

Supplementary Methods

Effects of gender on smokers' transcriptome. Two types of analyses were combined: (i) differentially expressed genes were identified separately for female and male smokers vs. never smokers, employing one-way ANOVA within GeneSpring GX 7 (as described above, except with no multiple testing correction), (ii) the interaction between smoking status and gender was investigated for four comparison groups: never smoker, male; never smoker, female; smoker, male; smoker, female using unbalanced two-way ANOVA with asymptotic P value computation within GeneSpring GX 9 (Agilent Technologies). Genes that were significant in analysis (ii) and at 1.5-fold cutoff in at least one gender group after analysis (i) are shown in Table 3 with associated smoking status and gender interaction P values.

Comparison of effects of smoking on the oral and bronchial epithelium.

Enrichment Analysis. The distribution of the airway epithelium differentially expressed genes within the entire rank-ordered oral epithelium gene set was scored to identify the relationship between the gene expression responses to tobacco smoke in the oral and airway epithelia, using GSEA v2 (30). Airway genes were collapsed to HUGO symbols and separated into up- and down-regulated gene sets comprising 110 and 21 genes, respectively. Oral data input were the normalized ratio values of all probe sets in HGU133 Plus2 arrays collapsed to HGU133A array with HUGO symbols (the smaller chip was used for previous airway experiments). A reverse analysis was performed by searching for enrichment of airway data in oral gene sets. A significant P value in this analysis suggests that the expression of genes that are affected by smoking status in oral biopsies are also altered in the bronchial epithelium, even if they are not significantly differentially expressed. The core of the airway (and oral) gene sets that account for the enrichment score (30) is defined as the leading-edge subset, and listed in Supplementary Table 5.

Supplementary Table 1. Demographic characteristics of never smokers and smokers.

Never Smokers	Female (N=20)	Male (N=20)	P value
Age (years)			
Median	45	45	0.51
Range	(27-66)	(30-55)	
Smokers	Female (N = 19)	Male (N=20)	P value
Age (years)			
Median	43	45.5	0.87
Range	27-63	30-54	
Pack Years			
Median	25	32.5	0.37
Range	15-66	15-60	

Supplementary Table 2. List of qPCR primers. F and R stand for Forward and Reverse, respectfully.

Gene Name	Primers	Name
S100A7	F: 5'-AAGCCTGCTGACGATGATGAAGGA-3' R: 5'-ATGGCTCTGCTTGTGGTAGTCTGT-3'	S100 calcium binding protein A7
CYP1B1	F: 5'-ACGTACCGGCCACTATCACT-3' R: 5'-CTCGAGTCTGCACATCAGGA-3'	Cytochrome P450, family 1, subfamily B, polypeptide 1
CYP1A1	F: 5'-CCTGCTAGGGTTAGGAGGTC-3' R: 5'-GCTCAGCCTAGTTCAAGCAG-3'	Cytochrome P450, family 1, subfamily A, polypeptide 1
CD207	F: 5'-CAACAATGCTGGGAACAATG-3' R: 5'-CTGTCCTGTCACGGTTCTGA-3'	CD 207 molecule, langerin
CHRNA3	F: 5'-CTGGGGAGGTGACTTGGATA-3' R: 5'-CAGGAACCGAACTTCATGGT-3'	Cholinergic receptor, nicotinic, alpha 3
NQO1	F: 5'-GCCGCAGACCTTGTGATATT-3' R: 5'-TGAACACTCGCTCAAACCAG-3'	NAD(P)H dehydrogenase, quinone 1
PTGES	F: 5'-TGGAGACCATCTACCCC-3' R: 5'-CCACGAGGAAGACCAGGA-3'	Prostaglandin E synthase
AHRR	F: 5'-GTGCGAATCGGAACTGCATGGAAA-3' R: 5'-TCAGTCTGTTCCCTGAGCACCAAA-3'	Aryl-hydrocarbon receptor repressor
CD1a	F: 5'-TGGCTGAGTGATTTGCAGAC-3' R: 5'-TTCCTTCCACTCCTCATTGC-3'	CD1a molecule
LEPR	F: 5'-AGCCAAACTCAACGACACTCTCCT-3' R: 5'-CAATGGTGGGCTGGACCAAGAAAT-3'	Leptin receptor
IGF2BP3	F: 5'-AGAAGTGCACGGGAAACCCATAGA-3' R: 5'-TGTTCACTTGCTCACAGCTCTCCA-3'	Insulin-like growth factor 2 mRNA binding protein 3

Supplementary Table 3. Microarray results (fold-change) indicated that the expression of each of the 11 genes shown below was affected by smoking status (see Table 1). In an attempt to validate the microarray findings, qPCR was carried out on RNA isolated from the oral mucosa of 10 randomly selected smokers and 10 randomly selected never smokers. Fold-changes (smokers vs. never smokers) and P values for qPCR results are shown. For each of the genes, the qPCR findings were qualitatively consistent with the microarray results.

	qPCR <i>P</i> value	qPCR Fold	Microarray Fold
S100A7	0.04	4.3	4.4
CYP1B1	0.0004	6.3	4.2
CYP1A1	0.04	2.5	2.5
CD207	0.006	2.4	1.8
CHRNA3	0.02	2.6	1.7
NQO1	0.02	2.3	1.7
PTGES	0.047	1.6	1.6
AHRR	0.001	3.6	1.5
CD1a	0.01	1.5	1.5
LEPR	0.049	-1.5	-1.6
IGF2BP3 (IMP3)	0.03	-1.4	-1.7

Supplementary Table 4. Differentially expressed genes in the oral mucosa of never smoker females vs. never smoker males and corresponding fold-differences and P values.

Gene Name	Affymetrix ID	Fold	P-value	Gene Title
XIST	224588_at	48.4	5.3E-05	X (inactive)-specific transcript (non-protein coding)
XIST	227671_at	31.6	5.3E-05	X (inactive)-specific transcript (non-protein coding)
XIST	221728_x_at	18.7	1.9E-05	X (inactive)-specific transcript (non-protein coding)
XIST	214218_s_at	17.1	1.9E-05	X (inactive)-specific transcript (non-protein coding)
XIST	224590_at	13.1	2.0E-05	X (inactive)-specific transcript (non-protein coding)
XIST	224589_at	12.3	2.0E-05	X (inactive)-specific transcript (non-protein coding)
---	228919_at	4.4	1.2E-03	---
KDM6A	203990_s_at	1.6	3.0E-02	lysine (K)-specific demethylase
KDM6A	203991_s_at	1.5	7.3E-03	lysine (K)-specific demethylase
---	234931_at	-1.9	2.0E-04	---
DDX3Y	205001_s_at	-2.1	2.2E-02	DEAD (Asp-Glu-Ala-Asp) box polypeptide 3, Y-linked
---	239858_at	-2.1	2.1E-03	---
CYorf15B	223645_s_at	-2.6	7.2E-04	chromosome Y open reading frame 15B
CYorf15B	214131_at	-2.9	7.8E-04	chromosome Y open reading frame 15B
CYorf15B	223646_s_at	-3.3	3.3E-04	chromosome Y open reading frame 15B
CYorf15A	232618_at	-3.9	7.9E-05	chromosome Y open reading frame 15A
CYorf14	207063_at	-4.0	9.8E-05	chromosome Y open reading frame 14
ZFY	230760_at	-4.0	9.8E-05	zinc finger protein, Y-linked
USP9Y	206624_at	-4.5	3.5E-04	ubiquitin specific peptidase 9, Y-linked
CYorf15A	236694_at	-5.0	1.6E-04	chromosome Y open reading frame 15A
KDM5D	206700_s_at	-6.5	8.7E-05	lysine (K)-specific demethylase 5D
EIF1AY	204410_at	-6.8	6.4E-05	eukaryotic translation initiation factor 1A, Y-linked
USP9Y	228492_at	-7.7	8.7E-05	ubiquitin specific peptidase 9, Y-linked
DDX3Y	205000_at	-10.6	2.5E-04	DEAD (Asp-Glu-Ala-Asp) box polypeptide 3, Y-linked
EIF1AY	204409_s_at	-11.6	1.1E-04	eukaryotic translation initiation factor 1A, Y-linked
RPS4Y1	201909_at	-20.0	5.4E-04	ribosomal protein S4, Y-linked

Supplementary Table 5. The core members of the airway and oral mucosa genesets that account for the enrichment signal (the ‘leading edge subset’) in the GSEA analysis. The geneset from the oral mucosa of smokers enriched in genes significantly induced in the airway of smokers (P <0.001; marked X in column Oral/Air). Similarly, the geneset from smokers’ airways show enrichment in genes both significantly induced or repressed in the oral mucosa of smokers (P<0.001, induced; P = 0.062, repressed; marked X in column Air/Oral). Opposite enrichment was also detected whereby a geneset from the oral mucosa appears to be induced in smokers but repressed in the smokers’ airway (P=0.005; marked X in section Opposite Enrichment).

GSEA INDUCED	P VALUE	Oral/Air P<0.001	Air/Oral P <0.001
GENE NAME	DESCRIPTION	GSEA	GSEA
ABCB6	ATP binding cassette, sub-family B, member 6	X	
AKR1B10	aldo-keto reductase family 1, member B10	X	
AKR1C1	aldo-keto reductase family 1, member C1	X	
AKR1C2	aldo-keto reductase family 1, member C2	X	
AKR1C3	aldo-keto reductase family 1, member C3	X	
ALDH3A1	aldehyde dehydrogenase 3 family, member A1	X	X
CA12	Carbonic anhydrase	X	
CEACAM5	carcinoembryonic antigen-related cell adhesion molecule	X	
CTSC	cathepsin C	X	
CXCL14	chemokine (C-X-C motif) ligand 14	X	
CYP1A1	cytochrome P450, family 1, subfamily A, polypeptide 1	X	X
CYP1B1	cytochrome P450, family 1, subfamily B, polypeptide 1	X	X
FTL	ferritin, light polypeptide	X	
GDF15	growth differentiation factor 15	X	
GNP	glucosamine (UDP-N-acetyl)-2-epimerase/N-acetylmannosamine kinase	X	
GPX2	glutathione peroxidase 2 (gastrointestinal)	X	X
IMPA2	inositol(myo)-1(or 4)-monophosphatase 2	X	
LAPTM4B	lysosomal associated protein transmembrane 4 beta	X	
MSMB	microseminoprotein, beta-	X	
MUC1	mucin 1, cell surface associated	X	X
NQO1	NAD(P)H dehydrogenase, quinone 1	X	X
PIR	pirin (iron-binding nuclear protein)	X	
SLC7A11	solute carrier family 7, member 11	X	
TKT	transketolase (Wernicke-Korsakoff syndrome)	X	
TXNDC5	thioredoxin domain containing 5	X	
TXNRD1	thioredoxin reductase 1	X	
UGT1A10 ///			
UGT1A8 /	Na	X	X
UGT1A6	UDP glucuronosyltransferase 1 family, polypeptide A6	X	
REPPRESSED	PVALUE	N/A	P=0.062
GENE NAME	DESCRIPTION		
CCL18	chemokine (C-C motif) ligand 18		X
LEPR	leptin receptor		X
PEG3	paternally expressed 3		X
SOX9	SRY (Sex determining region Y)-box 9 (campomelic dysplasia, autosomal sex-reversal)		X
Opposite Enrichment:			
REPPRESSED in Airway but	INDUCED in Oral Mucosa	Oral/Air P= 0.005	
	P VALUE	AIR	N/A
GENE NAME	DESCRIPTION	Fold	
FCGBP	Fc fragment of IgG binding protein	X	
TMEM45A	transmembrane protein 45A	X	
FAM107A	family with sequence similarity 107, member A	X	
C3 ///			
LOC653879	Na	X	
CYP2F1	cytochrome p450, family 2, subfamily F, poypeptide 1	X	
FMO2	flavin containing monooxygenase 2 (non-functional)	X	