Short-term transcriptome and microRNAs responses to exposure to different air pollutants in two population studies

Co-authors: Almudena Espín-Pérez, Julian Krauskopf, Marc Chadeau-Hyam, Karin van Veldhoven, Fan Chung, Paul Cullinan, Jolanda Piepers, Marcel van Herwijnen, Nadine Kubesch, Glòria Carrasco-Turigas, Mark Nieuwenhuijsen, Paolo Vineis, Jos C.S. Kleinjans, Theo M.C.M. de Kok

Abstract

Diesel vehicle emissions are the major source of genotoxic compounds in ambient air from urban areas. These pollutants are linked to risks of cardiovascular diseases, lung cancer, respiratory infections and adverse neurological effects. Biological events associated with exposure to some air pollutants are widely unknown but applying omics techniques may help to identify the molecular processes that link exposure to disease risk. Most data on health risks are related to long-term exposure, so the aim of this study is to investigate the impact of short-term exposure (two hours) to air pollutants on the blood transcriptome and microRNA expression levels.

We analyzed transcriptomics and microRNA expression using microarray technology on blood samples from volunteers participating in studies in London, the Oxford Street cohort, and, in Barcelona, the TAPAS cohort. Personal exposure levels measurements of particulate matter (PM_{10}, $PM_{2.5}$), ultrafine particles (UFPC), nitrogen oxides (NO_2, NO and NOx), black carbon (BC) and carbon oxides (CO and CO_2) were registered for each volunteer. Associations between air pollutant levels and gene/microRNA expression were evaluated using multivariate normal models (MVN).

MVN-models identified compound-specific expression of blood cell genes and microRNAs associated with air pollution despite the low exposure levels, the short exposure periods and the relatively small-sized cohorts. Hsa-miR-197-3p, hsa-miR-29a-3p, hsa-miR-15a-5p, hsa-miR-16-5p and hsa-miR-92a-3p are found significantly expressed in association with exposures. These microRNAs target also relevant transcripts, indicating their potential relevance in the research of omics-biomarkers responding to air pollution. Furthermore, these microRNAs are also known to be associated with diseases previously linked to air pollution exposure including several cancers such lung cancer and Alzheimer's disease. In conclusion, we identified in this study promising compound-specific mRNA and microRNA biomarkers after two hours of exposure to low levels of air pollutants during two hours that suggest increased cancer risks.

Introduction

Vehicle emissions have become the principal source of air pollutants in urban areas close to roadways. Diesel engines emit lower levels of carbon monoxide and carbon dioxide than gasoline engines but they contribute the most to atmospheric particulate pollution [1]. About 90% of traffic-generated particulate matter in urban environments originates from diesel engine emissions [2]. In addition, exhaust from diesel vehicles is the major source of genotoxic carcinogens in ambient air in urban areas [3] and is classified as IARC Group 1 [4].

Epidemiological studies have shown that these pollutants may have a serious impact on human health. Next to increasing cancer risks [5, 6] air pollution may induce several other diseases and medical conditions such as cardiovascular disease [7-9], stroke [10, 11], respiratory infections [12, 13] and adverse neurological effects [14-16]. Long-term exposure to air pollution has been identified as a leading cause of global disease burden according to "The Global Burden of Diseases, Injuries, and Risk Factors Study 2015", with a higher incidence in countries with low/middle-income [17]. In 2015, a total of 4.2 million deaths and 103.1 million lost years of healthy life was estimated as a consequence of long-term exposure to the particulate matter $PM_{2.5}$ [17]. Previous studies have linked mortality to hourly peak particulate matter $PM_{2.5}$ [18] and increased asthma risk and cardiovascular mortality to particulate matter $PM_{2.5}$ and PM_{10} [19,20]. Similar studies have monitored personal exposure measurement to ultrafine particles over time [21]. The impact of short-term exposure to particulate matter in human however is still uncertain and cannot be easily addressed using the classical epidemiological approaches. In order to study risks associated with short-term exposures we need rapidly responding biomarkers. For this, omics-based technologies may be considered as these allow the global and sensitive identification of molecular changes that can be relevant for monitoring the development of disease.

Air pollution exposure induces gene expression profiles identified through transcriptomics analysis [22,23], which can be used as biomarkers of exposure [24]. In addition, microRNAs can play a role in mediating transcriptomic responses to air pollution by controlling gene expression at the post-transcriptional level [25]. Studying microRNAs expression as complementary to the transcriptomics analysis may therefore provide a better understanding of the gene expression regulatory mechanisms in response to exposure to ambient air pollutants.

Therefore, in this study, we aim to investigate the impact of short-term air pollution exposure on whole-genome gene and microRNA expression, for the purpose of identifying potential biomarkers of short- term exposure response and exploring associated health risks. The transcriptomics and microRNA expressions were analyzed in blood samples of volunteers walking for 2 hr along traffic/dense streets in London, the Oxford Street cohort, and, in Barcelona, the TAPAS cohort. Personal exposure levels measurements were performed for each participant. Then, we used linear mixed models to correct for technical variables and multivariate normal models to find associations between the exposures and gene/microRNA expression while correcting for confounders. We examined each of the following exposures individually: NO_2, NO, NOx, PM_{10}, $PM_{2.5}$, BC, UFPC, CO and CO_2.

Methods

- Study design of the Oxford Street study

The Oxford Street cohort was a randomized, crossover study, from the EXPOsOMICS project) (population characteristics in Table S1 from Supplementary Material). A total number of 59 volunteers spent two hours walking along Oxford Street (from 10:30 to 12:30), one of London's busiest roads where only diesel-powered buses and taxis are allowed to circulate. On a different day (separated by 3-8 weeks) the same volunteers walked again for two hours in Hyde Park, one of the largest parks in London. The individuals walked around 6km at each location, resting every 30 minutes for 15 minutes. The pollution levels (particulate matter, black carbon and nitrogen dioxide), together with records of physical activity and location, were monitored using a portable kit (backpack with air pollution sensors, batteries, smartphones with GPS and accelerometer).

For each volunteer in each exposure situation, a blood sample was taken as input for transcriptomics analysis. Volunteers were driven after the experiment by an electric-powered car from Oxford Street or Hyde Park to the Royal Brompton Hospital to donate the blood samples. In order to preserve gene expression for omics analysis, blood samples were mixed with RNAlater after collection and stored at -80°C. Participants were randomly allocated to Oxford Street or Hyde Park as the first location. Age, sex and BMI were registered for each subject.

The 59 volunteers were non-smokers, comprising healthy individuals (20 participants) and subjects with heart or lung condition, more specifically Chronic Obstructive Pulmonary Disease (COPD) (20 participants) and Ischemic Heart Disease (IHD) (19 participants). We ran an unsupervised Principal Component Analysis (PCA) to assess potential differences in gene expression between groups. The PCA analysis does not cluster gene expression based on the disease groups (Figure S2 from Supplementary Material), so in order to ensure higher statistical power we did not stratify the population into subgroups but we added this variable as an additional confounder in the statistical model.

The study was approved by the ethics committee (Brompton Hospital, London) and written informed consent forms were provided by all subjects.

- Study design of TAPAS

The TAPAS cohort was also a crossover study and it comprised four different scenarios. In addition to a low and high exposure location (Barceloneta and Ronda in Barcelona, Spain), the study design included rest and intermittent exercise (15 minutes cycling and 15 minutes break) within the two locations. The physical activity was modulated by cycling at such a pace that the heart rate of the participants fell within the range of 50-70% of their maximum rate. A total of 30 individuals (15 females and 15 males), completed one of the four different scenarios per day. The inclusion criteria for the subjects were healthy non-smokers, in the age range of 18-60, not taking any medication and not suffering from any chronic disease. Age, sex and BMI were also registered for each participant.

The subjects were exposed from 8:00 to 10:00 at one of the two locations (heavily polluted air due to the traffic rush hour and low levels of air pollution because of low dense traffic). Immediately after two hours of exposure, participants were transported by a van with cycle-ventilated and windows closed to a nearby clinic where blood samples were collected. The blood samples were stored at -80°C within two hours after collection.

The study was approved by the Ethic Review Committee of the Institut Municipal d'Investigatió Mèdica and all participants gave written informed consent.

- Exposures

Ambient concentrations of a mixture of air pollutants were measured throughout the experiment. In the Oxford Street study, these pollutants were: 1) particulate matter with diameters of 10 micrometers or smaller (PM_{10}) and 2) 2.5 micrometers or smaller ($PM_{2.5}$) measured using an air sampler (16 liters per minute), 3) ultrafine particles counts (UFPC) measured with a real-time condensation particle counter (Model 3007, TSI), 4) nitrogen dioxide (NO_2) measured using C18Sep-Pak cartridges coated with potassium hydroxide and triethanolamine followed by ion chromatography and 5) black carbon (BC) measured according to the National Institute for Occupational Safety and Health guidelines (method 5040, Sunset Laboratory) [26].

In the TAPAS cohort more exposures were measured: PM_{10} and $PM_{2.5}$ (DustTrack, DRX,Model 8534, TSI, Minnesota, USA), UFPC (condensation particle counter, CPC, Model 3007, TSI, Minnesota, USA), BC (portable aetholometer, Model AE-51, McAgee Scientific, California, USA) and additionally nitrogen oxide (NO), generic measurement for the nitrogen oxides (NO_x=NO+NO_2) (nitric oxide monitor, Model 410 Nitric Oxide Monitor, 2B Technologies, Colorado, USA, in combination with a nitrogen dioxide converter, Model 401 NO2 Converter, 2B Technologies, Colorado, USA), carbon monoxide (CO) and carbon dioxide (CO_2) [27].

The different personal exposure levels were compared for both locations (low versus high exposure level) and for each cohort using a paired t-test (R package "stats", "t.test( )" function). Correlation among exposures was also performed (R package "stats", "cor ()" function).

- RNA isolation and microarray analysis

The RiboPureTM-Blood kit (Ambion) was used to isolate total RNA from the blood samples (400 μl of whole blood and 1600 μl of RNA later) following the manufacturer's instructions. The isolated RNA was hybridized on Agilent 8x60K Whole Human Genome microarrays for mRNA using 200ng of material and on Sureprint G3 Human V19 miRNA 8 x 60K microarrays for microRNA using 100ng of material. The Agilent Feature Extraction Software was used to extract raw data on pixel intensities.

- Pre-processing of the gene and microRNA expression

The normalization of the whole-genome gene expression dataset was performed using an R pipeline based on Bioconductor (github.com/BiGCAT-UM/arrayQC_Module), where all Agilent probes were normalized using the quantile method. Afterwards, genes with less than 30% flagged bad spots were

selected and imputation by k-nearest neighbors (k-NN; k-value 15) was performed. The microRNA pre-processing was performed using the AgiMicroRna pipeline with all Agilent probes as input [28].

Microarray analysis may be confounded by "batch effects" or systematic error introduced when samples are handled on different dates [29]. In order to allow for correction of such effects, identical biological samples were distributed across the different batches. By subtracting (since expression values are $\log_2$ transformed) the median expression of the quality control samples from one batch to the median expression of all quality control samples from the dataset, a ratio comprising information of that individual batch effect was calculated. This ratio was then added to the expression values of the study samples from the same batch.

Due to missing exposure data and bad quality of the RNA isolation (samples with 260/280 ratio below 2 and RNA Integrity Number below 6 were excluded) or microarray laboratory analysis, the final number of mRNA samples was reduced to 94 in the case of Oxford Street and 109 for TAPAS; for microRNA the numbers dropped to 90 for Oxford Street and 109 for TAPAS.

- Statistical analysis: LMM and MVN

The associations between exposure to air pollutants and gene or microRNA expression in the Oxford Street and TAPAS cohorts were analyzed independently. Residual technical noise introduced by microarray analysis was removed by applying linear mixed model analysis using the R package "lme4" ("lmer" function). The exposure covariate was used as variable of interest and the date of RNA isolation, microarray labelling and microarray hybridization as random variables. The parameters from the linear mixed model analysis such as residuals, intercept and estimated betas were used to calculate gene expression values without the random variables effect. A PCA of the de-noised gene expression data was performed in both cohorts, coloring the two locations differently (high exposure location in black and low exposure location in white).

After de-noising the expression values, we ran a multivariate normal model (MVN) using the R package "nlme" ("gls" function) in order to find associations between personal exposure levels and gene/microRNA expressions. Multivariate normal models consider all OMIC profiles at once as multivariate outcome Y. The variance-covariance matrix is modelled flexibly, so the covariance depends both on the location and physical activity status and it allows different coefficients for each experimental condition. Thefore, the gene/microRNA expression (Y) from each individual, in each scenario (Oxford Street/Hyde Park or Ronda/Barceloneta and rest/activity) followed a multivariate normal distribution with a mean vector (μ) and a covariance matrix (Σ) Y ~ MVN(μ, Σ). The fully unstructured variance-covariance matrix was modelled using the IDs from the different subjects as a grouping factor. The variable of interest was the measured air pollution data collected during the experiment (2 hours).

Levels of exposure to different pollutants were modelled in an independent MVN analysis. All models are adjusted for the confounders sex, age and BMI. In the case of Oxford Street, the health condition of the subjects (healthy, COPD or IHD) was used as an additional confounder. For TAPAS, since the experiment dates were not as distant in time as it was for Oxford Street, the order in which the four different scenarios

took place was used as an additional confounder. Afterwards, p-values were adjusted using the Bonferroni threshold at 5%.

Additionally, truncation at different levels (5%, 10% and 20%) of the gene expression distribution was performed as a stability analysis in order to assess the robustness of the results.

- Pathway analysis of the significant transcripts

The output from the MVN from the different cohorts for the different exposures was used as input for pathway analysis using the over-representation analysis tool from ConsensusPathDB [30] with the default settings (minimum overlap with input list: 2, p-value cutoff 0.01). A background list including the total list of genes from the pre-processed dataset was used.

- Integration microRNA/transcripts

The microRNAs significantly associated with the different exposures from the MVN model were used to identify target genes. The links between microRNAs and their corresponding targets were investigated using the miRTarBase database, which only includes relationships from curated literature and after experimental validation [31]. From the list of targeted genes, only those included in the list of significant hits from the MVN models (Bonferroni threshold at 5%) were selected for further analysis. Genes and microRNAs significantly expressed in association with the exposures (NO_2, NO, NOx, PM_{10}, $PM_{2.5}$, BC, UFPC, CO and CO_2) and the interactions between them were displayed using the Cytoscape tool [32]. Information on the direction of the expression in association with exposure (regression coefficient from the statistical analysis) and direction of the correlation between transcripts and microRNAs (R package "stats", "cor( )" function) was added to the figure.

- Combined analysis of the Oxford Street and TAPAS studies

Further data integration was performed by combining the transcriptomics samples from the two population studies. The exposure data was transformed to z-scores in order to normalize the potential differences in measurement techniques and/or equipment, allowing establishing gene expression effects of integrated air quality parameters. Z-scores were calculated as: Z-Score(exposure) = (X − μ)/ σ (X representing the value of the exposure, μ the mean of the population and σ the standard deviation of the population). The models were adjusted for the confounders cohort, sex, age, BMI, health condition of the subjects ("healthy", "COPD" or "IHD" for the subjects from the Oxford Street cohort and "healthy" for the subjects from the TAPAS cohort) and physical activity ("PA" for those samples corresponding to subjects cycling from the TAPAS cohort and "rest" from the remaining samples from the TAPAS cohort and the Oxford Street samples). Afterwards, p-values were adjusted using the Bonferroni threshold at 5%.

Results

- Exposures and PCA analysis of gene expression for the different locations

Table 1 shows the levels of exposure to ambient air pollutants for each location in each cohort. All the exposures levels are significantly different between the two locations within a cohort (paired t-test, p-

value from 4.08E-03 to 2.20E-16). The exposure levels from the TAPAS cohort (Table 1B) are higher than the Oxford Street exposure levels (Table 1A) and they present larger differences between low and high exposure locations. However, although the differences of exposure levels among locations are significant, the unsupervised PCA shows no clustering of the gene expression according to the different locations (Figure S1 from Supplementary Material). The correlations among exposures levels is high for chemically related compounds such PM_{10} and PM_2, and low correlation for others such PM_{10} and CO. (Table S2 from Supplementary Material).

Table 1. Exposure overview of the Oxford Street (A) and TAPAS (B) cohorts, comparing low and high exposure location.

A)

Compounds	Hyde Park	Oxford Street	Paired t-test
	Mean ± SD	Mean ± SD	P-value
NO_2 (ppb)	4.80 ± 3.60	8.58 ± 3.76	1.74E-07
PM_{10} ($\mu g/m^3$)	21.76 ± 14.42	33.41 ± 14.85	6.15E-04
$PM_{2.5}$ ($\mu g/m^3$)	10.15 ± 11.12	21.31 ± 13.17	1.68E-05
BC ($\mu g/m^3$)	1.71 ± 1.25	10.86 ± 3.15	2.20E-16
UFPC (particles/cm^3)	6787.49 ± 3326.23	25210.68 ± 7833.51	2.20E-16

B)

Compounds	Barceloneta	Ronda	Paired t-test
	Mean ± SD	Mean ± SD	P-value
NO (ppb)	77.83 ± 56.88	602.57 ± 197.68	2.20E-16
NOx (ppb)	101.75 ± 63.72	702.61 ± 207.43	2.20E-16
PM_{10} ($\mu g/m^3$)	63.77 ± 42.70	121.11 ± 37.91	7.11E-10
$PM_{2.5}$ ($\mu g/m^3$)	39.31 ± 13.98	81.86 ± 16.06	2.20E-16
BC ($\mu g/m^3$)	6.54 ± 4.73	23.67 ± 5.27	2.20E-16
UFPC (particles/cm^3)	46481.24 ± 21027.97	166667.30 ± 28759.33	2.20E-16
CO (ppm)	1.16 ± 1.84	3.11 ± 4.56	4.08E-03
CO_2 (ppm)	420.61 ± 42.09	498.72 ± 35.10	2.58E-15

- MVN hits from Oxford Street and TAPAS

Table 2 shows the number of Bonferroni significant hits associated with each exposure species using the MVN statistical model for the gene and microRNA expression data (Bonferroni threshold at 5%). The lists of transcripts and microRNAs significantly associated with PM_{10} and $PM_{2.5}$ present the largest numbers in Oxford Street, while in the TAPAS cohort this is the case for PM_{10}, $PM_{2.5}$, CO, and CO_2. The overlap for PM_{10}

and $PM_{2.5}$ in Oxford Street is quite substantial corresponding to around half of the hits (Figure S3 and S4 from Supplementary Material). The overlap across cohorts for significant transcripts associated with the exposures to NO-NO_x/NO_2, PM_{10} and $PM_{2.5}$ represents five genes significantly associated with particulate matter in common between subjects from the Oxford Street and TAPAS cohorts (Figure 1). These genes are XLOC_014512, MIS12, XLOC_014512, RAD51-AS1 and GOLGA8R.

Table 2. Number of hits significantly associated with the exposures of the Oxford Street (A) and TAPAS (B) cohorts.
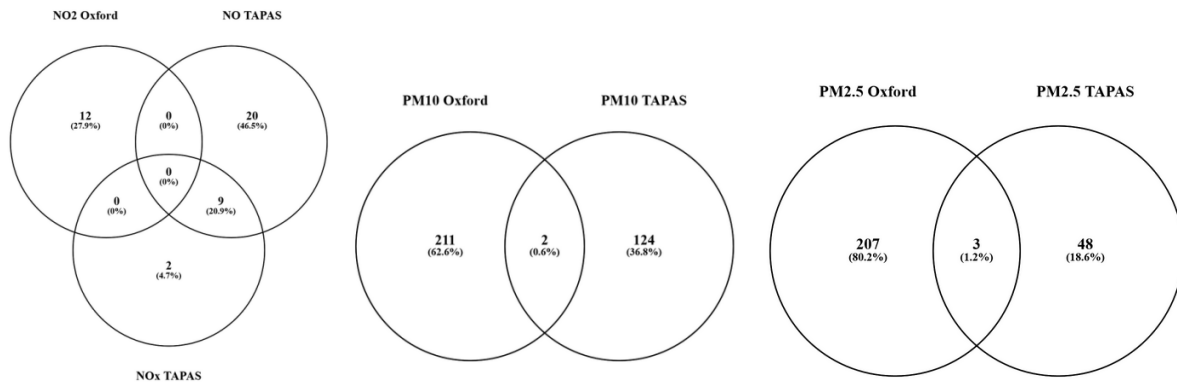
A) Oxford street London UK.

Exposure	Hits mRNA	Hits microRNA
NO_2	11	1
PM_{10}	229	9
PM_{25}	227	9
BC	0	1
UFPC	0	0

B) TAPAS Barcelona Spain

Exposure	Hits mRNA	Hits microRNA
NO	32	3
NO_x	13	1
PM_{10}	140	30
$PM_{2.5}$	59	10
BC	40	5
UFPC	34	3
CO	157	1
CO_2	65	20

Figure 1. Comparison between Oxford Street and TAPAS using the genes significantly associated with the different exposures in both cohorts.

The stability analysis performed by truncation of the gene expression distribution revealed moderate robust signals in the associations between gene expression and personal exposure levels after applying the capping parameters (5, 10, 20%) in the case of TAPAS (Table S3 B from Supplementary Material). However, this was not the case for Oxford Street, where most of the signals identified significantly associated with air pollution exposure are arising only from the extreme expression levels (Table S3 A from Supplementary Material). Subjects from the Oxford Street cohort that had extreme high/low expression levels are driving the observed results, probably because these hits are found significantly associated with low exposure levels in relatively small sample size after short-term exposure. The observed associations are subtle even if they remain significant after Bonferroni correction. In the TAPAS cohort the signals are more stable, which can be indicative of stronger associations due to uncontrolled burden such less noise in the expression or exposure measured values, potential confounding effects such dietary habits, physical activity, etc.

- Pathway analysis

Over-representation analyses of transcripts significantly associated with air pollution exposure by using ConsensusPathDB return lists of relevant deregulated signaling pathways (Table S4 from Supplementary Material). Cancer-related pathways such as regulation of TP53 and TGF-beta receptor signaling pathways are identified in the Oxford Street analysis. In the TAPAS study, other cancer-related pathways are found, such as p75 in association with NO exposure, WNT-beta-catenin in association with PM_{10} exposure and nonsense mediated decay pathways in association with CO exposure. Deregulated pathways derived from transcripts associated with NO and NO_x exposures in TAPAS demonstrate a high level of significance (range from 0.009 to 0.027) while for the other exposures the levels of significance are moderate (range from 0.093 to 0.220).

- Transcript/MicroRNA integration

The search of the miRTarBase database identified gene expressions that are targeted by significant microRNAs. In the Oxford Street study mRNA/microRNA integration analysis, the expression of 6 genes significantly associated with PM_{10} appeared to be targeted by 4 microRNAs also significantly expressed in association with PM_{10} exposure levels. In the TAPAS analysis this was the case for 18 significant transcripts and 8 significant microRNAs. Figure 2 presents the network of genes and microRNAs significantly associated with PM_{10} for both Oxford Street (A) and TAPAS (B).

Figure 2. Cytoscape figure of hits significantly associated with PM_{10} from Oxford Street (A) and TAPAS (B).

A)
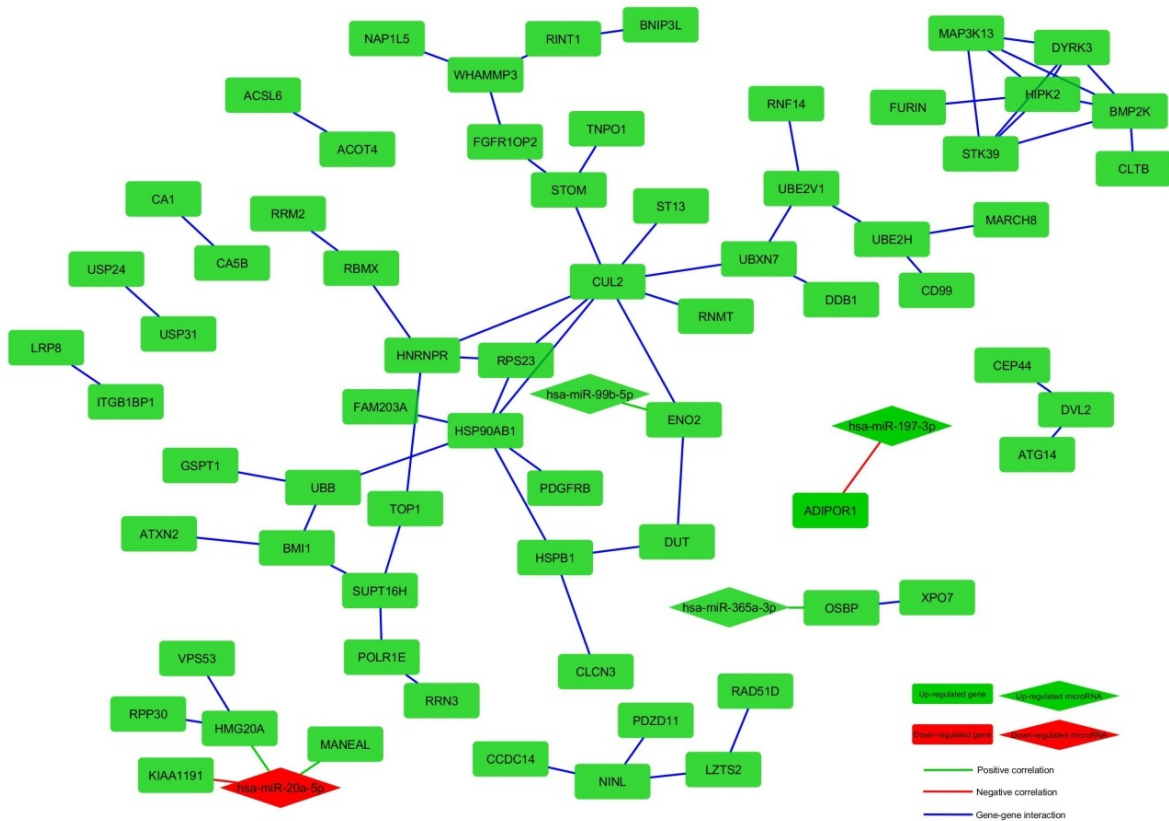
B)

Similarly, the integrative analysis of omics responses in relation to $PM_{2.5}$ exposure levels resulted in 5 gene expressions significantly associated with $PM_{2.5}$ exposure and targeted by 2 significant microRNAs in the Oxford Street analysis, and 3 significant transcripts targeted by 3 significantly expressed microRNAs in the TAPAS cohort (Figure S5 from Supplementary Material).

Finally, in the TAPAS cohort five genes significantly expressed in association with CO_2 are found that are targeted by three microRNAs also significantly expressed in association with CO_2 (Figure S6 from Supplementary Material). Integrative mRNA/microRNA analysis of the remaining exposures (NO_2, NO, NOx, BC, UFPC and CO) did not yield further associations between significantly expressed microRNAs and their target genes.

A summary of the functions of these targeted genes in association with the different exposures is presented in Tables S5, S6 and S7 in the Supplementary Material. Predominantly, cancer-related and neurological disease-related processes have been identified.

-   Combined analysis of the Oxford Street and TAPAS studies

The venn diagrams from Figure S7 in the Supplementary Material shows a high number of hits associated from the exposures from the combined analysis. However, the overlap of the genes significantly associated with the individual exposures from the combined analysis and the individual analyses from the Oxford Street and TAPAS cohorts is very limited. Overlap at the pathway level and not at the gene level is also very limited.

Discussion

In this study we assessed the transcriptome and microRNA responses from subjects exposed to different levels of air pollutants in order to identify potential biological mechanisms associated with such exposures. We identified transcripts and microRNAs for which the expressions were significantly

associated with exposure to NO_2, NO, NOx, PM_{10}, $PM_{2.5}$, BC, UFPC, CO and CO_2. Most of these biomarkers were found to be compound-specific.

We identified gene and microRNA expressions significantly associated with the different exposures for the two cohorts using the multivariate normal models (Table 2). We found little overlap between the two cohorts which might be a consequence of the substantial differences in exposure levels found in both populations. The overlapping genes are XLOC_014512 (non-annotated), MIS12 (involved in signaling by GPCR and cell cycle), XLOC_014512 (non-annotated), RAD51-AS1 (promotes proliferation, migration and invasion in ovarian cancer, and plays an important role in tumor progression) and GOLGA8R (non-annotated). Remarkably, the subjects from the low exposed scenario in the TAPAS cohort show higher exposure values than the subjects from the high exposed scenario in the Oxford Street cohort, which makes the interpretation of the comparison among the two cohorts more complex. Furthermore, exposure to dietary factors and life-style may be different between the two countries (dietary exposure, life style, etc.). The PM_{10}, $PM_{2.5}$, CO and CO_2 exposures appear to be the most significant ones in terms of number of hits (transcripts and microRNA) responding to exposure levels. The genes identified are involved in cancer-related pathways such as TP53, TGF-beta receptor signaling pathways, p75, WNT-beta-catenin and nonsense mediated decay pathways (Table S4 from Supplementary Material).

The combined analysis of the Oxford Street and TAPAS studies was performed in view of the similar study designs of both cohorts, the possibility of using Z-scores for the air quality parameters (correcting for inter-study differences in measurements) and the same outcome parameter. The high number of hits observed in the merged analysis can be the result of the increase in statistical power while the lack of consistency in the biological processes can be attributed to potential cohort-specific responses that disappear when you perform a cross-cohort analysis and correct for 'cohort'. However, these results should be interpreted with caution in view of the potential noise that could be generated by merging the two cohorts.

MicroRNAs are known to play an important role in the regulation of oncogenes and tumor suppressors [33]. From the total list of genes and microRNAs significantly associated with air pollutants, we therefore focused on those significant microRNAs that target significant genes.

The literature shows that hsa-mir-20a-5p is down-regulated in chronic myeloid leukemia [34] and breast cancer [35]. On the other hand, it is found up-regulated in lung cancer [36,37], hepatocellular carcinoma [38], B-cell lymphoma [39], medulloblastoma [40], pulmonary hypertension [41] and Alzheimer's [42]. In our study we found this miRNA down-regulated in association with PM_{10} exposure in the Oxford Street cohort whereas the rest of exposures did not show an effect. The significant genes targeted by hsa-mir-20a-5p are HMG20A, MANEAL and KIAA1191, having the two first ones a positive correlation with the microRNA. All the targeted genes were found up-regulated in association with PM_{10} exposure. HMG20A plays a role in neural differentiation but HMG20A-knockdown cells have shown a high proportion of malignant cancer progression-related genes being deregulated [43].

Hsa-miR-197-3p is found up-regulated in follicular thyroid carcinoma [44] and lung cancer [45]. It is found down-regulated in narcolepsy patients compared to controls [46]. We also found it down-regulated in association with PM_{10} and $PM_{2.5}$ exposures only in the Oxford Street cohort. The targeted gene is ADIPOR1,

which is up-regulated in association with PM_{10} exposure may be indicative of a favorable prognosis in non-small cell lung carcinoma [47]. RPL13A, STRN and AK3 are targets found positively associated with $PM_{2.5}$ exposure and negatively correlated to hsa-mir-197-3p. AK3 is reported as apoptosis inducer in colon cancer cells [48].

Hsa-miR-29a-3p is found down-regulated in Alzheimer's disease [49] and lung cancer [50, 51]. However, it is found up-regulated in association with PM_{10}, $PM_{2.5}$ and CO_2 exposures only in the TAPAS cohort. The gene SYNCRIP is targeted by hsa-miR-29a-3p and it is down-regulated and positively correlated with the microRNA. It is reported in literature a down-regulation of SYNCRIP in macrophages in response to apoptotic cells [52].

Hsa-miR-15a-5p is down-regulated in chronic lymphocytic leukemia [53-55], pituitary adenoma [56], prostate cancer [57], non-small cell lung cancer [58], ovarian cancer [59] and chronic lymphocytic leukemia [60] and up-regulated in acute promyelocytic leukemia [61]. In the TAPAS cohort it is found up-regulated only in association with PM_{10} exposure. MYBL1 and CENPJ are the targeted genes and are up-regulated and positively correlated with hsa-miR-15a-5p. Combined MYB and MYBL1 expression correlates with poor outcome in adenoid cystic carcinoma patients [62]. Also, expression levels of MYBL1 and MYBL2 are increased in late breast cancer development and are predictors of poor clinical outcome [63]. CENPJ is involved in brain growth [64] and CENPJ deficiency has consequences in the formation and function of the cerebral cortex in mice [65].

Hsa-miR-16-5p is found down-regulated in gastric cancer [66], chronic lymphocytic leukemia [67], non-small cell lung cancer [58], prostate cancer [68] and ovarian cancer [59]. On the other hand, it is found up-regulated in hepatocellular carcinoma [69]. In our study, it is found up-regulated in association with PM_{10} and $PM_{2.5}$ exposures in the TAPAS cohort. CENPJ, NOL11, CDC20, MYO19, MIS12 and VKORC1 are gene targets of hsa-miR-16-5p. CENPJ and NOL11 are up-regulated in association with PM_{10} exposure in TAPAS and positively correlated to the microRNA. MYO19, MIS12 and VKORC1 are also up-regulated but negatively correlated to the microRNA. The last target gene, CDC20, is down-regulated in association with the exposure and negatively correlated with the microRNA. Knockdown of NOL11 is linked to increased apoptosis and anomalous development of the craniofacial cartilages [70]. CDC20 is an oncogene that promotes the development and progression of human cancers [71].

Hsa-miR-92a-3p is found up-regulated in hepatocellular carcinoma [38] and medulloblastoma [40]. However, it is found down-regulated in association with PM_{10}, $PM_{2.5}$ and CO_2 exposures in the TAPAS cohort. Hsa-miR-92a-3p targets the genes CDC20, CCAR1, CDK11A, POGZ, THBS1, HIST1H1E, MYO19, DDRGK1, CRBN and HSPH1. The last one, MYO19, is found up-regulated in association with PM_{10} exposure in TAPAS and positively correlated to the microRNA. CCAR1, CDK11A and POGZ are found up-regulated in association with PM_{10} exposure in TAPAS but negatively correlated to the microRNA. HIST1H1E is down-regulated in association with the exposure and positively correlated with the microRNA. CDC20 and THBS1 are down-regulated in association with the exposure and negatively correlated with the microRNA. CCAR1 levels correlate with gastric cancer incidence [72], hepatocellular carcinoma poor prognosis [73] and growth of colon cancer cell lines [74]. CDK11A is upregulated in human ovarian cancer [75] and breast cancer tissues and cell lines [76]. POGZ disruption is linked to intellectual disability [77]. Mutations in

HIST1H1E are associated with adenoid cystic carcinoma [78]. Finally, THBS1 has been reported to promote invasion of oral squamous cell carcinoma [79], melanoma [80] and ovarian cancer [81]. Alternatively, high levels of THBS1 are related to better prognosis in non-small cell lung carcinoma [82]. DDRGK1 is up-regulated in association with $PM_{2.5}$ exposure in TAPAS and positively correlated to the microRNA. Depletion of DDRGK1 is found in literature to induce apoptosis [83, 84]. CRBN is also up-regulated in association with $PM_{2.5}$ exposure in TAPAS but negatively correlated to the microRNA. High levels of CRBN are reported as low risk for multiple myeloma [85]. HSPH1 is down-regulated in association with $PM_{2.5}$ exposure in TAPAS and positively correlated to the microRNA. HSPH1 promotes B-cell non-Hodgkin lymphomas [86].

The fact that the expressions of microRNAs and targeted genes mentioned in the paragraphs above are found significantly associated with short-term exposure to moderate and low air pollution levels in two cohorts with relatively small size populations suggests their potential relevance in the research of omics biomarkers responding to exposures. The diversity in the direction of gene and microRNA expression towards disease indicates complex biological interactions between air pollution exposure and omics signals. Some events reflect toxic risks while others may be indicative for adaptation and/or damage repair. The interpretation of changes in gene and microRNA expression in terms of disease risk remains a challenge since the identified intermediate biomarkers may originate from different tissues and because the signals may relate to processes that have not previously been linked to the disease or earlier stages during disease development. Furthermore, different diseases may show different directionalities for the same biomarkers. For instance, hsa-miR-29a-3p is reported down-regulated in Alzheimer's disease [49] whereas it is up-regulated in patients with Huntington's disease [87]. We found hsa-miR-15a-5p up-regulated in our analysis, opposite as in a wide range of cancers, while high levels of this microRNA are associated with schizophrenia [88] and acute ischemic stroke [89]. The different directions in the regulation of genes/microRNAs in association with exposure levels might be a consequence of the wide nature of the disorders linked to air pollutants, such as cancer, cardiovascular illnesses and neurological-related diseases.

Conclusion

We identified in this study promising compound-specific biomarkers (gene transcripts and microRNAs) associated with short term exposure to air pollution, despite the low levels of the short-term exposure and the relatively small-sized cohorts. We built networks for each compound integrating gene and microRNA interactions together with information on the direction of the association with the exposure that provides insights on the biological mechanistic processes. Most of the significant signals point towards increased cancer risks in association with particulate matter exposure. In view of the complexity of diseases derived from environmental exposures, studies with a more dedicated design are required to assign biological significance to particular up-/down-regulation of genes /microRNAs in association with air pollutants.

Acknowledgements

References

1. Polosa, R., S. Salvi, and G.U. Di Maria, *Allergic susceptibility associated with diesel exhaust particle exposure: Clear as mud*. Archives of Environmental Health, 2002. **57**(3): p. 188-193.

2. Manchester-Neesvig, J.B., J.J. Schauer, and G.R. Cass, *The distribution of particle-phase organic compounds in the atmosphere and their use for source apportionment during the southern California children's health study*. Journal of the Air & Waste Management Association, 2003. **53**(9): p. 1065-1079.

3. Aoki, Y., *Evaluation of in vivo mutagenesis for assessing the health risk of air pollutants*. Genes and environment : the official journal of the Japanese Environmental Mutagen Society, 2017. **39**: p. 16.

4. IARC/WHO, *Diesel and Gasoline Engine Exhausts and Some Nitroarenes, IARC Monographs on the Evaluation of Carcinogenic Risks to Human*. Lyon, 2014. **105**.

5. Yang, W.S., et al., *An evidence-based assessment for the association between long-term exposure to outdoor air pollution and the risk of lung cancer*. European Journal of Cancer Prevention, 2016. **25**(3): p. 163-172.

6. Pope, C.A., *Epidemiology of Lung Cancer Associated with Outdoor Air Pollution*. Environmental and Molecular Mutagenesis, 2015. **56**: p. S41-S41.

7. Gan, W.Q., et al., *Long-term exposure to traffic-related air pollution and the risk of coronary heart disease hospitalization and mortality*. Environ Health Perspect, 2011. **119**(4): p. 501-7.

8. Lee, B.J., B. Kim, and K. Lee, *Air pollution exposure and cardiovascular disease*. Toxicological research, 2014. **30**(2): p. 71-5.

9. Koulova, A. and W.H. Frishman, *Air pollution exposure as a risk factor for cardiovascular disease morbidity and mortality*. Cardiology in review, 2014. **22**(1): p. 30-6.

10. Wellenius, G.A., J. Schwartz, and M.A. Mittleman, *Air pollution and hospital admissions for ischemic and hemorrhagic stroke among medicare beneficiaries*. Stroke, 2005. **36**(12): p. 2549-2553.

11. Zanobetti, A. and J. Schwartz, *The effect of particulate air pollution on emergency admissions for myocardial infarction: A multicity case-crossover analysis*. Environ Health Perspect, 2005. **113**(8): p. 978-982.

12. Vimercati, L., *Traffic related air pollution and respiratory morbidity*. Lung India : official organ of Indian Chest Society, 2011. **28**(4): p. 238.

13. Laumbach, R.J. and H.M. Kipen, *Respiratory health effects of air pollution: update on biomass smoke and traffic pollution*. The Journal of allergy and clinical immunology, 2012. **129**(1): p. 3-11; quiz 12-3.

14. Xu, X., S.U. Ha, and R. Basnet, *A Review of Epidemiological Research on Adverse Neurological Effects of Exposure to Ambient Air Pollution*. Frontiers in public health, 2016. **4**: p. 157.

15. Woodward, N., C.E. Finch, and T.E. Morgan, *Traffic-related air pollution and brain development*. AIMS environmental science, 2015. **2**(2): p. 353-373.

16. Genc, S., et al., *The adverse effects of air pollution on the nervous system*. Journal of toxicology, 2012. **2012**: p. 782462.

17. Cohen, A.J., et al., *Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: an analysis of data from the Global Burden of Diseases Study 2015*. Lancet, 2017.

18. Jayeun, K., et al., *Hourly differences in air pollution on the risk of asthma exacerbation*. Environmental Pollution , 2015. **203**: p. 15-21.

19.   Hualiang, L., et al., *Particle size and chemical constituents of ambient particulate pollution associated with cardiovascular mortality in Guangzhou, China.* Environmental Pollution, 2016. **208**: p. 758-766.

20.   Hualiang, L., et al., *Particle size and chemical constituents of ambient particulate pollution associated with cardiovascular mortality in Guangzhou, China.* Environmental Pollution, 2016. **208**: p. 758-766.

21.   Buonanno, G., et al., *Personal exposure to ultrafine particles: The influence of time-activity patterns.* Science of the Total Environment, 2014. **468-469**. p. 903-907.

22.   Mostafavi, N., et al., *Associations Between Genome-wide Gene Expression and Ambient Nitrogen Oxides.* Epidemiology, 2017. **28**(3): p. 320-328.

23.   Chu, J.H., et al., *Gene expression network analyses in response to air pollution exposures in the trucking industry.* Environmental Health, 2016. **15**.

24.   Vrijens, K., et al., *Sex-Specific Associations between Particulate Matter Exposure and Gene Expression in Independent Discovery and Validation Cohorts of Middle-Aged Men and Women.* Environ Health Perspect, 2017. **125**(4): p. 660-669.

25.   Jardim, M.J., *microRNAs: Implications for air pollution research.* Mutation Research-Fundamental and Molecular Mechanisms of Mutagenesis, 2011. **717**(1-2): p. 38-45.

26.   McCreanor, J., et al., *Respiratory effects of exposure to diesel traffic in persons with asthma.* New England Journal of Medicine, 2007. **357**(23): p. 2348-2358.

27.   Matt, F., et al., *Acute respiratory response to traffic-related air pollution during physical activity performance.* Environ Int, 2016. **97**: p. 45-55.

28.   Lopez-Romero, P., *Pre-processing and differential expression analysis of Agilent microRNA arrays using the AgiMicroRna Bioconductor library.* BMC Genomics, 2011. **12**: p. 64.

29.   Leek, J.T., et al., *Tackling the widespread and critical impact of batch effects in high-throughput data.* Nature reviews. Genetics, 2010. **11**(10): p. 733-9.

30.   Kamburov, A., et al., *The ConsensusPathDB interaction database: 2013 update.* Nucleic Acids Res, 2013. **41**(Database issue): p. D793-800.

31.   Chou, C.H., et al., *miRTarBase 2016: updates to the experimentally validated miRNA-target interactions database.* Nucleic Acids Res, 2016. **44**(D1): p. D239-47.

32.   Cline, M.S., et al., *Integration of biological networks and gene expression data using Cytoscape.* Nature Protocols, 2007. **2**(10): p. 2366-2382.

33.   Zhang, B.H., et al., *microRNAs as oncogenes and tumor suppressors.* Developmental Biology, 2007. **302**(1): p. 1-12.

34.   Venturini, L., et al., *Expression of the miR-17-92 polycistron in chronic myeloid leukemia (CML) CD34(+) cells.* Blood, 2007. **109**(10): p. 4399-4405.

35.   Yu, Z.R., et al., *A cyclin D1/microRNA 17/20 regulatory feedback loop in control of breast cancer cell proliferation.* Journal of Cell Biology, 2008. **182**(3): p. 509-517.

36.   Hayashita, Y., et al., *A polycistronic microRNA cluster, miR-17-92, is overexpressed in human lung cancers and enhances cell proliferation.* Cancer research, 2005. **65**(21): p. 9628-9632.

37.   Matsubara, H., et al., *Apoptosis induction by antisense oligonucleotides against miR-17-5p and miR-20a in lung cancers overexpressing miR-17-92.* Oncogene, 2007. **26**(41): p. 6099-6105.

38.   Connolly, E., et al., *Elevated expression of the miR-17-92 polycistron and miR-21 in hepadnavirus-associated hepatocellular carcinoma contributes to the malignant phenotype.* American Journal of Pathology, 2008. **173**(3): p. 856-864.

39.   Tagawa, H., et al., *MicroRNA-17-92 downregulates expression of distinct targets in different B-cell lymphoma subtypes.* Cancer research, 2009. **69**.

40. Northcott, P.A., et al., *The miR-17/92 Polycistron Is Up-regulated in Sonic Hedgehog-Driven Medulloblastomas and Induced by N-myc in Sonic Hedgehog-Treated Cerebellar Neural Precursors.* Cancer research, 2009. **69**(8): p. 3249-3255.

41. Brock, M., et al., *Interleukin-6 Modulates the Expression of the Bone Morphogenic Protein Receptor Type II Through a Novel STAT3-microRNA Cluster 17/92 Pathway.* Circulation research, 2009. **104**(10): p. 1184-U139.

42. Satoh, J., Y. Kino, and S. Niida, *MicroRNA-Seq Data Analysis Pipeline to Identify Blood Biomarkers for Alzheimer's Disease from Public Data.* Biomarker Insights, 2015. **10**: p. 21-31.

43. Rivero, S., et al., *HMG20A is required for SNAI1-mediated epithelial to mesenchymal transition.* Oncogene, 2015. **34**(41): p. 5264-5276.

44. Weber, F., et al., *A limited set of human MicroRNA is deregulated in follicular thyroid carcinoma.* Journal of Clinical Endocrinology & Metabolism, 2006. **91**(9): p. 3584-3591.

45. Du, L.Q., et al., *miR-93, miR-98, and miR-197 Regulate Expression of Tumor Suppressor Gene FUS1.* Molecular Cancer Research, 2009. **7**(8): p. 1234-1243.

46. Liu, Z., et al., *Reproducibility of quantitative real-time PCR assay in microRNA expression profiling and comparison with microarray analysis in narcolepsy.* SpringerPlus, 2015. **4**: p. 812.

47. Abdul-Ghafar, J., et al., *Expression of Adiponectin Receptor 1 Is Indicative of Favorable Prognosis in Non-small Cell Lung Carcinoma.* Tohoku Journal of Experimental Medicine, 2013. **229**(2): p. 153-162.

48. Chopra, A.S., et al., *Identification of novel compounds that enhance colon cancer cell sensitivity to inflammatory apoptotic ligands.* Cancer biology & therapy, 2013. **14**(5): p. 436-49.

49. Hebert, S.S., et al., *Loss of microRNA cluster miR-29a/b-1 in sporadic Alzheimer's disease correlates with increased BACE1/beta-secretase expression.* Proceedings of the National Academy of Sciences of the United States of America, 2008. **105**(17): p. 6415-6420.

50. Fabbri, M., et al., *MicroRNA-29 family reverts aberrant methylation in lung cancer by targeting DNA methyltransferases 3A and 3B.* Proceedings of the National Academy of Sciences of the United States of America, 2007. **104**(40): p. 15805-15810.

51. Muniyappa, M.K., et al., *MiRNA-29a regulates the expression of numerous proteins and reduces the invasiveness and proliferation of human carcinoma cell lines.* European Journal of Cancer, 2009. **45**(17): p. 3104-3118.

52. Kuchler, L., et al., *SYNCRIP-Dependent Nox2 mRNA Destabilization Impairs ROS Formation in M2-Polarized Macrophages.* Antioxidants & Redox Signaling, 2014. **21**(18): p. 2483-2497.

53. Cimmino, A., et al., *miR-15 and miR-16 induce apoptosis by targeting BCL2.* Proceedings of the National Academy of Sciences of the United States of America, 2005. **102**(39): p. 13944-13949.

54. Calin, G.A., Y. Pekarsky, and C.M. Croce, *The role of microRNA and other non-coding RNA in the pathogenesis of chronic lymphocytic leukemia.* Best Practice & Research Clinical Haematology, 2007. **20**(3): p. 425-437.

55. Calin, G.A., et al., *MiR-15a and miR-16-1 cluster functions in human leukemia.* Proceedings of the National Academy of Sciences of the United States of America, 2008. **105**(13): p. 5166-5171.

56. Bottoni, A., et al., *miR-15a and miR-16-1 down-regulation in pituitary adenomas.* Journal of Cellular Physiology, 2005. **204**(1): p. 280-285.

57. Bonci, D., et al., *The miR-15a-miR-16-1 cluster controls prostate cancer by targeting multiple oncogenic activities.* Nature Medicine, 2008. **14**(11): p. 1271-1277.

58. Bandi, N., et al., *miR-15a and miR-16 Are Implicated in Cell Cycle Regulation in a Rb-Dependent Manner and Are Frequently Deleted or Down-regulated in Non-Small Cell Lung Cancer.* Cancer research, 2009. **69**(13): p. 5553-5559.

59. Bhattacharya, R., et al., *MiR-15a and MiR-16 Control Bmi-1 Expression in Ovarian Cancer.* Cancer research, 2009. **69**(23): p. 9090-9095.

60. Calin, G.A., et al., *MicroRNA profiling reveals distinct signatures in B cell chronic lymphocytic leukemias*. Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(32): p. 11755-11760.

61. Garzon, R., et al., *MicroRNA gene expression during retinoic acid-induced differentiation of human acute promyelocytic leukemia*. Oncogene, 2007. **26**(28): p. 4148-4157.

62. Brayer, K.J., et al., *Recurrent Fusions in MYB and MYBL1 Define a Common, Transcription Factor-Driven Oncogenic Pathway in Salivary Gland Adenoid Cystic Carcinoma*. Cancer Discovery, 2016. **6**(2): p. 176-187.

63. Liu, L.Y.D., et al., *A Supervised Network Analysis on Gene Expression Profiles of Breast Tumors Predicts a 41-Gene Prognostic Signature of the Transcription Factor MYB across Molecular Subtypes*. Computational and Mathematical Methods in Medicine, 2014.

64. Bond, J., et al., *A centrosomal mechanism involving CDK5RAP2 and CENPJ controls brain size*. Nature genetics, 2005. **37**(4): p. 353-355.

65. Garcez, P.P., et al., *Cenpj/CPAP regulates progenitor divisions and neuronal migration in the cerebral cortex downstream of Ascl1*. Nature communications, 2015. **6**.

66. Xia, L., et al., *miR-15b and miR-16 modulate multidrug resistance by targeting BCL2 in human gastric cancer cells*. International journal of cancer, 2008. **123**(2): p. 372-379.

67. Calin, G.A., et al., *Frequent deletions and down-regulation of micro-RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia*. Proceedings of the National Academy of Sciences of the United States of America, 2002. **99**(24): p. 15524-15529.

68. Takeshita, F., et al., *Systemic Delivery of Synthetic MicroRNA-16 Inhibits the Growth of Metastatic Prostate Tumors via Downregulation of Multiple Cell-cycle Genes*. Molecular Therapy, 2010. **18**(1): p. 181-187.

69. Tsang, W.P. and T.T. Kwok, *Epigallocatechin gallate up-regulation of miR-16 and induction of apoptosis in human cancer cells*. Journal of Nutritional Biochemistry, 2010. **21**(2): p. 140-146.

70. Griffin, J.N., et al., *The Ribosome Biogenesis Factor Nol11 Is Required for Optimal rDNA Transcription and Craniofacial Development in Xenopus*. Plos Genetics, 2015. **11**(3).

71. Wang, Z.W., et al., *Cdc20: A Potential Novel Therapeutic Target for Cancer Treatment*. Current pharmaceutical design, 2013. **19**(18): p. 3210-3214.

72. Chang, T.S., et al., *Inhibition of CCAR1, a Coactivator of beta-Catenin, Suppresses the Proliferation and Migration of Gastric Cancer Cells*. International journal of molecular sciences, 2017. **18**(2).

73. Ha, S.Y., et al., *The Overexpression of CCAR1 in Hepatocellular Carcinoma Associates with Poor Prognosis*. Cancer research and treatment : official journal of Korean Cancer Association, 2016. **48**(3): p. 1065-73.

74. Ou, C.Y., et al., *Requirement of cell cycle and apoptosis regulator 1 for target gene activation by Wnt and beta-catenin and for anchorage-independent growth of human colon carcinoma cells*. The Journal of biological chemistry, 2009. **284**(31): p. 20629-37.

75. Liu, X.Z., et al., *Cyclin-Dependent Kinase 11 (CDK11) Is Required for Ovarian Cancer Cell Growth In Vitro and In Vivo, and Its Inhibition Causes Apoptosis and Sensitizes Cells to Paclitaxel*. Molecular Cancer Therapeutics, 2016. **15**(7): p. 1691-1701.

76. Zhou, Y.B., et al., *Cyclin-dependent kinase 11(p110) (CDK11(p110)) is crucial for human breast cancer cell proliferation and growth*. Scientific reports, 2015. **5**.

77. Stessman, H.A.F., et al., *Disruption of POGZ Is Associated with Intellectual Disability and Autism Spectrum Disorders*. American Journal of Human Genetics, 2016. **98**(3): p. 541-552.

78. Ho, A.S., et al., *The mutational landscape of adenoid cystic carcinoma*. Nature genetics, 2013. **45**(7): p. 791-+.

79. Pal, S.K., et al., *THBS1 is induced by TGFB1 in the cancer stroma and promotes invasion of oral squamous cell carcinoma*. Journal of Oral Pathology & Medicine, 2016. **45**(10): p. 730-739.

80. Jayachandran, A., et al., *Thrombospondin 1 promotes an aggressive phenotype through epithelial-to-mesenchymal transition in human melanoma*. Oncotarget, 2014. **5**(14): p. 5782-5797.

81. Alvarez Secord A. et al., *TP53 status is associated with thrombospondin1 expression in vitro*. Front. Oncol, 2013.

82. Weng, T.Y., et al., *Differential Expression Pattern of THBS1 and THBS2 in Lung Cancer: Clinical Outcome and a Systematic-Analysis of Microarray Databases*. PloS one, 2016. **11**(8).

83. Liu, J., et al., *A critical role of DDRGK1 in endoplasmic reticulum homoeostasis via regulation of IRE1alpha stability*. Nature communications, 2017. **8**: p. 14186.

84. Xi, P., et al., *DDRGK1 regulates NF-kappaB activity by modulating IkappaBalpha stability*. PloS one, 2013. **8**(5): p. e64231.

85. Schuster, S.R., et al., *The clinical significance of cereblon expression in multiple myeloma*. Leukemia research, 2014. **38**(1): p. 23-8.

86. Zappasodi, R., et al., *HSPH1 inhibition downregulates Bcl-6 and c-Myc and hampers the growth of human aggressive B-cell non-Hodgkin lymphoma*. Blood, 2015. **125**(11): p. 1768-71.

87. Johnson, R., et al., *A microRNA-based gene dysregulation pathway in Huntington's disease*. Neurobiology of disease, 2008. **29**(3): p. 438-45.

88. Beveridge, N.J. and M.J. Cairns, *MicroRNA dysregulation in schizophrenia*. Neurobiology of disease, 2012. **46**(2): p. 263-71.

89. Wu, J., K. Du, and X. Lu, *Elevated expressions of serum miR-15a, miR-16, and miR-17-5p are associated with acute ischemic stroke*. International journal of clinical and experimental medicine, 2015. **8**(11): p. 21071-9.
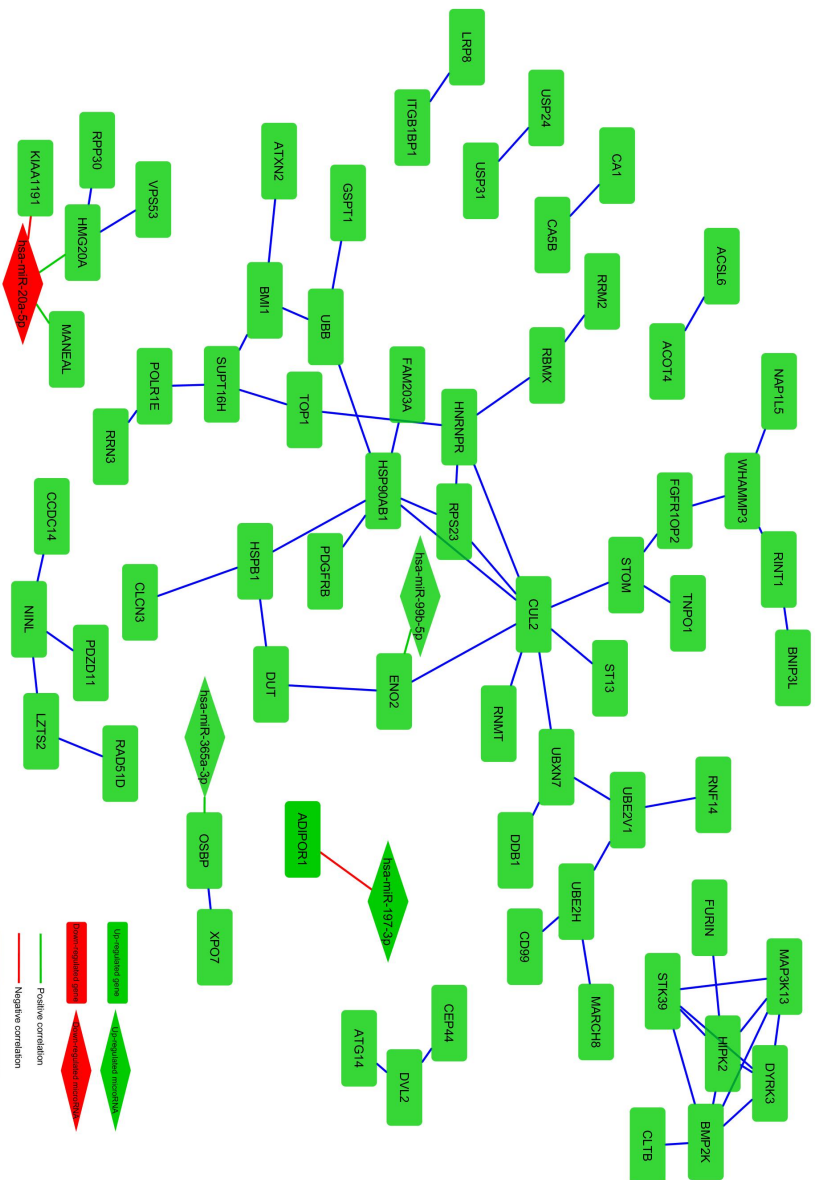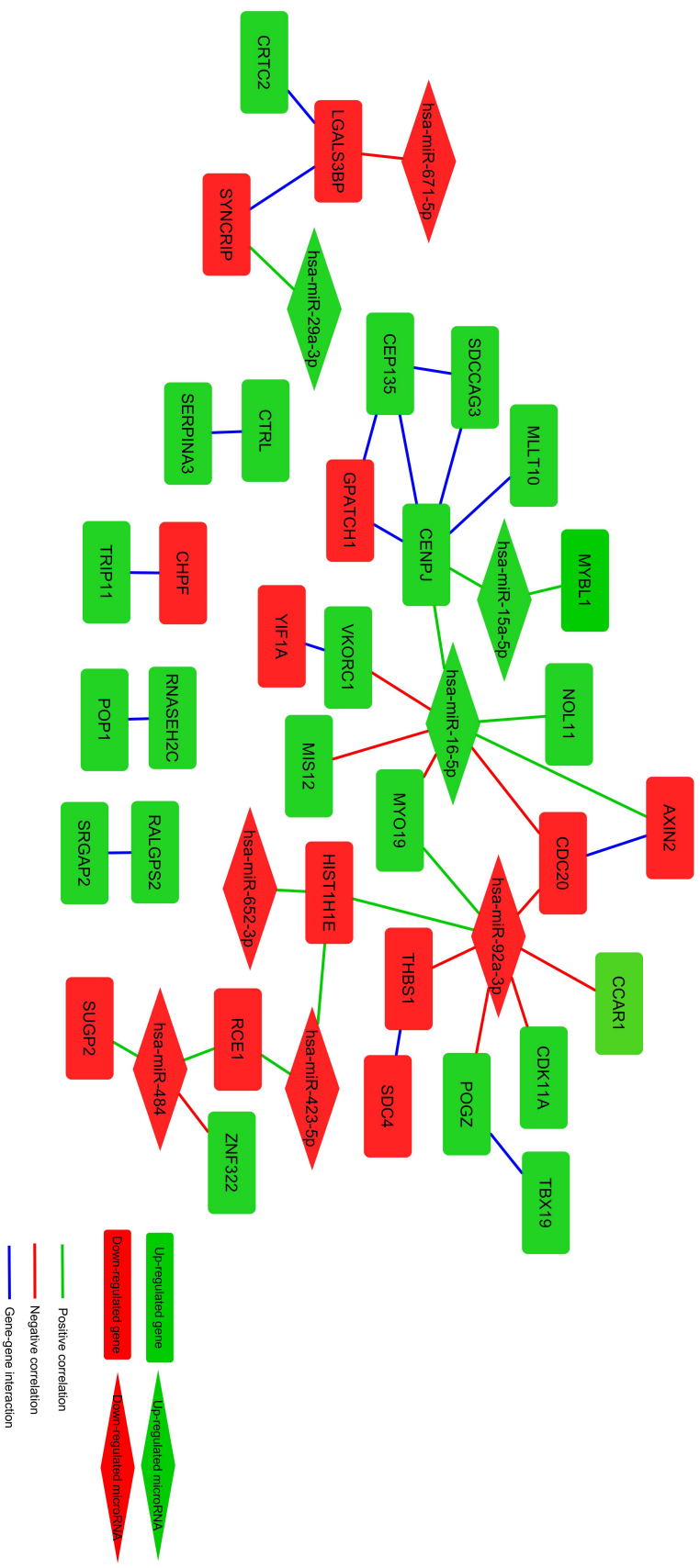
SUPPLEMENTARY MATERIAL

Table S1. Population characteristics from the Oxford Street and TAPAS cohorts

Oxford Street				
Variable	N		Mean	Std Dev
BMI	59		24.70	4.54
Age	59		65.43	6.43
Sex :	N	%	Cumulative Frequency	Cumulative Percent
Male	43	78	43	78
Female	16	12	59	100
TAPAS				
Variable	N		Mean	Std Dev
BMI	30		24.34	3.82
Age	30		36.18	11.64
Sex :	N	%	Cumulative Frequency	Cumulative Percent
Male	15	50	15	50
Female	15	50	30	100

Figure S1. PCA plot of gene expression from the Oxford Street (A) and TAPAS (B) cohorts
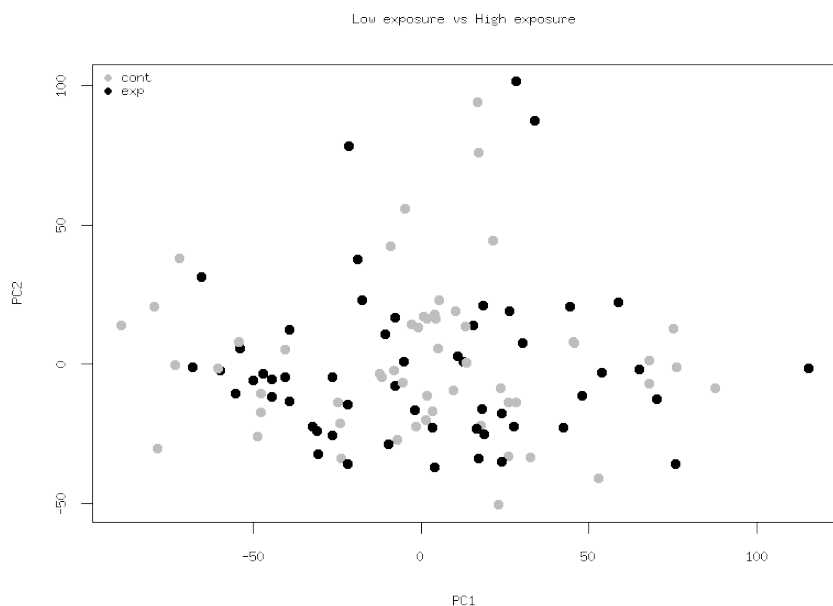
A)

B)

Low exposure vs High exposure
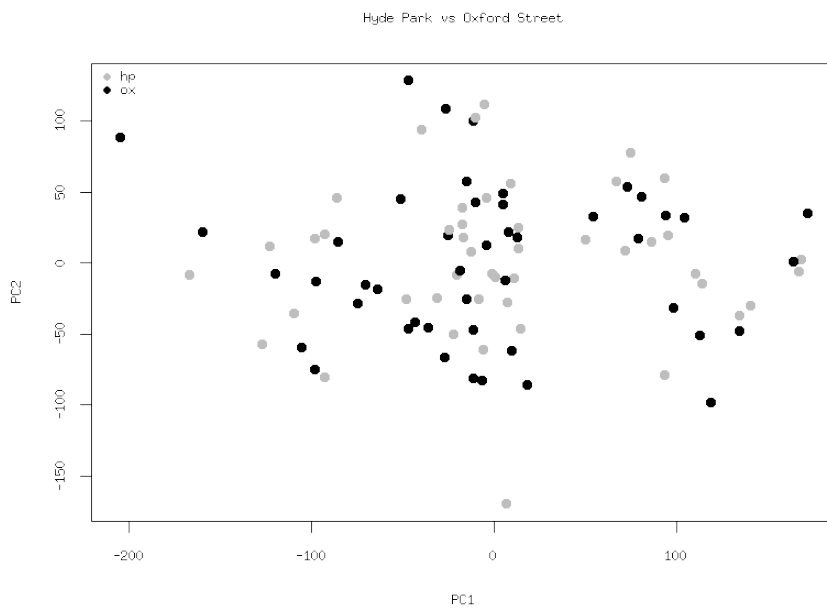
Table S2. Exposure correlations from the Oxford Street (A) and TAPAS (B) cohorts

A)

Compounds	NO_2	PM_{10}	PM_{25}	BC	UFPC
NO_2	1	0.3623144	0.4552182	-0.05911848	0.04976456
PM_{10}	0.36231444	1	0.9298075	0.48444264	0.41526547
PM_{25}	0.45521824	0.9298075	1	0.41730369	0.39308259
BC	-0.05911848	0.4844426	0.4173037	1	0.88437796
UFPC	0.04976456	0.4152655	0.3930826	0.88437796	1

B)

Compounds	NO	NOx	PM10	PM2.5	BC	UFPC	CO	CO2
NO	1	0.9930459	0.5552723	0.7855424	0.9460405	0.9789256	0.5495928	0.8417301
NOx	0.9930459	1	0.5418992	0.7741751	0.9202693	0.9701027	0.6317637	0.8362284
PM10	0.5552723	0.5418992	1	0.8990650	0.7145419	0.5934033	0.2035742	0.7979145
PM2.5	0.7855424	0.7741751	0.8990650	1	0.8789331	0.8244449	0.3178516	0.8970416
BC	0.9460405	0.9202693	0.7145419	0.8789331	1	0.9472066	0.3913952	0.9091338
UFPC	0.9789256	0.9701027	0.5934033	0.8244449	0.9472066	1	0.5045155	0.8618783
CO	0.5495928	0.6317637	0.2035742	0.3178516	0.3913952	0.5045155	1	0.4346168
CO2	0.8417301	0.8362284	0.7979145	0.8970416	0.9091338	0.8618783	0.4346168	1

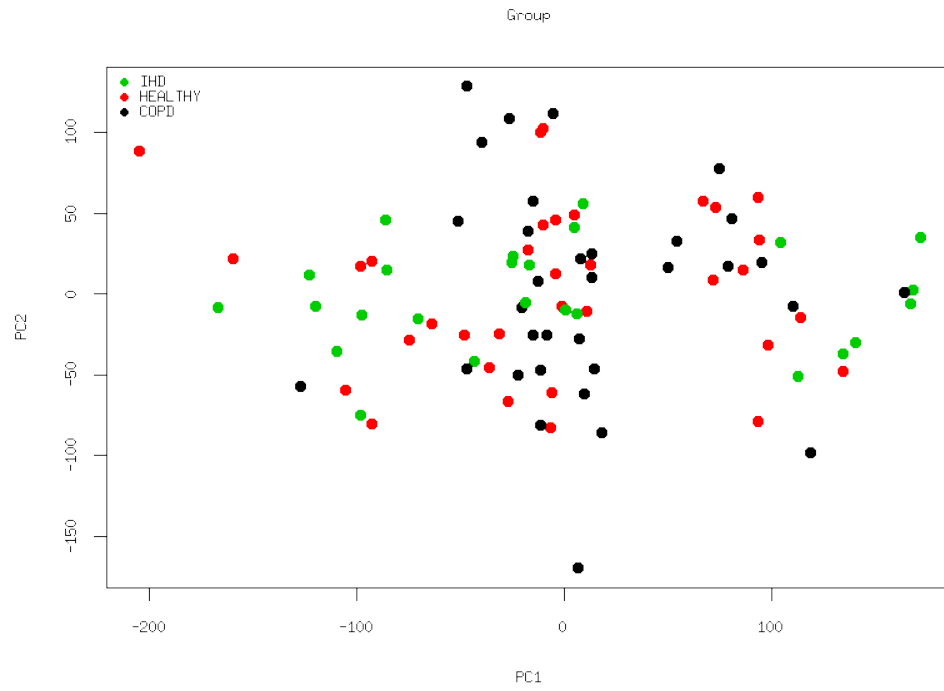Figure S2. PCA plot of gene expression from the Oxford Street color by disease group.

Figure S3. Overlap of hits significantly associated with NO_x, PM_{10} and $PM_{2.5}$ from the Oxford Street (A) and TAPAS (B) cohorts for transcriptomics.

A)

B)
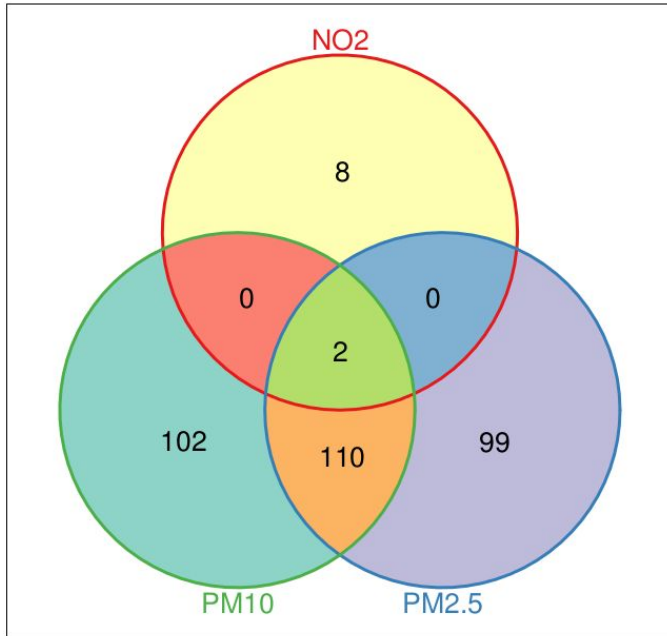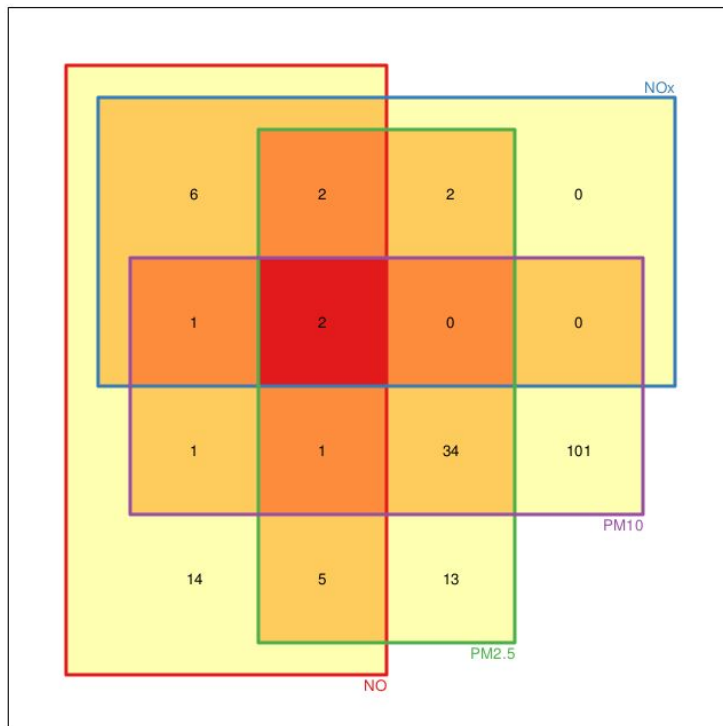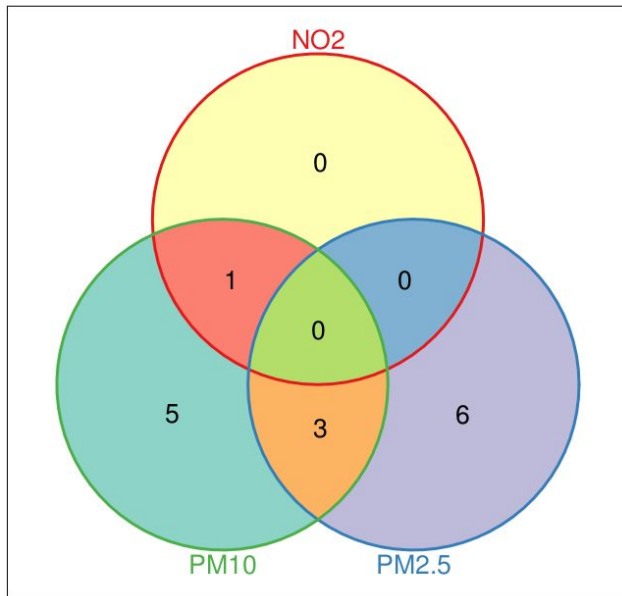
Figure S4. Overlap of hits significantly associated with NO_x, PM_{10} and $PM_{2.5}$ from the Oxford Street (A) and TAPAS (B) cohorts for microRNA.
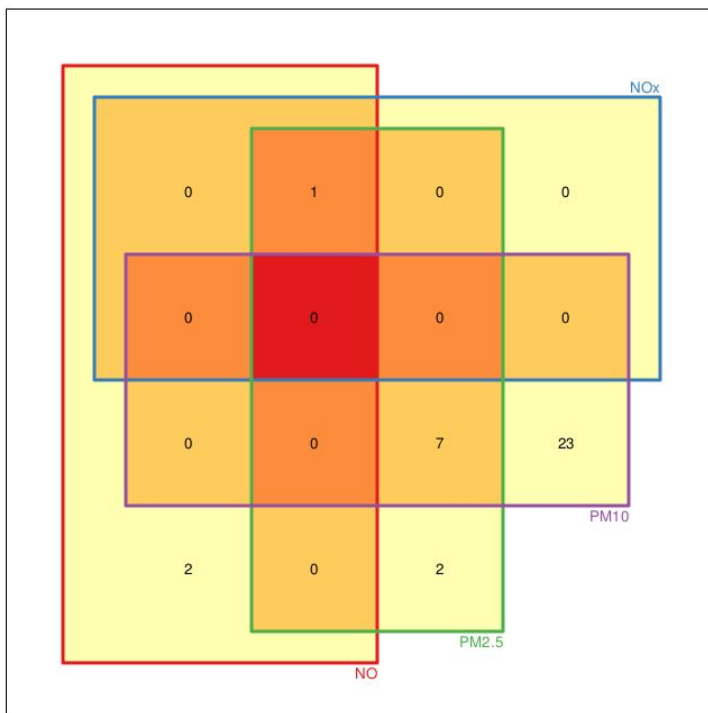
A)

B)

Table S3. Stability analysis of hits from gene expression from the Oxford Street (A) and TAPAS (B) cohorts

A)

Exposure	All transcripts Bonf<0.05	Capping 95%-5% Bonf<0.05	Capping 90%-10% Bonf<0.05	Capping 80%-20% Bonf<0.05	Full vs 95%-5% Bonf<0.05	Full vs 90%-10% Bonf<0.05	Full vs 80%-20% Bonf<0.05
NO_2	11	13	11	13	0	0	0
PM_{10}	229	13	16	19	4	6	5
$PM_{2.5}$	227	22	25	29	5	6	7
BC	0	4	4	6	0	0	0
UFPC	0	0	0	0	0	0	0

B)

Exposure	All transcripts Bonf<0.05	Capping 95%-5% Bonf<0.05	Capping 90%-10% Bonf<0.05	Capping 80%-20% Bonf<0.05	Full vs 95%-5% Bonf<0.05	Full vs 90%-10% Bonf<0.05	Full vs 80%-20% Bonf<0.05
NO	32	49	76	184	23	18	15
NO_x	13	16	16	55	8	6	3
PM_{10}	140	175	247	325	97	85	54
$PM_{2.5}$	59	92	131	206	41	33	18
BC	40	56	75	190	23	19	17
UFPC	34	46	62	139	22	18	14
CO	157	67	70	81	53	41	27
CO_2	65	85	115	178	39	31	24

Table S4. Pathway analysis of the genes significantly associated with the different exposures from the Oxford Street (A) and TAPAS (B) cohorts

A)

Exposure	q-value	Pathways
PM_{10}	0.09363881	Degradation of DVL
PM_{10}	0.09363881	Asymmetric localization of PCP proteins
PM_{10}	0.09363881	Assembly Of The HIV Virion
PM_{10}	0.09363881	Oxygen-dependent proline hydroxylation of Hypoxia-inducible Factor Alpha
PM_{10}	0.09363881	Regulation of TP53 Activity through Methylation
PM_{10}	0.09363881	PCP/CE pathway
PM_{10}	0.09363881	Ubiquitin-dependent degradation of Cyclin D1
PM_{10}	0.09363881	Regulation of Hypoxia-inducible Factor (HIF) by oxygen
PM_{10}	0.09363881	Cellular response to hypoxia
PM_{10}	0.09363881	CDK-mediated phosphorylation and removal of Cdc6
PM_{10}	0.09363881	Gene Expression

PM$_{10}$	0.09363881	Autodegradation of the E3 ubiquitin ligase COP1
PM$_{10}$	0.09363881	Vpu mediated degradation of CD4
PM$_{10}$	0.09363881	Downregulation of ERBB4 signaling
PM$_{10}$	0.09363881	InlA-mediated entry of Listeria monocytogenes into host cells
PM$_{10}$	0.09363881	IKK complex recruitment mediated by RIP1
PM$_{10}$	0.09363881	Antigen processing: Ubiquitination & Proteasome degradation
PM$_{10}$	0.09363881	Attenuation phase
PM$_{10}$	0.09363881	Regulation of TP53 Activity through Phosphorylation
PM$_{10}$	0.09363881	Ubiquitin Mediated Degradation of Phosphorylated Cdc25A
PM$_{10}$	0.09363881	p53-Independent DNA Damage Response
PM$_{10}$	0.09363881	p53-Independent G1/S DNA damage checkpoint
PM$_{10}$	0.09363881	Vif-mediated degradation of APOBEC3G
PM$_{10}$	0.09363881	Reversible hydration of carbon dioxide
PM$_{10}$	0.09730681	Late Phase of HIV Life Cycle
PM$_{10}$	0.09730681	PTK6 Regulates RTKs and Their Effectors AKT1 and DOK1
PM$_{10}$	0.09730681	Degradation of AXIN
PM$_{10}$	0.09730681	HSF1 activation
PM$_{10}$	0.09730681	Recognition of DNA damage by PCNA-containing replication complex
PM$_{10}$	0.09730681	TGF-beta receptor signaling activates SMADs
PM$_{10}$	0.09730681	Synthesis of active ubiquitin: roles of E1 and E2 enzymes
PM$_{10}$	0.09730681	Fluoropyrimidine Activity
PM$_{10}$	0.09730681	Ganglio Sphingolipid Metabolism
PM$_{10}$	0.09730681	Stabilization of p53
PM$_{10}$	0.09730681	FBXL7 down-regulates AURKA during mitotic entry and in early mitosis
PM$_{10}$	0.09730681	SCF-beta-TrCP mediated degradation of Emi1
PM$_{10}$	0.09730681	Association of licensing factors with the pre-replicative complex
PM$_{10}$	0.09740097	HSF1-dependent transactivation
PM$_{10}$	0.09740097	MAPK6/MAPK4 signaling
PM$_{10}$	0.09740097	HIV Life Cycle
PM$_{10}$	0.09740097	Membrane binding and targetting of GAG proteins
PM$_{10}$	0.09740097	Synthesis And Processing Of GAG
PM$_{10}$	0.09740097	Josephin domain DUBs
PM$_{10}$	0.09740097	Hh mutants that don
PM$_{10}$	0.09740097	Hh mutants abrogate ligand secretion
PM$_{25}$	0.219517381	Methionine De Novo and Salvage Pathway
PM$_{25}$	0.219517381	Spermidine and Spermine Biosynthesis

Exposure	q-value	Pathways
PM_{25}	0.219517381	Betaine Metabolism
PM_{25}	0.219517381	Reversible hydration of carbon dioxide
PM_{25}	0.219517381	Trans-sulfuration pathway
PM_{25}	0.219517381	Ganglio Sphingolipid Metabolism
PM_{25}	0.219517381	Generic Transcription Pathway
PM_{25}	0.219517381	sarcosine oncometabolite pathway

B)

Exposure	q-value	Pathways
NO	0.022157	Regulation of RAC1 activity
NO	0.022157	GPCR downstream signaling
NO	0.022157	NRAGE signals death through JNK
NO	0.026749	Cell death signalling via NRAGE
NO	0.026749	G alpha (12/13) signalling events
NO	0.027347	p75 NTR receptor-mediated signalling
NO	0.027347	Mesodermal Commitment Pathway
NO	0.027347	Signaling by GPCR
NO_x	0.009017	mRNA Splicing - Major Pathway
NO_x	0.009017	mRNA Splicing
NO_x	0.009731	Processing of Capped Intron-Containing Pre-mRNA
PM_{10}	0.108216	Syndecan interactions
PM_{10}	0.108216	Wnt-beta-catenin Signaling Pathway in Leukemia
PM_{10}	0.108216	ECM-receptor interaction - Homo sapiens (human)
PM_{10}	0.108216	O-glycosylation of TSR domain-containing proteins
PM_{10}	0.108216	HS-GAG degradation
PM_{10}	0.108216	Pancreatic secretion - Homo sapiens (human)
PM_{10}	0.108216	O-linked glycosylation
BC	0.123427	Glycosaminoglycan metabolism
CO	0.14675	Arginine and proline metabolism - Homo sapiens (human)
CO	0.14675	Nonsense Mediated Decay (NMD) enhanced by the Exon Junction Complex (EJC)
CO	0.14675	Nonsense-Mediated Decay (NMD)
CO	0.14675	Translation
CO	0.14675	SRP-dependent cotranslational protein targeting to membrane
CO	0.14675	Synthesis of glycosylphosphatidylinositol (GPI)

Table S5. Genes significantly associated with PM$_{10}$ targeted by significant microRNAs from the Oxford Street (A) and TAPAS (B) cohorts.

A)

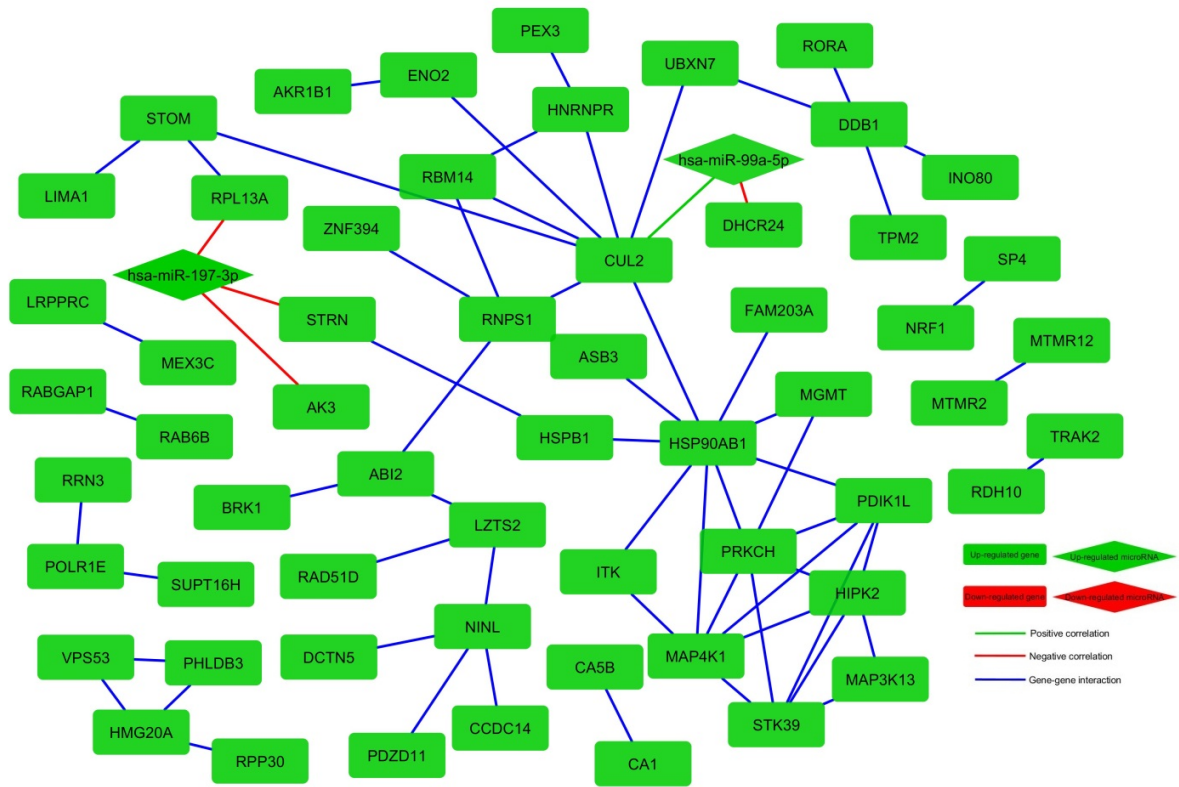| Gene	Function
ENO2	Isoenzyme found in mature neurons and cells of neuronal origin. Among its related pathways are HIF-1 signaling pathway and Glucose metabolism.
HMG20A	High Mobility Group 20A. Plays a role in neuronal differentiation as chromatin-associated protein. Induces the activation of neuronal-specific genes.
KIAA1191	Potential NADPH-dependent oxidoreductase. May be involved in the regulation of neuronal survival, differentiation and axonal outgrowth.
MANEAL	Mannosidase Endo-Alpha Like, hydrolase activity.
ADIPOR1	Receptor for adiponectin. It activates an AMP-activated kinase signaling pathway which affects levels of fatty acid oxidation and insulin sensitivity
OSBP	Intracellular protein that is believed to transport sterols from lysosomes to the nucleus where the sterol down-regulates the genes for the LDL receptor, HMG-CoA reductase, and HMG synthetase.

B)

Gene	Function
AXIN2	Important role in the regulation of the stability of beta-catenin in the Wnt signaling pathway. The deregulation of beta-catenin is an important event in the genesis of a number of malignancies
CCAR1	Cell Division Cycle And Apoptosis Regulator 1. May play an important role in the growth and tumorigenesis of prostate cancer cells
CDC20	Regulatory protein interacting with several other proteins at multiple points in the cell cyclE
CDK11A	Serine/threonine protein kinase family. Essential for eukaryotic cell cycle control. May play a role in cell apoptosis
CENPJ	During cell division, this protein plays a structural role in the maintenance of centrosome integrity and normal spindle morphology
HIST1H1E	Responsible for nucleosome structure of the chromosomal fiber. Among its related pathways are Cellular Senescence and Signaling events mediated by HDAC Class III
LGALS3BP	Beta-galactoside-binding proteins implicated in modulating cell-cell and cell-matrix interactions. Found elevated in the serum of patients with cancer and in those infected by HIV
MIS12	Kinetochore Complex Component. Related pathways are Signaling by GPCR and Cell Cycle, Mitotic
MYBL1	MYB Proto-Oncogene Like 1. Diseases associated with MYBL1 include Burkitt Lymphoma. Among its related pathways are HTLV-I infection and Mitotic Prophase
MYO19	Myosin XIX, involved in mitochondrial motility
NOL11	Nucleolar Protein 11. Among its related pathways are rRNA processing and Gene Expression
POGZ	Zinc finger protein containing a transposase domain at the C-terminus. Plays a role in mitotic cell cycle progression and is involved in kinetochore assembly and mitotic sister chromatid cohesion
RCE1	Member of the metalloproteinase family. This enzyme is thought to function in the maintenance and processing of CAAX-type prenylated proteins
SUGP2	Arginine/serine-rich family of splicing factors, functions in mRNA processing
SYNCRIP	Ribonucleoprotein. Role in mRNA maturation. Associated with several multiprotein complexes including the apoB RNA editing-complex and survival of motor neurons (SMN) complex.
THBS1	Adhesive glycoprotein that mediates cell-to-cell and cell-to-matrix interactions. Shown to play roles in platelet aggregation, angiogenesis, and tumorigenesis
VKORC1	Catalytic subunit of the vitamin K epoxide reductase complex, which is responsible for the reduction of inactive vitamin K 2,3-epoxide to active vitamin K
ZNF322	Zinc-finger transcription factor family and may regulate transcriptional activation in MAPK

Figure S5. Cytoscape figure of hits significantly associated with PM$_{2.5}$ from Oxford Street (A) and TAPAS (B).
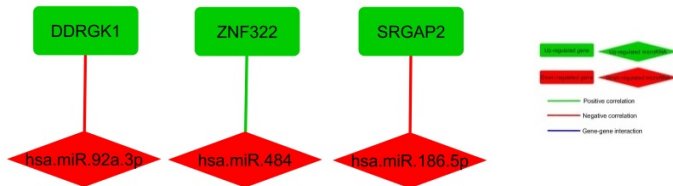
A)

B)

Table S6. Genes significantly associated with PM$_{2.5}$ targeted by significant microRNAs from the Oxford Street (A) and TAPAS (B) cohorts.

A)

Gene	Function
AK3	Involved in maintaining the homeostasis of cellular nucleotides. Among its related pathways are Factors involved in megakaryocyte development and platelet production and Purine metabolism
CUL2	Mediates ubiquitination of target proteins. Diseases associated are Renal Cell Carcinoma and Kidney Cancer. Among its related pathways are Hypoxic and oxygen homeostasis regulation of HIF-1-alpha and Immune System
DHCR24	Flavin adenine dinucleotide. Protects cells from oxidative stress during apoptosis. Reduced expression found in the temporal cortex of Alzheimer disease patients and overexpression in adrenal gland cancer cells.
RPL13A	Mediates interferon-gamma-induced transcript-selective translation inhibition in inflammation processes. Among its related pathways are Metabolism and Viral mRNA Translation.
STRN	Calmodulin-binding protein. Among its related pathways are Neurophysiological process Glutamate regulation of Dopamine D1A receptor signaling and Plasma membrane estrogen receptor signaling.

B)

Gene	Function
DDRGK1	Interacts with components of the ubiquitin fold modifier 1 conjugation pathway and helps prevent apoptosis in ER-stressed secretory tissues
SRGAP2	Encodes a member of the SLIT-ROBO Rho GTPase activating protein family. Plays an important role in different aspects of neuronal morphogenesis and migration
ZNF322	Zinc-finger transcription factor family and may regulate transcriptional activation in MAPK

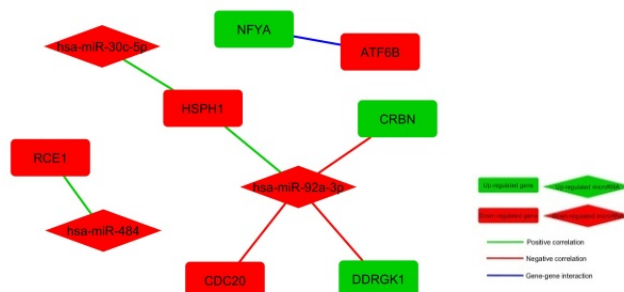Figure S6. Cytoscape figure of hits significantly associated with CO$_2$ from the TAPAS cohort.

Table S7. Genes significantly associated with CO_2 targeted by significant microRNAs from the TAPAS cohort.

Gene	Function
CDC20	Regulatory protein interacting with several other proteins at multiple points in the cell cyclE
CRBN	Mediates ubiquitination of target proteins. May play a role in memory and learning by regulating the assembly and neuronal surface expression of potassium channels
DDRGK1	Interacts with components of the ubiquitin fold modifier 1 conjugation pathway and helps prevent apoptosis in ER-stressed secretory tissues
HSPH1	Heat Shock Protein Family H. Among its related pathways are Vesicle-mediated transport and Cellular response to heat stress
RCE1	Classified as a member of the metalloproteinase family. Among its related pathways are G-protein signaling_Rap1B regulation pathway and Metabolism of proteins