A Systems Biology Approach to Investigate Human Lung Cell Response to Air Pollutants

Julia Elizabeth Rager

A thesis submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Master of Science in the Department of Environmental Sciences and Engineering of the Gillings School of Global Public Health

Chapel Hill 2010

Approved by:

Advisor: Dr. Rebecca C. Fry

Dr. David Leith

Dr. Ilona Jaspers

Dr. Kenneth G. Sexton

Dr. William Vizuete

THESIS ABSTRACT

Julia Elizabeth Rager

A Systems Biology Approach to Investigate Human Lung Cell Response to Air Pollutants

(Under the direction of Dr. Rebecca C. Fry)

Exposure to air pollution is associated with many diseases, such as asthma, bronchitis, and lung cancer. Despite these adverse health effects, the cellular mechanisms underlying air pollution-associated diseases remain largely unknown. In this study we set out to investigate the genome-wide responses of human lung cells exposed to multiple air pollutant conditions. We first employ a toxicogenomic approach to compare transcripts and molecular networks modulated upon exposure to freshly emitted air pollutants and photochemically altered pollutant mixtures, containing secondary pollutants. The results demonstrate that secondary pollutants initiate a more robust genomic response. After identifying this trend, we investigate potential mechanisms underlying responses to individual secondary pollutants. Here, we evaluate global microRNA expression modifications resulting from formaldehyde exposure. Our analysis reveals that formaldehyde induces significant changes in microRNA levels, which may in turn, regulate genes associated with inflammation and cancer. Together, these investigations reveal novel mechanisms potentially underlying air pollutant-induced disease.

ACKNOWLEDGEMENTS

I would first like to thank my advisor, Dr. Rebecca C. Fry, for welcoming me into her research group at UNC. I greatly appreciate her ability to guide her students, as she provides help when we need it, but at the same time, allows us to learn and grow on our own. I am especially grateful to her for allowing me to become involved in multiple projects as I'm trying to learn as much as possible in the genetics research field. Not only has she acted as a great mentor, but she's given me confidence in myself, such that I can achieve any goals I set, however constantly changing they seem to be. I will forever be grateful for her making my graduate experience at UNC fun, gratifying, and meaningful.

I would also like to thank our Research Specialist and Lab Manager, Lisa Smeester, who dedicated much of her time to helping me plan experiments and learn valuable laboratory techniques. Most importantly, however, she has provided me with friendship and daily entertainment for the past two years.

I would like to give special recognition to UNC's One Atmosphere research team, who helped me immensely with the air exposures. In particular, Kim de Bruijne and Seth Ebersviller devoted much of their time helping me learn several laboratory techniques related to cell exposure and assessment. Dr. Kenneth G. Sexton taught me about air chemistry in the research setting, and at the same time, served as a valuable mentor. This is a unique team of generous and thoughtful people, and I feel lucky to have been a small part of it.

I would also like to thank my committee members, Dr. David Leith, Dr. Ilona Jaspers, Dr. Kenneth G. Sexton, and Dr. William Vizuete, for their valuable input and advice in preparing my research project and thesis.

Lastly, I would like to thank my parents, Brent and Diane Rager, who have been extremely supportive of me throughout the years. Their guidance has made me the person I am today, and I would like to thank them for allowing me to grow-up in an environment that provided me with limitless goals and possibilities.

PREFACE

Toxicology studies analyzing the health effects of air pollutant exposure are becoming increasingly important, as the number of epidemiology studies showing strong associations between air pollution exposure and adverse health effects are rapidly growing (Brunekreef and Holgate, 2002; Burnett *et al.*, 2001; Cohen *et al.*, 2005; Götschi *et al.*, 2008). Some air pollution health effects may occur over time due to chronic exposure, as with chronic bronchitis and lung cancer (Brunekreef and Holgate, 2002; Nyberg *et al.*, 2000). Other health effects occur more abruptly, as with increased hospital admissions, acute respiratory symptoms, cardiovascular events, and decreased lung function (Brunekreef *et al.*, 1995; Dab *et al.*, 1996). Despite these serious health effects, the cellular mechanisms underlying air pollutant-induced disease remain largely unknown.

Our study set out to investigate lung cell responses to multiple air pollutant exposure conditions on a genome-wide scale. We perform our studies using a systems level approach, where pollutant-induced lung cell responses are identified and related molecular networks are constructed. Our goal is to increase the mechanistic understanding of air pollutant-related diseases, particularly in response to multiple components of air pollution present in urban atmospheres.

Thesis Organization

This report is organized into two chapters. Chapter One evaluates the genome-wide response of human lung cells exposed to gaseous pollutant mixtures found in urban atmospheres. Specifically, cellular response, at the gene-transcript level, is evaluated by identifying genes and molecular pathways modulated upon exposure to freshly emitted air pollutants. These response mechanisms are then compared against genes and networks modulated upon exposure to photochemically altered pollutant mixtures, containing secondary pollutants. This study reveals that secondary pollutants produced from primary pollutants irradiating

throughout the day cause a much more robust cellular response than primary pollutants. After identifying this trend, we focus on the potential mechanisms underlying lung cell response from exposure to individual secondary pollutants.

As detailed in Chapter Two, we focus our analysis on formaldehyde, a secondary pollutant found in both indoor and outdoor air pollution. Here, we investigate human lung cell response to formaldehyde exposure by measuring microRNA expression profiles across the genome. We reveal that formaldehyde induces significant changes in microRNA expression patterns, which may in turn, regulate genes associated with inflammation and cancer. These two chapters, combined, reveal novel mechanisms potentially underlying air pollutant-induced disease.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ADDITIONAL FILES

CHAPTER 1
A Toxicogenomic Comparison of Primary and Photochemically Altered Air Pollutants in Human Lung Cells

ABSTRACT

Background: Outdoor air pollution contributes significantly to global increases in mortality, particularly within urban environments. Studies have investigated cellular responses in lung cells caused by single pollutants, such as individual gases or particulate matter. There is, however, limited knowledge of the mechanisms underlying health effects resulting from exposure to pollutant mixtures. Toxicogenomic analyses can increase our mechanistic understanding of exposure effects through observations of gene expression alterations and their associated molecular pathways.

Objectives: We set out to compare the cellular responses of lung cells exposed to primary pollutants relative to photochemically altered (PCA) pollutant mixtures found in urban atmospheres. We hypothesized that lung cells exposed to PCA pollutants will show increased modulation of inflammatory-associated genes and pathways relative to primary pollutants.

Methods: Human lung epithelial cells were exposed to either primary or PCA mixtures. Transcription changes were assessed using microarrays and confirmed using RT-PCR on a subset of genes.

Results: There was a large difference in cell response with primary air pollutants altering the expression levels of 19 genes, while PCA pollutants altered 709 genes. Functional and molecular analyses of the altered genes showed that pollutant exposure modifies the signaling of pathways associated with cancer and inflammation. To confirm the altered inflammatory response, interleukin-8 were measured at significantly increased levels.

Conclusions: Our study shows for the first time that exposure to PCA pollutants results in robust gene expression changes relative to primary components of air pollution. These transcription changes may, in part, explain the health effects related to air pollutant exposure.

1.1 INTRODUCTION

Outdoor air pollution contributes significantly to global increases in morbidity and mortality, particularly within urban environments (Burnett *et al.*, 2000; Cao *et al.*, 2009). Some air pollution health effects may occur over time due to chronic exposure, as with chronic bronchitis and lung cancer (Brunekreef and Holgate, 2002; Nyberg, *et al.*, 2000). Other health effects occur more abruptly, as with increased hospital admissions, acute respiratory symptoms, cardiovascular events, and decreased lung function (Brunekreef, *et al.*, 1995; Dab *et al.*, 1996). Despite these adverse health outcomes, the underlying molecular mechanisms associating air pollutant mixtures with disease remain largely unknown.

1.1.1 Toxicogenomic Analyses of Air Pollution Health Effects

Toxicogenomic analyses can increase our understanding of the molecular mechanisms that associate exposure to air pollution with disease. Specifically, transcriptional studies can be used to assess alterations in mRNA levels providing information on the genes and regulatory pathways that are modified through environmental exposures (McHale *et al.*, 2010). Knowledge of signaling pathways altered by air pollution mixtures is currently limited, with primarily exposures to diesel exhaust and particulate matter mixtures studied. For example, exposure to diesel exhaust and/or particulate matter alters the signaling of cyclooxygenase-2 (*COX-2*), interleukin-6 (*IL-6*), interleukin-8 (*IL-8*), and tumor necrosis factor (*TNF*) as well as signal transducer and activator of transcription 3 (STAT3) (Becker *et al.*, 2005; Cao *et al.*, 2007). Individual gas-phase pollutant exposures have also been studied. For example, exposure to ozone or sulfur dioxide results in the altered signaling of cytokines, such as IL-6 and IL-8, as well as activator protein-1 (AP-1) and nuclear factor kappa beta (NFκB) (Damera *et al.*, 2009; Jaspers *et al.*, 1997; Thompson *et al.*, 2009). However, studies evaluating responses to gaseous pollutant mixtures are limited. Furthermore, to our knowledge, no studies have evaluated the genome-wide transcriptional response of lungs cells exposed to gaseous primary emitted pollutant mixtures relative to photochemically altered (PCA) air mixtures.

1.1.2 Air Pollution Chemistry

In order to link air pollution exposure to health outcomes, it is important to accurately characterize exposure conditions by understanding air pollution chemistry. Within urban air environments, primary pollutants emitted into the atmosphere undergo photochemical reactions in sunlight and form various secondary pollutants giving rise to photochemical smog. The main primary pollutants emitted are nitrogen oxides (NO_x) and volatile organic compounds (VOCs). These primary pollutants then photochemically react during daylight to generate secondary products, including ozone, peroxyacetyl nitrate, formaldehyde, and other various carbonyls (EPA, 2006). This complex chemistry between primary emitted air pollutants and sunlight has created an environment where almost 175.3 million (58%) Americans currently live in areas with unhealthy levels of pollution (ALA, 2010).

1.1.3 Study Objectives

In this study, we set out to investigate the toxicogenomic response of lung cells exposed to gaseous mixtures that mimic urban atmospheric conditions. Specifically, we analyze the transcriptomic response of lung cells exposed to common air pollutants reacting photochemically with sunlight. Through the use of an environmental irradiation chamber, we compare gene expression patterns and inflammatory responses in cells exposed to primary pollutants with cells exposed to PCA pollutant mixtures.

1.2 MATERIALS AND METHODS

1.2.1 Generation of Pollutant Mixtures

An outdoor environmental irradiation chamber (120 m^3 volume) was used to prepare air exposure conditions. Outdoor environmental irradiation chambers are photochemical reactors that utilize natural sunlight to initiate the natural photochemical transformation chemistry of air pollutants (Jeffries *et al.*, 1976; Jeffries, 1995; Sexton *et al.*, 2004; Yu *et al.*, 1997). Synthetic Urban Mix, a volatile organic carbon (VOC) mixture, was used as the starting material for the test atmosphere. This mixture contains 55 different hydrocarbons at specific

ratios representing chemicals present in urban atmospheres (Jeffries and Sexton, 1995). On the morning of the exposure day, the volatile organics of Synthetic Urban Mix were drawn from a gas cylinder into the environmental irradiation chamber, while a liquid mixture containing the less volatile organics was injected into the chamber. In addition, oxides of nitrogen (NO_x) were drawn from a gas cylinder (Reference # 8857085) into the chamber to establish a test atmosphere containing 2 ppmC Synthetic Urban Mix and 0.2 ppm NO_x. This test atmosphere remained inside the environmental irradiation chamber throughout the day where sunlight induced photochemical reactions between the mixture's compounds, and secondary chemical products were generated.

1.2.2 Cell Culture

Human A549 type II lung epithelial cells, derived from a human lung adenocarcinoma, were cultured according to standard protocol (ATCC). Cells were grown in growth media containing F-12K plus 10% FBS plus 1% penicillin and streptomycin. Cells were plated onto 24 mm diameter collagen-coated membranes with 0.4 μm pores (Trans-CLR; Costar, Cambridge, MA). Upon confluence, cells were cultured in phenol red-free F-12K nutrient mixture without FBS. Immediately prior to exposure, apical media was removed in order to create direct air-liquid interface culture conditions. The media beneath each membrane remained to supply nourishment for cells throughout the exposure.

1.2.3 Exposure to Pollutant Mixtures

A coupled chamber-*in vitro* exposure system was used for this study (Sexton, *et al.*, 2004). Sample lines directly linked the environmental irradiation chamber's mixture to a cellular exposure chamber (Billups-Rothenberg, Modular Incubator Chamber, Del Mar, CA), where air exposures were continuously drawn through at 1.0 L/min. The cellular exposure chamber was positioned within an incubator (set at 37°C), where CO_2 was added to the exposure source stream at 0.05 L/min and a small water dish provided proper humidification.

The first exposure was in the morning, representing the primary pollutant mixture exposure. Lung cells were exposed for four hours, while mock-treated control cells were exposed to humidified air under similar conditions. The second exposure was in the evening,

representing the PCA pollutant mixture, which contained primary pollutants and secondary products of irradiation chemistry. For this treatment, prepared lung cells were exposed for four hours, while another set of mock-treated control cells were exposed to humidified air. For each exposure period, four exposed sample wells and four unexposed sample wells were used, resulting in a total of 16 total samples. Cells were incubated for nine hours after each respective exposure period. Cells were then scraped and stored at -80°C in TRIzol® Reagent (Invitrogen Life Technologies), and basolateral supernatants were aspirated and stored at -80°C.

1.2.4 Chemical Analysis of Atmospheric Mixtures

Gas measurement methods were used to assess the chemical constituents within the chamber during the experiment. Volatile organic hydrocarbon levels were measured five times during the experiment using a Varian STAR 3400 capillary gas chromatograph - flame ionization detection with a Varian Saturn 2000 ion trap mass spectrometer. Ozone was measured every minute using an EPA equivalent method (EQOA-0193-091) based on chemiluminescence with a Teledyne Monitor Labs Model 9811 monitor (Englewood, CO). NO and NO_2 levels were measured every minute using an EPA standard reference method (RFNA-1292-090) based on chemiluminescence with a Teledyne Monitor Labs Model 9841. Formaldehyde concentrations were measured continuously using a Dasgupta-diffusion-tube sampler (Dasgupta *et al.*, 1988). Peroxyacetyl nitrate levels were measured every 30 minutes using a Varian CP-3800 gas chromatograph.

1.2.5 Analysis of Cytotoxicity

To measure cytotoxicity, the enzyme lactate dehydrogenase (LDH) was measured within the supernatant of each sample. LDH levels were assessed using a coupled enzymatic assay, according to the supplier's instructions (Clontech Laboratories, Inc., Mountain View, CA). For each exposure period, four unexposed and four exposed sample wells were analyzed in technical triplicate for a total of 24 measurements. Scanned absorbance reading outliers were identified through the Grubbs' test where outliers were identified as those with less than a 5% probability of occurring as an outlier by chance alone, relative to a normal distribution (Grubbs, 1969). Fold increase was calculated as $\mu_{\text{LDH exposed}} / \mu_{\text{LDH unexposed}}$, where μ

represents the mean LDH levels. Statistical significance of the exposed versus unexposed LDH levels was calculated using an unpaired t-test with Welch's correction (Welch, 1938).

1.2.6 Microarray Processing

Total RNA was isolated from unexposed cells and cells exposed to either primary or PCA air pollutants using Qiagen's RNeasy® Kit according to the manufacturer's protocol (Valencia, CA). RNA was quantified with the NanoDrop™ 1000 Spectrophotometer (Thermo Scientific, Waltham, MA) and its integrity was verified with an Agilent Technologies 2100 Bioanalyzer (Santa Clara, CA). RNA was biotin-labeled according to the Affymetrix protocol and hybridized to Affymetrix GeneChip® Human Gene 1.0 ST arrays, which probe for 28,869 genes. Samples were assessed in biological duplicate for each exposure condition: primary pollutant exposure and time-matched unexposed control, PCA pollutant exposure and time-matched unexposed control, for a total of eight microarray samples.

1.2.7 Microarray Analysis

Microarray data were first normalized using Robust Multi-Chip Average (RMA) (Irizarry *et al.*, 2003) and filtered for expression levels above background noise (> abs[30]) which resulted in a reduction of probesets from 28,869 to 24,652 for primary mixture conditions and 24,830 for PCA mixture conditions. Differential gene expression was defined as a significant difference in mRNA levels between exposed versus unexposed samples, where the following three statistical requirements were set: (1) fold change of ≥ 1.5 or ≤ -1.5 (exposed versus unexposed); (2) p-value < 0.05 (ANOVA); and (3) a false discovery rate corrected q-value < 0.05. Analysis of variance (ANOVA) p-values were calculated using Partek® Genomics Suite™ software (St. Louis, MO). To control the rate of false positives, q-values were calculated as the minimum "positive false discovery rate" that can occur when identifying significant hypotheses (Storey, 2003). Microarray data have been submitted to NCBI's Gene Expression Omnibus (GEO) repository (http://www.ncbi.nlm.nih.gov/geo/) and are available under accession number GSE23735 (Edgar *et al.*, 2002).

1.2.8 Enriched Biological Functions and Network Analysis

Biological functions and molecular networks associated with air pollutant exposure were identified using the Ingenuity database (Ingenuity Systems, www.ingenuity.com, Redwood

City, CA). The Ingenuity database provides a collection of gene to phenotype associations, molecular interactions, regulatory events, and chemical knowledge accumulated to develop a global molecular network. The lists of differentially expressed genes were overlaid onto this global molecular network, where related networks were algorithmically constructed based on connectivity. Functional analysis was carried out to identify biological functions and disease signatures most significantly associated with the differentially expressed genes. Statistical significance of each biological function or disease was calculated using a Fischer's exact test. Functions with p-values $\leq 3.0 \times 10^{-6}$ were evaluated.

1.2.9 Transcription Factor Binding Site Analysis

Transcription factor binding site analysis was performed using the EXPANDER Software, version 5.1 (acgt.cs.tau.ac.il/expander/). For this analysis, Affymetrix probesets were linked to sequence data in regions 1,000 base pairs upstream and 200 base pairs downstream of the transcription start sites. These sequence data were then analyzed for significant enrichment of transcription factor binding site sequences. P-values of significance represent the probability of obtaining an equal or greater number of matched binding site sequences using a randomly drawn sample of the same size as the input sequence set.

1.2.10 Inflammatory Cytokine Release

The protein levels of the inflammatory cytokine interleukin-8 (IL-8) were measured using the basolateral supernatant samples. A BD OptEIA^{TM} human IL-8 enzyme-linked immunosorbent assay (ELISA) was performed and analyzed according to the manufacturers' protocol (BD Biosciences, San Jose, California). Scanned absorbance reading outliers were identified, and statistical significance was calculated using the same method as the LDH analysis. IL-8 fold increase was calculated as $\mu_{\text{IL-8 exposed}} / \mu_{\text{IL-8 unexposed}}$, where μ represents mean IL-8 levels.

1.2.11 Quantitative RT-PCR Verification of mRNA Expression

Expression levels of selected genes were validated using quantitative real-time PCR (qRT-PCR). QuantiTect Primer Assays for glutamine-fructose-6-phosphate transaminase 2 (*GFPT2*) (Cat No. QT00007854), 2',5'-oligoadenylate synthetase 1 (*OAS1*) (Cat No.

00099134), and ATPase, class I, type 8B, member 1 (*ATP8B1*) (Cat No. QT00038094) were used with the Qiagen QuantiTect SYBR® Green PCR kit and Roche Lightcycler 480. Fold changes between exposed and unexposed samples were calculated based on cycle time values and normalized against a β-actin housekeeping gene. Statistical significance of the exposed versus unexposed transcript levels was calculated using an unpaired Student's t-test.

1.3 RESULTS

1.3.1 Chemical Analysis of Exposure Conditions

Human lung cells were exposed to either a primary air pollutant mixture or a PCA air pollutant mixture, containing secondary chemical products (see Materials and Methods). In order to assess the gas composition the cells were being exposed to, chemical concentrations were measured within the environmental irradiation chamber throughout the day (see Materials and Methods). Average nitrogen dioxide (NO_2), nitric oxide (NO), and ozone (O_3) concentrations for the primary pollutant exposure were measured at 0.21, 0.14, and 0.02 ppm, respectively (Fig. 1A). Average NO_2, NO, and O_3 concentrations for the PCA pollutant exposure were 0.12 , 0.01, and 0.14 ppm, respectively (Figure 1). Most hydrocarbon levels decreased throughout the exposure day (Figure 1), such as the aromatic compound, toluene (Additional File 1). More stable hydrocarbons, including n-hexane and benzene, remained at stable levels across the day (Additional File 1). Formaldehyde levels increased to 0.03 ppm by the end of the afternoon exposure period, while peroxyacetyl nitrate levels increased to 0.1 ppm (Figure 1).

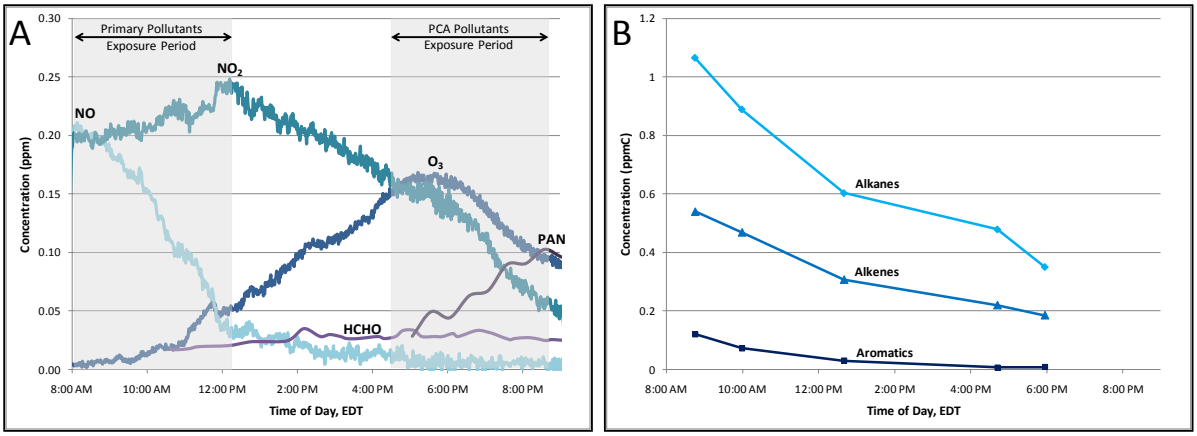

Figure 1: Chemical component analysis of exposure conditions. Figure (A) shows the concentrations for nitric oxide (NO), nitrogen dioxide (NO_2), ozone (O_3), formaldehyde (HCHO), and peroxyacetylnitrate (PAN) in parts per million (ppm). Exposure periods are shaded for the primary and photochemically altered air pollutant exposures. The decay of hydrocarbons throughout the day, in ppm carbon (ppmC) is shown (B), where chemicals are grouped according to species. EDT refers to Eastern Standard Time.

1.3.2 Cytotoxicity Measurements

After the exposures, lung cells were measured for cellular lactate dehydrogenase (LDH) release (see Materials and Methods). Investigation of the enzyme LDH showed that human lung cells experience more cell death after exposure to PCA pollutants (fold increase = 9, p-value < 0.001) than to primary pollutants (fold increase = 1.6, p-value = 0.07) (Figure 2).

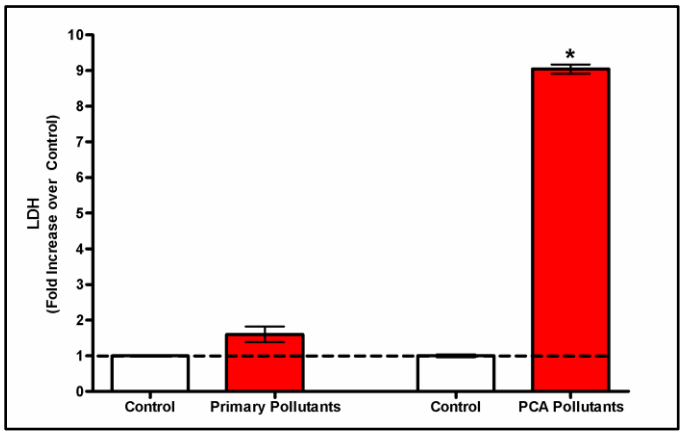

Figure 2: Levels of lactate dehydrogenase (LDH). LDH levels are shown for cells exposed to either primary or photochemically altered (PCA) pollutant mixtures relative to unexposed (control) cells. Results are displayed as fold increase over control +/- S.E.M. The (*) symbol indicates statistical significance (p-value < 0.05).

1.3.3 Gene Expression Analysis

Human lung cells were exposed to two different gaseous conditions: primary air pollutants in the morning, and PCA pollutants in the afternoon. Each exposure condition was performed alongside untreated, control samples. After exposure, mRNA was extracted and relative transcript abundance was assessed using Affymetrix Human Gene 1.0 ST microarrays (see Materials and Methods).

Lung cells exposed to the primary pollutant mixture showed differential expression in 19 genes, 15 of which showed increased expression levels and 4 of which showed decreased levels (Figure 3, Additional File 2). The PCA pollutant exposure resulted in changes in the expression levels of 709 genes. Of these, 190 showed increased expression levels and 519 showed decreased levels (Figure 3, Additional File 2). Among these two lists, 14 genes were identified as significantly differentially expressed in response to both pollutant mixtures (Table 1). Of this common set, 13 out of the 14 overlapping genes had higher expression fold change magnitudes due to PCA pollutant exposure when compared to the primary pollutant exposure.

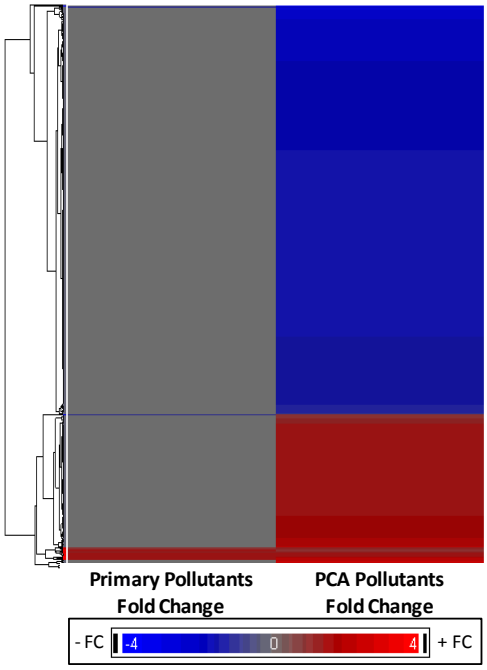

Figure 3: Heat map showing the average gene expression fold changes across 714 total genes modulated by primary and/or PCA pollutant mixture exposures. Red represents relative increase in expression, and blue represents relative decrease in expression.

Table 1: 14 overlapping genes with significantly (p-value < 0.05, q-value < 0.05) modified expression levels upon exposure to primary pollutants and photochemically altered (PCA) pollutant mixtures.

Gene Symbol	Primary Pollutants Fold Change (Exposed/Unexposed)	PCA Pollutants Fold Change (Exposed/Unexposed)
ACSM3	-1.50	-2.33
ACTA2	-1.58	-1.99
ATP8B1	-1.51	-2.95
CCL2	1.78	1.97
GCOM1	1.51	1.52
GFPT2	1.60	2.78
NFKBIA	1.51	1.53
OAS1	1.62	2.43
OR4A47	1.73	1.57
PAQR5	1.58	2.19
SLC5A3	1.75	1.80
TXNIP	-1.61	-2.42
USP17	1.53	1.57
USP17L2	1.53	1.53

1.3.4 Network Analysis and Biological Functions

In order to identify potential biological pathways in the lung cells that are affected by air pollutant exposure, the lists of differentially expressed genes were overlaid onto molecular network maps (see Materials and Methods). Networks containing the identified genes were algorithmically constructed based on connectivity and known relationships among proteins. The 19 genes with altered expression due to primary pollutant exposure generated one significant (p-value < 10^{-25}) network (Figure 4A). This network consists of 35 total proteins, nine of which are encoded by differentially expressed genes following primary pollutant exposure (Additional File 3). Within this network, there are gene products related to cancer, respiratory disease, and inflammation, such as chemokine (C-C motif) ligand 2 (*CCL2*).

The 709 genes altered in response to PCA pollutant mixture exposure generated 25 significant networks (p-values ranging from 10^{-12} to 10^{-52}) (Additional File 3). These 25 networks consist of 838 total proteins, 458 of which are encoded by differentially expressed

genes following PCA pollutant exposure (Additional File 3). Interestingly, a large interacting protein network (i.e. interactome) was identified containing 23 of the 25 networks modulated by PCA pollutants (p-value < 10^{-12}). This interactome illustrates multiple networks associated with air pollution interacting together (Figure 4B). Within this interactome, a highly significant (p-value < 10^{-52}) smaller network was identified (Figure 4C) containing proteins related to cancer, inflammation, and respiratory disease. Another network identified within the interactome shows a significant enrichment of proteins involved in inflammation linked to IL-8 and activator protein-1 (AP-1) signaling (p-value < 10^{-27}) (Figure 4D). This network also contains proteins related to cancer, inflammation, and respiratory disease.

To identify biological processes that may be influenced upon exposure to air pollutant mixtures, the constructed networks were queried for biological processes and disease signatures (see Materials and Methods). The network associated with primary pollutant exposure showed significant association with one disease signature, which was cancer (p-value = 3 x 10^{-6}) (Figure 5). PCA pollutant exposure, on the other hand, modified the expression of genes with protein products significantly (p-value < 3.0 x 10^{-6}) associated with ten different biological processes (Figure 5). These biological processes and signatures included cancer (p-value < 1.9 x 10^{-12}) and cellular growth and proliferation (p-value < 4.2 x 10^{-7}).

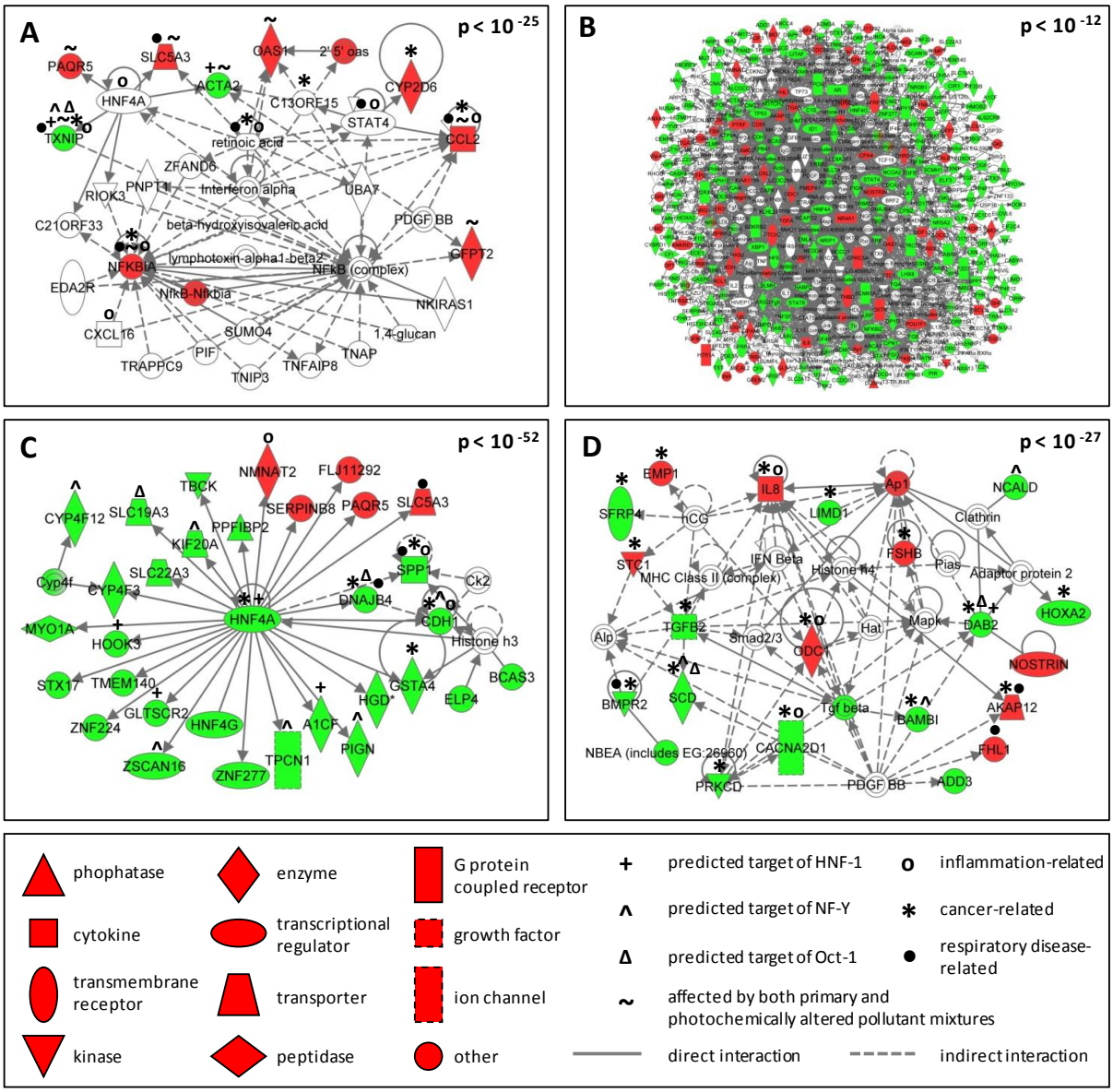

Figure 4: Molecular networks modulated by exposure to air pollution. Protein networks display (A) the network identified as associated with primary pollutant mixture exposure. A large interactome (B) displays multiple interacting networks associated with PCA pollutant exposure. Smaller, more focused networks of this interactome are shown for (C) a highly significant network, and (D) a significant network showing possible inflammatory response through IL-8 and AP-1 signaling. P-values are shown in the top right corners of each network. Networks are displayed with symbols representing protein products of genes that are up-regulated (red symbols), down-regulated (green symbols), or associated with the differentially expressed genes (clear symbols).

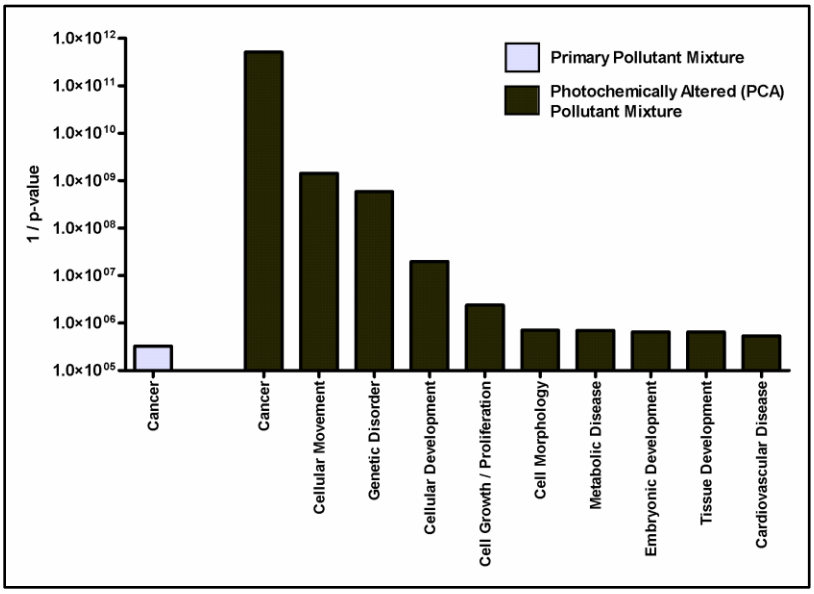

Figure 5: Biological functions significantly associated with primary and PCA pollutant exposure.

1.3.5 Predicted Transcription Factors

Transcription factor binding site analysis was performed to predict regulatory mechanisms that potentially underlie the gene expression modifications associated with the air pollution exposures (see Materials and Methods). For the primary pollutant mixture, analysis of the promoter regions of the 19 differentially expressed genes identified significant (p-value < 0.05) enrichment for binding sites of 17 transcription factors (Additional File 4). In the PCA pollutant mixture gene set, 53 total transcription factors with significantly (p-value < 0.05) enriched binding sites were predicted (Additional File 4). Comparison of the transcription factors predicted to control the responses to primary and PCA pollutant exposure revealed six common, overlapping transcription factors predicted to regulate pathway response to both exposure conditions (Table 2). All six of these transcription factors were associated with genes down-regulated by air pollutant exposure. The transcription factors with the most significant p-values, averaged across both exposure results, are hepatocyte nuclear factor 1 (HNF-1) (p-value = 0.003), nuclear transcription factor Y (NF-Y) (p-value = 0.005), and POU class 2 homeobox 1 (Oct-1) (p-value = 0.014). Networked genes associated with air pollution exposure that were identified as predicted targets of the these three transcription factors are shown in Figure 4, and all predicted targets are detailed in Additional File 4.

Table 2: Transcription factors predicted to regulate genes modified by exposure to primary pollutants and genes modified by exposure to photochemically altered pollutant mixtures.

Transcription Factor	TRANSFAC Accession Number	Average p-value
HNF-1	M00132	0.003
NF-Y	M00287	0.005
Oct-1	M00137	0.014
GATA-1	M00127	0.017
FOXO4	M00472	0.022
Evi-1	M00078	0.039

1.3.6 Inflammatory Cytokine IL-8 Release

After exposure to the gaseous pollutant mixtures, the release of inflammatory response protein interleukin-8 (IL-8) was assessed in human lung cells (see Materials and Methods). Analysis revealed that lung cells exposed to primary pollutants released an insignificant (p-value = 0.50) change in IL-8 levels, with an average fold increase of 1.14 compared to unexposed cells. Cells exposed to PCA pollutants showed a significant (p-value < 0.001) increase in IL-8 levels, with an average fold increase of 3.79 (Figure 6).

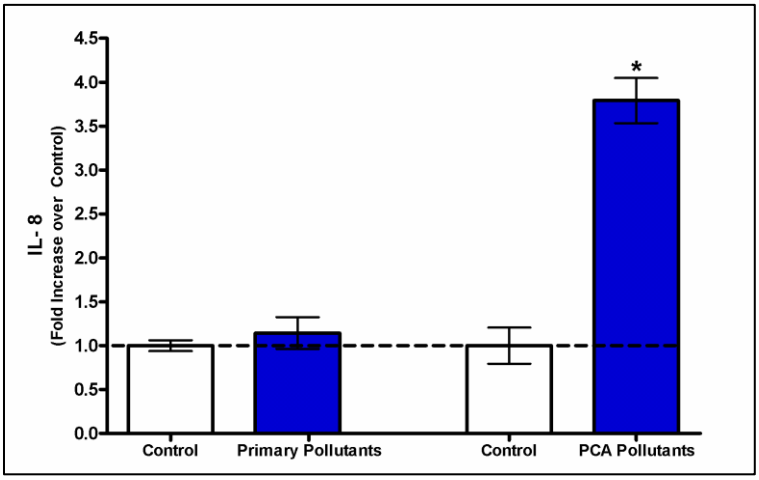

Figure 6: Levels of interkeukin-8 (IL-8) release. Levels of IL-8 are shown for cells exposed to either primary or photochemically altered (PCA) pollutant mixtures relative to unexposed (control) cells. Results are displayed as fold increase over control +/- S.E.M. The star symbol (*) indicates statistical significance (p-value < 0.05).

1.3.7 Validation of Expression Changes through qRT-PCR

Quantitative real-time RT-PCR (qRT-PCR) was used to validate the expression levels of three genes across both exposures. From microarray analysis, *GFPT2* and *OAS1* were significantly up-regulated in expression, while *ATP8B1* was significantly down-regulated due to both primary and PCA pollutant exposures (Table 1). PCR analysis confirmed significant (p-value < 0.05) up-regulation due to air pollutant exposure in *GFPT2* (Fold Change (FC) 1.85 primary pollutants, FC 4.52 PCA pollutants) and *OAS1* (FC 1.95 primary, FC 3.54 PCA) (Figure 7). PCR analysis also confirmed significant (p-value < 0.05) down-regulation of *ATP8B1* (FC -1.46 primary, FC -3.12 PCA). Furthermore, PCR analysis confirmed that these genes show higher magnitudes of expression fold changes after exposure to PCA pollutants in comparison to primary pollutants (Figure 7).

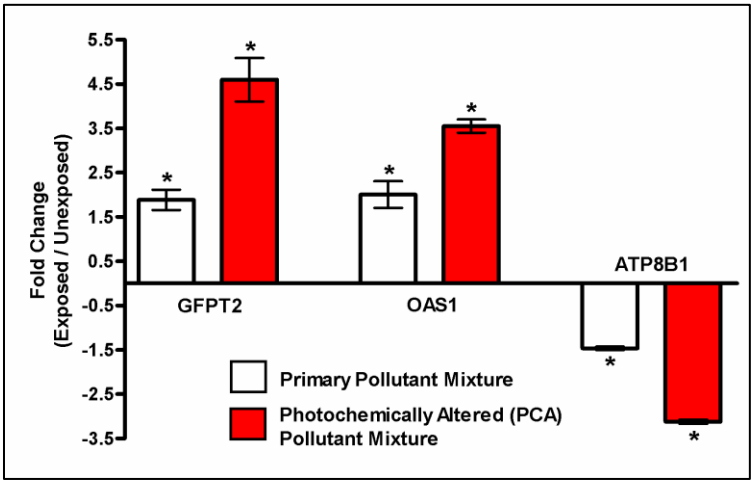

Figure 7: Quantitative real-time PCR results. Results are displayed as fold change in transcript level +/- S.E.M. The star symbol (*) indicates statistical significance (p-value < 0.05).

1.4 DISCUSSION

In this study, we compared the genome-wide response of human lung epithelial cells exposed to either primary air pollutant mixtures or PCA pollutant mixtures. The primary pollutant mixture includes compounds humans inhale during the morning, while the PCA pollutant mixture contains compounds humans are exposed to during the afternoon in outdoor urban environments (Jeffries and Sexton, 1995; Sexton, *et al.*, 2004). This study simulates real-life conditions using an *in vitro* exposure system that physically mimics *in vivo* human lung gas exposures (Bakand *et al.*, 2005). Furthermore, chemical component analysis verifies that the pollution chemistry analyzed within the outdoor environmental irradiation chamber is similar to urban air photochemistry and exposure conditions for humans living in cities (Jeffries and Sexton, 1995). More specifically, primary air pollutants, including oxides of nitrogen (NO_x) and multiple volatile organic compounds (VOCs), showed decreasing concentrations throughout the experiment day. At the same time, chemical reactions between NO_x, VOCs, and sunlight generated secondary products, including ozone, formaldehyde, and peroxyacetyl nitrate.

We find through microarray analysis that exposure to primary pollutants alters the expression of only 19 genes, while exposure to PCA pollutants significantly alters the expression of 709 genes. While there was differential cell survival between the two exposures, the changes in gene expression are unlikely to be the result of cytotoxicity (Jelinsky *et al.*, 2000). Further validating this claim, a study using human lung A549 cells shows that the difference in the number of genes differentially expressed upon exposure to air mixtures, at varying dilutions, is not proportional to changes in cell viability (Tsukue *et al.*, 2010).

To identify molecular pathways that may be influenced by exposure to air pollutants, the differentially expressed genes were integrated with their protein products and queried for known interactions to construct associated molecular networks. Again, a significant difference was seen between the number of altered networks associated with primary and PCA pollutant exposure. Specifically, exposure to the primary pollutant mixture was associated with the modulation of one molecular network, whereas PCA pollutant exposure

resulted in the modification of a massive interactome containing multiple overlapping networks. Altogether, 25 networks were identified as associated with PCA pollutant exposure. These results suggest that exposure to secondary products created through photochemical reactions involving VOCs and NO_x induce more substantial changes in lung cell signaling than primary emitted air pollutants.

As revealed through network analysis, genes associated with inflammatory response pathways showed modulation in human lung cells exposed to air pollution. For example, *CCL2*, *IL-8*, and *AP-1* expression levels were all increased, illustrating potential increases in molecular signaling within their associated pathways. Interestingly, *CCL2* gene products can signal for the accumulation of monocytes and dendritic cells, which promote the build-up of inflammatory microenvironments that can lead to possible tumor progression (Mantovani *et al.*, 2008). To note, increased *IL-8* expression is a biomarker of air pollutant-induced lung inflammation (Damera, *et al.*, 2009; Jaspers, *et al.*, 1997). Here the up-regulation of *IL-8* was verified at the protein level, where IL-8 protein abundance also showed significantly increased levels due to PCA exposure. A similar study also shows increased IL-8 levels in lung cells exposed to air pollutants undergoing chemical reactions under sunlight (Sexton, *et al.*, 2004). AP-1 is shown through network analysis as possibly regulating the altered IL-8 inflammatory response pathway. This finding coincides with a previous study showing that AP-1 activation increases *IL-8* transcription in human A549 lung epithelial cells (Adam *et al.*, 2006). Furthermore, ozone, one of the secondary products within the PCA pollutant mixture, has been shown to activate AP-1 in A549 lung cells, which potentially regulates ozone-induced IL-8 release (Jaspers, *et al.*, 1997). From our results, we show that air pollution exposures, such as those in urban environments, may influence lung cell function by altering the levels of gene transcripts and proteins associated with inflammatory response pathways.

Using a biological process enrichment analysis, the networks identified as modulated by air pollutant exposure were queried for known involvement with biological processes and disease signatures. Here, the primary pollutant exposure modified the expression levels of genes that encode protein products significantly related to one function, cancer. On the other

hand, exposure to PCA pollutants significantly modified the expression levels of genes that encode protein products associated with a large number of biological processes. Specifically, ten biological processes were identified as modified in lung cells exposed to PCA pollutants. These processes include cancer and cellular growth/ profileration. The increased number of biological processes affected by PCA pollutants further confirms our finding that PCA pollutants induce a more robust response in gene expression patterns and their associated biological functions than primary pollutants.

In order to analyze potential regulatory mechanisms underlying changes in gene expression resulting from air pollutant exposures, putative mediating transcription factors were computationally predicted. Six common transcription factors were identified as potential regulators of the observed gene expression changes in response to both primary and PCA pollutant mixture exposures. These transcription factors include Oct-1, FOXO4, and HNF-1, which were all predicted to regulate genes with decreased expression after exposure. Oct-1 is a transcription factor associated with cancer malignancy, and loss of Oct-1 can cause cells to become hypersensitive to genotoxic and oxidative stress agents (Kang *et al.*, 2009). Another transcription factor, FOXO4, is related to tumorigenesis with a role in apoptosis (Myatt and Lam, 2007). The most significant of these six transcription factors, HNF-1, has been primarily studies for mutations which are linked to diabetes susceptibility (Ellard and Colclough, 2006), hepatocellular carinoma (Bluteau *et al.*, 2002), and inflammatory pathways underlying coronary heart disease (Armendariz and Krauss, 2009). HNF-1 is also known to control *HNF-4α* transcription (Hatzis and Talianidis, 2001). HNF-4α is one of the highlighted molecules present in the networks constructed using the proteins encoded by genes with altered expression in primary and PCA pollutant-exposed lung cells. This finding connects our transcription factor binding site analysis with the network analysis. Interestingly, there is limited knowledge on how HNF-1 is specifically involved in lung cell function. With our results, we show that promoter sites of genes associated with air pollutant exposures are potentially enriched for transcription factors involved in cell cycle regulation, cancer, and cellular stress.

Two exposure atmospheres were used in our study, where each atmosphere contained a very large number of chemical compounds. When inhaled, these compounds are known to influence lung function (e.g. ozone) (EPA, 2006) and cause cancer (e.g. benzene, formaldehyde, 1,3-butadiene) (IARC, 2010). Due to the complexity of our exposure atmospheres, it is difficult to extrapolate which chemicals within the air mixtures are contributing the most to the observed changes in gene expression and molecular pathways. Future research will investigate the effects of individual air pollutants on lung cells, and compare these responses against air mixtures.

1.5 CONCLUSION

In conclusion, our study reveals potential mechanisms that may underlie health effects induced by primary emitted air pollutants and PCA pollutant mixtures. In this genome-wide comparison study, we find there is a significantly more robust response in lung cells exposed to PCA pollutants in comparison to primary pollutants. Specifically, there were 37-fold more genes dysregulated by exposure to PCA pollutants. Mapping the genes affected by pollutant exposure to their encoded protein products and analyzing their biological functions reveals the association of air pollution exposure to cellular stress, inflammation, and cancer pathways in human lung cells. Future research will investigate the differences between lung cell response to air pollutant mixtures and individual components within air mixtures.

CHAPTER 2
Epigenetic Changes Induced by Air Toxics: Formaldehyde Exposure Alters miRNA Expression Profiles in Human Lung Cells

ABSTRACT

Background: Exposure to formaldehyde, a known air toxic, is associated with cancer and lung disease. Despite its adverse health effects, the mechanisms underlying formaldehyde-induced disease remain largely unknown. Research investigations have uncovered microRNAs (miRNAs) as key post-transcriptional regulators of gene expression that may influence cellular disease state. While studies have compared different miRNA expression patterns between diseased and healthy tissue, this is the first study to examine perturbations in global miRNA levels resulting from formaldehyde exposure.

Objectives: We set out to investigate whether cellular miRNA expression profiles are modified by formaldehyde exposure in human lung cells. We hypothesized that formaldehyde exposure disrupts miRNA expression levels within lung cells, representing a novel epigenetic mechanism through which formaldehyde may induce disease.

Methods: Human lung cells were grown at air-liquid interface and exposed to gaseous formaldehyde at 1 ppm for 4 hours. Small RNAs and protein were collected and analyzed for miRNA expression using microarray analysis or IL-8 protein levels by ELISA, respectively.

Results: Gaseous formaldehyde exposure altered the miRNA expression profiles in human lung cells. Specifically, 89 miRNAs were significantly down-regulated in formaldehyde exposed samples versus controls. Functional and molecular network analysis of the predicted miRNA transcript targets revealed that formaldehyde exposure potentially alters signaling pathways associated with cancer, inflammatory response, and endocrine system regulation. IL-8 release was increased in cells exposed to formaldehyde, and results were confirmed by Real Time PCR.

Conclusions: Formaldehyde alters miRNA patterns which regulate gene expression, potentially leading to the initiation of a variety of diseases.

2.1 INTRODUCTION

Current indoor and outdoor air quality contributes significantly to global increases in morbidity and mortality (Brunekreef and Holgate, 2002; Burnett, *et al.*, 2001; Smith and Mehta, 2003). Epidemiological studies have shown that formaldehyde, a known air toxic, causes increased risk of childhood and adult asthma (Rumchev *et al.*, 2002; Wieslander *et al.*, 1997), acute respiratory tract illness (Smith *et al.*, 2000; Tuthill, 1984), nasopharyngeal cancer (Vaughan *et al.*, 2000), and possibly leukemia (Zhang *et al.*, 2009). In animal studies, strong links have been made between formaldehyde exposure and nasal carcinoma (Kerns *et al.*, 1983). Furthermore, the International Agency for Research on Cancer (IARC) has classified formaldehyde as a known human carcinogen (IARC, 2006).

2.1.1 Formaldehyde Exposure Sources

In outdoor environments, formaldehyde is present due to direct emissions from anthropogenic and biogenic sources, and is also formed as a secondary chemical product through hydrocarbon atmospheric chemistry (WHO, 2001). Anthropogenic sources of formaldehyde include automobile exhaust, power plants, manufacturing facilities, and incinerators (EPA, 2007; WHO, 2001). Ambient air is estimated to contain formaldehyde at levels between 0.0008 and 0.02 ppm (WHO, 2001). Higher formaldehyde exposure occurs within indoor environments, where humans inhale levels estimated between 0.02 and 0.3 ppm, depending on the presence of tobacco smoke (WHO, 2001). The highest formaldehyde levels are found in certain occupational environments, such as industries related to resin, plastics, wood, paper, insulation, textile, and chemical productions, as well as medical institutions using disinfectants and embalming products. In these high exposure cases, occupational workers are exposed to, on average, approximately 0.74 ppm (WHO, 2001).

2.1.2 Formaldehyde Dosimetry

As formaldehyde is highly reactive and water soluble, more than 95% of inhaled formaldehyde is predicted to be absorbed within the human respiratory tract (Overton *et al.*, 2001). While most inhaled formaldehyde is absorbed in the nasal and upper airways

(Overton, *et al.*, 2001), much remains uncertain about the dosimetry and mechanisms underlying pulmonary responses to formaldehyde (Thompson *et al.*, 2008). As a result, it is important to study the effects of gas exposure to the lower respiratory region. This has been studied recently where the effects of formaldehyde exposure on the DNA-damage protection of human A549 alveolar epithelial cells were established (Speit *et al.*, 2010).

2.1.3 Gene-Transcript Regulation through miRNAs

Previous research has shown that altered gene expression patterns exist in nasal and lung cells exposed to formaldehyde (Kim *et al.*, 2002; Li *et al.*, 2007). These changes in gene transcript profiles, which likely translate to changes in protein levels, could arise from altered microRNA expression. MicroRNAs (miRNAs) are an abundant class of regulatory molecules that have received scientific attention for their ability to alter mRNA abundance. miRNAs are non-coding single stranded RNA molecules approximately 22 nucleotides in length (Bartel, 2004). One of the more established functions of miRNAs is translational inhibition of target messenger RNA (mRNA) molecules. Translational inhibition occurs when miRNAs base pair with 3'-untranslated regions (UTRs) of mRNAs causing decreases in protein production (Filipowicz *et al.*, 2008). Once paired with mRNAs, miRNAs can destabilize mRNAs and target their degradation through deadenylation (Filipowicz, *et al.*, 2008; O'Hara *et al.*, 2009). The study of the dysregulation and modification of miRNA abundance has revealed miRNAs' important roles in many diseases, including heart failure, hematological malignancies, and neurodegenerative disease (Divakaran and Mann, 2008; Fabbri *et al.*, 2008; Hébert and De Strooper, 2009). In addition, studies have shown that miRNA expression profiling in tumor cells can aid in classifying cancer type, cancer state, and cellular response to treatment (Calin and Croce, 2006; Lu *et al.*, 2005; Ma *et al.*, 2007).

Mammalian miRNAs are estimated to regulate 30% of all protein-coding genes through post-transcriptional modification (Filipowicz, *et al.*, 2008). Because miRNAs play such a pivotal role in human cellular gene regulation, more research is needed to understand the effects of environmental exposures on miRNA levels. To our knowledge, only one other study has investigated the effects of environmental air pollution on cellular miRNA abundance. In that

study, diesel exhaust particles were shown to affect miRNA expression related to inflammatory response pathways and tumorigenesis (Jardim *et al.*, 2009).

2.1.4 Study Objectives

In this study, we hypothesized that formaldehyde exposure can disrupt miRNA levels within lung cells. We tested this hypothesis by exposing human lung epithelial cells to formaldehyde using a direct air-liquid interface that physically mimics the human respiratory tract. Using microarray analysis, we assessed the expression levels of more than 500 known miRNAs. We demonstrate that formaldehyde significantly alters the expression profiles of 89 miRNAs which are predicted to target mRNAs associated with numerous biological pathways related to cancer, inflammatory response, and endocrine system regulation. Confirming one of the most dysregulated inflammation-associated pathways, IL-8 showed significantly increased protein expression levels in formaldehyde exposed cells. Taken together, this research suggests a novel epigenetic mechanism by which formaldehyde may induce disease.

2.2 MATERIALS AND METHODS

2.2.1 Cell Culture

Human A549 type II lung epithelial cells derived from a human lung adenocarcinoma were cultured according to standard protocol (ATCC). Cells were grown in growth media containing F-12K plus 10% FBS plus 1% penicillin and streptomycin. Cells were plated onto 24 mm diameter collagen-coated membranes with 0.4 μM pores (Trans-CLR; Costar, Cambridge, MA). Upon confluence, cells were cultured in phenol red-free F-12K nutrient mixture without FBS. Immediately prior to exposure, media above each membrane was

aspirated in order to create direct air-liquid interface culture conditions. The media beneath each membrane remained to supply nourishment for cells throughout the exposure.

2.2.2 Formaldehyde Treatment

Gaseous formaldehyde was generated by heating 143 mg paraformaldehyde (Aldrich Chemical Company, Inc., Milwaukee Wi, lot no. 05910EI) in an air-flushed "U-tube" until the powder was completely vaporized within a dark un-irradiated 120 m^3 environmental chamber. The walls of the chamber are made of chemically non-reactive film, as detailed previously (Sexton, *et al.*, 2004). The chamber was naturally humidified from pre-flushing with HEPA filtered ambient air during cloudy conditions. This resulted in a formaldehyde concentration of 1 ppm (1.2 mg/m^3) which was then drawn through a cellular exposure chamber (Billups-Rothenberg, Modular Incubator Chamber, Del Mar, CA) at 1.0 L/min. The exposure chamber was positioned within an incubator where CO_2 was added to the formaldehyde exposure source stream at 0.05 L/min and a small water dish provided proper humidification. Prepared lung cells were exposed to 1 ppm formaldehyde for 4 hours, while mock-treated control cells were exposed to humidified air under similar conditions. Experiments were carried out with six technical replicates for each exposure condition, generating a total of 12 samples. After nine hours, cells were scraped and stored at -80°C in TRIzol® Reagent (Invitrogen Life Technologies), and basolateral supernatants were aspirated and stored at -80°C.

2.2.3 Cytotoxicity Analysis

To measure formaldehyde exposure's cytotoxicity, the enzyme lactate dehydrogenase (LDH) was measured within the supernatant of each sample. Measurements were acquired using a coupled enzymatic assay (Takara), according to the supplier's instructions (Takara Bio Inc., Japan). LDH fold increase was calculated as $\mu_{LDH, FE} / \mu_{LDH\ C}$, where μ represents the mean LDH activity, FE represents formaldehyde exposed samples, and C represents controls.

2.2.4 Microarray Processing

RNA molecules of at least 18 nucleotides in length were isolated using Qiagen's miRNeasy® Kit according to the manufacturer's protocol (Qiagen, Valencia CA). RNA was quantified with the NanoDrop™ 1000 Spectrophotometer (Thermo Scientific, Waltham MA) and its integrity was verified with an Agilent Technologies 2100 Bioanalyzer (Santa Clara, CA). RNA was labeled and hybridized to the human miRNA microarray (version 1) manufactured by Agilent Technologies (Santa Clara, CA). This microarray measures expression levels of 534 human miRNAs. Three of the six total samples from each exposure condition, three formaldehyde-exposed and three mock-treated samples, were hybridized using 400 ng of input RNA per sample. RNA labeling and hybridization were performed according to the manufacturer's protocol, and microarray results were extracted using Agilent Feature Extraction Software. Data were submitted to NCBI's Gene Expression Omnibus (GEO) database (http://www.ncbi.nlm.nih.gov/geo/) and are available under accession #GSE22365 (Edgar, *et al.*, 2002).

2.2.5 Microarray Analysis

The resulting expression levels for each of the miRNAs measured by the microarrays were calculated and used to filter for miRNAs expressed above a background level (background was set at 30, approximating the median signal per array). This resulted in a reduction of probesets from 12033 to 4900 records. Differential miRNA expression was defined as a significant difference in miRNA expression levels between treated samples and untreated samples, where the following three statistical requirements were set: (1) fold change of ≥ 1.5 or ≤ - 1.5 (treated versus untreated); (2) p-value < 0.005; and (3) false discovery rate (FDR) < 0.005. P-values and FDRs were generated using the Comparative Marker Selection tool in GenePattern (www.broadinstitute.org/cancer/software/genepattern/) (Reich *et al.*, 2006). Here, 2000 permutation tests were carried out using the signal-to-noise (SNR) ratio analysis and smoothed p-values were determined for each miRNA. SNR is defined by the equation $SNR = (\mu_A - \mu_B) / (\sigma_A + \sigma_B)$, where μ represents average sample intensity and σ represents standard deviation (Golub *et al.*, 1999). SNRs have been shown to provide one of the most accurate classification prediction methods (Cho and Ryu, 2002). False discovery rates (FDRs) were calculated as the expected fraction of false positives among probesets reported

as significant using the Benjamini and Hochberg procedure (Benjamini and Hochberg, 1995). Targets for the most differentially expressed miRNAs were identified using miRDB (www.mirdb.org) (Wang, 2008; Wang and El Naqa, 2008) where targets with a score of >70 were investigated.

2.2.6 Enriched Biological Functions and Network Analysis

Enriched biological functions and molecular network analyses were performed using the Ingenuity database (Ingenuity Systems, www.ingenuity.com, Redwood City, CA). The Ingenuity database provides a collection of gene to phenotype associations, molecular interactions, regulatory events, and chemical knowledge accumulated to develop a global molecular network. The lists of putative targets for each miRNA were overlaid onto this global molecular network, where protein networks significantly associated with the targets were algorithmically constructed based on connectivity. Associated enriched canonical pathways within these networks were also identified. Functional analysis was carried out to identify biological functions and disease signatures most significantly associated with the input targets. Statistical significance of each biological function or disease was calculated using a Fischer's exact test. This test generated a p-value signifying the probability that each function or disease was associated with the miRNA targets by chance alone. Only enriched functions with p-values < 0.005 were assessed.

2.2.7 Quantitative RT-PCR Verification of miRNA Expression

Expression levels of the five most significantly modified miRNAs were also tested using quantitative real-time PCR. The TaqMan® MicroRNA Primer Assays for hsa-miR-33 (ID 002135), hsa-miR-450 (ID 2303), hsa-miR-330 (ID 000544), hsa-miR-181a (ID 000516), and hsa-miR-10b (ID 002218) were used in conjunction with the TaqMan® Small RNA Assays PCR kit (Applied Biosystems). The Bio-Rad MyCycler Thermal Cyler was used for the reverse transcription step, and the Roche Lightcycler 480 was used for the real-time step. The same three control and three formaldehyde exposed samples from the microarray were used for qRT-PCR, which was performed in technical duplicate. Statistical significance was evaluated using a t-test.

2.2.8 Interleukin-8 Measurement

The protein abundance of the cytokine interleukin-8 (IL-8) was measured using the basolateral supernatant from all 12 samples. A BD OptEIA^{TM} human IL-8 enzyme-linked immunosorbent assay (ELISA) was performed and analyzed according to the manufacturers' protocol (BD Biosciences, San Jose, California). Experiments were carried out with 12 technical replicates for each exposure condition. Scanned absorbance reading outliers were identified through the Grubbs' test (www.graphpad.com) where outliers were identified as those with less than a 5% probability of occurring as an outlier by chance alone, as based off a normal distribution (Grubbs, 1969). IL-8 fold increase was calculated as $\mu_{IL\text{-}8\ FE} / \mu_{IL\text{-}8\ C}$, where μ represents the mean, FE represents formaldehyde exposed samples, and C represents controls. Statistical significance of the treated versus untreated IL-8 levels was calculated using a t-test with Welch's correction.

2.3 RESULTS

2.3.1 Formaldehyde Exposure Modulates miRNAs in Human Lung Cells

In this study, we set out to identify whether formaldehyde exposure alters the expression levels of miRNAs in lung cells. Human lung epithelial cells (A549) were exposed to gaseous formaldehyde drawn directly from an un-irradiated (dark) environmental chamber into an exposure chamber or were mock-treated (see Materials and Methods). This exposure resulted in a 6.68 fold increase in LDH release in the formaldehyde treated cells. After exposure, small RNAs were collected and their relative abundance measured using microarrays. A total of 343 unique miRNAs were detectable above background in these cells. The 343 miRNAs were further assessed for formaldehyde-induced changes in expression level (see Materials and Methods). A total of 89 miRNAs showed a significant decrease in expression in the formaldehyde exposed lung samples compared to control samples (Figure 8, Additional File 5). There were no miRNAs identified with significantly increased expression levels in

response to formaldehyde. The five most significantly differentially expressed miRNAs, as determined through microarray analysis, were miR-33 (FC -5.48), miR-450 (FC -3.57), miR-330 (FC -2.43), miR-181a (FC -2.11), and miR-10b (FC -2.11).

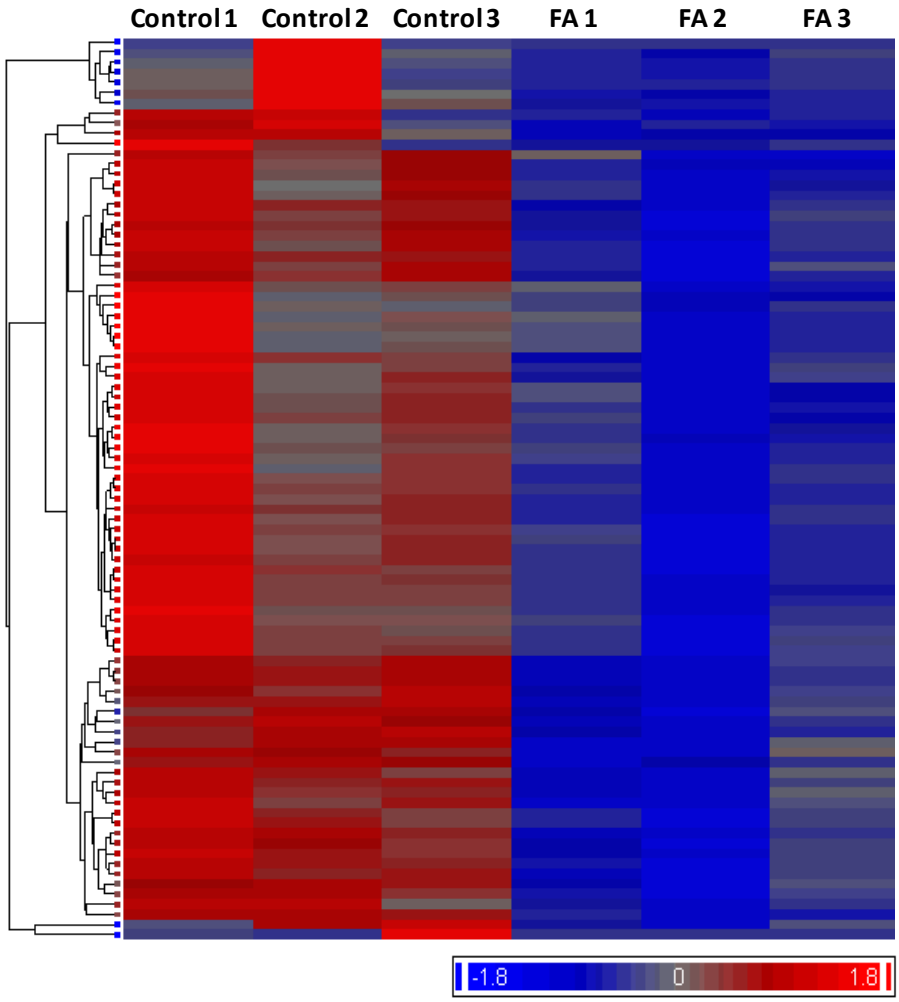

Figure 8: Heat map of 89 formaldehyde-modulated miRNAs. Data were mean standardized and hierarchical clustering was performed. Blue indicates relative low expression while red indicates relative high expression. Formaldehyde-treated samples are abbreviated as FA.

2.3.2 MicroRNA Expression Changes are Validated through qRT-PCR

Quantitative real-time RT-PCR was used to confirm the findings of the array-based results. The qRT-PCR validated the findings of the decreased miRNA expression induced by formaldehyde exposure. Specifically, miR-330 showed a fold change (FC) of -1.32, and miR-

181a showed a FC of -7.39 (Figure 9). Likewise, miR-33 displayed a FC of -1.2 and miR-10b displayed a FC of -1.48. miR-450 showed minimal expression changes with a FC of -1.04. Because it could not be validated, further analysis on miR-450 was not performed. To assess the similarity of the array-based quantification of the miRNA expression changes and the qRT-PCR, the relative miRNA abundances were compared against the raw microarray expression data. This analysis shows a high correlation (0.8) between miRNA abundance measured with both qRT-PCR and microarray (Figure 9). More specifically, these analyses support that the direction of differential expression of the miRNAs induced by exposure to formaldehyde was consistent between the qRT-PCR and microarray analyses. It is important to note, however, that there is a difference in the magnitude of expression change with the microarray results generally greater than those obtained with the qRT-PCR.

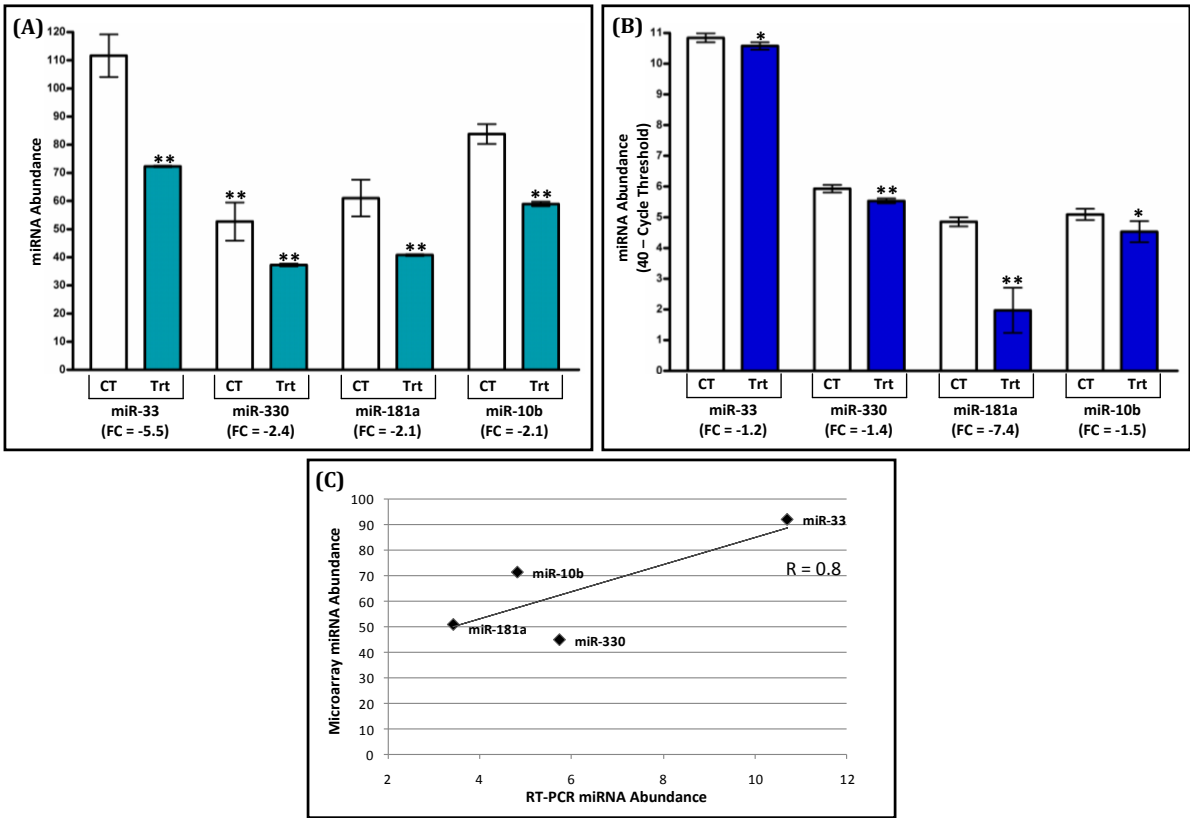

Figure 9: Comparison of Microarray and qRT-PCR Results. miRNA microarray results (A) are displayed as miRNA abundance obtained from raw microarray data. miRNA qRT-PCR results (B) are displayed in terms of miRNA abundance relative to 40, the maximum cycle threshold. Star symbols represent significance, where one star represents p-value < 0.1, and two stars represent p-value < 0.05. Each column represents either control samples (CT) or treated samples (Trt). Average fold changes (FC) are shown, and error bars represent S.E.M. Correlation (C) between miRNA abundance measured by microarrays and RT-PCR is illustrated.

2.3.3 miRNA Targets are Integrated into Biological Networks

In order to identify potential biological pathways affected by formaldehyde exposure, the 89 miRNAs that showed significant changes in expression levels were ranked according to their fold-changes, p-values, and qRT-PCR results (Additional File 5). Here, the four miRNAs with the most significant formaldehyde-induced changes in expression were further investigated: miR-33, miR-330, miR-181a, and miR-10b. For each of these four miRNAs, we identified their putative mRNA targets (see Materials and Methods). Using a stringent cutoff of a match score between each miRNA and its mRNA targets followed by analysis for unique mRNAs per target list (see Materials and Methods), we identified a total of 67 targets of miR-33, 217 targets of miR-330, 334 targets of miR-181a, and 25 targets of miR-10b (Additional File 6). Among this list of 647 mRNAs, there are 42 that are common to at least two of the modulated miRNAs (Additional File 6).

Once the predicted mRNA targets were identified for the most significant miRNAs, they were overlaid onto molecular pathway maps enabled through the Ingenuity Pathways' Knowledge Base (see Materials and Methods). Networks containing miRNA targets were algorithmically constructed based on connectivity and the known relationships among proteins. The predicted targets of the four modulated miRNAs (miR-33, miR-330, miR-181a, and miR-10b) resulted in the generation of a total of 40 networks (

Additional File **7**). For each of the miRNA gene targets, the most significant (p-values range from 10^{-23} to 10^{-43}) network has been highlighted for further evaluation (Figure 10). The proteins identified within these networks were queried for their enrichment for various canonical pathways (see Materials and Methods). A comparison of the canonical pathways highlighted the conservation of a cancer-associated pathway common to all four miRNA-generated networks (Additional File 8). Overlaying the pathway information onto the most significant networks resulted in the identification of enrichment for the nuclear factor kappa beta (NFκB) pathway and the interleukin-8 (IL-8) signaling pathway, among others (Figure 10, Additional File 8).

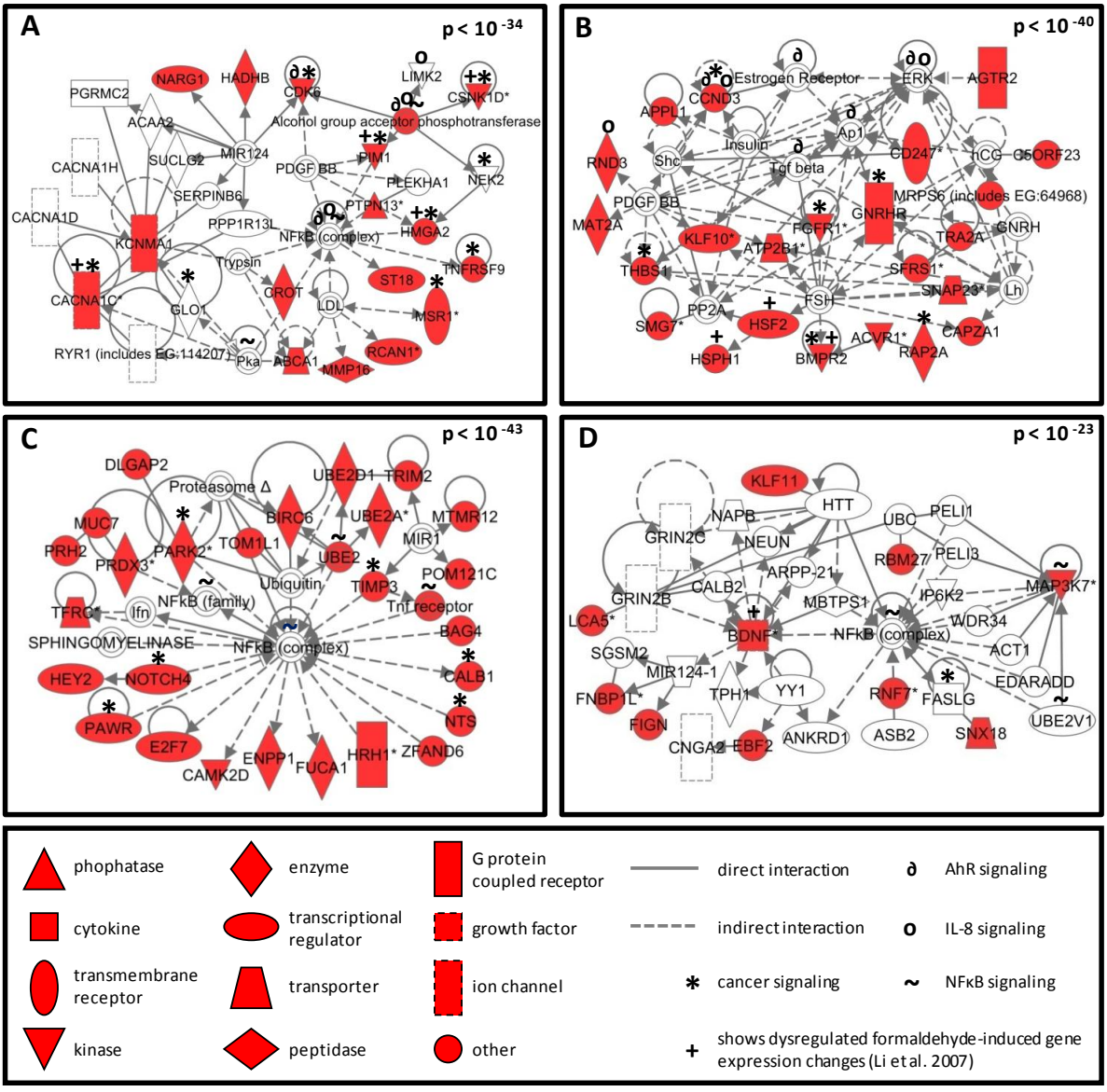

Figure 10: Most significant molecular networks of modulated miRNAs affected by formaldehyde exposure. Protein networks display interactions using the mRNA targets of (A) miR-33, (B) miR-330, (C) miR-181a, and (D) miR-10b. P-values representing the probability of these interactions occurring by chance are shown in the top right corners of each network. Networks are displayed with symbols representing predicted miRNA targets (red symbols) or proteins associated with the predicted targets (clear symbols).

Using a biological process enrichment analysis, the 40 networks encoded by the mRNA targets for each miRNA were queried for biological processes that were most significantly modulated by formaldehyde exposure. A total of 71 unique biological processes were found (Additional File 9). Across the mRNA targets, common enrichment was found for 13 different cellular biological processes. These processes included inflammatory response (p-value < 0.0029) and endocrine system development/function (p-value < 0.0018) which were enriched within the targets of all four miRNAs (Table 3).

Table 3: Biological functions significantly associated with all predicted target sets of miR-33, miR-330, miR-181a, and miR-10b.

Enriched Functions	Average p-value
Cellular Development	0.0011
Small Molecule Biochemistry	0.0015
Nervous System Development and Function	0.0016
Cell-To-Cell Signaling and Interaction	0.0017
Cell Morphology	0.0017
Tissue Development	0.0017
Cellular Function and Maintenance	0.0017
Cellular Movement	0.0017
Endocrine System Development and Function	0.0018
Gene Expression	0.0018
Cellular Growth and Proliferation	0.0021
Inflammatory Response	0.0029
Hematological System Development and Function	0.0033

2.3.4 Conservation of Predicted and Observed mRNA Targets

In our analysis, we used a stringent computational metric to match miRNAs to their predicted mRNA targets to better understand the biological implications of formaldehyde exposure. As

these mRNA targets were computationally predicted, we also compared our results with those of an existing genomic database established from a study that analyzed human tracheal fibroblast cells exposed to formaldehyde (Li, *et al.*, 2007). In this comparison, we found overlap between the predicted mRNA targets of the formaldehyde-modulated miRNAs and the tested formaldehyde-responsive genes identified (Li, *et al.*, 2007). Specifically, brain-derived neurotrophic factor (*BDNF*), bone morphogenetic protein receptor, type II (serine/threonine kinase) (*BMPR2*), calcium channel voltage-dependent L type, alpha 1C subunit (*CACNA1C*), casein kinase 1 delta (*CSNK1D*), high mobility group AT-hook 2 (*HMGA2*), heat shock transcription factor 2 (*HSF2*), heat shock 105kDa/110kDa protein 1 (*HSPH1*), and Pim-1 oncogene (*PIM1*), are found within the four most significant networks associated with the identified miRNA targets (Figure 10).

We expanded our comparison by performing network analysis on the formaldehyde-associated genes identified by Li et. al (2007). Here, networks were constructed and related biological functions were identified, as done with the miRNA predicted target network analysis (see Materials and Methods). Networks related to cancer (p-value = 1.9×10^{-19}), inflammation (p-value = 1.1×10^{-8}), and endocrine system disorders (p-value = 3.15×10^{-4}) were generated (Additional File 10).

2.3.5 Inflammatory Cytokine IL-8 is Released in Response to Formaldehyde

Based on our findings from the canonical pathway and biological process enrichment analyses that showed the IL-8 pathway as potentially dysregulated by formaldehyde exposure, we set out to confirm whether IL-8 protein levels may be influenced by such exposure. After cells were exposed to formaldehyde, IL-8 protein release was assessed (see Materials and Methods). The investigation of the inflammatory response protein interleukin-8 (IL-8) showed that human lung cells activate an inflammatory response after exposure to formaldehyde. Specifically, an average 16.9 fold increase (p-value < 0.05) in cytokine release was observed in formaldehyde exposed cells relative to control samples (Figure 11).

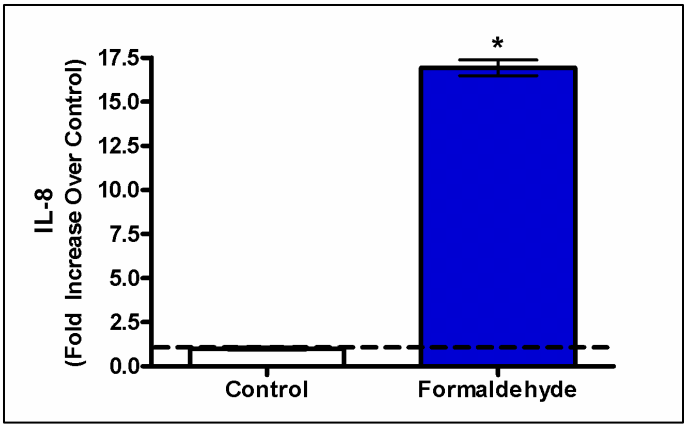

Figure 11: Interleukin-8 levels in formaldehyde-treated samples compared to untreated samples. Results are displayed as fold increase over control +/- S.E.M. The star symbol (*) indicates statistical significance compared to the control (p-value < 0.05).

2.4 DISCUSSION

In this study, we exposed human A549 lung epithelial cells to formaldehyde using an *in vitro* exposure system that physically replicates *in vivo* human lung gas exposures (Bakand, *et al.*, 2005). A549 cells are routinely used to study the effects of environmental air exposures (Jaspers, *et al.*, 1997; Sexton, *et al.*, 2004; Speit, *et al.*, 2010), and they have been proposed as a standardized model to study the lung epithelium (Foster *et al.*, 1998). Moreover, when exposed to gases at an air-liquid interface, A549 cells secrete enough surfactant to mimic airway surface tension (Blank *et al.*, 2006).

Our microarray analysis revealed that formaldehyde exposure resulted in the down-regulation of 89 miRNAs. It was interesting that all of the modulated miRNAs were down-regulated by formaldehyde exposure. This general trend of miRNA down-regulation has been observed in rat lung cells exposed to cigarette smoke (Izzotti *et al.*, 2009), as well as in multiple tumor cell types, including lung cancer, breast cancer, and leukemia (Lu, *et al.*, 2005)

We focused a detailed analysis on the four most significantly down-regulated miRNAs, as determined through microarray analysis and qRT-PCR: miR-33, miR-330, miR-181a, and miR-10b. These miRNAs have been studied, to some extent, and knowledge about their

regulation and association to disease is growing. For example, miR-33 shows decreased expression levels in tissues from patients with lung carcinomas (Yanaihara *et al.*, 2006). Also, miR-330 expression levels have been measured at significantly lower levels in human prostate cancer cells when compared against nontumorigenic prostate cells (Lee *et al.*, 2009). Furthermore, miR-330 has been suggested to act as a tumor suppressor by regulating apoptosis of cancer cells (Lee, *et al.*, 2009). In addition, miR-10b shows altered expression levels within breast cancer tissue, and is one of the most consistently dysregulated miRNAs able to predict tumor classification (Iorio *et al.*, 2005; Ma, *et al.*, 2007). These findings suggest that miR-33, miR-330, and miR-10b may influence cellular disease state, specifically related to cancer.

Formaldehyde exposure also altered the expression level of miR-181a, which has known associations with leukemogenesis (Marcucci *et al.*, 2009). The specific link between formaldehyde exposure and leukemia is currently debated, as numerous epidemiological studies show evidence for possible association to this disease (Pinkerton *et al.*, 2004; Zhang *et al.*, 2010), as well as against it (Bachand *et al.*, 2010; Marsh and Youk, 2004). However, it is important to note that our study evaluates miRNA expression in lung cells, which likely differ from leukemia target cells' responses to formaldehyde exposure, or exposure to formaldehyde's metabolic products. Nevertheless, it is worth highlighting the observation of the dysregulation of miR-181a upon exposure to formaldehyde.

To expand our analysis, we used a systems biology approach to understand the potential biological implications of the miRNA expression changes induced by acute formaldehyde exposure. For this analysis, we used a stringent computational matching approach to identify predicted mRNA targets for miR-33, miR-330, miR-181a, and miR-10b. The identified mRNA targets were used to construct associated molecular networks and were analyzed for their known involvement in signaling pathways and biological functions. The identified networks showed enrichment for various canonical pathways including nuclear factor kappa beta (NFκB) and interleukin-8 (IL-8) signaling. Although very few predicted targets overlapped between the four miRNAs, proteins involved with cancer mechanisms including that of the NFκB pathway were found within the miRNA target networks. Importantly, NFκB

has clear links to inflammation and cancer development (Karin and Greten, 2005; Schmid and Birbach, 2008). Also related to inflammation, IL-8 signaling was present in the miRNA target networks. Previous studies have shown IL-8 release in lungs cells representing inflammatory response after exposure to other air pollutants (Jaspers, *et al.*, 1997; Sexton, *et al.*, 2004). In addition, investigations have shown increased IL-8 levels in lungs of patients with diseases such as acute lung injury (McClintock *et al.*, 2008), adult respiratory distress syndrome (Jorens *et al.*, 1992), and asthma (Bloemen *et al.*, 2007). Inflammation is a recognized formaldehyde-induced response, as formaldehyde is known to irritate the respiratory system (Smith, *et al.*, 2000) and increase asthmatic response (Rumchev, *et al.*, 2002; Wieslander, *et al.*, 1997). Our findings suggest that the canonical pathways associated with formaldehyde-induced miRNA profile changes may affect the regulation of biological pathways associated with various disease states, including cancer and inflammation.

As a method to further verify our results, we compared the protein levels of cytokine interkeukin-8 (IL-8) in formaldehyde-exposed cells versus mock-treated controls. We found that, indeed, IL-8 showed significantly increased protein expression levels in the formaldehyde-exposed cells. These results support our findings that IL-8 signaling is altered in lung cells exposed to formaldehyde. Our network analyses suggest that cytokine signaling may be altered through changes in miRNA expression levels. Supporting this is a recent study that shows microRNAs can repress inflammatory response signaling by promoting the decay of related Interleukin transcripts (Anderson, 2010). Future research will test whether miRNA expression changes are directly associated with IL-8 signaling.

In an effort to gain further understanding of formaldehyde's effects on gene expression, we compared our results with those of an existing genomics database (e.g. mRNA) from a study that evaluated human lung cells exposed to formaldehyde (Li, *et al.*, 2007). Using the predicted targets in our most significant miRNA networks, we found the following genes overlap with the existing database: *BDNF, BMPR2, CACNA1C, CSNK1D, HMGA2, HSF2, HSPH1*, and *PIM1*. These genes have been shown to play a role in various diseases. For example, *BDNF*, or brain-derived neurotrophic factor, modulates neurogenesis after injury to the central nervous system (Ming and Song, 2005). *CSNK1D*, or casein kinase 1 delta, has

been identified as up-regulated in breast cancer tissue (Abba *et al.*, 2007). *HMGA2*, or high mobility group AT-hook 2, is oncogenic in many tumor cells, including lung carcinoma cells, and is regulated by the tumor-suppressive miRNA let-7 (Lee and Dutta, 2007). Lastly, *PIM1*, or Pim-1 oncogene, is found at increased levels within prostate cancer tissue (Dhanasekaran *et al.*, 2001). Network analysis of all formaldehyde-responsive genes identified through the Li et. al. (2007) study revealed significant associations with cancer, inflammation, and endocrine system regulation, which also overlap with our findings. These genes are therefore linked with formaldehyde-induced changes in miRNA abundance as well as mRNA alterations, and they are related to a diverse range of cellular responses including tumorigenesis.

2.5 CONCLUSIONS

Our study provides evidence of a potential mechanism that may underlie the cellular effects induced by formaldehyde, namely the modification of miRNA expression. We identify a set of 89 miRNAs that are dysregulated in human lung cells exposed to formaldehyde. Mapping the most significantly changed miRNAs to their predicted mRNA targets and their network interactomes within the cell reveals the association of formaldehyde exposure to inflammatory response pathways. We also validate our findings by: (1) performing qRT-PCR; (2) integrating our predicted networks with known formaldehyde-induced mRNA expression changes; and (3) examining protein expression changes of a key inflammatory response mediator, IL-8. Future research will investigate whether the expression levels of these miRNAs may serve as potential biomarkers of formaldehyde exposure in humans. Such biomarkers can be utilized to better monitor human exposure to environmental toxicants and relate them to health effects. Based on our findings, we believe that miRNAs likely play an important role in regulating formaldehyde-induced gene expression and may represent a possible link between exposure and disease.

3 THESIS CONCLUSION

Together, our two studies reveal novel biological mechanisms that may influence lung cell disease state after exposure to common air pollutants. We first investigate gene-transcript responses of lung cells exposed to air pollutant mixtures found in urban atmospheres. After discovering that secondary pollutants induce a significantly more robust response in lung cells in comparison to primary pollutants, we focus on lung cell response to formaldehyde, a common secondary pollutant. Here, we identify microRNAs at altered expression levels due to formaldehyde exposure. Mapping these miRNAs to their predicted transcript targets and associated pathways reveals networks and biological functions possibly affected by the formaldehyde-associated miRNAs. Some pathways identified as associated with air pollution exposure confirm findings from previous studies, such as air pollutant-induced IL-8 release through pathways potentially involving NFκB and AP-1 regulation. Other mechanisms are more novel, and have yet to be strongly associated with air pollution effects. These proposed mechanisms include, for example, HNF-1 and HNF-4 transcriptional regulation. Both studies identify genes that are influenced by air pollution exposure in human lung cells. We propose that these genes can potentially be used as biomarkers of air pollutant exposure, upon further validation. Such biomarkers can be utilized to effectively monitor human exposure to environmental toxicants and link them to health effects. Altogether, our research shows that air pollution heavily influences multiple signaling pathways within lung cells, potentially initiating a variety of diseases.

4 ADDITIONAL FILES

Additional File 1: Volatile organic compounds detected through gas chromatography throughout the experiment day.

Dashes (-) represent flame ionization detection peaks that were below detection limits. Concentration is abbreviated as Conc.

Compound	Chemical Group	Conc at 8:45 AM (ppmC)	Conc at 9:59 AM (ppmC)	Conc at 12:40 PM (ppmC)	Conc at 4:42 PM (ppmC)	Conc at 5:57 PM (ppmC)
butane	alkane	0.147	0.109	0.078	0.079	0.028
cyclohexane	alkane	0.017	0.016	0.012	0.006	0.006
isobutane	alkane	0.066	0.034	-	0.022	0.027
isopentane	alkane	0.173	0.146	0.071	0.056	0.036
methylcyclohexane	alkane	0.010	0.009	0.007	0.006	0.003
methylcyclopentane	alkane	0.023	0.018	0.013	-	-
n-decane	alkane	0.031	0.029	0.023	0.015	0.009
n-heptane	alkane	0.038	0.033	0.026	0.018	0.017
n-hexane	alkane	0.015	0.015	0.017	0.016	0.015
n-nonane	alkane	0.036	0.029	0.022	0.012	0.012
n-octane	alkane	0.024	0.021	0.018	0.012	0.010
n-pentane	alkane	0.092	0.097	0.056	0.057	0.013
propane/propene	85% alkane, 15% alkene	0.109	0.105	0.083	0.048	0.057
2,2,4-trimethylpentane	alkane	0.033	0.028	0.025	0.018	0.019
2,3,4-trimethylpentane	alkane	0.020	0.018	0.015	0.010	0.009
2,3-dimethyl-pentane	alkane	0.034	0.026	0.023	0.017	0.016
2,5-dimethylhexane	alkane	0.023	0.020	0.016	0.011	0.009
2-methylpentane	alkane	0.044	0.031	0.023	0.019	0.020
3-methylhexane	alkane	0.070	0.068	0.055	0.041	0.034
3-methylpentane	alkane	0.034	0.024	0.011	0.009	0.010
4-methylnonane	alkane	0.042	0.031	0.023	0.015	0.012
a-pinene	alkene	0.010	0.006	-	-	-
c-2-pentene	alkene	0.015	-	-	-	-
1-nonene	alkene	0.012	0.011	0.005	-	-
1-octene	alkene	0.017	0.010	0.005	-	-
2,3,3-trimethyl-1-	alkene	0.049	0.030	0.008	-	-
benzene	aromatic	0.044	0.042	0.008	0.033	0.030
ethylbenzene	aromatic	0.024	0.022	0.017	0.014	0.010
m-ethyltoluene	aromatic	0.016	0.013	0.009	0.004	-
m-xylene	aromatic	0.074	0.059	0.038	0.022	0.016
n-propylbenzene	aromatic	0.022	0.019	0.015	0.009	0.008
o-xylene	aromatic	0.025	0.023	0.017	0.010	0.009
p-ethyltoluene	aromatic	0.025	0.022	0.016	0.009	0.010
sec-butylbenzene	aromatic	0.018	0.018	0.015	0.008	0.006
toluene	aromatic	0.138	0.121	0.101	0.078	0.072
1,2,3,5-	aromatic	0.007	0.007	-	-	-
1,2,4-trimethylbenzene	aromatic	0.113	0.094	0.052	0.023	0.018
1,3-diethylbenzene	aromatic	0.034	0.029	0.018	0.010	0.007

Additional File 2: Genes identified as significantly* differentially expressed due to primary or photochemically altered (PCA) pollutant exposure.

Dashes (-) indicate insignificant changes in expression.
* Fold change ≥ 1.5 or ≤ - 1.5, p-value < 0.05, q-value < 0.05

Gene Symbol	Primary Pollutant Fold Change (Exposed / Unexposed)	p-value	PCA Pollutant Fold Change (Exposed / Unexposed)	p-value
A1CF	-	-	-1.81	0.00276
ABCA1	-	-	-1.56	0.00494
ABCA12	-	-	-1.60	0.00261
ABCA5	-	-	-1.90	0.00149
ABCB6	-	-	-1.70	0.00001
ABCC4	-	-	-1.59	0.00002
ABCG2	-	-	-1.57	0.00124
ACAD10	-	-	-1.78	0.00075
ACAD11	-	-	-1.55	0.00078
ACSM3	-1.50	0.00606	-2.33	0.00038
ACSS2	-	-	-1.54	0.00001
ACTA2	-1.58	0.01055	-2.00	0.00239
ADD3	-	-	-1.98	0.00022
ADH1C	-	-	-1.71	0.04174
ADH6	-	-	-2.12	0.00168
ADHFE1	-	-	-1.51	0.01654
AHCYL1	-	-	-1.60	0.00109
AK3L1	-	-	-1.99	0.00053
AK7	-	-	-2.18	0.00040
AKAP12	-	-	1.67	0.00192
AKAP9	-	-	-1.73	0.00219
AKR1B1	-	-	1.55	0.00102
ALDH5A1	-	-	-1.51	0.00686
ALDH6A1	-	-	-2.50	0.00005
ALPK1	-	-	-1.82	0.00084
ALS2CR8	-	-	-1.55	0.03130
AMPD1	-	-	1.49	0.00723
ANG	-	-	-1.92	0.00069
ANKRA2	-	-	-1.82	0.00414
ANKRD1	-	-	1.68	0.01161
ANKRD18A	-	-	-1.51	0.01560
ANKRD22	-	-	2.88	0.00045
ANKRD30A	-	-	-1.87	0.00110
ANKS4B	-	-	-1.65	0.00036
ANO5	-	-	-1.52	0.00979
ANXA10	-	-	1.69	0.00491
ANXA13	-	-	-1.68	0.00414

Gene Symbol	Primary Pollutant Fold Change (Exposed / Unexposed)	p-value	PCA Pollutant Fold Change (Exposed / Unexposed)	p-value
ANXA3	-	-	1.91	0.00053
ANXA4	-	-	-1.64	0.00035
ANXA9	-	-	-1.70	0.00337
AP1S3	-	-	1.80	0.00005
APAF1	-	-	-1.52	0.00074
APH1B	-	-	-1.52	0.00342
APOBEC3C	-	-	-1.52	0.00212
APOH	-	-	-1.56	0.00202
AQP3	-	-	1.79	0.00023
AR	-	-	-1.79	0.00008
AREG	-	-	3.22	0.00010
ARFGAP2	-	-	-1.55	0.00108
ARHGAP1	-	-	-1.54	0.00038
ARID4A	-	-	-1.67	0.00244
ARID4B	-	-	-1.53	0.00671
ARID5B	-	-	-1.60	0.00007
ARL15	-	-	-1.69	0.00037
ARMCX3	-	-	-1.58	0.00332
ARRB1	-	-	-1.63	0.00000
ARSD	-	-	-1.60	0.00019
ARSE	-	-	-1.93	0.00001
AS3MT	-	-	-1.76	0.00281
ASAM	-	-	1.91	0.00731
ASPM	-	-	-1.83	0.00018
ATF6B	-	-	-1.52	0.00112
ATG2B	-	-	-1.58	0.00018
ATP8B1	-1.52	0.02013	-2.95	0.00063
ATP9A	-	-	-1.80	0.00014
AXL	-	-	1.72	0.00217
BAMBI	-	-	-1.72	0.00027
BBS9	-	-	-1.59	0.01465
BCAS3	-	-	-1.71	0.00032
BCL2L11	-	-	-2.10	0.00011
BCL2L15	-	-	-1.85	0.01002
BCMO1	-	-	-2.49	0.00009
BDH2	-	-	-1.58	0.00797
BDKRB1	-	-	-1.57	0.00003
BDKRB2	-	-	-1.67	0.00396
BLMH	-	-	-1.51	0.00475
BMPR2	-	-	-1.82	0.00290
BNIP3L	-	-	-1.55	0.00010
BTBD11	-	-	-1.76	0.00003
BTN3A1	-	-	-1.73	0.00562
BTN3A3	-	-	-1.58	0.00015

Gene Symbol	Primary Pollutant Fold Change (Exposed / Unexposed)	p-value	PCA Pollutant Fold Change (Exposed / Unexposed)	p-value
C10orf114	-	-	1.61	0.00440
C10orf57	-	-	-1.70	0.00847
C12orf27	-	-	-1.91	0.00166
C14orf106	-	-	-1.55	0.02065
C15orf51	-	-	-1.57	0.01539
C18orf58	-	-	-1.61	0.00275
C1orf63	-	-	-1.64	0.00036
C1RL	-	-	-2.01	0.00000
C1S	-	-	-1.59	0.00003
C20orf19	-	-	-1.60	0.00380
C20orf194	-	-	-1.75	0.00203
C20orf74	-	-	-1.87	0.00071
C4orf18	-	-	-2.25	0.00219
C4orf34	-	-	-1.61	0.00130
C5	-	-	-2.07	0.00044
C5orf26	-	-	-1.50	0.00269
C5orf42	-	-	-1.64	0.00225
C6orf130	-	-	-1.54	0.00364
C6orf191	-	-	1.58	0.00075
C7orf11	-	-	1.63	0.00256
C7orf68	-	-	-1.53	0.00114
C9orf3	-	-	-2.13	0.00103
CA12	-	-	-1.52	0.00010
CABYR	-	-	-1.67	0.00169
CACNA1D	-	-	-1.87	0.00156
CACNA2D1	-	-	-1.52	0.00198
CALB1	-	-	1.62	0.02927
CALCOCO1	-	-	-1.54	0.00013
CAMK2D	-	-	-1.82	0.00122
CASP4	-	-	-1.63	0.00331
CASP6	-	-	-1.60	0.00474
CCBE1	-	-	1.95	0.00131
CCBL2	-	-	-1.74	0.00035
CCDC144A	-	-	-1.55	0.00054
CCDC28A	-	-	-1.56	0.00767
CCDC34	-	-	-1.59	0.00058
CCDC80	-	-	-1.64	0.00112
CCDC99	-	-	1.60	0.00285
CCL2	1.79	0.00249	1.96	0.00143
CCNG1	-	-	-1.52	0.00018
CCNG2	-	-	-1.98	0.00375
CCPG1	-	-	-1.58	0.00182
CD177	-	-	1.95	0.03989
CD209	-	-	1.50	0.02235

Gene Symbol	Primary Pollutant Fold Change (Exposed / Unexposed)	p-value	PCA Pollutant Fold Change (Exposed / Unexposed)	p-value
CD55	-	-	1.79	0.00225
CD99L2	-	-	-1.61	0.00007
CDC14B	-	-	-1.52	0.01131
CDC25C	-	-	-1.53	0.00153
CDCA7L	-	-	-1.54	0.00104
CDCP1	-	-	1.90	0.00178
CDH1	-	-	-2.16	0.00002
CDK5RAP3	-	-	-1.67	0.00070
CDRT1	1.59	0.01225	-	-
CEACAM1	-	-	-1.84	0.00148
CEACAM5	-	-	1.52	0.00019
CENPF	-	-	-1.55	0.00048
CEP152	-	-	-1.55	0.00683
CEP70	-	-	-1.75	0.00426
CFH	-	-	-1.71	0.00089
CFHR1	-	-	-2.19	0.02652
CFHR3	-	-	-1.85	0.00017
CFI	-	-	-1.66	0.00250
CIR1	-	-	-1.55	0.00185
CIRBP	-	-	-1.88	0.00006
CLDN1	-	-	1.65	0.00159
CLMN	-	-	-1.66	0.00034
CNNM2	-	-	-1.53	0.00028
CORO2A	-	-	-1.72	0.00068
COTL1	-	-	1.51	0.00228
CPA4	-	-	1.59	0.00746
CPB2	-	-	-1.69	0.01898
CPN1	-	-	-2.02	0.00108
CRBN	-	-	-1.52	0.00023
CSGALNACT2	-	-	1.59	0.00058
CSRP1	-	-	1.70	0.00046
CST1	-	-	1.77	0.01150
CTDSP2	-	-	-1.76	0.00009
CTNND1	-	-	-1.56	0.00021
CTPS	-	-	1.80	0.00187
CTTNBP2	-	-	-1.57	0.00047
CXCL5	-	-	2.43	0.00092
CYB5A	-	-	-1.55	0.00014
CYBRD1	-	-	-1.54	0.00007
CYFIP2	-	-	-1.66	0.00007
CYHR1	-	-	-1.91	0.00028
CYP2D6	1.63	0.04999	-	-
CYP4F11	-	-	-1.66	0.00038
CYP4F12	-	-	-1.54	0.00544

Gene Symbol	Primary Pollutant Fold Change (Exposed / Unexposed)	p-value	PCA Pollutant Fold Change (Exposed / Unexposed)	p-value
CYP4F3	-	-	-1.75	0.00058
DAB2	-	-	-1.70	0.00070
DAPK1	-	-	-1.77	0.00001
DCDC2	-	-	-1.56	0.00069
DCDC5	-	-	-1.76	0.00020
DCLK1	-	-	1.90	0.00068
DEPDC4	-	-	-1.53	0.03365
DEPDC6	-	-	-2.18	0.00003
DET1	-	-	-1.58	0.00028
DGCR6	-	-	1.52	0.02151
DHCR24	-	-	-1.51	0.00004
DHFR	-	-	1.81	0.01212
DHRS3	-	-	-3.69	0.00017
DHRS9	-	-	1.56	0.02798
DHX37	-	-	1.60	0.00006
DIAPH2	-	-	-1.53	0.00949
DMXL2	-	-	-1.71	0.00070
DNAJB4	-	-	-1.55	0.01249
DND1	-	-	1.52	0.02165
DPT	-	-	1.53	0.02035
DPYD	-	-	-1.53	0.00512
DTX3L	-	-	-1.86	0.00046
DUSP1	-	-	1.65	0.00089
DUSP4	-	-	1.52	0.00138
DUSP5	-	-	1.79	0.00028
DYNC2H1	-	-	-1.63	0.00337
DZIP3	-	-	-1.55	0.01019
ECE1	-	-	-1.59	0.00002
EFHC1	-	-	-1.55	0.00639
EFNA1	-	-	-1.79	0.00405
EFNB2	-	-	1.61	0.00046
EHHADH	-	-	-1.61	0.00032
EIF2C4	-	-	-1.52	0.00280
EIF4B	-	-	-1.68	0.00510
ELF3	-	-	-1.77	0.00003
ELMO1	-	-	-1.94	0.00006
ELOVL6	-	-	-1.56	0.00533
ELP4	-	-	-1.52	0.00125
EML4	-	-	-1.52	0.00001
EMP1	-	-	1.84	0.00096
ENTPD5	-	-	-1.62	0.00211
EPB41	-	-	-1.51	0.00242
EPB41L4A	-	-	-1.80	0.00023
EPHA2	-	-	1.60	0.00280

Gene Symbol	Primary Pollutant Fold Change (Exposed / Unexposed)	p-value	PCA Pollutant Fold Change (Exposed / Unexposed)	p-value
EPHX2	-	-	-1.90	0.00134
ERAP1	-	-	-1.53	0.00020
ERBB2	-	-	-1.77	0.00002
ERBB3	-	-	-2.33	0.00023
EREG	-	-	2.83	0.00013
ESSPL	-	-	1.60	0.01585
FAM105A	-	-	-1.55	0.03236
FAM111A	-	-	-1.69	0.00133
FAM149A	-	-	-1.76	0.00333
FAM149B1	-	-	-1.52	0.00208
FAM175A	-	-	-1.55	0.00181
FAM185A	-	-	-1.56	0.04457
FAM38B	-	-	-1.88	0.00154
FAM55C	-	-	-1.55	0.00081
FAM74A3	-	-	1.53	0.00197
FARP2	-	-	-1.72	0.00120
FBXO32	-	-	-1.84	0.03748
FBXO40	-	-	1.51	0.02732
FCHSD2	-	-	-1.66	0.00440
FGA	-	-	-1.81	0.00031
FGB	-	-	-1.88	0.00010
FGFBP1	-	-	3.85	0.00009
FGFR4	-	-	-1.71	0.00013
FGG	-	-	-1.80	0.00209
FHL1	-	-	1.66	0.00166
FKBP5	-	-	-1.53	0.00162
FLI1	-	-	1.52	0.04567
FLJ11292	-	-	1.52	0.00547
FLJ35848	-	-	-1.54	0.03243
FLJ41484	-	-	-1.56	0.02264
FLJ44124	-	-	-1.55	0.00060
FLOT1	-	-	-1.65	0.00036
FMN1	-	-	-1.53	0.02740
FMO5	-	-	-2.14	0.00015
FNBP1L	-	-	-1.54	0.00003
FNIP1	-	-	-1.84	0.00347
FOXN3	-	-	-1.72	0.01025
FRAS1	-	-	-1.53	0.00459
FRK	-	-	-2.00	0.00024
FRMD3	-	-	1.69	0.00082
FSHB	-	-	1.51	0.01644
FSTL5	-	-	1.70	0.00515
FZD7	-	-	-1.82	0.00001
GAB1	-	-	-1.53	0.02435

Gene Symbol	Primary Pollutant Fold Change (Exposed / Unexposed)	p-value	PCA Pollutant Fold Change (Exposed / Unexposed)	p-value
GABARAPL1	-	-	-1.69	0.00401
GABPA	-	-	2.01	0.00008
GABRA5	-	-	1.84	0.00003
GABRE	-	-	-1.76	0.01005
GALNT12	-	-	-1.56	0.00227
GALNT4	-	-	-1.65	0.00138
GATM	-	-	-1.53	0.00008
GATS	-	-	-1.71	0.00017
GATSL1	-	-	-1.68	0.00014
GCA	-	-	-1.66	0.02017
GCOM1	1.51	0.01963	1.52	0.01866
GDF15	-	-	1.60	0.01375
GFPT2	1.60	0.00208	2.78	0.00010
GIP	-	-	-1.59	0.00122
GK	-	-	-2.16	0.00005
GLIPR1	-	-	2.20	0.00018
GLS	-	-	1.54	0.00035
GLTSCR2	-	-	-1.59	0.00000
GPAM	-	-	2.35	0.02748
GPRC5A	-	-	1.68	0.00057
GPRC5B	-	-	-2.33	0.00020
GRAMD1A	-	-	-1.59	0.00031
GREM2	-	-	1.56	0.00011
GRIP1	-	-	-1.54	0.00589
GSTA4	-	-	-1.54	0.00194
GSTM4	-	-	-1.76	0.00138
GTF2IRD2	-	-	-1.53	0.00006
GUCY1B2	-	-	-1.55	0.02367
HABP2	-	-	-1.53	0.00125
HAO1	-	-	-1.52	0.01414
HAS2	-	-	2.19	0.00276
HBEGF	-	-	1.69	0.00043
HBP1	-	-	-1.88	0.00002
HERC6	-	-	-1.68	0.00007
HFE	-	-	-1.93	0.00003
HGD	-	-	-1.86	0.00098
HIST1H1C	-	-	-1.54	0.00009
HIST1H2AB	-	-	-1.63	0.00019
HIST1H2AC	-	-	-1.57	0.00023
HIST1H2AG	-	-	-1.56	0.00076
HIST1H2AI	-	-	-1.65	0.00306
HIST1H2BM	-	-	-1.51	0.00152
HIST1H3A	-	-	-1.61	0.00256
HIST1H3E	-	-	-1.54	0.01168

Gene Symbol	Primary Pollutant Fold Change (Exposed / Unexposed)	p-value	PCA Pollutant Fold Change (Exposed / Unexposed)	p-value
HIST1H3H	-	-	-1.75	0.00148
HIST1H3J	-	-	-1.68	0.00053
HIST1H4E	-	-	-1.60	0.00275
HIST1H4H	-	-	-1.71	0.00050
HIST2H2AA3	-	-	-1.52	0.00014
HIST2H2AB	-	-	-1.62	0.00076
HIST2H2BF	-	-	-1.52	0.00093
HIST2H4A	-	-	-1.72	0.00023
HMGB2	-	-	-1.58	0.00755
HMGCL	-	-	-2.08	0.00131
HNF4A	-	-	-1.71	0.00002
HNF4G	-	-	-1.66	0.00086
HOOK1	-	-	-1.61	0.00180
HOOK3	-	-	-1.79	0.00433
HOXA2	-	-	-1.56	0.00149
HP1BP3	-	-	-1.57	0.00116
HSD17B11	-	-	-1.62	0.00054
HSD17B6	-	-	-1.50	0.04876
HSPH1	-	-	1.57	0.00072
HTR1A	-	-	1.64	0.01235
HTR3D	-	-	1.54	0.01543
ID1	-	-	-1.55	0.00006
IER3	-	-	1.56	0.01359
IFI35	-	-	-1.66	0.00049
IFIT1	-	-	-2.12	0.00579
IFT81	-	-	-1.62	0.00776
IGHA1	-	-	1.50	0.00025
IGKC	-	-	1.69	0.00480
IL11	-	-	1.85	0.01698
IL8	-	-	1.60	0.00007
INA	-	-	1.64	0.00039
INADL	-	-	-1.64	0.00020
ING4	-	-	-1.55	0.00015
IP6K2	-	-	-1.55	0.00025
IQGAP2	-	-	-1.97	0.00035
ITGA3	-	-	1.67	0.00115
ITGA6	-	-	1.61	0.00050
ITGB8	-	-	1.52	0.03105
ITGBL1	-	-	1.56	0.00139
ITPR2	-	-	-1.86	0.00137
JUN	-	-	1.69	0.00013
KCNK5	-	-	-1.53	0.00055
KCNT2	-	-	-1.75	0.00295
KDM3A	-	-	-1.50	0.00034

Gene Symbol	Primary Pollutant Fold Change (Exposed / Unexposed)	p-value	PCA Pollutant Fold Change (Exposed / Unexposed)	p-value
KIAA0922	-	-	-1.73	0.00115
KIAA1109	-	-	-1.59	0.00023
KIAA1147	-	-	-1.91	0.00001
KIAA1161	-	-	-1.66	0.00119
KIAA1199	-	-	1.66	0.00915
KIAA1370	-	-	-2.70	0.00016
KIAA1377	-	-	-1.52	0.00437
KIAA1618	-	-	-1.53	0.00043
KIAA1632	-	-	-1.55	0.00050
KIAA1712	-	-	-1.60	0.00490
KIF13B	-	-	-1.56	0.00016
KIF20A	-	-	-1.57	0.00062
KIF20B	-	-	-1.52	0.02541
KIR2DL5A	-	-	1.51	0.04814
KLHDC2	-	-	-2.08	0.00016
KLHL24	-	-	-2.44	0.00010
KRT38	-	-	1.57	0.01687
KRT80	-	-	1.68	0.00320
LAMC2	-	-	2.74	0.00002
LARGE	-	-	-1.53	0.00001
LBA1	-	-	-1.93	0.00099
LETMD1	-	-	-1.72	0.00011
LHX8	-	-	-1.68	0.00030
LIMA1	-	-	-1.70	0.00084
LIMD1	-	-	-1.51	0.00395
LITAF	-	-	-1.60	0.00000
LMO7	-	-	1.61	0.00085
LOC100130581	-	-	-1.51	0.01258
LOC100289668	-	-	-1.73	0.00711
LOC162632	-	-	-1.52	0.00058
LOC345258	-	-	1.59	0.00970
LOC440518	-	-	1.51	0.00644
LOC644714	1.72	0.03988	-	-
LOC652493	-	-	1.66	0.00876
LOC729724	-	-	1.51	0.00076
LOXL2	-	-	1.55	0.00171
LRBA	-	-	-1.53	0.00005
LRIG1	-	-	-1.53	0.00104
LRP1	-	-	-1.60	0.00103
LRRFIP1	-	-	1.75	0.00191
LXN	-	-	-3.17	0.00008
LYAR	-	-	1.52	0.01154
LYRM5	-	-	-1.64	0.01013
MAFG	-	-	-1.50	0.00030

Gene Symbol	Primary Pollutant Fold Change (Exposed / Unexposed)	p-value	PCA Pollutant Fold Change (Exposed / Unexposed)	p-value
MANBA	-	-	-1.55	0.00047
MANSC1	-	-	-1.76	0.00055
MAOA	-	-	-2.04	0.00003
MARCH4	-	-	1.61	0.00392
MARCH8	-	-	-1.57	0.00542
MARCKS	-	-	-1.64	0.00055
MATN2	-	-	-2.00	0.00034
MBOAT1	-	-	-2.24	0.00006
MCCC1	-	-	-1.66	0.00242
MCTP1	-	-	1.61	0.00115
MEIS2	-	-	-1.80	0.00026
MGAM	-	-	-1.54	0.00001
MIA2	-	-	-2.10	0.00458
MICAL2	-	-	1.54	0.00571
MIR21	-	-	-1.56	0.00251
MLEC	-	-	-1.77	0.00001
MLF1	-	-	-1.57	0.01458
MLLT4	-	-	-1.55	0.00085
MR1	-	-	-1.85	0.00105
MRAP2	-	-	-1.58	0.00189
MSI2	-	-	-1.56	0.00022
MST131	-	-	2.14	0.00035
MTMR11	-	-	-1.56	0.00059
MUT	-	-	-1.95	0.00019
MYCNOS	-	-	1.56	0.00044
MYEOV	-	-	1.62	0.00574
MYO1A	-	-	-1.66	0.00277
NAGA	-	-	-1.56	0.00399
NAP1L2	-	-	-1.57	0.00312
NAV3	-	-	2.05	0.00442
NBEA	-	-	-1.73	0.00019
NBEAL1	-	-	-1.77	0.00311
NBR1	-	-	-1.55	0.00122
NCALD	-	-	-1.51	0.00015
NCAPD2	-	-	-1.64	0.00007
NCEH1	-	-	1.52	0.00011
NCOA2	-	-	-1.51	0.00336
NDC80	-	-	-1.79	0.00277
NDRG1	-	-	-1.59	0.00516
NDRG2	-	-	-1.51	0.00030
NEB	-	-	-1.76	0.00001
NEDD4L	-	-	-1.98	0.00001
NEK11	-	-	-1.54	0.00645
NFIA	-	-	-1.95	0.00006

Gene Symbol	Primary Pollutant Fold Change (Exposed / Unexposed)	p-value	PCA Pollutant Fold Change (Exposed / Unexposed)	p-value
NFKBIA	1.51	0.00016	1.53	0.00014
NFKBIZ	-	-	-1.56	0.00523
NID2	-	-	1.77	0.00950
NIPAL3	-	-	-1.54	0.00599
NIPSNAP3A	-	-	-1.67	0.00309
NMNAT2	-	-	1.68	0.00052
NNT	-	-	-1.52	0.00002
NOSTRIN	-	-	1.62	0.00233
NOTCH2	-	-	-1.68	0.00081
NOTCH2NL	-	-	-1.57	0.00049
NPY1R	-	-	-1.74	0.00596
NR0B1	-	-	-1.52	0.00120
NR1D2	-	-	-1.53	0.00500
NR4A1	-	-	1.69	0.00307
NR5A2	-	-	-1.87	0.00024
NRG1	-	-	1.73	0.00057
NRG4	-	-	-1.81	0.00017
NRIP1	-	-	-1.66	0.00006
NRM	-	-	-1.57	0.00020
NUDT7	-	-	-1.76	0.00017
NUSAP1	-	-	-1.59	0.00367
OAS1	1.62	0.00085	2.43	0.00008
ODC1	-	-	1.66	0.00054
OPHN1	-	-	-1.96	0.00109
OR10G2	1.55	0.01301	-	-
OR2T1	-	-	1.55	0.01790
OR4A47	1.73	0.00898	1.56	0.01815
OR4C6	-	-	1.91	0.00808
OR4K1	-	-	1.53	0.03161
OR51B4	-	-	1.53	0.02610
OR51D1	1.51	0.01660	-	-
OR51L1	-	-	1.54	0.00304
OR52I2	-	-	1.50	0.04879
OR5AP2	-	-	1.58	0.03335
OR5I1	-	-	1.51	0.02120
OR9A2	-	-	1.50	0.04432
OR9Q2	-	-	1.60	0.00867
OSBPL9	-	-	-1.70	0.00006
OTUD1	-	-	-1.54	0.00205
OXTR	-	-	1.68	0.00105
P2RX4	-	-	-1.73	0.00037
P2RY4	-	-	1.54	0.01409
PAIP2B	-	-	-2.02	0.00175
PAN2	-	-	-1.61	0.00035

Gene Symbol	Primary Pollutant Fold Change (Exposed / Unexposed)	p-value	PCA Pollutant Fold Change (Exposed / Unexposed)	p-value
PAQR5	1.58	0.00319	2.19	0.00041
PAR5	-	-	-1.63	0.02688
PARP14	-	-	-1.85	0.00301
PARP9	-	-	-2.18	0.00046
PBLD	-	-	-2.57	0.00011
PCCA	-	-	-1.52	0.00462
PCDH9	-	-	-1.56	0.00041
PCMTD1	-	-	-1.66	0.00712
PCMTD2	-	-	-1.66	0.01622
PDCD4	-	-	-2.83	0.00020
PDE3A	-	-	-1.59	0.00000
PDGFC	-	-	-1.62	0.00169
PDGFRL	-	-	-1.90	0.00023
PDK2	-	-	-1.76	0.00028
PDPR	-	-	-1.52	0.00073
PDXDC2	-	-	-1.53	0.03412
PDZK1	-	-	-2.25	0.00069
PECR	-	-	-1.55	0.00431
PER2	-	-	-1.55	0.00044
PFKFB3	-	-	-1.83	0.00003
PGAP2	-	-	-1.56	0.00005
PHKB	-	-	-1.65	0.00006
PHLDA1	-	-	1.57	0.00787
PIGN	-	-	-1.57	0.00143
PIR	-	-	-1.55	0.00011
PLAU	-	-	1.75	0.00069
PLCD4	-	-	-1.53	0.03026
PLCH1	-	-	-1.99	0.00005
PLD1	-	-	-1.98	0.00015
PLEK2	-	-	1.75	0.00089
PLEKHH2	-	-	-1.64	0.00354
PMEPA1	-	-	1.51	0.00154
POF1B	-	-	-1.78	0.00037
POU1F1	-	-	1.53	0.01782
PP13439	-	-	-1.57	0.01707
PPFIBP2	-	-	-1.58	0.00008
PRKCD	-	-	-1.67	0.00020
PRKD1	-	-	-1.51	0.00039
PSG8	-	-	1.67	0.02451
PTCH2	-	-	3.64	0.00156
PTGR2	-	-	-1.55	0.00099
PTPLAD2	-	-	-1.73	0.01052
PTRF	-	-	1.65	0.00203
RAB3B	-	-	1.80	0.01271

Gene Symbol	Primary Pollutant Fold Change (Exposed / Unexposed)	p-value	PCA Pollutant Fold Change (Exposed / Unexposed)	p-value
RAP1GAP	-	-	-1.50	0.00055
RARB	-	-	-2.43	0.00014
RBKS	-	-	-1.61	0.00033
RBM14	-	-	1.51	0.00420
RBPMS	-	-	-1.51	0.00151
REXO1L1	-	-	1.50	0.00607
RFC1	-	-	2.82	0.01137
RFX5	-	-	-1.60	0.00110
RHOBTB1	-	-	-1.72	0.00326
RHOBTB3	-	-	-2.60	0.00060
RMRP	-	-	1.51	0.00148
RND1	-	-	-1.70	0.00000
RNF182	-	-	1.80	0.00674
RNF213	-	-	-1.64	0.00071
RNU11	-	-	3.67	0.00002
RNU1A	-	-	2.73	0.00053
ROBO1	-	-	-1.63	0.00025
RPPH1	-	-	2.27	0.00040
RPS27L	-	-	-1.54	0.00908
RPS6KA5	-	-	-1.61	0.03867
S100A3	-	-	1.82	0.00190
SAMD7	-	-	1.56	0.01411
SAMD9	-	-	-1.58	0.00962
SASH1	-	-	-1.54	0.00060
SCAPER	-	-	-1.55	0.00306
SCARB1	-	-	-1.59	0.00036
SCARNA17	-	-	-1.71	0.00735
SCARNA9L	-	-	-1.52	0.02066
SCD	-	-	-1.59	0.00047
SCMH1	-	-	-1.88	0.00078
SCNN1A	-	-	-1.50	0.00613
SELENBP1	-	-	-1.64	0.00061
SEMA3C	-	-	1.51	0.00066
SEMA3E	-	-	-1.62	0.00373
SEPT14	-	-	-1.57	0.00871
SERPINA6	-	-	-1.52	0.00385
SERPINB1	-	-	-1.60	0.00030
SERPINB8	-	-	1.53	0.00242
SERPINE2	-	-	2.12	0.00056
SESN3	-	-	-2.96	0.00075
SFRP1	-	-	1.98	0.00127
SFRP4	-	-	-3.15	0.00197
SFRS18	-	-	-1.64	0.00018
SH3BGRL2	-	-	-1.51	0.00173

Gene Symbol	Primary Pollutant Fold Change (Exposed / Unexposed)	p-value	PCA Pollutant Fold Change (Exposed / Unexposed)	p-value
SHMT1	-	-	-1.62	0.00007
SKAP2	-	-	-1.68	0.00063
SLC16A7	-	-	-1.87	0.01007
SLC19A3	-	-	-1.55	0.00100
SLC22A3	-	-	-1.63	0.00131
SLC23A1	-	-	-1.77	0.00003
SLC23A2	-	-	-1.82	0.00014
SLC25A27	-	-	-1.63	0.00072
SLC29A3	-	-	-1.51	0.00096
SLC2A12	-	-	-1.62	0.00055
SLC35D2	-	-	-1.57	0.00050
SLC40A1	-	-	-1.77	0.00003
SLC41A2	-	-	-1.95	0.00481
SLC44A2	-	-	-1.64	0.00225
SLC46A3	-	-	-2.19	0.00010
SLC5A3	1.58	0.00021	1.80	0.00050
SLC7A2	-	-	-1.55	0.00019
SLC9A3R1	-	-	-1.51	0.00002
SLCO4A1	-	-	1.71	0.00150
SLFN5	-	-	-1.93	0.00126
SMOX	-	-	1.71	0.00078
SMPD1	-	-	-1.91	0.00009
SNORA23	-	-	2.06	0.00929
SNORA3	-	-	3.62	0.00002
SNORA42	-	-	6.35	0.00030
SNORA52	-	-	1.56	0.00372
SNORA56	-	-	1.77	0.00060
SNORA71D	-	-	1.60	0.00162
SNORA73A	-	-	1.92	0.01085
SNORD113-3	-	-	1.55	0.02204
SNORD114-2	-	-	1.52	0.00117
SNORD115-11	-	-	1.64	0.02566
SOAT1	-	-	-1.70	0.00056
SOCS2	-	-	1.64	0.00506
SORL1	-	-	-1.73	0.00003
SPANXE	-	-	1.54	0.00193
SPATA18	-	-	-1.57	0.00132
SPATA7	-	-	-1.54	0.00089
SPG11	-	-	-1.53	0.00373
SPP1	-	-	-1.51	0.00065
SPRR2B	-	-	1.73	0.04629
SPTLC3	-	-	-1.85	0.00091
SSBP2	-	-	-1.66	0.00118
SSFA2	-	-	1.55	0.00011

Gene Symbol	Primary Pollutant Fold Change (Exposed / Unexposed)	p-value	PCA Pollutant Fold Change (Exposed / Unexposed)	p-value
ST6GAL1	-	-	-2.01	0.00014
ST8SIA4	-	-	-1.83	0.00152
STAMBPL1	-	-	2.17	0.00034
STAT4	-	-	-2.20	0.00111
STAT6	-	-	-1.51	0.00003
STC1	-	-	1.69	0.00281
STEAP2	-	-	-1.53	0.00016
STRA6	-	-	-1.56	0.00983
STX17	-	-	-1.50	0.00142
SULT2B1	-	-	-1.69	0.00114
SVEP1	-	-	-1.85	0.00007
SYCP2L	-	-	-1.59	0.00681
SYNE2	-	-	-2.41	0.00008
TAF9B	-	-	-1.74	0.00036
TAS2R5	-	-	1.92	0.00015
TBC1D5	-	-	-1.56	0.00100
TBC1D8B	-	-	-1.63	0.00073
TBCK	-	-	-1.83	0.00024
TC2N	-	-	-1.96	0.00089
TCP11L2	-	-	-2.19	0.00000
TFDP2	-	-	-1.79	0.00115
TFPI2	-	-	1.77	0.00075
TGFA	-	-	1.53	0.00335
TGFB2	-	-	-1.67	0.00049
THBD	-	-	1.53	0.00012
THG1L	-	-	-1.56	0.00083
TIGD2	-	-	-1.85	0.00235
TJP2	-	-	-1.52	0.00181
TLR1	-	-	-2.18	0.00124
TLR3	-	-	-1.65	0.04429
TM4SF20	-	-	-2.48	0.00007
TMC5	-	-	-1.51	0.01022
TMC7	-	-	-1.64	0.00134
TMEM136	-	-	-1.52	0.00435
TMEM140	-	-	-1.56	0.01972
TMEM144	-	-	-1.81	0.00091
TMEM171	-	-	1.52	0.00509
TMEM202	-	-	1.59	0.04147
TMEM37	-	-	-2.39	0.00015
TMEM50B	-	-	-1.60	0.01620
TMEM60	-	-	-1.59	0.00078
TMEM74	-	-	1.53	0.03446
TNFRSF12A	-	-	1.83	0.00138
TNFSF10	-	-	-2.26	0.00008

Gene Symbol	Primary Pollutant Fold Change (Exposed / Unexposed)	p-value	PCA Pollutant Fold Change (Exposed / Unexposed)	p-value
TNS4	-	-	2.08	0.00110
TP53	-	-	-1.59	0.00023
TP53INP1	-	-	-2.68	0.00001
TPCN1	-	-	-1.86	0.00077
TRIM31	-	-	-1.84	0.00213
TRIM52	-	-	-1.56	0.02026
TRIML2	-	-	1.60	0.01377
TSKU	-	-	-1.59	0.00001
TSPAN15	-	-	-1.56	0.00043
TST	-	-	-1.54	0.00005
TTC28	-	-	-1.68	0.00006
TTC39B	-	-	-1.63	0.01453
TTLL6	-	-	-1.53	0.00987
TUBB2C	-	-	1.51	0.01846
TXNDC16	-	-	-1.67	0.00328
TXNIP	-1.61	0.01468	-2.43	0.00160
TYR	-	-	1.73	0.03422
UBASH3B	-	-	1.82	0.00084
UGT2B15	-	-	-1.77	0.00178
UIMC1	-	-	1.93	0.00243
UNC119B	-	-	-1.58	0.00219
UNC13B	-	-	-1.59	0.00049
USP17	1.52	0.04976	1.59	0.03729
USP17L2	1.52	0.03727	1.53	0.03761
USP3	-	-	-1.50	0.00022
VCAN	-	-	-1.62	0.00012
VIL1	-	-	-1.65	0.00193
VPS13C	-	-	-1.53	0.00114
VPS39	-	-	-1.55	0.00068
VRK2	-	-	-1.53	0.00264
VWC2L	-	-	1.57	0.03979
WDR19	-	-	-1.51	0.00064
WDR69	-	-	1.61	0.02320
WEE1	-	-	-1.93	0.00169
WSB1	-	-	-1.53	0.00269
WWP1	-	-	-1.51	0.00065
XBP1	-	-	-1.54	0.00334
XCL1	-	-	1.66	0.01526
XDH	-	-	1.52	0.00199
YPEL2	-	-	-2.04	0.00002
YPEL5	-	-	-1.93	0.00001
ZBTB20	-	-	-1.86	0.00095
ZC3H6	-	-	-1.52	0.01515
ZFP14	-	-	-1.51	0.01861

Gene Symbol	Primary Pollutant Fold Change (Exposed / Unexposed)	p-value	PCA Pollutant Fold Change (Exposed / Unexposed)	p-value
ZFYVE1	-	-	-1.54	0.00139
ZKSCAN1	-	-	-1.58	0.00080
ZMYM3	-	-	-1.55	0.00055
ZNF224	-	-	-1.53	0.00287
ZNF234	-	-	-1.52	0.00360
ZNF277	-	-	-1.54	0.00502
ZNF287	-	-	-1.65	0.01718
ZNF292	-	-	-1.62	0.00032
ZNF479	-	-	1.62	0.01404
ZNF594	-	-	-1.57	0.01350
ZNF608	-	-	-1.52	0.01259
ZNF626	-	-	1.87	0.03625
ZNF654	-	-	-1.69	0.01354
ZNF704	-	-	-1.61	0.00041
ZRSR1	-	-	1.51	0.00283
ZSCAN16	-	-	-1.52	0.00316
ZSWIM6	-	-	-1.59	0.00588

Additional File 3: Networks and proteins within networks constructed using genes and related gene products associated with exposure to (A) primary or (B) photochemically altered (PCA) pollutant mixtures.

Network Number	Molecules in Network	p-value
(A) Primary Pollutants Network		
1	1,4-glucan,2' 5' oas,ACTA2,beta-hydroxyisovaleric acid,C13ORF15,C21ORF33,CCL2,CXCL16, CYP2D6, EDA2R,GFPT2,HNF4A,Interferon alpha,lymphotoxin-alpha1-beta2,NFkB (complex),NfkB-Nfkbia,NFKBIA, NKIRAS1,OAS1,PAQR5, PDGF BB,PIF,PNPT1,retinoic acid,RIOK3,SLC5A3,STAT4, SUMO4,TNAP, TNFAIP8, TNIP3,TRAPPC9,TXNIP,UBA7,ZFAND6	1E-25
(B) PCA Pollutants Networks		
1	A1CF,BCAS3,CDH1,Ck2,Cyp4f,CYP4F3,CYP4F12,DNAJB4,ELP4,FLJ11292,GLTSCR2,GSTA4,HGD,Histone h3,HNF4A,HNF4G,HOOK3,KIF20A,MYO1A,NMNAT2,PAQR5,PIGN,PPFIBP2,SERPINB8, SLC19A3,SLC22A3, SLC5A3,SPP1,STX17,TBCK,TMEM140,TPCN1,ZNF224,ZNF277,ZSCAN16	1E-52
2	ANXA3,APOH,ASPM,C14ORF106,CALCOCO1,CCDC99,CCNG1,CCNG2,CDC14B,CENPF,CYFIP2 (includes EG:26999),FAM175A,HIST1H1C,Histone H1,HMGB2, HSPH1,ING4,LETMD1,NBR1, NCAPD2,NDC80, Nuclear factor 1,NUSAP1,P2RX4,PHLDA1,PMEPA1,Ppp2c,PRKAC,ROBO1,Rsk, CMH1,SH3BGRL2,TP53, TP53INP1,UIMC1	1E-43
3	ADH6,ADH1C (includes EG:126),ADHFE1,alcohol dehydrogenase,AQP3,AR, ATP9A,C7ORF68,CA12, CCDC80,CDCA7L,CIR1,CTDSP2,DHRS3,DHRS9,EFNB2,FSH,FZD7,GPRC5B,KIF20B,KLHL24,Lh,MAOA, MIR124,NDRG1,NID2,NPY1R,oxidoreductase,PTGR2,RNA polymeraseII,SELENBP1,SSFA2,STAT4, TGFA,Vegf	1E-41
4	APH1B,BLMH,C1S,CASP4,CASP6,DTX3L,ECE1,FGFBP1,GLS,GPRC5A,HABP2,IFI35,IFIT1,IgG,IL1,Immunoglobulin,Interferon alpha,IRG,MICAL2,Mmp,NFKBIA, NGF,OAS1,PARP,PARP9,PARP14, peptidase,PIR,RPS27L (includes EG:51065), S100A3,SERPINE2,SMOX,STAT,STAT6,TNFSF10	1E-35
5	20s proteasome,26s Proteasome,Actin,Alpha tubulin,ARRB1,DIAPH2,DPT,DUSP1,EIF2C4,EIF4B, EPB41, ERBB2,ERBB3,GPAM,HIST2H2AB,HTR1A,ID1,INA,Insulin,LRIG1,MAP2K1/2,MARCKS (includes EG:4082), MATN2,MCCC1, NDRG2,Notch,NOTCH2,OPHN1,Pkc(s),PP2A,PTCH2,PTRF, SFRP1,TCR,XBP1	1E-35
6	ANKRD1,ANXA13,Cbp,CDK5RAP3 (includes EG:80279),CFHR1,CPB2,CPN1,CXCL5,DUSP5, Elastase,FGA, FGB,FGG,Fibrin,Fibrinogen,GFPT2,HAS2,IER3,IFN TYPE 1,IGKC,IP6K2,LITAF,NFkB (complex),NfkB-RelA, NFKBIZ,PPAR-RXR, Pro-inflammatory Cytokine,RAP1GAP (includes EG:5909),SLC2A12,SOAT1, Stat3-Stat3,THBD,Tlr,TNFRSF12A,TXNIP	1E-32
7	ABCC4,AXL,BCMO1,CD209,CEACAM1,CEACAM5 (includes EG:1048),EML4,Estrogen Receptor, Ferritin, Gm-csf,Growth hormone,HFE,IL12 (complex), JAK,Ldh,LDL,LRP,MIR1,MIR21 (includes EG:406991), NEDD4L,NRG,NRG1, PDCD4,PDE3A,PI3K,PRKD1,SCNN1A,SLC9A3R1,SOCS2,ST6GAL1, STAT5a/b,TNS4, WWP1,XCL1,XDH	1E-29
8	14-3-3,AHCYL1,AKAP9,ANKRA2,Calmodulin,CAMK2D,CaMKII,Creb,DET1,DGCR6,EPHX2,FOXN3, FRAS1, GABARAPL1,GABPA,GRIP1,hydrolase,ITPR2, JUN,KIF13B,LRRFIP1,MAFG,NCEH1,Nfat (family), NIPSNAP3A,PER2,PHKB,Pka,Pka catalytic subunit,Pkg,RFC1,TUBB2C,Tubulin, UNC13B,USP3	1E-29
9	Adaptor protein 2,ADD3,AKAP12,Alp,Ap1,BAMBI,BMPR2,CACNA2D1,Clathrin,DAB2,EMP1, FHL1,FSHB, Hat,hCG,Histone h4,HOXA2,IFN Beta,IL8,LIMD1, Mapk,MHC Class II (complex),NBEA (includes EG:26960),NCALD,NOSTRIN, ODC1,PDGF BB,Pias,PRKCD,SCD,SFRP4,Smad2/3,STC1,Tgf beta,TGFB2	1E-27
10	ABCA1,C3-Cfb,CALB1,Cbp/p300,CD55,CFH,CFI,collagen,DUSP4,EHHADH,FRK,GC-GCR dimer, GDF15,HDL, ITGB8,JINK1/2,Jnk,KDM3A,LXR ligand-LXR-Retinoic acid-RXR±,N-cor,NCOA2,NR0B1, NR4A1,NR5A2, NRIP1,POU1F1, RAB3B,Retinoic acid-RAR-RXR,Rxr,SCARB1,SERPINB1,T3-TR-RXR,Thyroid hormone receptor,VitaminD3-VDR-RXR,VRK2	1E-27

Network Number	Molecules in Network	p-value
11	ACTA2,Alpha catenin,Cadherin,Calpain,CIRBP,Collagen type I,Collagen(s),CPA4,CTNND1,DAPK1, DCLK1, ERK1/2,Esr1-Esr1-estrogen-estrogen,Fgf,Fgfr, FGFR4,Focal adhesion kinase,Integrin, IntegrinI,ITGA3, ITGA6,LAMC2, Laminin,Laminin1,LARGE,LMO7,LXN,MLLT4,OXTR,PDGFC,SMPD1, TFPI2,TSKU,VCAN,Vla-4	1E-25
12	Adaptor protein 1,ALS2CR8,AP1S3 (includes EG:130340),BRF2,CABYR, CDCA7L,CDKN2A,CES2 (includes EG:8824),CYP4F3,CYP4F11,DDX10,ELF3,FAM111A,FETUB,GABRE,HNF4A,HSD17B11, JAK1,LOXL2,MUT, PAN2,PDXDC2,PLEKHA8,retinoic acid,SERPINB8,SLC5A3,STRA6,SUZ12,TBRG1, TCF19, TMEM49,TRIM52, USP30,USP36,ZNF133	1E-23
13	APOL1,ATXN1,AZU1,BTN3A3,C20ORF194,CD86,CD209,CFHR3,CNN1,DPYD,DZIP3,EIF4B,EIF4E,EIF4 ENIF1,EMP1,HABP2,heparin,HIVEP1,LAMC2,LHX8,MARCH8,MIR17 (includes EG:406952), phosphatidylinositol 3,4-diphosphate, SAA2,SERPINA6,SERPINE2,SVIL,TBC1D5,TNF,TNFRSF18, TRAF2,TRIM31,TST,XCL1, ZFYVE1	1E-23
14	ARSA,ARSD,ARSE,Aryl Sulfatase,CCNG1,CHST2,CLEC7A,CYBRD1,DMXL2, ELOVL6,ethanol,GREM2, HAMP, HFE2,HIST1H3A,hydrocortisone,IL2,IL13,IL13RA2,iron,MAOA,NCAPH,norepinephrine, NPY1R,PHLDA1,PHLDA2,progesterone,SERPINA6,SESN3,SLC40A1,SLC7A2,SUMF1,TC2N,TP53INP1	1E-20
15	ALDH2,ALDH5A1,ARPC2,BDKRB2,C9ORF3,CCDC28A,CCNG2,CLMN,CSRP1,EPHA2,ERCC5,FDXR,FMN 1,FNBP1,HSD17B6,IER3,JAG2,KCNJ4,KIAA1199,LIMK2,LRBA,MIA2,PBLD,PIGF,PLK2,PPP1R13B,RHO C,RHOD,SHISA5,SHMT1,SLC23A2,SNRPD3,STRAP,TGFB1,TP73	1E-20
16	Alpha actin,AMPK,APAF1,ARID4A,BCL2L11,BNIP3L,Calcineurin protein(s), Caspase,Cdc2,CDC25C, Cyclin A,Cyclin B,Cyclin E,Cytochrome c,DHCR24, DHFR,E2f,EFHC1,FBXO32,GIP,Hdac,HIST1H2AB, HIST1H2AG, Hsp27,Hsp70,Hsp90,IQGAP2,MEF2,Mek,P38 MAPK,PFKFB3,Rb,RPS6KA5,TFDP2,WEE1	1E-20
17	ACTR5,ACTR8,ANK3,ARID5B,ATF7IP,BAZ1A,ERVK6,FCHSD2,GATS,HOOK1,INO80,INO80B,INO80D,I NO80E,KIAA1370,KIAA1377,KIAA1632,KIF20B,MARCH4,MGAM,MIR292 (includes EG: 100049711),MIR30E (includes EG:407034), MTMR11,PEX6,PLCH1,RHOBTB1,RRBP1,RUVBL1, SCAPER,SMAD2,SMAD9,SUMO1, SVEP1,ZBTB20,ZMYND11	1E-18
18	Akt,Angiotensin II receptor type 1,AREG,BDKRB1,BDKRB2,CCL2,CDCP1, CHEMOKINE,CLDN1,EGFR ligand,ERBB,ERBB4 ligand,EREG,GAB1,Gpcr, HBEGF,IFN alpha/beta,Ifn gamma,Ifnar,Ikb,IKK (complex),IL11,IL12 (family), INADL,NfkB1-RelA,NRG4,P2RY4,p70 S6k,Pik3r,Sfk,Shc,TJP2,TLR1, TLR3,Tnf	1E-18
19	BBS9,BTN3A1,C1R,CABP7,CORO2A,CYB561,EFCAB6,EGR2,ERAP1,GCA,GIP2,HERC6,Hla-abc,IFI30, IFI35, IFI44,IFI47,IFNA2,IFNG,KDM5B,LARGE,MR1,MT1X, NFE2L3,OAS3 (includes EG:4940),PARP9, RNF213,SAMD9,SP110,STAMBPL1, TBC1D10A,TMEM50B,TRIM22,ZFP36,ZKSCAN1	1E-17
20	ABCG2,ARG1,BBS1,BCL6,beta-estradiol,BLNK,CEP152,CYB5A,CYP17A1, CYP2C9,DUSP6,EIF3A,ELF1, EPB41L2,EPB41L3,FLOT1,FMO5,GSR,HBEGF,HLA-DR,HLA-DRB1,ITGBL1,KIAA0922,MED23,MEIS2, Mhc ii (family),MIR200A (includes EG:406983),PBX1,PCDH9,RFX5,SAP30,SHBG,SSBP2,TM4SF20, TMEM37	1E-15
21	ARHGAP1,BCR,C5,C1q,COTL1,EFNA1,ELMO1,EPHA2,ERK,F Actin,G protein alpha,G protein alphai, G-Actin,G-protein beta,G-protein gamma,Igm,LIMA1, LRP1,p85 (pik3r),Pdgf,Pi3-kinase,PLC,PLC gamma, PLCD4,Pld,PLD1,Rac,Rap1, Ras,Ras homolog,RND1,Sapk,SORL1,SYNE2,VIL1	1E-14
22	ADCY10,COPS8,CST1,CTPS,DDAH2,E2F1,FNIP1,GUCY1A2,GUCY1B2,GUCY1B3,GUCY2C,GUCY2D,GU CY2E,GUCY2F,HIST1H2AC,HIST1H4H (includes EG:8365),HMGCL,HSP90AA1,KRT38,LIMA1,MLEC, MLXIP,NAGA, NFKBIL1,PECR,PNN,PRPF19,SFRS18,SLC25A3,SLC2A4,SOD2,Soluble guanylate cyclase, SRRM2,TXNL1, YWHAG	1E-13
23	ASF1B,BMP1,BTBD11,COL6A1,collagen,CX3CL1,dihydrotestosterone,ETV1,FGF3,GRAMD1A,HIST1 H3E,HIST1H3J,Histone h4,HOXA9,KITLG,KLK2,MME, MMP2,MSI2,NEB,NNT,PARK7,PDGFRL, poly(ADP-ribose), RBM39,RDX, RPS6KB1,SFRP1,SFRS11,SLC16A7,SLC44A2,TERT,UBASH3B,UGCG, WSB1	1E-13
24	adenylate kinase,AK1,AK5,AK7,AK3L1,APP,ARL15,C1ORF63,C1RL,CNNM2, CRCT1 (includes EG:54544), Cytoplasmic Dynein,DYNC1LI1,DYNC2H1, FAM38B,HP,KLK6,MIR98 (includes EG:407054),MIRLET7B (includes EG:406884),NAV3,prostaglandin E2,PROZ,PRSS1 (includes EG:5644),PRSS2 (includes EG:25052),PRSS3 (includes EG:5646),PTN,PTPLAD2,REG1A, SERPINE2, SLC31A1,TCOF1,Trypsin,YPEL5, ZNF654	1E-13

Network Number	Molecules in Network	p-value
25	AHSG,ALB,AMBP,APOC3,ATG2B,CDC45L,CDKN1B,CLCA2 (includes EG:9635), COPS3,CYHR1,HAO1, HNF1A,HPX,LGALS3,MIR26A1,MIR291B,MLF1,NFIA,NFIB,NFIC,NFIX,NR1D1,NR1D2,OSBPL9,PLEK2, PRSS1 (includes EG:5644),PRSS3 (includes EG:5646),SERPING1,SLC25A27,TBC1D8B,Tcf 1/3/4, TCF7L2 (includes EG:6934),TMOD2,ZNF292,ZNF608	1E-12

Additional File 4: Transcription factors predicted to regulate genes modified by exposure to (A) primary pollutants and (B) photochemically altered (PCA) pollutant mixtures

---

Transcription Factor	TRANSFAC Accession Number	Predicted Gene Targets (Targets are Genes Identified as Differentially Expressed)	Targets' Expression Direction	p-value
(A) Primary Pollutant Mixture				
PPARalpha	M00242	*ACSM3, ACTA2, TXNIP*	Decreased	0.001
PAX	M00808	*CYP2D6, OAS1, PAQR5, SLC5A3*	Increased	0.002
HNF-1	M00132	*ACSM3, TXNIP*	Decreased	0.006
UF1H3BETA	M01068	*CYP2D6, GFPT2, NFKBIA, PAQR5, SLC5A3*	Increased	0.006
TBX5	M01020	*NFKBIA, PAQR5, SLC5A3*	Increased	0.006
SRF	M00186	*ACTA2*	Decreased	0.007
NF-Y	M00287	*ACSM3, TXNIP*	Decreased	0.009
GATA-1	M00127	*ACSM3, ACTA2*	Decreased	0.015
MZF1	M00083	*CYP2D6, SLC5A3*	Increased	0.017
Oct-1	M00137	*ACSM3, TXNIP*	Decreased	0.027
Hmx3	M00433	*ACTA2, TXNIP*	Decreased	0.029
RREB-1	M00257	*CYP2D6, GFPT2, NFKBIA, OAS1*	Increased	0.029
Evi-1	M00078	*TXNIP*	Decreased	0.035
HNF-4	M00134	*CYP2D6, OAS1, SLC5A3*	Increased	0.036
COUPTF	M01036	*CYP2D6, GFPT2, SLC5A3*	Increased	0.041
FOXO4	M00472	*ACTA2, TXNIP*	Decreased	0.043
Freac-3	M00291	*TXNIP*	Decreased	0.045
(B) PCA Pollutant Mixture				
FOXO4	M00472	*A1CF, ACTA2, ALPK1, APAF1, APH1B, ARSE, ASPM, BCL2L11, BCMO1, BDH2, BNIP3L, C4orf18, CCDC28A, CCDC34, CCDC80, CCPG1, CDC25C, CDCA7L, CENPF, CFHR3, CFI, CIR1, CORO2A, CRBN, CTDSP2, CTNND1, CTTNBP2, DEPDC4, DIAPH2, EFNA1, EIF2C4, ELOVL6, EPHX2, FBXO32, FGB, FRK, GATM, HSD17B11, ID1, IFT81, IQGAP2, KIAA1370, KIF20A, MANSC1, MARCKS, MCCC1, MGAM, MUT, NAP1L2, NCOA2, NDRG1, NR0B1, NRG4, OSBPL9, PAIP2B, PBLD, PCMTD1, PDCD4, PDE3A, PDZK1, POF1B, RBKS, RND1, RPS6KA5, SESN3, SKAP2, SLC2A12, SLC35D2, SPTLC3, ST8SIA4, SYCP2L, TFDP2, TGFB2, TLR3, TMEM140, TMEM37, TNFSF10, TXNDC16, TXNIP, VCAN, VRK2, WEE1, WWP1, ZBTB20, ZNF608, ZNF654*	Decreased	9.45E-09
HNF-1	M00132	*A1CF, ABCA12, ABCG2, ACSM3, ADH6, ALPK1, ANG, ANKRA2, ANKS4B, ANXA13, APOH, ARSE, C4orf18, C5, C7orf68, CASP4, CEACAM1, CEP152, CFHR3, CFI, CPB2, CYB5A, DAB2, ELMO1, FAM38B, FGA, FGB, FGG, FRK, GLTSCR2, HABP2, HAO1, HIST2H2BA, HNF4A, HOOK3, IP6K2, MANSC1, MARCKS, MGAM, MIA2, MTMR11, NAP1L2, NEB, NIPAL3, NIPSNAP3A, NNT, NR5A2, NRM, OPHN1, PLCH1, RPS6KA5, SAMD9, SEMA3E, SERPINA6, SLC25A27, SLC41A2, SLC7A2, SPATA7, STEAP2, TBC1D5, THG1L, TLR1, TLR3, TM4SF20, TMEM136, TMEM144, TMEM37, TXNIP, VPS13C, VRK2, WSB1, YPEL2, ZNF654, ZNF704*	Decreased	6.23E-06
TEF	M00672	*ABCA12, ADHFE1, AKAP9, AR, ARID5B, ASPM, BCAS3, BTBD11, BTN3A1, C1RL, C1S, C4orf34, C5, C5orf42, CCDC28A, CCNG2, CCPG1, CENPF, CFH, CFHR1, CFHR3, CRBN, DAB2, DMXL2, DNAJB4, DZIP3, ELOVL6, FAM149A, FAM38B, FGB, FLOT1, GRIP1, GSTM4,*	Decreased	1.18E-05

Transcription Factor	TRANSFAC Accession Number	Predicted Gene Targets (Targets are Genes Identified as Differentially Expressed)	Targets' Expression Direction	p-value
		HIST2H2AA3, HIST2H4A, HSD17B6, ING4, KIAA0922, KIAA1377, KIF13B, MANSC1, MIA2, NAP1L2, NEB, NFKBIZ, NIPAL3, NIPSNAP3A, NR0B1, NRG4, NRM, P2RX4, PBLD, PLCD4, RFX5, SEMA3E, SESN3, SFRS18, SLC16A7, SPTLC3, SYCP2L, TBC1D8B, TIGD2, TLR1, TLR3, TRIM31, TSKU, TTC28, UGT2B15, UNC119B, VPS13C, ZC3H6, ZNF292, ZNF594, ZNF654, ZSCAN16		
NF-Y	M00287	*ABCA5, ABCB6, ACSM3, ADHFE1, ALDH6A1, ARFGAP2, ARHGAP1, ARID4A, ARID4B, ASPM, ATF6B, BAMBI, BLMH, BTBD11, BTN3A1, C4orf18, C6orf130, CABYR, CCBL2, CCDC28A, CCDC34, CCNG1, CCNG2, CDC14B, CDH1, CDK5RAP3, CENPF, CEP152, CFI, CIRBP, CNNM2, CPB2, CTDSP2, CYFIP2, CYP4F12, DAPK1, DCDC2, DHCR24, ECE1, EHHADH, ELP4, FAM105A, FAM55C, FARP2, FNIP1, GCA, GK, HBP1, HFE, HIST1H2AB, HIST1H2AC, HIST2H2AA3, HIST2H2BA, HMGB2, HOOK1, HP1BP3, ID1, ITPR2, KCNT2, KDM3A, KIAA1377, KIF20A, KIF20B, KLHL24, LARGE, LETMD1, LRBA, LRIG1, MAOA, MARCKS, MBOAT1, MLEC, NCALD, NCAPD2, NDRG2, NEB, NEDD4L, NRIP1, NUSAP1, PAIP2B, PAN2, PCDH9, PDCD4, PDK2, PDZK1, PER2, PGAP2, PIGN, PLEKHH2, RFX5, RND1, RPS6KA5, SAMD9, SCD, SERPINA6, SESN3, SHMT1, SKAP2, SLC25A27, SLC46A3, SLC9A3R1, ST8SIA4, TFDP2, TMEM50B, TP53, TPCN1, TSPAN15, TXNDC16, TXNIP, WEE1, XBP1, YPEL2, YPEL5, ZBTB20, ZNF287, ZNF594, ZSCAN16*	Decreased	2.31E-05
LEF1	M00805	*A1CF, ABCA1, ABCA12, ABCA5, ACAD10, ACAD11, ACSS2, ACTA2, ADD3, ADH1C, ALDH5A1, ANG, ANXA9, APOBEC3C, APOH, AR, ARHGAP1, ARMCX3, ARSD, ARSE, BCAS3, BCL2L11, BDH2, BDKRB1, BDKRB2, BLMH, BTN3A1, BTN3A3, C14orf106, C20orf74, C5, C6orf130, C9orf3, CABYR, CCDC28A, CCDC34, CCDC80, CCNG1, CCPG1, CD99L2, CDC25C, CDK5RAP3, CEACAM1, CFHR1, CFHR3, CIR1, CNNM2, CPB2, CPN1, CTNND1, CTTNBP2, CYBRD1, CYP4F11, DAB2, DEPDC4, DET1, DHRS3, DIAPH2, DMXL2, DNAJB4, EFHC1, EFNA1, ELF3, ELMO1, ELOVL6, ELP4, EPHX2, ERAP1, ERBB2, FAM111A, FAM149A, FAM55C, FGA, FGB, FKBP5, FNIP1, FRK, FZD7, GABARAPL1, GATSL1, GCA, GK, GLTSCR2, HABP2, HAO1, HIST1H2AB, HIST1H2AG, HIST2H2AA3, HOXA2, HSD17B11, ID1, KCNT2, KDM3A, KIAA0922, KIAA1377, KIF13B, KIF20A, KLHDC2, KLHL24, LETMD1, LRBA, LRIG1, LRP1, LXN, LYRM5, MANBA, MANSC1, MARCKS, MATN2, MCCC1, MIA2, MLEC, MR1, MRAP2, MSI2, MTMR11, NAP1L2, NCALD, NEB, NEDD4L, NFIA, NIPAL3, NIPSNAP3A, NOTCH2, NOTCH2NL, NPY1R, NR1D2, NR5A2, NRIP1, NUDT7, NUSAP1, PAIP2B, PAN2, PARP14, PARP9, PBLD, PDCD4, PDE3A, PDGFC, PDGFRL, PDK2, PFKFB3, PLCD4, PLCH1, POF1B, PRKCD, RAP1GAP, RARB, RBKS, RHOBTB1, RNF213, SAMD9, SCAPER, SELENBP1, SERPINB1, SESN3, SFRS18, SH3BGRL2, SLC16A7, SLC19A3, SLC23A2, SLC25A27, SLC29A3, SLC2A12, SLC40A1, SLC41A2, SLC44A2, SPATA7, ST8SIA4, STEAP2, STX17, SVEP1, SYCP2L, SYNE2, TBC1D8B, TBCK, TC2N, TFDP2, TGFB2, TLR3, TM4SF20, TMEM136, TMEM37, TNFSF10, TRIM31, TTC39B, TTLL6, USP3, VCAN, VPS13C, VRK2, WEE1, WSB1, YPEL5, ZFP14, ZKSCAN1, ZNF224, ZNF287, ZNF594, ZNF608, ZNF654, ZNF704, ZSCAN16*	Decreased	4.37E-04
PLZF	M01075	*A1CF, ACAD11, ACSM3, ADHFE1, ANO5, APH1B, AR, ASPM, BCAS3, BCL2L15, C1S, C4orf18, C4orf34, C5orf42, CCDC34, CDK5RAP3, CENPF, CFH, CFI, CLMN, CPB2, CRBN, CYB5A, DAB2, DAPK1, ELP4, FAM149B1, GABARAPL1, GCA, GSTA4, HAO1, HIST2H4A, HMGB2, HOOK3, IFIT1, KDM3A, KIAA0922, KIAA1370, KIAA1632, KIF13B, KLHDC2, KLHL24, LYRM5, MCCC1, MGAM, MIA2, NBEAL1, NCALD,*	Decreased	4.58E-04

Transcription Factor	TRANSFAC Accession Number	Predicted Gene Targets (Targets are Genes Identified as Differentially Expressed)	Targets' Expression Direction	p-value
		NEB, NFIA, NFKBIZ, NIPSNAP3A, NUDT7, NUSAP1, PAIP2B, PBLD, PCMTD1, RHOBTB3, SEMA3E, SESN3, SH3BGRL2, SLC16A7, SLC41A2, SSBP2, TBC1D5, TIGD2, TM4SF20, TMEM136, ZNF292, ZNF608, ZNF654		
Oct-1	M00161	*ABCB6, ACSM3, ADH6, ADHFE1, ANXA13, ARFGAP2, ASPM, C4orf18, CASP4, CCDC34, CEACAM1, CEP70, CFHR1, CORO2A, CRBN, DAB2, DNAJB4, ELF3, ENTPD5, FAM105A, FGFR4, FOXN3, GABARAPL1, GATM, HABP2, HIST1H2AC, HIST2H2AA3, HIST2H2BA, KCNT2, KIAA1109, LHX8, LYRM5, MARCKS, MBOAT1, MGAM, MTMR11, NBEAL1, NEB, OSBPL9, PARP9, PDZK1, PLCD4, POF1B, RHOBTB1, RHOBTB3, SCD, SCNN1A, SEMA3E, SLC19A3, STAT4, TBC1D5, TBC1D8B, TLR1, TSPAN15, WSB1, WWP1, YPEL5, ZNF292*	Decreased	6.30E-04
HOXA4	M00640	*ANKRD1, CALB1, CCBE1, CLDN1, CXCL5, EREG, FBXO40, ITGA3, ITGBL1, LMO7, SERPINB8*	Increased	8.03E-04
HNF-1	M00206	*A1CF, ABCA12, ADH6, ALPK1, ANG, ANKS4B, ANXA13, BAMBI, BDKRB1, C4orf18, C5orf42, CALCOCO1, CCNG2, CTNND1, DAB2, DIAPH2, DYNC2H1, FGA, FRK, HIST2H4A, HP1BP3, HSD17B11, ING4, IQGAP2, LXN, LYRM5, MIA2, NAP1L2, NEB, NIPAL3, NNT, NRM, PCCA, PLCH1, RARB, SAMD9, SLC19A3, SLC23A1, SLC2A12, SLC35D2, SPATA18, SPTLC3, STRA6, THG1L, TLR3, TM4SF20, TMEM37, UGT2B15, VPS13C, YPEL2, ZNF654*	Decreased	0.001
CDX	M00991	*A1CF, ABCA5, ACAD11, ACSM3, ALDH6A1, ANG, ARID4B, ARID5B, ARMCX3, ASPM, BBS9, BCMO1, BDH2, BTN3A3, C5orf42, CCDC28A, CENPF, CFH, CFHR1, CFHR3, CRBN, DEPDC6, DIAPH2, DNAJB4, DPYD, DYNC2H1, ENTPD5, FAM149B1, FAM175A, FGB, HBP1, HFE, HIST1H2AC, HIST2H4A, HNF4G, HOOK3, HOXA2, KCNK5, LYRM5, MARCKS, MCCC1, MIA2, NAP1L2, NBEAL1, NCALD, NRG4, PCMTD1, POF1B, SESN3, SHMT1, SKAP2, SLC16A7, SLC19A3, SLC41A2, SPATA18, SPP1, ST8SIA4, TBC1D5, TLR1, TLR3, TMEM136, TXNIP, UGT2B15, WWP1, XBP1, ZBTB20, ZC3H6, ZNF224, ZNF287, ZNF292, ZNF608, ZNF654, ZSCAN16*	Decreased	0.001
SRY	M00148	*A1CF, ABCA1, ABCG2, ADH1C, ANG, ANKRA2, ANXA13, APH1B, APOH, ARFGAP2, ARID5B, ARSD, ARSE, AS3MT, ASPM, BAMBI, BBS9, BCMO1, BDH2, BLMH, BMPR2, BNIP3L, BTN3A1, BTN3A3, C1orf63, C1S, C5, C7orf68, CCDC28A, CCNG2, CDC14B, CDC25C, CDCA7L, CEP152, CFHR3, CIR1, CORO2A, CRBN, CTTNBP2, DNAJB4, EFHC1, EFNA1, EIF4B, ELMO1, ELOVL6, EPHX2, ERAP1, FAM111A, FAM149B1, FAM38B, FBXO32, FGB, FGG, FOXN3, FRK, FZD7, GABARAPL1, GATM, GATSL1, GK, HABP2, HBP1, HFE, HIST1H2AC, HIST2H4A, HP1BP3, IFIT1, IFT81, INADL, ING4, IP6K2, IQGAP2, KIAA1632, LITAF, LYRM5, MARCKS, MCCC1, MGAM, MR1, NAP1L2, NCALD, NDRG1, NFIA, NPY1R, NRG4, NUDT7, PAIP2B, PBLD, PCMTD1, PDCD4, PPFIBP2, PTPLAD2, RFX5, RND1, RNF213, SAMD9, SASH1, SERPINA6, SFRP4, SFRS18, SH3BGRL2, SKAP2, SLC16A7, SLC2A12, SLC35D2, SLC7A2, SMPD1, SPATA18, SPATA7, SPP1, SPTLC3, STAT6, STX17, SYCP2L, SYNE2, TBC1D8B, TFDP2, THG1L, TLR1, TLR3, TM4SF20, TMEM136, TMEM144, TMEM37, TMEM50B, TPCN1, TRIM31, TST, TXNIP, UNC13B, VPS13C, VRK2, ZBTB20, ZNF287, ZNF704*	Decreased	0.002
Bach2	M00490	*CCDC99, CEACAM5, COTL1, CST1, DCLK1, FHL1, FLI1, IL11, IL8, KRT80, LAMC2, NID2, NR4A1, OR4C6, PSG8, RAB3B, RNF182, SEMA3C, SERPINB8, SNRPN, TFPI2, TMEM171*	Increased	0.003
S8	M00099	*A1CF, ACAD10, ADH1C, ANKS4B, ASPM, BAMBI, C1S, C5, C5orf42, CACNA1D, CALCOCO1, CDCA7L, CENPF, CYFIP2, DAB2, DCDC2, DIAPH2, DMXL2, DNAJB4, ELMO1, ELP4, FAM55C, FNIP1, HAO1,*	Decreased	0.003

Transcription Factor	TRANSFAC Accession Number	Predicted Gene Targets (Targets are Genes Identified as Differentially Expressed)	Targets' Expression Direction	p-value
		HERC6, HFE, HIST1H2AC, IFIT1, KCNT2, KIF13B, LYRM5, MARCKS, MGAM, NAP1L2, NBEAL1, NCALD, NCAPD2, NIPSNAP3A, NRG4, PDZK1, PECR, PIR, PLCH1, RHOBTB3, SAMD9, SASH1, SEMA3E, SH3BGRL2, SLC16A7, TIGD2, TLR1, TMEM60, TTLL6, YPEL2, ZC3H6, ZFYVE1, ZNF704		
C/EBPgamma	M00622	*ADH1C, ADH6, ALPK1, ANXA13, APOH, ARSD, BCMO1, C9orf3, CABYR, CACNA1D, CALCOCO1, CASP4, CCDC28A, CDC14B, CEP70, CFHR1, CRBN, CYBRD1, DAB2, DCDC2, DMXL2, DZIP3, FAM111A, FGB, IQGAP2, KIAA0922, KIAA1377, KIF13B, LHX8, MCCC1, MEIS2, NEK11, NRG4, PAN2, PBLD, PCMTD2, PDCD4, PECR, PLCH1, PLEKHH2, PTGR2, RPS6KA5, SEMA3E, SLC2A12, SLC41A2, SPG11, SPP1, STAT4, SYCP2L, TMEM140, UGT2B15, VPS13C, ZNF292*	Decreased	0.003
Nkx6-2	M00489	*ADH1C, ADH6, ANG, ANKS4B, ASPM, BBS9, C5, C5orf42, CALCOCO1, CCDC34, CEACAM1, CEP152, CYB5A, DMXL2, ELMO1, ENTPD5, EPHX2, FAM38B, FAM55C, FGB, FGG, FLOT1, FNIP1, FRK, HFE, HIST2H4A, HNF4G, KIAA1109, KIAA1370, KIAA1712, LHX8, LYRM5, MANSC1, MARCKS, MCCC1, MIA2, NAP1L2, NFIA, NIPAL3, NIPSNAP3A, NNT, NR0B1, PAN2, PDCD4, PLCH1, RPS6KA5, SCAPER, SESN3, SH3BGRL2, SHMT1, SORL1, TBC1D5, TMEM136, TNFSF10, TXNIP, UGT2B15, ZC3H6, ZKSCAN1, ZNF292*	Decreased	0.005
FOXJ2	M00423	*A1CF, ABCG2, ADH1C, ANXA13, ARID5B, ARMCX3, ASPM, BLMH, BTN3A3, C4orf18, C4orf34, C5orf42, CCDC28A, CCDC80, CFHR3, CFI, CPB2, DPYD, DTX3L, DYNC2H1, DZIP3, FGB, FGG, FNIP1, HSD17B11, KIAA1370, KIF20A, MANSC1, MARCKS, MGAM, MUT, NAP1L2, NBEAL1, NFIA, NFKBIZ, NIPSNAP3A, NRG4, NUDT7, PBLD, PCCA, POF1B, SHMT1, SKAP2, SLC16A7, SORL1, SPP1, SYCP2L, TBC1D5, TM4SF20, YPEL5, ZMYM3, ZNF292, ZNF654*	Decreased	0.005
C/EBPalpha	M00116	*ADH1C, ADHFE1, ALPK1, ARMCX3, C9orf3, CCNG2, CFH, CFHR1, CPB2, DEPDC4, DHRS3, DMXL2, ELP4, ENTPD5, FAM38B, FGA, GCA, HIST1H2AB, HNF4G, INADL, ING4, KIAA0922, LHX8, LRBA, LXN, MANBA, MANSC1, MR1, MUT, NBEAL1, NDRG2, NEK11, PBLD, PCMTD2, PFKFB3, SLC7A2, ST8SIA4, STEAP2, TIGD2, TLR3, TSKU, ZC3H6, ZNF292, ZNF608, ZNF654, ZSCAN16*	Decreased	0.006
STAT5A	M00499	*ANKRD1, ANXA10, C6orf191, CCL2, CLDN1, CPA4, FSTL5, GABRA5, GPRC5A, HBEGF, ITGA3, ITGBL1, KRT38, OR4C6, OR51B4, OXTR, SAMD7, SERPINE2, SLC5A3, SPRR2B, TFPI2, WDR69*	Increased	0.007
NF-AT	M00302	*ABCA1, ACSS2, ADHFE1, ALS2CR8, BTN3A1, C20orf74, C5orf42, CCBL2, CCDC80, CEP152, CFHR3, CYB5A, DIAPH2, EIF2C4, ELOVL6, ELP4, FAM105A, FAM55C, FGFR4, HOXA2, HP1BP3, LIMA1, LRBA, MARCKS, NFKBIZ, NR5A2, PAIP2B, PER2, SFRP4, SLC25A27, SLC29A3, STAT4, STX17, SYNE2, TBC1D5, TGFB2, TLR1, TM4SF20, TMEM144, TNFSF10, TP53INP1, TTC39B, UGT2B15, ZNF287, ZNF654*	Decreased	0.009
NF-AT	M00935	*ABCA5, ARID4B, ARRB1, C4orf18, CEP152, CEP70, CYP4F3, DMXL2, FBXO32, HFE, HIST2H2BA, KIAA1377, LRBA, MARCKS, NIPSNAP3A, NNT, NUSAP1, OPHN1, PBLD, PDGFRL, RARB, SEMA3E, SLC16A7, SYNE2, TBC1D5, TCP11L2, TMEM144, TP53, TTC39B, YPEL2, ZNF654*	Decreased	0.009
Freac-4	M00292	*AMPD1, ANKRD1, ANXA10, CCL2, CD55, DHRS9, DUSP4, EMP1, GDF15, HAS2, LMO7, NAV3, NMNAT2, OXTR, PHLDA1, RNF182, SNRPN, SPRR2B*	Increased	0.010
TFIIA	M00707	*AKAP12, ASAM, AXL, CLDN1, CXCL5, DGCR6, DHX37, ITGA3, KIAA1199, ODC1, PMEPA1, STAMBPL1, TGFA, XCL1*	Increased	0.011
PLZF	M01075	*AKAP12, ANKRD22, ANXA10, CCBE1, CPA4, DHX37, DUSP1, EMP1, FBXO40, FSHB, GPRC5A, HTR3D, IL11, IL8, ITGBL1, KRT38, OR4C6,*	Increased	0.013

Transcription Factor	TRANSFAC Accession Number	Predicted Gene Targets (Targets are Genes Identified as Differentially Expressed)	Targets' Expression Direction	p-value
		OR51B4, OXTR, POU1F1, SAMD7, SOCS2		
HFH-1	M00129	*A1CF, ACTA2, ADHFE1, APH1B, APOH, ARID4B, ARID5B, BDH2, BDKRB2, C14orf106, C4orf18, CASP4, CCDC28A, CLMN, CRBN, DAB2, DEPDC6, DMXL2, DNAJB4, DYNC2H1, ENTPD5, ERAP1, FAM105A, FAM175A, FGB, FGG, FNIP1, HAO1, HIST1H2AC, HIST2H2AA3, HSD17B11, IFT81, ITPR2, LITAF, LXN, MGAM, MSI2, NAP1L2, NCOA2, NIPSNAP3A, NR5A2, NRG4, NUDT7, PARP9, PBLD, PCMTD1, RHOBTB1, RHOBTB3, SASH1, SEMA3E, SKAP2, SLC16A7, SLC19A3, SLC2A12, SLC7A2, SORL1, SPTLC3, SYCP2L, SYNE2, TBC1D8B, TM4SF20, TMEM136, TMEM140, UNC119B, VRK2, WWP1, ZMYM3, ZNF654*	Decreased	0.013
TBP	M00471	*AMPD1, ANKRD22, ANXA10, CALB1, CCBE1, CPA4, DHRS9, DPT, FBXO40, FLI1, ITGBL1, NR4A1, OR4C6, OR51B4, POU1F1, SPRR2B, SSFA2, STAMBPL1, TFPI2*	Increased	0.014
ZID	M00085	*AP1S3, CDCP1, CEACAM5, CPZ, CTPS, DND1, FSTL5, GFPT2, ITGB8, ITGBL1, LRRFIP1, LYAR, MICAL2, NMNAT2, PTRF, RAB3B, SLCO4A1, SOCS2, TMEM171, TMEM74, XDH*	Increased	0.014
HMGIY	M01010	*ABCG2, ACTA2, ANKS4B, ANXA4, ARFGAP2, BCL2L11, BDKRB1, BTN3A1, CCBL2, CCNG1, CFH, CPB2, DHRS3, DYNC2H1, ELOVL6, ELP4, FGB, HIST2H2BA, IQGAP2, KCNT2, KDM3A, KIAA0922, KIAA1377, KIF20A, LHX8, LYRM5, MCCC1, NFIA, NPY1R, NR5A2, NUDT7, PAIP2B, PLCH1, SAMD9, SFRS18, SLC35D2, SLC41A2, SYCP2L, TBC1D8B, TIGD2, TLR3, TM4SF20, TPCN1, WEE1, XBP1, ZNF608*	Decreased	0.014
TATA	M00252	*AMPD1, DPT, FGFBP1, GDF15, GLIPR1, IL8, JUN, KIAA1199, NR4A1, SERPINB8, SNRPN, STAMBPL1, STC1 ,*	Increased	0.015
SOX	M01014	*C6orf191, CALB1, CCBE1, DUSP4, FGFBP1, FLI1, FRMD3, ITGB8, ITGBL1, NAV3, NMNAT2, NRG1, OAS1, SOCS2, TNFRSF12A*	Increased	0.015
NF-E2	M00037	*ANKRD22, AQP3, CCDC99, COTL1, CPZ, DGCR6, FGFBP1, GLIPR1, HAS2, HTR3D, IL11, KIAA1199, KRT80, NAV3, OR51B4, RAB3B, SERPINB8, TNS4*	Increased	0.017
DEC	M00997	*AKAP12, ASAM, C10orf114, CD55, CEACAM5, CSRP1, DND1, FHL1, IER3, ITGA6, MYEOV, SERPINE2, SLCO4A1, SSFA2, THBD, TNS4*	Increased	0.018
GATA-1	M00346	*CXCL5, FGFBP1, FSTL5, GPRC5A, GREM2, HAS2, HSPH1, LAMC2, POU1F1, SERPINE2, TGFA, WDR69*	Increased	0.018
FAC1	M00456	*A1CF, ACSM3, ALS2CR8, APOH, ARID5B, BCMO1, C14orf106, C1orf63, C20orf19, CALCOCO1, CCDC34, CCNG2, CDCA7L, CFHR1, CLMN, CRBN, CYB5A, DAB2, DIAPH2, EFNA1, EML4, ENTPD5, ERAP1, FAM111A, FAM38B, GABARAPL1, GATSL1, GLTSCR2, HBP1, HFE, HIST1H2AG, HNF4G, IFIT1, LITAF, MUT, NCAPD2, NEDD4L, NR5A2, NRG4, NRIP1, PDCD4, PECR, POF1B, SFRP4, SFRS18, SH3BGRL2, SKAP2, SLC23A2, SLC7A2, SPP1, SPTLC3, STEAP2, STX17, SYCP2L, SYNE2, TBC1D5, TBC1D8B, TGFB2, TLR1, TMC7, TMEM144, TMEM50B, VCAN, VPS13C, YPEL5, ZFYVE1, ZNF287, ZNF608*	Decreased	0.018
SREBP-1	M00221	*AMPD1, AQP3, CEACAM5, EREG, FGFBP1, GABRA5, GPRC5A, IL8, ITGBL1, JUN, LRRFIP1, PMEPA1, PSG8, SLCO4A1, TFPI2, THBD, TMEM171, XCL1*	Increased	0.021
HNF4	M01032	*ABCA1, ACSM3, AHCYL1, AKAP9, ALS2CR8, ANKRA2, ANKS4B, ANXA4, ANXA9, APOBEC3C, APOH, ARID4B, ARID5B, BCAS3, BCL2L15, BDKRB1, BDKRB2, C1RL, C4orf18, C4orf34, C5, C6orf130, CABYR, CASP4, CCDC34, CD99L2, CDH1, CEP70, CFHR1, CFHR3, CFI, CIR1, CLMN, CYHR1, CYP4F11, CYP4F12, CYP4F3, DAB2, DEPDC4, DHCR24, DNAJB4, DPYD, EFHC1, ELF3, ELP4, EPB41L4A, FAM111A, FAM55C, FGG, FKBP5, GATM, GCA, GIP, GRIP1, HAO1, HBP1,*	Decreased	0.023

Transcription Factor	TRANSFAC Accession Number	Predicted Gene Targets (Targets are Genes Identified as Differentially Expressed)	Targets' Expression Direction	p-value
		HERC6, HFE, HIST1H2AG, HMGB2, HSD17B11, ID1, IFI35, IFIT1, IFT81, KCNT2, KIAA1161, KIAA1370, KIF20A, KLHDC2, KLHL24, LHX8, LIMD1, LXN, MARCKS, MATN2, MCCC1, MEIS2, MLEC, MRAP2, MUT, NAP1L2, NBEA, NCAPD2, NDRG1, NEB, NEDD4L, NOTCH2, NOTCH2NL, NPY1R, NRG4, P2RX4, PAN2, PARP14, PARP9, PBLD, PCMTD1, PDGFC, PECR, PER2, PTGR2, RAP1GAP, RARB, RFX5, RHOBTB1, RND1, RNF213, SCNN1A, SEMA3E, SERPINB1, SESN3, SFRS18, SH3BGRL2, SLC23A2, SLC29A3, SLC35D2, SLC40A1, SLC41A2, SOAT1, SPATA7, SPP1, SPTLC3, STAT6, STRA6, SYCP2L, TBC1D5, TGFB2, TM4SF20, TMEM140, TMEM37, TMEM50B, TPCN1, TRIM31, TSKU, TTC28, TTLL6, USP3, VCAN, VPS13C, ZFYVE1, ZKSCAN1, ZNF277, ZNF287, ZNF292, ZNF608, ZNF704		
Pax-4	M00377	*ACAD11, ADH6, ARMCX3, BCAS3, BTN3A1, C20orf19, C5, CACNA1D, CCDC34, CCPG1, CEACAM1, DCDC2, DEPDC6, DET1, DMXL2, ENTPD5, ERBB3, FAM105A, FNIP1, HIST1H2AG, HIST2H2BA, KIAA1712, LHX8, MIA2, NBEAL1, NR0B1, PFKFB3, PLCH1, POF1B, SASH1, SFRP4, SLC2A12, SPP1, STEAP2, TBC1D5, TFDP2, UGT2B15, WEE1, YPEL2, ZNF292, ZNF704*	Decreased	0.023
GATA-1	M00128	*ACSS2, ADH1C, ADH6, APOBEC3C, AR, BCL2L15, BDKRB2, C5, CASP4, CCDC80, CCNG1, CDCA7L, CYFIP2, CYP4F11, CYP4F12, CYP4F3, DPYD, EFHC1, ELP4, FRK, GIP, HNF4G, KIAA1109, KIF20A, LYRM5, MYO1A, NAGA, RHOBTB3, SLC40A1, STAT6, TBC1D8B, TIGD2, TNFSF10, ZNF292*	Decreased	0.023
Ncx	M00484	*ABCA1, ACTA2, APAF1, C20orf194, CASP4, CDCA7L, CENPF, CEP152, CNNM2, CYB5A, CYP4F11, DIAPH2, DTX3L, EFHC1, GPRC5B, GSTA4, HBP1, ITPR2, KIAA1712, LIMA1, MANBA, NNT, NPY1R, NRG4, OSBPL9, PGAP2, PIGN, RHOBTB3, RNF213, SFRS18, SLC25A27, STEAP2, SVEP1, TMEM60, USP3, WWP1, YPEL2, ZBTB20, ZKSCAN1, ZNF704*	Decreased	0.023
FOXP1	M00987	*ABCA1, ABCA12, ADH1C, ADH6, ALPK1, ANKRA2, ANXA13, APH1B, AR, ARFGAP2, ARID5B, ARMCX3, ASPM, BBS9, BDH2, BNIP3L, BTN3A3, C1S, C4orf18, C5orf42, CACNA1D, CCDC28A, CCPG1, CEP152, CFH, CFI, CRBN, CTNND1, DAB2, DEPDC6, DYNC2H1, EFNA1, ELP4, EML4, ENTPD5, ERAP1, ERBB3, FAM111A, FAM149B1, FGG, FNBP1L, FNIP1, GABARAPL1, GRIP1, HAO1, HIST1H2AC, HIST2H4A, HOOK3, HSD17B11, HSD17B6, IFIT1, IFT81, INADL, ING4, ITPR2, KCNK5, KIAA1370, KIAA1632, LIMD1, LITAF, LYRM5, MANSC1, MARCKS, MCCC1, MIA2, MR1, NAP1L2, NBEA, NBEAL1, NCALD, NPY1R, NR5A2, NRG4, OSBPL9, PBLD, PCMTD1, PDCD4, PECR, PLCH1, POF1B, PTGR2, PTPLAD2, SASH1, SKAP2, SLC16A7, SLC25A27, SLC35D2, SLC41A2, SOAT1, SPATA18, SPP1, SPTLC3, STEAP2, SYCP2L, SYNE2, TBC1D5, TBC1D8B, TGFB2, THG1L, TIGD2, TLR1, TLR3, TM4SF20, TMC7, TMEM144, TMEM37, TMEM50B, TXNIP, UGT2B15, WEE1, WWP1, YPEL2, ZNF224, ZNF287, ZNF292, ZNF608, ZNF654*	Decreased	0.023
ELF-1	M00746	*ABCA12, ADH1C, AHCYL1, ARHGAP1, ARID4B, ARRB1, BCL2L15, BNIP3L, C1RL, C20orf74, C5orf42, CABYR, CCBL2, CCNG1, CFH, CFHR1, CTNND1, DAB2, DHCR24, DMXL2, DZIP3, EHHADH, ELMO1, ELOVL6, FGA, FRK, IQGAP2, LARGE, MARCKS, MCCC1, MR1, MUT, NAP1L2, NEK11, NR0B1, NUDT7, OPHN1, PAN2, PDK2, PGAP2, PLCD4, PLCH1, RHOBTB1, SAMD9, SESN3, SFRP4, SLC35D2, SOAT1, STAT6, TBC1D5, TGFB2, TLR1, TM4SF20, TTLL6, ZBTB20, ZNF608*	Decreased	0.026
HNF-6	M00639	*ABCA12, ABCG2, ADHFE1, ALDH5A1, ANO5, APOH, ARID4B, ARID5B, BLMH, BTN3A3, CACNA1D, CCBL2, CFH, DAB2, DEPDC6, DET1, DYNC2H1, EFHC1, ENTPD5, FGB, FGG, GCA, HABP2, HBP1,*	Decreased	0.027

Transcription Factor	TRANSFAC Accession Number	Predicted Gene Targets (Targets are Genes Identified as Differentially Expressed)	Targets' Expression Direction	p-value
		HIST1H2AC, HNF4A, HNF4G, KCNK5, KIAA0922, KIAA1377, KIAA1712, MARCKS, MCCC1, MTMR11, NAP1L2, NCALD, NEB, NNT, NRG4, PDZK1, PLCH1, POF1B, RFX5, SLC16A7, SLC25A27, SLC41A2, SSBP2, STX17, UGT2B15, ZNF292		
Alx-4	M00619	*ADH1C, BTBD11, CEP152, FGFR4, FNBP1L, HIST2H2AA3, HMGB2, HOOK3, MARCKS, NR1D2, PDCD4, PIGN, PLEKHH2, ZNF654*	Decreased	0.029
RP58	M00532	*AQP3, CCL2, CSRP1, DHX37, DPT, GREM2, IL8, ITGA3, ITGA6, ITGBL1, KRT38, SMOX, SNRPN, XDH*	Increased	0.034
Lyf-1	M00141	*ABCB6, APOH, ARSE, ATG2B, BLMH, CCBL2, CDCA7L, ERBB2, FAM111A, FAM38B, FGG, FRAS1, GRIP1, HNF4A, KIAA0922, LYRM5, MYO1A, NBEA, NRG4, PBLD, PGAP2, PLD1, PTGR2, SLC25A27, SLC40A1, SULT2B1, SYNE2, TC2N, TJP2, TSPAN15, ZKSCAN1, ZMYM3, ZNF608*	Decreased	0.036
En-1	M00396	*ABCA1, ACSM3, ALPK1, C1S, C20orf194, CASP4, CDCA7L, CEP152, CRBN, DIAPH2, DTX3L, EHHADH, ELOVL6, FAM105A, FAM38B, FMO5, GCA, HABP2, HIST2H4A, INADL, MANBA, MUT, NPY1R, PAIP2B, PARP14, SH3BGRL2, SLC25A27, SPTLC3, SVEP1, SYCP2L, TMEM60, UGT2B15, YPEL2, ZKSCAN1*	Decreased	0.038
ER	M00191	*AQP3, CLDN1, FBXO40, GABRA5, GFPT2, GLS, GPRC5A, LAMC2, LRRFIP1, P2RY4, PAQR5, PHLDA1, SEMA3C, SLCO4A1, WDR69, XDH*	Increased	0.039
AFP1	M00616	*ALDH5A1, ASPM, BTN3A1, C10orf57, C6orf130, CACNA1D, CALCOCO1, CCNG1, CORO2A, CTNND1, EHHADH, FGB, FOXN3, FRAS1, ID1, KCNT2, KIAA1712, KIF13B, LIMA1, LYRM5, MARCKS, MLEC, NAP1L2, NIPAL3, NIPSNAP3A, NUDT7, PBLD, PFKFB3, PLCH1, POF1B, RARB, RFX5, SFRP4, SFRS18, SLC16A7, SLC41A2, TIGD2, TLR1, TMEM144, TXNIP, UGT2B15, ZC3H6, ZMYM3, ZNF292, ZNF654*	Decreased	0.041
Evi-1	M00082	*ABCG2, ADH1C, ADH6, ALS2CR8, ANG, ANXA13, APOH, ASPM, BDKRB2, BNIP3L, C1orf63, C5orf42, CCNG2, CEP152, DYNC2H1, FAM55C, HIST2H2AA3, HIST2H2BA, KIAA1632, LYRM5, MANSC1, MBOAT1, MCCC1, MR1, MSI2, MYO1A, NCALD, NEB, NRG4, PIGN, PLCH1, RFX5, SAMD9, SLC41A2, SLC7A2, STAT6, TM4SF20, TMEM144, TXNDC16, TXNIP, WWP1, ZNF292, ZNF608*	Decreased	0.042
Gfi-1	M00250	*ACSS2, AS3MT, BCL2L11, BDH2, BLMH, CABYR, CDK5RAP3, CEP152, CTNND1, CTTNBP2, DIAPH2, ERBB3, FARP2, FGB, FGFR4, GCA, GRIP1, HBP1, HIST2H4A, HNF4A, INADL, KDM3A, KIAA1377, LIMD1, LRP1, LXN, NNT, NRG4, PAIP2B, PECR, PLCD4, POF1B, RHOBTB3, SASH1, SCAPER, SH3BGRL2, SKAP2, SLC2A12, STEAP2, TMC7, VPS39, YPEL5, ZBTB20, ZNF292*	Decreased	0.042
LEF1	M00805	*AKAP12, AKR1B1, ANKRD1, ANXA10, AQP3, C10orf114, CD55, CLDN1, CPA4, CPZ, CST1, CTPS, DND1, DUSP5, EMP1, EPHA2, EREG, FGFBP1, FRMD3, FSHB, FSTL5, GABRA5, GDF15, GFPT2, GLIPR1, GPRC5A, HAS2, HBEGF, HTR3D, IL8, ITGA3, ITGBL1, KIAA1199, KRT38, KRT80, LAMC2, NAV3, NOSTRIN, NRG1, ODC1, PMEPA1, POU1F1, PSG8, RAB3B, S100A3, SAMD7, SEMA3C, SERPINB8, SLCO4A1, SOCS2, SPRR2B, STC1, THBD, TNS4, UBASH3B, XCL1, XDH*	Increased	0.043
OTX	M01117	*ANKRA2, C14orf106, CASP6, CRBN, CTNND1, CTTNBP2, DIAPH2, ELF3, FAM149B1, FGB, HIST1H2AG, HIST2H4A, IP6K2, KDM3A, KLHDC2, MARCKS, NFKBIZ, NIPSNAP3A, NUDT7, PCDH9, PCMTD1, PECR, PLCD4, SESN3, TBCK, TSPAN15, WEE1, ZNF292*	Decreased	0.047
CDP	M00102	*ABCA1, ACAD11, ALPK1, ALS2CR8, APOH, ARFGAP2, ARID5B, ARMCX3, BAMBI, BCMO1, BNIP3L, BTN3A3, C4orf34, CCBL2, DAB2, DEPDC6, DET1, DYNC2H1, DZIP3, EFHC1, FGB, FGG, HIST1H2AG, HIST2H4A, HNF4A, HNF4G, HOXA2, INADL, KIAA1109, MARCKS,*	Decreased	0.047

Transcription Factor	TRANSFAC Accession Number	Predicted Gene Targets (Targets are Genes Identified as Differentially Expressed)	Targets' Expression Direction	p-value
		MCCC1, MTMR11, NIPSNAP3A, NRG4, NUSAP1, PDGFC, PDZK1, PIGN, SASH1, SEMA3E, SERPINA6, SFRS18, SLC16A7, SLC41A2, STEAP2, STX17, TBC1D5, TBCK, TLR3, TM4SF20, TMEM136, TXNDC16, UGT2B15, ZNF654		
c-Ets-1	M00339	*ANXA3, CALB1, CD55, CSRP1, CXCL5, DUSP1, LAMC2, LMO7, NCEH1, NMNAT2, NOSTRIN, NR4A1, NRG1, PHLDA1, PLEK2, PSG8, S100A3, SOCS2, SSFA2, TGFA, THBD, TMEM171, TNS4*	Increased	0.049

Additional File 5: miRNAs significantly (p-value < 0.005, FDR < 0.005) changed ≥ 1.5-fold due to formaldehyde exposure

---

miRNA	Formaldehyde/Control Ratio
miR-33	-5.48
miR-450	-3.57
miR-330	-2.43
miR-181a	-2.11
miR-10b	-2.11
miR-422b	-2.02
miR-532	-1.84
miR-501	-1.82
miR-487b	-1.80
miR-20a	-1.80
miR-34a	-1.73
miR-93	-1.72
miR-106b	-1.71
miR-137	-1.71
miR-103	-1.70
miR-301	-1.70
miR-10a	-1.70
miR-126	-1.70
miR-17-5p	-1.69
miR-107	-1.69
miR-454-3p	-1.69
miR-140	-1.68
miR-101	-1.68
miR-130a	-1.68
miR-19a	-1.67
miR-26a	-1.67
miR-19b	-1.67
miR-106a	-1.66
miR-99a	-1.66
miR-18a	-1.66
miR-424	-1.65
let-7a	-1.65
miR-20b	-1.65
miR-25	-1.64
miR-590	-1.64
miR-15b	-1.64
let-7b	-1.63
miR-660	-1.63
miR-27b	-1.63
miR-194	-1.62
miR-361	-1.62
miR-192	-1.62
miR-215	-1.62

miRNA	Formaldehyde/Control Ratio
miR-374	-1.62
miR-15a	-1.62
let-7c	-1.61
miR-148b	-1.60
miR-181b	-1.60
miR-425-5p	-1.60
miR-23b	-1.60
let-7d	-1.59
miR-28	-1.58
miR-125a	-1.58
miR-181d	-1.58
miR-130b	-1.58
miR-185	-1.58
miR-324-5p	-1.58
miR-9*	-1.57
miR-452	-1.57
miR-565	-1.57
miR-26b	-1.57
miR-152	-1.57
miR-16	-1.57
miR-650	-1.56
miR-21	-1.56
miR-9	-1.56
miR-186	-1.56
miR-151	-1.56
miR-582	-1.55
let-7e	-1.55
let-7g	-1.55
miR-98	-1.55
miR-224	-1.55
miR-23a	-1.54
miR-27a	-1.54
miR-362	-1.54
let-7f	-1.54
miR-17-3p	-1.53
miR-550	-1.53
miR-29b	-1.53
miR-182	-1.53
miR-100	-1.51
miR-509	-1.51
miR-652	-1.51
miR-331	-1.51
miR-34b	-1.51
miR-189	-1.51
let-7i	-1.51
miR-24	-1.50

Additional File 6: Predicted miRNA targets for miR-33, miR-330, miR-181a, and miR-10b

Symbol	GenBank	Entrez Gene ID	miRNAs with Overlapping Targets
		1. miR-33	
ABCA1	NM_005502	19	miR-330
ABCE1	NM_001040876	6059	
APPBP2	NM_006380	10513	
ARID5B	NM_032199	84159	
ASAP1	NM_018482	50807	
B3GALT2	NM_003783	8707	
C11ORF41	NM_012194	25758	
CACNA1C	NM_001129834	775	
CDC42BPA	NM_003607	8476	
CDK6	NM_001259	1021	
CLSPN	NM_022111	63967	
CROT	NM_021151	54677	
CSNK1D	NM_001893	1453	
DPY19L1	NM_015283	23333	
DSC3	NM_001941	1825	miR-181a
DYRK3	NM_001004023	8444	
EBF1	NM_024007	1879	
EEA1	NM_003566	8411	
EN2	NM_001427	2020	
ESCO1	NM_052911	114799	
FAM46C	NM_017709	54855	
FGA	NM_000508	2243	
FUT9	NM_006581	10690	miR-181a
GLCCI1	NM_138426	113263	
GOPC	NM_020399	57120	
GRIA3	NM_007325	2892	miR-330
HADHB	NM_000183	3032	
HBS1L	NM_006620	10767	
HIPK2	NM_001113239	28996	
HMGA2	NM_003483	8091	
ING3	NM_019071	54556	
KCNMA1	NM_001014797	3778	
KIAA2018	NM_001009899	205717	
KIF3C	NM_002254	3797	
LCA5	NM_181714	167691	miR-10b
LIPI	NM_198996	149998	
LOC152742	XM_001128848	152742	
LPP	NM_005578	4026	
MMP16	NM_005941	4325	
MSR1	NM_138715	4481	
NAP1L2	NM_021963	4674	
NARG1	NM_057175	80155	
NAT8	NM_003960	9027	
NAT12	NM_001011713	122830	miR-330
PIM1	NM_002648	5292	

Symbol	GenBank	Entrez Gene ID	miRNAs with Overlapping Targets
PNMA1	NM_006029	9240	
PRDM2	NM_001007257	7799	
PTPN13	NM_080685	5783	
RCAN1	NM_203417	1827	miR-330
RIMBP2	NM_015347	23504	
SATB2	NM_015265	23314	
SCN8A	NM_014191	6334	
SEC24C	NM_004922	9632	miR-181a
SIX4	NM_017420	51804	
SLC25A25	NM_001006642	114789	
SLC26A7	NM_052832	115111	
SLC39A14	NM_015359	23516	
SLITRK3	NM_014926	22865	
SLU7	NM_006425	10569	
SPAST	NM_199436	6683	
ST18	NM_014682	9705	
TNFRSF9	NM_001561	3604	
UBE2V2	NM_003350	7336	
ZNF140	NM_003440	7699	
ZNF148	NM_021964	7707	miR-330
ZNF281	NM_012482	23528	
ZNF300	NM_052860	91975	
		2. miR-330	
ABCA1	NM_005502	19	miR-33
ACVR1	NM_001105	90	
ADAMTS5	NM_007038	11096	
AFF2	NM_002025	2334	
AFF4	NM_014423	27125	
AGTR2	NM_000686	186	
AK7	NM_152327	122481	
ANGEL2	NM_144567	90806	
ANKH	NM_054027	56172	
AP2M1	NM_004068	1173	
API5	NM_006595	8539	
APPL1	NM_012096	26060	
ARFGEF2	NM_006420	10564	
ARHGAP12	NM_018287	94134	
ARHGAP20	NM_020809	57569	
ARL17P1	NM_001113738	51326	
ATL2	NM_022374	64225	
ATP2B1	NM_001001323	490	miR-181a
ATP2C1	NM_014382	27032	
AZIN1	NM_148174	51582	
BCL9	NM_004326	607	
BCL11B	NM_022898	64919	
BFSP1	NM_001195	631	
BMPR2	NM_001204	659	
BTRC	NM_033637	8945	
C10ORF10	NM_007021	11067	
C14ORF129	NM_016472	51527	miR-181a

Symbol	GenBank	Entrez Gene ID	miRNAs with Overlapping Targets
C2CD2	NM_015500	25966	
C5ORF15	NM_020199	56951	
C5ORF23	NM_024563	79614	
C9ORF5	NM_032012	23731	
C9ORF64	NM_032307	84267	
CALCR	NM_001742	799	
CAPZA1	NM_006135	829	
CBX5	NM_012117	23468	
CCND3	NM_001760	896	
CD247	NM_000734	919	
CDR2	NM_001802	1039	
CHP	NM_007236	11261	
CLCN5	NM_001127899	1184	
CLDN8	NM_199328	9073	miR-181a
CLDN18	NM_016369	51208	
CMPK1	NM_016308	51727	
CNBP	NM_001127195	7555	
CRLS1	NM_019095	54675	
CSNK1G3	NM_001044722	1456	
CYP7A1	NM_000780	1581	
D4S234E	NM_014392	27065	
DAG1	NM_004393	1605	
DCAF7	NM_005828	10238	
DICER1	NM_030621	23405	
DLX1	NM_001038493	1745	
DNM3	NM_015569	26052	
DNM1L	NM_005690	10059	
DOCK5	NM_024940	80005	
DPP10	NM_001004360	57628	
EDEM1	NM_014674	9695	
EEF1A1	NM_001402	1915	
EFHC1	NM_018100	114327	
EIF5	NM_183004	1983	
EIF4E	NM_001968	1977	
EPM2A	NM_005670	7957	
ERAP1	NM_016442	51752	
ERBB4	NM_005235	2066	
ERC1	NM_178039	23085	
ERLIN2	NM_001003791	11160	
EXOC8	NM_175876	149371	
FAM107B	NM_031453	83641	
FAM72D	NM_207418	728833	
FGFR1	NM_023107	2260	
FMO2	NM_001460	2327	
FOXK1	NM_001037165	221937	miR-181a
FRK	NM_002031	2444	
GJC1	NM_005497	10052	
GNRHR	NM_001012763	2798	
GPRASP1	NM_014710	9737	
GRB10	NM_001001550	2887	miR-181a

Symbol	GenBank	Entrez Gene ID	miRNAs with Overlapping Targets
GRIA3	NM_007325	2892	miR-33
HELZ	NM_014877	9931	
HIVEP2	NM_006734	3097	
HNRNPU	NM_031844	3192	
HSF2	NM_004506	3298	
HSPH1	NM_006644	10808	
ID2	NM_002166	3398	
IMPACT	NM_018439	55364	
INO80D	NM_017759	54891	miR-181a
INSL5	NM_005478	10022	
ITM2C	NM_030926	81618	
JPH1	NM_020647	56704	
KANK2	NM_015493	25959	
KAT2B	NM_003884	8850	
KDM4C	NM_015061	23081	
KDSR	NM_002035	2531	
KIAA1012	NM_014939	22878	
KLF10	NM_005655	7071	
KLHL24	NM_017644	54800	
LAPTM4B	NM_018407	55353	
LNX2	NM_153371	222484	
LRPPRC	NM_133259	10128	
MARK1	NM_018650	4139	miR-181a
MAT2A	NM_005911	4144	
MBNL2	NM_144778	10150	
MECOM	NM_001105078	2122	
METAP2	NM_006838	10988	miR-181a
MMD	NM_012329	23531	
MOBKL1A	NM_173468	92597	
MRPS6	NM_032476	64968	
MYEF2	NM_016132	50804	
MYPN	NM_032578	84665	
NAT12	NM_001011713	122830	miR-33
NEFL	NM_006158	4747	
ONECUT2	NM_004852	9480	miR-181a
OTUD3	NM_015207	23252	
PAFAH1B1	NM_000430	5048	
PCDHA1	NM_018900	56147	miR-181a
PCDHA2	NM_018905	56146	miR-181a
PCDHA3	NM_018906	56145	miR-181a
PCDHA4	NM_018907	56144	miR-181a
PCDHA5	NM_018908	56143	miR-181a
PCDHA6	NM_031849	56142	miR-181a
PCDHA7	NM_018910	56141	miR-181a
PCDHA8	NM_018911	56140	miR-181a
PCDHA9	NM_031857	9752	miR-181a
PCDHA10	NM_018901	56139	miR-181a
PCDHA11	NM_018902	56138	miR-181a
PCDHA12	NM_018903	56137	miR-181a
PCDHAC1	NM_018898	56135	miR-181a

Symbol	GenBank	Entrez Gene ID	miRNAs with Overlapping Targets
PCDHAC2	NM_018899	56134	miR-181a
PCK1	NM_002591	5105	
PCMT1	NM_005389	5110	
PCTP	NM_001102402	58488	
PHAX	NM_032177	51808	
PLSCR1	NM_021105	5359	
PLSCR4	NM_001128304	57088	
PLXNA2	NM_025179	5362	
PQLC1	NM_025078	80148	
PRKAB2	NM_005399	5565	
PRKCB	NM_002738	5579	
PSD3	NM_015310	23362	
PTBP2	NM_021190	58155	
RAI2	NM_021785	10742	
RAP2A	NM_021033	5911	
RAVER2	NM_018211	55225	
RBM12	NM_152838	10137	
RCAN1	NM_203417	1827	miR-33
RGS10	NM_002925	6001	
RND3	NM_005168	390	
RNF212	NM_194439	285498	
RNF144B	NM_182757	255488	
RUFY2	NM_001042417	55680	
SAMD12	NM_001101676	401474	miR-181a
SCG3	NM_013243	29106	
SCP2	NM_001007100	6342	
SELI	NM_033505	85465	
SEMA4D	NM_006378	10507	
SEP15	NM_203341	9403	
SERINC3	NM_006811	10955	
SFRS1	NM_006924	6426	
SH3TC2	NM_024577	79628	miR-181a
SIN3A	NM_015477	25942	
SLAIN1	NM_001040153	122060	
SLC2A2	NM_000340	6514	
SLC5A3	NM_006933	6526	
SMG7	NM_173156	9887	
SMNDC1	NM_005871	10285	
SNAP23	NM_130798	8773	
SNX2	NM_003100	6643	
SORL1	NM_003105	6653	
SOSTDC1	NM_015464	25928	
STAU1	NM_001037328	6780	
STEAP4	NM_024636	79689	
STK3	NM_006281	6788	
SUDS3	NM_022491	64426	
SUMF1	NM_182760	285362	
TAPT1	NM_153365	202018	
TBL1XR1	NM_024665	79718	miR-181a
TBX5	NM_080717	6910	miR-10b

Symbol	GenBank	Entrez Gene ID	miRNAs with Overlapping Targets
TEAD1	NM_021961	7003	
TGFBR3	NM_003243	7049	
THBS1	NM_003246	7057	
TJP1	NM_003257	7082	
TMEM59	NM_004872	9528	
TNKS	NM_003747	8658	
TNRC6B	NM_001024843	23112	
TNS1	NM_022648	7145	
TOX	NM_014729	9760	miR-181a
TRA2A	NM_013293	29896	
TRIM2	NM_015271	23321	miR-181a
TRIM37	NM_015294	4591	
TRIP12	NM_004238	9320	
TROVE2	NM_004600	6738	
TSFM	NM_005726	10102	
TSHR	NM_000369	7253	
UBE2Q1	NM_017582	55585	
UBTD2	NM_152277	92181	
UBXN4	NM_014607	23190	
USP15	NM_006313	9958	
USP37	NM_020935	57695	
VAPA	NM_003574	9218	
VASH2	NM_024749	79805	
VGLL3	NM_016206	389136	
VPS54	NM_016516	51542	
WDR37	NM_014023	22884	
XK	NM_021083	7504	
YIPF5	NM_001024947	81555	
YTHDC1	NM_133370	91746	
ZBTB34	NM_001099270	403341	miR-181a
ZC3HAV1	NM_020119	56829	
ZCCHC24	NM_153367	219654	
ZFC3H1	NM_144982	196441	
ZFR	NM_016107	51663	
ZNF148	NM_021964	7707	miR-33
ZNF410	NM_021188	57862	
ZNF423	NM_015069	23090	
ZNF490	NM_020714	57474	
ZNF706	NM_016096	51123	
ZNF280D	NM_001002843	54816	
3. miR-181a			
ACSL1	NM_001995	2180	
ACVR2A	NM_001616	92	
ACVR2B	NM_001106	93	
ACYP1	NM_001107	97	
ADAM28	NM_014265	10863	
ADAMTSL1	NM_001040272	92949	
ADRBK1	NM_001619	156	
AFTPH	NM_203437	54812	
AHCTF1	NM_015446	25909	

Symbol	GenBank	Entrez Gene ID	miRNAs with Overlapping Targets
AKAP5	NM_004857	9495	
AKAP6	NM_004274	9472	
AKAP7	NM_004842	9465	
ARHGAP26	NM_015071	23092	
ARHGEF3	NM_001128615	50650	
ARSJ	NM_024590	79642	miR-10b
ATG5	NM_004849	9474	
ATM	NM_138292	472	
ATP11C	NM_173694	286410	
ATP2A2	NM_170665	488	
ATP2B1	NM_001001323	490	miR-330
ATXN3	NM_001127696	4287	
BAG2	NM_004282	9532	
BAG4	NM_004874	9530	
BAI3	NM_001704	577	
BAZ2B	NM_013450	29994	
BBS7	NM_018190	55212	
BEND3	NM_001080450	57673	
BHLHE40	NM_003670	8553	
BIRC6	NM_016252	57448	
BOLL	NM_197970	66037	
BRAP	NM_006768	8315	
BRD1	NM_014577	23774	
BRWD1	NM_033656	54014	
BTBD3	NM_014962	22903	
C10ORF104	NM_173473	119504	
C14ORF129	NM_016472	51527	miR-330
C15ORF29	NM_024713	79768	
C16ORF87	NM_001001436	388272	
C19ORF12	NM_031448	83636	
C20ORF12	NM_001099407	55184	
C21ORF66	NM_016631	94104	
C2ORF69	NM_153689	205327	
C5ORF41	NM_153607	153222	
C5ORF47	XM_376444	133491	
C6ORF89	NM_152734	221477	
CABC1	NM_020247	56997	
CALB1	NM_004929	793	
CALM1	NM_006888	801	
CAMK2D	NM_172128	817	
CAPRIN1	NM_005898	4076	miR-10b
CARD8	NM_014959	22900	
CBX7	NM_175709	23492	
CCAR1	NM_018237	55749	
CCDC14	NM_022757	64770	
CCDC117	NM_173510	150275	
CCNB1	NM_031966	891	
CCNJ	NM_019084	54619	
CCNL2	NM_001039577	81669	
CD302	NM_014880	9936	

Symbol	GenBank	Entrez Gene ID	miRNAs with Overlapping Targets
CDON	NM_016952	50937	
CHMP2B	NM_014043	25978	
CLASP1	NM_015282	23332	
CLDN8	NM_199328	9073	miR-330
CLIP1	NM_002956	6249	
CLVS1	NM_173519	157807	
CNTN4	NM_175612	152330	
CPNE2	NM_152727	221184	
CPOX	NM_000097	1371	
CREB5	NM_004904	9586	
CSF2RB	NM_000395	1439	
CTDSPL	NM_005808	10217	
CTTNBP2NL	NM_018704	55917	
CUL3	NM_003590	8452	
DARS	NM_001349	1615	
DCN	NM_133504	1634	
DDX52	NM_007010	11056	
DDX3X	NM_001356	1654	
DDX3	NM_004660	8653	
DEPDC6	NM_022783	64798	
DIRAS3	NM_004675	9077	
DLGAP2	NM_004745	9228	
DNAJC13	NM_015268	23317	
DNAL1	NM_031427	83544	
DOCK10	NM_014689	55619	
DSC3	NM_001941	1825	miR-33
DYNC1LI2	NM_006141	1783	
E2F5	NM_001951	1875	
E2F7	NM_203394	144455	
EIF4A2	NM_001967	1974	
ENPP1	NM_006208	5167	
EPC2	NM_015630	26122	
ETV6	NM_001987	2120	
EXD1	NM_152596	161829	
FAM13B	NM_001101801	51306	
FAM160A2	NM_032127	84067	
FBXO33	NM_203301	254170	
FBXO34	NM_017943	55030	
FIGN	NM_018086	55137	miR-10b
FKBP1A	NM_000801	2280	
FMNL2	NM_052905	114793	
FNDC3B	NM_022763	64778	
FOXK1	NM_001037165	221937	miR-330
FOXP1	NM_032682	27086	
FUCA1	NM_000147	2517	
FUT9	NM_006581	10690	miR-33
G3BP2	NM_012297	9908	
GABRA1	NM_001127648	2554	
GAPVD1	NM_015635	26130	
GATA6	NM_005257	2627	

Symbol	GenBank	Entrez Gene ID	miRNAs with Overlapping Targets
GATM	NM_001482	2628	
GCC2	NM_181453	9648	
GHITM	NM_014394	27069	
GPD2	NM_000408	2820	
GPD1L	NM_015141	23171	
GPRIN3	NM_198281	285513	
GPX8	NM_001008397	493869	
GRB10	NM_001001555	2887	miR-330
HEY2	NM_012259	23493	
HIC2	NM_015094	23119	
HOOK1	NM_015888	51361	
HOXB4	NM_024015	3214	
HOXC8	NM_022658	3224	
HOXD1	NM_024501	3231	
HRH1	NM_001098212	3269	
HSPC159	NM_014181	29094	
IL2	NM_000586	3558	
INO80D	NM_017759	54891	miR-330
IPO8	NM_006390	10526	
ITGA2	NM_002203	3673	
ITSN1	NM_001001132	6453	
KANK1	NM_015158	23189	
KCNH8	NM_144633	131096	
KCTD3	NM_016121	51133	
KDM5A	NM_005056	5927	
KIAA0528	NM_014802	9847	
KIAA1219	NM_020336	57148	
KIAA1239	XM_940885	57495	
KIAA2022	NM_001008537	340533	
KIF3A	NM_007054	11127	
KLHL2	NM_007246	11275	
KLHL5	NM_199039	51088	
KRAS	NM_033360	3845	
LAMP2	NM_001122606	3920	
LARP4	NM_199190	113251	
LCLAT1	NM_182551	253558	
LHFPL3	NM_199000	375612	
LIFR	NM_002310	3977	
LMO3	NM_001001395	55885	
LOC161527	XM_929030	161527	
LONRF2	NM_198461	164832	
LRRC8D	NM_018103	55144	
MAP1B	NM_005909	4131	
MAPK1	NM_002745	5594	
MARK1	NM_018650	4139	miR-330
MATN3	NM_002381	4148	
MBOAT2	NM_138799	129642	
MED8	NM_201542	112950	
MEGF9	NM_001080497	1955	
METAP1	NM_015143	23173	

Symbol	GenBank	Entrez Gene ID	miRNAs with Overlapping Targets
METAP2	NM_006838	10988	miR-330
MINA	NM_032778	84864	
MKLN1	NM_013255	4289	
MORC3	NM_015358	23515	
MPP5	NM_022474	64398	
MTF2	NM_007358	22823	
MTMR12	NM_001040446	54545	
MTMR15	NM_014967	22909	
MTX3	NM_001010891	345778	
MUC7	NM_152291	4589	
NCOA2	NM_006540	10499	
NFAT5	NM_001113178	10725	
NLN	NM_020726	57486	
NOTCH4	NM_004557	4855	
NOVA1	NM_002515	4857	
NR3C1	NM_001018076	2908	
NR6A1	NM_033334	2649	
NRAS	NM_002524	4893	
NTS	NM_006183	4922	
NUDT12	NM_031438	83594	
ONECUT2	NM_004852	9480	miR-330
OSBPL3	NM_145320	26031	
OSBPL8	NM_001003712	114882	
OTUD4	NM_001102653	54726	
PAM	NM_138822	5066	
PAPD5	NM_001040285	64282	
PAPOLG	NM_022894	64895	
PARK2	NM_013987	5071	
PARP11	NM_020367	57097	
PAWR	NM_002583	5074	
PAX9	NM_006194	5083	
PCDHA1	NM_031411	56147	miR-330
PCDHA2	NM_018905	56146	miR-330
PCDHA3	NM_018906	56145	miR-330
PCDHA4	NM_018907	56144	miR-330
PCDHA5	NM_018908	56143	miR-330
PCDHA6	NM_031849	56142	miR-330
PCDHA7	NM_018910	56141	miR-330
PCDHA8	NM_018911	56140	miR-330
PCDHA9	NM_031857	9752	miR-330
PCDHA10	NM_018901	56139	miR-330
PCDHA11	NM_018902	56138	miR-330
PCDHA12	NM_018903	56137	miR-330
PCDHAC1	NM_018898	56135	miR-330
PCDHAC2	NM_018899	56134	miR-330
PCNP	NM_020357	57092	
PDE5A	NM_001083	8654	
PER3	NM_016831	8863	
PGAP1	NM_024989	80055	
PHC3	NM_024947	80012	

Symbol	GenBank	Entrez Gene ID	miRNAs with Overlapping Targets
PHLPP2	NM_015020	23035	
PHTF2	NM_001127358	57157	
PI4K2B	NM_018323	55300	
PITPNB	NM_012399	23760	
PKNOX2	NM_022062	63876	
PLAC1L	NM_173801	219990	
PLEKHA3	NM_019091	65977	
PNRC2	NM_017761	55629	
POLQ	NM_199420	10721	
POLR3G	NM_006467	10622	
POM121	NM_172020	9883	
POM121C	NM_001099415	100101267	
PPP1R12B	NM_002481	4660	
PPP1R9A	NM_017650	55607	
PPP2R5E	NM_006246	5529	
PRDM4	NM_012406	11108	
PRDX3	NM_006793	10935	
PRH2	NM_001110213	5555	
PRKCD	NM_006254	5580	
PRTG	NM_173814	283659	
PSG5	NM_002781	5673	
PSRC1	NM_001032291	84722	
PTGER3	NM_198715	5733	
RAB3IP	NM_022456	117177	
RAD21	NM_006265	5885	
RAN	NM_006325	5901	
RAP1B	NM_001010942	5908	
RASSF2	NM_014737	9770	
RBM26	NM_022118	64062	
RBM25	NM_021239	58517	
REPS2	NM_001080975	9185	
RFC1	NM_002913	5981	
RIN2	NM_018993	54453	
RLF	NM_012421	6018	
RNF8	NM_183078	9025	
RNF34	NM_025126	80196	
ROD1	NM_005156	9991	
RP5-1022P6.2	NM_019593	56261	
RPAP2	NM_024813	79871	
RPE65	NM_000329	6121	
RPS6KB1	NM_003161	6198	
RRP15	NM_016052	51018	
S1PR1	NM_001400	1901	
SAMD12	NM_207506	401474	miR-330
SCD	NM_005063	6319	
SCOC	NM_032547	60592	
SEC24C	NM_004922	9632	miR-33
SEMA3C	NM_006379	10512	
SEMA4G	NM_017893	57715	
SFRS7	NM_001031684	6432	

Symbol	GenBank	Entrez Gene ID	miRNAs with Overlapping Targets
SH3TC2	NM_024577	79628	miR-330
SIPA1L2	NM_020808	57568	
SIRT1	NM_012238	23411	
SLC19A2	NM_006996	10560	
SLC24A1	NM_004727	9187	
SLC25A24	NM_213651	29957	
SLC7A11	NM_014331	23657	
SLITRK1	NM_052910	114798	
SPIN1	NM_006717	10927	
SPOCK1	NM_004598	6695	
SPP1	NM_001040058	6696	
SPRY4	NM_030964	81848	
SRPK2	NM_182692	6733	
ST8SIA4	NM_005668	7903	
STX7	NM_003569	8417	
SUCLG2	NM_003848	8801	
SYNE1	NM_133650	23345	
TADA2B	NM_152293	93624	
TBC1D1	NM_015173	23216	
TBC1D4	NM_014832	9882	
TBL1X	NM_005647	6907	
TBL1XR1	NM_024665	79718	miR-330
TBPL1	NM_004865	9519	
TCERG1	NM_006706	10915	
TET2	NM_017628	54790	
TFEC	NM_012252	22797	
TFRC	NM_003234	7037	
TGFBR1	NM_004612	7046	
TGFBRAP1	NM_004257	9392	
TIFA	NM_052864	92610	
TIMP3	NM_000362	7078	
TLL1	NM_012464	7092	
TMEM26	NM_178505	219623	
TMEM27	NM_020665	57393	
TMEM131	NM_015348	23505	
TMEM165	NM_018475	55858	
TMF1	NM_007114	7110	
TNF	NM_000594	7124	
TNFRSF11B	NM_002546	4982	
TNFSF4	NM_003326	7292	
TNPO1	NM_153188	3842	
TOM1L1	NM_005486	10040	
TOR1AIP2	NM_145034	163590	
TOX	NM_014729	9760	miR-330
TRDMT1	NM_176081	1787	
TRIM2	NM_015271	23321	miR-330
TSPAN8	NM_004616	7103	
TSPYL4	NM_021648	23270	
UBE2A	NM_003336	7319	
UBE2D1	NM_003338	7321	

Symbol	GenBank	Entrez Gene ID	miRNAs with Overlapping Targets
UBP1	NM_001128160	7342	
USP42	NM_032172	84132	
VBP1	NM_003372	7411	
VCAN	NM_001126336	1462	
WHSC2	NM_005663	7469	
WNK1	NM_018979	65125	
XRN1	NM_019001	54464	
ZBTB34	NM_001099270	403341	miR-330
ZBTB44	NM_014155	29068	
ZFAND6	NM_019006	54469	
ZFP36L1	NM_004926	677	
ZFP36L2	NM_006887	678	
ZIC3	NM_003413	7547	
ZNF83	NM_001105549	55769	
ZNF124	NM_003431	7678	
ZNF439	NM_152262	90594	
ZNF440	NM_152357	126070	
ZNF441	NM_152355	126068	
ZNF454	NM_182594	285676	
ZNF468	NM_001008801	90333	
ZNF559	NM_032497	84527	
ZNF594	NM_032530	84622	
ZNF655	NM_138494	79027	
ZNF673	NM_001129898	55634	
ZSCAN23	NM_001012455	222696	
4. miR-10b			
ARG2	NM_001172	384	
ARSJ	NM_024590	79642	miR-181a
BDNF	NM_170731	627	
C1ORF71	XM_001717264	163882	
CAPRIN1	NM_005898	4076	miR-181a
CLCC1	NM_001048210	23155	
EBF2	NM_022659	64641	
FIGN	NM_018086	55137	miR-181a
FNBP1L	NM_001024948	54874	
GALNT1	NM_020474	2589	
HOXA3	NM_153631	3200	
HOXB3	NM_002146	3213	
KLF11	NM_003597	8462	
LCA5	NM_001122769	167691	miR-33
MAP3K7	NM_003188	6885	
NONO	NM_007363	4841	
RB1CC1	NM_014781	9821	
RBM27	NM_018989	54439	
RNF7	NM_183237	9616	
SNX18	NM_001102575	112574	
TBX5	NM_181486	6910	miR-330
TFAP2C	NM_003222	7022	
TRNT1	NM_182916	51095	
USP46	NM_022832	64854	
WDR26	NM_001115113	80232	

Additional File 7: 40 networks significantly associated with the predicted targets of miR-33, miR-330, miR-181a, and miR-10b

---

Network No.	p-value	No. of Predicted Transcripts	Molecules in Network
1. miR-33			
1	1E-34	16	ABCA1,ACAA2,Alcohol group acceptor phosphotransferase,CACNA1C, CACNA1D, CACNA1H,CDK6,CROT,CSNK1D,GLO1,HADHB,HMGA2,KCNMA1, LDL,LIMK2,MIR124, MMP16,MSR1,NARG1,NEK2,NFkB (complex),PDGF BB,PGRMC2,PIM1,Pka, PLEKHA1,PPP1R13L,PTPN13,RCAN1,RYR1 (includes EG:114207),SERPINB6,ST18, SUCLG2,TNFRSF9,Trypsin
2	1E-27	14	AHSG,ARD1A,BAT1,CHMP1B,CLSPN,CP110,CPB2,CTNNB1,DSC3,FBXW11,FG B,FH,GLCCI1,GOT1,Groucho,HBS1L,HNF1A,HNF4A,HPX,ING3,LIN7C,MIR9-1 (includes EG:407046),NAT8,NFYB,PCBD1,PLK1,PNMA1,PTPN13,PZP, RIMBP2,SLC26A7, SLITRK3,SPAST,ZNF281,ZNF300
3	1E-26	13	ABCE1,ANP32A,ARID5B,ATXN1,B3GALT2,BACE1,BAT2,BECN1,C11ORF41,CC DC85B,CFL1,CHI3L1,CSF1R,DAZAP2 (includes EG:9802),DZIP3,EEA1, FAM46C(includes EG:54855),FGA,KIAA2018,KIF3C,KRT18,LPA,MIR17 (includes EG:406952),MIR212 (includes EG:406994),MIR29A (includes EG:407021),MIR29B1,MIR29B2,NAT12,NEDD4L,SATB2,SLC39A14, SLU7,TARDBP,TGFB1,TRIP6
4	1E-21	11	ASAP1,CADM1,CCL9,CDC42,CDC42BPA,CDC42BPB,CXCL12,EBF1,EFS,ERK,FC GR2B,FUT9,FZD5,GOPC,IL1B,ITSN2,LIMK2,LPP,MATK,PALLD,POSTN,progest erone,PRRX1,PTK2,RRS1,SEC23A,SEC24C (includes EG:9632),SHPRH,SIX4, SLC25A25,SLC6A4, SSTR3,UBE2V2,VIM,ZNF140
5	1E-13	8	amino acids,APLP2,APP,APPBP2,beta-estradiol,CA2,CXCL5,DYRK3,EN2, EPB41L3,EPO,FLOT1,FSTL1,GRI,GRIA3,GRIA4,GRIP2,HIPK2,HIST4H4 (includes EG:121504), HSP90B1,KISS1,MLLT3,nitric oxide,NSF,PARD6B, PPP3R1,PRDM2,PRKCA,PRKG1, PTPN13,SLC6A4,SYT11,WNT5B,ZNF148, ZNRF1
2. miR-330			
1	1E-40	23	ACVR1,AGTR2,Ap1,APPL1,ATP2B1,BMPR2,C5ORF23,CAPZA1,CCND3,CD247, ERK,Estrogen Receptor,FGFR1,FSH,GNRH,GNRHR,hCG,HSF2,HSPH1,Insulin, KLF10,Lh, MAT2A,MRPS6 (includes EG:64968),PDGF BB,PP2A,RAP2A,RND3, SFRS1,Shc, SMG7,SNAP23,Tgf beta,THBS1,TRA2A
2	1E-37	24	ABCA1,ARFGEF2,BCL11B,BTRC,Calcineurin protein(s),CALCR,CBX5,CHP, cldn,CLDN8,CLDN18,Creb,CYP7A1,GJC1,HELZ,HNRNPU,LDL,N-cor,Nfat (family),NFkB (complex),PCK1,PRKAC,PRKCB,RCAN1,RNA polymerase II, Rxr,SEP15,SLC2A2, STAU1,TBL1XR1,TEAD1,TGFBR3,TJP1,TSHR,Vegf
3	1E-33	20	Akt,BCL9,Caspase,Ck2,EEF1A1,EIF5,EIF4E,ERBB4,ERK1/2,F Actin,FRK, GRB10, Histone h3,Histone h4,ID2,Jnk,KAT2B,KDM4C,MAGED2,Mapk, MECOM,MYEF2, NRG2 (includes EG:381149),P38 MAPK,PAFAH1B1, PCDHA4,PCDHA11,PI3K,Pkc(s), Rb,SIN3A,STK3,SUDS3,TNS1,XK
4	1E-26	17	AFF2,C10ORF10,CBLC,CD19,CD82,CNTFR,CRK,DAB1,EDEM1,EFEMP2,ERBB,F CGR1A/2A/3A,INPPL1,KHDRBS1,KRTAP4-12,LYN,NEDD9,PCDHA1,PCDHA2, PCDHA3, PCDHA5,PCDHA6,PCDHA7,PCDHA8,PCDHA9,PCDHA10,PCDHA12, PLSCR1,PLSCR4,RBM12,SSR1,STX18,XBP1,YIPF5
5	1E-25	16	ADIPOR1,ANGEL2,APOC3,ARL17P1,C9ORF5,C9ORF64,CIAO1,CNBP,EEF2K,FA M107B,FOXA2,glycogen,HNF1B,HNF4A,HNRNPR,KIAA1012,LRP5,MAGOH,M ETAP2,MLXIPL,NCBP1,ONECUT2,PCK1,PEPCK,PRKAB2,RAI2,RPL31,SLC2A2,S YTL4,TADA3L,TMEM59 (includes EG:9528),TOE1,TROVE2,TTR,USP15

Network No.	p-value	No. of Predicted Transcripts	Molecules in Network
6	1E-25	16	BMP2,C14ORF129,C2CD2,C5ORF15,CRADD,D4S234E,DNM1L,EPM2A (includes EG:7957),FRK,GSK3B,HNRNPA3,HOXA13,ITM2C,KLF10,MIR17 (includes EG:406952),MIR27B (includes EG:407019),MIRN330,MNT,MYPN, NAT12,PAX8, PLXNA2,PPP1CA,PPP1R3C,RBL2,SERPINB2,SFRP1,SFRS5, SMYD2,SOSTDC1,TBX2,TNF,ZBTB34,ZNF410,ZNF423
7	1E-23	15	ATL2,BCL2,BCLAF1,CGN,CMPK1,DCAF7,DERL1,DHX15,DPP10,E2F4,EFHC1,HIST2H2BE,IFNA2,KIF1B,LAPTM4B,MARK1,MIR122,MYCBP2,NDE1,NEFL,OFD1 ,PTPN3,SCG3,TBC1D4,TP53BP2,TRIM2,TSFM (includes EG:10102),UBXN4, USP18,USP37 (includes EG:57695),VCP,VPS54,YWHAG,YWHAZ,ZC3HAV1
8	1E-23	15	ABL1,AFF4,amino acids,APEX2,ARHGAP12,CSNK1G3,DLGAP5,DNM3,DYRK2, ERLIN2,EWSR1,GNS,hydrogen peroxide,MBNL2,MERTK,MIR133A,MIR133A-1, MIR133A-2,MIRN140,MOBKL1A,NBEA (includes EG:26960),PCDHAC2, PCMT1, PTBP2,RIPK4,SELI,SMNDC1,SNX2,SRF,STK38L (includes EG:23012), TCEB3B,TGFB1, TRAF2,TRIM37,UBTD2
9	1E-22	15	AGT,ANKH,AZIN1,BOK,CLNS1A,DICER1,DOCK5,EIF5,EIF2C2,FMR1,GRIA3,HSF2,LNX2,MIRLET7B (includes EG:406884),MIRN346,NR3C1,OTUD3,PIWIL4, PPP1R3C, PPP2CA,PRKRA,RNF144B,SERINC3,STXBP5,TARBP2,TBX2,TBX5, TNRC6B,UBE2A,UBE2I,UBE2L6,UBE2Q1,UBE2T,UGCG,WDR37
10	1E-21	14	AGR2,ARSA,ARSB,ARSD,ARSE,ARSF,ARSG,ARSI,ARSJ,BFSP1,BMX,CDKN2A,D4S234E,DAG1,DLX1,ERBB2,ERC1,KANK2,MMD,PLXNB2,PQLC1,PTRF,RASA3,RUFY2,SCP2,SLCO1A1,STEAP4,SULF1,SULF2,SUMF1,TLN1,TNF,VIM,ZFR,ZNF148
11	1E-19	13	ASS1,ATP2C1,BTRC,CCDC85B,CDKN1A,CDR2,CLCN5,Cofilin,ERAP1,EXOC7,EXOC8,GCN1L1,HIVEP2,HOXB4,IGF2BP1,IMPACT,KRT20,LRPPRC,MAZ,MBIP,MCM10,MIRLET7A1,MYC,PKN1,PNN,POLR1B,POLR2L (includes EG:5441), PRC1,RFX1,TADA2L, TRIP12,UXT,VASH2,YTHDC1,ZFC3H1
12	1E-19	13	ADAMTS5,AP1S2,API5,DHX15,DRD2,FMO,FMO2,FOXK1,GH1,GPRASP1,HTT, KCNK3,KLF16,MIR103-1 (includes EG:406895),MLH1,PCTP,PSD3,PTPN22, RALB,retinoic acid,RPL6,RPS19,SATB1,SFRS2IP,SLAIN1,SLC5A3,TOX,UBA7, VAMP1,VAPA,VAPB,VGLL3,YWHAB,ZNF133,ZNF706
13	1E-13	10	ACTR5,ACTR8,AP2M1,APEX2,ARHGAP29,ATP6V0E1,ERVK6,FGD1,FIGF,FURIN,GATS,HEBP2,INO80,INO80B,INO80D,INO80E,INS1,ITGB1,JPH1,KATNA1,KLHL24,LDLRAP1,MIR124,MIR124-1,phosphatidylinositol-3,4,5-triphosphate, PREX1,Rac,RAVER2, RGS10,SEMA4D,SLC16A1,SORL1,SURF4,TNKS,ZCCHC24
3. miR-181a			
1	1E-43	26	BAG4,BIRC6,CALB1,CAMK2D,DLGAP2,E2F7,ENPP1,FUCA1,HEY2,HRH1,Ifn,MIR1,MTMR12,MUC7,NFkB (complex),NFkB (family),NOTCH4,NTS,PARK2, PAWR, POM121C,PRDX3,PRH2,Proteasome,SPHINGOMYELINASE,TFRC, TIMP3,Tnf receptor,TOM1L1,TRIM2,UBE2,UBE2A,UBE2D1,Ubiquitin, ZFAND6
2	1E-41	25	ACVR2A,ACVR2B,Alcohol group acceptor phosphotransferase,ATM, BHLHE40, DDX3X,EIF4A2,FBXO33,FBXO34,FKBP1A,GABRA1,Histone H1,IFN Beta,Importin beta,IPO8,KLHL2,MAPK1,p70 S6k,PDE5A,POLQ,PP1,PP1-C,PPP1R9A,PPP2R5E, PRKCD,RAN,REPS2,Smad,SPRY4,Tgf beta,TGFBR, TGFBR1,TGFBRAP1,TNFRSF11B, TNPO1
3	1E-34	22	Akt,Ap1,ATXN3,BAG2,BRWD1,CABC1,Cbp/p300,CBX7,CCAR1,E2F5,Estrogen Receptor,ETV6,GRB10,Growth hormone,Histone h4,HOXB4,HOXC8,Hsp90, ITSN1, KDM5A,N-cor,NCOA2,NR3C1,Pias,RAD21,Rxr,SPP1,SRPK2,STAT5a/b, TBC1D4, TBL1X,TBL1XR1,TET2,tyrosine kinase,VitaminD3-VDR-RXR
4	1E-32	21	ACSL1,AKAP,AKAP5,AKAP6,AKAP7,AMPK,ATG5 (includes EG:9474),ATP2A2, CaMKII,Caspase,CCNL2,CDON,Creb,CREB5,Cytochrome c,DNAJC13, DYNC1LI2, FSH,GATA6,hCG,Histone h3,Hsp70,Mmp,P38 MAPK,PCDHA4, Pka,RFC1,RNA polymerase II,SCD,SFRS7,SIRT1,TBPL1,TCERG1,ZBTB44, ZNF83

Network No.	p-value	No. of Predicted Transcripts	Molecules in Network
5	1E-21	16	ADAM11,ADAM28,C2ORF69,C5ORF41,CCNJ,CD40LG,CTDSPL,FBXO33,FRK,ILK,KCNH8,MEGF9,MIR140 (includes EG:406932),MIR153-1,MIR153-2, MIR181B1, MIR181B2,MIR21 (includes EG:406991),MIR217 (includes EG:406999),MT1G, NEFH,NFIB,OTUD4,PAPOLA,PAPOLG,PAX8,PCDHAC2, PHLPP2,RAB9A,RB1,SEMA4G,SLITRK1,SPIN1,TERT,ZIC3
6	1E-20	17	14-3-3,Angiotensin II receptor type 1,Calcineurin A,Calcineurin protein(s), CARD8, CD3,CLASP1,CLIP1,DOCK10,DSC3,Eotaxin,Fcer1,GPD2,Ifn gamma, IKK (complex), Importin alpha,LRRC8D,MAP2K1/2,MEF2,MHC Class II,MKLN1,NFAT5,NFAT (complex),Nfat (family),NR6A1,PLEKHA3,Ptger, PTGER3,SEMA3C,TCR,TIFA,TNF, TNFSF4,VAV,ZFP36L1
7	1E-20	15	ADRBK1,CRK,EFNA1,ENG,ERBB4,FLNA,GNB2L1,INHBA,KAT2B,MIR34A (includes EG:407040),MMP13,MTF2,NEDD9,NFYC,NKX2-1,PCDHA1, PCDHA2,PCDHA3, PCDHA5,PCDHA6,PCDHA7,PCDHA8,PCDHA9,PCDHA10, PCDHA11,PCDHA12,PDPK1,PRDM4,SMAD3,sphingosine-1-phosphate, TMEM165,TP53BP2,WEE1,ZEB2
8	1E-20	15	ACTR5,ACTR8,ALDOB,BBS7,BEND3,CD302,CDCA7L,CHMP2B,ERVK6,FOXA2, GATS,HNF1A,HOXD1,INO80B,INO80D,INO80E,KANK1,METAP1,MIR124,MIR 124-1, ONECUT2,OSBPL8,PCNP,POLR3G,POU2F1,retinoic acid,RPE65, RUVBL1,SUCLG1, SUCLG2,SUMO1,TMEM109,TNFRSF21,TWIST2,ZNF673
9	1E-19	15	ACTB,AFTPH,ANKH,APBB1,beta-estradiol,C19ORF12,CPNE2,EIF3M,ENC1, FUT9, GH1,Glutathione peroxidase,GPX8,GSTM3 (includes EG:2947),GSTT1, hydrogen peroxide,MGST2,PARP,PARP4,PARP11,PER3,PNRC2 (includes EG:55629), progesterone,PSG5,PTGER3,PTP4A2,ROD1,SLC7A11,SMARCA4, SMPDL3A,SPOCK1,SRD5A2,TOR1AIP2,TOX,TSC22D1
10	1E-18	14	ATYPICAL PROTEIN KINASE C,BCR,C1q,Collagen type I,Collagen(s),DCN,ERK, Focal adhesion kinase,Ige,IgG,Igm,Integrin,ITGA2,KIF3A,Laminin,LAMP2, LIFR,Mek,NRAS, Pdgf,PDGF BB,PLA2,Pld,PP2A,RAB3IP,Rac,Rap1,Ras, RASSF2,RIN2,RPS6KB1,S1PR1, TLL1,TSPAN8,VCAN
11	1E-18	14	AP3B2,ARRB1,ATP11C,BRD1,BRF1,cldn,CLDN4,CLDN6,CLDN8,CNTN4,CSDA, DNAJB6,FIGN,FOXK1,GAPVD1,HNRNPAB,KCTD3,LOC161527,MIR31 (includes EG:407035),MIRN341,MTX3,NFYB,OGG1,OSBPL3,PLEC1,RAB8B, RBM26,SON,SSTR3,ST8SIA4,TJP1,VIM,VPS39,YWHAB,ZNF468
12	1E-18	14	AHCTF1,ARHGEF3,BTBD3,CDC2L1 (includes EG:984),COBRA1,DDX52, FAM13B, HEATR1,IPO7,KIAA1239,KIAA2022,KPNB1,MED8,MED25,MED26, MED28,MED29,MIR195,MIR373,MIR181C (includes EG:406957),MIR199A1, MLH1,MTMR15,MYC, NUP133,PAPD5,PHTF2,POLR2C,POLR2D,RDBP,Rfc, RPAP2,SR140,WHSC2,XRN1
13	1E-16	13	BRAP,C15ORF29,C20ORF12,CDC14A,CDK5,CTSB,DNAL1,FAM160A2,GFI1B,GHITM,GOLGA2,GPS2,HIC2,HLA-B,HNF4A,LARP4,MINA,MIRN336,MYST2, NDRG1, ONECUT1,PDK2,PPARGC1A,PPP1CA,PPP1R12B,PRDM5,PRKCE, RAB33B,RABAC1,ROCK1,SEC23A,SEC24C (includes EG:9632),SLC6A4,TMF1, ZNF594
14	1E-16	13	ADAMTS4,ADAMTS5,ARF4,C21ORF33,CAPRIN1,CASP14,CCDC14,CCDC85B,COMP,DNAJB4,FNDC3B,GCC2,MATN3,MATN4,NNMT,PARVA,PAX3,PHGDH,PI4K2B,RAB18,RBM25 (includes EG:58517),RNF8,RNH1,SDPR,SF3A3, SLC2A4,STOML2,STX7, TBC1D1,TGFB1,TSPYL4,VPS11,WBP4,WNK1,ZFP36L2
15	1E-16	13	ADAMTSL1,AHSG,ALP,ALPI,ATP2A1,B4GALNT1,BMYO,C16ORF87,CD274,CD 1A,CSF1,DBT,DDX3Y (includes EG:8653),DGKA,EPC2,GATM,GRB2,HMGN3, HSPC159,IL4,L-triiodothyronine,LIF,MIRN328,PAX9,RP5-1022P6.2,SHH, SLC24A1,SLC29A1,ST5, TFEC,TMEM131,VEGFA,ZBTB34,ZFP36,ZNF124
16	1E-15	13	ACADVL,AIMP2,ALB,BAI3,BAZ2B,BBC3,C10ORF104,CBX4,CPOX,CTBP2,DENR (includes EG:8562),EEF2K,EIF4A3 (includes EG:9775),GPRIN3,HIPK2,HP, ITGA2, LMO3,METAP2,MORC3,MTA1,PAX5,PCGF2,PDE4B,PEG3,PHC2,PHC3, PSRC1,RBBP6 (includes EG:5930),S100A4,SERPING1,SLC19A2,SUPT3H (includes EG:8464), TADA2B,TP53

Network No.	p-value	No. of Predicted Transcripts	Molecules in Network
17	1E-15	12	ACYP1,APOBEC3B,C14ORF129,CDC5L,CHIC2,CMBL,CRB3,CYP2D9,DISC1,ETNK2,EXOSC4,G3BP2,GSK3B,HNF4A,HOOK1,Hydrolase,KIF3C,LTA4H,MDFI (includes EG:4188),MPP1,MPP5,NLN,NPDC1,PARP4,PRCC,PTPN7,REXO2, SCOC,SFI1,SYNE1,UBP1,VPS29,ZNF439,ZNF440,ZNF559
18	1E-14	12	ADCY,ADRBK1,ARHGAP26,ATP2B1,CCNB1,Cdc2,CSF2RB,CUL3,Cyclin A, Cyclin B,Cyclin E,DARS,DIRAS3,ERK1/2,G protein beta gamma,Gpcr,IL1,IL2, Insulin, Interferon alpha,Jnk,KRAS,LDL,Lh,Mapk,PI3K,PITPNB,Pkc(s), PLC,RAP1B,Ras homolog,Sapk,Shc,STAT,Vegf
19	1E-11	10	amino acids,ASPH,C7ORF16,CDKL1,CTTNBP2NL,DAPK2 (includes EG:23604), DEPDC6 (includes EG:64798),EEF1A1,FMNL2 (includes EG:114793),FNBP4, FOXP1,HAT1,HTT,IL12 (complex),MARK1,MBOAT2,MLST8,MTOR,NXN, PDK1,PGGT1B, PPP2CA,PQBP1,PRPF40A,PTPN7,RIPK4,RLF,RNF34,RP6-213H19.1,STAT4,STRADA, TPP2,TRDMT1,TRIM30,ZNF655
20	1E-11	10	Actin,AFAP1,AKAP5,Alpha tubulin,APLNR,ARPP-21,C21ORF66,CALM1, Calmodulin,CAMK2N2,Ck2,CYB5R3,F Actin,FGD4,GABRR1,GRIN1,KCNQ2, KCNQ3,KIAA1219, MAP1B,NOVA1,PAM,PCP4,PFDN4,PFDN6,POM121, PPEF2,sodium chloride,SRC, TMEM27,TRIM13,Tubulin,UNC13A,USP6,VBP1
			4. miR-10b
1	1E-23	10	ACT1,ANKRD1,ARPP-21,ASB2,BDNF,CALB2,CNGA2,EBF2,EDARADD,FASLG, FIGN, FNBP1L,GRIN2B,GRIN2C,HTT,IP6K2,KLF11,LCA5,MAP3K7,MBTPS1, MIR124-1,NAPB,NEUN,NFkB (complex),PELI1,PELI3,RBM27,RNF7,SGSM2, SNX18,TPH1, UBC,UBE2V1,WDR34,YY1
2	1E-22	10	AGT,ANKRD1,ARG2,Arginase,beta-estradiol,CAPRIN1,CREB1,G3BP1, GALNT1, GRIN2C,GRM3,HOXA2,HOXA3 (includes EG:3200),HOXB3,HOXD3, LOR,MAFB, MYLPF,NR3C1,PBX1,PPP3R1,RB1CC1,RPS6KA5,SLC19A1,STAT3, TBX2,TBX3,TBX5,TFAP2C,TGFB1,TP53,TRNT1,WDR26,WWOX,WWTR1

Additional File 8: Canonical pathways involving at least three molecules present in top networks significantly associated with predicted targets of miR-33, miR-330, miR-181a, and miR-10b

---

Canonical Pathways	Present in miR-33 Network?	Present in miR-330 Network?	Present in miR-181a Network?	Present in miR-10b Network?	Total miRNA Networks with Canonical Pathway
Molecular Mechanisms of Cancer	Yes	Yes	Yes	Yes	4
Acute Phase Response Signaling	Yes	Yes	Yes	No	3
Colorectal Cancer Metastasis Signaling	Yes	Yes	Yes	No	3
Glucocorticoid Receptor Signaling	Yes	Yes	Yes	No	3
GNRH Signaling	Yes	Yes	Yes	No	3
G-Protein Coupled Receptor Signaling	Yes	Yes	Yes	No	3
Hepatic Cholestasis	Yes	Yes	Yes	No	3
NF-kB Signaling	Yes	No	Yes	Yes	3
PPAR Signaling	Yes	Yes	Yes	No	3
Production of Nitric Oxide and Reactive Oxygen Species in Macrophages	Yes	Yes	Yes	No	3
Role of Macrophages, Fibroblasts and Endothelial Cells in Rheumatoid Arthritis	Yes	Yes	Yes	No	3
Tight Junction Signaling	Yes	Yes	Yes	No	3
Type I Diabetes Mellitus Signaling	No	Yes	Yes	Yes	3
Actin Cytoskeleton Signaling	Yes	Yes	No	No	2
Androgen Signaling	Yes	Yes	No	No	2
Aryl Hydrocarbon Receptor Signaling	Yes	Yes	No	No	2
Axonal Guidance Signaling	Yes	Yes	No	No	2
B Cell Receptor Signaling	No	Yes	Yes	No	2
BMP signaling pathway	Yes	Yes	No	No	2
Breast Cancer Regulation by Stathmin1	Yes	Yes	No	No	2
Calcium Signaling	Yes	Yes	No	No	2
Cardiac Hypertrophy Signaling	Yes	Yes	No	No	2
Ceramide Signaling	No	Yes	Yes	No	2
Ephrin Receptor Signaling	Yes	Yes	No	No	2
Glioblastoma Multiforme Signaling	Yes	Yes	No	No	2
Glioma Signaling	Yes	Yes	No	No	2
Hepatic Fibrosis / Hepatic Stellate Cell Activation	No	Yes	Yes	No	2
HMGB1 Signaling	No	Yes	Yes	No	2
Huntington's Disease Signaling	No	Yes	No	Yes	2
iCOS-iCOSL Signaling in T Helper Cells	Yes	No	Yes	No	2
IL-12 Signaling and Production in Macrophages	Yes	Yes	No	No	2
IL-6 Signaling	No	Yes	Yes	No	2
IL-8 Signaling	Yes	Yes	No	No	2
ILK Signaling	No	Yes	Yes	No	2
LXR/RXR Activation	Yes	No	Yes	No	2
PAK Signaling	Yes	Yes	No	No	2
PPARalpha/RXRalpha Activation	Yes	Yes	No	No	2

Canonical Pathways	Present in miR-33 Network?	Present in miR-330 Network?	Present in miR-181a Network?	Present in miR-10b Network?	Total miRNA Networks with Canonical Pathway
Protein Kinase A Signaling	Yes	No	Yes	No	2
PTEN Signaling	No	Yes	Yes	No	2
RAR Activation	Yes	Yes	No	No	2
Renin-Angiotensin Signaling	Yes	Yes	No	No	2
Role of NFAT in Cardiac Hypertrophy	Yes	Yes	No	No	2
Role of NFAT in Regulation of the Immune Response	Yes	Yes	No	No	2
Role of Osteoblasts, Osteoclasts and Chondrocytes in Rheumatoid Arthritis	No	Yes	Yes	No	2
Thrombin Signaling	No	Yes	Yes	No	2
Type II Diabetes Mellitus Signaling	Yes	No	Yes	No	2
4-1BB Signaling in T Lymphocytes	Yes	No	No	No	1
Acute Myeloid Leukemia Signaling	Yes	No	No	No	1
AMPK Signaling	No	Yes	No	No	1
Amyloid Processing	Yes	No	No	No	1
Amyotrophic Lateral Sclerosis Signaling	Yes	No	No	No	1
Apoptosis Signaling	No	No	Yes	No	1
Atherosclerosis Signaling	Yes	No	No	No	1
Autoimmune Thyroid Disease Signaling	No	Yes	No	No	1
Cardiac B-adrenergic Signaling	Yes	No	No	No	1
CCR5 Signaling in Macrophages	No	Yes	No	No	1
CD28 Signaling in T Helper Cells	No	Yes	No	No	1
CD40 Signaling	No	No	No	Yes	1
Cdc42 Signaling	No	Yes	No	No	1
Cellular Effects of Sildenafil (Viagra)	Yes	No	No	No	1
Cholecystokinin/Gastrin-mediated Signaling	No	Yes	No	No	1
Chronic Myeloid Leukemia Signaling	Yes	No	No	No	1
Crosstalk between Dendritic Cells and Natural Killer Cells	No	No	Yes	No	1
CXCR4 Signaling	No	Yes	No	No	1
Death Receptor Signaling	No	No	Yes	No	1
Dendritic Cell Maturation	No	No	Yes	No	1
EGF Signaling	No	Yes	No	No	1
EIF2 Signaling	No	Yes	No	No	1
Endothelin-1 Signaling	No	Yes	No	No	1
ERK/MAPK Signaling	No	Yes	No	No	1
Erythropoietin Signaling	No	Yes	No	No	1
Estrogen Receptor Signaling	No	Yes	No	No	1
Estrogen-Dependent Breast Cancer Signaling	No	Yes	No	No	1
Factors Promoting Cardiogenesis in Vertebrates	No	Yes	No	No	1
Germ Cell-Sertoli Cell Junction Signaling	No	Yes	No	No	1
Human Embryonic Stem Cell Pluripotency	No	Yes	No	No	1
Hypoxia Signaling in the Cardiovascular System	No	No	Yes	No	1
IGF-1 Signaling	No	Yes	No	No	1
IL-1 Signaling	Yes	No	No	No	1

Canonical Pathways	Present in miR-33 Network?	Present in miR-330 Network?	Present in miR-181a Network?	Present in miR-10b Network?	Total miRNA Networks with Canonical Pathway
IL-2 Signaling	No	Yes	No	No	1
IL-3 Signaling	No	Yes	No	No	1
Induction of Apoptosis by HIV1	No	No	Yes	No	1
Inositol Phosphate Metabolism	Yes	No	No	No	1
Insulin Receptor Signaling	No	Yes	No	No	1
Integrin Signaling	No	Yes	No	No	1
mTOR Signaling	No	Yes	No	No	1
Natural Killer Cell Signaling	No	Yes	No	No	1
Neuropathic Pain Signaling in Dorsal Horn Neurons	No	No	No	Yes	1
Neurotrophin/TRK Signaling	No	Yes	No	No	1
Nicotinate and Nicotinamide Metabolism	Yes	No	No	No	1
Nitric Oxide Signaling in the Cardiovascular System	Yes	No	No	No	1
Ovarian Cancer Signaling	No	Yes	No	No	1
P13K/AKT Signaling	No	Yes	No	No	1
p70S6K Signaling	No	Yes	No	No	1
PDGF Signaling	No	Yes	No	No	1
Phospholipase C Signaling	No	Yes	No	No	1
Prolactin Signaling	No	Yes	No	No	1
Protein Ubiquitination Pathway	No	No	Yes	No	1
PXR/RXR Activation	Yes	No	No	No	1
Rac Signaling	Yes	No	No	No	1
Regulation of eIF4 and p70S6K Signaling	No	Yes	No	No	1
Regulation of IL-2 Expression in Activated and Anergic T Lymphocytes	No	Yes	No	No	1
Relaxin Signaling	Yes	No	No	No	1
Renal Cell Carcinoma	No	Yes	No	No	1
Role of NANOG in Mammalian Embryonic Stem Cell Pluripotency	No	Yes	No	No	1
Role of PKR in Interferon Induction and Antiviral Response	No	No	Yes	No	1
SAPK/JNK Signaling	No	Yes	No	No	1
Small Cell Lung Cancer Signaling	Yes	No	No	No	1
Sphingosine-1-phosphate Signaling	No	Yes	No	No	1
Synaptic Long Term Potentiation	Yes	No	No	No	1
Systemic Lupus Erythematosus Signaling	No	Yes	No	No	1
T Cell Receptor Signaling	No	Yes	No	No	1
TGF-B Signaling	No	Yes	No	No	1
Thrombopoietin Signaling	No	Yes	No	No	1
VDR/RXR Activation	Yes	No	No	No	1
Wnt/B-catenin Signaling	No	Yes	No	No	1
Xenobiotic Metabolism Signaling	No	No	Yes	No	1

Additional File 9: Biological functions significantly (p-value < 0.005) associated with predicted miRNA targets of miR-33, miR-330, miR-181a, and miR-10b

---

Functions	p-value
1. miR-33	
Amino Acid Metabolism	0.0034
Behavior	0.0001
Cancer	0.0041
Carbohydrate Metabolism	0.0041
Cardiovascular Disease	0.0046
Cell Cycle	0.0009
Cell Death	0.0041
Cell Morphology	0.0003
Cell-mediated Immune Response	0.0041
Cell-To-Cell Signaling and Interaction	0.0036
Cellular Assembly and Organization	0.0036
Cellular Compromise	0.0041
Cellular Development	0.0003
Cellular Function and Maintenance	0.0041
Cellular Growth and Proliferation	0.0041
Cellular Movement	0.0041
Connective Tissue Development and Function	0.0041
Drug Metabolism	0.0041
Embryonic Development	0.0003
Endocrine System Development and Function	0.0041
Gene Expression	0.0041
Genetic Disorder	0.0002
Hair and Skin Development and Function	0.0017
Hematological Disease	0.0041
Hematological System Development and Function	0.0041
Hematopoiesis	0.0041
Immunological Disease	0.0041
Inflammatory Response	0.0041
Lipid Metabolism	0.0041
Metabolic Disease	0.0041
Molecular Transport	0.0041
Nervous System Development and Function	0.0041
Neurological Disease	0.0002
Organ Development	0.0041
Organ Morphology	0.0041
Organismal Injury and Abnormalities	0.0041
Post-Translational Modification	0.0034
Protein Synthesis	0.0025
Psychological Disorders	0.0002
Renal and Urological System Development and Function	0.0009
Reproductive System Development and Function	0.0041
Skeletal and Muscular Disorders	0.0014

Functions	p-value
Skeletal and Muscular System Development and Function	0.0040
Small Molecule Biochemistry	0.0034
Tissue Development	0.0040
Vitamin and Mineral Metabolism	0.0041
2. miR-330	
Carbohydrate Metabolism	0.0006
Cell Cycle	0.0043
Cell Morphology	0.0038
Cell-mediated Immune Response	0.0038
Cell-To-Cell Signaling and Interaction	0.0011
Cellular Development	0.0028
Cellular Function and Maintenance	0.0002
Cellular Growth and Proliferation	0.0028
Cellular Movement	0.0002
Embryonic Development	0.0019
Endocrine System Development and Function	0.0028
Gene Expression	0.0028
Hematological System Development and Function	0.0038
Immune Cell Trafficking	0.0038
Inflammatory Response	0.0019
Lipid Metabolism	0.0006
Molecular Transport	0.0018
Nervous System Development and Function	0.0004
Organismal Development	0.0013
RNA Trafficking	0.0028
Small Molecule Biochemistry	0.0006
Tissue Development	0.0019
3. miR-181a	
Cancer	0.0012
Carbohydrate Metabolism	0.0032
Cardiovascular Disease	0.0000
Cell Cycle	0.0001
Cell Death	0.0001
Cell Morphology	0.0012
Cell-mediated Immune Response	0.0038
Cell-To-Cell Signaling and Interaction	0.0004
Cellular Assembly and Organization	0.0015
Cellular Development	0.0012
Cellular Function and Maintenance	0.0012
Cellular Growth and Proliferation	0.0004
Cellular Movement	0.0012
Connective Tissue Development and Function	0.0012
Connective Tissue Disorders	0.0012
Dermatological Diseases and Conditions	0.0012
DNA Replication, Recombination, and Repair	0.0038
Drug Metabolism	0.0036
Endocrine System Development and Function	0.0002
Endocrine System Disorders	0.0020

Functions	p-value
Gastrointestinal Disease	0.0005
Gene Expression	0.0001
Genetic Disorder	0.0003
Hair and Skin Development and Function	0.0012
Hematological Disease	0.0012
Hematological System Development and Function	0.0038
Hematopoiesis	0.0038
Immune Cell Trafficking	0.0038
Immunological Disease	0.0012
Infection Mechanism	0.0010
Inflammatory Disease	0.0005
Inflammatory Response	0.0012
Lipid Metabolism	0.0023
Metabolic Disease	0.0020
Molecular Transport	0.0012
Nervous System Development and Function	0.0003
Neurological Disease	0.0005
Nucleic Acid Metabolism	0.0012
Organ Development	0.0017
Organismal Development	0.0030
Organismal Functions	0.0023
Organismal Injury and Abnormalities	0.0019
Organismal Survival	0.0020
Respiratory Disease	0.0023
Respiratory System Development and Function	0.0004
Skeletal and Muscular Disorders	0.0012
Skeletal and Muscular System Development and Function	0.0002
Small Molecule Biochemistry	0.0004
Tissue Development	0.0002
Tissue Morphology	0.0004
Tumor Morphology	0.0012
4. miR-10b	
Amino Acid Metabolism	0.0015
Auditory and Vestibular System Development and Function	0.0046
Behavior	0.0030
Cardiovascular Disease	0.0015
Cardiovascular System Development and Function	0.0001
Cell Death	0.0015
Cell Morphology	0.0015
Cell-To-Cell Signaling and Interaction	0.0015
Cellular Assembly and Organization	0.0015
Cellular Compromise	0.0017
Cellular Development	0.0000
Cellular Function and Maintenance	0.0015
Cellular Growth and Proliferation	0.0013
Cellular Movement	0.0015
Connective Tissue Development and Function	0.0016
Connective Tissue Disorders	0.0030

Functions	p-value
Developmental Disorder	0.0046
Digestive System Development and Function	0.0015
Embryonic Development	0.0010
Endocrine System Development and Function	0.0002
Endocrine System Disorders	0.0046
Gastrointestinal Disease	0.0046
Gene Expression	0.0003
Genetic Disorder	0.0046
Hair and Skin Development and Function	0.0030
Hematological Disease	0.0015
Hematological System Development and Function	0.0015
Hematopoiesis	0.0015
Humoral Immune Response	0.0015
Immune Cell Trafficking	0.0046
Immunological Disease	0.0046
Inflammatory Disease	0.0030
Inflammatory Response	0.0046
Lymphoid Tissue Structure and Development	0.0030
Nervous System Development and Function	0.0015
Neurological Disease	0.0002
Organ Development	0.0002
Organ Morphology	0.0015
Organismal Development	0.0002
Organismal Functions	0.0015
Organismal Injury and Abnormalities	0.0015
Reproductive System Development and Function	0.0046
RNA Post-Transcriptional Modification	0.0030
Skeletal and Muscular Disorders	0.0030
Skeletal and Muscular System Development and Function	0.0016
Small Molecule Biochemistry	0.0015
Tissue Development	0.0010
Tissue Morphology	0.0001
Visual System Development and Function	0.0015

Additional File 10: Biological functions significantly (p-value < 0.005) associated with formaldehyde-responsive genes, as identified through pathway analysis of the Li et. al. (2007) genomic database

---

Functions	p-value
Cell Death	1.65E-22
Cellular Growth and Proliferation	1.17E-19
Cancer	1.87E-19
Gene Expression	4.40E-16
Cell Cycle	1.51E-15
Cellular Development	6.49E-13
Developmental Disorder	7.45E-12
Reproductive System Disease	2.93E-11
Cardiovascular System Development and Function	2.38E-10
Organismal Development	2.38E-10
Organismal Survival	3.94E-10
Hematological System Development and Function	1.93E-09
Hematopoiesis	1.93E-09
Tissue Development	5.73E-09
Cell-mediated Immune Response	1.09E-08
Cellular Function and Maintenance	1.09E-08
Connective Tissue Disorders	1.10E-08
Immunological Disease	1.10E-08
Inflammatory Disease	1.10E-08
Skeletal and Muscular Disorders	1.10E-08
Tissue Morphology	5.86E-08
Gastrointestinal Disease	5.86E-08
Skeletal and Muscular System Development and Function	7.68E-08
DNA Replication, Recombination, and Repair	1.47E-07
Cardiovascular Disease	3.01E-07
Organ Development	4.23E-07
Genetic Disorder	4.39E-07
Neurological Disease	4.39E-07
Embryonic Development	4.92E-07
Reproductive System Development and Function	4.92E-07
Connective Tissue Development and Function	7.51E-07
Cellular Compromise	9.51E-07
Cellular Movement	9.84E-07
Cell Morphology	1.74E-06
Hematological Disease	4.32E-06
Infection Mechanism	5.28E-06
Nervous System Development and Function	5.28E-06
Hair and Skin Development and Function	1.06E-05

Functions	p-value
Molecular Transport	1.53E-05
Protein Synthesis	1.53E-05
Post-Translational Modification	2.79E-05
Protein Folding	2.79E-05
Cell Signaling	5.05E-05
Small Molecule Biochemistry	5.05E-05
Digestive System Development and Function	8.26E-05
Hepatic System Development and Function	8.26E-05
Cell-To-Cell Signaling and Interaction	1.31E-04
Behavior	1.40E-04
Organ Morphology	1.40E-04
Organismal Injury and Abnormalities	1.47E-04
Cellular Assembly and Organization	1.59E-04
Lymphoid Tissue Structure and Development	2.00E-04
Respiratory Disease	2.03E-04
Metabolic Disease	2.40E-04
Tumor Morphology	2.84E-04
Endocrine System Disorders	3.15E-04
Dermatological Diseases and Conditions	3.39E-04
Lipid Metabolism	6.94E-04
RNA Trafficking	6.94E-04
Renal and Urological Disease	6.94E-04

5 REFERENCES

Abba, M. C., Sun, H., Hawkins, K. A., Drake, J. A., Hu, Y., *et al.* (2007). Breast Cancer Molecular Signatures as Determined by SAGE: Correlation with Lymph Node Status. *Mol Cancer Res* **5**, 881-890.

Adam, E., Hansen, K. K., Astudillo, O. F., Coulon, L., Bex, F., *et al.* (2006). The House Dust Mite Allergen Der p 1, Unlike Der p 3, Stimulates the Expression of Interleukin-8 in Human Airway Epithelial Cells via a Proteinase-activated Receptor-2-independent Mechanism. *J Biol Chem* **281**, 6910-6923.

American Lung Association (2010). State of the Air. Available at: www.stateoftheair.org. Accessed July 6, 2010.

Anderson, P. (2010). Post-transcriptional regulons coordinate the initiation and resolution of inflammation. *Nat Rev Immunol* **10**, 24-35.

Armendariz, A. D. and Krauss, R. M. (2009). Hepatic nuclear factor 1-[alpha]: inflammation, genetics, and atherosclerosis. *Curr Opin Lipidol* **20**, 106-111.

Bachand, A. M., Mundt, K. A., Mundt, D. J. and Montgomery, R. R. (2010). Epidemiological studies of formaldehyde exposure and risk of leukemia and nasopharyngeal cancer: A meta-analysis. *Crit Rev Toxicol* **40**, 85-100, 10.3109/10408440903341696.

Bakand, S., Winder, C., Khalil, C. and Hayes, A. (2005). Toxicity Assessment of Industrial Chemicals and Airborne Contaminants: Transition from In Vivo to In Vitro Test Methods: A Review. *Inhalation Toxicology* **17**, 775-787.

Bartel, D. P. (2004). MicroRNAs: Genomics, Biogenesis, Mechanism, and Function. *Cell* **116**, 281-297.

Becker, S., Mundandhara, S., Devlin, R. B. and Madden, M. (2005). Regulation of cytokine production in human alveolar macrophages and airway epithelial cells in response to ambient air pollution particles: Further mechanistic studies. *Toxicol Appl Pharmacol* **207**, 269-275.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J R Stat Soc B* **57**, 289-300.

Blank, F., Rothen-Rutishauser, B. M., Schurch, S. and Gehr, P. (2006). An Optimized In Vitro Model of the Respiratory Tract Wall to Study Particle Cell Interactions. *J Aerosol Med* **19**, 392-405.

Bloemen, K., Verstraelen, S., Van Den Heuvel, R., Witters, H., Nelissen, I., *et al.* (2007). The allergic cascade: Review of the most important molecules in the asthmatic lung. *Immunology Letters* **113**, 6-18.

Bluteau, O., Jeannot, E., Bioulac-Sage, P., Marques, J., Blanc, J., *et al.* (2002). Bi-allelic inactivation of TCF1 in hepatic adenomas. *Nat Genet* **32**, 312-315.

Brunekreef, B., Dockery, D. W. and Krzyzanowski, M. (1995). Epidemiologic studies on short-term effects of low levels of major ambient air pollution components. *Environ Health Perspect* **103(suppl 2)**, 3-13.

Brunekreef, B. and Holgate, S. T. (2002). Air pollution and health. *The Lancet* **360**, 1233-1242.

Burnett, R. T., Brook, J., Dann, T., Delocla, C., Philips, O., *et al.* (2000). Association between particulate- and gas-phase components of urban air pollution and daily mortality in eight Canadian cities. *Inhalation Toxicol* **12**, 15-39.

Burnett, R. T., Brook, J., Dann, T., Delocla, C., Philips, O., *et al.* (2001). Association between particulate- and gas-phase components of urban air pollution and daily mortality in eight canadian cities. *Inhalation Toxicology* **12**, 15-39, 10.1080/08958370050164851.

Calin, G. and Croce, C. (2006). MicroRNA signatures in human cancers. *Nat Rev Cancer* **6**, 857-866.

Cao, D., Tal, T. L., Graves, L. M., Gilmour, I., Linak, W., *et al.* (2007). Diesel exhaust particulate-induced activation of Stat3 requires activities of EGFR and Src in airway epithelial cells. *Am J Physiol Lung Cell Mol Physiol* **292**, L422-429.

Cao, J., Li, W., Tan, J., Song, W., Xu, X., *et al.* (2009). Association of ambient air pollution with hospital outpatient and emergency room visits in Shanghai, China. *Sci Total Environ* **407**, 5531-5536.

Cho, S.-B. and Ryu, J. (2002). Classifying gene expression data of cancer using classifier ensemble with mutually exclusive features. *Proceedings of the IEEE* **90**, 1744-1753.

Cohen, A. J., Anderson, H. R., Ostro, B., Pandey, K. D., Krzyzanowski, M., *et al.* (2005). The Global Burden of Disease Due to Outdoor Air Pollution. *Journal of Toxicology and Environmental Health, Part A: Current Issues* **68**, 1301-1307.

Dab, W., Medina, S., Quenel, P., Le Moullec, Y., Le Tertre, A., *et al.* (1996). Short term respiratory health effects of ambient air pollution: results of the APHEA project in Paris. *J Epidemiol Community Health* **50**, s42-s46.

Damera, G., Zhao, H., Wang, M., Smith, M., Kirby, C., *et al.* (2009). Ozone modulates IL-6 secretion in human airway epithelial and smooth muscle cells. *Am J Physiol Lung Cell Mol Physiol* **296**, L674-683.

Dasgupta, P. K., Dong, S., Hwang, H., Yang, H.-C. and Genfa, Z. (1988). Continuous liquid-phase fluorometry coupled to a diffusion scrubber for the real-time determination of atmospheric formaldehyde, hydrogen peroxide and sulfur dioxide. *Atmos Environ* **22**, 949-963.

Dhanasekaran, S. M., Barrette, T. R., Ghosh, D., Shah, R., Varambally, S., *et al.* (2001). Delineation of prognostic biomarkers in prostate cancer. *Nature* **412**, 822-826.

Divakaran, V. and Mann, D. L. (2008). The Emerging Role of MicroRNAs in Cardiac Remodeling and Heart Failure. *Circulation Research* **103**, 1072-1083.

Edgar, R., Domrachev, M. and Lash, A. E. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* **30**, 207-10.

Ellard, S. and Colclough, K. (2006). Mutations in the genes encoding the transcription factors hepatocyte nuclear factor 1 alpha (HNF1A) and 4 alpha (HNF4A) in maturity-onset diabetes of the young. *Hum Mutat* **27**, 854-869.

Environmental Protection Agency (2006). Air Quality Criteria for Ozone and Related Photochemical Oxidants (2006 Final). Washington, DC. EPA/600/R-05/004aF-cF.

EPA (2007). Formaldehyde Hazard Summary. Available at: http://www.epa.gov/ttn/atw/hlthef/formalde.html. Accessed Jan 27.

Fabbri, M., Garzon, R., Andreeff, M., Kantarjian, H. M., Garcia-Manero, G., *et al.* (2008). MicroRNAs and noncoding RNAs in hematological malignancies: molecular, clinical and therapeutic implications. *Leukemia* **22**, 1095-1105.

Filipowicz, W., Bhattacharyya, S. N. and Sonenberg, N. (2008). Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? *Nat Rev Genet* **9**, 102-114.

Foster, K. A., Oster, C. G., Mayer, M. M., Avery, M. L. and Audus, K. L. (1998). Characterization of the A549 Cell Line as a Type II Pulmonary Epithelial Cell Model for Drug Metabolism. *Experimental Cell Research* **243**, 359-366.

Golub, T. R., Slonim, D. K., Tamayo, P., Huard, C., Gaasenbeek, M., *et al.* (1999). Molecular Classification of Cancer: Class Discovery and Class Prediction by Gene Expression Monitoring. *Science* **286**, 531-537.

Götschi, T., Heinrich, J., Sunyer, J. and Künzli, N. (2008). Long-Term Effects of Ambient Air Pollution on Lung Function: A Review. *Epidemiology* **19**, 690-701.

Grubbs, F. E. (1969). Procedures for Detecting Outlying Observations in Samples. *Technometrics* **11**, 1-21.

Hatzis, P. and Talianidis, I. (2001). Regulatory Mechanisms Controlling Human Hepatocyte Nuclear Factor 4alpha Gene Expression. *Mol Cell Biol* **21**, 7320-7330.

Hébert, S. S. and De Strooper, B. (2009). Alterations of the microRNA network cause neurodegenerative disease. *Trends in Neurosciences* **32**, 199-206.

International Agency for Research on Cancer (2006). IARC Monographs on the Evaluation of Carcinogenic Risks to Humans: Formaldehyde, 2-butoxyethanol and 1-tert-butoxypropan-2-ol.

World Health Organization (2010). Agents Classified by the IARC Monographs. Available at: http://monographs.iarc.fr/ENG/Classification/index.php. Accessed July 21 2010.

Iorio, M. V., Ferracin, M., Liu, C. G., Veronese, A., Spizzo, R., *et al.* (2005). MicroRNA Gene Expression Deregulation in Human Breast Cancer. *Cancer Research* **65**, 7065-7070.

Irizarry, R. A., Bolstad, B. M., Collin, F., Cope, L. M., Hobbs, B., *et al.* (2003). Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res* **31**, e15.

Izzotti, A., Calin, G. A., Arrigo, P., Steele, V. E., Croce, C. M., *et al.* (2009). Downregulation of microRNA expression in the lungs of rats exposed to cigarette smoke. *The FASEB Journal* **23**, 806-812.

Jardim, M. J., Fry, R. C., Jaspers, I., Dailey, L. and Diaz-Sanchez, D. (2009). Disruption of MicroRNA Expression in Human Airway Cells by Diesel Exhaust Particles Is Linked to Tumorigenesis-Associated Pathways. *Environ Health Perspect* **117**.

Jaspers, I., Flescher, E. and Chen, L. C. (1997). Ozone-induced IL-8 expression and transcription factor binding in respiratory epithelial cells. *Am J Physiol Lung Cell Mol Physiol* **272**, L504-511.

Jeffries, H., Fox, D. and Kamens, R. (1976). Outdoor smog chamber studies: light effects relative to indoor chambers. *Environmental Science & Technology* **10**, 1006-1011, 10.1021/es60121a016.

Jeffries, H. E. (1995). Photochemical air pollution (H. B. Singh), pp. 308-348. Van Nostrand-Reinhold, New York.

Jeffries, H. E. and Sexton, K. G. (1995). The relative ozone forming potential of methanol-fueled vehicle emissions and gasoline-fueled vehicle emissions in outdoor smog chambers. Chapel Hill, NC. Final report to the Coordinating Research Council for CRC Project No. ME-1.

Jelinsky, S. A., Estep, P., Church, G. M. and Samson, L. D. (2000). Regulatory Networks Revealed by Transcriptional Profiling of Damaged Saccharomyces cerevisiae Cells: Rpn4 Links Base Excision Repair with Proteasomes. *Molecular and Cellular Biology* **20**, 8157-8167.

Jorens, P., Damme, J. V., Backer, W. D., Bossaert, L., DeJongh, R. F., *et al.* (1992). Interleukin 8 (IL-8) in the bronchoalveolar lavage fluid from patients with the adult respiratory distress syndrome (ARDS) and patients at risk for ARDS. *Cytokine* **4**, 592-7.

Kang, J., Shakya, A. and Tantin, D. (2009). Stem cells, stress, metabolism and cancer: a drama in two Octs. *Trends Biochem Sci* **34**, 491-499.

Karin, M. and Greten, F. R. (2005). NF-kB: linking inflammation and immunity to cancer development and progression. *Nat Rev Immunol* **5**, 749-759.

Kerns, W. D., Pavkov, K. L., Donofrio, D. J., Gralla, E. J. and Swenberg, J. A. (1983). Carcinogenicity of Formaldehyde in Rats and Mice after Long-Term Inhalation Exposure. *Cancer Research* **43**, 4382-4392.

Kim, W. J., Terada, N., Nomura, T., Takahashi, R., Lee, S. D., *et al.* (2002). Effect of formaldehyde on the expression of adhesion molecules in nasal microvascular endothelial cells: the role of formaldehyde in the pathogenesis of sick building syndrome. *Clinical & Experimental Allergy* **32**, 287-295.

Lee, K. H., Chen, Y. L., Yeh, S. D., Hsiao, M., Lin, J. T., *et al.* (2009). MicroRNA-330 acts as tumor suppressor and induces apoptosis of prostate cancer cells through E2F1-mediated suppression of Akt phosphorylation. *Oncogene* **28**, 3360-3370.

Lee, Y. S. and Dutta, A. (2007). The tumor suppressor microRNA let-7 represses the HMGA2 oncogene. *Genes Dev* **21**, 1025-1030.

Li, G. Y., Lee, H. Y., Shin, H. S., Kim, H. Y., Lim, C. H., *et al.* (2007). Identification of gene markers for formaldehyde exposure in humans. *Environ Health Perspect* **115**, 1460-1466.

Lu, J., Getz, G., Miska, E. A., Alvarez-Saavedra, E., Lamb, J., *et al.* (2005). MicroRNA expression profiles classify human cancers. *Nature* **435**, 834-838.

Ma, L., Teruya-Feldstein, J. and Weinberg, R. A. (2007). Tumour invasion and metastasis initiated by microRNA-10b in breast cancer. *Nature* **449**, 682-688.

Mantovani, A., Allavena, P., Sica, A. and Balkwill, F. (2008). Cancer-related inflammation. *Nature* **454**, 436-444.

Marcucci, G., Radmacher, M., Mrózek, K. and Bloomfield, C. (2009). MicroRNA expression in acute myeloid leukemia. *Curr Hematol Malig Rep* **4**, 83-88.

Marsh, G. M. and Youk, A. O. (2004). Reevaluation of mortality risks from leukemia in the formaldehyde cohort study of the National Cancer Institute. *Regul Toxicol Pharmacol* **40**, 113-124.

McClintock, D., Zhuo, H., Wickersham, N., Matthay, M. and Ware, L. (2008). Biomarkers of inflammation, coagulation and fibrinolysis predict mortality in acute lung injury. *Critical Care* **12**, R41.

McHale, C. M., Zhang, L., Hubbard, A. E. and Smith, M. T. (2010). Toxicogenomic profiling of chemically exposed humans in risk assessment. *Mutat Res* **doi: 10.1016/j.mrrev.2010.04.001**.

Ming, G. and Song, H. (2005). Adult Neurogenesis in the Mammalian Central Nervous System. *Annu Rev Neurosci* **28**, 223-250, 10.1146/annurev.neuro.28.051804.101459.

Myatt, S. S. and Lam, E. W. F. (2007). The emerging roles of forkhead box (Fox) proteins in cancer. *Nat Rev Cancer* **7**, 847-859.

Nyberg, F., Gustavsson, P., Jarup, L., Bellander, T., Berglind, N., *et al.* (2000). Urban Air Pollution and Lung Cancer in Stockholm. *Epidemiology* **11**, 487-495.

O'Hara, S. P., Mott, J. L., Splinter, P. L., Gores, G. J. and LaRusso, N. F. (2009). MicroRNAs: Key Modulators of Posttranscriptional Gene Expression. *Gastroenterology* **136**, 17-25.

Overton, J. H., Kimbell, J. S. and Miller, F. J. (2001). Dosimetry Modeling of Inhaled Formaldehyde: The Human Respiratory Tract. *Toxicological Sciences* **64**, 122-134.

Pinkerton, L. E., Hein, M. J. and Stayner, L. T. (2004). Mortality among a cohort of garment workers exposed to formaldehyde: an update. *Occup Environ Med* **61**, 193-200.

Reich, M., Liefeld, T., Gould, J., Lerner, J., Tamayo, P., *et al.* (2006). GenePattern 2.0. *Nat Genet* **38**, 500-501.

Rumchev, K. B., Spickett, J. T., Bulsara, M. K., Phillips, M. R. and Stick, S. M. (2002). Domestic exposure to formaldehyde significantly increases the risk of asthma in young children. *Eur Respir J* **20**, 403-408.

Schmid, J. A. and Birbach, A. (2008). IkB kinase B (IKKB/IKK2/IKBKB) - A key molecule in signaling to the transcription factor NF-kB. *Cytokine & Growth Factor Reviews* **19**, 157-165.

Sexton, K. G., Jeffries, H. E., Jang, M., Kamens, R. M., Doyle, M., *et al.* (2004). Photochemical Products in Urban Mixtures Enhance Inflammatory Responses in Lung Cells. *Inhalation Toxicol* **16**, 107-114.

Smith, K. R. and Mehta, S. (2003). The burden of disease from indoor air pollution in developing countries: comparison of estimates. *Int J Hyg Environ Health* **206**, 279-289.

Smith, K. R., Samet, J. M., Romieu, I. and Bruce, N. (2000). Indoor air pollution in developing countries and acute lower respiratory infections in children. *Thorax* **55**, 518-532.

Speit, G., Neuss, S. and Schmid, O. (2010). The human lung cell line A549 does not develop adaptive protection against the DNA-damaging action of formaldehyde. *Environmental and Molecular Mutagenesis* **51**, 130-137.

Storey, J. D. (2003). The Positive False Discovery Rate: A Bayesian Interpretation and the q-Value. *The Annals of Statistics* **31**, 2013-2035.

Thompson, A. M. S., Zanobetti, A., Silverman, F., Schwartz, J., Coull, B., *et al.* (2009). Baseline Repeated Measures from Controlled Human Exposure Studies: Associations between Ambient Air Pollution Exposure and the Systemic Inflammatory Biomarkers IL-6 and Fibrinogen. *Environ Health Perspect* **118**.

Thompson, C. M., Subramaniam, R. P. and Grafström, R. C. (2008). Mechanistic and dose considerations for supporting adverse pulmonary physiology in response to formaldehyde. *Toxicol Appl Pharmacol* **233**, 355-359.

Tsukue, N., Okumura, H., Ito, T., Sugiyama, G. and Nakajima, T. (2010). Toxicological evaluation of diesel emissions on A549 cells. *Toxicology in Vitro* **24**, 363-369.

Tuthill, R. W. (1984). Woodstoves, formaldehyde, and respiratory disease. *Am. J. Epidemiol* **120**, 952-955.

Vaughan, T. L., Stewart, P. A., Teschke, K., Lynch, C. F., Swanson, G. M., *et al.* (2000). Occupational exposure to formaldehyde and wood dust and nasopharyngeal carcinoma. *Occup and Environ Med* **57**, 376-384.

Wang, X. (2008). miRDB: A microRNA target prediction and functional annotation database with a wiki interface. *RNA* **14**, 1012-1017.

Wang, X. and El Naqa, I. M. (2008). Prediction of both conserved and nonconserved microRNA targets in animals. *Bioinformatics* **24**, 325-332.

Welch, B. L. (1938). The Significance of the Difference Between Two Means when the Population Variances are Unequal. *Biometrika* **29**, 350-362.

WHO (2001). Chapter 5.8 Formaldehyde. Copenhagen, Denmark. Air Quality Guidelines, Second Edition.

Wieslander, G., Norbäck, D., Björnsson, E., Janson, C. and Boman, G. (1997). Asthma and the indoor environment: the significance of emission of formaldehyde and volatile organic compounds from newly painted indoor surfaces. *Int Arch Occup Environ Health* **69**, 115-124.

Yanaihara, N., Caplen, N., Bowman, E., Seike, M., Kumamoto, K., *et al.* (2006). Unique microRNA molecular profiles in lung cancer diagnosis and prognosis. *Cancer Cell* **9**, 189-198.

Yu, J., Jeffries, H. E. and Sexton, K. G. (1997). Atmospheric photooxidation of alkylbenzenes--I. Carbonyl product analyses. *Atmospheric Environment* **31**, 2261-2280.

Zhang, L., Freeman, L. E. B., Nakamura, J., Hecht, S. S., Vandenberg, J. J., *et al.* (2009). Formaldehyde and leukemia: Epidemiology, potential mechanisms, and implications for risk assessment. *Environ Mol Mutagen* **51**, 181-191.

Zhang, L., Tang, X., Rothman, N., Vermeulen, R., Ji, Z., *et al.* (2010). Occupational Exposure to Formaldehyde, Hematotoxicity, and Leukemia-Specific Chromosome Changes in Cultured Myeloid Progenitor Cells. *Cancer Epidemiol Biomarkers Prev* **19**, 80-88.