· Advances in Medical Sciences · Vol. 58(2) · 2013 · pp 196-206 · DOI: 10.2478/ams-2013-0025

Molecular profiles of non-small cell lung cancers in cigarette smoking and never-smoking patients

Szymanowska-Narloch A[1], Jassem E[1], Skrzypski M[2], Muley T[3], Meister M[3], Dienemann H[3], Taron M[4], Rosell R[4], Rzepko R[5], Jarząb M[6], Marjański T[7], Pawłowski R[8], Rzyman W[7], Jassem J[2,*]

1 Department of Allergology, Medical University of Gdansk, Gdansk, Poland
2 Department of Oncology and Radiotherapy, Medical University of Gdansk, Gdansk, Poland
3 Thoraxklinik, University Hospital Heidelberg, Translational Research Center Heidelberg, Heidelberg, Germany
4 Catalan Institute of Oncology, Hospital Germans Trias i Pujol, Barcelona, Spain
5 Department of Pathology, Medical University of Gdansk, Gdansk, Poland
6 Radiotherapy and Chemotherapy Clinic, Maria Skłodowska-Curie Memorial Cancer Center and Institute of Oncology, Gliwice Branch, Gliwice, Poland
7 Department of Thoracic Surgery, Medical University of Gdansk, Gdansk, Poland
8 Department of Forensic Medicine, Medical University of Gdansk, Gdansk, Poland

* CORRESPONDING AUTHOR:
Department of Oncology and Radiotherapy
Medical University of Gdansk,
Dębinki 7,
80-211 Gdansk,Poland
Tel.: +4858 349 2270
Fax: +4858 349 2210
E-mail: jjassem@gumed.edu.pl (Jacek Jassem)

Received: 29.05.2013
Accepted: 19.06.2013
Advances in Medical Sciences
Vol. 58(2) 2013 · pp 196-206
DOI: 10.2478/ams-2013-0025

ABSTRACT

Purpose: Molecular features of non-small cell lung cancer (NSCLC) in never-smokers are not well recognized. We assessed the expression of genes potentially related to lung cancer etiology in smoking vs. never-smoking NSCLC patients.

Methods: We assayed frozen tumor samples from surgically resected 31 never-smoking and 54 clinically pair-matched smoking NSCLC patients, and from corresponding normal lung tissue from 27 and 43 patients, respectively. Expression of 21 genes, including cell membrane kinases, sex hormone receptors, transcription factors, growth factors and others was assessed by reverse transcription – quantitative PCR.

Results: Expression of 5 genes was significantly higher in tumors of non-smokers vs. smokers: *CSF*1*R* (p<0.0001), *RRAD* (p<0.0001), *PR* (p=0.0004), *TGFBR*2 (p=0.0027) and *EPHB*6 (p=0.0033). Expression of *AKR*1*B*10 (p<0.0001), *CDKN*2*A* (p<0.0001), *CHRNA*6 (p<0.0001), *SOX*9 (p<0.0001), *survivin* (p<0.0001) and *ER*2 (p=0.002) was significantly higher in tumors compared to normal lung tissue. Expression of *AR* (p<0.0001), *EPHB*6 (p<0.0001), *PR* (p<0.0001), *TGFBR*2 (p<0.0001), *TGFBR*3 (p<0.0001), *ER*1 (p=0.0006) and *DLG*1 (p=0.0016) was significantly lower in tumors than in normal lung tissue. Expression of *IGF*2 was higher in tumors than in healthy lung tissue in never-smokers (p=0.003), and expression of *AHR* (p<0.0001), *CSF*1*R* (p<0.0001) and *RRAD* (p<0.0001) was lower in tumors than in healthy lung tissue in smokers.

Conclusion: Expression of several genes in NSCLC is strongly related to smoking history. Lower expression of *PR* and higher expression of *ER*2 in tumors suggests a possibility of hormonal therapeutic intervention in selected NSCLC patients. Distinct molecular features of NSCLC in never-smokers, e.g. *CHRNA*6 upregulation, may prompt new treatment strategies.

Key words: Non-small cell lung cancer, gene expression, smoking

INTRODUCTION

Lung cancer is the leading cause of cancer-related mortality [1]. Tobacco smoke is a well-recognized cause of non-small cell lung cancer (NSCLC), however about 15% of men and 50% of women diagnosed with lung cancer worldwide have never smoked [2]. The tumors of non-smokers are predominantly located in the peripheral lung areas and are more likely to originate from bronchoalveolar stem cells [3].

The never-smokers with lung cancer are more frequently women than men.

NSCLCs of non-smokers are also distinct in respect to molecular pathology, e.g. they are more likely to harbor *EGFR* activating mutations [4] and a fusion oncogene *EML*4-*ALK* [5]. There are also differences in gene expression in NSCLCs and in normal lung tissue of smokers and never-smokers [6,7]. Further molecular characterization of NSCLC in never-smoking patients might aid the development of new therapeutic strategies.

In this study, we assessed the expression of selected genes potentially driving tumorigenesis and biology of NSCLC in tumors and normal lung tissue of never-smoking patients, using reverse transcription – quantitative polymerase chain reaction (RT-qPCR). The genes were selected based on a thorough current literature search. Hormone receptor genes were investigated, as the use of hormone replacement therapy (HRT) in women was shown to be associated with an earlier onset of NSCLC and shorter overall survival (OS) [8]. Another study indicated that combined analysis of estrogen receptor beta (ERbeta1) and progesterone receptor (PR) expression identified lung cancer patients with poor outcome [9]. Thus, the sex hormones emerge not only as factors triggering lung oncogenesis but they may also impact the course of the disease. Lung is also a target tissue for androgen receptor (AR) which is expressed in different lung cancer types [10]. Given the above, we hypothesized that the sex hormones background might elicit the NSCLC in never-smokers. To verify such correlation, we analyzed the expression of major sex hormone receptors (*ER*1*, ER*2*, PR* and *AR*). Carcinogens in tobacco smoke were found to interact with estrogen and androgen receptors and to affect hormonal proliferative signaling balance [11]. Previous studies demonstrated that aryl hydrocarbon receptor (AHR) is overexpressed in lung adenocarcinomas [12]. Therefore, *AHR*, a transcription factor that is activated by cigarette smoke was included into the analysis. *AKR*1*B*10*, CHRNA*6, *CHRNB*3 and *CDKN*2*A/P*16 genes were included, as their expression is affected by the tobacco smoke [13-17]. *AKR*1*B*10 expression is upregulated by smoking in airway epithelia in healthy smokers and this event might constitute an early step in lung carcinogenesis [13]. In a previous study by Lam et al. [14], the expression of *CHRNA*6 and *CHRNB*3 was higher in lung cancer cells from non-smokers compared to smokers. *CDKN*2*A*/*P*16 is a tumor suppressor gene which is inactivated in many tumors including lung cancer [16]. On the other hand, Zhao et al. [17] demonstrated that in one third of immunohistochemistry P16-positive NSCLC cases, expression of this protein was caused by increased gene copy number of *CDKN*2*A*. This diverse and apparently molecular context-dependent expression of P16 in NSCLC prompted the inclusion of this transcript into the pool of analyzed genes. Crizotinib, *EML*4-*ALK* fusion gene inhibitor, was recently shown to be efficacious in metastatic NSCLC harboring this abnormality, which is typical for never-smokers [18]. Oncogenic potential of ALK kinase may be acquired through gene fusion and possibly by gene amplification [19]. Thus, the transcript for *ALK* (covering all potential variants of oncogenic fusion transcripts containing *ALK*) [20] was assessed, together with 3 other factors of insulin receptor superfamily (*IGF*1*, IGF*1*R* and *IGF*2), potentially relevant to biology of NSCLC of non-smoking patients [21-23]. Finally, a number of genes (*survivin*, *CSF*1*R, DLG*1*, EPHB*6*, RRAD, SOX*9*, TGFBR*2*, TGFBR*3*)* were included in the analysis to validate previous findings [7, 24-30].

Due to a large size of the full article its Supplementary Files: Figures 1S – 8S (Appendix A) and Tables 1S–3S (Appendix B), are available only online at: http://dx.doi.org/10.2478/ams-2013-0025.

MATERIALS AND METHODS

Study subjects

This study was approved by the institutional review boards of all participating institutions. All patients underwent between 2002 and 2008 pulmonary resection with curative intent in two institutions (Department of Thoracic Surgery, Medical University of Gdansk, Poland, and Department of Surgery, Thoraxklinik at University Hospital Heidelberg, Germany), a member of the accredited National Center of Tumor Diseases tissue bank. Patients for whom snap-frozen lung tumor was available were enrolled in the study; for most of these patients also corresponding healthy lung tissue samples were available. None of the patients was administered preoperative or postoperative chemotherapy or radiotherapy (*Tab. 1*). Tumor staging was performed using AJCC/UICC classification from 1997. By the time of this analysis (February 2012), 42 out of 85 patients deceased.

The non-smoking group included 31 patients with pathologically confirmed NSCLC. Patients were defined as never-smokers if they declared smoking fewer than 100 cigarettes in their life-time. Each never-smoker was pair-matched to 1 or 2 smoking patients with similar clinical characteristics (age, sex, histological tumor type, pTNM). The reference group included 54 NSCLC patients with smoking history. Most of these patients had a history of more than 20 pack-years, with mean pack-years of 47.9.

Methods

Using RT-qPCR we compared expression of 21 genes (*Tab. 2*) in NSCLCs of never-smokers vs. smokers, and in tumors vs. corresponding normal lung tissue. These included genes associated with smoking (e.g. *RRAD, CHRNA*6*, CHRNB*3*, AHR, P*16/*CDKN*2*A*), kinase receptors (*CSF*1*R, EPHB*6, *ALK*), hormonal receptors (e.g. *PR, AR, ER*1-2), growth factor receptors (e.g. *TGFBR*2*, TGFBR*3*, IGF*1*R*), apoptosis inhibitors (*survivin*) and others. Normalization genes

Table 1. **Patient characteristics.**

Variable	Never-smokers (n=31)	Smokers (n=54)
Age (mean/range)	64 (37-81 years)	59 (40-78 years)
Origin		
Poland	13 (42%)	54 (100%)
Germany	18 (58%)	-
Sex		
Male	5 (16%)	19 (35%)
Female	26 (84%)	35 (65%)
Stage		
IA + IB	20 (65%)	34 (64%)
IIA + IIB	2 (6%)	10 (18%)
IIIA + IIIB	9 (29%)	10 (18%)
Histology		
Adenocarcinoma	24 (77%)	45 (83%)
Squamous cell carcinoma	4 (13%)	7 (13%)
Other	3 (10%)	2 (4%)
Grade		
1	3 (10%)	9 (17%)
2	17 (55%)	33 (61%)
3	11 (35%)	12 (22%)

included 18*S rRNA subunit, RNA polymerase II* (*POLR*2*A*) and *esterase D* (*ESD*).

Analysis was performed on snap frozen tumor samples from 31 never-smoking and 54 smoking patients and on corresponding frozen healthy lung tissue samples from 27 and 43 patients, respectively. All samples (blocks of around 1 cm^3) were obtained during pulmonary resection, snap frozen in liquid nitrogen and stored at -80°C until the time of analysis.

The RNA was isolated using commercially available kit (AllPrep kit, Applied Biosystems) only from tumor areas containing at least 60% of non-necrotic tumor cells in the microscopic section. The RNA concentration was assessed in Nano-Drop, and the quality of obtained RNA was confirmed on agarose gel. The workflow of sample processing for samples obtained from the Thoraxklinik at University Hospital Heidelberg, Germany, was reported previously [31].

Reverse transcription was combined with the polymerase chain reaction rewriting RNA to cDNA (High Capacity cDNA Archive Kit, Applied Biosystems) with the use of hexameric primers. One µg of RNA was transcribed into cDNA. Quantitative analysis of mRNA of selected genes by RT-qPCR was performed with the use of microfluidic cards (MFC), TaqMan probes, polymerase with 5'-exonuclease activity and specific primers. Analysis with TaqMan Low Density Arrays was performed using ABI PRISM 7900 HT Sequence Detection System (Applied Biosystems). Each channel of a microfluidic card was loaded with a 100 µl mix of equal volume of Taqman Universal PCR Master Mix (Applied Biosystems) and cDNA template (corresponding to 100 ng of total RNA).

Statistical analysis

RT-qPCR results were obtained using the Sequence Detection System 2.1 Software. Gene expression was calculated with $2^{-\Delta\Delta C}{}_T$ method [32], relative to the pool of RNA containing equal amounts of each study sample. Statistical analysis was performed with the use of statistical package SPSS 17.0.

Gene expression in both study groups was compared by Wilcoxon test (paired comparisons between tumor and normal tissue) and Mann-Whitney/Kruskal-Wallis tests (testing differences between independent groups). Chi-square test was used for comparison of proportions. Microarray data were analyzed by routinely applied Student's t-test. In the case of *ALK*, a threshold for "overexpression" was adopted. The most common activating translocation for *ALK* in NSCLC – *EML*4-*ALK* – is detected in 1.6–9% of surgical series and amplification of *ALK* was found in 10% of cases [19]. Therefore, we adopted the 90th percentile expression level for the entire cohort of analyzed NSCLC samples (n=85) as the overexpression threshold, so that samples with the expression from the highest 10 percentiles (n=9) were considered to show "overexpression". For a gene expressed only in a fraction of samples, i.e. *AKR*1*B*10*,* the positivity threshold was defined as expression above the expression in calibrator sample (pool of RNA).

Metastasis-free survival was calculated with Kaplan-Meier method from pulmonary surgery to the diagnosis of any metastasis, or censored due to end of follow up or death. Survival curves were compared using log-rank test. Owing to the multiple comparisons, a conservative significance threshold of p=0.005 was applied for all statistical analyses except for survival comparisons.

RESULTS

Tumor gene expression in smokers vs. never-smokers

Tumor expression of 5 genes was significantly higher in never-smokers compared to smokers: *CSF*1*R* (p<0.0001), *RRAD* (p<0.0001), *PR* (p=0.0004), *TGFBR*2 (p=0.0027) and *EPHB*6 (p=0.0033), whereas the expression of *AR* was at the borderline level (p=0.008) (*Tab. 3*; *Fig. 1*). The expression of *RRAD* (p=0.0003), *CSF*1*R* (p=0.0008) and *PR* (p=0.0009) in adenocarcinomas of never-smoking females was higher compared to their smoking counterparts. The expression of *CHRNA*6 tended to be higher (p=0.023) and the expression of *survivin* lower (p=0.007) in adenocarcinomas of never-smoking compared to smoking women.

Among 9 patients with overexpression of *ALK* transcript, 4 were never-smokers, 4 were former-smokers and 1 was an active smoker. This group included 7 females and 2 males;

Table 2. **Brief characteristics of the analyzed genes.**

#	Gene symbol	Assay-on-Demand ID used for analysis	Full gene name	Gene product function	References
1	*AHR*	*AHR*-Hs00169233_m1	aryl hydrocarbon receptor	cytosolic transcription factor	[12, 40]
2	*AKR1B10*	*AKR1B10*-Hs00252524_m1	aldo-keto reductase family 1, member B10 (aldose reductase)	enzyme reducing aliphatic and aromatic aldehydes	[13, 41, 42]
3	*ALK*	*ALK*-Hs00608289_m1	anaplastic lymphoma kinase	receptor tyrosine kinase; can form fusion genes	[5, 18-20, 48]
4	*AR*	*AR*-Hs00171172_m1	androgen receptor	a DNA-binding transcription factor regulating gene expression	[10, 33, 35]
5	*BIRC5*	*BIRC5*-Hs00978503_m1	baculoviral inhibitor of apoptosis repeat-containing 5, survivin	inhibit caspase activation, thereby leading to negative regulation of apoptosis or programmed cell death	[24, 46]
6	*CDKN2A*	*CDKN2A*-Hs00233365_m1	cyclin-dependent kinase inhibitor 2A (melanoma, p16, inhibits CDK4)	tumor suppressor protein, cell cycle regulation	[16, 17]
7	*CHRNA6*	*CHRNA6*-Hs00610231_m1	Cholinergic receptor, nicotinic, alpha 6	regulates cellular proliferation and apoptosis via Akt pathway	[14, 15]
8	*CHRNB3*	*CHRNB3*-Hs00181269_m1	Neuronal acetylcholine receptor subunit beta-3	regulates cellular proliferation and apoptosis via Akt pathway	[14, 15]
9	*CSF1R*	*CSF1R*-Hs00911250_m1	Colony stimulating factor 1 receptor; cell-surface protein	cytokine which controls the production, differentiation, and function of macrophages	[25, 26, 44]
10	*DLG1*	*DLG1*-Hs00938204_m1	Discs, large homolog 1 (Drosophila)	involved in the regulation of cell polarity and proliferation control	[27]
11	*EPHB6*	*EPHB6*-Hs00270052_m1	ephrin type-B receptor 6	receptor for transmembrane protein	[26, 43]
12	*ER1*	*ER1*-Hs00174860_m1	estrogen receptor 1	member of the nuclear hormone family of intracellular receptors; DNA-binding transcription factor that regulates gene expression	[11, 33, 36, 37]
13	*ER2*	*ER2*-Hs00230957_m1	estrogen receptor 2	member of the nuclear hormone family of intracellular receptors; DNA-binding transcription factor that regulates gene expression	[11, 33, 36, 37]
14	*IGF1*	*IGF1*-Hs01547657_m1	Insulin-like growth factor 1 tyrosine kinase	polypeptide protein hormone similar in molecular structure to insulin	[21]
15	*IGF1R*	*IGF1R*-Hs00609566_m1	insulin-like growth factor 1 receptor	transmembrane receptor that is activated by IGF-1 and by the related growth factor IGF-2	[22]
16	*IGF2*	*IGF2*-Hs01005963_m1	Insulin-like growth factor 2 tyrosine kinase	protein hormone sharing structural similarity to insulin	[23]
17	*PR*	*PR*-Hs00172183_m1	progesterone receptor	intracellular steroid receptor specifically binding progesterone	[9, 33, 34]
18	*RRAD*	*RRAD*-Hs00188163_m1	Ras-related associated with diabetes	GTP-binding protein RAD	[28, 29]
19	*SOX9*	*SOX9*-Hs00165814_m1	SRY (sex determining region Y)-box 9 (campomelic dysplasia, autosomal sex-reversal)	important role in respiratory epithelium development; has an impact on AR protein expression	[30, 47]
20	*TGFBR2*	*TGFBR2*-Hs00234253_m1	transforming growth factor, beta receptor II	polypeptide growth factor	[7]
21	*TGFBR3*	*TGFBR3*-Hs00234257_m1	transforming growth factor, beta receptor III	polypeptide growth factor	[7]

Table 3. **Differences in expression of analyzed genes between smokers and non-smokers. Significant differences are marked in bold letters.**

Gene	Expression in tumors of never-smokers and smokers Fold change (median expression) p value	Expression in tumors and normal lungs of never-smokers Fold change (median expression) p value	Expression in tumors and normal lungs of smokers Fold change (median expression) p value
AHR	1.16 (0.89/0.77); 0.0078	1.09 (0.89/0.82); 0.5	**0.86** (0.773/0.897); <0.0001
AKR1B10	0.23 (1.43/6.14); 0.9091	**23.8** (1.43/0.06); <0.0001	**458.32** (6.1415/0.0134); <0.0001
ALK	1.12 (2.61/2.34); 0.3541	3.39 (2.61/0.77); 0.105	3.82 (2.33/0.61); 0.601
AR	1.57 (0.74/0.47); 0.0078	**0.53** (0.74/1.39); <0.0001	**0.38** (0.47/1.25); <0.0001
BIRC5	0.45 (0.91/2.03); 0.0091	**5.35** (0.91/0.17); <0.0001	**12.0** (2.03/0.17); <0.0001
CDKN2A	1.17 (2.60/2.22); 0.9454	**18.57** (2.60/0.14); 0.001	**11.94** (2.22/0.19); <0.0001
CHRNA6	1.22 (2.81/2.30); 0.0346	**6.39** (2.81/0.44); <0.0001	**3.24** (2.30/0.71); <0.0001
CHRNB3	2.35 (2.19/0.93); 0.0990	3.59 (2.19/0.61); 0.36	0.45 (0.93/2.06); 0.816
CSF1R	**1.89** (0.72/0.38); <0.0001	1.29 (0.72/0.56); 0.181	**0.63** (0.38/0.60); <0.0001
DLG1	1.04 (0.76/0.73); 0.5080	0.86 (0.76/0.88); 0.028	0.85 (0.73/0.86); 0.014
EPHB6	**2.06** (0.99/0.48); 0.0033	**0.63** (0.99/1.58); <0.0001	**0.31** (0.48/1.55); <0.0001
ER1	0.95 (1.30/1.37); 0.7459	1.38 (1.30/0.94); 0.039	1.47 (1.37/0.93); 0.007
ER2	1.07 (1.62/1.52); 0.8230	1.69 (1.62/0.96); 0.029	1.6 (1.52/0.95); 0.027
IGF1	1.42 (1.38/0.97); 0.9600	1.77 (1.38/0.78); 0.914	1.11 (0.97/0.87); 0.97
IGF1R	0.80 (0.61/0.76); 0.1153	0.82 (0.61/0.74); 0.031	1.04 (0.76/0.73); 0.509
IGF2	0.79 (0.27/0.34); 0.7321	**1.5** (0.27/0.18); 0.003	2.19 (0.339/0.155); 0.062
PR	**2.11** (0.40/0.19); 0.0004	**0.39** (0.40/1.02); <0.0001	**0.18** (0.19/1.03); <0.0001
RRAD	**2.75** (0.55/0.20); <0.0001	0.65 (0.55/0.85); 0.025	**0.15** (0.20/1.30); <0.0001
SOX9	1.08 (1.28/1.18); 0.3933	**3.05** (1.28/0.42); <0.0001	**2.62** (1.18/0.45); 0.002
TGFBR2	**1.52** (0.41/0.27); 0.0027	**0.42** (0.41/0.97); <0.0001	**0.28** (0.27/0.97); <0.0001
TGFBR3	1.83 (0.22/0.12); 0.0275	**0.19** (0.22/1.17); <0.0001	**0.13** (0.12/0.92); <0.0001

Figures to Table 3 available from authors by request.

8 adenocarcinomas and 1 squamous cell carcinoma. Four of these patients developed distant metastases including 2 brain metastases. Overexpression of *ALK* was also found in 7 normal lung specimens. Interestingly, *ALK* overexpression coincided in lung and tumor tissue in only one case. If 75th percentile of expression was applied as a threshold for overexpression, *ALK* was found in 8 normal lung samples, with no expression in the corresponding tumor sample; 7 cases exhibited overexpression in both tumor and the corresponding normal lung and in 14 cases *ALK* was overexpressed in NSCLC and not in the corresponding normal lung.

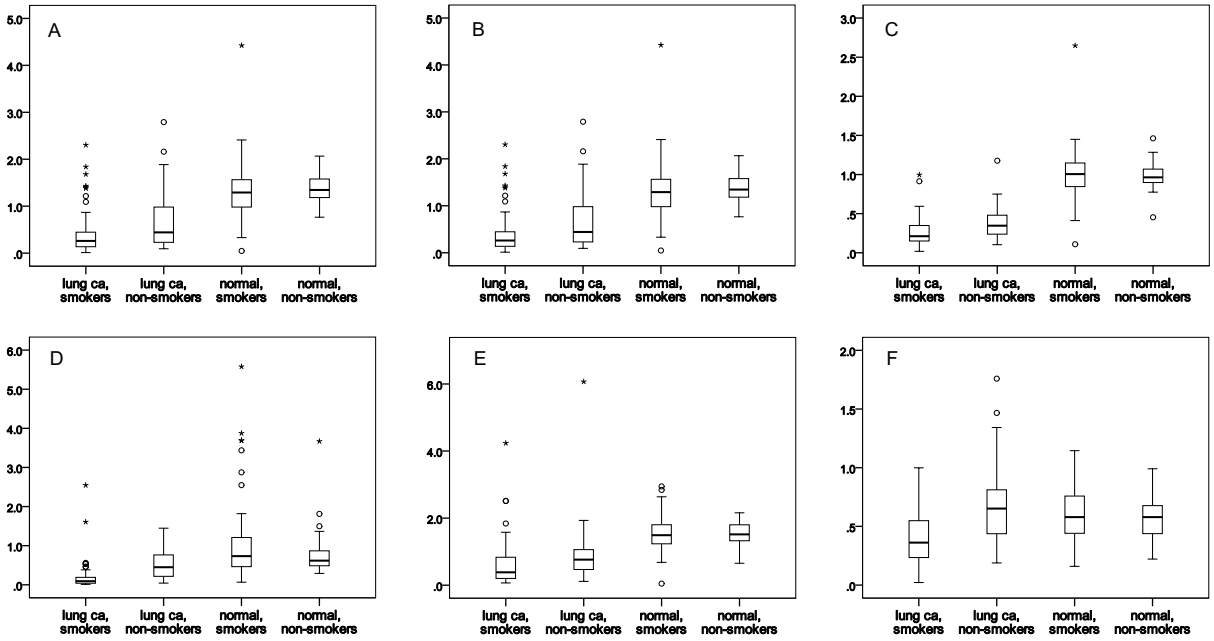
AKR1B10 was expressed in 23 NSCLCs, in a similar proportion of smokers (31%) and never-smokers (19%; p=0.31).

Gene expression in tumor samples and in corresponding normal lung tissue in smokers vs. never-smokers

Expression of 7 genes was significantly lower in tumors than in normal lung tissue: *AR* (p<0.0001), *EPHB6* (p<0.0001), *PR* (p<0.0001), *TGFBR2* (p<0.0001), *TGFBR3* (p<0.0001), *ER1* (p=0.0006) and *DLG1* (p=0.0016). When the gene expression in tumor samples and corresponding normal lung tissue was analyzed separately for smokers and never-smokers, there was no statistically significant difference in gene expression for *ER1* (p=0.007 and p=0.039, respectively) and *DLG1* (p=0.014 and p=0.028, respectively). Tumor expression of three genes: *AHR* (p<0.0001), *CSF1R* (p<0.0001) and *RRAD* (p<0.0001) was lower than in healthy lung tissue in smokers but not in never-smokers.

The expression of 6 genes was significantly higher in tumors compared to normal lung tissue: *AKR1B10* (p<0.0001*), CDKN2A* (p<0.0001), *CHRNA6* (6-fold in never-smokers and 3-fold in smokers; p<0.0001), *SOX9* (p<0.0001), *survivin* (p<0.0001) and *ER2* (p=0.002). Mean expression of *IGF2* was higher in tumor compared to healthy lung tissue in never-smokers (p=0.003) but not in smokers. The fold change in the expression level of *ER2* in tumor samples and in corresponding normal lung tissue both in smokers and never-smokers was above the established threshold of p=0.005 (p=0.027 and p=0.029, respectively).

Figure 1. **Expression of genes in tumors and in normal lungs of smokers and never-smokers: A) progesterone receptor – *PR* (p=0.0004); B) androgen receptor – *AR* (p=0.008); C) transforming growth factor beta receptor 2 – *TGFBR2* (p=0.0027); D) Ras-related associated with diabetes – *RRAD* (p<0.0001); E) ephrin type-B receptor 6 - *EPHB6* (p=0.0033); F) colony stimulating factor 1 receptor - *CSF1R* (p<0.0001).**



Gene expression in normal lung of smokers vs. never-smokers

Irrespective of smoking status, no expression of *AKR1B10* was found in normal lung tissue. All other genes were expressed at the same levels in normal lungs of smokers and never-smokers. In the analysis confined to women, expression of *CDKN2A* was higher (p=0.0013) in normal lung of smoking patients**.** Intriguingly, although not significant, an increased expression of lung *RRAD* was found in 20% of smoking women. The expression of *SOX9* was slightly higher (p=0.06) in lungs of smoking women, whereas the expression of *TGFBR3* tended to be lower in normal lung of smoking females (p=0.0051).

Prognostic impact of gene expression

There was no difference in metastasis-free survival between smokers and never-smokers (p=0.74). *ER2* (p=0.01) and *AKR1B10* (p=0.013) were significantly correlated with metastasis-free survival in the entire cohort. However, none of the analyzed genes was correlated with metastasis-free survival after correction for multiple comparisons.

DISCUSSION

This study showed distinct tumor expression of several genes potentially involved in NSCLC carcinogenesis in never-smoking compared to smoking patients, whereas none of the analyzed genes showed different expression in normal lung of smokers compared to never-smokers. This suggests that the differences in tumor expression between smokers and never-smokers reflect activation of smoking-related pathways in NSCLC patients rather than a mere impact of cigarette smoke insult on gene expression in pulmonary epithelium. The same holds true in the analysis restricted to women, which provides a strong suggestion that expression of sex hormone receptors in the normal lung is not affected by smoking status.

Our data suggest that aberrant sex hormone receptor signaling may be among the causes of NSCLC in both smokers and never-smokers. As in the study of Marquez-Garban [33], expression of *PR* and *AR* was lower in NSCLC than in normal lung. As progesterone was shown to inhibit the growth of NSCLC cell lines with expression of *PR* in vitro and in vivo [34], low expression of *PR* in tumors may be viewed as a means of averting the antiproliferative effect of progesterone by NSCLC. Similarly, androgens (*e.g.* fluoxymesterone) were demonstrated to inhibit the growth of breast cancer through the activation of AR and consequent induction of apoptosis [35]. It may be speculated therefore that the administration of PR or AR agonists might provide clinical benefit in both smoking and non-smoking NSCLC patients [34].

Importantly, the tumor expression of *PR* was significantly lower in smokers than in never-smokers both in all patients and in women. A similar trend was noted for *AR*. This finding might suggest that biology of NSCLC in smokers involves a higher degree of intolerance to inhibitory effects of progesterone. It was previously shown that carcinogens from cigarette smoke (BaP, benzo[a]pyrene) may disrupt protective

(anti-cancer) androgen function by reducing AR levels [35]. This might further explain the relative low expression of *AR*, and possibly also *PR* in smokers. Further, NSCLC expressed *ER*2 (beta) at higher levels and *ER*1 (alpha) at lower levels than the normal lung. These results are in accordance with previous report [36], where *ER*2 (beta) was found to be expressed in NSCLC cell lines, whose proliferation was shown to be inhibited by fulvestrant (an exclusive ER antagonist). Moreover, fulvestrant inhibited the growth of lung tumor xenograft by about 40% in immunodeficient mice [37]. A potential carcinogenic effect of estrogens was suggested after two trials had demonstrated increased risk of lung cancer in women administered HRT [38].

In previous study *RRAD* expression was found to be downregulated in tumors compared to normal lung. Interestingly, in our study, *RRAD* expression compared to normal lung was clearly downregulated only in tumors of smokers. This finding suggests that *RRAD* is downregulated specifically in cancerogenesis induced by tobacco smoke. Moreover, this gene is not expected to be involved in the pulmonary epithelium response to tobacco insult, as its expression was similar in normal lung of smokers and never-smokers. This gene is a negative regulator of glucose uptake. It may be therefore envisaged that in NSCLCs which are particularly dependent on glucose metabolism this gene is downregulated as part of the molecular background of Warburg effect. The question of whether NSCLCs of never-smokers are less dependent on glucose metabolism remains to be answered. Some data suggest that this gene is silenced in smokers through promoter hypermethylation [28]. This kind of epigenetic silencing may be therefore considered a potential target for treatment with a demethylating agent 5-aza-2'-deoxycitidine.

In contrast to previously described more frequent hypermethylation of *P*16 in tumors of smokers [39], we did not demonstrate apparent differences in expression of this gene between smokers and never-smokers. In non-expressing cases (equally common in smokers and never-smokers) no transcript was detected whatsoever. This implies that both alleles of *P*16 were functionally inactivated, possibly through homozygous deletion or LOH and hypermethylation, or other epigenetic mechanisms [16]. Detectable *P*16 expression in healthy lung tissue of some smoking females may be viewed as a response to genotoxic insult of tobacco smoke.

In our material, the nicotine receptor subunit *CHRNA*6 expression was higher in tumors compared to healthy lung tissue. Along the same lines as in the study of Lam et al. [14], NSCLC of never-smokers seems to be highly dependent on the expression of this gene (6-fold higher expression vs. normal lung). NSCLC of smokers is also dependent on this subunit, albeit to a lesser extent (3-fold higher expression vs. normal lung). Our results suggest that this receptor subunit may be a potential target of future NSCLC therapies in never-smokers.

AHR is a transcriptional factor involved in detoxification responses to pollutants and in intrinsic biological processes of multicellular organisms. AHR, upon activation by benzopyrenes (BPs), translocates into the nucleus and induces expression of several genes, including *CYP*1*A*1, that further activate BPs' carcinogenic potential. It was also previously shown that AHR can shuttle polycyclic aromatic hydrocarbons to the nucleus and enhance oxidative DNA damage [44]. Relatively lower expression of *AHR* in tumors of smokers may suggest their intolerance to undue genetic insult. *AKR*1*B*10 was expressed in a proportion of tumors of both smokers and never-smokers. Reduction of benzopyrenes by AKR1B10 in lung cancer cells leads to ROS-mediated genotoxicity and contributes to lung carcinogenesis. The lack of expression of this enzyme in normal lung of smokers suggests that upregulation of *AKR*1*B*10 may represent a specific tumor-promoting aberration [41]. Experimental data suggest that inhibitors of AKR1B10 can enhance the chemopreventive effect of retinoids [42].

In our series, expression of *EPHB*6, a kinase receptor, was significantly lower in tumors compared to healthy lung tissue, and its tumor expression was significantly higher in non-smokers compared to smokers. This gene is frequently silenced by promoter hypermethylation [43] and its low expression was found to correlate with increased risk of metastasis in breast cancer. This may suggest a potential therapeutic role of demethylating agents in these patients.

We confirmed the previously reported downregulation of transforming growth factor receptors, *TGFBR*2 and *TGFBR*3, in transition from normal lung to NSCLC [7]. However, in our series low expression of these genes was also found in NSCLCs of smokers, possibly through hypermethylation, as an early event in carcinogenesis. Further, tumor expression of *TGFBR*2 was significantly lower in smokers compared to never-smokers. It may be envisaged that NSCLC carcinogenesis in both smokers and never-smokers is dependent on silencing of antigrowth effect of TGF, albeit to a different extent. The downregulation of *EPHB*6*, TGFBR*2 and *TGFBR*3 was confirmed in *in silico* validation of our results. In one of the datasets *TGFRB*3 gene was confirmed to be a very sensitive malignancy marker (a fold change of 0.18).

Tumor expression of *CSF*1*R*, a colony stimulating factor for macrophage receptor, in smokers was significantly lower compared to normal lung and to tumors of never-smokers. In cell cultures, CSF1R was shown to immortalize the epithelial cells through upregulation of telomerase [44]. Upregulation of *CSF*1 is correlated with poor prognosis in squamous cell and in other types of lung cancer [25]. In our study, tumor *CSF*1*R* expression was dependent on the smoking status. This finding requires further investigation, as the biological role of macrophages (growth promoting and inhibiting) is tumor dependent [45]. Agents affecting CSF1R activity could have potential therapeutic effect.

Tumor expression of *survivin* and *SOX*9, a transcription factor, was higher in tumors compared to healthy lung tissue, and this difference was more pronounced in smokers. The expression of *SOX*9 was insignificantly higher in lungs of smokers compared to lungs of never-smokers, suggesting its role in early carcinogenesis in smokers. Survivin inhibitors are currently being tested in clinical trials [46]. A relatively high expression of *SOX*9 in adenocarcinomas of smoking women as compared to their never-smoking counterparts might guide patient selection in future studies. In colorectal cancer, the down-regulation of SOX9 inhibited cell proliferation *in vitro* and the growth of tumor xenografts [47]. Therefore, *SOX*9 gene might also be a potential target for future therapies.

*IGF*2 expression was increased in tumors compared to healthy lung tissue in never-smoking NSCLC patients. This gene is commonly overexpressed in many human cancers, e.g. breast, colon or prostate cancer, and its higher expression has been attributed to a loss of imprinting [23]. Biological significance of these findings warrants further studies.

ALK overexpression was detected in NSCLCs of both smokers and never-smokers and was not tumor-specific. This is an interesting finding, suggesting that *ALK* overexpression may be among early events in lung cancerogenesis, however a question of whether the transcript found in normal lung of cancer patients is canonical or one of the fusion variants, remains to be established. Similarly to our findings, Martelli et al. [48] found expression of fusion gene *EML*4-*ALK* in 15% of histologically normal lung samples, with no corresponding tumoral expression of this fusion transcript in the same patient.

To validate the obtained results, the expression of relevant genes was examined in the publicly available gene expression microarray datasets. First, we evaluated genes expressed in pulmonary epithelium as a response to tobacco smoke insult. The potentially relevant dataset GSE23546 [49] included 1349 samples of non-tumorous human lung tissue, hybridized on Rosetta/Merck Human RSTA Custom Affymetrix 2.0 microarrays. However, GEO data repository did not include clear annotation of the smoking status for the individual cases. The data set GSE7895 [50] describes gene expression of 104 NSCLC patients: 21 never-smokers, 52 current smokers and 31 former smokers. This experiment was carried on Affymetrix HG U133a microarrays and covered expression of *CSF*1*R*, *RRAD*, *TGFBR*2, *EPHB*6, *AR*, and *CHRNA*6. Three genes (*RRAD*, *TGFBR*2 and *EPHB*6) were confirmed to be downregulated in current smokers (*Appendix A: Fig. 1S-3S; Appendix B: Tab. 1S*). Importantly, for *RRAD* gene both probes indicated the difference in expression with high significance (Student's t-test; $p<0.0005$). We also investigated the differences in gene expression in cancer vs. healthy lung tissue in two datasets published in GEO. The dataset GSE12428 [51] includes information on gene expression of 62 patients: 31 lung cancer and 31 non-tumorous tissues. Experiments were performed with Agilent two channel (Cy3/Cy5) Whole Human Genome Oligo Microarrays, design number 012391 (GPL1708). In this dataset we were able to evaluate 7 genes: *AR*, *EPHB*6, *PR* (*PGR*), *TGFBR*2 (*TGFBR*1), *TGFBR*3, *ER*1 (*ESR*1) and *DLG*1, three of which were confirmed to be downregulated in tumors: *PR* (p=1.78E-21), *TGFBR*2 (p=0.009) and *TGFBR*3 (p=4.70E-11) (*Appendix A: Fig. 4S; Appendix B: Tab. 2S*). Other genes did not show significant expression differences. Another database GSE10072 [52] includes gene expression information obtained from 107 samples: 58 lung adenocarcinomas and 49 normal lung tissues. Tissue mRNAs were hybridized to Affymetrix HG U133a microarrays, therefore for many genes of interest several probe sets were tested. Data confirmed *TGFRB*3 gene to be the most significantly downregulated in tumors vs. normal lung with a fold change of 0.18 (p=0.18E-34). The results for all other genes are included in the supplemental files, together with appropriate figures (*Appendix A: Fig. 5S-8S; Appendix B: Tab. 3S*).

CONCLUSIONS

In conclusion, we found significant differences in tumor expression of several genes between smoking and never-smoking NSCLC patients. Some of these differences (e.g. related *PR*, *CSF*1*R*) have not been previously reported, whereas others (for *RRAD*, *TGFBR*2, *EPHB*6) were confirmed. Downregulation of *PR* and upregulation of *ER*2 in tumors compared to normal lungs suggests possible hormonal dependence of NSCLC in both smokers and never-smokers. Thus, therapeutic potential of progesterone and estrogen inhibitors warrants further investigations. *AKR*1*B*10 is not expressed in lung parenchyma, which makes it a potentially highly specific, however moderately sensitive, NSCLC marker. *ALK* overexpression seems not to be lung cancer specific. Distinct molecular features of NSCLC of never-smokers (e.g. *CHRNA*6 upregulation) may prompt new treatment strategies.

ACKNOWLEDGEMENTS

This study was financially supported by National Science Centre, Poland, grant no. N402 238137.

Conflict of interest

The authors declare no conflict of interest.

REFERENCES

1. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. CA Cancer J Clin. 2011 Mar-Apr;61(2):69-90.
2. Wakelee HA, Chang ET, Gomez SL, Keegan TH, Feskanich D, Clarke CA, et al. Lung cancer incidence in never smokers. J Clin Oncol. 2007 Feb;25(5):472-8.
3. Kim CF. Paving the road for lung stem cell biology: bronchioalveolar stem cells and other putative distal lung stem cells. Am J Physiol Lung Cell Mol Physiol. 2007 Nov;293(5):L1092-8.
4. Rosell R, Moran T, Queralt C, Porta R, Cardenal F, Camps C, et al. Screening for epidermal growth factor receptor mutations in lung cancer. N Engl J Med. 2009 Sep;361(10):958-67.
5. Shaw AT, Yeap BY, Mino-Kenudson M, Digumarthy SR, Costa DB, Heist RS, et al. Clinical features and outcome of patients with non-small-cell lung cancer who harbor EML4-ALK. J Clin Oncol. 2009 Sep;27(26):4247-53.
6. Dutu T, Michiels S, Fouret P, Penault-Llorca F, Validire P, Benhamou S, et al. Differential expression of biomarkers in lung adenocarcinoma: a comparative study between smokers and never-smokers. Ann Oncol. 2005 Dec;16(12):1906-14.
7. Powell CA, Spira A, Derti A, DeLisi C, Liu G, Borczuk A, et al. Gene expression in lung adenocarcinomas of smokers and nonsmokers. Am J Respir Cell Mol Biol. 2003 Aug;29(2):157-62.
8. Ganti AK, Sahmoun AE, Panwalkar AW, Tendulkar KK, Potti A. Hormone replacement therapy is associated with decreased survival in women with lung cancer. J Clin Oncol. 2006 Jan;24(1):59-63.
9. Stabile LP, Dacic S, Land SR, Lenzner DE, Dhir R, Acquafondata M, et al. Combined analysis of estrogen receptor beta-1 and progesterone receptor expression identifies lung cancer patients with poor outcome. Clin Cancer Res. 2011 Jan;17(1):154-64.
10. Mikkonen L, Pihlajamaa P, Sahu B, Zhang FP, Jänne OA. Androgen receptor and androgen-dependent gene expression in lung. Mol Cell Endocrinol. 2010 Apr;317(1-2):14-24.
11. Ohtake F, Fujii-Kuriyama Y, Kawajiri K, Kato S. Cross-talk of dioxin and estrogen receptor signals through the ubiquitin system. J Steroid Biochem Mol Biol. 2011 Oct;127(1-2):102-7.
12. Chuang CY, Chang H, Lin P, Sun SJ, Chen PH, Lin YY, et al. Up-regulation of osteopontin expression by aryl hydrocarbon receptor via both ligand-dependent and ligand-independent pathways in lung cancer. Gene. 2012 Jan;492(1):262-9.
13. Wang R, Wang G, Ricard MJ, Ferris B, Strulovici-Barel Y, Salit J, et al. Smoking-induced upregulation of AKR1B10 expression in the airway epithelium of healthy individuals. Chest. 2010 Dec;138(6):1402-10.
14. Lam DC, Girard L, Ramirez R, Chau WS, Suen WS, Sheridan S, et al. Expression of nicotinic acetylcholine receptor subunit genes in non-small-cell lung cancer reveals differences between smokers and nonsmokers. Cancer Res. 2007 May;67(10):4638-47.
15. Zeiger JS, Haberstick BC, Schlaepfer I, Collins AC, Corley RP, Crowley TJ, et al. The neuronal nicotinic receptor subunit genes (CHRNA6 and CHRNB3) are associated with subjective responses to tobacco. Hum Mol Genet. 2008 Mar;17(5):724-34.
16. Washimi O, Nagatake M, Osada H, Ueda R, Koshikawa T, Seki T, et al. In vivo occurrence of p16 (MTS1) and p15 (MTS2) alterations preferentially in non-small cell lung cancers. Cancer Res. 1995 Feb;55(3):514-7.
17. Zhao W, Huang CC, Otterson GA, Leon ME, Tang Y, Shilo K, et al. Altered p16(INK4) and RB1 Expressions Are Associated with Poor Prognosis in Patients with Nonsmall Cell Lung Cancer. J Oncol. 2012;2012:957437.
18. Choi YL, Soda M, Yamashita Y, Ueno T, Takashima J, Nakajima T, et al. EML4-ALK mutations in lung cancer that confer resistance to ALK inhibitors. N Engl J Med. 2010 Oct;363(18):1734-9.
19. Salido M, Pijuan L, Martínez-Avilés L, Galván AB, Cañadas I, Rovira A, et al. Increased ALK gene copy number and amplification are frequent in non-small cell lung cancer. J Thorac Oncol. 2011 Jan;6(1):21-7.
20. Sanders HR, Li HR, Bruey JM, Scheerle JA, Meloni-Ehrig AM, Kelly JC, et al. Exon scanning by reverse transcriptase-polymerase chain reaction for detection of known and novel EML4-ALK fusion variants in non-small cell lung cancer. Cancer Genet. 2011 Jan;204(1):45-52.
21. Smith GD, Gunnell D, Holly J. Cancer and insulin-like growth factor-I. A potential mechanism linking the environment with cancer risk. BMJ. 2000 Oct;321(7265):847-8.
22. Dziadziuszko R, Merrick DT, Witta SE, Mendoza AD, Szostakiewicz B, Szymanowska A, et al. Insulin-like growth factor receptor 1 (IGF1R) gene copy number is associated with survival in operable non-small-cell lung cancer: a comparison between IGF1R fluorescent in situ hybridization, protein expression, and mRNA expression. J Clin Oncol. 2010 May;28(13):2174-80.
23. Kohda M, Hoshiya H, Katoh M, Tanaka I, Masuda R, Takemura T, et al. Frequent loss of imprinting of IGF2 and MEST in lung adenocarcinoma. Mol Carcinog. 2001 Aug;31(4):184-91.
24. Jin Q, Menter DG, Mao L, Hong WK, Lee HY. Survivin expression in normal human bronchial epithelial cells: an early and critical step in tumorigenesis induced by tobacco exposure. Carcinogenesis. 2008 Aug;29(8):1614-22.
25. Skrzypski M, Jassem E, Taron M, Sanchez JJ, Mendez P, Rzyman W, et al. Three-gene expression signature

predicts survival in early-stage squamous cell carcinoma of the lung. Clin Cancer Res. 2008 Aug;14(15):4794-9.

26. Müller-Tidow C, Diederichs S, Bulk E, Pohle T, Steffen B, Schwäble J, et al. Identification of metastasis-associated receptor tyrosine kinases in non-small cell lung cancer. Cancer Res. 2005 Mar;65(5):1778-82.

27. Cavatorta AL, Fumero G, Chouhy D, Aguirre R, Nocito AL, Giri AA, et al. Differential expression of the human homologue of drosophila discs large oncosuppressor in histologic samples from human papillomavirus-associated lesions as a marker for progression to malignancy. Int J Cancer. 2004 Sep;111(3):373-80.

28. Suzuki M, Shigematsu H, Shames DS, Sunaga N, Takahashi T, Shivapurkar N, et al. Methylation and gene silencing of the Ras-related GTPase gene in lung and breast cancers. Ann Surg Oncol. 2007 Apr;14(4):1397-404.

29. Heguy A, O'Connor TP, Luettich K, Worgall S, Cieciuch A, Harvey BG, et al. Gene expression profiling of human alveolar macrophages of phenotypically normal smokers and nonsmokers reveals a previously unrecognized subset of genes modulated by cigarette smoking. J Mol Med (Berl). 2006 Apr;84(4):318-28.

30. Jiang SS, Fang WT, Hou YH, Huang SF, Yen BL, Chang JL, et al. Upregulation of SOX9 in lung adenocarcinoma and its involvement in the regulation of cell growth and tumorigenicity. Clin Cancer Res. 2010 Sep;16(17):4363-73.

31. Muley T, Herth FJF, Schnabel P, Dienemann H, Meister M. From tissue to molecular phenotyping: Pre-analytical requirements Heidelberg Experience. Transl Lung Cancer Res. 2012;1(2):111-21.

32. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods. 2001 Dec;25(4): 402-8.

33. Marquez-Garban DC, Mah V, Alavi M, Maresh EL, Chen HW, Bagryanova L, et al. Progesterone and estrogen receptor expression and activity in human non-small cell lung cancer. Steroids. 2011 Aug;76(9):910-20.

34. Ishibashi H, Suzuki T, Suzuki S, Niikawa H, Lu L, Miki Y, et al. Progesterone receptor in non-small cell lung cancer—a potent prognostic factor and possible target for endocrine therapy. Cancer Res. 2005 Jul;65(14):6450-8.

35. Lin P, Chang JT, Ko JL, Liao SH, Lo WS. Reduction of androgen receptor expression by benzo[alpha]pyrene and 7,8-dihydro-9,10-epoxy-7,8,9,10-tetrahydrobenzo[alpha] pyrene in human lung cells. Biochem Pharmacol. 2004 Apr;67(8):1523-30.

36. Hershberger PA, Vasquez AC, Kanterewicz B, Land S, Siegfried JM, Nichols M. Regulation of endogenous gene expression in human non-small cell lung cancer cells by estrogen receptor ligands. Cancer Res. 2005 Feb;65(4): 1598-605.

37. Jones SE. Fulvestrant: an estrogen receptor antagonist that downregulates the estrogen receptor. Semin Oncol. 2003 Oct;30(5 Suppl 16):14-20.

38. Slatore CG, Chien JW, Au DH, Satia JA, White E. Lung cancer and hormone replacement therapy: association in the vitamins and lifestyle study. J Clin Oncol. 2010 Mar;28(9):1540-6.

39. Zhang B, Zhu W, Yang P, Liu T, Jiang M, He ZN, et al. Cigarette smoking and p16INK4α gene promoter hypermethylation in non-small cell lung carcinoma patients: a meta-analysis. PLoS One. 2011;6(12):e28882.

40. Park JH, Mangal D, Frey AJ, Harvey RG, Blair IA, Penning TM. Aryl hydrocarbon receptor facilitates DNA strand breaks and 8-oxo-2'-deoxyguanosine formation by the aldo-keto reductase product benzo[a]pyrene-7,8-dione. J Biol Chem. 2009 Oct;284(43):29725-34.

41. Park JH, Mangal D, Tacka KA, Quinn AM, Harvey RG, Blair IA, et al. Evidence for the aldo-keto reductase pathway of polycyclic aromatic trans-dihydrodiol activation in human lung A549 cells. Proc Natl Acad Sci U S A. 2008 May;105(19):6846-51.

42. Ruiz FX, Gallego O, Ardèvol A, Moro A, Domínguez M, Alvarez S, et al. Aldo-keto reductases from the AKR1B subfamily: retinoid specificity and control of cellular retinoic acid levels. Chem Biol Interact. 2009 Mar;178(1-3):171-7.

43. Yu J, Bulk E, Ji P, Hascher A, Tang M, Metzger R, et al. The EPHB6 receptor tyrosine kinase is a metastasis suppressor that is frequently silenced by promoter DNA hypermethylation in non-small cell lung cancer. Clin Cancer Res. 2010 Apr;16(8):2275-83.

44. Li NF, Kocher HM, Salako MA, Obermueller E, Sandle J, Balkwill F. A novel function of colony-stimulating factor 1 receptor in hTERT immortalization of human epithelial cells. Oncogene. 2009 Feb;28(5):773-80.

45. Lewis CE, Pollard JW. Distinct role of macrophages in different tumor microenvironments. Cancer Res. 2006 Jan;66(2):605-12.

46. Miyazaki A, Kobayashi J, Torigoe T, Hirohashi Y, Yamamoto T, Yamaguchi A, et al. Phase I clinical trial of survivin-derived peptide vaccine therapy for patients with advanced or recurrent oral cancer. Cancer Sci. 2011 Feb;102(2):324-9.

47. Lü B, Fang Y, Xu J, Wang L, Xu F, Xu E, et al. Analysis of SOX9 expression in colorectal cancer. Am J Clin Pathol. 2008 Dec;130(6):897-904.

48. Martelli MP, Sozzi G, Hernandez L, Pettirossi V, Navarro A, Conte D, et al. EML4-ALK rearrangement in non-small cell lung cancer and non-tumor lung tissues. Am J Pathol. 2009 Feb;174(2):661-70.

49. Bossé Y, Postma DS, Sin DD, Lamontagne M, Couture C, Gaudreault N, et al. Molecular signature of smoking in human lung tissues. Cancer Res. 2012 Aug;72(15):3753-63.

50. Beane J, Sebastiani P, Liu G, Brody JS, Lenburg ME, Spira A. Reversible and permanent effects of tobacco smoke exposure on airway epithelial gene expression. Genome Biol. 2007;8(9):R201.

51. Boelens MC, van den Berg A, Fehrmann RS, Geerlings M, de Jong WK, te Meerman GJ, et al. Current smoking-specific gene expression signature in normal bronchial epithelium is enhanced in squamous cell lung cancer. J Pathol. 2009 Jun;218(2):182-91.

52. Landi MT, Chatterjee N, Yu K, Goldin LR, Goldstein AM, Rotunno M, et al. A genome-wide association study of lung cancer identifies a region of chromosome 5p15 associated with risk for adenocarcinoma. Am J Hum Genet. 2009 Nov;85(5):679-91.