



Transcriptomic changes in the nasal epithelium associated with diesel engine exhaust exposure



E. Drizik^{a,1}, S. Corbett^{b,1}, Y. Zheng^{e,*,1}, R. Vermeulen^{d,1}, Y. Dai^e, W. Hu^f, D. Ren^g, H. Duan^e, Y. Niu^e, J. Xu^h, W. Fu^g, K. Meliefste^d, B. Zhou^e, Xiaohui Zhang^a, J. Yang^g, Bryan Bassig^f, Hanqiao Liu^a, M. Ye^e, Gang Liu^a, X. Jia^e, T. Meng^e, P. Bin^e, J. Zhang^{i,j}, D. Silverman^{f,2}, A. Spira^{a,b,c,2}, N. Rothman^{f,2}, M.E. Lenburg^{a,b,*,2}, Q. Lan^{f,2}

^a Section of Computational Biomedicine, Department of Medicine, Boston University School of Medicine, Boston, MA, USA

^b Bioinformatics Program, Boston University, Boston, MA, USA

^c The Lung Cancer Initiative at Johnson & Johnson, Cambridge, MA, USA

^d Institute for Risk Assessment Sciences, Utrecht University, Utrecht, the Netherlands

^e Key Laboratory of Chemical Safety and Health, National Institute of Occupational, Health and Poison Control, Chinese Center for Disease Control and Prevention, Beijing, China

^f Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, MD, USA

^g Chaoyang Center for Disease Control and Prevention, Chaoyang, China

^h Hong Kong University, Hong Kong, China

ⁱ Nicholas School of the Environment and Duke Global Health Institute, Duke University, Durham, NC, USA

^j Global Health Research Center, Duke Kunshan University, Kunshan City, Jiangsu Province, China

ARTICLE INFO

Keywords:

Diesel engine exhaust
Microarray
Lung cancer
Occupational exposure

ABSTRACT

Background: Diesel engine exhaust (DEE) exposure causes lung cancer, but the molecular mechanisms by which this occurs are not well understood.

Objectives: To assess transcriptomic alterations in nasal epithelium of DEE-exposed factory workers to better understand the cellular and molecular effects of DEE.

Methods: Nasal epithelial brushings were obtained from 41 diesel engine factory workers exposed to relatively high levels of DEE (17.2–105.4 $\mu\text{g}/\text{m}^3$), and 38 unexposed workers from factories without DEE exposure. mRNA was profiled for gene expression using Affymetrix microarrays. Linear modeling was used to identify differentially expressed genes associated with DEE exposure and interaction effects with current smoking status. Pathway enrichment among differentially expressed genes was assessed using EnrichR. Gene Set Enrichment Analysis (GSEA) was used to compare gene expression patterns between datasets.

Results: 225 genes had expression associated with DEE exposure after adjusting for smoking status ($\text{FDR } q < 0.25$) and were enriched for genes in pathways related to oxidative stress response, cell cycle pathways such as MAPK/ERK, protein modification, and transmembrane transport. Genes up-regulated in DEE-exposed individuals were enriched among the genes most up-regulated by cigarette smoking in a previously reported bronchial airway smoking dataset. We also found that the DEE signature was enriched among the genes most altered in two previous studies of the effects of acute DEE on PBMC gene expression. An exposure-response relationship was demonstrated between air levels of elemental carbon and the first principal component of the DEE signature.

Conclusions: A gene expression signature was identified for workers occupationally exposed to DEE that was altered in an exposure-dependent manner and had some overlap with the effects of smoking and the effects of acute DEE exposure. This is the first study of gene expression in nasal epithelial cells of workers heavily exposed to DEE and provides new insights into the molecular alterations that occur with DEE exposure.

* Corresponding authors at: Key Laboratory of Chemical Safety and Health, National Institute of Occupational Health and Poison Control, Chinese Center for Disease Control and Prevention, Beijing, China (Y. Zheng). Section of Computational Biomedicine, Department of Medicine, Boston University School of Medicine, Boston, MA, USA (M.E. Lenburg).

E-mail addresses: yxzheng@qdu.edu.cn (Y. Zheng), mleburg@bu.edu (M.E. Lenburg).

¹ Co-first authors.

² Co-senior authors.

<https://doi.org/10.1016/j.envint.2020.105506>

Received 22 August 2019; Received in revised form 19 December 2019; Accepted 17 January 2020

0160-4120/ © 2020 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The use of diesel engines is widespread due to their robustness, efficiency, and low operating costs, which explains their prevalence in industrial settings such as mining, commercial transport, and construction as well as in the general environment. Because of their ubiquity, but also due to their tendency to produce aerosols, diesel engines contribute to 65–90% of vehicular secondary organic aerosol in urban areas worldwide (Gentner et al., 2012; Resitoglu et al., 2015).

Diesel engine exhaust (DEE) is made up of gaseous and particulate matter that includes nitro-polycyclic aromatic hydrocarbons (nitro-PAHs), nitrogen oxides (NOx), and sulfides, as well as other hydrocarbons. DEE and its components is classified by the International Agency for Research on Cancer (IARC) as a human carcinogen based on sufficient evidence of carcinogenicity to the lung (IARC, 2014). Mechanistic studies of how DEE exposure may lead to lung cancer or other airway diseases have focused on cellular and molecular effects. DEE and its various components have genotoxic effects that occur as a result of oxidative DNA damage and are thought to be important in the development of an inflammatory response and lung cancer (Attfield et al., 2012; Benbrahim-Tallaa et al., 2012; IARC, 2014; Neumeyer-Gromen et al., 2009). In a recent cross-sectional epidemiology study of factory workers in China, we showed that the number of CD4+, CD8+, and B lymphocyte subsets were increased in DEE-exposed workers from a diesel engine factory compared to those from factories with no known diesel exposures (Lan et al., 2015). Additional analyses from this study have also shown significant alterations in specific inflammatory markers and serum cytokine levels that have been linked to an increased lung cancer risk (Bassig et al., 2017; Dai et al., 2018).

The broad study of gene expression is also an important approach to identifying potential mechanisms of action for DEE in lung cancer pathogenesis. Experimental studies (Grilli et al., 2018; Jaguin et al., 2015; Koike et al., 2002; Kowalska et al., 2017; Li et al., 2009; Rossner et al., 2016; Verheyen et al., 2004; Zarcone et al., 2016) have shown that epithelial cells exposed to both the organic and particulate phases of DEE, or solely the particulate phase of DEE, known as diesel exhaust particulates (DEP), have altered expression of genes involved in xenobiotic metabolism (e.g., cytochrome P450 1A1 and 1B1), the DNA damage response (e.g., GADD45), response to hypoxia (e.g., HIF1a) and responsive to oxidative damage (e.g., NFE2L2). Previous gene expression studies of humans exposed to DEE have largely focused on Peripheral Blood Mononuclear Cell (PBMC) gene expression with short-term exposure (Peretz et al., 2007; Pettit et al., 2012) or relatively low-level environmental exposure (Espín-Pérez et al., 2018) that may not be directly transferrable to what would be observed in an occupational setting. A previous study of the effect of DEE and allergen exposure on the bronchial airway epithelium profiled the expression of a panel of immune-related genes (Rider et al., 2016).

To our knowledge, no study has yet performed genome-wide transcriptome profiling of the airway epithelium of humans chronically exposed to relatively high levels of both organic and particulate phases of DEE in an occupational setting. Based on the previously described concept of the airway field of injury (Steiling et al., 2008), we have observed that sampling the nasal epithelium reflects the transcriptomic changes that occur throughout the airway in response to smoking (Zhang et al., 2010), lung cancer (Perez-Rogers et al., 2017), and Chronic Obstructive Pulmonary Disease (COPD) (Boudewijn et al., 2017). We previously conducted a study of workers in China exposed to a wide range of DEE levels compared to a group of unexposed controls. Exposure was characterized in detail and nasal epithelial cells were collected and stored to optimize analysis of mRNA. Here, our purpose is to identify transcriptomic alterations in nasal epithelium of DEE-exposed factory workers relative to subjects without occupational DEE exposure to better characterize the cellular and molecular effects of DEE.

2. Methods

2.1. Study population

The design of the study has been described (Lan et al., 2015). In brief, 54 healthy male subjects were recruited from the testing facility of a factory that manufactured diesel engines, and 55 healthy male control subjects were recruited from the same geographic region from a beer bottling plant, a water treatment plant, a meat packing facility, and an administrative facility. A detailed walk-through survey was performed to determine that none of the latter workplaces contained any DEE sources. Subjects unexposed to DEE (controls) were frequency-matched to exposed workers on age (± 5 years) and smoking status (current, former, never). Former smoker status was defined as not being a current smoker and having smoked at least 20 cigarettes. Demographic and lifestyle characteristics were obtained for each worker through a questionnaire as part of a health examination conducted by the local Center for Disease Control.

Participation in the study was voluntary for all sampled subjects enrolled in the study, and written informed consent was obtained. The demographic data and nasal brushing sample collection was performed by the local Center for Disease Control (CDC) in China during the administration of a regular health exam. This study was approved by the US National Cancer Institute (NCI), as well as by the National Institute of Occupational Health and Poison Control in the Chinese Center for Disease Control and Prevention (CCDC).

2.2. Exposure analysis

Air monitoring was used to assess exposure to DEE. Assessment of repeated full-shift personal air exposure was measured using personal cyclone air sampling equipment attached to the lapels of diesel factory workers near the breathing zone. The methods for obtaining measurements of elemental carbon (EC), a major component of DEE particulate matter that has been used as a quantitative measure of DEE exposure in occupational settings, organic carbon (OC), soot, and particulate matter less than 2.5 μm in diameter ($\text{PM}_{2.5}$) has been previously described for this cohort (Lan et al., 2015). For DEE-exposed workers, exposures were quantified as the amount of each exposure metric (EC, OC, soot, and $\text{PM}_{2.5}$) measured in each individual. For workers without occupational DEE exposure, each exposure metric was quantified as an average across all control factories. Association between the DEE exposed and unexposed groups with respect to Never, Former, and Current smoker status groups was determined using a Fisher's Exact Test p-value. Wilcoxon Test was used to calculate the p-value between the DEE exposed and unexposed group for age, BMI, average cigarette use per day, and the four exposure metrics.

2.3. Nasal turbinate sample collection

One nasal brushing from each nostril (2 total) was obtained from each participant as previously described (Zhang et al., 2010). Each brush that was used to gently scrape the nasal epithelium of the volunteer subjects was placed into an individual tube with 1 ml of RNA-Later (Qiagen, Valencia, CA) immediately after sample collection and then frozen at -80°C and transported on dry ice. Material from the two brushes from each subject were combined, and high molecular-weight RNA was isolated using the miRNeasy Mini Kit (Qiagen, Valencia, CA). A NanoDrop spectrophotometer was used to assess RNA purity, while the Agilent BioAnalyzer was used to assess RNA integrity (RIN). After Quality Control (QC) analysis, 94 samples with at least 200 ng of RNA were selected for microarray analysis.

2.4. Microarray processing

At least 200 ng of analyzable RNA was obtained from 94 of the 109

study subjects and hybridized to Affymetrix Human Gene 1.0 ST GeneChips (Affymetrix, Santa Clara, CA) according to the manufacturer's protocol (Gene Expression Omnibus accession: GSE124267). Samples were split into 2 batches, which maintained an even distribution of samples handled by each of the two technicians processing the RNA, as well as evenly matched RIN scores in both batches. Three of the samples processed on microarrays were collected during field training and used to assess quality of laboratory assay and were not included in further analysis, leaving 91 total samples. Affymetrix Expression Console software (version 1.4.1.46) was used to normalize the arrays using the Robust Multiarray Average (RMA) procedure with the default Affymetrix probe-set mappings in order to compute detection above background (DABG) and Area Under the Receiver Operating Characteristics Curve (AUC) statistics for QC analysis. For the main analysis, a Chip Definition File (CDF) containing 19,718 Entrez Gene identifiers (hugene10stv1hsentrezgcdf and hugene10stv1hsentrezg.db packages) (Dai et al., 2005) was utilized for RMA (Gautier et al., 2004; Irizarry et al., 2003) normalization and probe-level summarization using the Affy package (Gautier et al., 2004), and using the R programming language (<http://r-project.org>) version 2.15.3. Additional data processing and statistical analysis was also done using this version of R.

2.5. Quality control and statistical analyses

Quality of the 91 arrays was determined using Relative Log Expression (RLE), Normalized Unscaled Standard Error (NUSE), and the AUC values derived from the Affymetrix Expression Console software. Arrays were deemed to have been of good quality if they had RLE values < 0.1, NUSE values < 1.05, and AUC values > 0.8. 12 samples were excluded from further analysis for not meeting these criteria. The remaining 79 samples (41 exposed, 38 controls) were renormalized. Adjustment for batch effects was made using ComBat (Johnson et al., 2007) in the Surrogate Variable Analysis (SVA) package for R.

A linear modeling approach was used to assess how gene expression changes were associated with DEE exposure (categorical, $n = 2$ levels) after adjusting for RIN (continuous), batch (categorical $n = 2$ levels), and smoking status (current smoker vs. former or never smoker, categorical, $n = 2$ levels). The model was further used to assess the gene expression changes associated with continuous diesel exposure metrics, EC, OC, soot, and $PM_{2.5}$. To recover gene expression effects specifically associated with cigarette smoking, we performed additional linear modeling for smoking status, adjusting for RIN, batch, and DEE exposure. This linear modeling approach was also used to explore the interaction effect between DEE exposure and smoking status by including a DEE-exposure*smoking interaction term with diesel, smoking, RIN and batch as covariates. t-statistics for each of the linear model coefficients and their corresponding p-values were calculated for each gene in each linear model using the `lmFit` function in LIMMA (Ritchie et al., 2015; Smyth, 2005). The False Discovery Rate (FDR) at each observed p-value was calculated using the method of Benjamini and Hochberg (Benjamini and Hochberg, 1995).

Elemental Carbon is commonly used as a marker of diesel particulate matter exposure both in laboratory and occupational settings (Schauer, 2003). Additionally, EC levels strongly correlated with levels of Organic Carbon and soot for each individual ($\rho = 0.98$). As DEE was the only notable source of EC exposure, EC was considered for further analysis. EC levels were divided into four groups based on ranges of raw values described previously, with the unexposed control group separate from the three subdivided DEE exposed groups and assigned a value of $11.1 \mu\text{g}/\text{m}^3$ as described (Lan et al., 2015).

The dataset from a previously published acute diesel exhaust (DE) exposure of PBMCs from healthy non-smokers in Peretz et al. (2007) was used for comparison to our findings. A subset of 9 participants at 6-hours post-DE exposure time point was selected for comparison of exposure to either $200 \mu\text{g}/\text{m}^3$ or $0 \mu\text{g}/\text{m}^3$ of diesel exhaust. Similarly, we also compared our results to another gene expression dataset from

Pettit et al. (2012) derived from PBMCs collected from participants exposed acutely to either $300 \mu\text{g}/\text{m}^3$ DE or clean air. In both datasets, a mixed linear modeling approach was used to assess exposure-associated gene expression, accounting for study participants as a random effect.

2.6. Enrichment and pathway analyses

Genes differentially expressed with respect to DEE exposure were divided into up and down-regulated gene-sets, which were used in EnrichR (Kuleshov et al., 2016) to perform pathway enrichment analysis with pathway gene sets from BioCarta (NCI), Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa and Goto, 2000), WikiPathways (Slenter et al., 2017), Gene Ontology (GO) Biological Process and GO Molecular Function (Ashburner et al., 2000) databases.

Gene Set Enrichment Analysis (GSEA) was used to determine the relationship between gene expression changes associated with DEE and other diesel exposure as well as cigarette smoke exposure datasets. Gene expression data from Peretz et al. (2007) and Pettit et al. (2012) were used to generate respective ranked lists of genes, sorted by their degree of differential expression in PBMCs between DE exposed and unexposed study participants based on the mixed-effects model coefficients as described previously. A ranked list was similarly generated for Beane et al. (2007) gene expression data in bronchial airway epithelium comparing current and never smokers via Student's t-statistic. Enrichment of gene sets comprised of the genes increased or decreased in DEE-exposed individuals at the top or bottom of this ranked list was then evaluated by GSEA (Subramanian et al., 2005).

2.7. Exposure-response relationship analysis

Principal component analysis was performed on the 225 gene Diesel signature to obtain the first principal component (PC1), which explains 35% of the variance in expression of the DEE signature genes, and summarizes the expression of the entire signature in single value. The PC1 scores were then analyzed relative to binned EC levels using pairwise Student's t-tests, and the Multinomial Cochran-Armitage test for trend (via the `multiCA` package for R).

2.8. Quantitative real-time polymerase chain reaction (qRT-PCR)

qRT-PCR was performed on selected genes using 16 DEE exposed and 16 unexposed control samples that were chosen based on availability of RNA. The genes selected were CYP1B1, CREBRF, CIR1, OSGIN2, CDRT1, and GAPDH as an endogenous reference gene for normalization. The RNA samples were reverse transcribed to cDNA with a mix of random hexamers and oligo-dT primers using RT² first strand kit (Qiagen, Valencia, CA). The PCR amplification mixture consisted of 9 ng of template cDNA, 12.5 μl of 2x RT² SYBR Green master mix, and 400 nM RT² qPCR Primer Assays (Qiagen). Amplification was performed for 40 cycles in a StepOnePlus Real-time PCR thermocycler (Applied Biosystems), and data acquisition was completed with StepOne software (version 2.2.2, Applied Biosystems), with threshold determination performed automatically for each reaction. The comparative CT method was used to obtain the gene expression levels for each gene relative to GAPDH, and fold changes were computed from the average expression values across each experimental group (Schmittgen and Livak, 2008). A linear modeling approach including DEE exposure and smoking status was used to assess the effect of DEE on gene expression in the qRT-PCR data. A linear modeling approach including DEE exposure, smoking status, and their interaction was used to explore the DEE-exposure*smoking interaction effect in the qRT-PCR data, specifically for CYP1B1 and CREBRF. Of the genes selected for qRT-PCR analysis, the microarray data for these two genes displayed a significant DEE-exposure*smoking interaction effect, and were representative of the two major trends observed among all 8 genes with a significant interaction effect.

3. Results

3.1. Study population

After quality control, 79 microarrays from individual subjects were retained for further gene expression analysis. Demographic and exposure information for the subjects are shown in Table 1. The subjects from the diesel factory worked an average of 18.5 years there, while the subjects from other factories worked an average of 12.8 years at their respective facilities. The subjects in each group were comparable based on age, body mass index (BMI) and smoking status. EC air levels were strikingly higher in DEE exposed workers compared to controls (median 60.7 vs 11.1 $\mu\text{g}/\text{m}^3$, respectively), whereas $\text{PM}_{2.5}$ levels were strongly elevated in the DEE exposed vs. unexposed workers (median 0.4 vs. 0.2 mg/m^3).

3.2. Differential gene expression associated with diesel engine exhaust exposure

We identified 225 genes at an FDR q-value cutoff of 0.25 that were differentially expressed in the nasal epithelium between the exposed and unexposed subjects (Diesel signature; Fig. 1 and Supplemental Tables 1A–C). Genes observed to be up-regulated include CYP1B1, NFE2L2, and HIF1a, which are involved in the xenobiotic metabolism, oxidative stress response, and hypoxia response pathways, respectively. The up-regulated gene set is also enriched for genes with functions in cell cycle maintenance, DNA repair, and circadian rhythm maintenance. Among the downregulated genes were the chloride channel, CFTR, as well as other genes responsible for transmembrane movement of solutes. Furthermore, the downregulated genes included calcium dependent genes and modification mechanisms, such as phospholipases, ribonucleases, and sulfotransferases.

Three genes from the DEE signature with the greatest statistical significance in the linear model, CIR1, CDRT1, OSGIN2, and CYP1B1 were selected for analysis with RT-PCR in 32 subjects. The expression of these genes was analyzed using the linear modeling approach as described for the microarray data above. Genes OSGIN2 and CYP1B1 had $p < 0.05$ in the linear model for the effect of DEE, while genes CIR1 and CDRT1 showed a trend consistent with results from the microarray analysis (Supplemental Fig. 1).

With the same linear modelling approach and covariates, the microarray data was analyzed for gene expression differences associated with other exposure metrics, including EC and $\text{PM}_{2.5}$. Notably, at FDR $q < 0.25$, there were no genes associated with $\text{PM}_{2.5}$, while there were 120 genes associated with EC (Elemental Carbon signature). Of the 120

genes in the Elemental Carbon signature, 86 genes were also present in the Diesel signature. Among the DEE-exposed subjects only, no genes were associated with increasing magnitude of exposure at FDR $q < 0.25$ in the analysis of any of the exposure metrics, suggesting that beyond some minimum threshold of exposure, the physiological response to DEE is saturated (Table 2).

3.3. Exposure-response relationship between the diesel signature and elemental carbon levels

We evaluated the exposure-response relationship of the DEE signature and levels of elemental carbon, by summarizing the expression of the genes in the DEE signature into a single value using the first principal component. As expected, the values of the first principal component are substantially higher in the DEE exposed workers vs. the unexposed controls. Moreover, each tertile of DEE exposure is significantly different from the control samples, and there is a significant increasing trend across all four groups ($p = 8.11\text{E} - 11$; Fig. 2).

3.4. Validation of gene expression differences associated with DEE

The gene expression findings were compared to two previously published studies that compared PBMC gene expression of acutely DE exposed and unexposed volunteers Peretz et al. (2007) and Pettit et al. (2012). We observed that the genes that increased in expression with DEE exposure in our dataset were significantly enriched among the genes that increased in expression in the PBMCs of the acutely DE exposed study participants in the Peretz et al. dataset (GSEA $q < 0.001$), but not in the Pettit et al dataset (GSEA $q = 0.1$) [Supplemental Fig. 2a and c]. The genes that decreased in expression with DEE exposure were enriched among the genes that decreased in expression with DE exposure in both Peretz et al dataset (GSEA $q = 0.04$) and Pettit et al dataset (GSEA $q = 0.03$) [Supplemental Fig. 2b and d].

3.5. Comparisons between diesel engine exhaust and cigarette smoke exposures

To examine the relationship between the effects of DEE and cigarette smoke exposures on airway gene expression, we used GSEA to determine whether genes in our nasal DEE signature are among the genes that have previously been found to be altered in the bronchial airway of smokers (Beane et al., 2007). We found that the genes increased with DEE exposure are significantly enriched among the genes that are most increased in current smokers (GSEA $q < 0.001$, Fig. 3a), while the genes decreased with DEE exposure are not significantly

Table 1

Descriptive characteristics of factory workers enrolled in the study and that were either exposed to diesel engine exhaust (Diesel) or unexposed (Control).

Subjects (n = 79)	Control (n = 38)		Diesel (n = 41)		p-value
Age, mean (SD)	41.3 (6.8)		41.0 (7.0)		0.95 [◇]
BMI, mean (SD)	25.1 (4.2)		24.6 (3.2)		0.78 [◇]
	Smoking Status				
Never, n (%)	5 (13.2)		8 (19.5)		0.72 [▽]
Former, n (%)	9 (23.7)		10 (24.4)		
Current, n (%)	24 (63.2)		23 (56.1)		
Avg. cigs/day (SD), Current	11.4 (6.6)		11.1 (5.5)		1 [◇]
	Exposure Metrics				
	Mean (SD)	Median (range)	Mean (SD)	Median (range)	
Elemental Carbon, µg/m ³	11.1 (1.34)	11.1 (9.9–12.6)	57.54 (21.1)	60.7 (17.2–105.4)	5.38E–16 [◇]
Organic Carbon, µg/m ³	68.7 (4.03)	68.7 (65.0–73.1)	136.66 (25.7)	131.3 (95.6–209.1)	5.38E–16 [◇]
Soot, abs coefficient in 10 ^{–5} /m	6.8 (0.64)	6.8 (6.2–7.5)	49.57 (14.6)	52.3 (23.0–77.3)	5.38E–16 [◇]
PM _{2.5} , mg/m ³	0.2 (0.07)	0.2 (0.2–0.3)	0.36 (0.1)	0.4 (0.2–0.5)	1.27E–14 [◇]

[◇] Wilcoxon Test.

[▽] Fisher's Exact Test.

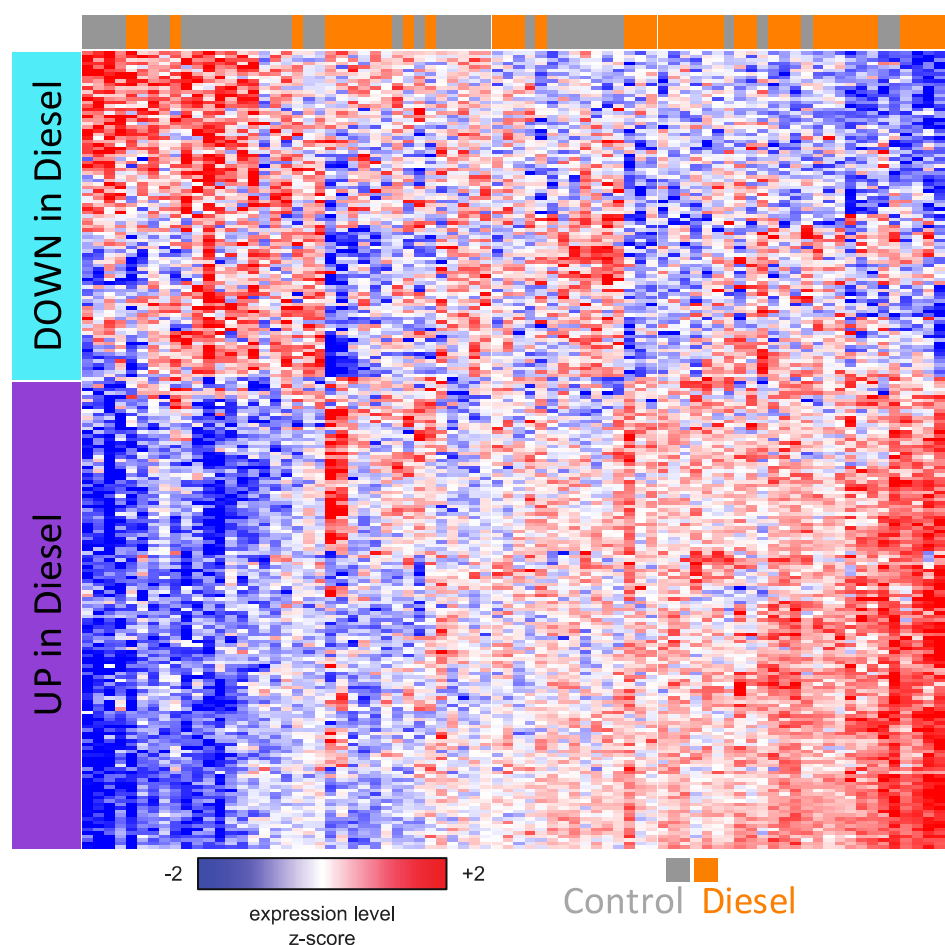


Fig. 1. Differential gene expression in subjects exposed to DEE ($n = 41$) compared to DEE unexposed controls ($n = 38$). Unsupervised hierarchical clustering of 225 genes significantly associated with DEE-exposure status after adjusting for RIN, batch, and smoking status (FDR $q < 0.25$). Each column represents a sample ($n = 79$) and each row represents each one of the 225 differentially expressed genes. The color bar on the left describes two clusters of gene expression patterns associated with DEE exposure. Genes with increased or decreased expression in DEE exposed individuals are designated as UP or DOWN in Diesel, respectively.

Table 2

Summary of the number of differentially expressed genes in linear models associated with diesel exposure, elemental carbon, or $PM_{2.5}$ at various statistical cutoffs in all samples or in diesel exposed samples only.

	n (Exposed/ Unexposed)	Statistical Cutoff	Diesel	Elemental Carbon	$PM_{2.5}$
All samples	41/38	$p < 0.05$	1869	1529	1101
		FDR	225	120	0
		$q < 0.25$			
Diesel exposed	41	$p < 0.05$	–	518	296
		FDR	–	0	0
		$q < 0.25$			

enriched among the genes down-regulated in the smoking dataset (FDR $q = 0.43$; Fig. 3b).

3.6. Interaction between diesel exhaust exposure and cigarette smoke exposure

Among all measured genes, we identified eight genes where the pattern of gene expression indicated a significant interaction between the effects of DEE and current smoking (Benjamini-Hochberg FDR $q < 0.25$; Supplemental Table 2). Of these, HOMEZ and CYP1B1 were also in the DEE signature. We observed two patterns of expression among the eight interaction-effect genes by exposure status (Fig. 4). First, both DEE and current smoking increased expression of CYP1B1,

CYP1A1, and MAP3K8, but the combined impact of both exposures resulted in expression levels that were lower than what would be expected if the independent effects of each exposure had a linearly additive effect. Second, and in contrast, DEE, but not smoking, altered the expression of CREBRF, HYAL2, ESRP2, HOMEZ, and FKBP4; moreover, the effect of DEE on the expression of these genes seemed to be attenuated in current smokers. We validated the interaction effect between smoking and DEE for CREBRF ($p = 0.03$), but not for CYP1B1 ($p = 0.13$) by RT-PCR using a subset of the RNA samples ($n = 32$), and observed similar expression patterns to those observed in microarray data for each gene across the different DEE and smoking subgroups (Supplemental Fig. 3).

4. Discussion

We conducted a cross-sectional study of healthy workers exposed to relatively high DEE levels with a wide range of exposure, compared to unexposed controls, and identified a DEE gene expression signature in epithelial samples from the nasal turbinate. To our knowledge, this is the first study to evaluate nasal airway epithelial gene expression of individuals heavily exposed to diesel engine exhaust.

Some of the genes that were increased with DEE exposure are involved in pathways known to be activated in response to oxidative and endoplasmic reticulum stress (CYP1B1, NFE2L2, GABPB1, MAPK8 [also known as JNK-1], CIR1, OSGIN2), hypoxia (HIF1a, BARD1, BHLHE40), and DNA damage (GADD45, RRM2B, SPRTN, SMC3). We also observed increased expression of genes involved in the circadian rhythm (FBXL3,

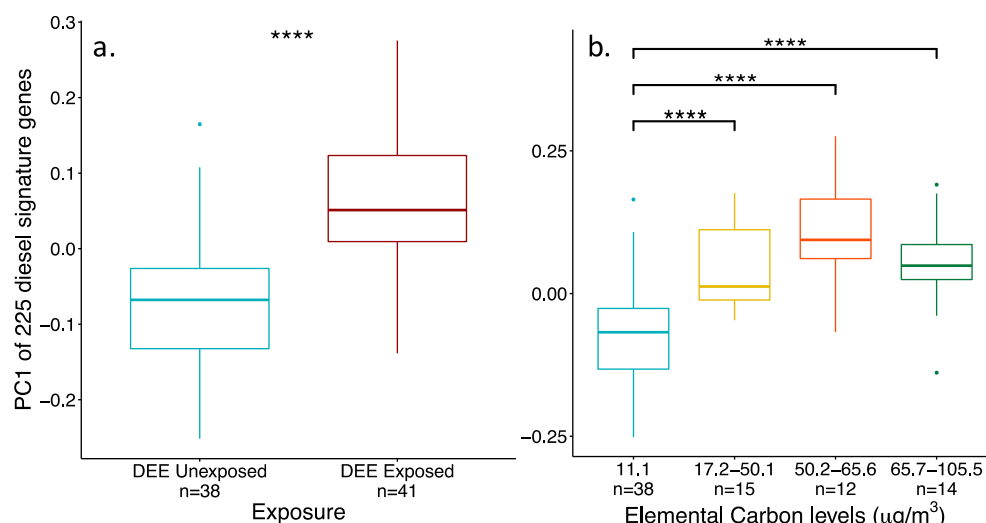


Fig. 2. DEE exposure alters gene expression in a dose-dependent manner. **a.** The first principal component (PC1) of the DEE signature genes differs between samples from diesel exposed ($n = 41$) and unexposed ($n = 38$) individuals. **b.** PC1 of the DEE signature genes increases with increasing elemental carbon (EC) exposure (Cochrane-Armitage trend test $p < 0.0001$). The DEE-exposed individuals were divided based on tertiles of EC exposure levels (with the ranges given below each box) as described previously (Lan et al. (2015)). The 38 non-DEE exposed individuals were assigned an EC exposure level of $11.1 \mu\text{g}/\text{m}^3$ based on the mean of values measured at the control factories (Lan et al., 2015). Pair-wise comparisons were made using Student's t -test. **** $p < 0.0001$. For each boxplot, the edges of the box are the 25th and 75th percentile of the data and the whiskers are the minimum and maximum values that do not exceed 1.5X the inter-quartile range.

LSM3, CRY1, BHLHE40), which is similar to effects that have been observed for cigarette and electronic cigarette exposures (Lechasseur et al., 2017). Finally, we observed increased expression in genes involved in the cell cycle and MAPK/ERK-related pathways (MAP3K2, MBIP, MAPK8, RRM2B, TBK1, ETS2, DUSP11, CAD). In particular, alteration in expression of each of the genes GADD45, MAPK8, NFE2L2, CYP1B1, and HIF1a have been previously reported in one or more experimental studies of DEE exposure (Grilli et al., 2018; Kowalska et al., 2017; Li et al., 2009; Rossner et al., 2016). Genes involved in molecular modification, such as phospholipases, ribonucleases, and sulfotransferases, were among those with decreased expression in DEE-exposed individuals. Genes that code for cross-membrane solute carriers and other transmembrane proteins, including the chlorine channel involved in cystic fibrosis, CFTR, were down-regulated as a result of DEE exposure. These findings suggest that DEE exposure affects the expression of genes whose expression is commonly observed to be altered in other environmental exposures and in lung cancer.

The gene expression changes that we observed in this study are significantly concordantly enriched among the genes previously observed to be most altered in PBMCs of human volunteers following acute $200 \mu\text{g}/\text{m}^3$ DE exposure (Peretz et al. (2007)), as well as with the genes most decreased in expression following a $300 \mu\text{g}/\text{m}^3$ acute exposure (Pettit et al., 2012). This suggests that the gene expression differences associated with long-term occupational exposure to DEE identified in our study at least in part reflect an ongoing acute response

to DEE exposure. Further analysis or time course studies would be required to parse out the differences between acute and chronic exposures at different doses.

Both diesel engine exhaust and cigarette smoking are causally associated with lung cancer (IARC, 2012, 2014). Additionally, some of the genes and pathways that we found to be altered as a result of DEE exposure are also altered in current smokers across various datasets. Consistent with this observation, we found that the genes that were increased with DEE exposure were enriched among the genes increased in current smokers relative to never smokers in a previously published dataset (Beane et al., 2007). We did not observe this relationship for the genes that were decreased with either exposure. This suggests that physiological responses corresponding to genes upregulated by DEE relative to controls, such as the oxidative stress response, are similar with both exposures. However, this finding also indicates that there may exist physiologic responses that are specific to DEE exposure represented among the genes downregulated by DEE, which were not similarly altered in comparison exposures such as cigarette smoke.

We identified several genes that showed an antagonistic interaction of DEE and tobacco smoking on gene expression, where the joint impact of both exposures was less than the sum of the effects observed for each exposure individually. Two of these genes, CYP1A1 and CYP1B1, were previously found to be induced by exposure to DEE or diesel particulate matter (Rossner et al., 2016). These results provide some potential biological insight into results from a nested case-control study of lung

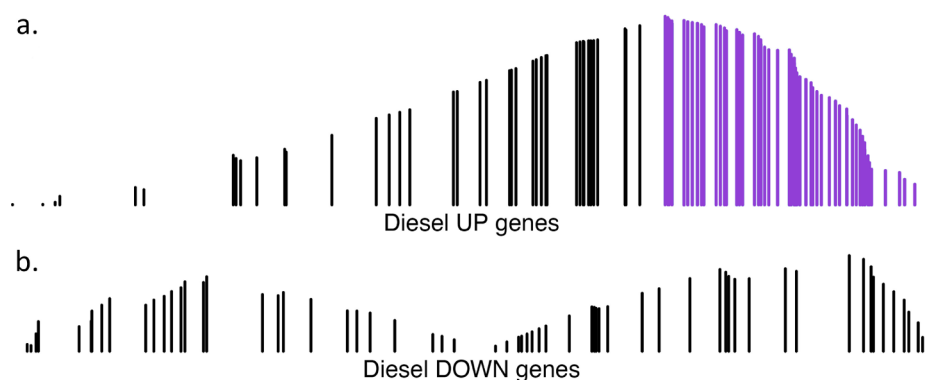


Fig. 3. Relationship between DEE and smoking-related gene expression. All of the genes measured on the microarray were ranked according to their degree of smoking-associated gene expression from most decreased (blue) to most increased (red). **a.** The position of the genes with increased expression in DEE-exposed individuals ($n = 41$) within the smoking-ranked list is indicated by the vertical lines, with the height of the line proportional to the GSEA running-enrichment score. The genes with increased expression in DEE-exposed individuals are significantly enriched among the genes most increased in current smokers (GSEA $q < 0.001$). The lines colored in purple represent the leading-edge gene set which comprised the most enriched genes. **b.** As in (a) but showing the position of the genes with decreased expression in DEE-exposed individuals.

individuals ($n = 41$) within the smoking-ranked list. Significant enrichment of the genes with decreased expression in DEE-exposed individuals was not detected among the genes with smoking-associated expression levels (GSEA $q = 0.43$).

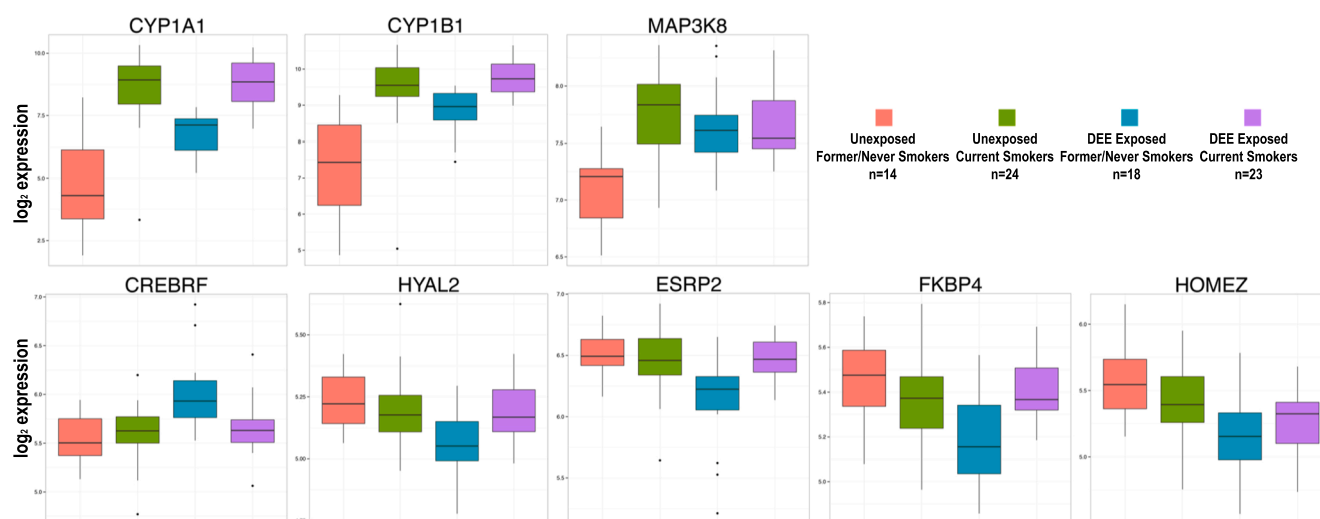


Fig. 4. Expression pattern of 8 genes across 79 samples in which there is a synergistic effect of DEE and cigarette smoking on gene expression levels (linear-model interaction effect FDR < 0.25). For the identified genes, the magnitude of the difference between the observed expression level of the DEE-exposed smokers and the unexposed non-smokers is less than would be predicted by the effects of DEE-exposure and smoking alone. For each boxplot, the edges of the box are the 25th and 75th percentile of the data and the whiskers are the minimum and maximum values that do not exceed 1.5X the inter-quartile range.

cancer among miners exposed to DEE, which found evidence of negative interaction between tobacco smoking and DEE in that each exposure attenuated the lung cancer risk of the other (Attfield et al., 2012). A similar negative interaction between tobacco and particulate exposure was also suggested in a study of the exceptionally high lung cancer incidence in Xuanwei, China, which has been causally linked to use of bituminous (i.e., “smoky coal”), where use of coal for home cooking and heating attenuated the effects of tobacco smoking on lung cancer risk (Kim et al., 2014). Our observations here may provide some insight into molecular mechanisms that play a role in the apparent antagonistic relationship between tobacco and other particulate exposures on risk of lung cancer. However, these interaction effects between DEE and smoking may be specific to the context of high amounts of DEE exposure as observed in the present study’s participants, and it is possible such patterns would not be observable at lower levels of exposure.

The sample size of our study was relatively small, but was composed of workers who were exposed to relatively high levels of DEE compared to many current workplace settings and well above environmental levels, as noted above. Important benefits of the study lie in the profiling of tissues that are directly exposed to DEE, as well as the minimally-invasive nature of sample collection in the nasal epithelium. The ability to obtain findings using the present techniques in this population that validate and extend previous *in vitro* and *in vivo* findings demonstrates the feasibility of the approach in understanding the effects of occupational exposures on the airway epithelium. Larger studies of individuals with the same exposure patterns would help to evaluate the presence and nature of interactions between DEE and other exposures, such as cigarette smoke, which may play a significant role in determining the likelihood of developing an airway malignancy. A larger study of workers exposed to lower levels of DEE would also be helpful to determine to what extent this signature is present in less exposed populations.

All of the workers at the diesel factory from which DEE exposed individuals were recruited were male. This limits the generalizability of our study, and future studies should include both men and women.

In conclusion, we identified a DEE gene signature in nasal epithelial cells among workers exposed to DEE compared to controls in China. The study has particular strengths in that the exposure was characterized in detail, exposure levels were relatively high compared to the typical levels studied in cross-sectional molecular epidemiology studies of DEE conducted in the West (Chiu et al., 2016), and unmeasured co-

exposures that could have contributed to gene expression patterns were likely to be minimal.

CRediT authorship contribution statement

E. Drizik: Data curation, Formal analysis, Writing - original draft, Writing - review & editing, Visualization. **S. Corbett:** Data curation, Formal analysis, Validation, Visualization. **Y. Zheng:** Conceptualization, Investigation, Supervision. **R. Vermeulen:** Conceptualization, Investigation, Supervision, Writing - review & editing. **Y. Dai:** Investigation. **W. Hu:** Investigation. **D. Ren:** Investigation. **H. Duan:** Investigation. **Y. Niu:** Investigation. **J. Xu:** Investigation. **W. Fu:** Investigation. **K. Meliefste:** Investigation. **B. Zhou:** Investigation. **Xiaohui Zhang:** Investigation. **J. Yang:** Investigation. **Bryan Bassig:** Investigation. **Hanqiao Liu:** Investigation. **M. Ye:** Investigation. **Gang Liu:** Investigation. **X. Jia:** Investigation. **T. Meng:** Investigation. **P. Bin:** Investigation. **J. Zhang:** Investigation. **D. Silverman:** Conceptualization, Investigation. **A. Spira:** Conceptualization, Supervision, Funding acquisition, Writing - review & editing. **N. Rothman:** Conceptualization, Supervision, Funding acquisition, Investigation, Writing - review & editing. **M.E. Lenburg:** Conceptualization, Supervision, Writing - review & editing. **Q. Lan:** Conceptualization, Supervision, Funding acquisition, Investigation, Writing - review & editing.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Dr. Avrum Spira is an employee of Johnson and Johnson.

Acknowledgements

This work is partly supported by National Natural Science Foundation of China (NSFC 91643203) and the National Institutes of Health/National Cancer Institute (HHSN261201500229P).

Appendix A. Supplementary material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.envint.2020.105506>.

References

- Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., et al., 2000. Gene ontology: Tool for the unification of biology. *Nat. Genet.* 25, 25.
- Attfield, M.D., Schleiff, P.L., Lubin, J.H., Blair, A., Stewart, P.A., Vermeulen, R., et al., 2012. The diesel exhaust in miners study: a cohort mortality study with emphasis on lung cancer. *J. Natl Cancer Inst.* 104, 869–883.
- Bassig, B.A., Dai, Y., Vermeulen, R., Ren, D., Hu, W., Duan, H., et al., 2017. Occupational exposure to diesel engine exhaust and alterations in immune/inflammatory markers: a cross-sectional molecular epidemiology study in China. *Carcinogenesis* 38, 1104–1111.
- Beane, J., Sebastiani, P., Liu, G., Brody, J.S., Lenburg, M.E., Spira, A., 2007. Reversible and permanent effects of tobacco smoke exposure on airway epithelial gene expression. *Genome Biol.* 8, R201.
- Benbrahim-Tallaa, L., Baan, R.A., Grosse, Y., Lauby-Secretan, B., El Ghissassi, F., Bouvard, V., et al., 2012. Carcinogenicity of diesel-engine and gasoline-engine exhausts and some nitroarenes. *Lancet Oncol.* 13, 663–664.
- Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Roy. Stat. Soc.: Ser. B (Methodol.)* 57, 289–300.
- Boudewijn, I.M., Faiz, A., Steiling, K., van der Wiel, E., Telenga, E.D., Hoonhorst, S.J., et al., 2017. Nasal gene expression differentiates copd from controls and overlaps bronchial gene expression. *Respirat. Res.* 18, 213.
- Chiu, Y.-H.M., Garshick, E., Hart, J.E., Spiegelman, D., Dockery, D.W., Smith, T.J., et al., 2016. Occupational vehicle-related particulate exposure and inflammatory markers in trucking industry workers. *Environ. Res.* 148, 310–317.
- Dai, M., Wang, P., Boyd, A.D., Kostov, G., Athey, B., Jones, E.G., et al., 2005. Evolving gene/transcript definitions significantly alter the interpretation of genechip data. *Nucl. Acids Res.* 33, e175.
- Dai, Y., Ren, D., Bassig, B.A., Vermeulen, R., Hu, W., Niu, Y., et al., 2018. Occupational exposure to diesel engine exhaust and serum cytokine levels. *Environ. Mol. Mutagen.* 59, 144–150.
- Espín-Pérez, A., Krauskopf, J., Chadeau-Hyam, M., van Veldhoven, K., Chung, F., Cullinan, P., et al., 2018. Short-term transcriptome and microRNAs responses to exposure to different air pollutants in two population studies. *Environ. Poll.* 242 (Pt A), 182–190. <https://doi.org/10.1016/j.envpol.2018.06.051>.
- Gautier, L., Cope, L., Bolstad, B.M., Irizarry, R.A.J.B., 2004. Affy—analysis of affymetrix genechip data at the probe level. *Bioinformatics* 20, 307–315.
- Gentner, D.R., Isaacman, G., Worton, D.R., Chan, A.W.H., Dallmann, T.R., Davis, L., et al., 2012. Elucidating secondary organic aerosol from diesel and gasoline vehicles through detailed characterization of organic carbon emissions. *P. Natl. Acad. Sci. USA* 109, 18318–18323.
- Grilli, A., Bengalli, R., Longhin, E., Capasso, L., Proverbio, M.C., Forcato, M., et al., 2018. Transcriptional profiling of human bronchial epithelial cell beas-2b exposed to diesel and biomass ultrafine particles. *BMC Genom.* 19, 302.
- IARC, 2012. IARC monographs on the evaluation of carcinogenic risks to humans: Personal habits and indoor combustions. Lyon, France.
- IARC, 2014. IARC monographs on the evaluation of carcinogenic risks to humans: Diesel and gasoline engine exhausts and some nitroarenes. Lyon, France.
- Irizarry, R.A., Hobbs, B., Collin, F., Beazer-Barclay, Y.D., Antonellis, K.J., Scherf, U., et al., 2003. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* 4, 249–264.
- Jaguin, M., Fardel, O., Lecureur, V., 2015. Exposure to diesel exhaust particle extracts (depe) impairs some polarization markers and functions of human macrophages through activation of ahr and nrf2. *PLoS ONE* 10.
- Johnson, W.E., Li, C., Rabinovic, A.J.B., 2007. Adjusting batch effects in microarray expression data using empirical bayes methods. *Biostatistics* 8, 118–127.
- Kanehisa, M., Goto, S., 2000. Kegg: Kyoto encyclopedia of genes and genomes. *Nucl. Acids Res.* 28, 27–30.
- Kim, C., Chapman, R.S., Hu, W., He, X., Hosgood, H.D., Liu, L.Z., et al., 2014. Smoky coal, tobacco smoking, and lung cancer risk in xuanwei, china. *Lung Cancer* 84, 31–35.
- Koike, E., Hirano, S., Shimojo, N., Kobayashi, T., 2002. Cdna microarray analysis of gene expression in rat alveolar macrophages in response to organic extract of diesel exhaust particles. *Toxicol. Sci.* 67, 241–246.
- Kowalska, M., Wegierek-Ciuk, A., Brzoska, K., Wojewodzka, M., Meczynska-Wielgosz, S., Gromadzka-Ostrowska, J., et al., 2017. Genotoxic potential of diesel exhaust particles from the combustion of first-and second-generation biodiesel fuels—the fuelhealth project. *Environ. Sci. nd Poll. Res.* 24, 24223–24234.
- Kuleshov, M.V., Jones, M.R., Rouillard, A.D., Fernandez, N.F., Duan, Q., Wang, Z., et al., 2016. Enrichr: A comprehensive gene set enrichment analysis web server 2016 update. *Nucl. Acids Res.* 44, W90–W97.
- Lan, Q., Vermeulen, R., Dai, Y., Ren, D., Hu, W., Duan, H., et al., 2015. Occupational exposure to diesel engine exhaust and alterations in lymphocyte subsets. *Occup. Environ. Med.* 72, 354–359.
- Lechasseur, A., Jubinville, E., Routhier, J., Berube, J.C., Hamel-Auger, M., Talbot, M., et al., 2017. Exposure to electronic cigarette vapors affects pulmonary and systemic expression of circadian molecular clock genes. *Physiol. Rep.* 5.
- Li, R., Ning, Z., Cui, J., Khalsa, B., Ai, L., Takabe, W., et al., 2009. Ultrafine particles from diesel engines induce vascular oxidative stress via JNK activation. *Free Rad. Biol. Med.* 46, 775–782.
- NCI. Cancer genome anatomy project: Biocarta pathways.
- Neumeyer-Gromen, A., Razum, O., Kersten, N., Seidler, A., Zeeb, H., 2009. Diesel motor emissions and lung cancer mortality—results of the second follow-up of a cohort study in potash miners. *Int. J. Cancer* 124, 1900–1906.
- Peretz, A., Peck, E.C., Bammler, T.K., Beyer, R.P., Sullivan, J.H., Trenga, C.A., et al., 2007. Diesel exhaust inhalation and assessment of peripheral blood mononuclear cell gene transcription effects: an exploratory study of healthy human volunteers. *Inhal. Toxicol.* 19, 1107–1119.
- Perez-Rogers, J.F., Gerrein, J., Anderlind, C., Liu, G., Zhang, S., Alekseyev, Y., et al., 2017. Shared gene expression alterations in nasal and bronchial epithelium for lung cancer detection. *JNCI: J. Natl. Canc. Instit.* 109.
- Pettit, A.P., Brooks, A., Laumbach, R., Fiedler, N., Wang, Q., Strickland, P.O., et al., 2012. Alteration of peripheral blood monocyte gene expression in humans following diesel exhaust inhalation. *Inhalation Toxicol.* 24, 172–181.
- Resitoglu, I.A., Altinisik, K., Keskin, A., 2015. The pollutant emissions from diesel-engine vehicles and exhaust aftertreatment systems. *Clean Technol. Environ.* 17, 15–27.
- Rider, C.F., Yamamoto, M., Günther, O.P., Hirota, J.A., Singh, A., Tebbutt, S.J., et al., 2016. Controlled diesel exhaust and allergen coexposure modulates microRNA and gene expression in humans: effects on inflammatory lung markers. *J. Aller. Clin. Immunol.* 138, 1690–1700.
- Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., et al., 2015. Limma powers differential expression analyses for rna-sequencing and microarray studies. *Nucl. Acids Res.* 43, e47.
- Rossner, P., Strapacova, S., Stolcpartova, J., Schmuczerova, J., Milcova, A., Neca, J., et al., 2016. Toxic effects of the major components of diesel exhaust in human alveolar basal epithelial cells (a549). *Int. J. Mol. Sci.* 17, 1393.
- Schauer, J.J., 2003. Evaluation of elemental carbon as a marker for diesel particulate matter. *J. Exposure Sci. Environ. Epidemiol.* 13, 443.
- Schmittgen, T.D., Livak, K.J., 2008. Analyzing real-time pcr data by the comparative c t method. *Nat. Protoc.* 3, 1101.
- Slenter, D.N., Kutmon, M., Hanspers, K., Riutta, A., Windsor, J., Nunes, N., et al., 2017. Wikipathways: A multifaceted pathway database bridging metabolomics to other omics research. *Nucl. Acids Res.* 46, D661–D667.
- Smyth, G.K., 2005. Limma: Linear models for microarray data. In: *Bioinformatics and computational biology solutions using r and bioconductor*. Springer, pp. 397–420.
- Steiling, K., Ryan, J., Brody, J.S., Spira, A., 2008. The field of tissue injury in the lung and airway. *Cancer Prev. Res. (Phila.)* 1, 396–403.
- Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., et al., 2005. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci.* 102, 15545–15550.
- Verheyen, G.R., Nuijten, J.M., Van Hummelen, P., Schoeters, G.R., 2004. Microarray analysis of the effect of diesel exhaust particles on in vitro cultured macrophages. *Toxicol. In Vitro* 18, 377–391.
- Zarcone, M.C., Duistermaat, E., van Schadewijk, A., Jedynska, A., Hiemstra, P.S., Kooter, I.M., 2016. Cellular response of mucociliary differentiated primary bronchial epithelial cells to diesel exhaust. *Am. J. Physiol. Lung Cell. Mol. Physiol.* 311, L111–L123.
- Zhang, X., Sebastiani, P., Liu, G., Schembri, F., Zhang, X., Dumas, Y.M., et al., 2010. Similarities and differences between smoking-related gene expression in nasal and bronchial epithelium. *Physiol. Genom.* 41, 1.