



Published in final edited form as:

Cancer Prev Res (Phila). 2010 March ; 3(3): 266–278. doi:10.1158/1940-6207.CAPR-09-0192.

Effects of Cigarette Smoke on the Human Oral Mucosal Transcriptome

Jay O. Boyle^{1,*}, Zeynep H. Gümüş^{2,8,*}, Ashutosh Kacker³, Vishal L. Choksi¹, Jennifer M. Bocker¹, Xi Kathy Zhou⁴, Rhonda K. Yantiss⁵, Duncan B. Hughes⁶, Baoheng Du⁷, Benjamin L. Judson¹, Kotha Subbaramaiah⁷, and Andrew J. Dannenberg⁷

¹ Department of Surgery (Head and Neck Service), Memorial Sloan-Kettering Cancer Center, New York, NY

² Department of Physiology and Biophysics, Weill Medical College of Cornell University, New York, NY

³ Department of Otorhinolaryngology, Weill Medical College of Cornell University, New York, NY

⁴ Department of Public Health, Weill Medical College of Cornell University, New York, NY

⁵ Department of Pathology and Laboratory Medicine, Weill Medical College of Cornell University, New York, NY

⁶ Department of Surgery, Weill Medical College of Cornell University, New York, NY

⁷ Department of Medicine, Weill Medical College of Cornell University, New York, NY

⁸ HRH Prince Alwaleed Bin Talal Bin Abdulaziz Alsaud Institute for Computational Biomedicine, Weill Medical College of Cornell University, New York, NY

Abstract

Use of tobacco is responsible for approximately 30% of all cancer-related deaths in the United States including cancers of the upper aerodigestive tract. In the current study, 40 current and 40 age- and gender-matched never smokers underwent buccal biopsies to evaluate the effects of smoking on the transcriptome. Microarray analyses were carried out using Affymetrix HGU 133 Plus2 arrays. Smoking altered the expression of numerous genes: 32 genes showed increased expression and 9 genes showed reduced expression in the oral mucosa of smokers vs. never smokers. Increases were found in genes involved in xenobiotic metabolism, oxidant stress, eicosanoid synthesis, nicotine signaling and cell adhesion. Increased numbers of Langerhans cells were found in the oral mucosa of smokers. Interestingly, smoking caused greater induction of aldo-keto reductases, enzymes linked to polycyclic aromatic hydrocarbon induced genotoxicity, in the oral mucosa of women than men. Striking similarities in expression changes were found in oral compared to the bronchial mucosa. The observed changes in gene expression were compared to known chemical signatures using the Connectivity Map database, and suggested that geldanamycin, an Hsp90 inhibitor, might be an anti-mimetic of tobacco smoke. Consistent with this prediction, geldanamycin caused dose-dependent suppression of tobacco smoke extract-mediated induction of CYP1A1 and CYP1B1 *in vitro*. Collectively, these results provide new insights into the carcinogenic effects of tobacco smoke, support the potential use of oral epithelium as a surrogate tissue in future lung cancer chemoprevention trials and illustrate the potential of computational biology to identify chemopreventive agents.

Requests for Reprints: Andrew J. Dannenberg, Department of Medicine and Weill Cornell Cancer Center, 525 East 68th Street, Room F-206, New York, NY 10065. Phone: (212) 746-4403; FAX: (212) 746-4885; ajdann@med.cornell.edu.

*J.O. Boyle and Z. H. Gumu contributed equally to this work

Keywords

tobacco; smoking; microarray; aryl hydrocarbon receptor; heat shock protein 90

Introduction

More than a billion people smoke cigarettes daily worldwide. Tobacco use is responsible for approximately 30% of all cancer-related deaths in the United States (1). Exposure to tobacco causes multiple human malignancies including cancers of the lung, oral cavity, pharynx, esophagus, stomach, liver, pancreas, kidney, bladder, and cervix (2). More than 60 carcinogens are found in mainstream cigarette smoke and most of these are also found in sidestream smoke (3). In addition to being a major cause of cancer, smoking alters the activity of chemopreventive agents (4,5), stimulates the clearance of selected targeted anticancer therapies (6), reduces the efficacy of cancer treatment (7–10) and increases the risk of second primary tumors (11). Women have been suggested to be at increased risk of lung, oral and oropharyngeal cancer compared with men who had similar cigarette smoking exposure levels (12–14). The mechanisms underlying this apparent gender-dependent difference in risk are poorly understood.

Numerous studies have been carried out to elucidate the carcinogenic effects of tobacco smoke on the bronchial epithelium. In histologically normal airway epithelial cells, smoking causes a range of abnormalities including *P53* mutations (15), changes in promoter methylation (16, 17), and allelic loss (18). Transcriptome profiling showed that smoking induced the expression of genes involved in xenobiotic metabolism and redox stress in large airway epithelial cells (19). Importantly, a profile of bronchial airway gene expression in cytologically normal large airway epithelial cells was found to be potentially useful as a biomarker of lung cancer (20). In theory, the successful development of a transcriptome-based biomarker to identify high-risk smokers could provide the basis for risk reduction strategies including chemoprevention. Although sampling the bronchial epithelium to identify potential biomarkers of cancer risk has yielded significant insights, it would be very useful if similar information could be obtained using less invasive tissue collection methods. Recently, Spira and colleagues (21) compared the effects of smoking on the transcriptome of extrathoracic (buccal and nasal) vs. intrathoracic (bronchial) epithelium. The results of gene expression profiles from buccal (n=10) and nasal (n=15) epithelial cells indicated that many of the smoking-related changes in the bronchial epithelium were also present in buccal and nasal epithelium. Possibly, sampling of extrathoracic epithelial cells will yield information that can help to define individual susceptibility to smoking-related diseases of the upper aerodigestive tract including the lung.

In the current study, 40 current smokers and 40 age- and gender-matched never smokers underwent buccal biopsies. We had four objectives: (a) to define the effects of smoking on the transcriptome of oral epithelial cells; (b) to determine if any of the effects of tobacco smoke on the transcriptome are gender-dependent; (c) to compare the effects of tobacco smoke exposure on the transcriptome in oral vs. bronchial epithelium and (d) to identify agents with the potential to suppress the effects of tobacco smoke on the transcriptome. We show that smoking altered the expression of genes involved in xenobiotic metabolism, oxidant stress, eicosanoid synthesis, nicotine signaling and cell adhesion. Smoking-mediated induction of aldo-keto-reductases (AKRs), enzymes linked to polycyclic aromatic hydrocarbon (PAH) induced genotoxicity (22), was greater in women than in men. Most smoking-related changes in gene expression in oral epithelial cells also occur in airway epithelial cells. Collectively, these data provide new insights into the carcinogenic effects of tobacco smoke and offer insights that may prove useful in developing preventive strategies.

Materials and Methods

Materials

Keratinocyte basal and growth media were obtained from Lonza. Antibody to β -actin and Lowry protein assay kits were obtained from Sigma Chemical. Antiserum to CYP1B1 was a gift of Dr. Craig B. Marcus (Oregon State University, Corvallis, OR). Antibody to CYP1A1 was obtained from Santa Cruz Biotechnology. CD1a mouse monoclonal antibody (clone MTB1) was from Novocastra Laboratories Ltd. Western blot analysis detection reagents (enhanced chemiluminescence) were from Amersham Biosciences. Nitrocellulose membranes were from Schleicher and Schuell. Geldanamycin was purchased from Calbiochem. MuLV reverse transcriptase, oligo(dt)16, and RNase inhibitor were from Roche Applied Science, and Taq polymerase was from Applied Biosystems. HGU133 Plus 2 microarrays were from Affymetrix.

Study design

40 never smokers (<100 cigarettes per lifetime) and 40 active smokers (≥ 15 pack year exposure) were recruited (see Supplementary Table S1). Subjects were age- and gender-matched. Eligible subjects were healthy volunteers recruited from the community and hospital. Subjects were excluded if they had gross evidence of oral inflammation, a history of heavy alcohol consumption, or recent use of nonsteroidal anti-inflammatory drugs or other anti-inflammatory medications. The study was approved by the Weill Cornell Medical College Institutional Review Board and the Clinical and Translational Science Center. All subjects provided written informed consent for participation.

Human tissue

After topical anesthesia, 5-mm punch biopsies were obtained from grossly normal appearing buccal mucosa. Tissue samples were immediately divided into two parts. Approximately, two-thirds of each specimen was snap frozen in liquid nitrogen. Total RNA was then isolated with an RNeasy Mini Kit (Qiagen Inc.) and stored at -80°C until analysis. The remaining one-third of the biopsy was formalin fixed for immunohistochemical analysis.

Microarray procedures

Biotinylated cRNA were prepared according to the standard Affymetrix protocol from 2.5 μg total RNA (<http://www.affymetrix.com>). Following fragmentation, 10 μg of cRNA were hybridized for 16 h at 45°C on GeneChip HGU133 Plus 2 arrays. GeneChips were washed and stained in the Affymetrix Fluidics Station 450 and scanned using the Affymetrix GeneChip Scanner 3000 7G.

Microarray data analysis

Each array's scanned image was checked for significant artifacts. One sample was excluded from the study based on this quality measure, leaving 79 arrays for analysis.

Preprocessing—Raw image data were background corrected, normalized and summarized into probeset expression values using the Robust Multichip Average (RMA) algorithm (23, 24) within GeneSpring 7.2 Software (Agilent Technologies). Data from each chip were normalized for inter-array comparisons by first setting measurements of <0.01 to 0.01 and then normalizing to 50% of the measurements taken from that array. Probesets that were not reliably detected were filtered out. From the complete set of $\sim 54,675$ probe sets on the HGU133 Plus 2 array, genes were filtered for minimum raw expression level of 50 in at least 16 out of 79 conditions. Genes with low confidence were filtered out based on *t*-test P value <0.05 in at least 1 out of 2 conditions (smoker or never smoker). The cross-gene error model was active.

The ~24,103 probesets that passed these tests were defined as expressed and were statistically analyzed.

Statistical analysis—To identify differentially expressed gene groups between smoker and never smoker groups, one-way ANOVA was performed using parametric test, variances not assumed equal (Welch t-test) with P value cutoff of 0.05 and Benjamini-Hochberg multiple testing correction to maintain False Discovery Rate (FDR) at 5%. Genes with normalized smoker vs. never smoker expression values that changed by a factor of 1.5-fold were deemed significant and listed in Table 1.

Clustering—An unsupervised hierarchical clustering analysis across all samples of the microarray data was performed for the probe sets found to be differentially expressed in the oral mucosa of smokers and never smokers (using log-transformed, normalized, gene median centered data). Pearson correlation (un-centered) similarity metric and average linkage clustering was performed with CLUSTER and TREEVIEW software obtained at <http://rana.lbl.gov/EisenSoftware.htm> and shown in Fig. 1A.

Functional analysis—The effects of tobacco smoke were examined in the context of detailed molecular interaction networks using Ingenuity Pathway Analysis (IPA), a web-delivered application used to discover, visualize and explore relevant networks (<http://www.ingenuity.com>). Affymetrix probe identifiers and fold-values were uploaded to IPA and each identifier was mapped to its corresponding gene object in the Ingenuity Pathways Knowledgebase. Interactions were then queried between these gene objects and all other gene objects stored within IPA to generate a set of direct interaction networks that were merged. Putative transcription regulator hubs that directly interact with a minimum of three differentially expressed genes were included in the network. Because UGTs and AKR1C probes map to multiple genes in these families, all members of these families were included to identify their individual interconnections. The regulation of ALDH3A1, UGT1A1, UGT1A3, UGT1A4 by the aryl hydrocarbon receptor (AHR) and AKR1Cs by Nrf2 were manually added (25–28) to Fig. 1B.

Significantly altered groups—Significantly differentially expressed genes between smokers and never smokers were mined for statistically overrepresented gene groups using EASE software (29). Functional gene groups in Gene Ontology (GO) (<http://www.geneontology.org>) database were queried and the likelihood of overrepresentation of each gene group in the differentially expressed gene set with respect to the HGU133 Plus 2 microarray was scored (using Affymetrix identifiers). Relevant gene sets with Bonferroni $p < 0.05$ are reported in Table 2.

Effects of gender on smokers' transcriptome—The approach that was used to carry out this analysis is detailed in the Supplementary methods.

Modest and consistent alterations—The entire 54,675 microarray probesets from each of the 79 subjects were mined for statistically significant, concordant functional gene group differences between smokers and never smokers using Gene Set Enrichment Analysis (GSEA v2) (30). GSEA helps functionally interpret modest but consistent changes in the gene expression data and focuses on groups of genes that share common biological function. Normalized ratio expression values were analyzed using default parameter settings. Relevant gene sets with $FDR < 0.25$ were deemed significant.

Comparison of effects of smoking on the oral and bronchial epithelium—Smoking-related changes in the transcriptome of the oral and airway epithelium were

compared, using the current data as well as previously reported smoker and never smoker airway transcriptome data (19) (analyzed as described in *Statistical Analysis*). Overlapping genes are listed in Table 4. The relationship between the gene expression patterns in response to tobacco smoke in the oral and bronchial epithelium was identified by Enrichment Analysis as described in the Supplementary methods.

Gene Expression Signature Based Chemical Genomic Prediction—Differentially expressed genes were separated into up- and down-regulated gene sets and converted to their HGU 133A identifiers (<http://www.affymetrix.com/analysis/netaffx/index.affx>) which were queried to identify drugs with anti-mimetic gene expression signatures within the Connectivity Map (<http://www.broad.mit.edu/cmap>) (31).

Additional information

The complete results from the gender and GSEA analyses are available through an interactive web site (<http://physiology.med.cornell.edu/go/smoke>) established as a resource of the Institute for Computational Biomedicine. The microarray data have been deposited at the National Center for Biotechnology Information Gene Expression Omnibus (GEO) (<http://www.ncbi.nlm.nih.gov/geo/>) under GEO Series accession no. GSE17913.

Quantitative PCR validation

Samples from 10 never smokers and 10 smokers were chosen at random. Total RNA was isolated using RNeasy Mini Kit (Qiagen Inc.). RNA (1 µg) was reverse transcribed using MuLV reverse transcriptase and oligo d(T)16 primer. The resulting cDNA was then used for amplification. Each PCR reaction was 20 µL and contained 5µL cDNA, 2x SYBR green PCR master mix, and forward and reverse primers (see Supplementary Table S2 for list of primers). Experiments were performed using a 7500 real time PCR system (Applied Biosystems). β -actin served as an endogenous normalization control. Relative-fold induction was determined by ddCT (Relative Quantification) analysis.

Immunohistochemistry

Formalin fixed, paraffin-embedded oral mucosal tissue sections from 54 subjects (27 smokers, 27 never smokers) were evaluated for the presence and distribution of Langerhans cells using antiserum directed against CD1a, a Langerhans cell marker. Four micron thick tissue sections were immunohistochemically stained with the CD1a mouse monoclonal antibody as described below. Unstained tissue sections were baked, deparaffinized, and rehydrated on the Vision Biosystems/Leica BondMax autostainer (Chicago, IL). Tissue sections were pretreated using the heat induced epitope retrieval solution-1 (Vision Biosystems/Leica) and incubated with the primary antibody (1:20 dilution) for 25 min. The Refine Detection Kit supplied by the manufacturer was used to block endogenous peroxidase activity and enhance the staining reaction. Positive (skin) and negative (replacement of the primary antibody with immunoglobulin) controls were included in the experiment. Cells that displayed moderate to strong cytoplasmic staining for CD1a in dendritic-type cellular processes were separately evaluated in three regions of the mucosa: the peripapillary, interpapillary, and superficial epithelium. The total and mean number of CD1a-positive cells present in the peripapillary mucosa of four well-oriented papillae, and four high-magnification (400X objective) fields of the interpapillary and superficial mucosa, were recorded for each of the 54 cases. Comparisons between smokers and never smokers were made by Student's t-test. A difference between groups of $P < 0.05$ was considered significant.

Tissue culture

The MSK-Leuk1 cell line was established from a dysplastic leukoplakia lesion adjacent to a squamous cell carcinoma of the tongue (32). Cells were routinely maintained in keratinocyte growth medium supplemented with bovine pituitary extract. Cells were grown in basal medium for 24 h before treatment.

Preparation of tobacco smoke extract

Cigarettes (2R4F, Kentucky Tobacco Research Institute) were smoked in a Borgwaldt piston-controlled apparatus (model RG-1) using a Federal Trade Commission standard protocol. Cigarettes were smoked one at a time in the apparatus and the smoke was drawn under sterile conditions into premeasured amounts of sterile PBS (pH 7.4). This smoke in PBS represents whole trapped mainstream smoke, abbreviated as TS. Quantitation of smoke content is expressed in puffs/mL of PBS with one cigarette yielding about 8 puffs drawn into a 5 mL volume. The final concentration of TS in the cell culture medium is expressed as puffs/mL medium. All treatments were carried out with 0.03 puffs/mL TS because this concentration was previously found to induce CYP1A1 and CYP1B1 (33).

Western blot analysis

Cell lysates were prepared by treating cells with lysis buffer (150 mmol/L NaCl, 100 mmol/L Tris, pH 8.0, 1% Tween 20, 50 mmol/L diethyldithiocarbamate, 1 mmol/L phenylmethylsulfonyl fluoride, 10 µg/mL aprotinin, 10 µg/mL trypsin inhibitor and 10 µg/mL leupeptin). Lysates were sonicated for 3 × 10 s on ice and centrifuged at 14,000 × g for 10 min at 4°C to sediment the particulate material. The protein concentration of the supernatant was measured by the method of Lowry (34). SDS-PAGE was performed under reducing conditions on 10% polyacrylamide gels. The resolved proteins were transferred onto nitrocellulose sheets and then incubated with antisera to CYP1A1, CYP1B1, and s-actin. Secondary antibody to immunoglobulin G conjugated to horseradish peroxidase was used. The blots were then reacted with the ECL western blot detection system, according to the manufacturer's instructions.

Results

Smoking status is a determinant of the transcriptome in the oral mucosa

A total of 80 subjects (40 smokers, 40 never smokers) underwent biopsies of the buccal mucosa. One female smoker was excluded from the study because of problems processing the biopsy sample. Hence, samples from 79 subjects were available for analysis. Demographic data for these 79 subjects are presented in Supplementary Table S1. The never smoker group included 20 males (median age, 45 years) and 20 females (median age, 45 years). The smoker group included 20 males (median age, 45.5 years; median pack years 32.5) and 19 females (median age, 43 years; median pack years 25). Messenger RNA from 79 subjects (40 never smokers, 39 smokers) was suitable in quantity and quality for microarray analysis. The gene probes that were differentially expressed at least 1.5-fold between smokers and never smokers are listed in Table 1. Smoking altered the expression of numerous genes. Forty probes representing 32 genes showed increased expression and 14 probes representing 9 genes showed reduced expression in the oral mucosa of smokers *vs.* never smokers. Increases were found for genes involved in xenobiotic metabolism (CYP1A1, CYP1B1, AKR1C1/C2, UGT1A, NQO1, AHRR), oxidant stress (ALDH3A1, GPX2), eicosanoid synthesis (PTGES, ALOX12B, ALOX15B), nicotine signaling (CHRNA3) and cell adhesion (CEACAM7). Decreased expression was detected for genes including CCL18, SOX9, IGF2BP3 and LEPR. Subsequently, quantitative PCR was used to validate the microarray findings for a subset of 11 differentially expressed genes. Importantly, the observed changes in expression were quantitatively consistent with the microarray results for all 11 genes evaluated (Supplementary

Table S3). Figure 1A shows the unsupervised hierarchical clustering analysis of smokers *vs.* never smokers based on genes that were differentially expressed in the two groups. The majority of subjects clustered accurately into the two groups.

Interpreting the global transcriptome changes in terms of biological pathways and functions

Several databases and tools were used to classify the differentially expressed genes into relevant molecular and physiological categories. Interactions within IPA knowledgebase (<http://www.ingenuity.com>) and other known literature (25–28) were used to define potential smoking-induced effects on molecular interaction networks (Fig. 1B). The likely role of the AHR, a PAH activated transcription factor, was evident because increased levels of CYP1A1, CYP1B1 and AHRR mRNAs were found in the oral mucosa of smokers. PAH activated AHR stimulates the transcription of each of these genes (35). NFE2L2 (Nrf2), a transcription factor activated by oxidative stress, can induce AKR1C1/2, NQO1, GPX2 and ALD3A1 (36–38). Each of these genes was overexpressed in the oral mucosa of smokers strongly suggesting the involvement of Nrf2 (Table 1, Fig. 1B). IPA network analysis also suggested the involvement of other regulators of transcription including ARNT, RELA and SP1. The genes were further classified in terms of relevant functional categories to identify additional effects of tobacco smoke. Pathways within the KEGG (<http://www.genome.jp/kegg>) and GenMAPP (<http://www.genmapp.org>) databases were queried using GSEA v2. The following pathways were enriched in smokers: metabolism of xenobiotics by cytochromes P450, androgen and estrogen metabolism, eicosanoid synthesis, prostaglandin and leukotriene metabolism, and glutathione metabolism (Table 2). Query of GO (<http://www.geneontology.org>) functional databases using EASE suggested that electron transporter activity and oxidoreductase activity were increased in the oral mucosa of smokers (Table 2).

Increased numbers of Langerhans cells are found in the oral mucosa of smokers

Generally, changes in the transcriptome reflect altered gene expression. We note, however, that changes in the cellular composition of a biopsy can also affect the transcriptome. Increased levels of both CD207 (langerin) and CD1a mRNAs, transcripts that are abundant in Langerhans cells, were found in the oral mucosa of smokers (Table 1, Supplementary Table S3). This suggested the possibility that the number of Langerhans cells might be increased in the oral mucosa of smokers. Because CD1a is a marker for Langerhans cells, immunohistochemistry was carried out to quantify the number of CD1a positive cells in the oral mucosa of smokers *vs.* never smokers. A significant increase in the number of CD1a positive cells was found in the oral mucosa of smokers *vs.* never smokers (Figs. 2A–E).

Gender-dependent differences in smoking-mediated changes in gene expression

Previously, a somewhat higher risk of cancers of the lung, oral cavity and oropharynx was found in women than men at comparable pack years of smoking (12–14). It was of interest, therefore, to determine if levels of gene expression differed in the oral mucosa of males *vs.* females. In never smokers, the genes that were differentially expressed in males *vs.* females primarily reflected gender-dependent differences in genes of X and Y chromosomes (Supplementary Table S4). The effects of smoking were also evaluated. Interestingly, smoking had a greater effect on both the induction (AKR1C2/3, UGT family members) and suppression (IGFL1) of several genes in women than in men (Table 3).

Comparison of oral mucosa and airway epithelial transcriptome of smokers *vs.* never smokers

Smoking modulates gene expression in the airway epithelium. Hence, we compared our findings for oral mucosa with previously reported data for airway epithelium (19). Striking similarities in expression changes were found in the oral mucosa and bronchial epithelial cells

of smokers (Table 4). For example, smoking was associated with increased expression of a variety of genes (CYP1A1, CYP1B1, NQO1, ALDH3A1, UGTs) involved in xenobiotic metabolism. Increased levels of GPX2 and CEACAM family members were found in both the oral and bronchial epithelium of smokers. Interestingly, smoking was associated with increased levels of FCGBP in oral mucosa but reduced expression in bronchial mucosa. Gene Set Enrichment Analysis also suggested that the inductive effects of smoking are similar in both the oral and bronchial epithelium (Supplementary Table S5).

Targeting Hsp90 can attenuate the activation of AHR-dependent gene expression

Agents that suppress tobacco smoke-mediated effects on the transcriptome are likely to possess chemopreventive properties. Accordingly, we next attempted to identify a small molecule with the potential to attenuate some of the changes in the transcriptome found in the oral mucosa of smokers. To achieve this goal, a computational approach was used in combination with an *in vitro* model that has been used in previous tobacco studies (33). The mRNA profile that was observed in the oral mucosa of smokers vs. never smokers was compared to known signatures of pharmaceutical and small molecule treatments using the Connectivity Map database (31). This computational analysis suggested that geldanamycin, an inhibitor of heat shock protein 90 (Hsp90), might be an anti-mimetic of tobacco smoke (P=0.0003). As detailed above, the AHR, a client protein of Hsp90, mediates the induction of *CYP1A1* and *CYP1B1* transcription in response to PAHs (39). *CYP1A1* and *CYP1B1* were among the genes most overexpressed in the oral mucosa of smokers (Table 1). Given this background, we determined whether geldanamycin suppressed the induction of *CYP1A1* and *CYP1B1* by TS in MSK-Leuk1 cells, a cellular model of oral leukoplakia (32). Consistent with the findings in the computational analysis, geldanamycin caused dose-dependent suppression of TS-mediated induction of both *CYP1A1* and *CYP1B1* (Fig. 2F).

Discussion

This study provides new insights into the mechanisms underlying the carcinogenic effects of tobacco smoke. Multiple genes encoding enzymes (CYP1A1, CYP1B1, AKRs, ALDH3A1, NQO1, UGTs) involved in carcinogen metabolism were overexpressed in the oral mucosa of smokers. PAHs, an important class of tobacco carcinogen, are likely to mediate some of these expression changes. The AHR, a ligand-activated transcription factor, binds with high affinity to PAHs. Following ligand binding, the AHR translocates to the nucleus where it forms a heterodimer with ARNT. The AHR-ARNT heterodimer then binds to xenobiotic responsive elements in the upstream regulatory region of target genes, resulting in the transcriptional activation of a network of genes, including *CYP1A1* and *CYP1B1* (33). The activation of AHR-mediated signaling leading to induction of xenobiotic metabolism provides a first line of defense against environmental carcinogens. However, the induction of xenobiotic metabolizing enzymes by ligand-activated AHR may also contribute to mutagenesis. PAHs are generally biologically inert and must be metabolically activated by inducible enzymes including CYP1A1 and CYP1B1 to exert their genotoxic actions. For example, benzo[a]pyrene (B[a]P), a potent ligand of the AHR, induces its own metabolism to noncarcinogenic B[a]P phenols (40) and a toxic metabolite anti-7,8-dihydroxy-9,10-epoxy-7,8,9,10-tetrahydrobenzo[a]pyrene, which covalently binds to DNA, forming bulky DNA adducts that induce mutations (41). In addition to CYP1A1 and CYP1B1, PAHs induce the AHR repressor (35). Notably, levels of AHR repressor (AHRR) mRNA were increased in the oral mucosa of smokers. The AHR and AHRR constitute a negative feedback loop of xenobiotic signal transduction. The liganded AHR induces *AHR repressor* transcription, whereas expressed AHRR, in turn, inhibits the function of AHR (35).

A second pathway of PAH activation that causes mutations involves members of the AKR superfamily. PAH *trans*-dihydrodiols are oxidized by the AKRs to redox active and electrophilic PAH *o*-quinones. The AKR-generated B[a]P-7,8-dione enters into futile redox cycles, which amplifies the formation of reactive oxygen species resulting in oxidative DNA damage (22). Oxidative stress is caused by the presence of heavy metals and benzoquinone in tobacco smoke and AKR-mediated production of PAH *o*-quinones. Nrf2, a transcription factor that binds to antioxidant response elements in gene promoters, induces the expression of AKRs, NQO1 and ALDH3A1 (36–38). Expression of AKRs, NQO1 and ALDH3A1 was increased in the oral mucosa of smokers suggesting a counter-response to oxidative stress. Induction of these genes may protect against the damaging effects of harmful quinones and lipid peroxidation break down products. Individuals who fail to mount a normal counter-response may be at increased risk of developing cancer. Thus, it appears that AKRs can both stimulate bioactivation of PAHs leading to increased mutagenesis and participate in a counter-response to oxidative stress.

Increased levels of PTGES (prostaglandin E synthase), ALOX12B (arachidonate 12-lipoxygenase, 12R type) and ALOX15B (arachidonate 15-lipoxygenase, type B) were found in the oral mucosa of smokers. Each of these enzymes is involved in eicosanoid synthesis. These findings are potentially significant because eicosanoids including prostaglandins have been implicated in the development of multiple epithelial malignancies including cancers of the upper aerodigestive tract (42). Notably, use of aspirin, a prototypic inhibitor of prostaglandin synthesis, has been associated with a reduced risk of oral cancer in smokers (43). Based on these findings, future studies are warranted to determine whether levels of eicosanoids including prostaglandin E₂ are increased in the oral mucosa of smokers.

Levels of CD1a and CD207 mRNAs, transcripts expressed in Langerhans cells, were increased in the oral mucosa of smokers compared with never smokers. Changes in transcript levels may occur either because of altered gene expression or a difference in cellular composition. Immunohistochemistry was carried out and revealed an increased number of Langerhans cells in the oral mucosa of smokers. This finding is consistent with previous reports (44) and may reflect a smoking-related change in mucosal immune function. PAH-mediated induction of prostaglandin E₂ has been suggested to stimulate the accumulation of Langerhans cells in skin (45). Lipoxygenase products, e.g., 12-HETE, have been reported to be chemotactic for Langerhans cells (46). It is reasonable to speculate, therefore, that the increased expression of enzymes involved in arachidonic acid metabolism may be causally linked to the increased number of Langerhans cells in the oral mucosa of smokers. Possibly, smoking cessation or treatment with inhibitors of prostaglandin synthesis will result in normalization of the number of Langerhans cells in the oral mucosa and improved immune function.

Levels of CHRNA3, the alpha 3 subunit of the nicotinic acetylcholine receptor, were increased in the oral mucosa of smokers. Nicotine binds to nicotinic acetylcholine receptors leading to activation of Akt signaling and increased epithelial cell survival (47). The alpha-3 subunit is important for mediating these effects of nicotine in epithelial cells (47). Common variants in the nicotinic acetylcholine receptor gene cluster on chromosome 15q24–25.1 have been associated with an increased risk of lung cancer in smokers (48). This region includes the nicotinic acetylcholine receptor subunit gene *CHRNA3*. In theory, nicotine-mediated increased cell survival might lead to the accumulation of DNA adducts and increased mutagenesis and thereby stimulate carcinogenesis. The fact that levels of CHRNA3 are increased in the oral mucosa of smokers underscores the possible role that altered nicotine signaling plays in carcinogenesis.

As mentioned above, women appear to be at increased risk of lung, oral and oropharyngeal cancer compared with men who had similar levels of cigarette smoking exposure (12–14). Our

results provide potential insights into the mechanisms underlying this gender-dependent difference in smoking-related cancer risk. *CYP1B1* was one of the genes most highly overexpressed in the oral mucosa of smokers. *CYP1B1* catalyzes the hydroxylation of estradiol to 4-hydroxy estradiol (4-OHE₂) (49). 4-hydroxycatechol estrogen is a highly reactive catechol estrogen, which is further oxidized to estrogen-3, 4-quinone that can react with DNA to form unstable, adducts, leading to depurination and mutations. Although the link between *CYP1B1*, estrogen metabolism and breast carcinogenesis has been intensively investigated (49), much less attention has been given to aerodigestive malignancies. Marked increases in levels of *CYP1B1* mRNA were found in the oral mucosa of both male and female smokers. Since menstruating females produce higher levels of estrogen than males, it's possible that increased *CYP1B1*-mediated catabolism of estradiol occurs in the aerodigestive tracts of female smokers resulting in enhanced mutagenesis and elevated cancer risk. As shown in Table 3, the magnitude of induction of AKR and UGT family members was greater in the oral mucosa of female than male smokers. By contrast, there was greater suppression of IGF-like family member 1 in the oral mucosa of female than male smokers. Although these findings need to be validated in larger studies, these differences could also help to explain gender-dependent differences in the risk of cancer. For example, as detailed above, activation of PAH-trans-dihydrodiols by AKRs leads to reactive oxygen species-mediated genotoxicity (22).

Our results also suggest that smoking induces similar changes in gene expression in the oral and bronchial epithelium (Table 4, Supplementary Table 5). For example, smoking is associated with increased expression of several genes (*CYP1A1*, *CYP1B1*, *NQO1*, *ALDH3A1*, UGTs) involved in xenobiotic metabolism in both oral and bronchial epithelium. In addition to being important for understanding carcinogenesis, smoking-related changes in xenobiotic metabolism may alter the activity of selected chemopreventive agents (4,5) and targeted anticancer therapies (6) resulting in reduced efficacy. Increased levels of CEACAM family members and *GPX2* were found in both the oral and bronchial epithelium of smokers. These findings are in agreement with other recent studies (21), and suggest that easily accessible oral epithelial cells provide insights into tobacco-induced molecular changes not only in the oral cavity but also in the bronchial epithelium. Use of oral epithelium should be considered as a surrogate tissue in future lung cancer prevention trials.

A powerful tool in computational biology is the ability to compare existing sets of expression data for patterns. The expression profile data from the current study were compared with expression profiles of drugs and small molecule inhibitors. This computational analysis suggested that geldanamycin, an Hsp90 inhibitor, might suppress the changes in the transcriptome induced by cigarette smoke. Consistent with this prediction, we showed that geldanamycin blocked tobacco smoke-mediated induction of *CYP1A1* and *CYP1B1* *in vitro*. These results are consistent with other evidence that Hsp90 inhibitors suppress AHR-mediated activation of *CYP1A1* and *CYP1B1* transcription (50). In addition to suppressing PAH-mediated induction of *CYP1A1* and *CYP1B1*, inhibitors of Hsp90 have multiple other effects. It is predictable, for example, that Hsp90 inhibitors will down-regulate levels of multiple other client proteins and suppress the induction of other AHR regulated genes. AHR-dependent genes play a role in both the activation and detoxification of tobacco carcinogens. Given the overall complexity of these effects, it is uncertain whether systemic or topical treatment with an Hsp90 inhibitor will suppress the mutagenic effects of tobacco smoke or have a chemopreventive effect. Additional studies will be needed to address these questions. More importantly, our findings illustrate the potential use of computational biology as a strategy to identify chemopreventive agents.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Grant Support: National Institutes of Health grants R25 CA105012, T32 CA09685, P01 CA106451 and CTSC UL1-RR024996

We thank Jenny Xiang from the Microarray Core of Weill Cornell Medical College for expert help, Professor Harel Weinstein and Piali Mukherjee for helpful discussions and Kevin C. Dorff from the HRH Prince Alwaleed Bin Talal Bin Abdulaziz Alsaud Institute for Computational Biomedicine at Weill Medical College of Cornell University for webpage design.

References

1. Gritz ER, Dresler C, Sarna L. Smoking, the missing drug interaction in clinical trials: ignoring the obvious. *Cancer Epidemiol Biomarkers Prev* 2005;14:2287–93. [PubMed: 16214906]
2. Tobacco smoke and involuntary smoking. *IARC Monogr Eval Carcinog Risks Hum* 2004;83:1–1438. [PubMed: 15285078]
3. Hecht SS. Tobacco carcinogens, their biomarkers and tobacco-induced cancer. *Nat Rev Cancer* 2003;3:733–44. [PubMed: 14570033]
4. Mayne ST, Lippman SM. Cigarettes: a smoking gun in cancer chemoprevention. *J Natl Cancer Inst* 2005;97:1319–21. [PubMed: 16174848]
5. The α -Tocopherol, β Carotene Cancer Prevention Study Group. The effect of vitamin E and β carotene on the incidence of lung cancer and other cancers in male smokers. *N Engl J Med* 1994;330:1029–35. [PubMed: 8127329]
6. Hamilton M, Wolf JL, Rusk J, et al. Effects of smoking on the pharmacokinetics of erlotinib. *Clin Cancer Res* 2006;12:2166–71. [PubMed: 16609030]
7. Fox JL, Rosenzweig KE, Ostroff JS. The effect of smoking status on survival following radiation therapy for non-small cell lung cancer. *Lung Cancer* 2004;44:287–93. [PubMed: 15140541]
8. Pantarotto J, Malone S, Dahrouge S, Gallant V, Eapen L. Smoking is associated with worse outcomes in patients with prostate cancer treated by radical radiotherapy. *BJU Int* 2007;99:564–9. [PubMed: 17166241]
9. Shepherd FA, Rodrigues Pereira J, Ciuleanu T, et al. Erlotinib in previously treated non-small-cell lung cancer. *N Engl J Med* 2005;353:123–32. [PubMed: 16014882]
10. Browman GP, Wong G, Hodson I, et al. Influence of cigarette smoking on the efficacy of radiation therapy in head and neck cancer. *N Engl J Med* 1993;328:159–63. [PubMed: 8417381]
11. Khuri FR, Lee JJ, Lippman SM, et al. Randomized phase III trial of low-dose isotretinoin for prevention of second primary tumors in stage I and II head and neck cancer patients. *J Natl Cancer Inst* 2006;98:441–50. [PubMed: 16595780]
12. Risch HA, Howe GR, Jain M, Burch JD, Holowaty EJ, Miller AB. Are female smokers at higher risk for lung cancer than male smokers? a case-control analysis by histologic type. *Am J Epidemiol* 1993;138:281–293. [PubMed: 8395141]
13. International Early Lung Cancer Action Program Investigators. Women's susceptibility to tobacco carcinogens and survival after diagnosis of lung cancer. *JAMA* 2006;296:180–184. [PubMed: 16835423]
14. Neugut AI, Jacobson JS. Women and lung cancer: gender equality at a crossroad? *JAMA* 296:218–19. [PubMed: 16835429]
15. Powell CA, Klares S, O'Connor G, Brody JS. Loss of heterozygosity in epithelial cells obtained by bronchial brushing: clinical utility in lung cancer. *Clin Cancer Res* 1999;5:2025–2034. [PubMed: 10473082]
16. Franklin WA, Gazdar AF, Haney J, et al. Widely dispersed p53 mutation in respiratory epithelium. A novel mechanism for field carcinogenesis. *J Clin Invest* 1997;100:2133–2137. [PubMed: 9329980]
17. Wistuba II, Mao L, Gazdar AF. Smoking molecular damage in bronchial epithelium. *Oncogene* 2002;21:7298–7306. [PubMed: 12379874]
18. Guo M, House MG, Hooker C, et al. Promoter hypermethylation of resected bronchial margins: a field defect of changes? *Clin Cancer Res* 2004;10:5131–5136. [PubMed: 15297416]

19. 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]
20. Spira A, Beane JE, Shah V, et al. Airway epithelial gene expression in the diagnostic evaluation of smokers with suspect lung cancer. *Nature Med* 2007;13:361–366. [PubMed: 17334370]
21. Sridhar S, Schembri F, Zeskind J, et al. Smoking-induced gene expression changes in the bronchial airway are reflected in nasal and buccal epithelium. *BMC Genomics* 2008;9:259. [PubMed: 18513428]
22. Park JH, Mangal D, Tacka KA, et al. Evidence for the aldo-keto reductase pathway of polycyclic aromatic *trans*-dihydrodiol activation in human lung A549 cells. *Proc Natl Acad Sci U S A* 2008;105:6846–6851. [PubMed: 18474869]
23. Irizarry RA, Bolstad BM, Collin F, Cope LM, Hobbs B, Speed TP. Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res* 2003;31:e15. [PubMed: 12582260]
24. Cope LM, Irizarry RA, Jaffee HA, Wu Z, Speed TP. A benchmark for Affymetrix GeneChip expression measures. *Bioinformatics* 2004;20:323–31. [PubMed: 14960458]
25. Lindros KO, Oinonen T, Kettunen E, Sippel H, Muro-Lupori C, Koivusalo M. Aryl hydrocarbon receptor-associated genes in rat liver: regional coinduction of aldehyde dehydrogenase 3 and glutathione transferase Ya. *Biochem Pharmacol* 1998;55:413–21. [PubMed: 9514075]
26. Erichsen TJ, Ehmer U, Kalthoff S, et al. Genetic variability of aryl hydrocarbon receptor (AhR)-mediated regulation of the human UDP glucuronosyltransferase (UGT) 1A4 gene. *Toxicol Appl Pharmacol* 2008;230:252–60. [PubMed: 18433817]
27. Landkisch TO, Gillman TC, Erichsen TJ, et al. Aryl hydrocarbon receptor-mediated regulation of the human estrogen and bile acid UDP-glucuronosyltransferase 1A3 gene. *Arch Toxicol* 2008;82:573–82. [PubMed: 18677463]
28. Olinga P, Elferink MG, Draaisma AL, et al. Coordinated induction of drug transporters and phase I and II metabolism in human liver slices. *Eur J Pharm Sci* 2008;33:380–9. [PubMed: 18328680]
29. Hosack DA, Dennis G Jr, Sherman BT, Lane HC, Lempicki RA. Identifying biological themes within lists of genes with EASE. *Genome Biol* 2003;4:R70. [PubMed: 14519205]
30. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 2005;102:15545–50. [PubMed: 16199517]
31. Lamb J, Crawford ED, Peck D, et al. The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* 2006;313:1929–1935. [PubMed: 17008526]
32. Sacks PG. Cell, tissue and organ culture as *in vitro* models to study the biology of squamous cell carcinomas of the head and neck. *Cancer Metastasis Rev* 1996;15:27–51. [PubMed: 8842478]
33. Gumu ZH, Du B, Kacker A, et al. Effects of tobacco smoke on gene expression and cellular pathways in a cellular model of oral leukoplakia. *Cancer Prev Res* 2008;1:100–111.
34. Lowry OH, Rosebrough NJ, Farr AL, Randall RJ. Protein measurement with the Folin phenol reagent. *J Biol Chem* 1951;193:265–75. [PubMed: 14907713]
35. Stevens EA, Mezrich JD, Bradfield CA. The aryl hydrocarbon receptor: a perspective on potential roles in the immune system. *Immunology* 2009;127:299–311. [PubMed: 19538249]
36. Lou H, Du S, Ji Q, et al. Induction of AKR1C2 by phase II inducers: identification of a distal consensus antioxidant response element regulated by NRF2. *Mol Pharmacol* 2006;69:1662–72. [PubMed: 16478829]
37. Penning TM, Drury JE. Human aldo-keto reductases: Function, gene regulation and single nucleotide polymorphisms. *Arch Biochem Biophys* 2007;464:241–50. [PubMed: 17537398]
38. Sreerama L, Sladek NE. Three different stable human breast adenocarcinoma sublines that overexpress ALDH3A1 and certain other enzymes, apparently as a consequence of constitutively up-regulated gene transcription mediated by transactivated EpREs (electrophile responsive elements) present in the 5'-upstream regions of these genes. *Chem Biol Interact* 2001;130–132:247–60.
39. Kekatpure VD, Dannenberg AJ, Subbaramaiah K. HDAC6 modulates Hsp90 chaperone activity and regulates activation of aryl hydrocarbon receptor signaling. *J Biol Chem* 2009;284:7436–7445. [PubMed: 19158084]
40. Conney AH, Miller EC, Miller JA. Substrate-induced synthesis and other properties of benzopyrene hydroxylase in rat liver. *J Biol Chem* 1957;228:753–66. [PubMed: 13475357]

41. Volk DE, Thiviyanathan V, Rice JS, et al. Solution structure of a *cis*-opened (10*R*)-*N*6-deoxyadenosine adduct of (9*S*,10*R*)-9,10-epoxy-7,8,9,10-tetrahydrobenzo[*a*]pyrene in a DNA duplex. *Biochemistry* 2003;42:1410–20. [PubMed: 12578353]
42. Dannenberg AJ, Subbaramaiah K. Targeting cyclooxygenase-2 in human neoplasia: rationale and promise. *Cancer Cell* 2003;4:431–6. [PubMed: 14706335]
43. Jayaprakash V, Rigual NR, Moysich KB, et al. Chemoprevention of head and neck cancer with aspirin: a case-control study. *Arch Otolaryngol Head Neck Surg* 2006;132:1231–6. [PubMed: 17116820]
44. Barrett AW, Williams DM, Scott J. Effect of tobacco and alcohol consumption on the Langerhans cell population of human lingual epithelium determined using a monoclonal antibody against HLADR. *J Oral Pathol Med* 1991;20:49–52. [PubMed: 2016696]
45. Andrews FJ, Halliday GM, Narkowicz CK, Muller HK. Indomethacin inhibits the chemical carcinogen benzo(a)pyrene but not dimethylbenz(a)anthracene from altering Langerhans cell distribution and morphology. *Br J Derm* 1991;124:29–36.
46. Arenberger P, Kemeny L, Rupec R, Bieber T, Ruzicka T. Langerhans cells of the human skin possess high-affinity 12 (S)-hydroxyicosatetraenoic acid receptors. *Eur J Immunol* 1992;22:2469–72. [PubMed: 1325349]
47. West KA, Brognard J, Clark AS, et al. Rapid Akt activation by nicotine and a tobacco carcinogen modulates the phenotype of normal human airway epithelial cells. *J Clin Invest* 2003;111:81–90. [PubMed: 12511591]
48. Amos CI, Wu X, Broderick P, et al. Genome-wide association scan of tag SNPs identifies a susceptibility locus for lung cancer at 15q25.1. *Nature Genet* 2008;40:616–622. [PubMed: 18385676]
49. Yager JD, Davidson NE. Estrogen carcinogenesis in breast cancer. *N Engl J Med* 2006;354:270–282. [PubMed: 16421368]
50. Hughes D, Guttenplan JB, Marcus CB, Subbaramaiah K, Dannenberg AJ. Heat shock protein 90 inhibitors suppress aryl hydrocarbon receptor-mediated activation of CYP1A1 and CYP1B1 transcription and DNA adduct formation. *Cancer Prev Res* 2008;1:485–493.

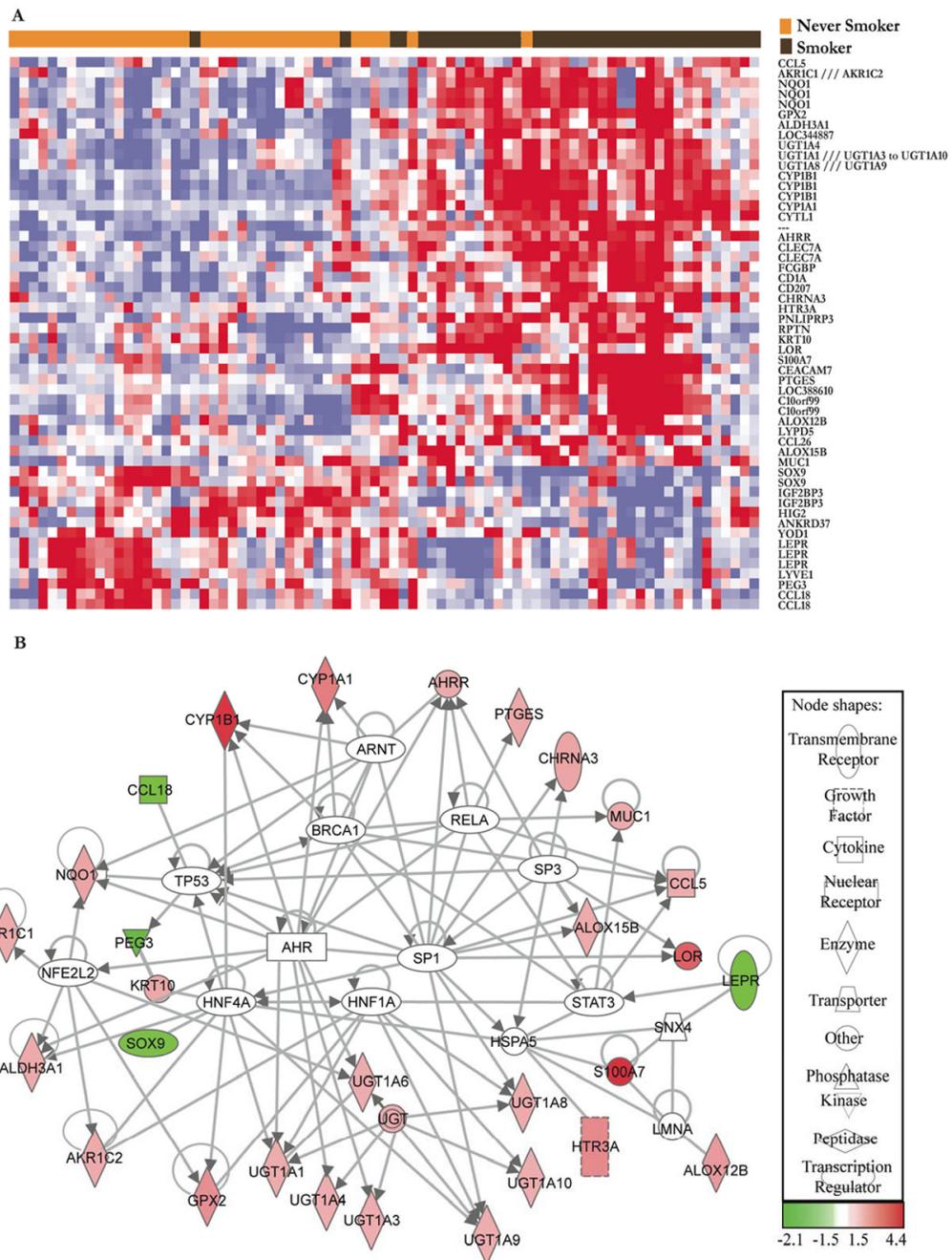


Figure 1.
A, Unsupervised hierarchical clustering of the expression of probesets differentially expressed in the oral mucosa of smokers vs. never smokers. Smokers and never smokers cluster primarily into two distinct groups. Each column corresponds to the expression profile of an oral mucosal biopsy, and each row corresponds to an mRNA. The color in each cell reflects the level of expression of the corresponding mRNA relative to its mean level of expression in the entire set of biopsy samples. In this heatmap, the increasing intensities of red signify that a specific mRNA has a higher expression in the given sample whereas the increasing intensities of blue mean that this mRNA has lower expression. White indicates mean level of expression. B, Direct interaction network of differentially expressed genes generated using IPA and other known

interactions. The white nodes represent genes with no significant expression change that potentially contribute to the effects of smoking.

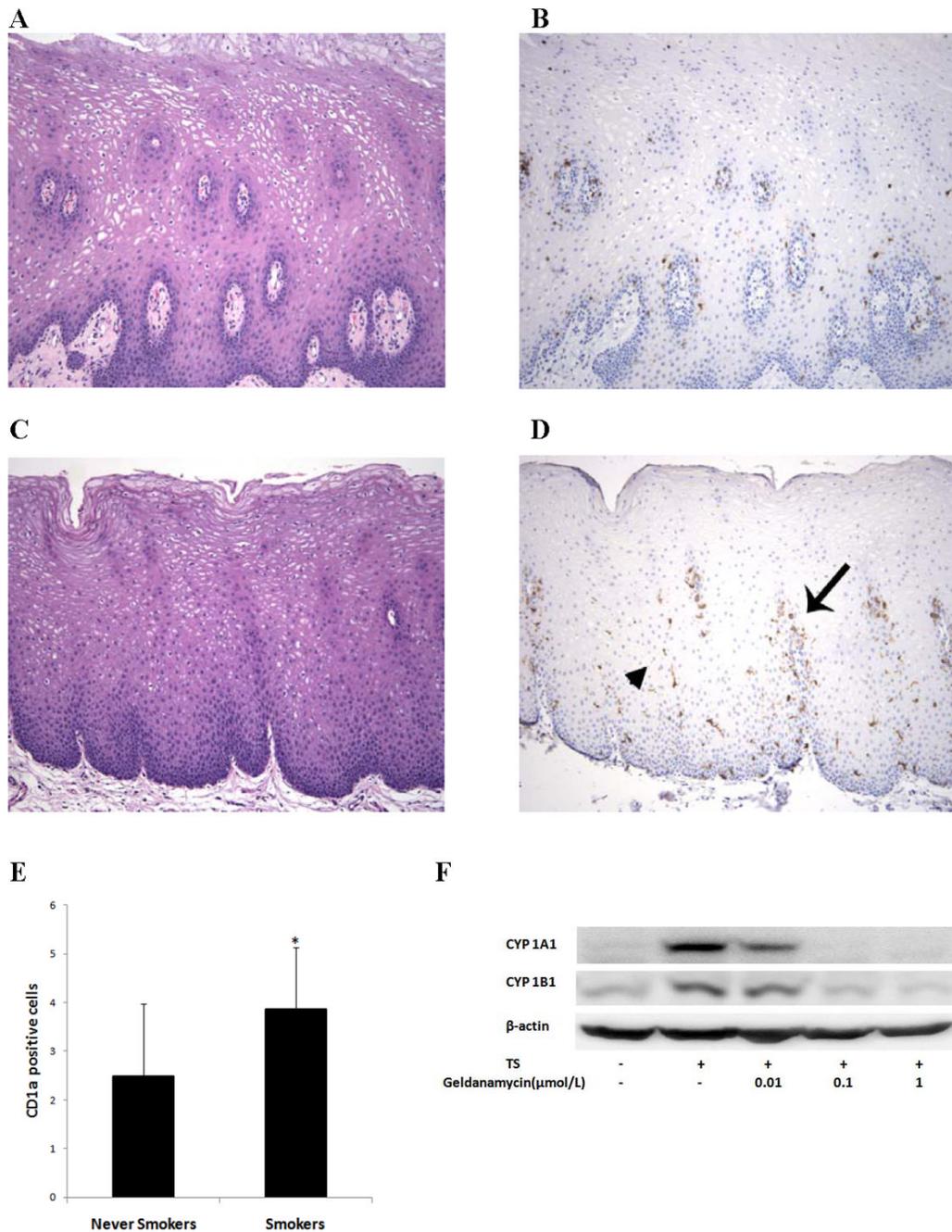


Figure 2.

Increased numbers of Langerhans cells were found in the oral mucosa of smokers. Non-neoplastic oral mucosae from never smokers (A) and smokers (C) were morphologically similar, but samples from never smokers showed relatively few Langerhans cells (B) compared to those from smokers (D), which contained numerous Langerhans cells in the peripapillary (arrow) and interpapillary mucosa (arrowhead). [Magnification 100X for panels A–D; panels A and C stained with hematoxylin and eosin; panels B and D stained with CD1a immunostain and hematoxylin]. E, Intraepithelial cells that displayed moderate to strong cytoplasmic staining for CD1a in dendritic-type cellular processes were quantified in the peripapillary, interpapillary and superficial epithelium. A statistically significant increase in the number of

CD1a positive cells was found in all three regions in smokers compared to never smokers ($P < 0.001$, < 0.001 and $= 0.032$ for peripapillary, interpapillary and superficial areas, respectively). Panel E reflects the total number of CD1a positive cells in the three regions. Columns, means; bars, S.E.; $n = 27/\text{group}$. *, $P < 0.001$. F, MSK-Leuk1 cells were pretreated with vehicle or the indicated concentration of geldanamycin for 2 h. Subsequently, cells received vehicle or TS for 5 h and were then harvested for Western blot analysis. Cellular lysate protein (100 $\mu\text{g}/\text{lane}$) was loaded onto a 10% SDS-polyacrylamide gel, electrophoresed and subsequently transferred onto nitrocellulose. Immunoblots were probed with antibodies specific for CYP1A1, CYP1B1 and β -actin.

Table 1

Differentially expressed genes in the oral mucosa of smokers vs. never smokers with corresponding fold-changes and P values. Detailed annotations are provided at <http://physiology.med.cornell.edu/go/smoke>.

Gene Name	Affymetrix ID	Fold	P value	Gene Title
S100A7	205916_at	4.4	3.1E-02	S100 calcium binding protein A7
RPTN	1553454_at	4.3	1.7E-02	repetin
CYP1B1	202437_s_at	4.2	1.4E-11	cytochrome P450, family 1, subfamily B, polypeptide 1
CYP1B1	202436_s_at	3.2	6.9E-10	cytochrome P450, family 1, subfamily B, polypeptide 1
LOR	207720_at	3.2	9.4E-03	loricrin
CEACAM7	206198_s_at	3.0	3.1E-02	carcinoembryonic antigen-related cell adhesion molecule 7
CYP1A1	205749_at	2.5	1.1E-07	cytochrome P450, family 1, subfamily A, polypeptide 1
CYP1B1	202435_s_at	2.5	2.7E-08	cytochrome P450, family 1, subfamily B, polypeptide 1
HTR3A	216615_s_at	2.2	7.7E-03	5-hydroxytryptamine (serotonin) receptor 3A
GPX2	202831_at	2.1	3.1E-04	glutathione peroxidase 2 (gastrointestinal)
FCGBP	203240_at	2.0	2.9E-05	Fc fragment of IgG binding protein
---	227452_at	2.0	7.8E-10	Full-length cDNA clone CS0DD005YM12 of Neuroblastoma Cot 50-normalized of Homo sapiens (human)
CCL26	223710_at	1.9	2.8E-02	chemokine (C-C motif) ligand 26
PNLIPRP3	1558846_at	1.9	6.2E-03	pancreatic lipase-related protein 3
ALOX12B	207381_at	1.9	1.2E-02	arachidonate 12-lipoxygenase, 12R type
LOC388610	227862_at	1.9	2.9E-02	hypothetical LOC388610
CD207	220428_at	1.8	1.3E-05	CD207 molecule, langerin
CHRNA3	210221_at	1.7	2.0E-04	cholinergic receptor, nicotinic, alpha 3
CYTL1	219837_s_at	1.7	1.6E-07	cytokine-like 1
NQO1	201468_s_at	1.7	6.7E-04	NAD(P)H dehydrogenase, quinone 1
NQO1	210519_s_at	1.6	2.1E-03	NAD(P)H dehydrogenase, quinone 1
NQO1	201467_s_at	1.5	4.9E-03	NAD(P)H dehydrogenase, quinone 1
CLEC7A	1555756_a_at	1.6	1.6E-02	C-type lectin domain family 7, member A
CLEC7A	221698_s_at	1.6	9.4E-04	C-type lectin domain family 7, member A
LOC344887	241418_at	1.6	6.2E-03	similar to hCG2041270
PTGES	210367_s_at	1.6	4.9E-02	prostaglandin E synthase
KRT10	207023_x_at	1.6	3.5E-02	keratin 10 (epidermolytic hyperkeratosis; keratosis palmaris et plantaris)
C10orf99	227736_at	1.6	9.4E-03	chromosome 10 open reading frame 99
C10orf99	227735_s_at	1.6	1.2E-02	chromosome 10 open reading frame 99
ALDH3A1	205623_at	1.6	2.9E-05	aldehyde dehydrogenase 3 family, member A1
ALOX15B	206714_at	1.6	3.5E-02	arachidonate 15-lipoxygenase, type B
UGT1A6 ///	221305_s_at	1.5	1.1E-02	UDP glucuronosyltransferase 1 family, polypeptide A6 ///
UGT1A8 ///				glucuronosyltransferase 1 family, polypeptide A8 ///
UGT1A9				UDP glucuronosyltransferase 1 family, polypeptide A9
MUC1	213693_s_at	1.5	4.7E-02	mucin 1, cell surface associated
AKR1C1 ///	1555854_at	1.5	1.9E-02	aldo-keto reductase family 1, member C1 ///
AKR1C2				aldo-keto reductase family 1, member C2
AHRR ///	229354_at	1.5	3.8E-08	aryl-hydrocarbon receptor repressor ///

Gene Name	Affymetrix ID	Fold	P value	Gene Title
PDCD6				programmed cell death 6
LYPD5	236039_at	1.5	4.7E-02	LY6/PLAUR domain containing 5
UGT1A1 ///	215125_s_at	1.5	2.3E-03	UDP glucuronosyltransferase 1 family, polypeptide UDP
UGT1A3 to				glucuronosyltransferase 1 family, polypeptide A3 to A10
UGT1A10				
CCL5	1405_i_at	1.5	1.7E-02	chemokine (C-C motif) ligand 5
UGT1A1 ///	207126_x_at	1.5	3.1E-04	UDP glucuronosyltransferase 1 family, polypeptide A1///
UGT1A4 ///				UDP glucuronosyltransferase 1 family, polypeptide A4///
UGT1A6 ///				UDP glucuronosyltransferase 1 family, polypeptide A6///
UGT1A8 to				UDP glucuronosyltransferase 1 family, polypeptide A8 to A10
UGT1A10				
CD1a	210325_at	1.5	9.6E-03	CD1a molecule
LYVE1	220037_s_at	-1.5	1.0E-02	lymphatic vessel endothelial hyaluronan receptor 1
YOD1	215150_at	-1.5	3.0E-02	YOD1 OTU deubiquinating enzyme 1 homolog (<i>S. cerevisiae</i>)
CCL18	209924_at	-1.5	3.2E-02	chemokine (C-C motif) ligand 18 (pulmonary and activation-regulated)
ANKRD37	227337_at	-1.5	3.7E-02	ankyrin repeat domain 37
SOX9	202936_s_at	-1.5	2.4E-03	SRY (sex determining region Y)-box 9
SOX9	202935_s_at	-1.5	1.1E-02	SRY (sex determining region Y)-box 9
LEPR	211355_x_at	-1.5	4.4E-03	leptin receptor
LEPR	211354_s_at	-1.6	4.3E-03	leptin receptor
LEPR	211356_x_at	-1.6	2.4E-03	leptin receptor
IGF2BP3	203819_s_at	-1.7	5.4E-04	insulin-like growth factor 2 mRNA binding protein 3
IGF2BP3	203820_s_at	-1.6	1.4E-02	insulin-like growth factor 2 mRNA binding protein 3
CCL18	32128_at	-1.6	3.4E-02	chemokine (C-C motif) ligand 18 (pulmonary and activation-regulated)
HIG2	1554452_a_at	-1.9	3.8E-02	hypoxia-inducible protein 2
PEG3	209242_at	-2.1	1.1E-02	paternally expressed 3

Table 2

Functional gene groups altered in the oral mucosa of smokers vs. never smokers.

Gene Set	
Pathways Enriched in Smokers (using GSEA v2)	FDR
Metabolism of Xenobiotics by Cytochromes P450	0.01↑ (KEGG)
Androgen and Estrogen Metabolism	0.110↑ (KEGG)
Eicosanoid Synthesis	0.075↑ (GenMAPP)
Prostaglandin and Leukotriene Metabolism	0.128↑ (GenMAPP)
Glutathione Metabolism	0.166↑ (KEGG), 0.093↑ (GENMAPP)
GO Groups Enriched in Smokers (using EASE)	Bonferroni p
GO Molecular Function: Electron transporter activity	0.00085 (8/27)↑
GO Molecular Function: Oxidoreductase activity	0.049 (8/27)↑

Gender-dependent differences in the effect of smoking on the expression of select genes in the oral mucosa. Differentially expressed genes (fold-changes) in the oral mucosa of smokers vs. never smokers for females and males, respectively. Interaction P values indicate that the magnitude of the change in gene expression induced by smoking was greater in females than males for each of the genes shown below. [Annotations from Nov 2008 NetAffx]

Table 3

Gene Name	Affymetrix ID	Female Fold	Male Fold	Interaction P value	Gene Title
AKRIC3	209160_at	1.6	1.1	0.0196	Aldo-keto reductase family 1, member C3 (3-alpha hydroxysteroid dehydrogenase, type II)
AKRIC2	209699_x_at	1.5	1.1	0.0266	aldo-keto reductase family 1, member C2
UGT1A1//	206094_x_at	1.7	1.3	0.0313	
UGT1A3 to					UDP glucuronosyltransferase I family, polypeptide A1, A3 to A10
UGT1A10					
UGT1A1//	204532_x_at	1.8	1.3	0.0346	
UGT1A4//					
UGT1A6//					
UGT1A8 to					UDP glucuronosyltransferase I family, polypeptide A1, A4, A6, A8 to A10
UGT1A10					
UGT1A1//	207126_x_at	1.8	1.3	0.0423	
UGT1A4//					
UGT1A6//					
UGT1A8 to					UDP glucuronosyltransferase I family, polypeptide A1, A4, A6, A8 to A10
UGT1A10					
IGFL1	239430_at	-2.1	-1.0	0.0210	IGF-like family member 1

Table 4

Genes that are overexpressed in the oral mucosa of smokers are also commonly overexpressed in the airways of smokers. Differentially expressed genes (fold-changes) in the oral and airway mucosa of smokers vs. never smokers with associated P values (Anova). The CEACAM family genes CEACAM5 and CEACAM6 are induced in the airways of smokers whereas CEACAM7 is induced in the oral mucosa of smokers. FCGBP is induced in the oral mucosa while repressed in the airway of smokers.

Oral vs. Airway				
Gene Name	Affymetrix ID	Oral Mucosa Fold (P)	Airway Fold (P)	Gene Title
CYP1B1	202437_s_at	4.2 (1.1e-11)	8.1 (1.2e-07)	cytochrome P450, family 1, subfamily B, polypeptide 1
CYP1B1	202436_s_at	3.2 (5.5e-10)	7.1 (2.4e-07)	cytochrome P450, family 1, subfamily B, polypeptide 1
CYP1A1	205749_at	2.5 (3.1e-05)	2.8 (5.6e-04)	cytochrome P450, family 1, subfamily A, polypeptide 1
GPX2	202831_at	2.1 (3.1e-04)	3.3 (5.8e-14)	glutathione peroxidase 2 (gastrointestinal)
NQO1	201468_s_at	1.7 (6.9e-04)	3.7 (4.1e-13)	NAD(P)H dehydrogenase, quinone 1
NQO1	210519_s_at	1.6 (1.9e-03)	3.6 (4.5e-14)	NAD(P)H dehydrogenase, quinone 1
NQO1	201467_s_at	1.5 (4.4e-03)	3.2 (6.7e-12)	NAD(P)H dehydrogenase, quinone 1
ALDH3A1	205623_at	1.6 (7.5e-04)	6.5 (3.4e-12)	aldehyde dehydrogenase 3 family, memberA1
UGT1A1 ///	215125_s_at	1.5 (2.1e-03)	2.2 (1.4e-08)	
UGT1A3 to UGT1A10				UDP glucuronosyltransferase 1 family, polypeptide A1 /// A3 to A10
UGT1A1 ///	207126_x_at	1.5 (3.1e-04)	1.8 (2.0e-08)	
UGT1A4 /// UGT1A6 ///				
UGT1A8 to UGT1A10				UDP glucuronosyltransferase 1 family, polypeptide A1 /// A4 /// A6 /// A8 to A10
MUC1	207847_s_at	(>0.05)	1.6 (0.021)	mucin 1, cell surface associated
MUC1	213693_s_at	1.5 (0.043)	1.4 (0.015)	mucin 1, cell surface associated
CEACAM5	201884_at	1.3 (0.011)	5.4 (3.8e-11)	Carcinoembryonic antigen-related cell adhesion molecule 5
CEACAM6	203757_s_at	(>0.05)	2.4 (1.3e-4)	Carcinoembryonic antigen-related cell adhesion molecule 6
CEACAM6	211657_at	(>0.05)	2.3 (2.1e-4)	Carcinoembryonic antigen-related cell adhesion molecule 6
CEACAM7	206198_s_at	3.0 (0.031)	(>0.05)	Carcinoembryonic antigen-related cell adhesion molecule 7
FCGBP	203240_at	2.0 (3.1e-05)	-2.3 (0.01)	Fc fragment of IgG binding protein