obtain limitless capacity for proliferation through dysregulated self-renewal. Given this premise, it is plausible that lung CSCs may still reflect the hallmarks and phenotypes of normal stem cells [1]. Our data, showing the increased expression levels of the above-mentioned CSC markers in the tumour tissues relative to the normal counterparts, strongly supports this model of carcinogenesis.

Contrary to the above, we found that levels of ALDH and ABCG2 were significantly lowered in the paired malignant tissues relative to their normal counterparts. Being a transporter of certain chemotherapeutic drugs, most of the studies on ABCG2 expression profiles have been associated with drug response and general disease outcome. The presence of ABCG2 in NSCLC has also been considered as a predictor of lower drug response, and shorter progression-free and overall survival [14]. In this study, the expression profile of ABCG2 in the LAC tumour was found to have no significant relevance to any of the clinicopathological factors evaluated; hence, we could not assign any pertinent role to this marker in the oncogenesis of LAC.

ALDH activity is a prevailing stem cell and CSC marker in both normal and malignant tissues, respectively. However, there are limited data on the comparison of expression of ALDH in cancers and their corresponding normal tissues. In the lung, most of the studies have been focused on the assessment of the presence or absence of ALDH in the tumour tissues. To the best of our knowledge, this is the first work to explore the differential expression of ALDH between matched normal and tumour biopsies. Patel *et al*. [15] have recently demonstrated reduced ALDH expression in pneumocytes derived from normal lungs of non-smokers compared with elevated levels in those derived from smokers. This group reasoned that chronic exposure to tobacco smoke may have induced the increased levels of ALDH as a cellular protective mechanism against this insult. We indeed observed a significantly lower ALDH expression profile in the normal tissues of non-smokers compared with the smoker group. However, in the absence of the smoking cohort, a comparison of tumour and paired normal tissues illustrates the expected up-regulation of ALDH in tumour tissues of LAC. The question of whether ALDH has a protective or oncogenic role in LAC remains to be answered.

We examined the correlation of CSC-associated gene expression in the tumour samples, with patient characteristics, tumour pathology and disease-free interval, and demonstrated direct associations of CD133, OCT4A and ALDH to the above

Table 2: Cox proportional hazards regression for disease-free interval

Patients	No. (%)	Hazard ratio	95% CI	*P*-value
OCTA expression				
Low	31	3.81	1.02–14.3	0.047
High	30			
CD133 expression				
Low	31	3.93	1.12–13.8	0.033
High	32			
ALDH expression				
Low	30	3.30	0.91–11.9	0.068
High	30			

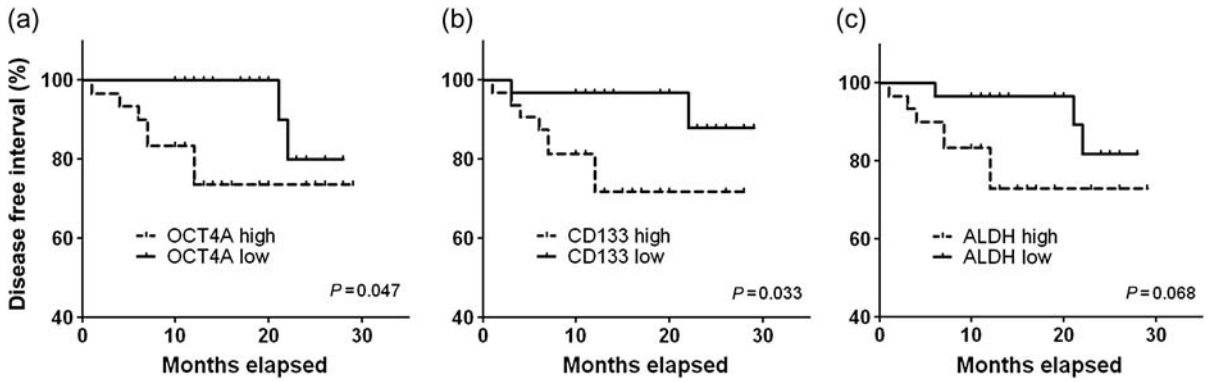

Figure 5: The disease-free interval curves based on (**a**) OCT4A, (**b**) CD133 and (**c**) ALDH cut-off values. Patients with high OCT4A and CD133 levels (above cut-off values) showed significantly (OCT4A, $P = 0.047$; CD133, $P = 0.033$) low disease-free intervals. Patients with increased ALDH show the same trend but did not reach a statistical significant (ALDH, $P = 0.068$). Kaplan–Meier survival analysis was used to create the disease-free interval curve.

parameters. Our data indicated a significant relationship between ALDH and tumour stage and tumour grade in LAC. In both parameters, we observed significantly increased levels of ALDH in stage I and G1 with a declining trend towards the progression of the tumour. One of the hallmarks of CSCs is the capacity to undergo asymmetrical division resulting in multi-lineage differentiation giving rise to phenotypically diverse tumour phenotypes. The majority of the studies focusing on ALDH in the malignant setting have shown that $ALDH^+$ cells are more tumorigenic *in vitro* and *in vivo*, clonogenic, and are able to produce generations of $ALDH^+$ and $ALDH^-$ cells [8, 16], thereby highlighting the CSC status of $ALDH^+$ cells in neoplastic tissues. Considering the above, we propose that the increased ALDH expression during the early stages and in poorly differentiated LAC tumours may be due to a capacity for aberrant self-renewal, a CSC hallmark, thus implicating $ALDH^+$ cells in the tumour initiation of LAC. Further, the reduced levels of ALDH in the later stages of LAC could be justified by the fact that ALDH-expressing cells undergo asymmetric division, generating a diverse phenotype and therefore resulting in reduced ALDH positivity.

ALDH activity is associated with poor prognosis in leukaemia, melanoma and breast, bladder, prostrate, human epithelial cancer patients [2]. A reduced patient survival has been correlated to ALDH positivity in stage I NSCLC and in NSCLC patients [8, 16]. Consistent with these findings, we observed a trend for poor prognosis for patients with high ALDH transcripts, but a significant association with a disease-free interval could not be depicted. These results, in part, may be attributed to a short follow-up period of 29 months. Hence, these findings would require further prospective validation of ALDH activity with respect to patient prognosis.

Our results provide evidence of a significant association of CD133 with both tumour stage and differentiation in LAC. The enhanced CD133 levels in the initial phases of tumour stage and grade of differentiation together with an observed reduction in both the later stages of these parameters highly indicate an involvement of CD133 in the tumour initiation of LAC. Evidence for a key role of CD133 in tumour initiation has been shown in several other malignancies. Singh and Dirks [17] first reported that $CD133^+$ cells from brain tumours showed self-renewal and differentiation and were able to recapitulate brain tumour formation *in vivo*. Since then, inoculation of tumour-derived $CD133^+$ cells into immune-deficient mice has demonstrated the capacity to re-form the original tumour in human brain, pancreas, colon and liver carcinomas [6]. Together with our results, these studies provide evidence in support of CD133 as a specific surface molecule of CSCs in solid tumours, suggesting that CD133 is likely to become an effective target of anti-tumour therapy. With respect to patient prognosis, our results show that tumours with high levels of CD133 transcripts are associated with shorter disease-free intervals in comparison with those expressing low CD133. The role of CD133 as a prognostic marker in lung cancer has not yet been fully elucidated [11, 18–20]. Woo *et al*. [20] and Xu *et al*. [18] noted an association between CD133 and poor patient prognosis, while others found no relationship between the clinicopathological parameters or patient prognosis [11, 19]. Our results fit in with the studies of the former groups, in which a high CD133 level correlates with diminished disease-free intervals, contributing to a poor disease outcome.

Our group has previously reported an aberrant up-regulation of embryonic transcription factor, OCT4A, in LAC tumour tissue biopsies [21]. In this study, we found a significant association of OCT4A to both tumour size and lymph node metastasis. OCT4A amplification was markedly higher in tumours <3 cm in comparison with larger tumours (>3 cm). In patients without lymph node metastasis (N0), OCT4A mRNA levels were high but significantly reduced in the tumours with evidence of lymph node metastasis (N+). Collectively, these findings point to an involvement of OCT4A during the initiation phase of LAC. Self-renewal and pluripotency are central features in tumour initiation and OCT4A is a key regulator in these processes [4]. In support of our studies, OCT4A expression has also been implicated in the tumorigenesis of breast cancer [22]. Additionally, we found that high OCT4A expression was significantly associated with a shorter disease-free interval. This is in congruence with two studies showing unfavourable prognosis and poor survival of LAC patients [4, 23].

Notably, we found a significant correlation between OCT4A and CD133, suggesting an interplay of these two biomarkers in some regulatory mechanisms in the tumorigenesis of LAC. Supporting these findings, Chen *et al*. [24] reported that knocking down of OCT4A expression blocks the ability of lung cancer-CD133 + cells to form spheres, colonies and tumours. As these are properties of CSCs, it is conceivable that OCT4 may have a basic regulatory role in the expression and behaviour of CD133 in LAC. This interesting regulatory mechanism warrants further analysis.

Our study highlights new insights into the cancer biology of LAC. We found an increased expression of OCT4A, CD133, BMI-1, SOX2 and uPAR in the tumour samples relative to their normal counterparts in LAC. It is important to note, however, that although we assessed the expression of CSC markers in the tumour as well as normal tissues of each patient, we cannot completely rule out the effects of other influencing factors such as inflammation [25]. A current review has summarized that inflammation may be involved in several stages of carcinogenesis, from tumour initiation to tumour promotion and even in metastatic progression. Particularly in lung cancer, high serum levels of inflammatory cytokines IL-6 and IL-8 have been associated with increased risk of this neoplasm (reviewed in [25]). Of the seven CSC markers investigated, only OCT4, CD133 and ALDH have significant associations in the early phases of tumour stage, differentiation and nodal metastasis. Our data strongly supports their identity as CSC markers in the lung, implication in the initiation of tumorigenesis and a negative prognostic effect in disease survival of LAC. We conclude that OCT4A, CD133 and ALDH can be used as biomarkers in the early stage of LAC and as a prognostic tool to predict the general disease outcome, thus facilitating better therapeutic strategies.

ACKNOWLEDGEMENT

The authors would like to acknowledge Andrew Welton for his excellent technical assistance in the preparation of this study.

Funding

This research was financially supported by the Bernese Cancer League.

Conflict of interest: none declared.

REFERENCES

[1] van Klaveren RJ, van't Westeinde SC, de Hoop BJ, Hoogsteden HC. Stem cells and the natural history of lung cancer: implications for lung cancer screening. Clin Cancer Res 2009;15:2215–8.

[2] Ma I, Allan AL. The role of human aldehyde dehydrogenase in normal and cancer stem cells. Stem Cell Rev 2011;7:292–306.

[3] Linn DE, Yang X, Sun F, Xie Y, Chen H, Jiang R *et al*. A role for OCT4 in tumor initiation of drug-resistant prostate cancer cells. Genes Cancer 2010;1:908–16.

[4] Chiou SH, Wang ML, Chou YT, Chen CJ, Hong CF, Hsieh WJ *et al*. Coexpression of Oct4 and Nanog enhances malignancy in lung adenocarcinoma by inducing cancer stem cell-like properties and epithelial-mesenchymal transdifferentiation. Cancer Res 2010;70:10433–44.

[5] Takahashi S, Kamiyama T, Tomaru U, Ishizu A, Shida T, Osaka M *et al*. Frequency and pattern of expression of the stem cell marker CD133 have strong prognostic effect on the surgical outcome of colorectal cancer patients. Oncol Rep 2010;24:1201–12.

[6] Wu Y, Wu PY. CD133 as a marker for cancer stem cells: progresses and concerns. Stem Cells Dev 2009;18:1127–34.

[7] Bertolini G, Roz L, Perego P, Tortoreto M, Fontanella E, Gatti L *et al*. Highly tumorigenic lung cancer CD133+ cells display stem-like features and are spared by cisplatin treatment. Proc Natl Acad Sci USA 2009;106:16281–86.

[8] Jiang F, Qiu Q, Khanna A, Todd NW, Deepak J, Xing L *et al*. Aldehyde dehydrogenase 1 is a tumor stem cell-associated marker in lung cancer. Mol Cancer Res 2009;7:330–8.

[9] Sholl LM, Barletta JA, Yeap BY, Chirieac LR, Hornick JL. Sox2 protein expression is an independent poor prognostic indicator in stage I lung adenocarcinoma. Am J Surg Pathol 2010;34:1193–8.

[10] Boonstra MC, Verspaget HW, Ganesh S, Kubben FJ, Vahrmeijer AL, van de Velde CJ *et al*. Clinical applications of the urokinase receptor (uPAR) for cancer patients. Curr Pharm Des 2011;17:1890–910.

[11] Li F, Zeng H, Ying K. The combination of stem cell markers CD133 and ABCG2 predicts relapse in stage I non-small cell lung carcinomas. Med Oncol 2010;28:1458–62.

[12] Almasi CE, Christensen IJ, Hoyer-Hansen G, Dano K, Pappot H, Dienemann H *et al*. Urokinase receptor forms in serum from non-small cell lung cancer patients: relation to prognosis. Lung Cancer 2011;74:510–5.

[13] Vrzalikova K, Skarda J, Ehrmann J, Murray PG, Fridman E, Kopolovic J *et al*. Prognostic value of Bmi-1 oncoprotein expression in NSCLC patients: a tissue microarray study. J Cancer Res Clin Oncol 2008;134:1037–42.

[14] Robey RW, Polgar O, Deeken J, To KW, Bates SE. ABCG2: determining its relevance in clinical drug resistance. Cancer Metastasis Rev 2007;26:39–57.

[15] Patel M, Lu L, Zander DS, Sreerama L, Coco D, Moreb JS. ALDH1A1 and ALDH3A1 expression in lung cancers: correlation with histologic type and potential precursors. Lung Cancer 2008;59:340–9.

[16] Sullivan JP, Spinola M, Dodge M, Raso MG, Behrens C, Gao B *et al*. Aldehyde dehydrogenase activity selects for lung adenocarcinoma stem cells dependent on notch signaling. Cancer Res 2010;70:9937–48.

[17] Singh S, Dirks PB. Brain tumor stem cells: identification and concepts. Neurosurg Clin N Am 2007;18:31–8, viii.

[18] Xu YH, Zhang GB, Wang JM, Hu HC. B7-H3 and CD133 expression in non-small cell lung cancer and correlation with clinicopathologic factors and prognosis. Saudi Med J 2010;31:980–6.

[19] Salnikov AV, Gladkich J, Moldenhauer G, Volm M, Mattern J, Herr I. CD133 is indicative for a resistance phenotype but does not represent a prognostic marker for survival of non-small cell lung cancer patients. Int J Cancer 2010;126:950–8.

[20] Woo T, Okudela K, Mitsui H, Yazawa T, Ogawa N, Tajiri M *et al*. Prognostic value of CD133 expression in stage I lung adenocarcinomas. Int J Clin Exp Pathol 2010;4:32–42.

[21] Karoubi G, Cortes-Dericks L, Gugger M, Galetta D, Spaggiari L, Schmid R. Atypical expression and distribution of embryonic stem cell marker, *OCT4*, in human lung adenocarcinoma. J Surg Oncol 2010;102:689–98.

[22] Kim RJ, Nam JS. OCT4 expression enhances features of cancer stem cells in a mouse model of breast cancer. Lab Anim Res 2011;27:147–52.

[23] Zhang XY, Dong QG, Huang JS, Huang AM, Shi CL, Jin B *et al*. The expression of stem cell-related indicators as a prognostic factor in human lung adenocarcinoma. J Surg Oncol 2010;102:856–62.

[24] Chen YC, Hsu HS, Chen YW, Tsai TH, How CK, Wang CY *et al*. Oct-4 expression maintained cancer stem-like properties in lung cancer-derived CD133-positive cells. PLoS One 2008;3:e2637.

[25] Lagiou P, Trichopoulos D. Inflammatory biomarkers and risk of lung cancer. J Natl Cancer Inst 2011;103:1073–5.

APPENDIX. CONFERENCE DISCUSSION

Dr A. Zuin *(Padua, Italy)*: The authors examined a very interesting and intriguing field of the prognostic impact of cancer stem cells in lung adenocarcinoma. I have two questions for you. You talk about disease-free survival in your paper, but did you correlate the expression of these biomarkers, these features, with long-term survival? This is the first question. Second, what could be the possible clinical implications of the expression of these biological features?

Dr Galetta: We began this study together with the University of Berne about two and a half years ago. To answer your first question, the time for the follow-up is very, very short to reach any significance in the evaluation in regards to long-term survival. We think that in the near future we will be able with mature results to answer your interesting question on the correlation between these biological markers and survival in these patients with lung adenocarcinoma.

With regard to your second question, regarding the clinical implication of these factors, as I showed in my presentation, we noticed that they were correlated in a significant way with the tumour size, the T, and the N factors. So we think that these main factors will be predictive of survival, recurrence and so on.

Dr G. Friedel *(Gerlingen, Germany)*: I have one question. Have you measured these stem cell markers also in patients without lung cancer?

Dr Galetta: No, only in patients with lung cancer, adenocarcinoma specifically. The study was conducted with two different kinds of tissue, on lung adenocarcinoma and on normal lung tissue of the same patients.

Dr K. Nowak *(Mannheim, Germany)*: If you were measuring stem cell markers in lung cancer, did you perform it on several biopsies in each tumour. The variation in the expression of stem cell markers in lung cancer might vary a lot between the vital tumour rim and the centre of the tumour. This has been shown in a lot of experimental studies.

Dr Galetta: No. Usually I myself performed the biopsies of the lung tumours, and I usually performed the same kind of biopsy, a little piece of tumour from the central part to the periphery of the nodule. We didn't perform an accurate analysis between the central part expressions or the peripheral part, because most of these cases were initial tumours, T1, so we didn't have much tissue. The pathologist requires a lot of tissue to perform other studies.

Dr Nowak: You have shown in stage III cancers that you had significantly lower expressions of CD133. What percentage of the patients had neoadjuvant therapy?

Dr Galetta: What group of patients?

Dr Nowak: The advanced cancer group had lower expression profiles for CD133.

Dr Galetta: Yes. It's about a quarter of patients.

Dr Nowak: And it's possible that after neoadjuvant therapy, these markers be affected.

Dr Galetta: We didn't evaluate the difference between the neoadjuvant and the adjuvant patients.

Dr R. Milton *(Leeds, UK)*: It was a great deal of work that you presented in a few slides really. I'm quite excited at the prospect that there was a correlation between the expression of some of those markers and the pathological staging, presumably based on the tumour and the nodal involvement. Can you detect any of these markers in the blood?

Dr Galetta: To-date, we didn't perform this evaluation in this study, but I think in the future we will be able to have sufficient kits to perform this evaluation also. I think that in the future this will be possible.

Dr Milton: Presumably that potentially would indicate those who may have metastatic disease.

Dr Galetta: Yes.