

## Molecular Signature of Smoking in Human Lung Tissues

Yohan Bossé<sup>1,2</sup>, Dirkje S. Postma<sup>3</sup>, Don D. Sin<sup>5,6</sup>, Maxime Lamontagne<sup>2</sup>, Christian Couture<sup>2</sup>, Nathalie Gaudreault<sup>2</sup>, Philippe Joubert<sup>2</sup>, Vivien Wong<sup>5</sup>, Mark Elliott<sup>5</sup>, Maarten van den Berge<sup>3</sup>, Corry A. Brandsma<sup>4</sup>, Catherine Tribouley<sup>7</sup>, Vladislav Malkov<sup>7</sup>, Jeffrey A. Tsou<sup>7</sup>, Gregory J. Opitck<sup>7</sup>, James C. Hogg<sup>5,8</sup>, Andrew J. Sandford<sup>5,6</sup>, Wim Timens<sup>4</sup>, Peter D. Paré<sup>5,6</sup>, and Michel Laviolette<sup>2</sup>

## Abstract

Cigarette smoking is the leading risk factor for lung cancer. To identify genes deregulated by smoking and to distinguish gene expression changes that are reversible and persistent following smoking cessation, we carried out genome-wide gene expression profiling on nontumor lung tissue from 853 patients with lung cancer. Gene expression levels were compared between never and current smokers, and time-dependent changes in gene expression were studied in former smokers. A total of 3,223 transcripts were differentially expressed between smoking groups in the discovery set ( $n = 344$ ,  $P < 1.29 \times 10^{-6}$ ). A substantial number of smoking-induced genes also were validated in two replication sets ( $n = 285$  and  $224$ ), and a gene expression signature of 599 transcripts consistently segregated never from current smokers across all three sets. The expression of the majority of these genes reverted to never-smoker levels following smoking cessation, although the time course of normalization differed widely among transcripts. Moreover, some genes showed very slow or no reversibility in expression, including SERPIND1, which was found to be the most consistent gene permanently altered by smoking in the three sets. Our findings therefore indicate that smoking deregulates many genes, many of which reverse to normal following smoking cessation. However, a subset of genes remains altered even decades following smoking cessation and may account, at least in part, for the residual risk of lung cancer among former smokers. *Cancer Res*; 72(15); 3753–63. ©2012 AACR.

## Introduction

Globally, more than one billion people smoke tobacco products (1). Tobacco smoke contains about 5,000 compounds including many carcinogens (2). Smoking increases the risk of lung cancer 20-fold (3). Smoking cessation reduces the relative risk of lung cancer, but individuals who quit smoking still have an increased risk even decades following smoking cessation compared with never smokers (4, 5).

The molecular effects of smoking and smoking cessation in the human lung are still not well understood. Previous

transcriptomic studies indicate that a large number of genes are modulated by smoking in airways (6–14), alveolar macrophages (15–17), lung tumors (7, 18–21), and peripheral leukocytes (22–27). Fewer transcriptomic studies have also examined the effect of the duration of smoking cessation on gene expression profiles (6, 8, 10, 12). All these studies were carried out in epithelial cells obtained from bronchial brushings with sample sizes that range from 24 to 104 individuals. These studies showed that the gene expression pattern is similar between former and never smokers, suggesting that most smoking-induced gene expression reverse following smoking cessation. However, the same studies also showed that the levels of expression of a small group of genes are permanently altered by cigarette smoke, but little overlap was found among studies. In addition, it is unknown whether these findings in human airway epithelial cells can be extrapolated to human lung parenchyma. Only a few transcriptomic studies have evaluated the impact of tobacco smoking on gene expression in histologically normal pulmonary parenchyma (18, 20, 21). The largest among the later studies was carried out with 15 never, 18 former, and 16 current smokers (18).

Here we reported the results of a large-scale whole-genome gene expression study evaluating the impact of smoking and smoking cessation in nontumor lung tissue. We hypothesized that the gene expression profile of the human lung would be altered by cigarette smoking and that most of these changes would revert to never-smoking levels following smoking cessation. We also hypothesized that the expression of a smaller

**Authors' Affiliations:** <sup>1</sup>Department of Molecular Medicine, Laval University; <sup>2</sup>Institut universitaire de cardiologie et de pneumologie de Québec, Québec, Canada; Departments of <sup>3</sup>Pulmonology and <sup>4</sup>Pathology and Medical Biology, University Medical Center Groningen, University of Groningen, GRIAC research institute, Groningen, The Netherlands; <sup>5</sup>UBC James Hogg Research Center, Providence Heart + Lung Institute, St. Paul's Hospital; <sup>6</sup>Respiratory Division, Department of Medicine, The University of British Columbia, Vancouver, British Columbia, Canada; <sup>7</sup>Merck & Co. Inc., Rahway, New Jersey; and <sup>8</sup>Department of Pathology and Laboratory Medicine, The University of British Columbia, Vancouver, British Columbia, Canada

**Note:** Supplementary data for this article are available at Cancer Research Online (<http://cancerres.aacrjournals.org/>).

**Corresponding Author:** Yohan Bossé, Laval University, Institut universitaire de cardiologie et de pneumologie de Québec, Pavillon Marguerite-d'Youville, Y4190, 2725, chemin Sainte-Foy, Québec (Québec), Canada, G1V 4G5. Phone: 418-656-8711 (ext. 3725); Fax: 418-656-4602; E-mail: yohan.bosse@crhl.ulaval.ca

doi: 10.1158/0008-5472.CAN-12-1160

©2012 American Association for Cancer Research.

number of genes would be permanently altered by smoking. To examine these hypotheses, we collected nontumor lung samples from a large group of patients who underwent lung surgery. The lung transcriptomes of never and current smokers were then compared. Genes differentially expressed among smoking groups were then validated in 2 independent replication sets of lung specimens to obtain a list of genes consistently altered by smoking. The genes reproducibly induced by smoking were further studied in relationship to years of smoking cessation in former smokers to identify early, late, and never-reversible genes.

## Materials and Methods

### Discovery set

Lung tissue was obtained from patients undergoing lung cancer surgery between April 2004 and December 2008. Specimens for the discovery set were taken from the Institut universitaire de cardiologie et de pneumologie de Québec (IUCPQ) site of the Respiratory Health Network Tissue Bank of the Fonds de la recherche en santé du Québec. Lung tissue samples were obtained in accordance with Institutional Review Board guidelines. All patients provided written informed consent, and the study was approved by the ethics committee of the IUCPQ.

### Smoking status

Smoking history included self-reported current smoking status, number of pack-years, and year of smoking cessation (for former smokers). Plasma cotinine quantification was determined by high-performance liquid chromatography–tandem mass spectrometry (ACQUITY UPLC System and the Quattro Premier XE; Waters). Current smokers were defined as subjects self-reported as smokers who also had a plasma cotinine concentration greater than or equal to 15 ng/mL (28). Never and former smokers were defined as subjects self-reported as lifelong nonsmokers and ex-smokers, respectively, whose plasma cotinine concentrations were below detection (<0.4 ng/mL). Never and former smokers with cotinine levels more than 0.4 ng/mL were excluded from analysis.

### Tissue processing

After surgical removal, lung specimens were immediately examined by a pulmonary pathologist (C.C.). After processing for pathologic diagnosis and staging, a nonneoplastic pulmonary parenchyma sample (2–5 cm<sup>3</sup>) was harvested from a site as far distant as possible from the tumor. The research specimens were immediately divided into smaller fragments (~0.5 cm<sup>3</sup>) placed in 5-mL cryovials and snap-frozen in liquid nitrogen. The cryovials were then transported in dry ice to the IUCPQ BioBank where they were stored at –80°C until further processing. The time from surgical removal to storage was between 15 to 30 minutes.

### Replication sets

Lung specimens collected at 2 other sites (University of British Columbia, Vancouver, Canada and University of Groningen, Groningen, The Netherlands) were used as replication

sets. At the UBC site, the study was approved by the ethics committee of the UBC Providence Health Care Research Institute Ethics Board. At the Groningen site, the study protocol was consistent with the Research Code of the University Medical Center Groningen and Dutch national ethical and professional guidelines ("Code of conduct; Dutch federation of biomedical scientific societies"; <http://www.federa.org>). To obtain replication sets similar to the discovery set, only subjects with lung neoplasm were considered. The final analyses were conducted with 285 and 224 lung specimens from UBC and Groningen, respectively.

### Assays

Whole-genome gene expression profiling was carried out using Affymetrix arrays at the same facility using the same methods for the discovery and replication sets. Microarray platform and preprocessing, quality controls, RNA extraction, quantitative real-time PCR, and immunohistochemistry are described in the Supplementary Data. Gene expression data are available through GSE23546.

### Statistical analyses

All analyses were carried out with the R statistical software version 2.9.0 and Bioconductor packages (29).

**Gene filtering.** The MAS5 call from the R affy package was used to remove genes that could not be adequately detected in the human lung tissues. Only probe sets called "present" in at least 20% of the samples were kept for subsequent analyses. A total 38,820 probe sets were available following the application of this filter.

**Selection of covariates.** The impact of standard demographic variables and possible confounders, such as COPD and types of cancer, on gene expression in the lung were tested independently to select covariates. Smoking had a much bigger impact on gene expression than COPD and lung cancer considered independently (see Supplementary Fig. S1), which is consistent with the idea that smoking induces molecular damages that in turn lead to lung diseases. Several clinical variables had an effect size similar to COPD or types of lung cancer on gene expression, and adjusting for all these variables would have made our model unstable. Accordingly, expression traits were only adjusted for age, sex, and height. Please note that we further rely on replication sets to identify genes reproducibly modulated by smoking.

**Expression trait processing.** The distribution of expression traits was largely non-Gaussian. Accordingly, we used robust residuals and nonparametric tests to conduct the association analyses. The *rlm* function in the R statistical package MASS was used to calculate residuals (M-estimation with Tukey bisquare weights). Residual values deviating from the median by more than 3 SDs were filtered out as outliers.

**Association tests.** Wilcoxon tests were used to compare adjusted expression traits between never and current smokers. We applied a Bonferroni correction to correct for multiple testing (0.05/38,820 probe sets,  $P$  value <  $1.29 \times 10^{-6}$ ). The fold changes were obtained by raising 2 to the power of the mean difference in expression between never and current smokers. Transcripts differentially expressed were those that passed the

**Table 1.** Clinical characteristic of patients in the discovery set that passed microarray quality control filters grouped by smoking status

	All subjects (n = 344)	Never smokers (n = 43)	Former smokers (n = 211)	Current smokers (n = 90)
Gender (male:female)	195:149 (56.7% male)	13:30 (30.2% male)	136:75 (64.5% male)	46:44 (51.1% male)
Age (y)	63.9 ± 10.1 (0)	56.6 ± 11.8 (0)	65.9 ± 9.1 (0)	62.5 ± 9.5 (0)
BMI (kg/m <sup>2</sup> )	27.2 ± 5.3 (0)	27.9 ± 6.4 (0)	27.7 ± 5.3 (0)	25.9 ± 4.6 (0)
FEV <sub>1</sub> % predicted	81.4 ± 18.9 (13)	91.3 ± 15.4 (4)	82.3 ± 19.9 (7)	75.1 ± 15.5 (2)
FVC% predicted	90.4 ± 16.0 (29)	95.5 ± 15.2 (6)	90.6 ± 16.8 (17)	87.7 ± 14.1 (6)
Cardiac diseases	107 (31.1%; 0)	9 (20.9%; 0)	72 (34.1%; 0)	26 (28.9%; 0)
Diabetes	43 (12.5%; 0)	2 (4.7%; 0)	30 (14.2%; 0)	11 (12.2%; 0)
COPD	164 (52.6%; 32)	7 (19.4%; 7)	97 (50.5%; 19)	60 (71.4%; 6)
Asthma	13 (3.8%; 0)	1 (2.3%; 0)	10 (4.7%; 0)	2 (2.2%; 0)
Primary diagnostic (n)				
Adenocarcinoma	191 (55.5%)	25 (58.1%)	115 (54.5%)	51 (56.7%)
Squamous cell carcinoma	95 (27.6%)	2 (4.7%)	63 (29.9%)	30 (33.3%)
NSCLC other	11 (3.2%)	1 (2.3%)	7 (3.3%)	3 (3.3%)
Carcinoid	17 (4.9%)	7 (16.3%)	9 (4.3%)	1 (1.1%)
Large cell carcinoma	12 (3.5%)	0 (0%)	9 (4.3%)	3 (3.3%)
Small cell lung carcinoma	4 (1.2%)	0 (0%)	2 (0.9%)	2 (2.2%)
Others	14 (4.1%)	8 (18.6%)	6 (2.8%)	0 (0%)

NOTE: Continuous variables are mean ± SD. The numbers of missing values are shown in parentheses.

Abbreviations: BMI, body mass index; FEV<sub>1</sub>, forced expiratory value in one second; FVC, forced vital capacity.

Bonferroni correction and had a fold change greater than or equal to 1.2.

**Replication sets.** Tests of association between gene expression and smoking status were carried out as described above. However association tests were limited to the 3,223 probe sets that were significantly altered by smoking in the discovery set. To adjust for multiple testing, we used both the Benjamini–Hochberg procedure to calculate the false discovery rate (FDR; ref. 30) and the Bonferroni correction (0.05/3,223 probe sets,  $P < 1.55 \times 10^{-5}$ ).

### Smoking cessation and normalization of gene expression

We estimated the time it takes for gene expression to revert to never-smoker levels after smoking cessation. These reversibility analyses were only carried out on probe sets significantly altered by smoking in the 3 data sets. To do this, we first determined the mean expression values for each probe set for both never and current smokers. Next, we determined the mean expression values for former smokers in relation to the elapsed time between smoking cessation and the time of surgery. The elapsed time intervals were divided such that there were at least 25 individuals per bin. The time intervals in the discovery set were (i) within 2 years before surgery, (ii) 2–5 years, (iii) 6–9 years, (iv) 10–14 years, (v) 15–19 years, (vi) 20–24 years, and (vii) 25–49 years, resulting in bin sizes of 26, 29, 32, 26, 31, 28, and 37 subjects, respectively. Probe sets were considered to have reverted to the "normal" value if, within a given time interval, the Wilcoxon test comparing former smokers to never smokers was no longer significant ( $P > 0.01$ ). These time course analyses generated 8 clusters of up- (U1–U8)

and 8 clusters of downregulated (D1–D8) probe sets based on the duration of smoking cessation to return to normal. U8 and D8 represent probe sets that do not return to normal.

Similar to the discovery set, the time intervals for the replication sets were divided such that there were at least 25 individuals per bin. The final time intervals strata differ across populations owing to differences in sample sizes and the distribution in the number of years of smoking cessation. The time intervals for the UBC set were (i) 0–2 years, (ii) 3–9 years, (iii) 10–19 years, and (iv) 20–50 years, resulting in bin sizes of 36, 38, 30, and 25 subjects, respectively. Five clusters of up- (U1–U5) and 5 clusters of downregulated (D1–D5) probes sets were formed for the UBC analyses. The time intervals for the Groningen set were (i) 0–2 years, (ii) 3–9 years, and (iii) 10–30 years, resulting in bin sizes of 36, 71, and 49 subjects, respectively. Four clusters of up- (U1–U4) and 4 clusters of downregulated (D1–D4) probes sets were formed for the Groningen analyses.

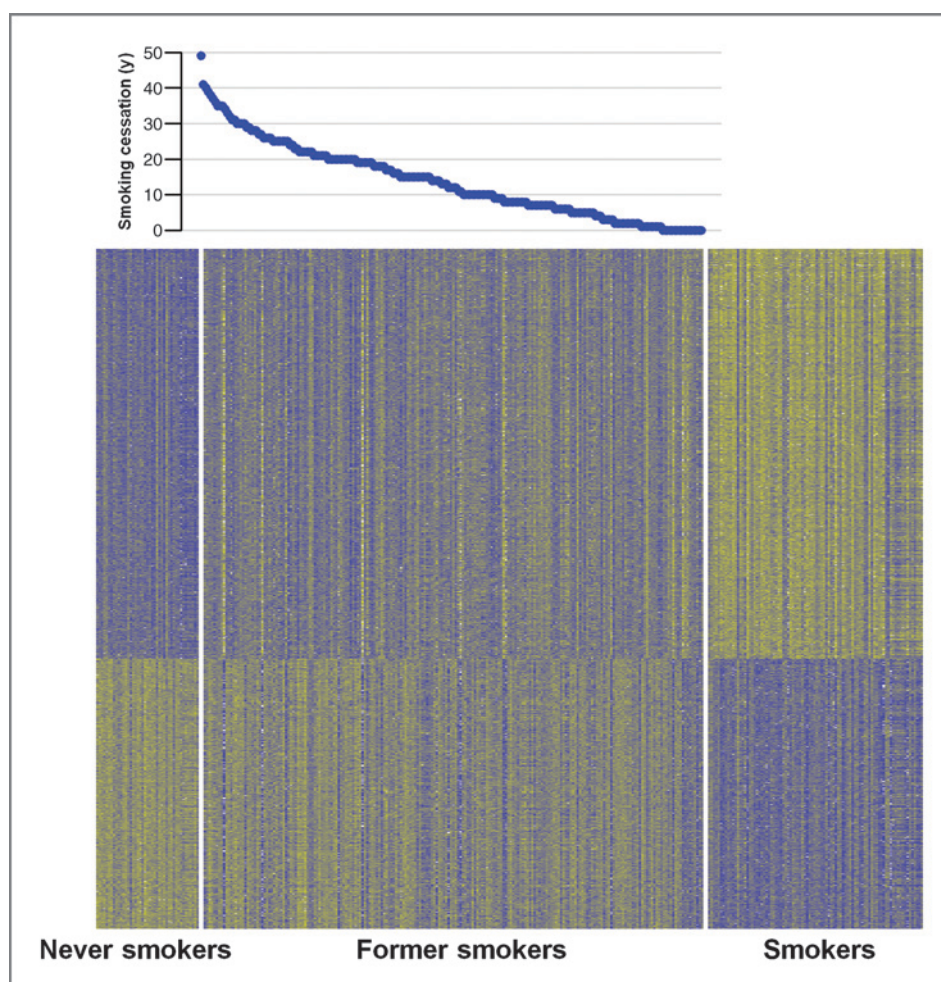
To identify slowly or never-reversible genes that were consistent across populations, we carried out  $\chi^2$  tests ( $2 \times 2$  frequency tables) comparing the distribution of probe sets returning or not returning to never-smoker levels within 10 years of smoking cessation between the discovery set and the replication sets. Probe sets not returning to never-smoker levels within 10 years of smoking cessation in the discovery set and in at least one of the 2 replication sets were identified.

## Results

### Discovery set population

The clinical characteristics of the 344 patients who passed microarray quality control and blood cotinine





**Figure 1.** The molecular signature of smoking in the discovery set. The samples ( $n = 342$ ) and probe sets ( $n = 3,223$ ) are illustrated in columns and rows, respectively. Yellow indicates a high level of expression; blue indicates a low level of expression. Never smokers are on the left and current smokers on the right. The vertical white lines separate the 3 smoking groups. The former smokers are ordered based on the number of years since cessation of smoking (top). A clear dichotomy of gene expression pattern is observed between never and current smokers.

filters in the discovery set are shown in Table 1. Most patients were former (61.3%) or current (26.2%) smokers. The vast majority of these patients underwent lung surgery for non-small-cell lung cancer (NSCLC; Table 1), predominantly adenocarcinoma (55.5%) and squamous cell carcinoma (27.6%).

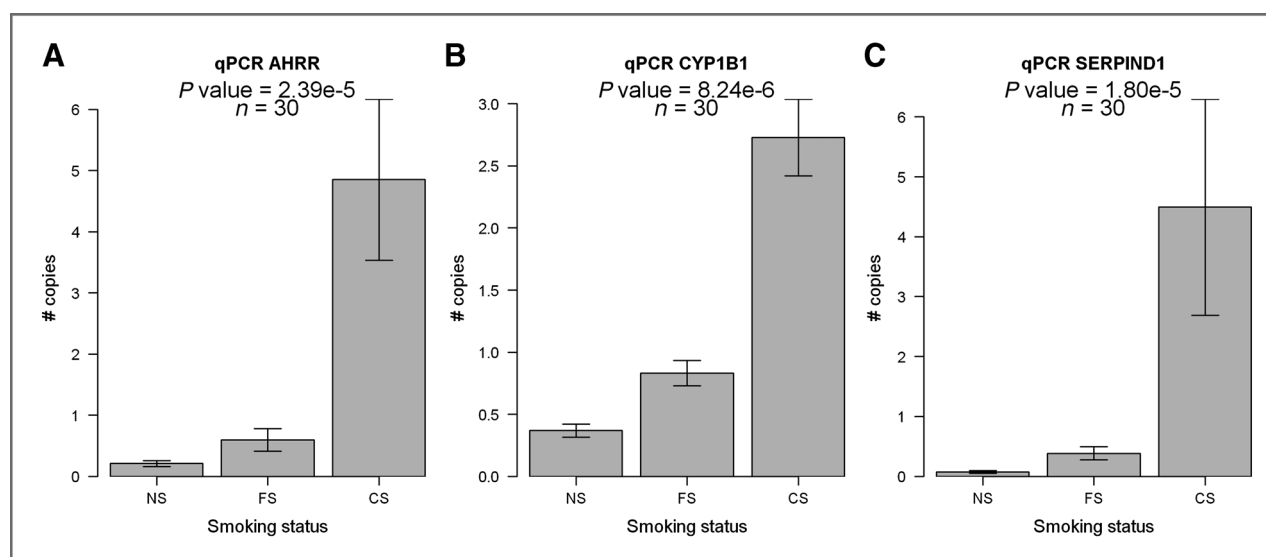
#### Transcripts associated with smoking status in the discovery set

A total of 38,820 adjusted expression traits were compared between never and current smokers. A total of 3,223 probe sets were differentially expressed between the 2 groups (i.e.,  $P < 1.29 \times 10^{-6}$  and fold change  $> 1.2$ ). These included 1,943 differentially upregulated and 1,280 downregulated probe sets in smokers (Supplementary Table S1). Figure 1 shows a heat map of these probe sets. Former smokers were ordered based on the duration of smoking cessation (Fig. 1, top panel). The date of smoking cessation was not available for 2 former-smokers. These 2 patients were excluded from the heat map. A clear separation in the pattern of gene expression was observed between never and current smokers. The most significant probe set interrogated the aryl hydrocarbon receptor repressor (*AHRR*) gene, which had a  $P$  value of  $3.3 \times 10^{-20}$  and a fold

change of 6.1. The next 3 most significant probe sets were all testing the *CYP1B1* gene ( $P$  values  $< 1.1 \times 10^{-19}$  and fold changes  $> 4.0$ ). *SERPIND1* was also among the top upregulated genes in smokers with a  $P$  value of  $5.2 \times 10^{-17}$  and fold change of 13.3. The changes in expression of *AHRR*, *SERPIND1*, and *CYP1B1* were confirmed by quantitative real-time PCR in a subset of samples (Fig. 2).

#### Replication sets

The UBC set included 30 never smokers, 158 former smokers, and 97 current smokers. The clinical characteristics of these subjects by smoking group are provided in Supplementary Table S2. Of the 3,223 probe sets found to be significant in the discovery set, 1,696 probe sets were significantly associated with smoking following the Benjamini-Hochberg correction. Considering the large number of probe sets, we also applied a more stringent multiple testing correction factor to focus on the most strongly replicated genes. Following Bonferroni correction ( $P < 1.55 \times 10^{-5}$ ), 144 probe sets were significantly associated with smoking in the UBC data set. There was 100% concordance in the orientation of the altered gene expression for the 1,696 probe sets between the 2 cohorts (i.e., genes upregulated



**Figure 2.** Confirmation of AHRR, CYP1B1, and SERPIND1 by quantitative real-time PCR. A, B, and C show the qPCR results for AHRR, CYP1B1, and SERPIND1, respectively. Results are derived from 10 never smokers, 10 former smokers, and 10 current smokers. *P* values are derived from Kruskal-Wallis tests comparing the number of copies in the 3 smoking groups, relative to *GAPDH*. Bars are means  $\pm$  SEs. NS, never smokers; FS, former smokers; CS, current smokers.

by smoking in the discovery samples were also upregulated in the UBC samples and similarly for genes that were downregulated).

The Groningen set included 16 never smokers, 164 former smokers, and 44 current smokers. The clinical characteristics of these subjects by smoking group are provided in Supplementary Table S3. Association tests were carried out between smoking status and 3,223 probe sets. A total of 910 and 30 probe sets were significantly associated with smoking status following Benjamini-Hochberg and Bonferroni ( $P < 1.55 \times 10^{-5}$ ) corrections, respectively. The orientation of the effect for the 30 probe sets was 100% concordant between the discovery and Groningen samples. For the Benjamini-Hochberg threshold 906 of 910 probe sets were concordant. The *P* values for the 3,223 probe sets in both replication samples are provided in Supplementary Table S1.

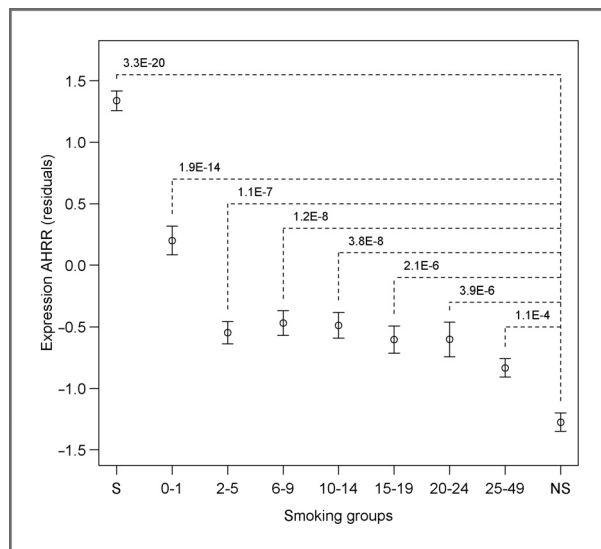
#### Genes associated with smoking across the three populations

Using stringent Bonferroni correction, 7 probe sets overlapped among significant probe sets in the discovery ( $n = 3,223$ ), UBC ( $n = 144$ ), and Groningen ( $n = 30$ ) populations. Table 2 shows these probe sets with annotation and association test results for the 3 populations. These 7 highly reproducible probe sets were all upregulated by smoking. Using a less stringent correction for multiple testing (i.e., Benjamini-Hochberg), 599 probe sets overlapped among significant probe sets in the discovery ( $n = 3,223$ ), UBC ( $n = 1,696$ ), and Groningen ( $n = 910$ ) populations. These included 558 and 41 up- and downregulated probe sets, respectively. The direction of expression was 100% concordant across the 3 populations for these probe sets. Supplementary Table S1 shows these reproducible probe sets.

**Table 2.** Highly replicated probe sets upregulated by smoking

Probe set ID	Gene symbol	Gene name	<i>P</i> Laval	<i>P</i> UBC	<i>P</i> GRNG
100123705_TGI_at	<i>PI4K2A</i>	Phosphatidylinositol 4-kinase type 2 $\alpha$	3.18E-12	1.51E-05	7.19E-06
100128259_TGI_at	<i>UBASH3B</i>	Ubiquitin associated and SH3 domain containing, B	1.62E-12	9.17E-06	2.53E-06
100130524_TGI_at <sup>a</sup>	NA	NA	3.32E-16	1.62E-12	3.44E-06
100133335_TGI_at	<i>SERPIND1</i>	Serpin peptidase inhibitor, clade D (heparin cofactor), member 1	5.20E-17	9.60E-10	1.82E-06
100136024_TGI_at	<i>DNASE2B</i>	Deoxyribonuclease II $\beta$	2.56E-15	6.92E-07	3.83E-06
100149468_TGI_at	<i>FUCA1</i>	Fucosidase, $\alpha$ -L-1, tissue	1.01E-15	5.74E-09	7.99E-06
100155403_TGI_at	<i>AHRR</i>	Aryl hydrocarbon receptor repressor	3.28E-20	5.03E-12	1.68E-07

<sup>a</sup>Human mRNAs from GenBank AK056826.



**Figure 3.** Expression of the *AHRR* gene in smoking groups. The y-axis represents mean residual gene expression values. The average and SE of gene expression values for current smokers (S,  $n = 90$ ) and never smokers (NS,  $n = 43$ ) are shown at the 2 extremes of the x-axis. Former smokers were grouped by years of smoking cessation in a way that ensured at least 25 individuals per bin ( $n$  for each bin = 26, 29, 32, 26, 31, 28, and 37).  $P$  values are Wilcoxon tests comparing current smokers and each group of former smokers to never smokers.

### Time course analyses

Figure 3 shows the time course analyses for the most significant probe set associated with smoking in the discovery set (i.e., *AHRR*). The expression of the *AHRR* gene falls substantially following smoking cessation, but remained significantly higher in former smokers who had quit smoking for more than 25 years compared with never smokers ( $P = 0.0001$ ). The same analyses were conducted for all of the reproducible probe sets ( $n = 599$ ) to identify genes that are slowly or never-reversible following smoking cessation. Among upregulated probe sets ( $n = 558$ ), 135 returned to never-smoker levels within 2 years of smoking cessation (cluster U1), 202 probe sets returned to normal after 2 to 5 years (U2), 29 after 6 to 9 years (U3), 57 after 10 to 14 years (U4), 44 after 15 to 19 years (U5), 59 after 20 to 24 years (U6), and 12 after 25 years (U7). Finally the level of expression of 20 probe sets (U8) remained significantly different ( $P > 0.01$ ) between former smokers who quit smoking for more than 25 years and never smokers. Among these 20 never reversed probe sets, the 3 that interrogate the *CYP1B1* gene were the most significant. For downregulated probe sets ( $n = 41$ ), 11 returned to never-smoker levels within 2 years of smoking cessation (D1), 14 after 2 to 5 years (D2), 6 after 6 to 9 years (D3), 0 after 10 to 14 years (D4), 3 after 15 to 19 years (D5), 6 after 20 to 24 years (D6), and 0 after 25 years (D7). One probe set downregulated by smoking did not returned to never-smoker levels (D8). For each cluster, the list of probe sets with gene annotation is provided in Supplementary Table S1.

The time course analyses of gene expression were also carried out in the 2 replication sets for the probe sets reproduced across the 3 populations. It should be noted that because

of a lesser number of subjects, the number of clusters for the duration of smoking cessation differ across the 3 populations. Accordingly, a head-to-head comparison with the discovery set is difficult for these analyses. Nonetheless, fast- and slow-responding genes were significantly concordant between the discovery and UBC sets. This was shown by  $\chi^2$  analysis for probe sets dichotomized into early and late reversibility clusters in the 2 patient groups ( $\chi^2 = 14.4$ ,  $P$  value = 0.0001, Supplementary Fig. S2). Interestingly, we observed a faster rate of gene expression recovery following smoking cessation in the Groningen data set. Out of 558 probe sets upregulated by smoking and replicated across the 3 data sets, only 7 probe sets including 6 known genes were still significantly higher than never smokers after 10 years of smoking cessation (*SERPIND1*, *AHRR*, *FASN*, *PI4K2A*, *ACSL5*, and *GANC*). Five of these 7 probe sets did not return to normal within 10 years of smoking cessation in the discovery set ( $\chi^2 = 4.3$ ,  $P$  value = 0.038, Supplementary Fig. S3). The list of slowly reversible probe sets found in the discovery set and validated in at least one of the replication cohort can be found in Table 3. *SERPIND1* was the only known gene overlapping among slowly or never-reversible genes in the 3 populations (Table 3 and Fig. 4). The upregulation of *SERPIND1* in smokers compared with never smokers was confirmed at the protein level by immunohistochemistry (Supplementary Fig. S4).

Gene set enrichment analysis (31) was used to test whether slowly reversible genes in the replication sets were enriched among the 599 reproduced genes altered by smoking in the discovery set (see Supplementary Data for more details). The 599 reproduced probe sets were preranked by their degree of reversibility in the discovery set. Slowly reversible genes in the 2 replication sets were then tested for enrichment against this preranked list. Supplementary Figs. S5 and S6 show the enrichment plots for both gene sets (i.e., UBC and Groningen). Genes slowly reversible in UBC ( $n = 49$  probe sets) cluster among the top slowly reversible genes in the discovery set (FDR  $Q$  value = 0.002, Supplementary Fig. S5). Although only 6 known genes were considered slowly reversible in the Groningen set, they also tend to cluster among slowly reversible genes in the discovery set (FDR  $Q$  value = 0.053, Supplementary Fig. S6).

### Discussion

This study provides a comprehensive description of the molecular signature of smoking in the human lung and the effects of smoking cessation on this signature. Thousands of genes were significantly associated with smoking status in the discovery set and the gene expression patterns of never and current smokers were clearly distinguishable. The impact of smoking was validated for 599 probe sets in 2 replication data sets. Although most of the differentially expressed genes returned to levels similar to never smokers following smoking cessation, many slowly reversible genes were identified and some never normalized despite prolonged smoking cessation. Some of the genes that showed enduring alterations in expression have never been reported to be involved in lung cancer, whereas others are targets of

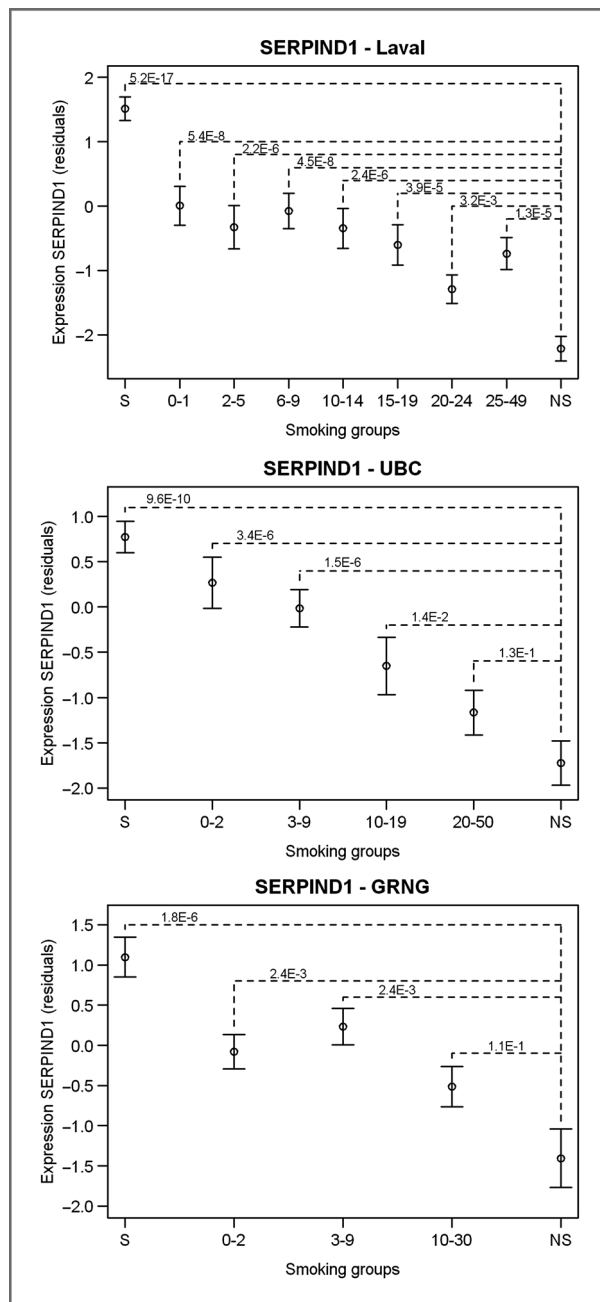
**Table 3.** Replicated probe sets upregulated by smoking and not returning to never-smoker levels within 10 years of smoking cessation in the discovery set and validated in at least one of the 2 replication sets

Probe set ID	Gene symbol	Gene name	p Laval <sup>a</sup>	p UBC <sup>a</sup>	p GRNG <sup>a</sup>
100123700_TGI_at <sup>b</sup>	<i>ACSL5</i>	Acyl-CoA synthetase long-chain family member 5	6.66E-13	2.05E-02	1.45E-06
100155403_TGI_at <sup>b</sup>	<i>AHRR</i>	Aryl hydrocarbon receptor repressor	3.28E-20	5.03E-12	1.68E-07
100159958_TGI_at	<i>AMICA1</i>	Adhesion molecule, interacts with CXADR antigen 1	5.31E-15	7.36E-07	6.07E-03
100312275_TGI_at	<i>AMICA1</i>	Adhesion molecule, interacts with CXADR antigen 1	9.12E-15	1.01E-06	8.24E-03
100154796_TGI_at	<i>AMICA1</i>	Adhesion molecule, interacts with CXADR antigen 1	4.54E-15	2.02E-06	6.22E-03
100307462_TGI_at	<i>ATP13A4</i>	ATPase type 13A4	3.18E-17	2.67E-07	3.58E-05
100150912_TGI_at	<i>C2orf58</i>	Chromosome 2 open reading frame 58	3.48E-17	3.63E-10	1.89E-05
100153378_TGI_at	<i>CD163L1</i>	CD163 molecule-like 1	2.23E-14	2.72E-06	6.07E-03
100130932_TGI_at	<i>CD86</i>	CD86 molecule	6.29E-14	1.10E-04	8.71E-03
100132353_TGI_at	<i>CLIP4</i>	CAP-GLY domain containing linker protein family, member 4	7.06E-16	4.37E-06	3.21E-04
100125484_TGI_at	<i>CYP1B1</i>	Cytochrome P450, family 1, subfamily B, polypeptide 1	1.14E-19	3.33E-10	2.81E-03
100131143_TGI_at	<i>CYP1B1</i>	Cytochrome P450, family 1, subfamily B, polypeptide 1	6.12E-20	3.30E-11	3.37E-03
100303658_TGI_at	<i>CYP1B1</i>	Cytochrome P450, family 1, subfamily B, polypeptide 1	6.69E-20	4.16E-10	1.70E-03
100134409_TGI_at	<i>DNAJC5B</i>	DnaJ (Hsp40) homolog, subfamily C, member 5 β	2.32E-14	2.17E-06	4.67E-05
100309944_TGI_at <sup>b</sup>	<i>FASN</i>	Fatty acid synthase	4.29E-11	5.18E-03	2.21E-05
100136438_TGI_at	<i>GPR110</i>	G-protein-coupled receptor 110	3.62E-18	1.07E-06	1.37E-03
100302763_TGI_at	<i>HABP2</i>	Hyaluronan-binding protein 2	1.13E-15	3.53E-08	1.43E-03
100305401_TGI_at	<i>ITPR2</i>	Inositol 1,4,5-triphosphate receptor, type 2	6.79E-15	6.37E-05	5.32E-04
100122391_TGI_at	<i>MT1F</i>	Metallothionein 1F	1.35E-09	1.98E-05	5.14E-03
100162545_TGI_at	<i>PLA2G4E</i>	Phospholipase A2, group IVE	1.28E-17	1.80E-11	2.13E-05
100121991_TGI_at	<i>PLAUR</i>	Plasminogen activator, urokinase receptor	1.13E-10	2.96E-04	1.20E-02
100303500_TGI_at	<i>PLAUR</i>	Plasminogen activator, urokinase receptor	1.76E-11	1.99E-04	4.34E-03
100138457_TGI_at	<i>PLAUR</i>	Plasminogen activator, urokinase receptor	5.52E-15	1.80E-05	1.38E-04
100300362_TGI_at	<i>PLAUR</i>	Plasminogen activator, urokinase receptor	1.53E-14	9.17E-06	1.61E-04
100143476_TGI_at	<i>PON1</i>	Paraoxonase 1	4.09E-08	1.16E-05	1.37E-03
<b>100133335_TGI_at</b>	<i>SERPIND1</i>	Serpin peptidase inhibitor, clade D (heparin cofactor), member 1	5.20E-17	9.60E-10	1.82E-06
100144500_TGI_at	<i>SFN</i>	Stratifin	6.77E-10	4.41E-04	2.41E-03
100160780_TGI_at	<i>SPINK5</i>	Serine peptidase inhibitor, Kazal type 5	3.81E-15	9.92E-06	3.21E-04
100139182_TGI_at <sup>c</sup>			2.34E-11	1.22E-05	3.49E-05
100156798_TGI_at <sup>d</sup>		Similar to phosphodiesterase 4D, cAMP specific	2.12E-12	2.09E-08	1.20E-03
<b>100130524_TGI_at<sup>e</sup></b>			3.32E-16	1.62E-12	3.44E-06

NOTE: Probe sets in bold are those not returning to normal within 10 years of smoking cessation in the 3 populations.

<sup>a</sup>P values for the discovery set (Laval) and the 2 replication sets (UBC and Groningen). P values are derived from a Wilcoxon test comparing current smokers to never smokers.<sup>b</sup>Probe sets not returning to normal within 10 years of smoking cessation in the discovery and Groningen sets.<sup>c</sup>Human mRNAs from GenBank BQ225423.<sup>d</sup>Human mRNAs from GenBank BX648604.<sup>e</sup>Human mRNAs from GenBank AK056826.





**Figure 4.** Expression of *SERPIND1* in the 3 populations by smoking groups and duration of smoking cessation among former smokers. Dots are means  $\pm$  SEs. S, current smokers; NS, never smokers. In the 3 panels, former smokers were grouped based on the duration of smoking cessation (see Materials and Methods).

current clinical trials or are used as biomarkers for the progression or prognosis of this disease.

Most whole-genome gene expression studies that have been done to investigate the impact of tobacco smoking were carried out on airways, (6–14) alveolar macrophages, (15–17) lung tumors, (7, 18–21), or peripheral leukocytes (22–27). The impact of tobacco smoking on gene expression in nontumor

lung tissue has been evaluated in only a few studies (18, 20, 21). Perhaps the best comparison with our study is the report by Landi and colleagues (18) who studied the gene expression profile of nontumor lungs of 15 never, 18 former, and 16 current smokers. They identified 25 up- and 73 downregulated genes between never and current smokers. Sixteen upregulated (*CD1A*, *CRTAM*, *CYBB*, *CYP1B1*, *DNASE2B*, *ELF5*, *FABP3*, *FGG*, *IGSF6*, *ITGAX*, *KMO*, *RRAGD*, *SERPIND1*, *SPINK5*, *TM7SF4*, and *TREM2*) and 2 downregulated (*TMSL8* and *GHR*) genes overlapped with our list of 599 replicated transcripts. When compared with the list of significant transcripts in our discovery set, 24 of 25 upregulated genes and 56 of 73 downregulated genes identified by Landi and colleagues (18) overlapped. They also confirmed 5 upregulated genes (*CEACAM5*, *CYPBB*, *CYLT1*, *FGG*, and *TM7SF4*) by real-time PCR in technical and biologic replicates. These 5 genes were significantly upregulated by smoking in our discovery set. This study provides strong external validation of our results.

The gene most strongly upregulated by smoking was the aryl hydrocarbon receptor repressor (*AHRR*), which is a negative regulator of the aryl hydrocarbon receptor (AHR) signaling pathway (32, 33). This pathway, also known as the xenobiotic or dioxin-signaling pathway, is an important regulator of cellular responses to exogenous ligands. Cigarette smoke contains agonists of the AHR-signaling pathway (34) that in turn regulate the expression of genes encoding cytochrome P450 enzymes (35). In this study, members of the CYP1 family, including *CYP1A1*, *CYP1A2*, and *CYP1B1* were among the top genes upregulated by smoking. These enzymes are essential for detoxification of environmental chemicals but are also known to generate mutagenic and toxic intermediates that are believed to be carcinogenic (34). Murine models suggest that the AHR pathway is required for carcinogen-induced cancer (36) and that the constitutive activation of this pathway may be responsible for the development of cancer (37, 38). Treatment of mice with benzopyrene, a major carcinogen found in cigarette smoke, resulted in an increase in *AHRR* expression in the lung (39). In humans, *AHRR* is ubiquitously expressed but is more abundant in some tissues including the lung (40). Recent evidence suggests that *AHRR* is a tumor suppressor gene (41). It is thus tempting to speculate that the upregulation of *AHRR* in the lung tissues related to cigarette smoking may be a defence response against possible malignant transformation.

There was large variation in the time since smoking cessation among former smokers in our data sets. This provided a unique opportunity to differentiate genes that respond favorably from those that are resistant to smoking cessation. Our data suggest that the smoking-related gene expression signature largely disappears within a few years of smoking cessation. More importantly, the time for gene expression to revert to never-smokers values is gene specific. Whereas groups of genes return to normal within the first years of smoking cessation, others only slowly revert or their expression levels are still higher than in never smoker after decades of smoking cessation. The latter genes are likely to



be important markers to aid in the understanding of the residual risk of lung cancer among former smokers and may also explain why some individuals with COPD have persistent accelerated lung function decline even after smoking cessation. Particularly relevant for lung cancer is *SERPIND1*, the most enduringly altered gene in our study. This gene encodes a protein that has been used as a biomarker for progression of NSCLC (NCT00155116, ClinicalTrials.gov). *SERPIND1* and the other slowly reversible genes may thus be extremely valuable in improving our understanding of the initial and persistent processes that eventually lead to lung diseases in smokers.

There are certain limitations to the study. First, we did not have longitudinal clinical or gene expression data on these patients. As such, we do not know the robustness of the molecular signature related to smoking within the same individual over time. Moreover, we could not adequately assess possible interactions of smoking and other time-dependent factors such as aging on gene expression. Second, the smoking-related gene expression pattern could be influenced by a different proportion of cells of different types in the lungs of smokers. There is cellular heterogeneity in lung tissue as in other organs, and this influences the gene expression pattern (42). Lung samples contain many cell types including organ-specific cells (i.e., pneumocytes), endothelial cells, and migratory inflammatory cells, and our analysis represents the amalgam of all these cells. Smoking could influence the relative proportion of cell types in the lung and because gene expression patterns are cell type specific, this could contribute to the smoking signals that we observed. For example, it is well known that the total number of alveolar macrophages is markedly increased in the lungs of smokers (43). We have also shown that inflammation persists in former smokers and COPD patients years after smoking cessation (44, 45). Accordingly our results must be interpreted with caution and follow-up experiments using specific cell types will be required. Considering the large number of future follow-up experiments that will be required, we believe that sharing our results with the scientific community is likely to accelerate the confirmation of preventive and therapeutic targets. Third, we do not know which of the genes altered by cigarette smoking and smoking cessation are critically important in the genesis of lung cancer. Future work is clearly needed to elucidate the salient pathways. In addition, gene expression is only one aspect of the molecular alterations induced by smoking in the lung (46–48). Multi-dimensional genomic profiling of lung specimens including somatic mutations and epigenetic marks will be required to understand lung carcinogenesis and the residual risk of lung cancer among former smokers. Finally, the list of slowly reversible genes following smoking cessation is derived from patients with lung cancer. It would be interesting to know whether these genes are also permanently deregulated in current and former smokers that never develop lung cancer. Identifying genes that do not normalize following smoking cessation in subjects who develop lung cancer compared with those who do not

develop lung cancer will be an important step to uncovering the drivers of disease.

In conclusion, we have shown that smoking has a major impact on the lung parenchymal transcriptome and that the altered molecular signature can return to near normal levels over time. However, there are genes affected by smoking that do not normalize or are slow to normalize even after decades of smoking cessation. It also provides molecular evidence of the devastating health impact of smoking in human lung and reinforces the importance of smoking cessation. This study is an important step to pinpoint the specific genes and pathways altered by smoking in the lung and provides new candidates that are likely involved in lung cancer and other respiratory diseases.

### Disclosure of Potential Conflicts of Interest

M. van den Berge has been a principal investigator of a study funded by Chiesi and is also a consultant/advisory board member of Novartis; J.C. Hogg has served as a principal investigator on a grant from Merck and also served as principal investigator on contract from BI. He is also a consultant/advisory board member of COPD. W. Timens has received other commercial research support from Merck, Sharp & Dohme.

### Authors' Contributions

**Conception and design:** Y. Bossé, D.S. Postma, D.D. Sin, C. Couture, C. Tribouley, G.J. Opiteck, P.D. Paré, M. Laviolette

**Development of methodology:** D.S. Postma, M. Lamontagne, C. Couture, G.J. Opiteck, M. Laviolette

**Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.):** D.S. Postma, D.D. Sin, C. Couture, P. Joubert, M. Elliott, M. van den Berge, C.A. Brandsma, C. Tribouley, J.A. Tsou, G.J. Opiteck, J.C. Hogg, W. Timens, P.D. Paré, M. Laviolette

**Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis):** Y. Bossé, D.S. Postma, D.D. Sin, M. Lamontagne, V. Wong, V. Malkov, W. Timens, M. Laviolette

**Writing, review, and/or revision of the manuscript:** Y. Bossé, D.S. Postma, D. Sin, C. Couture, M. Elliott, C. Tribouley, J.A. Tsou, G.J. Opiteck, A.J. Sandford, W. Timens, P.D. Paré, M. Laviolette

**Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases):** C. Couture, N. Gaudreault, V. Wong, M. van den Berge, C.A. Brandsma, V. Malkov, J.A. Tsou, G.J. Opiteck, J.C. Hogg, M. Laviolette

**Study supervision:** Y. Bossé, M. Laviolette

### Acknowledgments

The authors thank Christine Racine and Sabrina Biardel at the IUCPQ site of the Respiratory Health Network Tissue Bank of the FRQS for her valuable assistance and Rémy Thériault and the information technology team at the Institut universitaire de cardiologie et de pneumologie de Québec for the secure server-based data repository. They also thank the staff at the Consortium Laval, Université du Québec, McGill and Eastern Quebec (CLUMEQ) for IT support with the high-performance computer clusters.

### Grant Support

Y. Bossé is a research scholar from the Heart and Stroke Foundation of Canada. D.S. Postma is the holder of a KNAW (Royal Academy of Arts and Sciences) chair for research. D.D. Sin holds a Canadian Research Chair in COPD and is a Michael Smith Foundation Scholar. A.J. Sandford is the recipient of a Canada Research Chair in genetics and a Michael Smith Foundation for Health Research Senior Scholar Award. P.D. Paré is a Michael Smith Foundation Distinguished Scholar and the Jacob Churg Distinguished Researcher. This study was funded by Merck Research Laboratories. It was also partly supported by the Chaire de pneumologie de la Fondation JD Bégin de l'Université Laval, the Fondation de l'Institut universitaire de cardiologie et de pneumologie de Québec, the Respiratory Health Network of the FRQS, and the Cancer Research Society.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received March 26, 2012; revised May 22, 2012; accepted May 29, 2012; published OnlineFirst June 1, 2012.

## References

- Jha P. Avoidable global cancer deaths and total deaths from smoking. *Nat Rev Cancer* 2009;9:655–64.
- Hoffmann D, Hoffmann I, El-Bayoumy K. The less harmful cigarette: a controversial issue. a tribute to Ernst L. Wynder *Chem Res Toxicol* 2001;14:767–90.
- Alberg AJ, Samet JM. Epidemiology of lung cancer. *Chest* 2003;123:21S–49S.
- Halpern MT, Gillespie BW, Warner KE. Patterns of absolute risk of lung cancer mortality in former smokers. *J Natl Cancer Inst* 1993;85:457–64.
- Ebbert JO, Yang P, Vachon CM, Vierkant RA, Cerhan JR, Folsom AR, et al. Lung cancer risk reduction after smoking cessation: observations from a prospective cohort of women. *J Clin Oncol* 2003;21:921–6.
- Beane J, Sebastiani P, Liu G, Brody JS, Lenburg ME, Spira A. Reversible and permanent effects of tobacco smoke exposure on airway epithelial gene expression. *Genome Biol* 2007;8:R201.
- Boelens MC, van den Berg A, Fehrmann RS, Geerlings M, de Jong WK, te Meerman GJ, et al. Current smoking-specific gene expression signature in normal bronchial epithelium is enhanced in squamous cell lung cancer. *J Pathol* 2009;218:182–91.
- Chari R, Loneragan KM, Ng RT, MacAulay C, Lam WL, Lam S. Effect of active smoking on the human bronchial epithelium transcriptome. *BMC Genomics* 2007;8:297.
- Harvey BG, Heguy A, Leopold PL, Carolan BJ, Ferris B, Crystal RG. Modification of gene expression of the small airway epithelium in response to cigarette smoking. *J Mol Med* 2007;85:39–53.
- Spira A, Beane J, Shah V, Liu G, Schembri F, Yang X, et al. Effects of cigarette smoke on the human airway epithelial cell transcriptome. *Proc Natl Acad Sci U S A* 2004;101:10143–8.
- Strulovici-Barel Y, Omberg L, O'Mahony M, Gordon C, Hollmann C, Tilley AE, et al. Threshold of biologic responses of the small airway epithelium to low levels of tobacco smoke. *Am J Respir Crit Care Med* 2010;182:1524–32.
- Zhang L, Lee JJ, Tang H, Fan YH, Xiao L, Ren H, et al. Impact of smoking cessation on global gene expression in the bronchial epithelium of chronic smokers. *Cancer Prev Res* 2008;1:112–8.
- Zhang X, Sebastiani P, Liu G, Schembri F, Dumas YM, Langer EM, et al. Similarities and differences between smoking-related gene expression in nasal and bronchial epithelium. *Physiol Genomics* 2010;41:1–8.
- Sridhar S, Schembri F, Zeskind J, Shah V, Gustafson AM, Steiling K, et al. Smoking-induced gene expression changes in the bronchial airway are reflected in nasal and buccal epithelium. *BMC Genomics* 2008;9:259.
- Heguy A, O'Connor TP, Luetthich K, Worgall S, Cieciuch A, Harvey BG, et al. Gene expression profiling of human alveolar macrophages of phenotypically normal smokers and nonsmokers reveals a previously unrecognized subset of genes modulated by cigarette smoking. *J Mol Med* 2006;84:318–28.
- Shaykhiev R, Krause A, Salit J, Strulovici-Barel Y, Harvey BG, O'Connor TP, et al. Smoking-dependent reprogramming of alveolar macrophage polarization: implication for pathogenesis of chronic obstructive pulmonary disease. *J Immunol* 2009;183:2867–83.
- Woodruff PG, Koth LL, Yang YH, Rodriguez MW, Favoreto S, Dolganov GM, et al. A distinctive alveolar macrophage activation state induced by cigarette smoking. *Am J Respir Crit Care Med* 2005;172:1383–92.
- Landi MT, Dracheva T, Rotunno M, Figueroa JD, Liu H, Dasgupta A, et al. Gene expression signature of cigarette smoking and its role in lung adenocarcinoma development and survival. *PLoS One* 2008;3:e1651.
- Miura K, Bowman ED, Simon R, Peng AC, Robles AI, Jones RT, et al. Laser capture microdissection and microarray expression analysis of lung adenocarcinoma reveals tobacco smoking- and prognosis-related molecular profiles. *Cancer Res* 2002;62:3244–50.
- Powell CA, Spira A, Derti A, DeLisi C, Liu G, Borczuk A, et al. Gene expression in lung adenocarcinomas of smokers and nonsmokers. *Am J Respir Cell Mol Biol* 2003;29:157–62.
- Woenckhaus M, Klein-Hitpass L, Grepmeier U, Merk J, Pfeifer M, Wild P, et al. Smoking and cancer-related gene expression in bronchial epithelium and non-small-cell lung cancers. *J Pathol* 2006;210:192–204.
- Buttner P, Mosig S, Funke H. Gene expression profiles of T lymphocytes are sensitive to the influence of heavy smoking: A pilot study. *Immunogenetics* 2007;59:37–43.
- Charlesworth JC, Curran JE, Johnson MP, Goring HH, Dyer TD, Diego VP, et al. Transcriptomic epidemiology of smoking: the effect of smoking on gene expression in lymphocytes. *BMC Med Genomics* 2010;3:29.
- Lampe JW, Stepaniants SB, Mao M, Radich JP, Dai H, Linsley PS, et al. Signatures of environmental exposures using peripheral leukocyte gene expression: tobacco smoke. *Cancer Epidemiol Biomarkers Prev* 2004;13:445–53.
- Lodovici M, Luceri C, De Filippo C, Romualdi C, Bambi F, Dolara P. Smokers and passive smokers gene expression profiles: correlation with the DNA oxidation damage. *Free Radic Biol Med* 2007;43:415–22.
- Ryder MI, Hyun W, Loomer P, Haqq C. Alteration of gene expression profiles of peripheral mononuclear blood cells by tobacco smoke: implications for periodontal diseases. *Oral Microbiol Immunol* 2004;19:39–49.
- van Leeuwen DM, van Agen E, Gottschalk RW, Vlietinck R, Gielen M, van Herwijnen MH, et al. Cigarette smoke-induced differential gene expression in blood cells from monozygotic twin pairs. *Carcinogenesis* 2007;28:691–7.
- Soldin OP, Makambi KH, Soldin SJ, O'Mara DM. Steroid hormone levels associated with passive and active smoking. *Steroids* 2011;76:653–9.
- Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, et al. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* 2004;5:R80.
- Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Royal Stat Soc Ser B* 1995;57:289–300.
- Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, 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.
- Haarmann-Stemmann T, Abel J. The arylhydrocarbon receptor repressor (AhRR): structure, expression, and function. *Biol Chem* 2006;387:1195–9.
- Hahn ME, Allan LL, Sherr DH. Regulation of constitutive and inducible AHR signaling: complex interactions involving the AHR repressor. *Biochem Pharmacol* 2009;77:485–97.
- Kitamura M, Kasai A. Cigarette smoke as a trigger for the dioxin receptor-mediated signaling pathway. *Cancer Lett* 2007;252:184–94.
- Kawajiri K, Fujii-Kuriyama Y. Cytochrome P450 gene regulation and physiological functions mediated by the aryl hydrocarbon receptor. *Arch Biochem Biophys* 2007;464:207–12.
- Shimizu Y, Nakatsuru Y, Ichinose M, Takahashi Y, Kume H, Mimura J, et al. Benzo[a]pyrene carcinogenicity is lost in mice lacking the aryl hydrocarbon receptor. *Proc Natl Acad Sci U S A* 2000;97:779–82.
- Andersson P, McGuire J, Rubio C, Gradin K, Whitelaw ML, Pettersson S, et al. A constitutively active dioxin/aryl hydrocarbon receptor induces stomach tumors. *Proc Natl Acad Sci U S A* 2002;99:9990–5.
- Moennikes O, Loeppen S, Buchmann A, Andersson P, Ittrich C, Poellinger L, et al. A constitutively active dioxin/aryl hydrocarbon receptor promotes hepatocarcinogenesis in mice. *Cancer Res* 2004;64:4707–10.
- Bernshausen T, Jux B, Esser C, Abel J, Fritsche E. Tissue distribution and function of the Aryl hydrocarbon receptor repressor (AhRR) in C57BL/6 and Aryl hydrocarbon receptor deficient mice. *Arch Toxicol* 2006;80:206–11.
- Fujita H, Kosaki R, Yoshihashi H, Ogata T, Tomita M, Hasegawa T, et al. Characterization of the aryl hydrocarbon receptor repressor gene and association of its Pro185Ala polymorphism with micropenis. *Teratology* 2002;65:10–8.

41. Zudaire E, Cuesta N, Murty V, Woodson K, Adams L, Gonzalez N, et al. The aryl hydrocarbon receptor repressor is a putative tumor suppressor gene in multiple human cancers. *J Clin Invest* 2008;118:640–50.
42. Wang IM, Stepaniants S, Boie Y, Mortimer JR, Kennedy B, Elliott M, et al. Gene expression profiling in patients with chronic obstructive pulmonary disease and lung cancer. *Am J Respir Crit Care Med* 2008;177:402–11.
43. Harris JO, Swenson EW, Johnson JE 3rd. Human alveolar macrophages: comparison of phagocytic ability, glucose utilization, and ultrastructure in smokers and nonsmokers. *J Clin Invest* 1970;49:2086–96.
44. Lapperre TS, Sont JK, van Schadewijk A, Gosman MM, Postma DS, Bajema IM, et al. Smoking cessation and bronchial epithelial remodeling in COPD: a cross-sectional study. *Respir Res* 2007;8:85.
45. Willemse BW, ten Hacken NH, Rutgers B, Lesman-Leegte IG, Postma DS, Timens W. Effect of 1-year smoking cessation on airway inflammation in COPD and asymptomatic smokers. *Eur Respir J* 2005;26:835–45.
46. Breitling LP, Yang R, Korn B, Burwinkel B, Brenner H. Tobacco-smoking-related differential DNA methylation: 27K discovery and replication. *Am J Hum Genet* 2011;88:450–7.
47. Lee W, Jiang Z, Liu J, Haverly PM, Guan Y, Stinson J, et al. The mutation spectrum revealed by paired genome sequences from a lung cancer patient. *Nature* 2010;465:473–7.
48. Pleasance ED, Stephens PJ, O'Meara S, McBride DJ, Meynert A, Jones D, et al. A small-cell lung cancer genome with complex signatures of tobacco exposure. *Nature* 2010;463:184–90.

# Cancer Research

The Journal of Cancer Research (1916–1930) | The American Journal of Cancer (1931–1940)

## Molecular Signature of Smoking in Human Lung Tissues

Yohan Bossé, Dirkje S. Postma, Don D. Sin, et al.

*Cancer Res* 2012;72:3753-3763. Published OnlineFirst June 1, 2012.

<b>Updated version</b>	Access the most recent version of this article at: doi: <a href="https://doi.org/10.1158/0008-5472.CAN-12-1160">10.1158/0008-5472.CAN-12-1160</a>
<b>Supplementary Material</b>	Access the most recent supplemental material at: <a href="http://cancerres.aacrjournals.org/content/suppl/2012/06/01/0008-5472.CAN-12-1160.DC1">http://cancerres.aacrjournals.org/content/suppl/2012/06/01/0008-5472.CAN-12-1160.DC1</a>

<b>Cited articles</b>	This article cites 48 articles, 11 of which you can access for free at: <a href="http://cancerres.aacrjournals.org/content/72/15/3753.full#ref-list-1">http://cancerres.aacrjournals.org/content/72/15/3753.full#ref-list-1</a>
<b>Citing articles</b>	This article has been cited by 9 HighWire-hosted articles. Access the articles at: <a href="http://cancerres.aacrjournals.org/content/72/15/3753.full#related-urls">http://cancerres.aacrjournals.org/content/72/15/3753.full#related-urls</a>

<b>E-mail alerts</b>	<a href="#">Sign up to receive free email-alerts</a> related to this article or journal.
<b>Reprints and Subscriptions</b>	To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at <a href="mailto:pubs@aacr.org">pubs@aacr.org</a> .
<b>Permissions</b>	To request permission to re-use all or part of this article, use this link <a href="http://cancerres.aacrjournals.org/content/72/15/3753">http://cancerres.aacrjournals.org/content/72/15/3753</a> . Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.