Variability in Small Airway Epithelial Gene Expression Among Normal Smokers

Zeinab Ammous, BS[1], **Neil R. Hackett, PhD**[2], **Marcus W. Butler, MB, BCh, BAO**[3], **Tina Raman, BS**[2], **Igor Dolgalev, BS**[2], **Timothy P. O'Connor, PhD**[2], **Ben-Gary Harvey, MD**[3], and **Ronald G. Crystal, MD**[2,3]

[1]Weill Cornell Medical College in Qatar, Weill Cornell Medical College, New York, New York

[2]Department of Genetic Medicine, Weill Cornell Medical College, New York, New York

[3]Divison of Pulmonary and Critical Care Medicine, Weill Cornell Medical College, New York, New York

Summary

Background—Despite overwhelming data that cigarette smoking causes chronic obstructive pulmonary disease (COPD), only a minority of chronic smokers are affected, strongly suggesting that genetic factors modify susceptibility to this disease. We hypothesized that there are individual variations in the response to cigarette smoking, with variability among smokers in expression levels of protective / susceptibility genes.

Methodology—Affymetrix arrays and TaqMan PCR were used to assess the variability of gene expression in the small airway epithelium obtained by fiberoptic bronchoscopy of 18 normal non-smokers, 18 normal smokers and 18 smokers with COPD.

Results—We identified 201 probesets representing 152 smoking-responsive genes that were significantly up- or down-regulated, and assessed the coefficient of variation in expression levels among the study population. Variation was a reproducible property of each gene as assessed by different microarray probesets and realtime PCR and was observed in both normal smokers and smokers with COPD. There was greater individual variability in smokers with COPD than in normal smokers. The majority of these highly variable smoking responsive genes were in the functional categories of signal transduction, xenobiotic degradation, metabolism, transport, oxidant-related and transcription. A similar pattern of the same highly variable genes was observed in an independent data set.

Conclusions—We propose that there is likely genetic diversity within this subset of genes with highly variable individual to individual responses of the small airway epithelium to smoking, and this subset of genes represent putative candidates for assessment of susceptibility/protection from disease in future gene-based epidemiological studies of smokers' risk for COPD.

Introduction

Despite the overwhelming evidence of cigarette smoking as the major risk factor for the development of chronic obstructive pulmonary disease (COPD), the majority of chronic smokers remain healthy, strongly suggesting that genetic factors modify disease susceptibility to this environmental stress[1-5]. Multiple studies support the concept that there

Correspondence: Department of Genetic Medicine, Weill Medical College of Cornell University, 1300 York Avenue, Box 96, New York, New York 10065, Phone: (646) 962-4363, Fax: (646) 962-0220, geneticmedicine@med.cornell.edu.

are genes that give susceptibility to, or protect from, the stress that smoking places on the airways, and that together, mediate whether the individual smoker does or does not develop COPD[6-9]. However, while progress has been made, the identification of these genes has proven to be a significant challenge, as there are likely many genes involved, complicated by issues such as ancestral diversity, variability of smoking exposures, and the false discovery rate when assessing all 25,000 genes in the human genome[10-14].

The present study is directed toward a novel approach to identify possible candidate genes for susceptibility to smoking-induced lung disease that should be complementary to the strategies of family-based and genome wide association studies[6-17]. Our strategy, derived from the observation of variability of gene expression in the human airway epithelium[18-20], is based on the following concepts: (1) the airway epithelium is the cell population that takes the brunt of the stress of cigarette smoke; (2) disease of the small airway epithelium is an early component of the development of COPD[1-3,21-23]; and (3) the increasing evidence that genetic susceptibility to complex disorders is mediated by polymorphisms in multiple protective/susceptibility genes resulting in over-and under-expression of gene products that together, mediate protection from, or the development of, disease[11-14,24-26]. With the background of our prior studies documenting variability in oxidant-related genes in the large airway epithelium of cigarette smokers[18], we hypothesized that among the genes that are significantly over-or under-expressed in the small airway epithelium of cigarette smokers, there is a subset of genes that exhibit more individual variability in the levels of gene expression in response to smoking than other genes. Further, based on the knowledge that smokers with COPD are (by definition) more susceptible to smoking-induced lung abnormalities than comparable asymptomatic smokers, we hypothesized that, for individual genes, the variability of gene expression in the small airway epithelium of smokers with COPD will be different than that of asymptomatic smokers. Using Affymetrix Human Genome U133 Plus 2.0 arrays to evaluate gene expression of the small airway epithelium of 18 non-smokers and 18 phenotypic normal smokers and 18 current smokers with COPD, the data demonstrates that a subset of genes expressed by the small airway epithelium exhibit a large individual to individual variability in response to smoking. Further, this variability is a reproducible, intrinsic property of this subset of genes confirmed by two independent experimental approaches and in an independent dataset. Interestingly, smokers with COPD have a greater variability of expression of these genes than do asymptomatic smokers. We propose that this subset of genes represent a novel list of candidate genes for future genetic epidemiology studies regarding the risk for susceptibility to or protection from smoking-induced development of COPD.

Methods

Study Population

After signing informed consent, subjects were evaluated in the Weill Medical College of Cornell University NIH General Clinical Research Center under an Institutional Review Board approved clinical protocol. They were assessed by standard history, physical exam, complete blood count, coagulation studies, liver function tests, urine studies, chest X-ray, EKG, and pulmonary function tests. A total of 54 subjects in this study included 3 groups (Table I): normal non-smokers (n=18) and normal smokers (n=18) and smokers with COPD GOLD stages I or II (n=18). Normal non-smokers and normal smokers had no symptoms referable to the lungs, normal lung function and normal chest X-ray. All tests were within the normal ranges for normal non-smokers and normal smokers. Smokers with established COPD included current smokers who met the Global Initiative for Chronic Obstructive Lung Disease (GOLD) criteria for GOLD I or II only[27]. The demographic composition of the three groups was similar with respect to sex (p>0.4) and self-described race (p>0.2). The

mean age for the non-smokers (41±7) was slightly lower than for the smokers (46±5, p<0.05) and for the smokers with COPD (50±6, p<0.05).

Sampling Small Airway Epithelium RNA Extraction and Microarray Assessment

Pure populations of small airway epithelium were obtained by bronchial brushing as previously described[19]. After mild sedation with demerol and versed, and routine anesthesia of the vocal cords and bronchial airways with topical lidocaine, the fiberoptic bronchoscope (Pentax, EB-1530T3) was positioned distal to the opening of the desired lobar bronchus. To obtain small airway epithelial cells, a 1.2 mm diameter brush was advanced approximately 7 to 10 cm distally from the 3rd order bronchial branching. The distal end of the brush was wedged at about the 10th to12th generation branching of the right lower lobe, and small airway epithelial cells were obtained by gently gliding the brush back and forth on the epithelium 5 to 10 times in 10 different locations in the same general area. Cells were detached from the brush by flicking into 5 ml of ice-cold bronchial epithelial basal cell medium (BEBM, Clonetics, Walkersville, MD). An aliquot of 0.5 ml was used for differential cell count and the remainder (4.5 ml) was processed immediately for RNA extraction. Pelleted airway epithelial cells were lysed with the TRIzol reagent (InVitrogen, Carlsbad, CA), and after chloroform extraction the RNA was purified directly from the aqueous phase by RNeasy MinElute RNA purification kit (Qiagen, Valencia, CA). The samples were stored in RNA Secure (Ambion, Austin, TX) and concentration was determined using a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE). An aliquot was evaluated using an Agilent Bioanalyzer (Agilent Technologies, Palo Alto, CA) with three quality control criteria: (1) concentration greater than 0.2 $\mu g/\mu l$; (2) A_{260}/A_{280} ratio from 1.7 to 2.3; and (3) Agilent electropherogram displaying 2 distinct peaks corresponding to the 28S and 18S ribosomal RNA bands at a ratio of 28S/18S of >1.5. RNA processing, hybridizations to test chips and, if quality control was acceptable, to the HG U133 Plus 2.0 microarray were performed according to Affymetrix protocols, and using Affymetrix reagents processed by the Affymetrix GeneChip Fluidics Station 450, and scanned with an Affymetrix GeneChip Scanner 3000 7G (http://www.affymetrix.com/support/technical/manual/expression_manual.affx), as previously described[19]. Overall microarray quality was verified by the following criteria: (1) RNA Integrity Number (RIN) > 7.0; (2) $3'/5'$ ratio for GAPDH < 3; (3) scaling factor range no more than ± 2.5 standard deviations from the mean for all microarrays; and (4) expression level for all 100 housekeeping genes (as defined by Affymetrix, www.affymetrix.com) with coefficient of variation of <40%. For 4 highly variable genes (CYP1B1, UCHL1, HES6, AKR1B10) and two smoking responsive but not highly variable genes (NQO1 and TCF7L1), mRNA levels were also estimated by TaqMan realtime PCR.

Data Analysis and Statistics

The Microarray Suite 5.0 (MAS5) software was used to analyze the data for all samples. Probesets for which less than 50% of samples were scored "Present" were discarded. MAS5-analyzed data was normalized using GeneSpring followed by per-chip and per-gene normalization across all 54 samples. In the initial analysis, changes in gene expression between non-smokers and smokers were considered significant (p<0.05) using unpaired t test with Benjamini-Hochberg correction. An additional criterion of >2-fold change was introduced to confine analysis to the most robust differences between non-smokers and normal smokers. Coefficient of variation was assessed for the normal smokers. Highly variable genes were defined as those smoking-dependent probe sets with coefficient of variation >98th percentile of all probe sets. Functional annotation of genes was carried out using the NetAffx Analysis Center (www.affymetrix.com) to retrieve the Gene Ontology (GO) annotations from the National Center for Biotechnology NCBI databases. All

microarray data has been deposited at the Gene Expression Omnibus (GEO) site (http://www.ncbi.nlm.nih.gov/geo/; accession number GSE 8545).

Comparisons of parameters between groups were performed by ANOVA for continuous variables and by Chi square test for discrete variables. Linear regression was used to assess the correlation of expression level with phenotype, and the reproducibility of coefficient of variation among different phenotypic groups. ANOVA was used to assess the impact of smoking on different functional categories of genes.

An independent data set of gene expression profiles of airway epithelium in smokers from the studies of Spira et al[20] was used to validate our identification of genes with highly variable expression levels in normal smokers (GEO accession number GSE994). In the independent data set, the coefficient of variation was assessed for the normal smokers and highly variable genes were defined as those probe sets with coefficient of variation >98th percentile of all probe sets "present" in 50% of samples.

Results

Small Airway Epithelium

The cell number, purity and differential counts for the samples recovered from the small airway epithelium of the normal non-smokers, normal smokers and smokers with COPD were similar (Table I). A range of 2.5 to 10.1×10^6 cells were recovered from all groups (p>0.08 comparing all samples by ANOVA). There were fewer inflammatory cells (p<0.03) and secretory cells (p<0.01) and more basal cells (p<0.02) in non-smokers than in the other two groups, though the proportion of ciliated and undifferentiated cells were similar (p>0.1 all comparisons).

Smoking-dependent Genes

For the initial analysis, a group of smoking-dependent genes was identified for which the expression levels were significantly higher or lower in the 18 normal smokers than in the 18 demographically matched non-smokers. Of the 25,667 probesets expressed [Affymetrix call of "P" (present)] in >50% of all small airway samples, 201 were identified for which the change in expression level between smokers and non-smokers was >2-fold with a significance level of <0.05 after application of the Benjamini-Hochberg correction for multiple testing (Figure 1). These 201 probesets represented 152 unique named genes. Expression levels for approximately equal numbers of probesets were upregulated (n=96) and downregulated (n=105) by smoking.

Variability in Expression of Smoking-responsive Genes

The variability of expression level in normal smokers was assessed for the 201 probesets identified in the analysis of gene expression in the normal smokers *vs* normal non-smokers. The median coefficient of variation for the smoking responsive genes in normal smokers was 49.2% (5th to 95th percentile 28.6 to 102.5%). By contrast, the median coefficient of variation for all 25,667 probesets expressed was 31.2% (5th to 95th percentile 18.2 to 57.8%). A total of 31 of the 201 of the smoking-responsive probesets, representing 28 named genes, had a coefficient of variation greater than the 98th percentile for all probesets (69.9%). These 31 probesets genes were designated as the "highly variable smoking responsive gene list" (Table II).

To help verify that this list of highly variable smoking responsive genes was not an artifact of assessing expression of a large number of genes, we evaluated an independent data set of gene expression levels in the airway epithelium of smokers based on the work of Spira et

al [20] (gene expression omnibus accession number GSE994). For this data set, which comprise data collected using a previous generation of Affymetrix microarray, 7,625 probe sets were identified as "present" in >50% of the 57 normal smokers and non-smokers. The median coefficient for these probesets in the normal smokers (n=34) was 32.5% (5th to 95th percentile 21.1 to 56.1%). Of the 31 probesets shown in Table II, 20 are also present on the Affymetrix HG U133A microarray used in Spira et al[20]. Of these, 15 have coefficients of variation within the top 2% among all probe sets in this data set (Table II).

To assess possible functional significance of the genes for which there were highly variable probesets in our data set, the 201 variable smoking responsive probesets were categorized, including: xenobiotic metabolism, n=12; transport, n=14; transcription factors and chromatin structure, n=14; signal transduction, n=40; protease, n=10; oxidant-related, n=27; metabolism, n=20; immune and inflammatory response, n=3; cell cycle/cytoskeleton, n=8; and other/ unknown, n=53 (Figure 2). Among the 201 variable smoking-responsive probesets, the extent of variability as assessed by coefficient of variation, was dependent on the category, with the xenobiotic degradation probesets, primarily representing cytochrome p450 genes, having a higher coefficient of variation (mean 75.5%) compared to all other catagories ($p < 0.005$) and the oxidant related genes having a lower coefficient of variation (40.3%) compared to all other categories ($p < 0.005$).

To ensure that the measured variability was a genuine reflection of the mRNA levels, we compared the available replicate probesets for each of the variable genes. There was a strong correlation in the coefficient of variation between replicate measurements with different probesets (Figure 3A; $r^2 = 0.64$, $p < 0.0001$). To further exclude technical factors that may account for the result, the data was also assessed by Robust Multichip Average method, an independent method of determining expression levels from microarray scans. There was a strong correlation of the coefficient of variation by this method and the MAS5.0 method ($r^2 = 0.65$, $p < 0.005$). The data was further confirmed for selected genes by showing a strong correlation between the coefficient of variation determined by TaqMan realtime PCR with that obtained by microarrays (Figure 3B, $r^2 = 0.58$, $p < 0.01$ by linear regression). To examine potential causes of the variable expression level, we assessed if known difference among subjects could account for the changes. However, none of the probesets showed a strong correlation of expression level with any demographic or pulmonary function parameter ($r^2 < 0.3$ for all 201 probesets in independent linear regressions with age, sex, genetic ancestry, smoking history, FEV1, FEV1/FVC, and DLCO).

Similarities and Differences in the Variable Expression Levels Among Normal Smokers and Smokers with COPD

To assess the similarities and differences in the observed variability among normal smokers and smokers with COPD, the coefficient of variation was compared for the 2 groups. There was an overall correlation between the coefficient of variation in the two groups, but a majority (78%, 157/201) of the probesets had a higher coefficient of variation in the smokers with COPD than in the normal smokers (Figure 4A). Analysis of the frequency distribution of the coefficient of variation in the groups of smokers showed that the coefficient of variation for the smoking-responsive genes was greater in the smokers with COPD compared to the normal smokers ($p < 0.0001$, chi square test, Figure 4B) and the coefficients of variation were higher (median of 61.0 *vs* 49.3% for the smokers with COPD, $p < 0.0001$ comparing distribution by paired t test, Figure 4C). Individual examples of highly variable smoking up-regulated genes included those in the category of oxidant-related genes (Figure 5A-C), while there was highly variable suppression of gene expression in the category of signal transduction (Figure 5D-F). In all cases, the expression patterns among individual smokers with COPD most closely resembled the expression patterns in normal smokers and not that of non-smokers.

Discussion

In this study we have assessed individual to individual variability in cigarette smoking-induced changes in mRNA expression levels in the small airway epithelium, the cell population that takes the brunt of the stress of smoking and the initial site of pathological changes of COPD[1,21-24,28]. The data demonstrate that, among the 201 probesets representing 152 genes in the small airway epithelium that smoking induces to be significantly up- or down-regulated in normal smokers compared to normal non-smokers, 31 probesets (15%) representing 28 (14%) genes showed a coefficient of variation of expression level greater than the 98th percentile of coefficient of variation for all probesets and genes. Variability in gene expression levels was a reproducible property of each of these genes, not only in normal smokers but also in current smokers with COPD GOLD stages I or II. Interestingly, there was a higher degree of variation in gene expression in the smokers with COPD compared to the normal smokers. In the context that there is significant variability in whether or not cigarette smokers with comparable smoking history develop lung disease[1-5], we propose that this subset of smoking-induced variable response genes represent a novel set of candidate genes for future genetic epidemiology studies for risk / protection from smoking-induced disorders of the airways.

Reproducibility of Measured Parameters

There are a number of lines of evidence to suggest the observed variability in gene expression was not an experimental aberration. First, replicate probesets provided similar coefficient of variation for the same genes. Second, previous studies have shown that replicate determinations of the levels of gene expression in the airway epithelium of the same subject gave highly reproducible levels of expression when assessed in the same individual over time[18]. Third, there is extensive data from our group and others of TaqMan confirmation of expression levels of genes in the airway epithelium determined by microarray methods and selected genes were confirmed in this study[18-20,29-34]. Fourth, assessment of an independent data set of nonsmokers and smokers verified a similar pattern of increase variability in this subset of the same genes. However, while microarray methods allow a comprehensive survey of gene expression at the mRNA level, follow up would require proteomic methods to confirm.

In this study we have borrowed a concept from genome wide association studies and applied it to analysis of variability in gene expression levels[12,14,24]. Statistical methods such as the Benjamini-Hochberg correction applied in the present study help considerably in reducing the possibility of false positives. But, as in genome wide association studies, the replication of the initial observation in a second dataset is considered a prerequisite to confidence in the observed association. By finding high variability in expression level in two different populations both in our data set and in an independent data set, we reduce the probability that variability in expression level is a statistical anomaly. Interestingly, however, although there was an overall correlation of coefficient of variation in the normal smokers and smokers with COPD, the extent of variability was higher in the smokers with COPD. The reason for this is unclear, but may reflect pathogenic processes including inflammation and tissue remodeling in the airway epithelium of subjects as disease progresses and/or the inherent genetic diversity of smokers who have abnormalities compared to those that remain phenotypically normal.

Implication of Airway Gene Expression Variability for Genesis of Smoking-related Pulmonary Disease

Many of the genes discovered in this study could rationally play a role in the protection / susceptibility of the airway to smoking related lung diseases. For example, cytochromes

P450 are critical to metabolism of the xenobiotics in tobacco smoke and CYP1B1 in particular has been implicated in the oxidation of benzopyrene[35,36]. Polymorphisms of CYP1B1 have previously been implicated in susceptibility to various types of cancer including lung cancer[37,38]. Expression of MUC5AC has also been investigated in airway epithelial cells and shown to be induced by inflammatory stimuli[39]. Mucus hypersecretion is a characteristic of patients with COPD and contributes to their morbidity by causing airway obstruction.

There is extensive evidence that susceptibility / resistance to smoking-induced airway disorders is genetically determined[6-14,40]. Family studies confirm there is a heritable component of COPD[41]. Using a candidate gene approach with assumptions about the pathogenic mechanisms, polymorphisms in multiple genes with putative relevance to COPD have been identified in small studies, including microsomal epoxide hydroxylase (EPHX1)[6,7,40,42], angiotensin converting enzyme[43,44], β-defensin 1[45,46], glutathione S-transferase[42,47], matrix metalloproteinases[48,49] and tumor necrosis factor[50,51]. However, follow up studies in wider population have sometimes failed to confirm these association, possibly due to there likely being multiple genes associated with susceptibility/resistance to COPD, which may not be equally influential in all populations[11-14]. Genome wide association studies that are now ongoing should significantly contribute to this gene list. The present study proposes a new strategy to complement these approaches, based on the increasing evidence that variability on gene expression level is genetically determined and is a critical factor in the development of COPD.

Acknowledgments

References

1. Pauwels RA, Buist AS, Calverley PM, et al. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease. NHLBI/WHO Global Initiative for Chronic Obstructive Lung Disease (GOLD) Workshop summary. Am J Respir Crit Care Med. 2001; 163:1256–1276. [PubMed: 11316667]

2. Barnes PJ, Shapiro SD, Pauwels RA. Chronic obstructive pulmonary disease: molecular and cellular mechanisms. Eur Respir J. 2003; 22:672–688. [PubMed: 14582923]

3. Celli BR, MacNee W. Standards for the diagnosis and treatment of patients with COPD: a summary of the ATS/ERS position paper. Eur Respir J. 2004; 23:932–946. [PubMed: 15219010]

4. Lokke A, Lange P, Scharling H, et al. Developing COPD: a 25 year follow up study of the general population. Thorax. 2006; 61:935–939. [PubMed: 17071833]

5. Rennard SI, Vestbo J. COPD: the dangerous underestimate of 15%. Lancet. 2006; 367:1216–1219. [PubMed: 16631861]

6. Brogger J, Steen VM, Eiken HG, et al. Genetic association between COPD and polymorphisms in TNF, ADRB2 and EPHX1. Eur Respir J. 2006; 27:682–688. [PubMed: 16585076]

7. Hersh CP, Demeo DL, Lazarus R, et al. Genetic association analysis of functional impairment in chronic obstructive pulmonary disease. Am J Respir Crit Care Med. 2006; 173:977–984. [PubMed: 16456143]

8. Juul K, Tybjaerg-Hansen A, Marklund S, et al. Genetically increased antioxidative protection and decreased chronic obstructive pulmonary disease. Am J Respir Crit Care Med. 2006; 173:858–864. [PubMed: 16399992]

9. Matheson MC, Ellis JA, Raven J, et al. Beta2-adrenergic receptor polymorphisms are associated with asthma and COPD in adults. J Hum Genet. 2006; 51:943–951. [PubMed: 16946993]

10. Chappell S, Daly L, Morgan K, et al. Cryptic haplotypes of SERPINA1 confer susceptibility to chronic obstructive pulmonary disease. Hum Mutat. 2006; 27:103–109. [PubMed: 16278826]

11. Cookson WO. State of the art. Genetics and genomics of chronic obstructive pulmonary disease. Proc Am Thorac Soc. 2006; 3:473–475. [PubMed: 16921110]

12. Molfino NA. Current thinking on genetics of chronic obstructive pulmonary disease. Curr Opin Pulm Med. 2007; 13:107–113. [PubMed: 17255800]

13. Silverman EK. Progress in chronic obstructive pulmonary disease genetics. Proc Am Thorac Soc. 2006; 3:405–408. [PubMed: 16799082]

14. Wood AM, Stockley RA. The genetics of chronic obstructive pulmonary disease. Respir Res. 2006; 7:130. [PubMed: 17054776]

15. Cheung VG, Conlin LK, Weber TM, et al. Natural variation in human gene expression assessed in lymphoblastoid cells. Nat Genet. 2003; 33:422–425. [PubMed: 12567189]

16. Cheung VG, Spielman RS, Ewens KG, et al. Mapping determinants of human gene expression by regional and genome-wide association. Nature. 2005; 437:1365–1369. [PubMed: 16251966]

17. Spielman RS, Bastone LA, Burdick JT, et al. Common genetic variants account for differences in gene expression among ethnic groups. Nat Genet. 2007; 39:226–231. [PubMed: 17206142]

18. Hackett NR, Heguy A, Harvey BG, et al. Variability of antioxidant-related gene expression in the airway epithelium of cigarette smokers. Am J Respir Cell Mol Biol. 2003; 29:331–343. [PubMed: 12702543]

19. Harvey BG, Heguy A, Leopold PL, et al. Modification of gene expression of the small airway epithelium in response to cigarette smoking. J Mol Med. 2007; 85:39–53. [PubMed: 17115125]

20. Spira A, Beane J, Shah V, et al. Effects of cigarette smoke on the human airway epithelial cell transcriptome. Proc Natl Acad Sci U S A. 2004; 101:10143–10148. [PubMed: 15210990]

21. Hogg JC, Macklem PT, Thurlbeck WM. Site and nature of airway obstruction in chronic obstructive lung disease. N Engl J Med. 1968; 278:1355–1360. [PubMed: 5650164]

22. Niewoehner DE, Kleinerman J, Rice DB. Pathologic changes in the peripheral airways of young cigarette smokers. N Engl J Med. 1974; 291:755–758. [PubMed: 4414996]

23. Saetta M, Turato G, Maestrelli P, et al. Cellular and structural bases of chronic obstructive pulmonary disease. Am J Respir Crit Care Med. 2001; 163:1304–1309. [PubMed: 11371392]

24. Hirschhorn JN, Daly MJ. Genome-wide association studies for common diseases and complex traits. Nat Rev Genet. 2005; 6:95–108. [PubMed: 15716906]

25. Morley M, Molony CM, Weber TM, et al. Genetic analysis of genome-wide variation in human gene expression. Nature. 2004; 430:743–747. [PubMed: 15269782]

26. Cheung VG, Spielman RS. The genetics of variation in gene expression. Nat Genet. 2002; 32:522–525. Suppl. [PubMed: 12454648]

27. Rabe KF, Hurd S, Anzueto A, et al. Global Strategy for the Diagnosis, Management, and Prevention of chronic obstructive pulmonary disorder: GOLD executive summary. Am J Respir Crit Care Med. 2007; 176:532–555. [PubMed: 17507545]

28. Shapiro SD, Ingenito EP. The pathogenesis of chronic obstructive pulmonary disease: advances in the past 100 years. Am J Respir Cell Mol Biol. 2005; 32:367–372. [PubMed: 15837726]

29. Carolan BJ, Heguy A, Harvey BG, et al. Up-regulation of expression of the ubiquitin carboxyl-terminal hydrolase L1 gene in human airway epithelium of cigarette smokers. Cancer Res. 2006; 66:10729–10740. [PubMed: 17108109]

30. Gelbman BD, Heguy A, O'Connor TP, et al. Upregulation of pirin expression by chronic cigarette smoking is associated with bronchial epithelial cell apoptosis. Respir Res. 2007; 8:10. [PubMed: 17288615]

31. Heguy A, Harvey BG, Leopold PL, et al. Responses of the human airway epithelium transcriptome to in vivo injury. Physiol Genomics. 2007; 29:139–148. [PubMed: 17164391]

32. Kaplan R, Luettich K, Heguy A, et al. Monoallelic up-regulation of the imprinted H19 gene in airway epithelium of phenotypically normal cigarette smokers. Cancer Res. 2003; 63:1475–1482. [PubMed: 12670893]

33. Xu W, Zheng S, Goggans TM, et al. Cystic fibrosis and normal human airway epithelial cell response to influenza a viral infection. J Interferon Cytokine Res. 2006; 26:609–627. [PubMed: 16978065]

34. Kuperman DA, Lewis CC, Woodruff PG, et al. Dissecting asthma using focused transgenic modeling and functional genomics. J Allergy Clin Immunol. 2005; 116:305–311. [PubMed: 16083784]

35. Buters JT, Mahadevan B, Quintanilla-Martinez L, et al. Cytochrome P450 1B1 determines susceptibility to dibenzo*[a,l]*pyrene-induced tumor formation. Chem Res Toxicol. 2002; 15:1127–1135. [PubMed: 12230405]

36. Mahadevan B, Luch A, Atkin J, et al. Inhibition of human cytochrome p450 1b1 further clarifies its role in the activation of dibenzo*[a,l]*pyrene in cells in culture. J Biochem Mol Toxicol. 2007; 21:101–109. [PubMed: 17623886]

37. Yoon KA, Kim JH, Gil HJ, et al. CYP1B1, CYP1A1, MPO, and GSTP1 polymorphisms and lung cancer risk in never-smoking Korean women. Lung Cancer. 2007 doi:10.1016/j.physletb. 2003.10.071.

38. Cote ML, Wenzlaff AS, Bock CH, et al. Combinations of cytochrome P-450 genotypes and risk of early-onset lung cancer in Caucasians and African Americans: a population-based study. Lung Cancer. 2007; 55:255–262. [PubMed: 17174438]

39. Shao MX, Nakanaga T, Nadel JA. Cigarette smoke induces MUC5AC mucin overproduction via tumor necrosis factor-alpha-converting enzyme in human airway epithelial (NCI-H292) cells. Am J Physiol Lung Cell Mol Physiol. 2004; 287:L420–L427. [PubMed: 15121636]

40. Sandford AJ, Chagani T, Weir TD, et al. Susceptibility genes for rapid decline of lung function in the lung health study. Am J Respir Crit Care Med. 2001; 163:469–473. [PubMed: 11179124]

41. Gottlieb DJ, Wilk JB, Harmon M, et al. Heritability of longitudinal change in lung function. The Framingham study. Am J Respir Crit Care Med. 2001; 164:1655–1659. [PubMed: 11719305]

42. Cheng SL, Yu CJ, Chen CJ, et al. Genetic polymorphism of epoxide hydrolase and glutathione S-transferase in COPD. Eur Respir J. 2004; 23:818–824. [PubMed: 15218992]

43. Kanazawa H, Okamoto T, Hirata K, et al. Deletion polymorphisms in the angiotensin converting enzyme gene are associated with pulmonary hypertension evoked by exercise challenge in patients with chronic obstructive pulmonary disease. Am J Respir Crit Care Med. 2000; 162:1235–1238. [PubMed: 11029323]

44. Tkacova R, Joppa P, Stancak B, et al. The link between angiotensin-converting enzyme genotype and pulmonary artery pressure in patients with COPD. Wien Klin Wochenschr. 2005; 117:210–214. [PubMed: 15875760]

45. Matsushita I, Hasegawa K, Nakata K, et al. Genetic variants of human beta-defensin-1 and chronic obstructive pulmonary disease. Biochem Biophys Res Commun. 2002; 291:17–22. [PubMed: 11829455]

46. Demeo DL, Mariani TJ, Lange C, et al. The SERPINE2 Gene Is Associated with Chronic Obstructive Pulmonary Disease. Am J Hum Genet. 2006; 78:253–264. [PubMed: 16358219]

47. Ishii T, Matsuse T, Teramoto S, et al. Glutathione S-transferase P1 (GSTP1) polymorphism in patients with chronic obstructive pulmonary disease. Thorax. 1999; 54:693–696. [PubMed: 10413721]

48. Ito I, Nagai S, Handa T, et al. Matrix metalloproteinase-9 promoter polymorphism associated with upper lung dominant emphysema. Am J Respir Crit Care Med. 2005; 172:1378–1382. [PubMed: 16126934]

49. Joos L, He JQ, Shepherdson MB, et al. The role of matrix metalloproteinase polymorphisms in the rate of decline in lung function. Hum Mol Genet. 2002; 11:569–576. [PubMed: 11875051]

50. Huang SL, Su CH, Chang SC. Tumor necrosis factor-alpha gene polymorphism in chronic bronchitis. Am J Respir Crit Care Med. 1997; 156:1436–1439. [PubMed: 9372657]

51. Sakao S, Tatsumi K, Igari H, et al. Association of tumor necrosis factor alpha gene promoter polymorphism with the presence of chronic obstructive pulmonary disease. Am J Respir Crit Care Med. 2001; 163:420–422. [PubMed: 11179116]
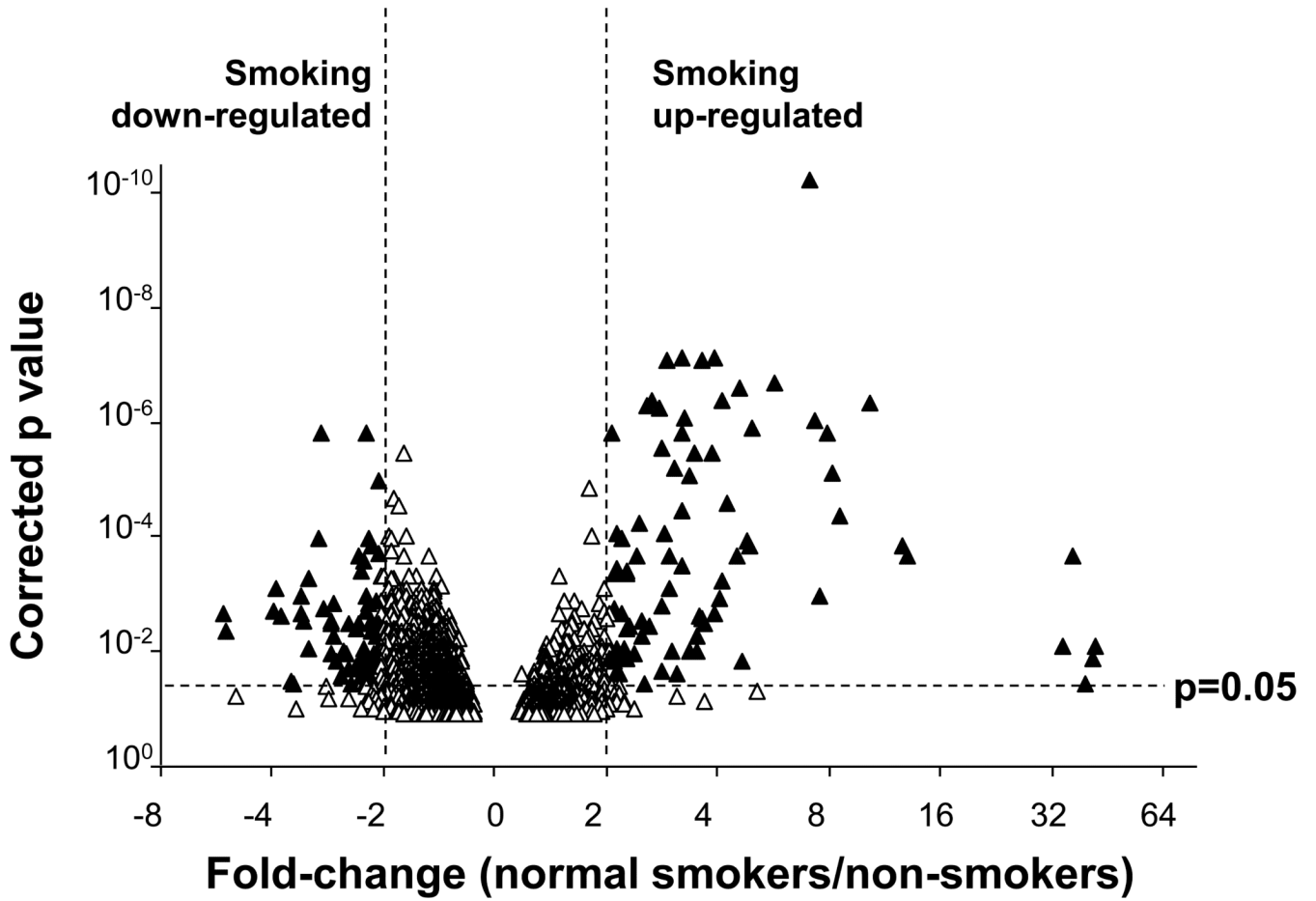
Figure 1.
Impact of smoking on the gene expression profile of the small airway epithelium. The normalized gene expression level for all 54 subjects was determined for all probesets scored as present in >50% of the samples. The natural log (ln) of the ratio of expression levels in normal smokers *vs* normal non-smokers was plotted against the -log of the p value (paired t test, all normal smokers *vs* normal nonsmokers; Benjamini-Hochberg corrected). The areas of the plot representing genes that are significantly smoking upregulated and smoking down-regulated genes are indicated.
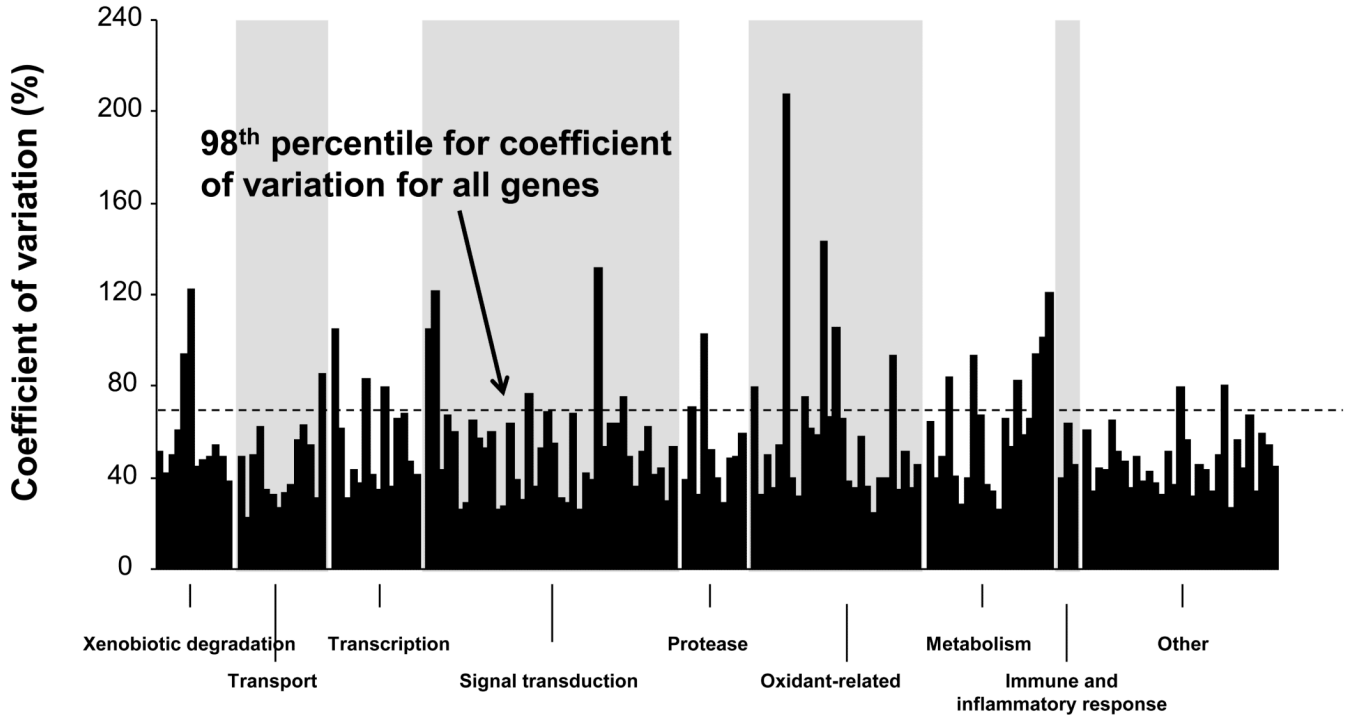
Figure 2.
Categories of smoking responsive genes in small airway epithelium. The set of 201 smoking responsive probesets (p<0.05, fold-change >2) in normal smokers compared to normal nonsmokers was divided into functional groups based on the Gene Ontology Biological Process Description supplemented with specific searches of the GenBank Gene descriptor. For the normal smokers, the coefficient of variation is plotted for each gene within a functional group in an arbitrary order based on Affymetrix probeset numbers.
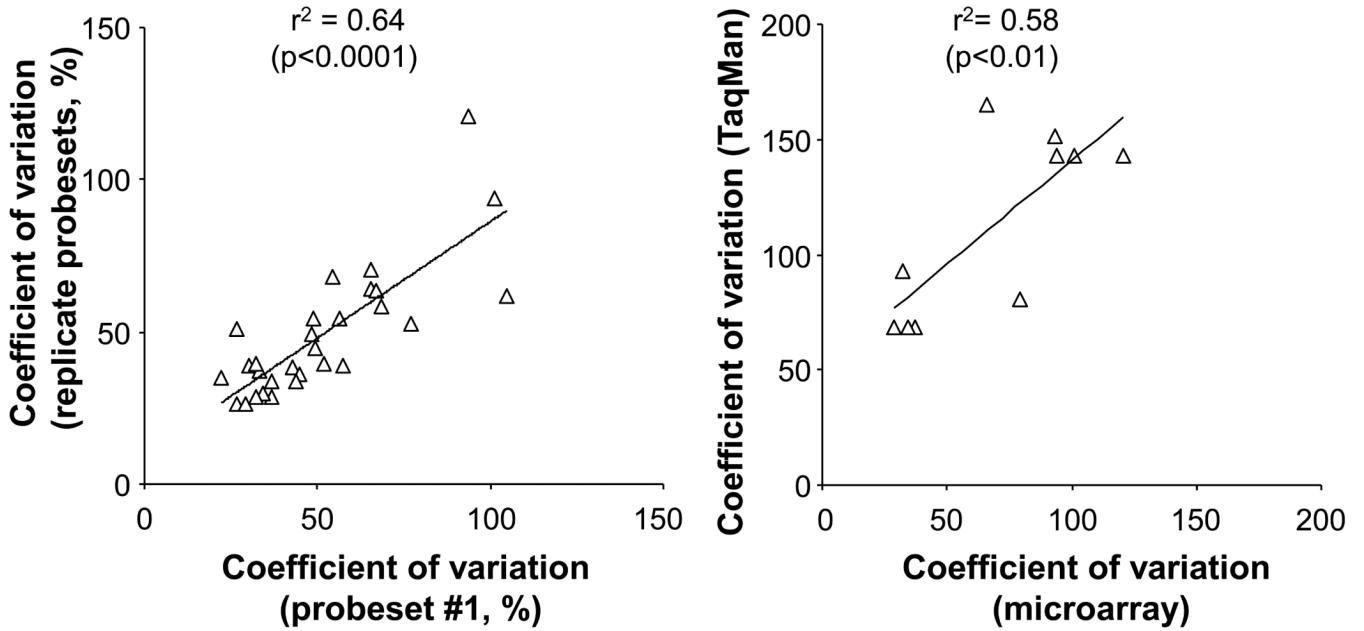
Figure 3.
Reproducibility of coefficient of variation. **A.** Replicate probesets for a single gene. The set of 201 smoking responsive probesets (p<0.05, fold change >2) was assessed for the normal smokers for all genes with duplicate probesets and the coefficient of variation for the first probeset (arbitrarily designated on the basis of Affymetrix probeset identifier) was plotted against the coefficient of variation for replicate probesets. **B.** Expression levels for six genes CYP1B1, UCHL1, HES6, AKR1B10, NQO1 and TCF7L1 were assessed by TaqMan realtime PCR for n=14 smokers by the ΔΔCt method and the coefficient of variation plotted against to coefficient of variation for all probesets corresponding to those genes on microarrays.

A.

B.

C.
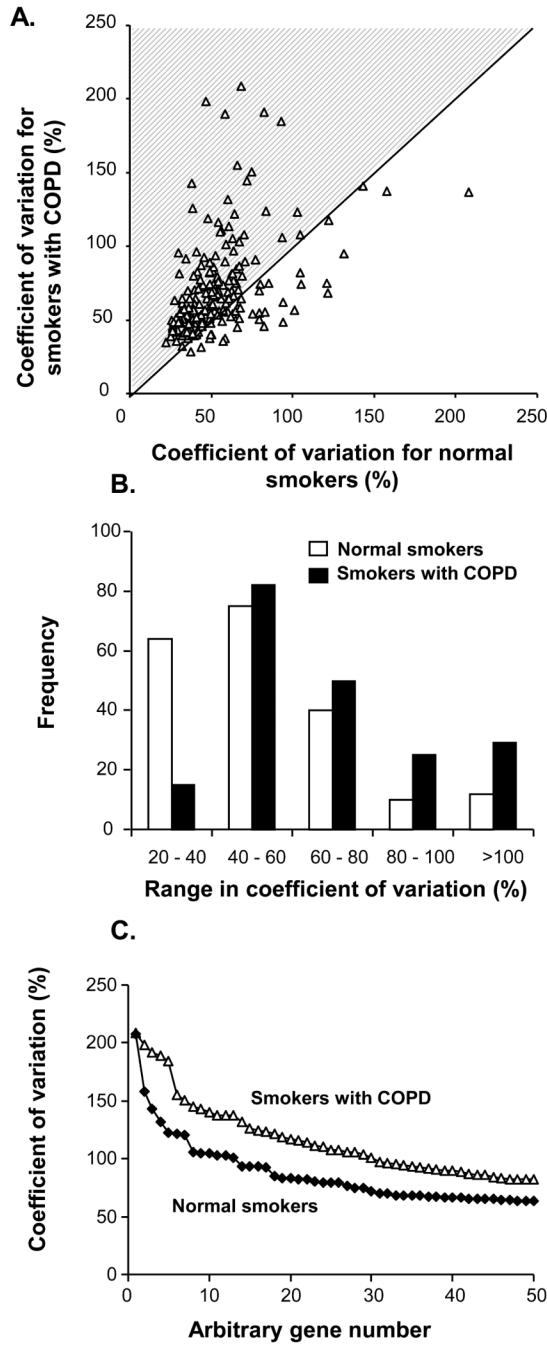
Figure 4.
Impact of COPD on coefficient of variation for the 201 smoking-dependent highly variable
probesets **A.** Coefficient of variation for smokers with COPD *vs* normal smokers. The
shaded area represents those probesets with a higher coefficient of variation in the smokers
with COPD. **B.** Frequency distribution of the coefficient of variation in normal smokers (□),
and smokers with COPD (■). **C.** Distribution of coefficient of variation plotted against gene
in rank order for the top 50 most variable genes for the normal smokers and the smokers
with COPD.
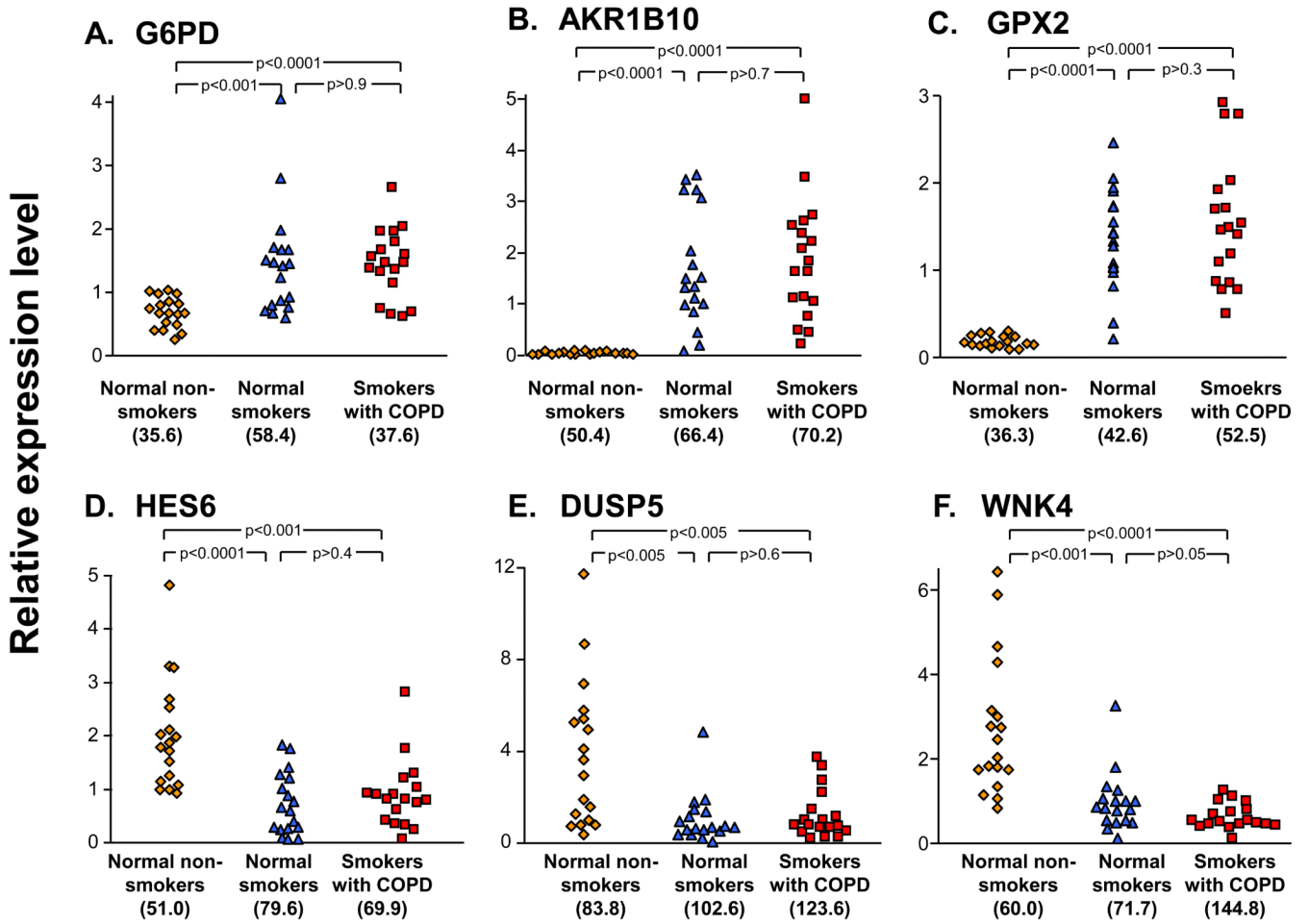
Figure 5.
Examples of genes up- and down-regulated by smoking with high coefficient of variation in normal smokers and smokers with COPD. The normalized expression levels for the non-smokers, normal smokers and smokers with COPD were compared. **A**. G6PD - Glucose-6-phosphate dehydrogenase; **B.** AKR1B10 - aldo-keto reductase family 1, member B10; **C.** GPX2 - glutathione peroxidase 2; **D.** HES6 - hairy and enhancer of split 6; **E.** DUSP5 - dual specificity phosphatase 5; and **F.** WNK4 - WNK lysine deficient protein kinase 4. The coefficient of variation for the genes in each of the study groups in shown in parentheses below the scatter plot; p values are based on unpaired t tests.

Table I

Demographic Characteristics of the Study Population and Small Airway Epithelial Cells Recovered

Parameter	Normal non-smokers[1]	Normal smokers[1]	Smokers with COPD (GOLD I / II)[1]
Number	18	18	18
Sex (male / female)	14/4	12/6	15/3
Self-described race (African / European / other)	10/6/2	10/8/0	6/8/4
Age	41 ± 7	46 ± 5	50 ± 6
Smoking history (pack-yr)	0	31 ± 16	34 ± 17
Urine nicotine (ng/ml)	0	1099 ± 1217	894 ± 930
Urine cotinine (ng/ml)	0	1419 ± 1008	1381 ± 732
Carboxyhemoglobin (%)	0.7 ± 1	3.4 ± 2.5	3.2 ± 1.6
Pulmonary function tests[2]			
FVC	105 ± 9	109 ± 15	103 ± 19
FEV1	106 ± 9	109 ± 15	82 ± 18
FEV1/FVC	82 ± 5	82 ± 4	64 ± 6
TLC	98 ± 9	102 ± 14	105 ± 23
DLCO	97 ± 10	93 ± 8	77 ± 16
Number of airway epithelial cells recovered ($\times 10^6$)	5.3×10^6	6.7×10^6	5.3×10^6
% Total cells epithelial	99.8 ± 0.4	99 ± 0.8	96 ± 1
Inflammatory	0.2 ± 0.4	0.5 ± 0.8	4 ± 1
% Epithelial cells[3]			
Ciliated	71.5 ± 6.0	72.6 ± 6.1	66 ± 8
Secretory	7.1 ± 3.4	10.2 ± 3.1	18 ± 6
Basal	15.2 ± 3.7	10.9 ± 3.2	7 ± 3
Undifferentiated	5.9 ± 1.8	5.8 ± 4.0	9 ± 1

[1] For the criteria for the study groups, see Methods.

[2] FVC - forced vital capacity; FEV1- forced expiratory volume in 1 sec; DLCO - diffusing capacity for carbon monoxide; TLC - total lung volume; all values are presented as % predicted except for FEV1/FVC presented as % observed.

[3] Compared to differential of 43% ciliated, 10% secretory, 27% basal and 20% undifferentiated in the large airway epithelium of non-smokers[31].

Table II

Smoking Responsive Genes with High Coefficient of Variation in Expression Levels in Small Airway Epithelium of Normal Smokers Compared to Nonsmokers[1]

Gene symbol	Probe set ID	Gene title	Coefficient of variation (%)	Coefficient of variation in independent study[2]
CYP1A1	205749_at	Cytochrome P450, family 1, subfamily A, polypeptide 1	207.7%	139.6%
---	238835_at	Transcribed locus	157.9%	NA
SLC26A4	206529_x_at	Solute carrier family 26, member 4	143.3%	257.7%
MUC5AC	217182_at	Mucin 5AC	131.6%	76.2%
---	242680_at	Transcribed locus	122.3%	NA
SPP1	209875_s_at	Osteopontin	121.5%	126.9%
CYP1B1	202437_s_at	Cytochrome P450, family 1, subfamily B, polypeptide 1	120.8%	107.5%
NR0B1	206645_s_at	Nuclear receptor subfamily 0, group B, member 1	105.4%	211.1%
H19	224646_x_at	H19, imprinted maternally expressed untranslated mRNA	104.8%	NA
CXCL2	209774_x_at	Chemokine (C-X-C motif) ligand 2	104.8%	106.3%
DUSP5	209457_at	Dual specificity phosphatase 5	102.6%	121.4%
CYP1B1	202436_s_at	Cytochrome P450, family 1, subfamily B, polypeptide 1	101.0%	108.1%
---	242054_s_at	Transcribed locus	93.9%	NA
CYP1B1	202435_s_at	Cytochrome P450, family 1, subfamily B, polypeptide 1	93.9%	110.5%
UCHL1	201387_s_at	Ubiquitin carboxyl-terminal esterase L1	93.3%	144.1%
SAA1	208607_s_at	Serum amyloid A1	93.0%	115.3%
MALAT1	224559_at	Metastasis associated lung adenocarcinoma transcript 1	85.2%	NA
SERPING1	200986_at	Serpin peptidase inhibitor, clade G, member 1	83.5%	63.2%
MSRB3	225782_at	Methionine sulfoxide reductase B3	83.0%	NA
LTF	202018_s_at	Lactotransferrin	82.3%	136.1%
UGT8	228956_at	UDP glycosyltransferase 8	82.2%	NA
SLC34A2	204124_at	Solute carrier family 34 member 2	80.2%	40.5%
VEGFB	203683_s_at	Vascular endothelial growth factor B	79.7%	40.1%
HES6	226446_at	Hairy and enhancer of split 6	79.6%	NA
GAD1	205278_at	Glutamate decarboxylase 1	79.3%	79.6%
SFTPB	213936_x_at	Surfactant, pulmonary-associated protein B	76.9%	34.6%
CYP4F11	206153_at	Cytochrome P450, family 4, subfamily F, polypeptide 11	75.2%	56.4%
LOC653879	217767_at	Similar to complement C3 precursor	74.9%	81.1%
WNK4	229158_at	WNK lysine deficient protein kinase 4	71.7%	NA
TFPI2	209278_s_at	Tissue factor pathway inhibitor 2	70.6%	NA
CDC20B	240161 s at	Cell division cycle 20 homolog B	70.3%	NA

[1] The table lists all highly variable probesets (coefficient of variation greater than the 98th percentile of all probesets) with a statistically significant (p<0.05 after Benjamini-Hochberg correction) change in expression level of >2-fold between normal smokers and normal non-smokers. If more than 1 probeset was available, the coefficient of variation for all probesets are presented.

[2] For the probesets that are also represented on the Affymetrix HG U133A microarray used in Spira et al [20]. Coefficient of variation is shown for the n=34 smokers. NA = not present on HG U133A microarray. Note that of the 20 probesets from this data set, 15 show a coefficient of variation greater than the 98th percentile (72.8%) of samples in this data set.