CANCER GENOMICS & PROTEOMICS *4*: 349-358 (2007)

Review

Molecular Profiling for the Identification of New Biomarkers in Smokers and Cancer Patients

KAJ ALBERMANN[1], STEPHAN LANG[2] and REINHARD ZEIDLER[3]

[1]*Biomax Informatics AG, D-82152 Martinsried,*
[2]*University of Essen, ENT-Department, D-45122 Essen,*
[3]*University of Munich, ENT-Department, D-81377 Munich, Germany*

Abstract. *Smoking is associated with different serious diseases, including cancer. The fact that only a minority of smokers develops tobacco-associated diseases suggests the contribution of other individual factors, which are still far from understood. New technologies that can be referred to as 'molecular profiling' allow for investigating the deregulation of thousands of genes simultaneously. Numerous such studies have investigated in vitro and in vivo the effects of smoking in different cell types aiming at a better understanding of smoking-induced diseases and the detection of new biomarkers of exposure and harm. This review is a short survey of these investigations and how they have contributed to the detection of new biomarkers and to a better understanding of smoking-induced harm.*

There is compelling evidence that cigarette smoking is associated with a plethora of diseases, notably cancer of the aerodigestive tract and chronic obstructive pulmonary disease (COPD). Smoking probably contributes to almost 30% of human cancers in the developed world, and it is estimated that more than 90% of lung cancer cases are attributable to tobacco abuse. The fact that only a minority of smokers develops cancer or other severe smoking-associated diseases indicates that the process of tobacco-induced diseases is complex and dependent on secondary individual factors such as nutrition, genetic polymorphism, life style, or exposure to other environmental toxins. Therefore, the identification of reliable biomarkers of exposure and harm would help to identify individuals at high risk and contribute to our understanding of smoking-associated diseases.

Correspondence to: Reinhard Zeidler, Ph.D., c/o GSF-Forschungszentrum, Marchioninistr. 25, 81377 Munich, Germany, Tel: +4989 7099296, Fax: +4989 7099-225, e-mail: zeidler@gsf.de

Key Words: Nicotine, profiling, expression profiling, tobacco smoking, review.

Recently, a number of new technologies have emerged for monitoring global cellular responses to perturbation due to disease, drug treatment, or toxicity. One can group these technologies by the type of molecules being measured. Serial analysis of gene expression (SAGE) and DNA microarrays allow systematic detection of changes in the mRNA levels of thousands of genes simultaneously. Different proteomic approaches allow determination of the expression, post-translational modification, localization, or interaction of proteins. In particular, proteomic analysis combining two-dimensional electrophoresis (2D-PAGE) and mass spectrometry (MS) as well as database searches provide a powerful high-throughput method for screening proteins responsive to a variety of stimuli.

The common objective of all these approaches is to capture unique fingerprints of molecular changes and to interpret them with the goal of class discovery, comparison or prediction. The identified fingerprints reflect cumulative responses of complex molecular interactions. Given the high interindividual variation as well as molecular differences in carcinogenesis, distinct molecular and/or immunological signatures may significantly contribute to improving the validity and reliability of biomarkers of exposure and harm.

A growing number of published *in vivo* and *in vitro* studies using SAGE, DNA microarray or proteomics strategies investigate the effects of tobacco smoke or nicotine on gene and/or protein expression profiles allowing the classification of lung cancer, or for prediction of patient survival. Currently there are no such studies using metabolomics and/or metabonomics approaches, but these methods are being applied with increasing frequency in other related areas. Examples are the evaluation of physiological influences on biofluid composition (1, 2) or in the diagnosis of liver or ovarian cancer (3, 4). These techniques will probably be applied to investigations of topics related to this review in the near future.

Table I. *Overview of in vitro and animal model studies of nicotine and cigarette smoke effects on gene expression.*

Organism	Model	Treatment	No. Genes studied	Technology	Reference
Human	Cell line	H_2O_2, smoke	9,600	EST membranes	(6, 7)
Human	Cell line	Nicotine	5,000	cDNA microarray	(8)
Human	Cell line	Nicotine	4,132	cDNA filters	(5)
Mouse	Cell line	Cigarette smoke	513	cDNA glass chip	(9)
Mouse	Tissue	Cigarette smoke	746	cDNA microarray	(10)
Mouse	Tissue	Cigarette smoke	1,185	cDNA microarray	(11)
Rat	Tissue	Cigarette smoke	2,031	cDNA microarray	(12)
Mouse	Tissue	Cigarette smoke	36,000	Oligo microarray	(13)
Mouse	Tissue	Cigarette smoke, LPS	39,000	Oligo microarray	(14)

Gene Expression Studies

In vitro and animal model studies of nicotine and cigarette smoke effects on gene expression. A number of studies investigated the effects of nicotine and cigarette smoke on gene expression in animal models or in human cell lines (Table I).

Zhang *et al.* (5) performed the first systematic study describing the differential gene expression related to smoking. They measured nicotine-induced changes in gene expression in the human coronary artery endothelial cell line HCAEC. Expression changes detected included up-regulation of type II phospatidylinositol-4-phosphate kinase and a decrease in diacylglycerol kinase suggesting an activation of the inositol phospholipid pathway by nicotine. Moreover, the transcription factors CREB and NF-kappaB were detected to be up- and down-regulated, respectively.

Several other studies investigated the effect of nicotine or cigarette smoke on cultured cells or in rodent or mouse models of smoking. Dunckley *et al.* (8) investigated the effects of nicotine exposure in human SH-SY5Y neuroblastoma cells. The expression of 17 genes was significantly altered following nicotine treatment. These genes can be functionally classified into transcription factors, protein processing, RNA binding, and plasma membrane-associated proteins. Bosio *et al.* (9) exposed mouse SWISS 3T3 cells to non-toxic concentrations of aqueous extracts of cigarette smoke. After 4-8 h of treatment, they observed a distinct expression pattern of up- and down-regulated genes. Up-regulated genes were mainly antioxidant response genes, transcription factors, cell cycle related genes, and genes coding for inflammatory/immune-regulatory response mediators. Yoneda *et al.* (6, 7) used primary human bronchial epithelial and immortalized human bronchial epithelial HBE1 cells to investigate gene expression and regulation associated with smoke- and hydrogen peroxide-induced injury and repair. In total, 22 genes were found to be induced in smoke- and H2O2-exposed cells. A time-course Northern blot study of 14 of these genes identified at least three phases of change in gene expression. The first immediate up-regulation was seen 1 h after exposure, including *MKP-1, mdm-2*, and *bcl-2*. The expression of the genes of the second phase, *HSP40, HSP70, HSP90a*, and *ubiquitin p62*, was transient and returned to normal levels 15 h after exposure. The third phase occurred 10 h after exposure and included *glutaredoxin, ferritin light chain, dihydrodiol dehydrogenase, MMP-1*, and *SPRR1B*.

Gebel *et al.* (12) analyzed the *in vivo* effects of cigarette smoke on gene expression in respiratory nasal epithelium (RNE) and lungs in a rat model. They exposed animals either acutely or subchronically to mainstream cigarette smoke (CS) with death immediately or 20 h after exposure. The number of genes differentially expressed by >2-fold was higher in rats immediately killed after exposure, with the gene most strongly differentially expressed being cyp1a1, which was induced 176-fold in the lung and 142-fold in the RNE of subchronically exposed rats. The genes of the RNE of rats exposed once to cigarette smoke showed most evident differential gene expression characterized by the significant up-regulation of genes coding for oxidative stress-responsive and phase II drug-metabolizing enzymes. The strength of expression of these genes was significantly reduced after 3 weeks of exposure suggesting for an adaptive response to cigarette smoke. In contrast, the induced expression for genes encoding phase I drug-metabolizing enzymes (cytochrome P450, cyp1a1; aldehyde dehydrogenase-3) was not altered after 3 weeks of exposure and showed similar levels in lungs and RNE. Almost all changes in gene expression returned to normal after 20 h in a CS-free environment.

Izzotti *et al.* (10) investigated the transcriptional effects of cigarette smoke in Swiss albino mouse fetal livers after transplacental exposure to environmental CS throughout the pregnancy period. Exposure up-regulated 116 genes involved in metabolism, response to oxidative stress, DNA and protein repair, and signal transduction. Additionally, a number of pro-apoptotic genes, cell cycle regulators, hypoxia-related genes, oncogenes, and leukocyte proliferation and differentiation receptors were found to be stimulated. Only one gene, encoding for multiplication stimulating polypeptide was found to be down-regulated after exposure to ECS.

Another study conducted by Izzotti *et al.* (15) studied the expression of cancer-related genes in the lungs of mice carrying a dominant-negative germ-line *p53* mutation. In mice exposed to ECS, this mutation resulted in a lack of induction of pro-apoptotic genes and in overexpression of genes involved in cell proliferation, signal transduction, angiogenesis, inflammation and immune response. In both wild-type and mutant mice, ECS had little effect on the expression of oncogenes, tumor suppressor genes, and DNA repair genes.

Rangasamy *et al.* (13) studied the effect of cigarette smoke on the induction of pulmonary emphysema in mice deficient in the *Nrf2* gene, a transcription factor involved in regulation of detoxification and antioxidant genes. Mice with disrupted *Nrf2* showed more pronounced and earlier onset of emphysema, accompanied with more severe bronchoalveolar inflammation. By microarray analysis, pulmonary expression profiles of cigarette smoke-exposed $Nrf2^{-/-}$ and $Nrf2^{+/+}$ animals were compared. About 50 antioxidant and cytoprotective Nrf2-dependent genes could be identified, which are suspected of providing resistance against the development of emphysema.

Meng *et al.* (14) studied differentially expressed genes of mice exposed to a combination of lipopolysaccharide (LPS) and cigarette smoke to evaluate the use of this LPS-compromised mouse model as a model for investigation of COPD. They identified a number of genes and functional modules that may serve as potential biomarkers for monitoring COPD progression.

In vivo human studies of cigarette smoke effects on gene expression in epithelial cells of the pulmonary airways. To date, few microarray studies have been performed in humans to determine the effect of smoking on gene expression in epithelial cells of the pulmonary airways (16-19). An overview of these studies is given in Table II.

Hackett *et al.* (16) studied the expression of antioxidant-related genes in airway epithelium cells in order to identify candidate genes linked to the risk for chronic bronchitis in association with cigarette smoking. From each subject in smokers and non-smoker cohorts two independent samples of airway epithelium cells were analyzed. Significant up-regulation of 16 out of 44 antioxidant-related genes was detected in smokers, whereas none of the antioxidant-related genes were found to be down-regulated. Analysis of the variability in expression of the antioxidant-related genes within and among individual smokers revealed a marked inter-individual variability for many of the upregulated genes. Determination of the inter-individual variability in expression levels among smokers demonstrated that the extent of inter-individual variability was greater for the 16 up-regulated genes. For one gene, glutathione peroxidase 3, a significant correlation between the gene expression level and smoking history was revealed ($r^2=0.32$, $p=0.024$).

Table II. *Overview of in vivo human studies of nicotine and cigarette smoke effects on gene expression in cells of pulmonary airways.*

Samples	No. of genes	Study population		Reference
		Smoker	Nonsmoker	
Airway epithelial cells	5,000	13	9	(16)
Airway epithelial cells	6,800	14	7	(17)
Airway epithelial cells	22,500	34 (current) 18 (former)	23	(18, 217)
Alveolar macrophages	5,600	5	5	(20)

Kaplan *et al.* (17) compared gene expression patterns in airway epithelium of healthy 20 pack-year smokers *versus* nonsmokers. In total, 47 genes were significantly up- or down-regulated in smokers ($p<0.01$). These included genes known to be affected by cigarette smoke such as genes involved in antioxidant mechanisms, xenobiotic detoxification and activation, and genes whose expression is altered as part of the normal epithelial airway response to cigarette smoke . In addition, a number of genes not previously linked to cigarette smoke exposure were significantly altered, *e.g.* intermediary metabolism genes (transaldolase, transketolase, phosphogluconate dehydrogenase, glucose 6-phosphate dehydrogenase, and all genes of the pentose phosphate pathway), genes related to carcinogenesis (hevin, tenascin-C) as well as the transcription factor pirin. The paternally imprinted gene H19 was found to have dramatically elevated RNA levels in smokers compared to nonsmokers ($p<0.00001$). The authors also showed that the up-regulation of *H19* was not due to loss of imprinting which is known to be associated with lung cancer. Thus, it was speculated that cigarette smoking induces up-regulation of the active H19 allele with eventual temporal progression to loss of imprinting as the burden of smoking increases, paralleling the sequential pathological transition of the airway epithelium.

Spira *et al.* investigated the effect of cigarette smoke in human pulmonary airway epithelial cells (18, 19). The study described gene subsets expressed in large-airway epithelial cells of healthy never smokers, genes whose expression was altered by cigarette smoking, genes whose expression correlated with cumulative pack-years of smoking, and genes whose smoking-induced expression changes were reversible when smoking was stopped. In total, 97 genes were found to be differentially expressed between current and never smokers ($p<1.06\text{x}10^{-5}$), with 68 genes being up-regulated in current smokers. Greatest increases were found in genes coding for xenobiotic functions, antioxidants, and genes involved in electron transport . Additionally cell adhesion molecules and putative oncogenes were also increased. Genes reduced in current smokers are generally involved in regulation of inflammation or are putative

tumor suppressor genes. Interestingly, three current smokers failed to show increased expression of a number of redox-related and xenobiotic genes and thus resembled never smokers. Genes for which induced gene expression could be correlated ($p<3.1x10^{-6}$) with pack-years included cystatin, involved in tumor growth and inflammation; HBP17, enhancing fibroblast growth factor activity; and BRD2, a transcription factor involved in activation of cell cycle genes. Genes showing reduced expression correlating with pack-years included genes involved in DNA repair. The authors also studied the effects of smoking cessation for the 97 genes distinguishing current smokers from never smokers. Former smokers discontinuing smoking <2 years before the study tend to cluster with current smokers, whereas those discontinuing >2 years before the study group cluster more closely to never smokers. Genes showing reversible expression were mainly drug-metabolizing and antioxidant genes. However, thirteen genes did not return to normal levels, even when subjects discontinued smoking for 20-30 years ($p<9.8x10^{-4}$). These genes included potential tumor suppressor genes (*TU3A, CX3CL1*) being permanently reduced, several putative oncogenes (*CEACAM6, HN1*) and three metallothionein genes, which were permanently increased.

Gene expression studies on lung cancer. A number of studies have been conducted in order to investigate the patterns of differential gene expression in lung cancer. Three studies have subclassified lung adenocarcinoma based on gene expression profiles with microarrays of large cohorts and studied the correlation of expression patterns with pathological clinical parameters (21-24). These studies observed a significant heterogeneity of lung adenocarcinoma, but different and overlapping conclusions on important subsets and genes were reached for disease classification and prognosis. Two other studies (22, 23) were able to separate squamous cell carcinoma, small cell lung carcinoma (SCLC) and adenocarcinoma by specific gene expression profiles.

Two studies (22, 23) used gene expression profiling on SCLC cell lines, resected SCLC tumors, xenograft tumors and cell lines of different normal tissues to search for genes that show a specific gene expression in tumor cells. Hierarchical cluster analysis revealed that the cell lines grouped into four clusters were clearly distinct from the cluster formed by the normal tissues. Two subclusters distinguished between classical and variant SCLC type with a number of genes showing differential gene expression. All xenografts clustered closest to the cell lines from which they originated and showed almost identical expression profiles to the cell culture cells. Many genes previously known to be highly expressed in SCLC, including neuroendocrine markers, oncogenes, cell proliferation and cell division genes, were confirmed. Additionally, a number of genes not previously associated with SCLC were identified. Most of them were expressed not only in SCLC cell lines and xenografts but also in the resected SCLC tumors.

Another group of studies analyzed differential gene expression in lung adenocarcinomas of smokers and nonsmokers (25-27). One study (22, 23) compared gene expression profiles in resected samples of nonmalignant lung tissue and tumor tissue of smokers and nonsmokers with adenocarcinoma. Four times as many genes changed expression in the transition from non-involved lung to tumor in nonsmokers as in smokers. This suggested that adenocarcinoma in nonsmokers evolves locally, whereas in smokers it evolves in a field of genetically altered tissue. Some similarities in gene expression between smokers and nonsmokers were found, but differences were significant, pointing to different pathways involved in cell transformation and tumor formation. Multidimensional scaling revealed that noninvolved lungs of smokers group with tumors rather than non-involved lungs of nonsmokers.

Miura *et al.* (25) studied tobacco smoking- and prognosis-related molecular profiles of lung adenocarcinoma. Forty-five genes were identified delineating smokers and nonsmokers ($p=0.0025$), of which 30 were already known genes and 15 were expressed sequence tags (ESTs) or genes of unknown function. Additionally, 27 genes (18 known and 9 unknown genes) were differentially expressed in survivors and nonsurvivors 5 years after surgery ($p<0.0025$).

The objective of another group of studies was to identify potential biomarkers and targets with diagnostic and therapeutic value (28-30). Sugita *et al.* (28, 29) used a combined approach using oligonucleotide and tissue microarrays. First they searched for biomarkers overexpressed in NSCLC and SCLC tumor cell lines. In a second step, these data were validated by immunohistologically testing a tissue microarray containing NSCLC clinical samples. Of the 20 most highly expressed genes in the tumor cell lines, 6 were cancer/testis antigens (CTAs) including 5 MAGE-A subfamily members and NY-ESO-1. The expression of these CTAs was concluded to be a useful marker for the early detection of lung cancer because of their high specificity for central airway squamous and small cell carcinomas. Feng and coworkers identified another potential biomarker of early stage human lung carcinoma (28). They screened for differentially expressed genes in normal human bronchial epithelial (NHBE) cells and in the tumorigenic cell line 1170-I. Beside other genes, they identified the S100A2 gene to be differentially expressed. Immunohistochemical staining identified the S100A2 protein in most tested normal and hyperplastic lung tissues, whereas it was absent from the majority of metaplastic lung tissues, squamous cell carcinoma and adenocarcinoma. The authors concluded that expression of S100A2 is suppressed early during lung carcinogenesis and that its loss may be a

biomarker of early stage disease. Gordon *et al.* (31) developed a prognostic test for patients with resected stage I lung adenocarcinoma using gene expression profiling data of two other published studies (21, 22). Fukumoto *et al.* (30) used microarray analysis to search for genes overexpressed in squamous cell carcinomas (SCC). Seven genes were identified, including the aldo-keto reductase *AKR1B10*. Further immunohistochemical analysis revealed *AKR1B10* to be overexpressed in most cases of investigated SCC and in many adenocarcinomas analyzed from smokers. Thus, *AKR1B10* might be involved in tobacco-related carcinogenesis and is a potential diagnostic marker specific for NSCLCs of smokers.

Only recently, the group of Spira and co-workers identified an 80-gene biomarker profile distinguishing smokers with and without lung cancer. The investigators used the fact that cigarette smoke creates a field of injury throughout the airway, and determined a set of differentially expressed genes in histologically normal large-airway epithelial cells (32). A validation set consisting of 35 biopsies revealed an approximate 90% sensitivity of their biomarker to identify stage 1 lung cancer. Combining cytopathology of lower airway cells with the biomarker yielded 95% sensitivity and a 95% negative predictive value. A functional classification of genes that differentiate lung cancer patients from smokers revealed that many genes up-regulated in cancer patients belong to the categories inflammation/immune function (like *IL8*, *CD55*), cell cycle (*TOB1*, *PPBP*) and cytoskeleton/cell adhesion (*DMD*, *NELL2*). Almost no genes of these categories were expressed at lower levels in cancer patients as compared to smokers without cancer. In contrast, differentially expressed genes belonging to the categories antioxidant, ubiquitinilation and DNA-repair were all down-regulated in cancer patients. These results support the notion that smoking provokes a cancer-specific airway-wide injury and concomitant cancer-specific alterations in gene expression that might precede the development of lung cancer by years.

Gene expression studies on COPD. Chronic obstructive pulmonary disease (COPD) is a slowly progressive and irreversible disorder with functional abnormality of airway obstruction. Largely, cigarette smoking causes COPD, but only 10-15% of all smokers develop clinically apparent disease. Currently two studies have been performed in order to systematically investigate gene expression profiles in patients with COPD.

Ning *et al.* analyzed lung tissue from 14 smokers with moderate COPD (GOLD-2-classification) and 12 control smokers, which exhibited no pulmonary obstruction (GOLD-0) (33). Global gene expression profiles were analyzed in a complementary approach using SAGE and microarrays. Comparison of gene expression patterns between GOLD-2 and GOLD-0 smokers revealed the differential expression of 327 genes. Genes overexpressed in GOLD-2 smokers encoded proteins associated with inflammation (*TGF-beta1, CX3CL1, CTGF, CYR61, TNFSF10, IL1R*), cell proliferation inhibitors (CDKN1A, CDC2L1), transcription factors (*FOS, EGR1, KLF2, HEYL, HAX1, ILF3*), and apoptosis-related proteins (*TEGT, TXNL, GRIM19, NCKAP1, BCAP31*). Genes found underrepresented in GOLD-2 smoker lungs were the ECM genes *COL-1A1, -3A1, -4A1, -6A1*, and *-18A1*. Additionally a number of genes involved in angiogenesis were differentially expressed in GOLD-2 *versus* GOLD-0 smokers. The expression of *EGR-1, CTGF,* and *CYR61* was also analyzed in independent lung tissue from six GOLD-2 and six GOLD-0 smokers. This analysis confirmed the differential expression of these genes. The same three genes were also up-regulated in severe COPD patients (GOLD-4; n=3).

Spira *et al.* performed a second gene expression profiling study on COPD (18). They investigated differentially expressed genes in lung tissues of smokers being severely emphysematous (n=20) and normal or mildly emphysematous (n=14). Due to the limited number of samples, several class prediction algorithms were used to identify genes whose expression in the lung samples allowed the distinction between severe emphysema from no or mild emphysema. In total 102 genes were identified by at least four of the used algorithms. A total of 75 genes were up-regulated in the lung samples of smokers with severe emphysema compared to samples of smokers with no or mild emphysema. A large proportion of these genes were extracellular matrix (ECM)-related genes. Immune and cell signaling-related genes were downregulated in severe emphysema lung samples . The authors analyzed the correlation of expression with pulmonary phenotypic expressions of the disease, such as the diffusion capacity for carbon monoxide (DL_{CO}) and with the forced expiratory volume at 1s (FEV_1). This analysis revealed 92 genes whose level of expression strongly correlated with DL_{CO} ($p<0.001$), 73 genes correlated significantly with FEV_1 ($p<0.001$), and 33 genes correlated with both DLCO and FEV_1. One of the striking findings of this study was the large number of ECM-related genes upregulated in severe emphysema.

A recent study by Heguy *et al.* aimed to identify new genes linked to early events in the molecular pathogenesis of COPD (20). Since cigarette smoke is suggested to activate the release of multiple COPD mediators in alveolar macrophages (AM) they applied gene expression profiling to investigate AM in phenotypically normal, 20 pack-year smokers. In total, 75 genes were identified that showed difference in expression in smokers when compared to healthy nonsmokers. Sixty-nine of these genes have not been previously known to show altered gene expression in AM in association with COPD or smoking. Most of these genes were associated with functions related to immune response/inflammation, signal transduction, cell adhesion, extracellular matrix, and transcription.

Table III. *Overview of protein expression studies.*

Tissue	Technology	Samples	References
Human CYP1A1 transfected amnion epithelial cell line	2D; MS	Cell line treated with B[a]P	(34)
Oral tongue carcinoma	2D; MS	10 patients	(35)
SCC of buccal mucosa	2D; MS	10 patients	(36)
Lung cancer	2D; Western Blot; MS	Serum: 163 cancer, 71 non cancer; Tissue: 82 lung tumors, 16 tumor adjacent normal tissue	(37)
Lung adenocarcinoma	2D; MS; DNA microarrays	93 Lung adenocarcinoma, 10 control lung samples	(38)
Lung adenocarcinoma	2D; MS	90 Lung adenocarcinomas, 10 Control lung samples	(39)
Lung cancer	MS	10 Lung cancer samples, 10 Normal lung tissue samples	(40).
COPD bronchoalveolar lavage fluids	MS	10 COPD smokers, 8 Non COPD smokers, 8 Non COPD nonsmokers	(41)

MS, mass spectrometry, 2D, 2-dimensional gel electrophoresis.

Protein expression studies. Only a few studies related to smoking using proteomic approaches have been performed so far (see Table III).

Using 2D-PAGE and MS, Gao *et al.* identified more than 40 protein spots, which were deregulated upon benz[a]pyrene (B[a]P) treatment in human amnion epithelial cells stably transfected with the *CYP1A1* cDNA (34). Among them, 20 spots were only present in controls, whilst 6 were only detected in B[a]P-treated cells. Statistical analysis indicated that an additional 21 protein spots were significantly changed ($p<0.05$), including 10 upregulated and 11 downregulated proteins following B[a]P treatment. A total of 22 proteins could be identified by MALDI-MS. Many of the deregulated genes were zinc finger proteins and other transcription regulators, indicating a general reaction in the response to B[a]P (34). Of interest, B[a]P also up-regulated PINCH, a component of integrin and growth factor signaling pathways, which is upregulated in the stroma of cancers (42) as well as SNF2L1, which is linked to chromatin remodeling processes and transcription regulation (43).

He *et al.* searched for tumor related proteins in oral tongue carcinoma, an aggressive tumor that particularly affects chronic smokers, drinkers and betel squid chewers. The protein expression profiles of 10 carcinomas and their matched normal mucosal resection margins were examined by 2D-PAGE and MALDI/MS. The study identified a number of tumor-associated proteins including heat-shock protein HSP60, HSP27, alpha B-crystalline, ATP synthase beta, calgranulin B, myosin, tropomyosin and galectin 1, which were consistently found to be significantly overexpressed in malignant tissues. The expression profile portrayed a global protein alteration that appeared specific to oral tongue cancer (35).

A second study conducted by the same group on squamous cell carcinoma (SCC) of the buccal mucosa revealed a number of proteins that were significantly overexpressed in cancer tissues (n=10) as compared to normal mucosa (36). These included glycolytic enzymes, heat-shock proteins, tumor antigens, cytoskeleton proteins, enzymes involved in detoxification and antioxidation systems, and proteins involved in mitochondrial and intracellular signaling pathways. Only one gene (*alpha B-crystalline*) was substantially down-regulated. This study did not identify a specific biomarker for buccal SCC since all identified proteins have been identified in other cancers as well.

Brichory *et al.* used a combined approach using 2D-PAGE and Western-blot analysis to identify proteins that induce an antibody response in patients with lung cancer (37). Proteins from tumor tissue or tumor cell lines were separated by 2D-PAGE, transferred to membranes and hybridized to sera from 64 cancer patients. Proteins that specifically reacted with patient sera were identified by MALDI/MS. Additionally the sera of 99 patients with other types of cancer, and 71 noncancer controls were tested (46 healthy nonsmokers, 15 chronic smokers, 10 chronic lung disease). In sera of 9 out of 64 lung cancer patients, autoantibodies against the neurospecific PGP 9.5 antigen, which was previously described as a marker for non-small cell lung carcinoma and which is identical to ubiquitin thiolesterase, were detected. Subsequently, immunohistochemistry revealed the expression of PGP 9.5 protein in 100% of all small cell carcinomas and neuroendocrine differentiated adenocarcinoma (22/22), 63% of adenocarcinoma (21/33), and 85% of SCC (23/27). In contrast, no PGP 9.5 protein was detectable in healthy mucosa (0/16). PGP 9.5 protein has also been detected in the secreted fraction of the adenocarcinoma cell line A549 and in the serum of 2 lung cancer patients with no detectable PGP 9.5 autoantibodies.

Two 2D-PAGE/MALDI-MS studies by Chen *et al.* analyzed lung adenocarcinomas (38, 39). In the first study, they compared malignant specimens (64 stage I and 29 stage III)

and uninvolved lung samples and also determined associations between overexpressed proteins and clinicopathological features. In addition, microarray analysis was performed on 76 of the tumors and 9 of the normal lung samples. In total isoforms of 9 enzyme proteins showing significantly increased expression in adenocarcinoma were identified. These enzymes included proteins involved in energy-related pathways and antioxidation or detoxification, with GSTM4 being the most consistently overexpressed protein in 96.8% of all tumors. Correlation analysis of overexpression with clinicopathological variables revealed that AOE372 and TPI correlated with tumor differentiation. PPase correlated with tumor subhistology and was increased in bronchial-derived adenocarcinomas ($p=0.02$). Together with TPI and UCHL1, PPase was also found to correlate with smoking history. P4HB was found to be reduced in tumors showing a positive lymphocytic response ($p=0.03$). Comparison of mRNA expression and protein expression values revealed only for GRP58 isoforms significant correlations ($p<0.05$), indicating that increase of protein expression in tumors of these isoforms is due to transcriptional regulation. Analysis of the relative mRNA expression of all these genes between lung adenocarcinomas and normal lung tissues revealed significant increase of mRNA levels for *AOE372, GRP58, P4HB, TPI,* and *UCHL1* ($p<0.005$), but not for *ATP5D, B4GALT, PPase,* and *GSTM4* (38).

The second study by Chen *et al.* analyzed protein profiles associated with patients survival (39). A total of 682 individual protein spots were quantified in 90 lung adenocarcinomas (62 stage I, 28 stage III) and 10 non-neoplastic lung tissues using 2D-PAGE analysis. Forty-six proteins were reported to predict patient survival when used together in protein expression profiles or as individual protein candidates. A leave-one-out cross-validation procedure using the top 20 survival-associated proteins indicated that protein profiles as a whole can predict survival in stage I patients ($p=0.01$). A combined analysis using microarray gene expression and protein expression data revealed that 11 components of the glycolytic pathway were associated with poor survival. For four components an increased protein expression and for seven components, an increased mRNA expression in lung adenocarcinomas and association with poor survival was shown. Only phosphoglycerate kinase (PKG1) was significantly deregulated at both mRNA and protein levels. An elevated level of PKG1 in the serum was also significantly associated with poor survival as found by the analysis of a validation set of 107 patients using ELISA analysis.

Campa *et al.* performed a MALDI/MS study in lung cancer to search for overexpressed proteins as potentially novel molecular targets (40). In total tumor tissue/normal tissue lysate pairs of 40 patients were used. The most statistically significant differentially expressed proteins were identified as MIF and CyP-A. MIF is a cytokine causing T-cell activation, and has been described as being overexpressed in lung adenocarcinoma previously. CyP-A is a member of the immunophilin family and possesses peptidyl *cis-trans* isomerase activity. Both proteins were detected in 27/34 tumor lysates and in only 1/40 nonmalignant tissue lysates (79% sensitivity, 98% specificity). When examined individually, MIF was seen in 30 of 34 tumor lysates and 3 of 40 nonmalignant lysates (88% sensitivity, 93 % specificity), and CyP-A was found in 27 of 34 tumor lysates and in 1 of 40 nonmalignant lysates (79% sensitivity, 98% specificity).

Merkel *et al.* investigated bronchoalveolar lavage fluids from smokers with COPD and identified a number of proteins and peptides as potential diagnostic markers for COPD (41). They applied a combination of surface-enhanced laser desorption/ionization-MS and reversed-phase chromatography with subsequent MALDI-MS or nanoliquid chromatography MS/MS analysis. Beside other differences, they found that smokers with COPD showed increased levels of neutrophil defensins 1 and 2, calgranulin A and B compared to smokers without COPD.

Discussion

Expression profiling studies of cigarette smoking-related diseases have already provided important information about molecular abnormalities, pathways mediating disease progression, and potential diagnostic and therapeutic markers. To date, several hundred-gene expression studies have been published that present analyses of human cancer samples. These analyses identified gene expression signatures for different cancer types and subtypes and uncovered gene expression patterns that correlate with various characteristics of tumors, such as tumor grade, differentiation, metastatic potential, or patient survival. Moreover, novel tissue and serum biomarkers as well as potential therapeutic targets have been identified using these screenings. Unfortunately, many authors tend to present only one interpretation of their data and selectively only reported on a subset of genes that were of interest for their hypothesis.

Taken together, most of the studies at present – as like studies using large clinically important datasets in general – are rarely reported or poorly done so that many of them are not comparable or even reproducible. This is supported by at least two publications that systematically reviewed molecular and biological tumor markers in neuroblastoma (44), and microarray studies with major clinical outcome in cancer (45), respectively. Both reviews concluded that results were often limited by small sample sizes, poor statistical reporting, and large heterogeneity. Ioannidis and coworkers thus recommended the classification of molecular prognostic studies into three phases: phase 1 being early exploratory probing associations, phase 2 being exploratory with extensive analysis, and phase 3 being large confirmatory studies with prestated hypotheses and a

Table IV. *Overview of online databases containing smoking (Smoking), cancer (Cancer) and normal tissue (Normal) gene expression profiling datasets, as well as databases serving as public repositories for expression data (General).*

Study area	Database	Organism	Dataset	Reference
Smoking	SIEGE	H	22,500 genes, 100 patients	(47)
Cancer	Oncomine	H	65 gene expression datasets, >4700 experiments	(48)
Cancer	CGED	H	478 cancer samples, 139 normal samples	(49)
Normal	Oncogenomics	H	18,927 genes, 158 samples, 19 organs, 30 individuals	(50)
Normal	HugeIndex	H	7,000 gene, 59 experiments, 19 tissues	(51)
Nomal	GeneNote	H	62,839 probes, 12 tissues	(52)
Normal	Symatlas	H, M, R	44,775 human transcripts, 79 human tissues, 36,182 mouse transcripts, 61 mouse tissues, 8,000 rat transcripts/ ESTs, 26 rat brain regions	(53-55)
General	GEO	>100	>30,000 datasets of mRNA, genomic DNA, proteomic data	(56)
General	ArrayExpress	35	12,000 datasets from > 300 studies	(57)
General	SMD	34	7,000 datasets	(58)

H, human, M, mouse, R, rat.

precise quantification of the effect. Only studies that had undergone phase 3 should be considered robust enough for use in clinical practice. Currently most studies can considered as phase 1 or some as phase 2.

Beside the poor design of the studies, data communication has also been limited. The Microarray Gene Expression Data Society has proposed a set of guidelines (MIAME) for the reporting of microarray data, and that all microarray data should be deposited in public databases. MIAME, the minimal information for the annotation of a microarray experiment, specifies which data and contextual information should be supplied. Some journals have already begun to endorse or encourage MIAME compliance for publications. Beside MIAME, a standardized data format (MAGE-ML) for exchange of data between microarray data producers and users has been defined (46).

But disclosure of data is not enough unless there is consensus on the appropriate statistical analyses. Many of them are developed on a case-by-case basis and may not even be reproducible by the authors themselves. There is currently no simple answer to this problem; a standardized statistical package would be one solution, which would allow users/readers of such data to quickly repeat an entire analysis. Some published studies are accompanied by supplements to the publication containing a transcript of the respective statistical analyses that allow reproduction of the performed analysis. Another solution to this problem is the generation of databases, hosting large datasets, which support the established common standards for data exchange and comparison. Ideally these databases support effective analysis, data mining, and visualization of the results in a way that users without computational or expression profiling related analytical expertise can access the data.

The full potential of cancer microarray data will only be assessed when the complete datasets of all published studies will be unified, analyzed and made available to the whole cancer research community. There are already a number of databases publicly available which aim for this goal. Three of them (Gene Expression Omnibus, ArrayExpress, and the Stanford Microarray Database) serve as large data repositories for general high-throughput expression data (see Table IV).

References

1 Bollard ME, Stanley EG, Lindon JC, Nicholson JK and Holmes E: NMR-based metabonomic approaches for evaluating physiological influences on biofluid composition. NMR in Biomedicine, 2004.

2 Lenz EM, Bright J, Wilson ID, Morgan SR and Nash AF: A 1H NMR-based metabonomic study of urine and plasma samples obtained from healthy human subjects. J Pharm Biomed Anal *33*: 1103-1115, 2003.

3 Odunsi K, Wollman RM, Ambrosone CB, Hutson A, McCann SE, Tammela J, Geisler JP, Miller G, Sellers T, Cliby W, Qian F, Keitz B, Intengan M, Lele S and Alderfer JL: Detection of epithelial ovarian cancer using (1)H-NMR-based metabonomics. Int J Cancer *113*: 782-788, 2005.

4 Yang J, Xu G, Zheng Y, Kong H, Pang T, Lv S and Yang Q: Diagnosis of liver cancer using HPLC-based metabonomics avoiding false-positive result from hepatitis and hepatocirrhosis diseases. J Chromatogr B Analyt Technol Biomed Life Sci *813*: 59-65, 2004.

5 Zhang S, Day IN and Ye S: Microarray analysis of nicotine-induced changes in gene expression in endothelial cells. Physiol Genomics *5*: 187-192, 2001.

6 Yoneda K, Chang MM, Chmiel K, Chen Y and Wu R: Application of high-density DNA microarray to study smoke- and hydrogen peroxide-induced injury and repair in human bronchial epithelial cells. J Am Soc Nephrol *14*: S284-289, 2003.

7 Yoneda K, Peck K, Chang MM, Chmiel K, Sher YP, Chen J, Yang PC, Chen Y and Wu R: Development of high-density DNA microarray membrane for profiling smoke- and hydrogen peroxide-induced genes in a human bronchial epithelial cell line. Am J Respir Crit Care Med *164*: S85-89, 2001.

8 Dunckley T and Lukas RJ: Nicotine modulates the expression of a diverse set of genes in the neuronal SH-SY5Y cell line. J Biol Chem *278*: 15633-15640, 2003.

9 Bosio A, Knorr C, Janssen U, Gebel S, Haussmann HJ and Muller T: Kinetics of gene expression profiling in Swiss 3T3 cells exposed to aqueous extracts of cigarette smoke. Carcinogenesis *23*: 741-748, 2002.

10 Izzotti A, Balansky RM, Cartiglia C, Camoirano A, Longobardi M and De Flora S: Genomic and transcriptional alterations in mouse fetus liver after transplacental exposure to cigarette smoke. FASEB J *17*: 1127-1129, 2003.

11 Izzotti A, Cartiglia C, Longobardi M, Balansky RM, D'Agostini F, Lubet RA and De Flora S: Alterations of gene expression in skin and lung of mice exposed to light and cigarette smoke. FASEB J *18*: 1559-1561, 2004.

12 Gebel S, Gerstmayer B, Bosio A, Haussmann HJ, Van Miert E and Muller T: Gene expression profiling in respiratory tissues from rats exposed to mainstream cigarette smoke. Carcinogenesis *25*: 169-178, 2004.

13 Rangasamy T, Cho CY, Thimmulappa RK, Zhen L, Srisuma SS, Kensler TW, Yamamoto M, Petrache I, Tuder RM and Biswal S: Genetic ablation of Nrf2 enhances susceptibility to cigarette smoke-induced emphysema in mice. J Clin Invest *114*: 1248-1259, 2004.

14 Meng QR, Gideon KM, Harbo SJ, Renne RA, Lee MK, Brys AM and Jones R: Gene expression profiling in lung tissues from mice exposed to cigarette smoke, lipopolysaccharide, or smoke plus lipopolysaccharide by inhalation. Inhal Toxicol *18*: 555-568, 2006.

15 Izzotti A, Cartiglia C, Longobardi M, Bagnasco M, Merello A, You M, Lubet RA and De Flora S: Gene expression in the lung of p53 mutant mice exposed to cigarette smoke. Cancer Res *64*: 8566-8572, 2004.

16 Hackett NR, Heguy A, Harvey BG, O'Connor TP, Luettich K, Flieder DB, Kaplan R and Crystal RG: Variability of antioxidant-related gene expression in the airway epithelium of cigarette smokers. Am J Respir Cell Mol Biol *29*: 331-343, 2003.

17 Kaplan R, Luettich K, Heguy A, Hackett NR, Harvey BG and Crystal RG: Monoallelic up-regulation of the imprinted H19 gene in airway epithelium of phenotypically normal cigarette smokers. Cancer Res *63*: 1475-1482, 2003.

18 Spira A, Beane J, Shah V, Liu G, Schembri F, Yang X, Palma J and Brody JS: Effects of cigarette smoke on the human airway epithelial cell transcriptome. Proc Natl Acad Sci USA *101*: 10143-10148, 2004.

19 Spira A, Schembri F, Beane J, Shah V, Liu G and Brody JS: Impact of cigarette smoke on the normal airway transcriptome. Chest *125*: 115S, 2004.

20 Heguy A, O'Connor TP, Luettich K, Worgall S, Cieciuch A, Harvey BG, Hackett NR and Crystal RG: 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 *84*: 318-328, 2006.

21 Beer DG, Kardia SL, Huang CC, Giordano TJ, Levin AM, Misek DE, Lin L, Chen G, Gharib TG, Thomas DG, Lizyness ML, Kuick R, Hayasaka S, Taylor JM, Iannettoni MD, Orringer MB and Hanash S: Gene-expression profiles predict survival of patients with lung adenocarcinoma. Nat Med *8*: 816-824, 2002.

22 Bhattacharjee A, Richards WG, Staunton J, Li C, Monti S, Vasa P, Ladd C, Beheshti J, Bueno R, Gillette M, Loda M, Weber G, Mark EJ, Lander ES, Wong W, Johnson BE, Golub TR, Sugarbaker DJ and Meyerson M: Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. Proc Natl Acad Sci USA *98*: 13790-13795, 2001.

23 Garber ME, Troyanskaya OG, Schluens K, Petersen S, Thaesler Z, Pacyna-Gengelbach M, van de Rijn M, Rosen GD, Perou CM, Whyte RI, Altman RB, Brown PO, Botstein D and Petersen I: Diversity of gene expression in adenocarcinoma of the lung. Proc Natl Acad Sci USA *98*: 13784-13789, 2001.

24 Giordano TJ, Shedden KA, Schwartz DR, Kuick R, Taylor JM, Lee N, Misek DE, Greenson JK, Kardia SL, Beer DG, Rennert G, Cho KR, Gruber SB, Fearon ER and Hanash S: Organ-specific molecular classification of primary lung, colon, and ovarian adenocarcinomas using gene expression profiles. Am J Pathol *159*: 1231-1238, 2001.

25 Miura K, Bowman ED, Simon R, Peng AC, Robles AI, Jones RT, Katagiri T, He P, Mizukami H, Charboneau L, Kikuchi T, Liotta LA, Nakamura Y and Harris CC: Laser capture microdissection and microarray expression analysis of lung adenocarcinoma reveals tobacco smoking- and prognosis-related molecular profiles. Cancer Res *62*: 3244-3250, 2002.

26 Powell CA, Bueno R, Borczuk AC, Caracta CF, Richards WG, Sugarbaker DJ and Brody JS: Patterns of allelic loss differ in lung adenocarcinomas of smokers and nonsmokers. Lung Cancer *39*: 23-29, 2003.

27 Powell CA, Spira A, Derti A, DeLisi C, Liu G, Borczuk A, Busch S, Sahasrabudhe S, Chen Y, Sugarbaker D, Bueno R, Richards WG and Brody JS: Gene expression in lung adenocarcinomas of smokers and nonsmokers. Am J Respir Cell Mol Biol *29*: 157-162, 2003.

28 Feng G, Xu X, Youssef EM and Lotan R: Diminished expression of S100A2, a putative tumor suppressor, at early stage of human lung carcinogenesis. Cancer Res *61*: 7999-8004, 2001.

29 Sugita M, Geraci M, Gao B, Powell RL, Hirsch FR, Johnson G, Lapadat R, Gabrielson E, Bremnes R, Bunn PA and Franklin WA: Combined use of oligonucleotide and tissue microarrays identifies cancer/testis antigens as biomarkers in lung carcinoma. Cancer Res *62*: 3971-3979, 2002.

30 Fukumoto S, Yamauchi N, Moriguchi H, Hippo Y, Watanabe A, Shibahara J, Taniguchi H, Ishikawa S, Ito H, Yamamoto S, Iwanari H, Hironaka M, Ishikawa Y, Niki T, Sohara Y, Kodama T, Nishimura M, Fukayama M, Dosaka-Akita H and Aburatani H: Overexpression of the aldo-keto reductase family protein AKR1B10 is highly correlated with smokers' non-small cell lung carcinomas. Clin Cancer Res *11*: 1776-1785, 2005.

31 Gordon GJ, Richards WG, Sugarbaker DJ, Jaklitsch MT and Bueno R: A prognostic test for adenocarcinoma of the lung from gene expression profiling data. Cancer Epidemiol Biomarkers Prev *12*: 905-910, 2003.

32 Spira A, Beane JE, Shah V, Steiling K, Liu G, Schembri F, Gilman S, Dumas YM, Calner P, Sebastiani P, Sridhar S, Beamis J, Lamb C, Anderson T, Gerry N, Keane J, Lenburg ME and Brody JS: Airway epithelial gene expression in the diagnostic evaluation of smokers with suspect lung cancer. Nat Med *13*: 361-366, 2007.

33 Ning W, Li CJ, Kaminski N, Feghali-Bostwick CA, Alber SM, Di YP, Otterbein SL, Song R, Hayashi S, Zhou Z, Pinsky DJ, Watkins SC, Pilewski JM, Sciurba FC, Peters DG, Hogg JC and Choi AM: Comprehensive gene expression profiles reveal pathways related to the pathogenesis of chronic obstructive pulmonary disease. Proc Natl Acad Sci USA *101*: 14895-14900, 2004.

34 Gao Z, Jin J, Yang J and Yu Y: Zinc finger proteins and other transcription regulators as response proteins in benzo[a]pyrene exposed cells. Mutat Res *550*: 11-24, 2004.

35 He QY, Chen J, Kung HF, Yuen AP and Chiu JF: Identification of tumor-associated proteins in oral tongue squamous cell carcinoma by proteomics. Proteomics *4*: 271-278, 2004.

36 Chen J, He QY, Yuen AP and Chiu JF: Proteomics of buccal squamous cell carcinoma: the involvement of multiple pathways in tumorigenesis. Proteomics *4*: 2465-2475, 2004.

37 Brichory F, Beer D, Le Naour F, Giordano T and Hanash S: Proteomics-based identification of protein gene product 9.5 as a tumor antigen that induces a humoral immune response in lung cancer. Cancer Res *61*: 7908-7912, 2001.

38 Chen G, Gharib TG, Huang CC, Thomas DG, Shedden KA, Taylor JM, Kardia SL, Misek DE, Giordano TJ, Iannettoni MD, Orringer MB, Hanash SM and Beer DG: Proteomic analysis of lung adenocarcinoma: identification of a highly expressed set of proteins in tumors. Clin Cancer Res *8*: 2298-2305, 2002.

39 Chen G, Gharib TG, Wang H, Huang CC, Kuick R, Thomas DG, Shedden KA, Misek DE, Taylor JM, Giordano TJ, Kardia SL, Iannettoni MD, Yee J, Hogg PJ, Orringer MB, Hanash SM and Beer DG: Protein profiles associated with survival in lung adenocarcinoma. Proc Natl Acad Sci USA *100*: 13537-13542, 2003.

40 Campa MJ, Wang MZ, Howard B, Fitzgerald MC and Patz EF, Jr.: Protein expression profiling identifies macrophage migration inhibitory factor and cyclophilin a as potential molecular targets in non-small cell lung cancer. Cancer Res *63*: 1652-1656, 2003.

41 Merkel D, Rist W, Seither P, Weith A and Lenter MC: Proteomic study of human bronchoalveolar lavage fluids from smokers with chronic obstructive pulmonary disease by combining surface-enhanced laser desorption/ionization-mass spectrometry profiling with mass spectrometric protein identification. Proteomics *5*: 2972-2980, 2005.

42 Wang-Rodriguez J, Dreilinger AD, Alsharabi GM and Rearden A: The signaling adapter protein PINCH is up-regulated in the stroma of common cancers, notably at invasive edges. Cancer *95*: 1387-1395, 2002.

43 Langst G and Becker PB: Nucleosome mobilization and positioning by ISWI-containing chromatin-remodeling factors. J Cell Sci *114*: 2561-2568, 2001.

44 Riley RD, Heney D, Jones DR, Sutton AJ, Lambert PC, Abrams KR, Young B, Wailoo AJ and Burchill SA: A systematic review of molecular and biological tumor markers in neuroblastoma. Clin Cancer Res *10*: 4-12, 2004.

45 Ntzani EE and Ioannidis JP: Predictive ability of DNA microarrays for cancer outcomes and correlates: an empirical assessment. Lancet *362*: 1439-1444, 2003.

46 Spellman PT, Miller M, Stewart J, Troup C, Sarkans U, Chervitz S, Bernhart D, Sherlock G, Ball C, Lepage M, Swiatek M, Marks WL, Goncalves J, Markel S, Iordan D, Shojatalab M, Pizarro A, White J, Hubley R, Deutsch E, Senger M, Aronow BJ, Robinson A, Bassett D, Stoeckert CJ, Jr. and Brazma A: Design and implementation of microarray gene expression markup language (MAGE-ML). Genome Biol *3*: RESEARCH0046, 2002.

47 Shah V, Sridhar S, Beane J, Brody JS and Spira A: SIEGE: Smoking Induced Epithelial Gene Expression Database. Nucleic Acids Res *33 Database Issue*: D573-579, 2005.

48 Rhodes DR, Yu J, Shanker K, Deshpande N, Varambally R, Ghosh D, Barrette T, Pandey A and Chinnaiyan AM: ONCOMINE: a cancer microarray database and integrated data-mining platform. Neoplasia *6*: 1-6, 2004.

49 Kato K, Yamashita R, Matoba R, Monden M, Noguchi S, Takagi T and Nakai K: Cancer gene expression database (CGED): a database for gene expression profiling with accompanying clinical information of human cancer tissues. Nucleic Acids Res *33 Database Issue*: D533-536, 2005.

50 Son CG, Bilke S, Davis S, Greer BT, Wei JS, Whiteford CC, Chen QR, Cenacchi N and Khan J: Database of mRNA gene expression profiles of multiple human organs. Genome Res *15*: 443-450, 2005.

51 Haverty PM, Weng Z, Best NL, Auerbach KR, Hsiao LL, Jensen RV and Gullans SR: HugeIndex: a database with visualization tools for high-density oligonucleotide array data from normal human tissues. Nucleic Acids Res *30*: 214-217, 2002.

52 Shmueli O, Horn-Saban S, Chalifa-Caspi V, Shmoish M, Ophir R, Benjamin-Rodrig H, Safran M, Domany E and Lancet D: GeneNote: whole genome expression profiles in normal human tissues. C R Biol *326*: 1067-1072, 2003.

53 Su AI, Cooke MP, Ching KA, Hakak Y, Walker JR, Wiltshire T, Orth AP, Vega RG, Sapinoso LM, Moqrich A, Patapoutian A, Hampton GM, Schultz PG and Hogenesch JB: Large-scale analysis of the human and mouse transcriptomes. Proc Natl Acad Sci USA *99*: 4465-4470, 2002.

54 Su AI, Wiltshire T, Batalov S, Lapp H, Ching KA, Block D, Zhang J, Soden R, Hayakawa M, Kreiman G, Cooke MP, Walker JR and Hogenesch JB: A gene atlas of the mouse and human protein-encoding transcriptomes. Proc Natl Acad Sci USA *101*: 6062-6067, 2004.

55 Walker JR, Su AI, Self DW, Hogenesch JB, Lapp H, Maier R, Hoyer D and Bilbe G: Applications of a rat multiple tissue gene expression data set. Genome Res *14*: 742-749, 2004.

56 Barrett T, Suzek TO, Troup DB, Wilhite SE, Ngau WC, Ledoux P, Rudnev D, Lash AE, Fujibuchi W and Edgar R: NCBI GEO: mining millions of expression profiles – database and tools. Nucleic Acids Res *33 Database Issue*: D562-566, 2005.

57 Parkinson H, Sarkans U, Shojatalab M, Abeygunawardena N, Contrino S, Coulson R, Farne A, Lara GG, Holloway E, Kapushesky M, Lilja P, Mukherjee G, Oezcimen A, Rayner T, Rocca-Serra P, Sharma A, Sansone S and Brazma A: ArrayExpress – a public repository for microarray gene expression data at the EBI. Nucleic Acids Res *33 Database Issue*: D553-555, 2005.

58 Ball CA, Awad IA, Demeter J, Gollub J, Hebert JM, Hernandez-Boussard T, Jin H, Matese JC, Nitzberg M, Wymore F, Zachariah ZK, Brown PO and Sherlock G: The Stanford Microarray Database accommodates additional microarray platforms and data formats. Nucleic Acids Res *33 Database Issue*: D580-582, 2005.

Received June 15, 2007
Revised August 1, 2007
Accepted August 27, 2007