MOLECULAR & CELLULAR TOXICOLOGY, Vol. 2, No. 3, 202-211, September 2006

Biological Effects of Smoking-induced Environmental Toxicity

Sung-Hwa Sohn[1], In Kyoung Kim[1], Ki-Nam Kim[1], Hye Won Kim[1], Sang-Hui Seo[1], Seung Ho Lee[1], Yu-Ri Kim[1], Eun Il Lee[2] & Meyoung-Kon Kim[1]

[1]*Department of Biochemistry & Molecular Biology, College of Medicine, Korea University, Seoul 136-705, Korea*
[2]*Department of Preventive, College of Medicine, Korea University, Seoul 136-705, Korea*
Correspondence and requests for materials should be addressed to M-K. Kim (jerrykim@korea.ac.kr)

Accepted 25 August 2006

Abstract

Our objective is to identify molecular factors which contribute to the increased risk of smoke in human. About 677 workers who had control and experimental groups according to their urinary Naphthol levels were enrolled in our study. In the present study, we investigated the effects of smoking on gene expression profiles in human. We determined differential gene expression patterns in smoker versus non-smoker using cDNA microarray. Specific genes were up- or down-regulated according to smoking and age. Inflammatory related genes such as cytokine, interleukin, and tumor necrosis factor were up-regulated, DNA repair related genes such as high-mobility group (nonhistone chromosomal) protein 1, and protein 2 were down-regulated, apoptosis related genes such as myeloperoxidase and Bcl-2-associated athanogene were down-regulated, and cell cycle related genes were down-regulated. In our epidemiological study, notably, inflammatory, DNA repair, apoptosis, signal transduction, metabolism, cell cycle, cell proliferation, transcription related genes were regulated.

Keywords: Smoking, Urinary Naphthol, Gene expression profiling, cDNA microarray

The numbers of deaths attributed by tobacco smoking have remarkably increased, reflecting the heavy smoking patterns of previous decades. It has been estimately 30% of all cancer deaths[1]. Tobacco smoking is estimated to account for approximately 4-5 million deaths a year worldwide. Thus, if current smoking patterns continue, there will be more than 1 billion deaths attributed by tobacco smoking in the 21st century compared with approximately 100 million deaths in the 20th century[1,2]. Tobacco smoke has been implicated in various human disease conditions including asthma, bronchitis, myocardial infarction, chronic obstructive pulmonary disease, intracranial aneurysms, and lung cancer[3-6]. In addition, it is known that active tobacco smoking was carcinogenic in humans, and that it caused cancer not only of the lung, but also of many other organs including nasal cavities, esophagus, paranasal sinuses, pharynx, larynx, liver, stomach, pancreas, kidney, pelvis, bladder, and uterine cervix[1,7]. Recently, the International Agency for Research on Cancer identified tobacco smoking as the cause of cancer at more organ sites than any other human carcinogen[8].

Many of the chemical carcinogens contained in tobacco smoke are polycyclic aromatic hydrocarbons (PAH), a family of ubiquitous environmental carcinogens that are known to have mutagenic and carcinogenic effects. Vineis *et al.*[1] reported that PAH adducts are present in DNA extracted from human tissues exposed to tobacco smoke, and that There is a precise correlation between mutational hotspots and hotspots of adduct formation by PAHs found in tobacco smoke. PAHs are activated to genotoxic intermediates through at least three primary pathways, all of which can lead to the production of G to T transversions[9]. For example, G to T transversions within the p53 gene have been linked to a molecular signature of tobacco mutagens in smoking associated lung cancers[1,2]. Another major group of chemicals in tobacco smoke are N-nitroso compounds are another major group of chemicals found in tobacco smoke, several of which are potent animal carcinogens. N-nitroso compounds are found in the urine of smokers. In particular, compounds known as 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanol (NNAL) and NNAL-Gluc are very useful biomarkers because they are derived from the carcinogen 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK), which is specific to tobacco products[1,10]. Also, Kellen *et al.*[11] reported that smoking is the main risk factor for bladder cancer, and the aromatic amines and PAH may play an important role in perhaps 10% of bladder cancers. The purpose of the present study was to identify gene expression profiling as molecular factors related with smoking using by cDNA microarray. Extraction of potential biological interaction networks among those iden-

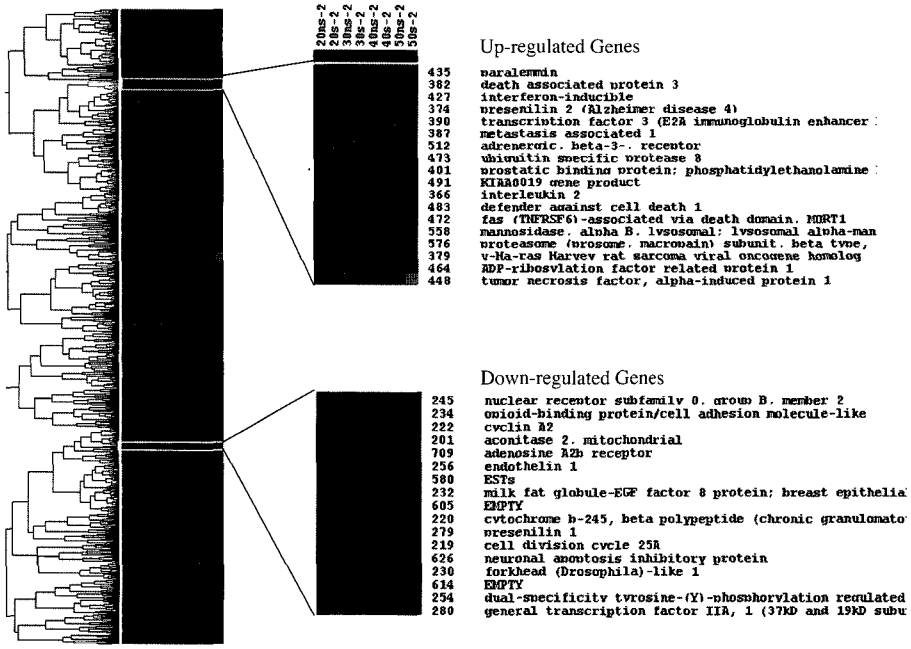


Fig. 1. Clustergram of up- and down-regulated gene in smoking and age groups. Microarray data from non-smoker group and smoker group were combined and clustered. Cluster analysis was performed on Z-transformed microarray data using two separate programs available as shareware from Michael Eisen's lab. Each gene is represented by a single row of colored boxes; each experimental sample is represented by a single column. The entire clustered image is shown on the left. These clusters contain uncharacterized genes and genes not involved in these processes.

Table 1. General characteristics of subjects.

	N (%)
Age (year old)	
20-29	200 (29.5%)
30-39	200 (29.5%)
40-49	189 (27.9%)
50-59	88 (13.1%)
Smoking	
Non-smoker	206 (29.4%)
Ex-smoker	166 (23.7%)
Smoker	329 (46.9%)

Table 2. Age distribution of control and experimental groups.

	Control groups (N=642)	Experiment groups (N=35)	Significance
Age (year old)			
20-29	192 (29.9%)	8 (22.9%)	χ^2=3.649 P=0.302 df=3
30-39	185 (28.8%)	15 (42.9%)	
40-49	182 (28.3 %)	7 (20.0%)	
50-59	83 (12.9%)	5 (14.3%)	

Table 3. Levels of 2-naphthol in control and experimental groups.

	Control groups (N=642)	Experiment groups (N=35)	P-value
2-Naphthol (μ mol/mol creatinine)	1.82±1.64	8.56±2.31	0.000

tified and validated genes also suggests novel mechanistic pathways to rationalize the smoking response in human.

(Note: although much of the evidence is based on cigarette smoking, many of the papers also contained information on other forms of tobacco smoking. Consequently, we use the generic term "tobacco" to include all forms of smoking. In addition, the term "non-smokers" as used by the authors, usually means the more appropriate term "never smokers.")

General Characteristics and Urinary Naphthol Levels of Subjects

Table 1 shows general characteristics of subjects such as age and smoking habits. There are no significant differences of age distribution between control and experimental groups (Table 2). More subjects in experimental group (100%) were surveyed as smoker than those in control group (non-smoker and ex-smoker, 53.1%).

As shown in Table 3, urinary Naphthol levels were significantly higher in experimental group (8.56± 2.31 μmol/mol creatinine) than those in control group (1.82±1.64 μmol/mol creatinine).

Gene Expression Profiles in Smoker

Gene expression profiles of interest were significantly up-regulated or down-regulated in smoker group when compared with non-smoker group. Genes showing highly altered expression levels were aligned

Table 4. Up- and down-regulated gene expression in smoking group compared with non-smoking group.

Genes	Abb.	Regulation profile and Z-ratio			
		20 s	30 s	40 s	50 s
<Inflammatory and immune related group>					
CD14 antigen	CD14	**2.05**	1.66	0.19	−0.36
interleukin 9	IL9	**−2.52**	0.16	1.74	0.25
small inducible cytokine subfamily A (Cys-Cys), member 16	SCYA16	0.05	**2.18**	−0.07	−0.28
pentaxin-related gene, rapidly induced by IL-1 beta	PTX3	0.88	**2.01**	−0.10	0.44
GRO3 oncogene	GRO3	0.62	**−4.82**	1.28	**−4.77**
major histocompatibility complex, class II, DQ beta 1	HLA-DQB1	−0.94	**−4.32**	−0.47	**−4.82**
pro-platelet basic protein	PPBP	−0.72	**−4.27**	−0.74	**−2.16**
interleukin 12A	IL12A	−0.88	**−3.28**	1.56	**−2.42**
major histocompatibility complex, class II, DN alpha	HLA-DNA	−1.13	**−3.06**	−1.54	**−2.35**
interleukin 10 receptor, beta	IL10RB	1.00	**−2.28**	−0.36	**−3.23**
tumor necrosis factor (ligand) superfamily, member 4	TNFSF4	−0.40	−0.86	**2.64**	0.01
Geminin	GEM	0.91	0.80	**2.50**	0.65
small inducible cytokine subfamily A (Cys-Cys), member 23	SYCA23	−0.24	−0.38	**2.13**	0.08
interferon (gamma)-induced cell line; protein 10 from	INP10	1.38	0.04	**2.05**	−0.09
TCR adaptor molecule cbl-b	CBLB	−0.02	0.52	**2.02**	1.01
interferon, alpha-inducible protein	GIP3	−1.83	0.03	**−2.57**	0.42
chemokine (C-C motif) receptor 2	CCR2	1.30	0.64	**−2.35**	−0.62
major histocompatibility complex, class II, DP beta 1	HLA-DPB1	−0.17	−0.74	**−2.28**	0.23
cytokine receptor induced by latent Epstein-Barr virus infection	EBI3	0.90	0.31	**−2.16**	0.44
interleukin 2 receptor, gamma	IL2RG	0.93	0.19	−1.56	**6.66**
interleukin 6 (interferon, beta 2)	IL6	−1.30	−0.96	−0.21	**6.35**
lymphotoxin beta (TNF superfamily, member 3)	LTB	0.13	**2.08**	0.00	0.93
<DNA repair>					
high-mobility group (nonhistone chromosomal) protein 1	HMG1	**−2.58**	−0.88	−1.96	1.69
high-mobility group (nonhistone chromosomal) protein 2	HMG2	1.84	−1.57	0.26	**−2.91**
<Apoptosis and anti-apoptosis related group>					
deleted in colorectal carcinoma	DCC	**2.63**	0.52	0.22	1.90
myeloperoxidase	MPO	**−2.24**	**−4.93**	−0.83	**−4.15**
TNF receptor-associated factor 3, CD40 binding protein	TRAF3	**−2.08**	−1.33	−0.27	0.80
Bruton agammaglobulinemia tyrosine kinase	BTK	−1.81	**2.20**	0.37	−0.64
apoptotic protease activating factor	APAF1	0.64	**2.16**	−0.34	1.94
Bcl-2-associated athanogene	BAG1	0.73	**−4.16**	−0.25	**−3.67**
caspase 10	CASP10	−0.33	0.07	**2.13**	−0.27
TIA1 cytotoxic granule-associated RNA-binding protein-like 1	TIAL1	0.24	0.29	**2.13**	0.74
<Cell cycle related group>					
suppressor of Ty (S.cerevisiae) 3 homolog	SUPT3H	−1.17	**−5.99**	1.77	**−4.76**
cyclin D1 (PRAD1: parathyroid adenomatosis 1)	CCND1	−0.34	**−4.02**	0.60	**−2.52**
cell division cycle 25A	CDC25A	−1.90	**−2.26**	−1.54	−0.15
protein kinase, mitogen-activated 6	ERF	0.40	−0.01	**−2.84**	0.34
<Signal transduction group>					
G protein-coupled receptor	TYMSTR	**3.10**	0.68	−1.82	−0.25
ras homolog gene family, member G (rho G)	ARHG	**2.99**	−0.05	−1.83	0.00
CD8 antigen, alpha polypeptide (p32)	CD8A	**2.40**	−.79	−0.15	−0.60
prostaglandin E receptor 2 (subtype EP2), 53 kD	PTGER2	**2.38**	0.60	1.75	−0.08
EphB2	EPHB2	**2.23**	0.71	0.78	−0.56
phosphatidylinositol-4-phosphate 5-kinase I-gamma	KIAA0589	**2.20**	0.78	**−2.09**	0.15
G protein-coupled receptor 37	GPR37	**2.06**	0.96	−0.09	−0.77
glutamate receptor, ionotropic, N-methyl D-aspartate 2C	GRIN2C	**−3.08**	**−2.90**	−1.26	**5.25**
transforming growth factor,beta receptorIII (betaglycan, 300 kD)	TGFBR3	**−2.88**	−1.97	−1.05	−0.06
platelet-derived growth factor receptor, beta polypeptide	PDGFRB	**−2.44**	−0.93	0.04	−0.49
v-Ki-ras2 Kirsten rat sarcoma 2 viral oncogene homolog	KRAS2	**−2.28**	−1.33	−1.33	0.42
anti-Mullerian hormone receptor, type II	AMHR2	**−2.27**	0.02	1.59	−1.18
v-akt murine thymoma viral oncogene homolog 2	AKT2	**−2.20**	−0.78	−0.94	−0.33
prodynorphin	PDYN	1.63	**−4.50**	0.48	**−4.82**
EphA1	EPHA1	0.96	1.25	**2.09**	0.84
chemokine (C-X3-C) receptor 1	CX3CR1	−0.42	−0.05	**2.08**	0.59
Janus kinase 1 (a protein tyrosine kinase)	JAK1	1.16	1.13	**2.08**	0.33
hypoxia-inducible factor 1, alpha subunit	HIF1A	1.24	0.37	**2.07**	−0.82

Table 4. Continued.

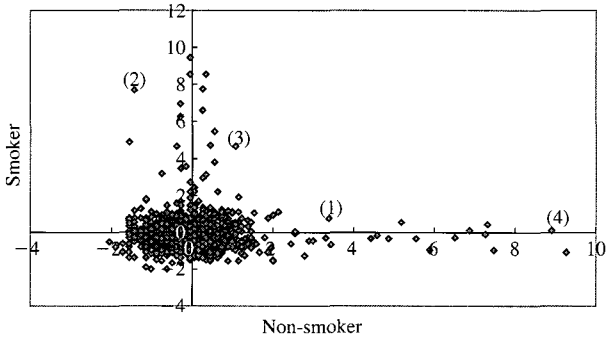
Genes	Abb.	Regulation profile and Z-ratio			
		20 s	30 s	40 s	50 s
v-yes-1 Yamaguchi sarcoma viral related oncogene homolog	LYN	0.41	−0.09	**−2.37**	0.20
CD69 antigen (p60, early T-cell activation antigen)	CD69	1.03	0.25	**−2.35**	0.45
mitogen-activated protein kinase kinase kinase 8	MAP3K8	0.15	−0.24	**−2.27**	−0.73
endothelin receptor type B	EDNRB	1.40	0.36	**−2.11**	−0.50
solute carrier family 9, isoform 3 regulatory factor 2	SLC9A3R2	−0.26	−0.56	−1.09	**4.57**
interferon (alpha, beta and omega) receptor 2	IFNAR2	−1.59	−0.10	−0.44	**3.54**
mitogen-activated protein kinase kinase kinase kinase 1	MAP4K1	0.69	0.96	−0.15	**2.28**
<Cell adhesion related group>					
claudin 3	CLDN3	**−2.22**	**−2.00**	−1.63	0.54
villin2	VIL2	0.26	**2.01**	0.12	−0.37
catenin (cadherin-associated protein), delta 2	CTNND2	0.66	0.47	**2.24**	0.15
Egf-like module containing, mucin-like, hormone receptor-like sequence 1	EMR1	1.61	0.29	**−2.23**	−0.82
pleckstrin homology, Sec7 and coiled/coil domains, binding protein	PSCDBP	1.39	0.46	**−2.19**	−1.57
integrin, alpha 6	ITGA6	−0.05	0.01	0.59	**5.99**
integrin, beta 2	ITGB2	**−2.04**	−0.84	−0.99	**2.99**
<Metabolism related group>					
3' mRNA for neurone-specific enolase (EC 4.2.1.11)	ENO2	−0.86	**−4.25**	0.10	**−3.38**
acyloxyacyl hydrolase (neutrophil)	AOAH	0.35	**−2.36**	−0.08	−1.84
phosphatidylinositol glycan, class A (paroxysmal nocturnal hemoglobinuria)	PIGA	−1.80	**−2.29**	−0.85	−0.32
phospholipase A2, group V	PLA2G5	0.73	−0.27	**2.25**	0.92
retinol-binding protein 1, cellular	RBP1	0.90	0.15	**2.01**	1.06
sterol regulatory element binding transcription factor 1	SREBF1	0.35	0.52	**−2.31**	0.08
phospholipase C, gamma 2 (phosphatidylinositol-specific)	PLCG2	−1.92	**−2.14**	**−2.11**	1.29
ubiquitin-like 1 (sentrin)	UBL1	0.49	1.32	−0.19	**2.51**
<Protein folding related group>					
cyclophilin H	USA-CYP	1.88	**2.21**	−0.17	0.90
tumor rejection antigen (gp96) pseudogene 1	TRAP1	0.81	−0.04	**−2.34**	−0.60
<Cell proliferation and differentiation related group>					
interleukin 1, beta	IL1B	**2.18**	0.83	0.84	−0.18
retinoblastoma-binding protein 7	RBBP7	**2.08**	0.76	1.80	−0.78
neurotrophic tyrosine kinase, receptor, type 3	NTRK3	**2.06**	0.88	1.46	−0.17
TGFB inducible early growth response	TIEG	**−3.19**	**−2.77**	−1.32	−0.26
v-ski avian sarcoma viral oncogene homolog	SKI	**−2.70**	−1.77	−0.71	0.66
pleiotrophin	PTN	**−2.65**	−1.83	−0.61	−0.38
singed (Drosophila)-like (sea urchin fascin homolog like)	SNL	−0.05	**2.16**	0.31	0.37
platelet-derived growth factor alpha polypeptide	PDGFA	−1.24	**−2.62**	−1.11	−0.75
endothelin 1	EDN1	−1.39	**−2.42**	−1.25	−1.01
angiopoietin 1	ANGPT1	1.13	−0.22	**2.32**	−1.24
feline sarcoma (Snyder-Theilen) viral (v-fes)/Fujinami avian sarcoma (PRCII) viral (v-fps) oncogene homolog	FES	1.73	0.35	**2.13**	−0.13
epidermal growth factor receptor pathway substrate 15	EPS15	0.90	0.37	**−2.33**	−0.13
insulin-like growth factor 2 (somatomedin A)	IGF2	−0.74	−0.49	−1.04	**2.73**
kinase insert domain receptor (a typeIII receptor tyrosine kinase)	KDR	1.21	0.71	−0.56	**2.02**
tissue inhibitor of metalloproteinase 1	TIMP1	0.90	0.07	1.96	**−2.44**
<Transcription group>					
SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 3	SMARCA3	**2.34**	0.05	−0.68	−0.86
nuclear factor of activated T-cells, cytoplasmic 1	NFATC1	**−2.78**	−1.51	−0.12	−0.21
early growth response 3	EGR3	**−2.76**	**−2.18**	−1.65	0.51
general transcription factor IIF, polypeptide 1 (74kD subunit)	GTF2F1	**−2.75**	−1.90	−1.04	0.11
jun B proto-oncogene	JUNB	**−2.50**	**−2.21**	−0.50	−0.10
nuclear transcription factor Y, alpha	NFYA	−0.07	**−6.41**	−0.52	**−7.24**
v-maf musculoaponeurotic fibrosarcoma (avian) oncogene homolog	MAF	−0.73	**−4.20**	−0.09	**−3.37**
TEA domain family member 4	TEAD4	−0.39	**−3.27**	0.05	−1.79
nuclear factor I/C (CCAAT-binding transcription factor)	NFIC	−1.39	−0.46	**2.35**	−0.40
TATA box binding protein (TBP)-associated factor, RNA polymerase II, J, 20 kD	TAF2J	0.94	0.54	**−2.61**	0.33
basic transcription factor 3	BTF3	0.59	1.01	**−2.47**	−0.71
GATA-binding protein 6	GATA6	0.95	0.37	**−2.39**	0.20

Table 4. Continued.

Genes	Abb.	Regulation profile and Z-ratio			
		20 s	30 s	40 s	50 s
Golgi autoantigen, golgin subfamily b, macrogolgin, 1	GOLGB1	−0.32	−0.36	1.25	**5.72**
signal transducer and activator of transcription 2, 113kD	STAT2	**−2.23**	0.23	0.74	−0.05
<Unknown group>					
microseminoprotein, beta-	MSMB	0.12	**−6.01**	0.97	**−6.18**
Ndr protein kinase	NDR	0.78	**−5.50**	0.15	**−5.20**
nuclear factor of kappa light polypeptide gene enhancer in Bcells inhibitor,epsilon	NFKBIE	−0.18	**−4.31**	0.48	**−5.93**
manic fringe (Drosophila) homolog	MFNG	1.32	1.00	**2.27**	−0.17
X (inactive)-specific transcript	XIST	−0.97	−0.59	**2.12**	0.65
new Ets-related factor	NERF-2	1.30	1.19	**2.10**	0.23
collagen, type III, alpha 1	COL3A1	0.80	−0.12	**−2.54**	−0.46
macrophage stimulating 1	MST1	0.26	0.30	**−2.42**	0.20
iron-responsive element binding protein 1	IREB1	−1.09	−1.07	−0.79	**3.45**
capping protein (actin filament) muscle Z-line, beta	CAPZB	−0.91	−0.57	−1.03	**2.05**
paraoxonase 3	PON3	1.50	−1.01	0.16	**−2.85**
grancalcin	GCL	1.07	−0.49	1.53	**−2.52**

[1)]$Z\text{-ratio}_{(gene1)}=Z\text{-difference}_{(gene1)}/Sdev_{(Z\text{-difference all genes})}$; Fold of expression change for an individual gene based on the ratio for the experimental group (smoking) compared to that of the control group (non-smoking).

in the order of magnitude of altered expression in smoker group. The up- and down-regulated genes are listed in Table 4. Table 4 showed that the regulation of genes related with age was not different in smoker compared to non-smoker group. Fig. 2 is a Scatter plot for comparing the expression profiles between smoker and non-smoker group. Expression profiles of smoker and nonsmoker group are shown as a scatter plot of 1,152 genes from the microarray. Regression analysis of Z scores from two independent samples of smoker and non-smoker group was performed and Z scores of individual genes were plotted. To obtain a molecular portrait of the relationships between the metabolism associated with smoker group, we used a hierarchical clustering algorithm to group genes on the basis of similar expression patterns and the data is presented in a matrix format (Fig. 1). Each row of Fig. 1 represents the hybridization results for a single DNA element of the array and each column represents the expression levels for all genes in a single hybridization sample. The expression level of each gene was visualized in color, relative to its median expression level across all samples. Red represented expression greater than the mean, green represents expression less than the mean and color intensity denotes the degree of deviation from the mean. Gray represented the median expression level. Distinct samples representing similar gene patterns from control cells were aligned in adjacent rows. The cells included in this map were samples from right smoker and non-smoker group. Coordinately expressed genes



(1) thyroid stimulating hormone receptor
(2) interleukin 6 (interferon, beta 2)
(3) ubiquitin-like 1(sentrin)
(4) microseminoprotein, beta

Fig. 2. Scatter plot for comparison of expression profile between non-smoker and smoker group. Expression profiles of non-smoker and smoker groups are shown as bivariated scatter plot of 1,152 genes from the microarray. The values are corrected intensities relative to control (non-smoker), representing levels of expression for the cDNA elements of the microarrays.

were grouped into clusters, which were named on the basis of the cellular process in which the component genes participated in. The cluster gram revealed that clusters of genes related to progression were up- and down-regulated in smoker group, as compared to non-smoker group (Fig. 1).

Discussion

Tobacco smoking is the main known cause of cancer-related death worldwide. Until recently, the health effects of tobacco smoking were confined largely to developed countries. It has been estimated that tobacco causes approximately 25% of all cancers in men and 4% in women, and in both sexes, approximately 16% of cancers in more developed countries and 10% in less developed countries. It is estimated that smoking killed four to five million people each year worldwide in the 1990s, and was responsible for some 60 million deaths in the second half of last century. In most countries, the worst consequences of the tobacco epidemic are yet to emerge, notably among women in the developing countries and among men in the developing countries. In fact, half of the current smokers will be killed by tobacco. Moreover, smokers are responsible for the increased incidence of tobacco-related diseases in non-smokers exposed to second-hand tobacco smoke[7]. Tobacco smoking is known to contribute to PAH uptake in general and naphthalene uptake in particular, our data confirm this finding that smokers exhibited higher naphthol concentrations than non-smokers (Table 3). Our data are in agreement with several previous reports[12,13]. Overall 2-naphthol levels were determined to be four-times higher in smoker (Table 3).

Our objective was also to determine response differences among the age in human following exposure to smoking. In this experiment, several genes were up- or down-regulated (Table 4). These genes were categorized depending on their functions, as follows, inflammatory and immune related group, DNA repair, apoptosis and anti-apoptosis related group, transport related group, signal transduction group, cell adhesion related group, metabolism related group, cell cycle related group, cell proliferation and differentiation related group, transcription group and protein folding related group in smoker. Also, we carried out the evaluation of the differences in gene expression profiles between smoker groups and non-smoker according to ages.

Genes involved in inflammatory and immune such as major histocompatibility complex, class II, DQ beta 1 (HLA-DQB1), major histocompatibility complex, class II, DO alpha (HLA-DNA), major histocompatibility complex, class II, DP beta 1 (HLA-DPB1), pro-platelet basic protein (chemokine (C-X-C motif) ligand 7) (PPBP), interleukin 12A (IL12A), lymphotoxin beta (TNF superfamily, member 3) (LTB) and interleukin 10 receptor, beta (IL10RB), DNA repair such as high-mobility group protein 1 (HMG1) and high-mobility group protein 2 (HMG2) Oxidative stress such as signal transducer and activator of transcription 2 (STAT2), apoptosis such as apoptotic peptidase activating factor (APAF1), signal transduction such as ras homolog gene family, member G (rho G) (ARHG), G protein-coupled receptor 37 (GPR37) and platelet-derived growth factor receptor, beta polypeptide (PDGFRB), metabolism such as acyloxyacyl hydrolase (AOAH), cell cycle such as cyclin D1 (CCND1), cell proliferation such as platelet-derived growth factor alpha polypeptide (PDGFA) and tissue inhibitor of metalloproteinase 1 (TIMP1) and genes involved in transcription related several genes in subjects with higher serum naphthol levels were up- and down-regulated in Smoker.

HLA-DQB1 belongs to the HLA class II beta chain paralogues. This class II molecule is a heterodimer consisting of an alpha (DQA) and a beta chain (DQB), both anchored in the membrane. It plays a central role in the immune system by presenting peptides derived from extracellular proteins. Class II molecules are expressed in antigen presenting cells (APC: B lymphocytes, dendritic cells, macrophages). The beta chain is approximately 26-28 kDa and it contains 6 exons. Exon one encodes the leader peptide, exons 2 and 3 encode the two extracellular domains, exon 4 encodes the transmembrane domain and exon 5 encodes the cytoplasmic tail. Within the DQ molecule both the alpha chain and the beta chain contain the polymorphisms specifying the peptide binding specificities, resulting in up to 4 different molecules. Typing for these polymorphisms is routinely done for bone marrow transplantation. In our experiments, we detected the down-regulation of HLA-DQB1, HLA-DNA and HLA-DPB1 in smoker.

LTB is a type II membrane protein of the TNF family. It anchors lymphotoxin-alpha to the cell surface through heterotrimer formation. The predominant form on the lymphocyte surface is the lymphotoxin-alpha 1/beta 2 complex (e.g. 1 molecule alpha/2 molecules beta) and this complex is the primary ligand for the lymphotoxin-beta receptor. The minor complex is lymphotoxin-alpha 2/beta 1. LTB is a mediator of inflammatory and immune defense mechanisms, and a regulator of lymphoid organ development. In our experiments, we detected the up-regulation of LTB in 30-age smoker[14].

HMG gene encodes a member of the non-histone chromosomal high mobility group protein family. The proteins of this family are chromatin-associated and ubiquitously distributed in the nucleus of higher eukaryotic cells. In vitro studies have demonstrated that this protein is able to efficiently bend DNA and form DNA circles. These studies suggest a role in

facilitating cooperative interactions between cis-acting proteins by promoting DNA flexibility. This protein was also reported to be involved in the final ligation step in DNA end-joining processes of DNA double-strand breaks repair[15]. In our experiments, we detected the down-regulation of HMG1 in 20-age smoker and down-regulation of HMG2 in 50-age smoker.

APAF1 gene encodes a cytoplasmic protein that initiates apoptosis. This protein contains several copies of the WD-40 domain, a caspase recruitment domain (CARD), and an ATPase domain (NB-ARC). Upon binding cytochrome c and dATP, this protein forms an oligomeric apoptosome. The apoptosome binds and cleaves caspase 9 preproprotein, releasing its mature, activated form. Activated caspase 9 stimulates the subsequent caspase cascade that commits the cell to apoptosis[16]. In our experiments, we detected the up-regulation of APAF1 in 30-age smoker.

ARHG is a member of the RAS superfamily of genes, which encode GTP-binding proteins that act in the pathway of signal transduction and play a key role in the regulation of cellular functions[17]. In our experiments, we detected the up-regulation of ARHG in 20-age smoker.

PDGFRB gene encodes a cell surface tyrosine kinase receptor for members of the platelet-derived growth factor (PDGF) family. The PDGF family comprises four members, PDGF-A, -B, -C and -D, which are assembled as disulfide-linked homo- or heterodimers of two distinct but related chains. Two receptor subtypes have been identified that can form mature dimeric receptor complexes, the α and β subunits (PDGFR-α, -β), which can bind to ligands with different affinities. PDGF-A and -B are widely distributed in CNS. PDGFR-α is largely expressed in oligodendroglial O-2A precursors and not significantly expressed in cortical and hippocampal neurons, while PDGFR-β is predominantly expressed in neurons[18]. In our experiments, we detected the down-regulation of PDGFRB in 20-age smoker.

The AOAH is a 2-subunit lipase which selectively hydrolyzes the secondary (acyloxyacyl-linked) fatty acyl chains from the lipid A region of bacterial endotoxins (lipopolysaccharide, LPS). Enzymatic deacylation of LPS by AOAH might thus function both to detoxify LPS and to generate an LPS antagonist in humans. AOAH may modulate host inflammatory responses to gram-negative bacterial invasion[19]. In our experiments, we detected the down-regulation of AOAH in 30-age smoker.

The CCND1 protein encoded by this gene belongs to the highly conserved cyclin family, whose members are characterized by a dramatic periodicity in protein abundance throughout the cell cycle. Cyclins function as regulators of CDK kinases. Different cyclins exhibit distinct expression and degradation patterns which contribute to the temporal coordination of each mitotic event. This cyclin forms a complex with and functions as a regulatory subunit of CDK4 or CDK6, whose activity is required for cell cycle G1/S transition. This protein has been shown to interact with tumor suppressor protein Rb and the expression of this gene is regulated positively by Rb. Mutations, amplification and overexpression of this gene, which alters cell cycle progression, are observed frequently in a variety of tumors and may contribute to tumorigenesis[20]. In our experiments, we detected the down-regulation of CCND1 in 30- and 50-age smoker.

The PDGFA protein encoded by this gene is a member of the platelet-derived growth factor family. The four members of this family are mitogenic factors for cells of mesenchymal origin and are characterized by a motif of eight cysteines. This gene product can exist either as a homodimer or as a heterodimer with the platelet-derived growth factor beta polypeptide, where the dimers are connected by disulfide bonds. Studies using knockout mice have shown cellular defects in oligodendrocytes, alveolar smooth muscle cells, and Leydig cells in the testis, knockout mice die either as embryos or shortly after birth. Two splice variants have been identified for this gene. In our experiments, we detected the down-regulation of PDGFA in smoker.

The STAT2 protein encoded by this gene is a member of the STAT protein family. The Stat proteins are a group of cytoplasmic transcription factors. The seven mammalian members of this family, Stat1, Stat2, Stat3, Stat4, Stat5a, Stat5b and Stat6 The Stat genes are localized to three chromosomal regions: Stats 1 and 4 are located on human chromosome 2 (murine chromosome 1), Stats 2 and 6 are on human chromosome 12 (murine chromosome 10) and Stats 3, Stat5a, and Stat5b are located on chromosome 17 (murine chromosome 11). The basic function of Stat complexes is to induce the transcriptional activity of target genes. Association of activated Stats with a specific promoter substantially increases the transcription rate from that promoter. The ability of Stat-containing complexes to regulate transcription is dependent largely on their recruitment and interaction with a number of nuclear coactivator and adaptor proteins[21]. In our experiments, we detected the down-regulation of STAT2 in 20-age smoker

In conclusion, there is ample evidence showing that the smoking-induced environmental toxicity and cancer is age dependent. We demonstrated that smok-

ing is environmental toxicants of concern due to potent toxic effects on humans, inducing multiple organ pathology and carcinogenesis. Consequently, the expression profiles associated with PAH could be useful in monitoring human populations in the environment for potential point sources of environmental contamination. The microarray-based genomic survey is a high-throughput approach that allows parallel study on expression patterns of numerous of genes. There has been interest in using array in toxicology to quickly classify toxicants based on characteristic expression profiles and to use these profiles as means of identifying putative mechanisms of action.

Methods

Subjects

In clinical study, 2000 workers who had undergone an annual health examination at Korea University Hospital (Seoul, South Korea) from June 1 to December 31, 2004 were enrolled in our study. All subjects completed a questionnaire, which include items on smoking, age, mediation, etc, and their blood and urine were collected. They had not eaten any grilled and smoked meats within the last 48h. We divided to control (Urinary Naphthol under 95th confidence upper limit) and Experimental groups (Urinary Naphthol above 95th confidence upper limit) according to their urinary Naphthol concentrations.

Determination of 2-naphthol in Urine

Urinary 2-naphthol was analyzed according to the method of Kim *et al.*[22]. Three milliliter of urine samples were buffered with 300 μL of 0.2 M sodium acetate buffer (pH 5.0), and then hydrolyzed enzymatically with 30 μL of β-glucuronidase and sulfatase (Sigma Co. St. Louis, MO, USA), for 16 h at 37°C in a shaking water bath. After hydrolysis, 5 ml of acetonitrile was added and centrifuged at 1,000 g for 10 min. A 20 μL of the supernatant was injected into the HPLC system. A standard stock solution of 2-naphthol was prepared by dissolving 2-naphthol (Sigma Co. St. Louis, MO, USA) in acetonitrile (Merck Co. Darmstadt, Germany). The urinary creatinine was determined with a Hitachi 747 Computer Directed Analyzer.

Blood Sample Preparation

Blood samples, 4-5 mL of heparinized whole blood, were collected by veinpuncture from each human subject. The human blood plasma and the buffy coat containing white blood cells were isolated from the remaining whole blood by centrifugation (1,000 g, 5 min) and then this was aliquoted and frozen at −70°C until RNA extraction.

RNA Preparation

Frozen whole human blood were transferred to 1.0 ml of TRIzol reagent™ (Invitrogen, Carlsbad, CA) were transferred to 1.0 mL of TRIzol reagent and homogenized in a polytron homogenizer (PRO Scientific Inc. Monroe. CT USA).

Total RNA was isolated using TRIzol reagent according to protocols provided by the manufacturer's recommendations (http://www.protocol-online.org/). In brief, after homogenization with TRIZOL, 0.2 mL chloroform was added for each 1 mL of TRIZOL usage, and the sample was centrifuged at 12,000 × g, 4°C for 15 min. Then, the aqueous phase was transferred into a new tube and 0.5 mL of isopropyl alcohol was added. After centrifugation at 12,000 × g, 4°C for 10 min, the RNA was precipitated as a white pellet at the bottom of the tube. The RNA pellets in 75% ethanol were washed and centrifuged at 7,500 × g, 4°C for 5 min. The RNA pellets were dried and dissolved in RNase-free water and incubated for 10 min. at 55 to 60°C. The total RNA was quantified using the NanoDropND-1000 Spectrophotometer (NanoDrop, Montchanin, USA). After quantification the total RNA was stored at −70°C. An aliquot of total RNA was resolved with the use of agarose gel electrophoresis and RNA integrity was assessed from a visual comparison of the relative intensities of the 18S and 28S rRNA bands. For all samples, the intensity of the 28S rRNA band exceeded that of the 18S band.

Radioactive cDNA Microarray

A human cDNA microarray was primarily derived from a commercially available master set of approximately 15,000 human verified sequences (Research Genetics, Inc., Huntsville, AL). The 15,000-human cDNA clone set was sorted for a list of genes (1,152 elements) representing families involved in differentiation, development, proliferation, transformation, cell-cycle progression, immune response, transcription and translation factors, oncogenes, and molecules involved in cell growth and maintenance. PCR-amplified cDNAs were spotted on nylon membranes. The general methodology of arraying is based on the procedures of DeRisi *et al*[23].

cDNA Radiolabeling

Total RNAs prepared from control and experimental group were used to synthesize ^{33}P-labeled cDNAs by reverse transcription. Briefly, 3-10 g of

RNA were labeled in a reverse transcription reaction containing 5 X first-strand PCR buffer, 1 g of 24-mer poly dT primer, 4 L of 20 mM each dNTP excluding dCTP, 4 L of 0.1 M DTT, 40 U of RNase inhibitor, 6 l of 3,000 Ci/mmol-^{33}P dCTP to a final volume of 40 L. The mixture was heated at 65°C for 5 min, followed by incubation at 42°C for 3 min. 2 L (specific activity: 200,000 U/mL) of Superscript II reverse transcriptase (Life Technologies, Inc., Rockville, MD) was then added and the samples were incubated for 30 min at 42°C, followed by the addition of 2 L of Superscript II reverse transcriptase and another 30 min of incubation. 5 L of 0.5 M EDTA was added to chelate divalent cations. After the addition of 10 L of 0.1 M NaOH, the samples were incubated at 65°C for 30 min to hydrolyze the remaining RNA. Following the addition of 25 L of 1 M Tris (pH 8.0), the samples were purified using Bio-Rad 6 purification columns (Hercules, CA). This resulted in 5×10^6 to 3×10^7 cpm per reaction[24].

Hybridization & Scanning

cDNA microarrays were pre-hybridized in a hybridization buffer containing 4.0 mL Microhyb (Research Genetics), 10 L of 10 mg/mL human Cot 1 DNA (Life Technologies), and 10 L of 8 mg/mL poly dA (Pharmacia, Peapack, NJ). Both Cot 1 and poly dA were denatured at 95°C for 5 min prior to use. After 4 h of prehybridization at 42°C, approximately 10^7 cpm/mL of heat-denatured (95°C, 5 min) probes were added and incubation continued for 17 h at 42°C. Hybridized arrays were washed three times in 2X SSC and 0.1 % SDS for 15 min at room temperature. The microarrays were exposed to phosphoimager screens for 1-5 days, and the screens were then scanned in a FLA-8000 (Fuji Photo Film Co., Japan) at 50-m resolution[24,25].

Data Analysis

The data were normalized with Z scores by subtracting each average of gene intensity and dividing with each standard deviation. A Z score represents the variability from the average intensity, expressed in units of standard deviation, for each of the 684 genes. A Z score provides flexibility to compare different sets of microarray experiments by adjusting differences in hybridization intensities. Gene expression differences between microarray experiments were calculated by comparing Z score differences for the same genes in different microarrays, which allowed gene expression to be compared in different samples. Z differences were calculated by subtracting the Z score of the control from each Z score of the control sample. These Z differences were further normalized to distribute their position by dividing with the average z difference and dividing with the standard deviation of the z differences. These distributions represented Z ratio values and allowed efficient comparisons to be made between microarray experiments[24]. Scatter plots of intensity values were produced in Spotfire (Spotfire, Inc., Cambridge, MA)[26]. Color overlay images were produced in Adobe Photoshop 5.0 (Adobe Systems, Inc., San Jose, CA). Cluster analysis was performed on Z-transformed microarray data by using two programs available as shareware from Michael Eisen's laboratory (http://rana.lbl.gov). Clusterings of changes in gene expression were determined by using public domain Cluster based on pair wise complete-linkage cluster analysis[27]. Raw gene expression, log values, and Z scores were averaged and are expressed as means ± standard deviation. Raw gene expression, log values and Z scores were averaged and are expressed as means ± standard deviation.

Statistical Analysis

Statistical analyses were performed using the SPSS Statistical Analysis System (Version 12.0). Data were expressed as means ± S.D or frequencies (percentage). All statistical tests were two-sided. Chi-square analysis was used to exam the distribution of age and smoking habits between groups, and Student's t-test was performed to determine the significant difference of urinary Naphthol between groups. Differences were considered significant at the level of $p < 0.05$.

Acknowledgements

We thank Dr. Yoon S. Cho-Chung (Cellular Biochemistry Section, Basic Research Laboratory, CCR, NCI, NIH, Bethesda, MD) and Dr. Kevin G. Becker (DNA Array Unit, NIA, NIH, Baltimore, MD) for valuable advices on cDNA microarray. This study was supported by a grant from Medical Research Center for Environmental Toxicogenomic and Proteomics, funded by Korea Science and Engineering Foundations and Ministry of Science & Technology, a grant from the Korea Health 21 R&D Project, Ministry of Health & Welfare, Republic of Korea (Hmp-00-GN-01-0002 & KPGRN-R-04), a Korea Institute of Science & Technology Evaluation and Planning (KISTEP) and Ministry of Science & Technology (MOST), Korean government, through its National Nuclear Technology Program, and a grant No. R01-2001-000-00212-0 from the Basic Research Program of the Korea Science & Engineering Foundation.

References

1. Vineis, P. *et al.* Tobacco and cancer: recent epidemiological evidence. *J Natl Cancer Inst.* **96**, 99-106 (2004).
2. Hainaut, P. & Pfeifer, G.P. Patterns of p53 G →T transversions in lung cancers reflect the primary mutagenic signature of DNA-damage by tobacco smoke. Carcinogenesis **22**, 367-374 (2001).
3. Gebremichael, A. *et al.* Ah-receptor-dependent modulation of gene expression by aged and diluted sidestream cigarette smoke. *Toxicol Appl Pharmacol.* **141**, 76-83 (1996).
4. Ortlepp, J.R. *et al.* Additive effects of the chemokine receptor 2, vitamin D receptor, interleukin-6 polymorphisms and cardiovascular risk factors on the prevalence of myocardial infarction in patients below 65 years. *Int J Cardiol.* **105**, 90-95 (2005).
5. Seifart, C. *et al.* TNF-alpha-, TNF-beta-, IL-6-, and IL-10-promoter polymorphisms in patients with chronic obstructive pulmonary disease. Tissue Antigens **65**, 93-100 (2005).
6. Krischek, B. & Inoue, I. The genetics of intracranial aneurysms. *J Hum Genet.* **51**, 587-594 (2006).
7. Sasco, A.J., Secretan, M.B. & Straif, K. Tobacco smoking and cancer: a brief review of recent epidemiological evidence. Lung Cancer **45** Suppl 2, S3-9 (2004).
8. DeMarini, D.M. Genotoxicity of tobacco smoke and tobacco smoke condensate: a review. *Mutat Res.* **567**, 447-474 (2004).
9. Nagaraj, N.S. *et al.* Cigarette smoke condensate induces cytochromes P450 and aldo-keto reductases in oral cancer cells. *Toxicol Lett.* **165**, 182-194 (2006).
10. Hecht, S.S. Human urinary carcinogen metabolites: biomarkers for investigating tobacco and cancer. Carcinogenesis **23**, 907-922 (2002).
11. Kellen, E. *et al.* Does occupational exposure to PAHs, diesel and aromatic amines interact with smoking and metabolic genetic polymorphisms to increase the risk on bladder cancer?; The Belgian case control study on bladder cancer risk. *Cancer Lett.* (2006).
12. Preuss, R. *et al.* Current external and internal exposure to naphthalene of workers occupationally exposed to polycyclic aromatic hydrocarbons in different industries. *Int Arch Occup Environ Health* **78**, 355-362 (2005).
13. Serdar, B., Egeghy, P.P., Gibson, R. & Rappaport, S.M. Dose-dependent production of urinary naphthols among workers exposed to jet fuel (JP-8). *Am J Ind Med.* **46**, 234-244 (2004).
14. Nagy, B. *et al.* Lymphotoxin beta expression is high in chronic lymphocytic leukemia but low in small lymphocytic lymphoma: a quantitative real-time reverse transcriptase polymerase chain reaction analysis. *Haematologica* **88**, 654-658 (2003).
15. Reeves, R. & Adair, J.E. Role of high mobility group (HMG) chromatin proteins in DNA repair. DNA Repair (Amst) **4**, 926-938 (2005).
16. Cheung, E.C. *et al.* Dissociating the dual roles of apoptosis-inducing factor in maintaining mitochondrial structure and apoptosis. *Embo J.* **25** (17), 4061-4073 (2006).
17. Taviaux, S.A., Vincent, S., Fort, P. & Demaille, J.G. Localization of ARHG, a member of the RAS homolog gene family, to 11p15.5-11p15.4 by fluorescence in situ hybridization. Genomics **16**, 788-790 (1993).
18. Ishii, Y. *et al.* Mouse brains deficient in neuronal PDGF receptor-beta develop normally but are vulnerable to injury. *J Neurochem.* **98**, 588-600 (2006).
19. Coulthard, M.G. *et al.* Adenovirus-mediated transfer of a gene encoding acyloxyacyl hydrolase (AOAH) into mice increases tissue and plasma AOAH activity. *Infect Immun.* **64**, 1510-1515 (1996).
20. Georgopoulou, N. *et al.* BM88 is a dual function molecule inducing cell cycle exit and neuronal differentiation of neuroblastoma cells via cyclin D1 down-regulation and pRB hypo-phosphorylation. *J Biol Chem.* (2006).
21. Brierley, M.M. & Fish, E.N. Stats: multifaceted regulators of transcription. *J Interferon Cytokine Res.* **25**, 733-744 (2005).
22. Kim, H. *et al.* Assay of 2-naphthol in human urine by high-performance liquid chromatography. *J Chromatogr B Biomed Sci Appl.* **734**, 211-217 (1999).
23. DeRisi, J. *et al.* Use of a cDNA microarray to analyse gene expression patterns in human cancer. *Nat Genet.* **14**, 457-460 (1996).
24. Vawter, M.P. *et al.* Application of cDNA microarrays to examine gene expression differences in schizophrenia. *Brain Res Bull* **55**, 641-650 (2001).
25. Park, G.H. *et al.* Genome-wide expression profiling of 8-chloroadenosine- and 8-chloro-cAMP-treated human neuroblastoma cells using radioactive human cDNA microarray. *Exp Mol Med.* **34**, 184-193 (2002).
26. Tanaka, T.S. *et al.* Genome-wide expression profiling of mid-gestation placenta and embryo using a 15,000 mouse developmental cDNA microarray. *Proc Natl Acad Sci USA* **97**, 9127-9132 (2000).
27. Eisen, M.B., Spellman, P.T., Brown, P.O. & Botstein, D. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA* **95**, 14863-14868 (1998).