Environment International 168 (2022) 107491

Contents lists available at ScienceDirect

Environment International

journal homepage: www.elsevier.com/locate/envint

Full length article

Impact of long-term exposure to $PM_{2.5}$ on peripheral blood gene expression pathways involved in cell signaling and immune response

Jelle Vlaanderen [a,*], Roel Vermeulen [a,b], Matthew Whitaker [c], Marc Chadeau-Hyam [c], Jouke-Jan Hottenga [d], Eco de Geus [d], Gonneke Willemsen [d], Brenda W.J.H. Penninx [e], Rick Jansen [e], Dorret I. Boomsma [d]

[a] Institute for Risk Assessment Sciences, Utrecht University, the Netherlands
[b] Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht, the Netherlands
[c] School of Public Health, Imperial College London, London, UK
[d] Department of Biological Psychology, Vrije Universiteit, Amsterdam, the Netherlands
[e] Department of Psychiatry, Amsterdam UMC Location Vrije Universiteit Amsterdam, Amsterdam Public Health Research Institute and Amsterdam Neuroscience, Amsterdam, the Netherlands

ARTICLE INFO

Handling Editor: Adrian Covaci

Keywords:
Air pollution
Transcriptomics
Molecular epidemiology
Biological pathways
Cell signaling
Immune system

ABSTRACT

Background: Exposure to ambient air pollution, even at low levels, is a major environmental health risk. The peripheral blood transcriptome provides a potential avenue for the elucidation of ambient air pollution related biological perturbations. We assessed the association between long-term estimates for seven priority air pollutants and perturbations in peripheral blood transcriptomics data collected in the Dutch National Twin Register (NTR) and Netherlands Study of Depression and Anxiety (NESDA) cohorts.

Methods: In both the discovery (n = 2438) and replication (n = 1567) cohort, outdoor concentration of 7 air pollutants (NO_2, NO_x, particulate matter ($PM_{2.5}$, $PM_{2.5}$abs, PM_{10}, PM_{coarse}), and ultrafine particles) was predicted with land use regression models. Gene expression was assessed by Affymetrix U219 arrays. Multi-variable univariate mixed-effect models were applied to test for an association between the air pollutants and the transcriptome. Functional analysis was conducted in DAVID.

Results: In the discovery cohort, we observed for 335 genes (374 probes with FDR < 5 %) a perturbation in peripheral blood gene expression that was associated with long-term average levels of $PM_{2.5}$. For 69 genes pooled effect estimates from the NTR and NESDA cohorts were significant. Identified genes play a role in biological pathways related to cell signaling and immune response. Sixty-two out of 69 genes had a similar direction of effect in an analysis in which we regressed the probes on differential $PM_{2.5}$ exposure within monozygotic twin pairs, indicating that the observed differences in gene expression were likely driven by differences in air pollution, rather than by confounding by genetic factors.

Conclusion: Our results indicate that $PM_{2.5}$ can elicit a response in cell signaling and the immune system, both hallmarks of environmental diseases. The differential effect that we observed between air pollutants may aid in the understanding of differential health effects that have been observed with these exposures.

1. Introduction

Over the years epidemiological studies have convincingly shown that exposure to ambient air pollution is a major environmental health risk, even at levels below European exposure limits (Liu et al., 2021; Mostafavi et al., 2017; Strak et al., 2021). Recent estimates indicate that long-term exposure to ambient air pollution (as measured by $PM_{2.5}$) contributed to more than 4 million deaths worldwide in 2016, primarily due to cardiovascular and respiratory diseases, and lung cancer (Health Effects Institute, 2018).

Although the exact mechanisms behind the observed associations have not clearly been delineated, long-term pulmonary oxidative stress and inflammation induced by chronic exposure to inhaled pollutants has been hypothesized to result in a systemic inflammatory state capable of

* Corresponding author at: Division of Environmental Epidemiology, Institute for Risk Assessment Sciences (IRAS), Utrecht University, PO Box 80178, 3508 TD Utrecht, the Netherlands.
E-mail address: J.J.Vlaanderen@uu.nl (J. Vlaanderen).

https://doi.org/10.1016/j.envint.2022.107491
Received 12 May 2022; Received in revised form 2 August 2022; Accepted 25 August 2022
Available online 29 August 2022

activating hemostatic pathways, impairing vascular function, and accelerating atherosclerosis (Brook et al., 2010; Peters et al., 2021).

Further understanding of the biological mechanisms of air pollution-related health effects allows for a better underpinning of the existing empirical evidence, and aids in the identification of previously undiscovered air pollutant related health effects. For example, O'Beirne et al. recently demonstrated that $PM_{2.5}$ exposure contributes to significant dysregulation of the small airway epithelium transcriptome of smokers thereby providing insights into the role of air pollution in the development of respiratory disease among susceptible individuals (O'Beirne et al., 2018). Croft et al. evaluated transcriptome patterns in peripheral blood and showed an air pollutant, timing, and infections specific effect of air pollution on the pathogenesis of respiratory infection (Croft et al., 2021).

The peripheral blood transcriptome provides a potential avenue for the elucidation of ambient air pollution related biological perturbations in epidemiological studies as it reflects physiological and pathological events occurring in different tissues of the body (Mohr and Liew, 2007). There is proof-of-principle that the peripheral blood transcriptome can serve to identify early, disease-related perturbations caused by environmental exposures; for example, from studies of cigarette smoking (an exposure in many ways similar to particulate air pollutants). Smoking-modified genome-wide gene expression profiles have been shown to reliably predict diseases and conditions independently known to be causally associated with cigarette smoking (Huan et al., 2016; Vink et al., 2017).

To date, a limited number of epidemiological studies has explored the impact of ambient air pollution on the peripheral blood transcriptome, most of which have focused on acute (transient) perturbations in response to variations in pollutant levels (Huang et al., 2010; Peretz et al., 2007; Pettit et al., 2012; Wittkopp et al., 2016). Four studies have focused on the impact of ambient air pollution on (non-transient) dysregulation of the peripheral blood transcriptome (Merid et al., 2021; Mostafavi et al., 2017; Vrijens et al., 2017; Winckelmans et al., 2017). Vrijens et al. (n = 98 discovery cohort; n = 175 validation cohort) reported associations between 2-year average concentrations of $PM_{2.5}$ and a range of dysregulated pathways including cell-cell communication (among men) and respiratory electron transport (among women). In the same cohort (n = 98), Winckelmans et al. reported associations between one-month average PM_{10} exposure and electron transport chain and the Tri Carbonic Acid cycle. Mostafavi et al. (n = 550) described subtle changes in gene expression related to 2-year average exposure to NO_x, though no pathways were significantly enriched. Merid et al. (n = 656) reported two differentially expressed transcript clusters associated with exposure to $PM_{2.5}$ at birth, one of them mapping to the MIR1296 gene. A main limitation of these studies is their comparatively small study size, limiting their statistical power to detect (likely subtle) ambient air pollution induced dysregulation of the peripheral blood transcriptome.

Here we present the largest study to date looking at the effects of air pollution on gene expression. Our study includes roughly four times as many individuals as have been included in the combined evidence base available so far.

We combine gene expression data from two large Dutch cohorts (Netherlands Twin Register (NTR), n = 2438) (Willemsen et al., 2013), discovery cohort and the Netherlands Study of Depression and Anxiety (NESDA), n = 1567, replication cohort) (Penninx et al., 2008), with state of the art assessment of seven priority ambient air pollution pollutants (PM2.5, NO2, NOx, PM2.5abs, PM10, PMcoarse, and UFP), and two non-air pollution indicators of the urban environment: the percentage greenspace, and the neighborhood socio-economic status.

In addition, the presence of a high number of monozygotic (MZ) twin pairs in the NTR cohort provided us with the possibility to perform a within-family analysis. MZ twins are perfectly matched for genotype and early (including prenatal) shared environmental exposures, such as maternal smoking during pregnancy and air pollution. As previously shown for smoking behavior, analyses within MZ pairs can help to differentiate whether the associations we observed between long-term exposure to ambient air pollution and dysregulated gene expression are the result of a reaction of gene expression to exposure to ambient air pollution, or are confounded by genetic factors (Vink et al., 2017). As differential gene expression between a highly exposed MZ twin and the genetically identical lower exposed co-twin cannot be caused by differences in genetic liability, observed differential expression is therefore likely to be reactive to air pollution exposure.

2. Methods

2.1. Study populations

This study incorporated data from two studies: the Netherlands Twin Register (NTR) (Willemsen et al., 2010) and the Netherlands Study of Depression and Anxiety (NESDA) (Penninx et al., 2008). Both studies were approved by the Central Ethics Committee on Research Involving Human Subjects of the VU University Medical Center, Amsterdam, and all subjects provided written informed consent. Blood samples were collected between January 2004 and July 2008. Data from participants without blood samples or geo-location information at the time of blood sampling were excluded. NTR contributed data from 2438 individuals (66 % women), including 1104 MZ twins (421 complete twin pairs) and 974 dizygotic twins (207 complete twin pairs). NESDA contributed data of 1567 individuals (66 % women). NTR was defined as the discovery cohort and NESDA as the replication cohort.

2.2. Exposure assessment

Using residential address information at the time of blood sampling, annual outdoor concentrations of seven air pollutants were generated for all study participants based on land use regression models published in the peer reviewed literature and used in previous health studies: $PM_{2.5}$ (Particulate matter with a diameter < 2.5 μm), NO_2 (Nitrogen dioxide), NO_x (Nitrogen oxides) (Beelen et al., 2013); $PM_{2.5abs}$ (Absorbance of $PM_{2.5}$), PM_{10} (Particulate matter with a diameter < 10 μm), Pm_{coarse} (Particulate matter with a diameter 2.5–10 μm) (Eeftens et al., 2012); and UFP (ultrafine particulates) (Van Nunen et al., 2017). Due to their skewed distribution, we log-transformed all air pollutants before inclusion into the statistical analyses. In addition, we generated estimates for two indicators of surrounding green space: Normalized difference vegetation index (NDVI) (Weier and Herring, 2000) and TOP10NL (a national land-use database of the Netherlands) (Kadaster), and included 'Percentage of households with a low purchasing power' (PLPP) as indicator of neighborhood socioeconomic status.

2.3. Blood sampling and gene expression assessment

Blood sampling and RNA extraction procedures have been described in detail previously. (Jansen, 2014; Ouwens, 2020; Willemsen, 2010; Wright, 2014) Briefly, venous blood samples were collected 7–10 AM (NTR) and 8–10 AM (NESDA) after an overnight fast. In NTR, blood collection in fertile women took place, when possible, at a fixed time of their menstrual cycle. Within 60 min after blood draw, blood collected in lithium heparin coagulant tubes was transferred into PAXgene Blood RNA tubes (Qiagen, Valencia, Florida, USA) and stored at −20 °C. Total RNA was extracted according to the manufacturer's protocol (Qiagen) (Spijker et al., 2004; Willemsen et al., 2010).

Gene expression was assessed at the Rutgers University Cell and DNA Repository. Samples were randomly assigned to plates. Samples were hybridized to Affymetrix U219 array plates (GeneTitan, Affymetrix, Santa Clara, California, USA). Array hybridization, washing, staining and scanning were carried out in an Affymetrix GeneTitan System per the manufacturer's protocol.

Probes were removed if they did not pass standard Affymetrix quality

control metrics (Affymetrix expression console) and when their location was uncertain or if their location intersected a polymorphic single nucleotide polymorphism (SNP), leaving 44 241 probe sets for analysis (Peters et al., 2021). Expression values were obtained using robust multi-array average normalization implemented in Affymetrix Power Tools (v 1.12.0). Data for samples that displayed an average Pearson correlation below 0.8 with the probe set expression values of other samples and samples with incorrect sex chromosome expression were removed.

2.4. Statistical methods

We applied linear mixed-models (LMM) to test for associations between each of the seven air pollutants and the 44 241 probes in NTR. Models were adjusted for: plate, well, family ID, month of blood sampling, sex, age, smoking status, BMI, total white blood cell counts, days between sampling and extraction, days between extraction and amplification, and sampling time. Observed associations with a Benjamini and Hochberg adjusted p-values (BH) < 0.05 were considered statistically significant. We conducted sensitivity analyses for the probes that were significantly associated with air pollution in the screening analysis: we windsorized the exposure distribution at the 5th and 95th percentiles of the original exposure distribution to assess the impact of outliers in the exposure distribution on the results. We conducted analyses among non-smokers and smokers separately to assess the impact of smoking behavior. Finally, we fit a multiple regression model that included all seven pollutants, two indicators for surrounding greenness, and neighborhood socio-economic status.

Analyses within MZ twin pairs were conducted by regressing residualized differential probe levels, using the confounder model from the linear mixed models as applied in the screening analyses, on differential air pollution exposure levels. Differential air pollution exposure levels were calculated by subtracting estimated exposures within twin pairs.

Replication analyses in NESDA were conducted by applying LMM. Models were adjusted for plate, well, month of blood sampling, lab (included as random effects), as well as sex, age, smoking status, BMI, time of blood sampling, red blood cell count, and hematocrit (included as fixed effects). We conducted a meta-analysis to combine results from NTR and NESDA for associations that were significantly associated in NTR and had a similar direction of effect in NESDA. We call probes that were significant in the meta-analysis 'robust'. No adjustment for multiple testing was applied in the replication analyses.

Statistical analyses were performed in R version 3.6.1 using the lme4 package v1.1.21 (Douglas et al., 2014) and meta package 5.1.1.

Robust genes were submitted to DAVID v6.8 (Huang et al., 2009) for gene functional classification, functional annotation, and functional annotation clustering. An enrichment score of 1.2 and a BH < 0.05 were defined as threshold for enrichment. As reference set for enrichment analysis we included all 18,238 genes targeted by the Affymetrix U219 arrays. Results from LMM analyses for all probes in both NTR and NESDA were also input for Gene Set Enrichment Analysis (GSEA). GSEA was assessed for all Gene-Ontology terms using the FGSEA package in R v 1.12 (Korotkevich et al.). Five hundred permutations were run to estimate the p-value and a q-value < 0.05 was used a threshold for significant enrichment. We retained the terms with significant enrichment in both NTR and NESDA.

3. Results

General descriptive statistics for NTR and NESDA participants are provided in Table 1. The two cohorts are highly comparable in terms of gender balance, smoking and BMI status, and age. Exposure distributions for the seven air pollutants are very similar for NTR and NESDA (Fig. 1). As observed in other studies, the air pollutants are moderately to highly correlated (Fig. S1). $PM_{2.5}$ stands out with more moderate correlation with the other pollutants, ranging from 0.39 (with PM_{coarse}) to 0.79 (with $PM_{2.5}$absorbance). Between-pollutant correlations in NESDA are slightly higher than in NTR.

Table 1
Main characteristics of participants from the NTR[a] and NESDA[b] cohorts that are part of this study.

Variable	NTR[a]	NESDA[b]
N	2438	1567
Sex		
Men	825	528
Women	1613	1039
Smoking status (%)		
Never smoked	1337 (54.8)	956
Current smoker	515 (21.1)	611
Former smoker	586 (24.0)	
BMI (%)		
BMI 20–30	1987 (81.5)	1161
BMI < 20	250 (10.3)	129
BMI > 30	201 (8.2)	277
Age (mean (SD))	36.9 (13.04)	42.3 (12.7)

[a] Netherlands Twin Register.
[b] Netherlands Study of Depression and Anxiety.

3.1. Screening analysis

In the univariate screening analysis within NTR we observed a signal for $PM_{2.5}$ (374 probes (335 genes) associated at BH < 0.05), and no strong associations for any of the other air pollutants (Table 2). Sensitivity analyses are presented in Table S1. The 374 probes associated with $PM_{2.5}$ were robust in a minimal model (only corrected for well, plate, family ID, and month of sampling) and when the exposure distribution was winsorized at the 5th and 95th percentiles of the exposure distribution. In a model in which we additionally corrected for the other pollutants (NO_2, NO_X, $PM_{2.5abs}$, PM_{10}, PM_{coarse}, and UFP), and the two urban exposome factors (proportion greenspace and neighborhood socio-economic status), the number of probes associated with $PM_{2.5}$ at BH < 0.05 dropped to 253. Among non-smokers only (1337 individuals), 357 probes remained nominally associated with $PM_{2.5}$, while among smokers (515 individuals) only 137 probes remained nominally associated. The direction of effect of all 374 observed associations (286 upregulated, and 88 downregulated) was constant across all sensitivity analyses.

3.2. Replication analysis

Out of the 374 probes significantly associated with PM2.5 in NTR, 198 (52.9 %) had a similar direction of effect in the replication analysis in NESDA. 71 of these probes (69 genes) were significantly associated in the meta-analysis, including 5 probes that were independently significantly associated with $PM_{2.5}$ in NESDA as well ($p < 0.05$). In Table 3 we report the top 25 probes with coefficient concordance between NTR and NESDA and for which the confidence interval from the meta-analysis excluded zero (all 71 probes are reported in Table S2).

3.3. Within MZ twin pair analysis

There was moderate concordance in $PM_{2.5}$ exposure levels within MZ twin pairs (Pearson correlation coefficient 0.56, Fig. 2). Among the 71 probes (69 genes) that were robustly associated with $PM_{2.5}$ in both NTR and NESDA in the meta-analysis, 62 had a similar direction of effect in the within MZ twin pair analysis in the NTR study (421 pairs). Five of those were nominally significant ($p < 0.05$) associated with $PM_{2.5}$ (Table S3).

3.4. Pathway analysis

Gene functional classification in DAVID based on the 69 robustly

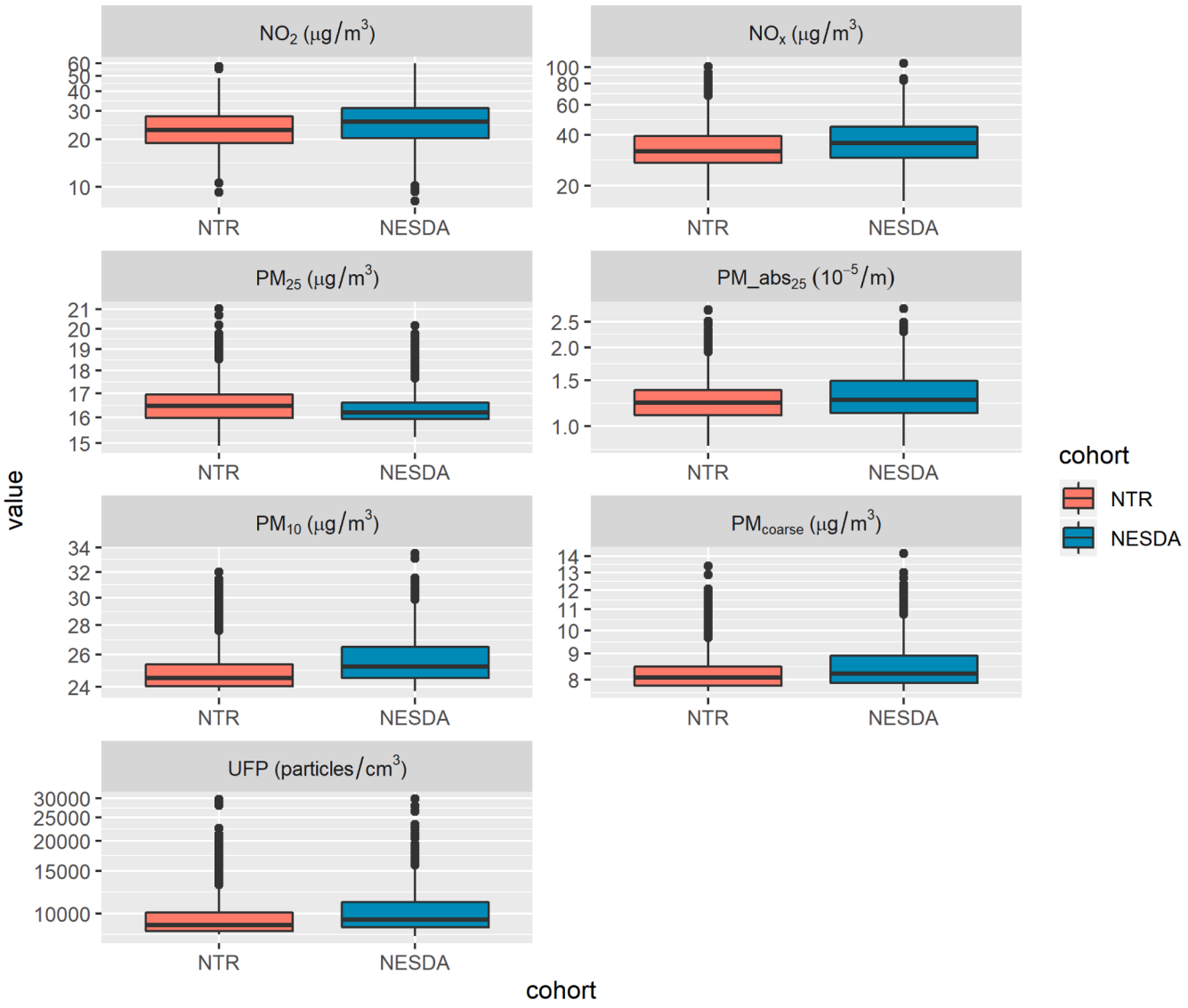

Fig. 1. Exposure distributions for the seven air pollutants included in the current analysis within the NTR and NESDA cohorts.

Table 2
Number of associations observed between air pollutants and gene expression in the NTR cohort; results from univariate screening.[a]

	$PM_{2.5}$[b]	NO2[b]	NOx[b]	$PM_{2.5abs}$[b]	PM_{10}[b]	PM_{coarse}[b]	UFP[b]
P < 0.05	5940	2810	2082	3011	2107	2385	2001
BH20[c]	2560	0	0	1	0	0	1
BH5[d]	374	0	0	0	0	0	0
BH1[e]	8	0	0	0	0	0	0

[a] Univariate NTR models adjusted for: plate, well, family ID, month of sampling, sex, age, smoking status, BMI, white blood cell counts, days between sampling and extraction, days between extraction and amplification, sampling time.
[b] $PM_{2.5}$ (Particulate matter with a diameter < 2.5 μm), NO_2 (Nitrogen dioxide), NO_x (Nitrogen oxides); $PM_{2.5abs}$ (Absorbance of $PM_{2.5}$), PM_{10} (Particulate matter with a diameter < 10 μm), Pm_{coarse} (Particulate matter with a diameter 2.5–10 μm), and UFP (ultrafine particulates).
[c] Benjamini and Hochberg adjusted p-values (BH) < 0.20.
[d] Benjamini and Hochberg adjusted p-values (BH) < 0.05.
[e] Benjamini and Hochberg adjusted p-values (BH) < 0.01.

associated genes resulted in one group of highly related genes, identified with an enrichment score of 1.19 (marginally below our predefined threshold for enrichment). The group includes RNA binding motif protein 23 (RBM23), RNA binding motif protein 12B (RBM12B), serine and arginine rich splicing factor 11(SRSF11), and heterogeneous nuclear ribonucleoprotein L like (HNRNPLL). These genes are involved in RNA binding. We identified one annotation cluster (Enrichment score 1.55, Table S4) consisting of pathways related to cell signaling and identified one significantly enriched term: "phosphoprotein" (overlap of 43 genes, BH p-value = 0.022), related to post translational modification (Table S5).

In Fig. 3 we present results from GSEA based on the complete results from screening analysis in NTR and NESDA. We identified three GO-terms that were consistently enriched (q < 0.05) in both cohorts:

Table 3
Top 25 probes with coefficient concordance between NTR and NESDA and CI meta analysis excluding 0, ordered by p-value NTR.[a]

probe	Gene	Chrom	β NTR[a]	SENTR[a]	p-value NTR[a]	B NTR[b]	SE NTR[b]	*p-value NTR*[b]	B NESDA[c]	SE NESDA[c]	p-val NESDA[c]	I2[d]	β meta[e]	Lower Meta[e]	Upper Meta[e]
11741727_a_at	NEDD1	chr12	0.879	0.179	1.24E−06	0.060	0.030	8.47E−04	0.317	0.259	2.32E−01	68.8 %	0.629	0.081	1.177
11719226_at	MOB1B	chr4	1.616	0.355	7.02E−06	0.077	0.029	3.06E−02	0.515	0.510	3.15E−01	68.1 %	1.127	0.055	2.198
11742903_at	CETN3	chr5	0.790	0.180	1.26E−05	0.014	0.021	4.23E−01	0.242	0.245	3.29E−01	69.3 %	0.542	0.007	1.076
11725311_s_at	TMED7	chr5	1.127	0.258	1.42E−05	0.064	0.045	1.37E−02	0.560	0.322	8.64E−02	46.8 %	0.876	0.325	1.428
11715925_a_at	TRAM1	chr8	1.169	0.269	1.49E−05	0.114	0.034	2.25E−05	0.553	0.400	1.74E−01	38.9 %	0.932	0.345	1.520
11744405_a_at	C2orf76	chr2	0.767	0.178	1.75E−05	0.033	0.020	6.70E−02	0.484	0.258	6.24E−02	0.0 %	0.676	0.389	0.963
11730899_s_at	WAPAL	chr10	1.267	0.295	1.94E−05	0.075	0.021	1.08E−02	0.419	0.393	2.90E−01	66.3 %	0.883	0.056	1.710
11741816_s_at	TAB3	chrX	1.176	0.275	2.35E−05	0.083	0.022	2.37E−03	0.506	0.402	2.15E−01	47.1 %	0.905	0.261	1.549
11722432_at	C1orf27	chr1	1.691	0.400	2.71E−05	0.106	0.032	8.16E−03	0.515	0.599	3.93E−01	62.5 %	1.188	0.047	2.328
11743651_a_at	DEK	chr6	0.888	0.210	2.83E−05	0.022	0.023	2.90E−01	0.391	0.280	1.68E−01	50.4 %	0.674	0.192	1.156
11731270_a_at	MGAM	chr7	−0.880	0.207	2.91E−05	−0.086	0.016	3.48E−05	−0.911	0.380	1.85E−02	0.0 %	−0.887	−1.244	−0.530
11758686_s_at	CD164	chr6	1.090	0.262	3.36E−05	0.079	0.016	2.45E−03	0.486	0.320	1.32E−01	53.1 %	0.816	0.227	1.405
11735999_a_at	ZNF791	chr19	0.907	0.220	4.05E−05	0.038	0.023	8.24E−02	0.610	0.318	5.83E−02	0.0 %	0.811	0.457	1.165
11720430_a_at	ORMDL1	chr2	0.864	0.211	4.49E−05	0.035	0.037	9.24E−02	0.339	0.280	2.30E−01	55.6 %	0.634	0.123	1.145
11739130_a_at	ABAT	chr16	−0.941	0.230	4.51E−05	−0.053	0.037	2.14E−02	−0.436	0.362	2.34E−01	27.9 %	−0.766	−1.237	−0.295
11754191_x_at	SRSF11	chr1	1.259	0.307	4.64E−05	0.073	0.012	1.70E−02	0.422	0.404	2.99E−01	63.1 %	0.882	0.066	1.697
11751862_a_at	KDM1B	chr6	−0.659	0.163	5.75E−05	−0.023	0.010	1.65E−01	−0.381	0.235	1.27E−01	0.0 %	−0.568	−0.831	−0.305
11758658_s_at	EPM2AIP1	chr3	1.496	0.368	5.84E−05	0.057	0.035	1.23E−01	0.579	0.422	1.85E−01	62.7 %	1.060	0.163	1.958
11736830_s_at	MAP3K7	chr6	1.548	0.385	6.43E−05	0.073	0.031	5.61E−02	0.848	0.478	7.96E−02	23.2 %	1.255	0.579	1.932
11719758_a_at	RP1-178F10.3	chr17	−0.491	0.123	6.88E−05	−0.041	0.048	9.58E−04	−0.257	0.180	1.75E−01	13.2 %	−0.410	−0.628	−0.193
11756273_a_at	RBM12B	chr8	0.725	0.183	7.58E−05	0.039	0.023	3.31E−02	0.318	0.210	1.31E−01	53.2 %	0.535	0.137	0.932
11755245_x_at	C3orf17	chr3	0.854	0.215	7.84E−05	0.059	0.016	6.30E−03	0.264	0.292	3.71E−01	62.2 %	0.592	0.017	1.166
11760412_a_at	EIF3K	chr19	−0.702	0.177	7.93E−05	−0.033	0.030	6.56E−02	−0.229	0.218	2.97E−01	64.8 %	−0.483	−0.945	−0.020
11727087_a_at	PTS	chr11	0.906	0.228	8.11E−05	−0.002	0.034	9.40E−01	0.285	0.343	4.06E−01	56.1 %	0.648	0.048	1.248
11744850_a_at	SSH2	chr17	−0.652	0.166	8.67E−05	−0.036	0.027	2.77E−02	−0.286	0.266	2.86E−01	26.3 %	−0.528	−0.867	−0.190

[a] Univariate NTR models adjusted for: plate, well, family ID, month of sampling, sex, age, smoking status, BMI, white blood cell counts, days between sampling and extraction, days between extraction and amplification, sampling time.

[b] Models additionally corrected for NO_2, NO_X, $PM_{2.5abs}$, PM_{10}, PM_{coarse}, UFP, and top10NL, NDVI, PLPP.

[c] Univariate NESDA models adjusted for: sex, age, smoking status, BMI, rbc, hgb, hct, sampling time, plate, well, month lab.

[d] Results from meta-analysis Univariate results NTR and NESDA.

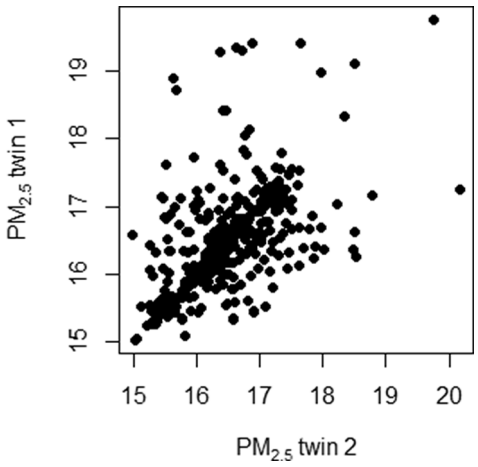

Fig. 2. Concordance in $PM_{2.5}$ exposure levels in monozygotic twin pairs (n = 421 pairs, Pearson correlation coefficient = 0.56).

'response to type I interferon' (biological process related to *immune response*), platelet activation (biological process related to *intercellular communication and mediation of inflammatory activities*), vesicle lumen (cellular component).

4. Discussion

We conducted a study in two large Dutch cohorts for which assessment of both air pollutants and the genome-wide transcriptome was harmonized. We observed a perturbation in peripheral blood gene expression associated with long-term average levels $PM_{2.5}$ and no associations with any of the other included air pollutants. Genes that were associated with $PM_{2.5}$ in both cohorts play a role in biological pathways related to cell signaling and immune response. All sixty-nine genes identified in this study have not previously been identified in one of the four existing studies of long-term exposure to air pollution and perturbations in the genome-wide transcriptome and are therefore novel findings.

Among the robustly associated genes, a group of genes involved in RNA binding was enriched. Interactions between air pollution and RNA binding have been reported before, for example in Baldridge et al. (2015) who demonstrated that cellular RNA is chemically modified by exposure to air pollution mixtures (Baldridge et al., 2015). Juan C. Gonzalez-Rivera et al. 2020 show that air pollution induces oxidative modifications such as 8-oxo-7,8-dihydroguanine in RNAs of lung cells, which could be associated with premature lung dysfunction (Gonzalez-Rivera et al., 2020). Kupsco et al. presented results in which total blood RNA m6A was positively associated with 8-hour black carbon exposure (Kupsco et al., 2020).

Uniprot keyword 'phosphoprotein' was enriched in our study as well. Phosphoproteins play a role in the respiratory chain. Our results are complementary to findings from Vrijens (Vrijens et al., 2017), Winckelmans (Winckelmans et al., 2017), and Rossner (Rossner et al., 2015), who identified several respiratory chain related pathways in relation to long-term exposure to particulate air pollution.

We observed a downregulation of the response to type 1 interferon in our study. Interferons play a key role in innate immunity (Bauer et al., 2012). As such our findings would point toward a potential role of $PM_{2.5}$ in dysregulation of the immune system and systemic inflammation. Similar findings were reported by Bauer et al., who demonstrated that the presence of inhaled PM in alveolar macrophages significantly decreases interleukin 1β and TNF-α production in bronchoalveolar cells in response to *M tuberculosis* infection as well as interferon-γ-production in peripheral blood mononuclear cells in response to purified protein derivative (Bauer et al., 2012), in a study by Lilian Calderón-Garcidueñas et al. where exposure to air pollution resulted in increased CD8+ T cells and mCD14+ monocytes and reduced numbers of natural killer cells, interferon gamma and granulocyte-macrophage colony-stimulating factor (GM-CSF) (Calderón-Garcidueñas et al., 2009), and in a study by Tao et al. who reported that exposure to $PM_{2.5}$ compromises antiviral immunity in influenza infection by inhibiting activation of NLRP3 inflammasome and expression of interferon-β (Tao et al., 2020).

We observed evidence for an impact of $PM_{2.5}$ on the downregulation of platelet activation. Platelet activation likely plays a role in air pollution induced cardiovascular disease (Bourdrel et al., 2017). In

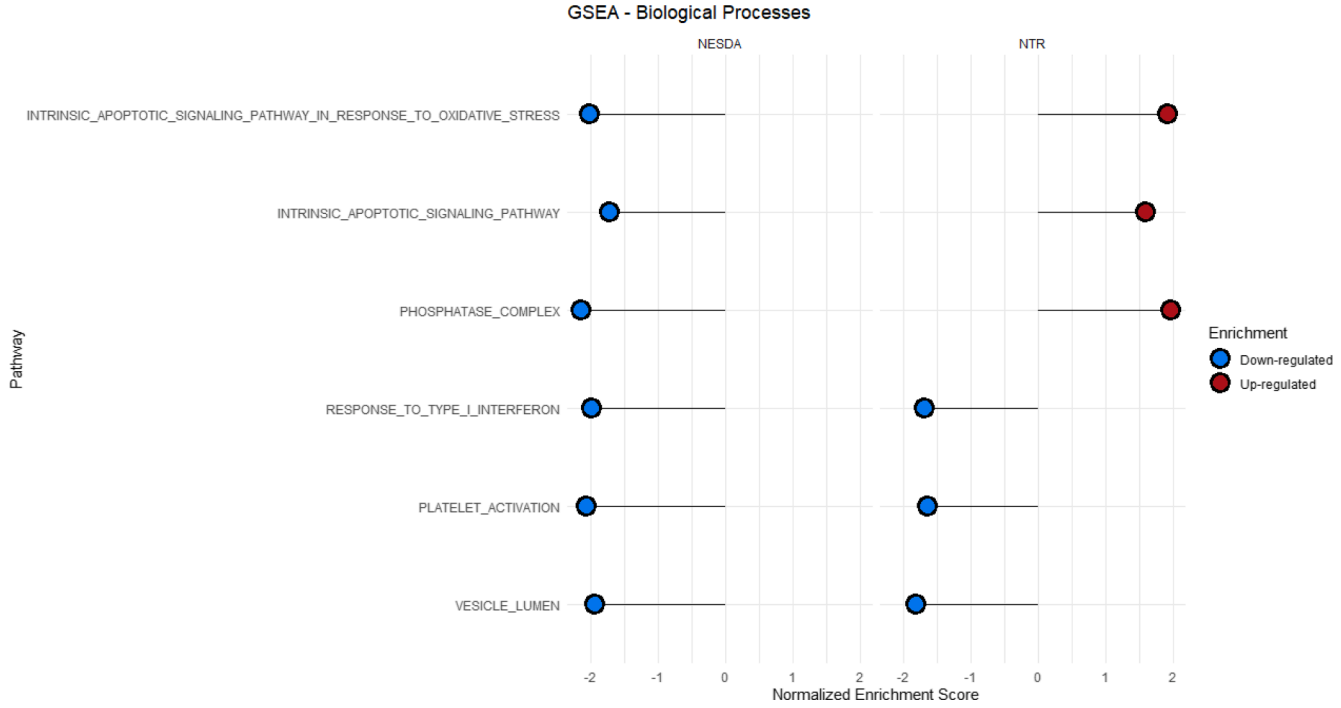

Fig. 3. Gene Ontology terms enriched (q < 0.05) in results observed in both the NTR and NESDA cohorts.

contrast to our findings, in the literature there is consistent signal for an increase in platelet activation and its role in enhanced thrombosis in response to exposure to air pollution (Robertson and Miller, 2018).

We observed a signal specific to $PM_{2.5}$ and only very weak indications for an association with any of the other pollutants that were included in the analysis. $PM_{2.5}$ has been identified as primary culprit in other epidemiological air pollution studies as well (Landrigan et al., 2017). The absence of a signal for air pollutants that are typically considered to be largely driven by traffic (ie. NO_2, BC, UFP) in our study would indicate the perturbations observed in our study are not primarily driven by traffic related air pollution, but might derive from other sources such as livestock, and industries (Strak et al., 2011). To further understand the impact of air pollution on perturbations in gene expression, a broader exposome approach is warranted where not only include physical-chemical exposures (such as air pollution) are included, but that also includes more extensive information on socio-economic (partly covered in this analysis through 'neighbourhood purchasing power'), dietary, 'built environment' factors as well (Vlaanderen et al., 2021). In the current analysis neighborhood purchasing power was very weakly correlated with $PM_{2.5}$ (Pearson correlation coefficient: 0.08) and did therefore not confound our results, yet more extensive characterization of socio economic status in follow-up analyses might provide additional insights.

4.1. Strengths and weaknesses

Our study has several strengths. This is the largest study to date to assess the association between air pollution and transcriptomic perturbations. We used state of the art exposure models for air pollutants and several key exposures that define the urban exposome. We were able to assess the replication of findings from our detection cohort (NTR) in a large replication cohort (NESDA). Assessment of exposures and transcriptomics was the same for both cohorts, while covariates included in the analyses were either the same or harmonized between the cohorts. The twin design of the NTR study has allowed us to assess the potential impact of confounding by genetic factors on our results. Our finding that $PM_{2.5}$ levels were moderately concordant within MZ twin pairs, indicates that both members of the twin pair resided at relatively similar locations with respect to their $PM_{2.5}$ exposure. This may have reduced the statistical power to detect whether the observed gene expression patterns were confounded by genetic factors.

Even though our study is the largest of its kind to date, rather modest signals were observed in both detection and replication cohort as well as in the MZ twin pair analysis. Several factors could have contributed to this observation. It is possible that the effect of air pollution on the peripheral blood transcriptome is of such subtle and transient nature that even larger studies or exposure contrasts are needed to further increase statistical evidence. A pan-European comparison in $PM_{2.5}$ concentration distributions indeed indicated that the exposure contrast within the Netherlands is rather modest compared to the contrast in exposure levels that can be observed across Europe (Eeftens et al., 2012). With regards to the likely transient nature of transcriptomic perturbations due to air pollution, the timing of blood sampling in relation to the timing of episodes of (high) exposure likely has an impact on the ability to detect associations. Further insight into kinetics and stability environmentally induced transcriptomic perturbations are needed, which require study designs that incorporate repeated blood sampling as well as detailed characterization of patterns in environmental exposure. Methods that allow for more personalized exposure estimates, such as those that incorporate insights from personal monitoring or time-activity data will contribute to improved characterization of environmental exposures, including air pollution (Vlaanderen et al., 2021). In this study we applied predictions from a land-use regression model to a time period (2004–2008) before the period of air quality monitoring used for model development (2008–2011). Studies have documented that land use regression models (as applied in our study) can be utilized successfully to estimate air pollution concentrations several years forwards or backwards in time (Wang et al., 2013). We therefore assume minimal impact of this extrapolation on the degree of air pollutant measurement error in our study.

While multicollinearity of the seven air pollutants included in this study is a potential issue, $PM_{2.5}$, the only pollutant for which we observed an effect in the univariate analysis, was only moderately correlated with the other pollutants. As such, the impact of multicollinearity on the effects estimates for $PM_{2.5}$ from the multiple regression models which included all seven air pollutants was limited. Other avenues to improve the power to detect associations between the air pollution and transcriptomic perturbations would include scaling up to cross-continental analyses. In comparison to our study, such analyses would yield larger contrasts in environmental exposures including $PM_{2.5}$ (Eeftens et al., 2012). Yet, comparisons between countries are often complicated by the large differences in culture, diet, environment, etc., and would require harmonization of the available data. Methodological improvements in terms of statistical methods (e.g. moving toward machine learning approaches better capable of handling OMICs data, especially by integrating signals from multiple OMICs layers (Merid, 2021; Mostafavi, 2018), and better integration of existing insights from toxicology and mechanistic studies in the epidemiological analysis (Scholten et al., 2021) will likely yield further insights into the impact of the environment on gene expression perturbations.

To conclude, our results indicate that $PM_{2.5}$ can elicit a response in cell signaling and the immune system, both hallmarks of environmental insults (Peters et al., 2021). Next steps in this line of research include replication of our results in large cohorts with state-of the art assessment of air pollution and other aspects in the exposome and additional mechanistic research to assess the implications of our findings.

CRediT authorship contribution statement

Jelle Vlaanderen: Conceptualization, Methodology, Formal analysis, Data curation, Writing – original draft. **Roel Vermeulen:** Conceptualization, Methodology, Writing – review & editing. **Matthew Whitaker:** Methodology, Formal analysis, Writing – review & editing. **Marc Chadeau-Hyam:** Methodology, Writing – review & editing. **Jouke Jan Hottenga:** Writing – review & editing, Resources. **Eco de Geus:** Writing – review & editing, Resources. **Gonneke Willemsen:** Writing – review & editing, Resources. **Brenda W.J.H. Penninx:** Writing – review & editing, Resources. **Rick Jansen:** Writing – review & editing, Resources. **Dorret I. Boomsma:** Conceptualization, Writing – review & editing, Resources.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgements

Funding: This work was supported by the European Union's Horizon 2020 research and innovation programme under grant agreement No 874627 (EXPANSE) and by the US National Institute of Mental Health (RC2 MH089951, PI Sullivan) as part of the American Recovery and Reinvestment Act of 2009. The Netherlands Study of Depression and Anxiety (NESDA) and the Netherlands Twin Register (NTR) acknowledge funding by the Netherlands Organization for Scientific Research (MagW/ZonMW grants 904-61-090, 985-10-002,904-61-193,480-04-004, 400-05-717, 912-100-20; Spinozapremie 56-464-14192;

Geestkracht program grant 10-000-1002); Biobanking and Biomolecular Resources Research Infrastructure (BBMRI-NL); the European Science Foundation (EU/QLRT-2001-01254); the European Community's Seventh Framework Program (FP7/2007-2013); ENGAGE (HEALTH-F4-2007-201413); and the European Research Council (ERC, 230374).

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.envint.2022.107491.

References

Wright, F., et al., 2014. Heritability and genomics of gene expression in peripheral blood. Nat Genet 46 (5), 430–437.

Baldridge, K.C., Zavala, J., Surratt, J., Sexton, K.G., Contreras, L.M., 2015 Jan. Cellular RNA is chemically modified by exposure to air pollution mixtures. Inhal Toxicol. 27 (1), 74–82. https://doi.org/10.3109/08958378.2014.987361. PMID: 25600141.

Beelen, R., et al., 2013. Development of NO2 and NOx land use regression models for estimating air pollution exposure in 36 study areas in Europe – The ESCAPE project. Atmos. Environ. 72, 10–23.

Bourdrel, T., Bind, M.A., Béjot, Y., Morel, O., Argacha, J.F., 2017. Cardiovascular effects of air pollution. Arch. Cardiovasc. Dis. 110, 634–642.

Brook, R., Rajagopalan, S., Pope, C., Circulation, J.B., 2010. undefined. Particulate matter air pollution and cardiovascular disease: an update to the scientific statement from the American Heart Association. Am Hear. Assoc.

Calderón-Garcidueñas, L., et al., 2009. Immunotoxicity and environment: Immunodysregulation and systemic inflammation in children. Toxicol. Pathol. 37, 161–169.

Bauer, R.N., Diaz-Sanchez, D., Jaspers, I., 2012 Jan. Effects of air pollutants on innate immunity: the role of Toll-like receptors and nucleotide-binding oligomerization domain-like receptors. J Allergy Clin Immunol. 129 (1), 14–24. https://doi.org/10.1016/j.jaci.2011.11.004 quiz 25-6. PMID: 22196521; PMCID: PMC4341993.

Croft, D., Burton, D., Nagel, D., Bhattacharya, S., Falsey, A., 2021. The Effect of Air Pollution on the Transcriptomics of the Immune Response to Respiratory Infection. doi: 10.21203/rs.3.rs-440415/v1.

Douglas, B., Maechler, M., Bolker, B., Walker, S., 2014. lme4: Linear mixed-effects models using Eigen and S4. *R Packag. version.*

Eeftens, M., et al., 2012. Spatial variation of PM2.5, PM10, PM2.5 absorbance and PMcoarse concentrations between and within 20 European study areas and the relationship with NO2 – Results of the ESCAPE project. Atmos. Environ. 62, 303–317.

Eeftens, M., et al., 2012. Development of Land Use Regression models for PM(2.5), PM (2.5) absorbance, PM(10) and PM(coarse) in 20 European study areas; results of the ESCAPE project. Environ. Sci. Technol. 46, 11195–11205.

Gonzalez-Rivera, J.C., et al., 2020. Post-transcriptional air pollution oxidation to the cholesterol biosynthesis pathway promotes pulmonary stress phenotypes. Commun. Biol. 3.

Health Effects Institute, 2018. *State of Global Air 2018. Special Report.*

Huan, T., et al., 2016. A whole-blood transcriptome meta-analysis identifies gene expression signatures of cigarette smoking. Hum. Mol. Genet. 25, 4611–4623.

Huang, Y.-C.-T., et al., 2010. Gene expression profile in circulating mononuclear cells after exposure to ultrafine carbon particles. Inhal. Toxicol. 22, 835–846.

Huang, D.W., Sherman, B.T., Lempicki, R.A., 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. 4, 44–57.

Jansen, R., et al., 2014. Sex differences in the human peripheral blood transcriptome. BMC Genomics 15.

Kadaster, C. Bestand Bodemgebruik. https://www.kadaster.com/automatic-generalisation.

Korotkevich, G., Sukhov, V., Sergushichev, A. Fast gene set enrichment analysis. *biorxiv.org.* doi: 10.1101/060012.

Kupsco, A., et al., 2020. Associations of smoking and air pollution with peripheral blood RNA N6-methyladenosine in the Beijing truck driver air pollution study. Environ. Int. 144, 106021.

Landrigan, P. et al., 2017. The Lancet Commission on pollution and health. *thelancet.com.* doi: 10.1016/S0140-6736(17)32345-0.

Liu, S., et al., 2021. Long-term exposure to low-level air pollution and incidence of asthma: the ELAPSE project. Eur. Respir. J. 57, 2003099.

Merid, S.K., et al., 2021. Integration of gene expression and DNA methylation identifies epigenetically controlled modules related to PM2.5 exposure. Environ. Int. 146, 106248.

Mostafavi, N., et al., 2017. Associations between genome-wide gene expression and ambient nitrogen oxides (NOx). Epidemiology. https://doi.org/10.1097/EDE.0000000000000628.

Mostafavi, N., et al., 2018. Acute changes in DNA methylation in relation to 24hr personal air pollution exposure measurements: A panel study in four European countries. Environ Int 120, 11–21.

O'Beirne, S.L., et al., 2018. Ambient pollution-related reprogramming of the human small airway epithelial transcriptome. Am. J. Respir. Crit. Care Med. 198, 1413–1422.

Ouwens, K.G., et al., 2020. A characterization of cis- and trans-heritability of RNA-Seq-based gene expression. Eur. J. Hum. Genet. 28, 253–263.

Penninx, B.W.J.H., et al., 2008. The Netherlands Study of Depression and Anxiety (NESDA): Rationale, objectives and methods. Int. J. Methods Psychiatr. Res. 17, 121–140.

Peretz, A., et al., 2007. Diesel exhaust inhalation and assessment of peripheral blood mononuclear cell gene transcription effects: an exploratory study of healthy human volunteers. Inhal. Toxicol. 19, 1107–1119.

Peters, A., Nawrot, T.S., Baccarelli, A.A., 2021. Hallmarks of environmental insults. Cell 184, 1455–1468.

Pettit, A.P., et al., 2012. Alteration of peripheral blood monocyte gene expression in humans following diesel exhaust inhalation. Inhal. Toxicol. 24, 172–181.

Robertson, S., Miller, M.R., 2018. Ambient air pollution and thrombosis. Part. Fibre Toxicol. 15.

Mohr, S., Liew, C.C., 2007 Oct. The peripheral-blood transcriptome: new insights into disease and risk assessment. Trends Mol Med. 13 (10), 422–432. https://doi.org/10.1016/j.molmed.2007.08.003. Epub 2007 Oct 4 PMID: 17919976.

Rossner Jr, P., Tulupova, E., Rossnerova, A., Libalova, H., Honkova, K., Gmuender, H., Pastorkova, A., Svecova, V., Topinka, J., Sram, R.J., 2015 Oct. Reduced gene expression levels after chronic exposure to high concentrations of air pollutants. Mutat Res. 780, 60–70. https://doi.org/10.1016/j.mrfmmm.2015.08.001. Epub 2015 Aug 11 PMID: 26298100.

Scholten, B., et al., 2021. Estimation of the exposure response relation between benzene and acute myeloid leukemia by combining epidemiological, human biomarker, and animal data. AACR. https://doi.org/10.1158/1055-9965.EPI-21-0287.

Spijker, S., Van De Leemput, J.C.H., Hoekstra, C., Boomsma, D.I., Smit, A.B., 2004. Profiling gene expression in whole blood samples following an in-vitro challenge. Twin Res. 7, 564–570.

Strak, M., et al., 2011. Variation in characteristics of ambient particulate matter at eight locations in the Netherlands – The RAPTES project. Atmos. Environ. 45, 4442–4453.

Strak, M., et al., 2021. Long term exposure to low level air pollution and mortality in eight European cohorts within the ELAPSE project: Pooled analysis. BMJ 374.

Tao, R., et al., 2020. PM2.5 compromises antiviral immunity in influenza infection by inhibiting activation of NLRP3 inflammasome and expression of interferon-β. Mol. Immunol. 125, 178–186.

Van Nunen, E., et al., 2017. Land Use Regression Models for Ultrafine Particles in Six European Areas. Environ. Sci. Technol. 51.

Vink, J.M., et al., 2017. Differential gene expression patterns between smokers and non-smokers: cause or consequence? Addict. Biol. 22, 550–560.

Vlaanderen, J., et al., 2021. Developing the building blocks to elucidate the impact of the urban exposome on cardiometabolic-pulmonary disease: The EU EXPANSE project. Environ. Epidemiol. 5.

Vrijens, K., et al., 2017. 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. 125, 660–669.

Wang, R., Henderson, S.B., Sbihi, H., Allen, R.W., Brauer, M., 2013. Temporal stability of land use regression models for traffic-related air pollution. Atmos. Environ. 64, 312–319.

Weier, J., Herring, D., 2000. Measuring Vegetation (NDVI & EVI) Normalized Difference Vegetation Index (NDVI). NASA Earth Observatory.

Willemsen, G., et al., 2010. The Netherlands Twin Register Biobank: A Resource for Genetic Epidemiological Studies. Twin Res. Hum. Genet. 13, 231–245.

Willemsen, G., et al., 2010. The Netherlands twin register biobank: A resource for genetic epidemiological studies. Twin Res. Hum. Genet. 13, 231–245.

Willemsen, G., et al., 2013. The Adult Netherlands Twin Register: Twenty-Five Years of Survey and Biological Data Collection. Twin Res. Hum. Genet. 16, 271–281.

Winckelmans, E., et al., 2017. Transcriptome-wide analyses indicate mitochondrial responses to particulate air pollution exposure. Environ. Heal. A Glob. Access Sci. Source 16, 1–15.

Wittkopp, S., Staimer, N., Tjoa, T., et al., 2016 Mar- Apr. Nrf2-related gene expression and exposure to traffic-related air pollution in elderly subjects with cardiovascular disease: An exploratory panel study. J Expo Sci Environ Epidemiol. 26 (2), 141–149.