DOI: http://doi.org/10.1016/j.envint.2020.106248

Integration of gene expression and DNA methylation identifies epigenetically controlled modules related to $PM_{2.5}$ exposure

Simon Kebede Merid[1], Mariona Bustamante[2,3,4], Marie Standl[5], Jordi Sunyer [2,3,4,6], Joachim Heinrich[7,8], Nathanaël Lemonnier[9], Daniel Aguilar[10], Josep Maria Antó[2,3,4,6], Jean Bousquet[11,12,13,] Loreto Santa-Marina[14,15,16], Aitana Lertxundi[14,15,17], Anna Bergström[18,19], Inger Kull[1,20], Åsa M. Wheelock[21], Gerard H. Koppelman[22,23], Erik Melén[1,20] and Olena Gruzieva*[18,19]*

[1] Department of Clinical Sciences and Education, Karolinska Institutet, Södersjukhuset, Stockholm, Sweden

[2] ISGlobal, Institute for Global Health, Barcelona, Spain

[3] Universitat Pompeu Fabra (UPF), Barcelona, Spain

[4] Spanish Consortium for Research on Epidemiology and Public Health (CIBERESP), Madrid, Spain

[5] Institute of Epidemiology, Helmholtz Zentrum München - German Research Center for Environmental Health, Ingolstädter Landstraße 1, 85764, Neuherberg, Germany

[6] IMIM (Hospital del Mar Medical Research Institute), Barcelona, Spain

[7] Institute and Clinic for Occupational, Social and Environmental Medicine, University Hospital, LMU Munich, Ziemssenstraße 1, 80336, Munich, Germany

[8] Allergy and Lung Health Unit, Melbourne School of Population and Global Health, The University of Melbourne, Melbourne, Australia

[9] Institute for Advanced Biosciences, UGA-INSERM U1209-CNRS UMR5309, Allée des Alpes, France

[10] Biomedical Research Networking Center in Hepatic and Digestive Diseases (CIBEREHD), Instituto de Salud Carlos III, Barcelona, Spain

[11]Charité, Universitätsmedizin Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Comprehensive Allergy Center, Department of Dermatology and Allergy, Berlin, Germany

[12]University Hospital, Montpellier, France

[13]MACVIA-France, Montpellier, France

[14] Health Research Institute-BIODONOSTIA, Basque Country, Spain

[15] CIBER Epidemiología y Salud Pública (CIBERESP), Spain

[16] Health Department of Basque Government, Sub-directorate of Public Health of Gipuzkoa, 20013 San Sebastian, Spain

[17] Preventive Medicine and Public Health Department, University of Basque Country (UPV/EHU), Spain

[18] Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden

[19] Centre for Occupational and Environmental Medicine, Region Stockholm, Sweden

[20] Sachs Children's Hospital, Stockholm, Sweden

[21] Respiratory Medicine Unit, Department of Medicine and Center for Molecular Medicine, Karolinska Institutet, Solna, Stockholm, Sweden.

[22] University of Groningen, University Medical Center Groningen, Beatrix Children's Hospital, Department of Pediatric Pulmonology and Pediatric Allergology, Groningen, the Netherlands

[23] University of Groningen, University Medical Center Groningen, Groningen Research Institute for Asthma and COPD (GRIAC), Groningen, the Netherlands

* Shared last authors

Corresponding author:

Dr Olena Gruzieva

Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden

E-mail address: olena.gruzieva@ki.se

Competing interests

None

Acknowledgments

BAMSE: We would like to thank the families who participated in the BAMSE study and the staff for their hard work and effort. We additionally acknowledge Alexandra Ek, Sandra Ekström, André Lauber and Niklas Andersson for excellent data management.

GINIplus: The authors thank all the families for their participation in the GINIplus study. Furthermore, we thank all members of the GINIplus Study Group for their excellent work. The GINIplus Study group consists of the following: Institute of Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg (Heinrich J, Brüske I, Schulz H, Flexeder C, Zeller C, Standl M, Schnappinger M, Ferland M, Thiering E, Tiesler C); Department of Pediatrics, Marien-Hospital, Wesel (Berdel D, von Berg A); Ludwig-Maximilians-University of Munich, Dr von Hauner Children's Hospital (Koletzko S); Child and Adolescent Medicine, University Hospital rechts der Isar of the Technical University Munich (Bauer CP, Hoffmann U); IUF- Environmental Health Research Institute, Düsseldorf (Schikowski T, Link E, Klümper C, Krämer U, Sugiri D).

INMA: The authors thank all the families for their participation in the INMA project. A full list of INMA researchers can be found at http://www.proyectoinma.org/presentacion-inma/listado-investigadores/listado-investigadores.html.

Funding

The research leading to these results has received funding from the European Community's Seventh Framework Program under grant agreement numbers: 211250 (ESCAPE), and 261357 (MeDALL), and the European Research Council (grant agreement number 757919, TRIBAL). NL was recipient of ANR-15- IDEX-02 grant.

BAMSE: BAMSE was supported by The Swedish Research Council, The Swedish Heart-Lung Foundation, Region Stockholm, the Strategic Research Programme (SFO) in Epidemiology at

Karolinska Institutet, The Swedish Research Council Formas and the Swedish Environment Protection Agency, the Swedish Asthma and Allergy Research Foundation, the Cancer and Allergy Foundation. O. Gruzieva is supported by the Swedish Research Council for Health, Working life and Welfare (FORTE 2017-01146).

GINIplus: The GINIplus study was mainly supported for the first 3 years of the Federal Ministry for Education, Science, Research and Technology (interventional arm) and Helmholtz Zentrum Munich (former GSF) (observational arm). The 4 year, 6 year, 10 year and 15 year follow-up examinations of the GINIplus study were covered from the respective budgets of the 5 study centres (Helmholtz Zentrum Munich (former GSF), Research Institute at Marien-Hospital Wesel, LMU Munich, TU Munich and from 6 years onwards also from IUF - Leibniz Research-Institute for Environmental Medicine at the University of Düsseldorf) and a grant from the Federal Ministry for Environment (IUF Düsseldorf, FKZ 20462296). Further, the 15-year follow-up examination of the GINIplus study was supported by the Commission of the European Communities, the 7th Framework Program: MeDALL project, and as well by the companies Mead Johnson and Nestlé.

INMA: INMA project was mainly supported by the Commission of the European Community, the 7th Framework Program: MeDALL project and by grants from Instituto de Salud Carlos III (Red INMA G03/176; CB06/02/0041; PI04/1436; PI06/0867; PI08/1151; and PI18/01142 incl. FEDER funds), and Generalitat de Catalunya-CIRIT 1999SGR 00241, Department of Health of the Basque Government (2005111093), Provincial Government of Gipuzkoa (DFG06/002), and annual agreements with the municipalities of the study area (Zumarraga, Urretxu , Legazpi, Azkoitia y Azpeitia y Beasain).

Abstract

Air pollution has been associated with adverse health effects across the life-course. Although underlying mechanisms are unclear, several studies suggested pollutant-induced changes in transcriptomic profiles. In this meta-analysis of transcriptome-wide association studies of 656 children and adolescents from three European cohorts participating in the MeDALL Consortium, we found two differentially expressed transcript clusters (FDR $p<0.05$) associated with exposure to particulate matter < 2.5 micrometers in diameter ($PM_{2.5}$) at birth, one of them mapping to the *MIR1296* gene. Further, by integrating gene expression with DNA methylation using Functional Epigenetic Modules algorithms, we identified 9 and 6 modules in relation to $PM_{2.5}$ exposure at birth and at current address, respectively (including *NR1I2, MAPK6, TAF8* and *SCARA3).* In conclusion, $PM_{2.5}$ exposure at birth was linked to differential gene expression in children and adolescents. Importantly, we identified several significant interactome hotspots of gene modules of relevance for complex diseases in relation to $PM_{2.5}$ exposure.

Key words: air pollution, DNA methylation, gene expression, integration, children

1. Introduction

Air pollution exposure at birth has been associated with different types of health effects, such as adverse pregnancy outcomes (Pedersen et al. 2013), childhood airway disease (Gehring et al. 2015), and neurodevelopmental disorders (Sram et al. 2017). Although the precise mechanisms responsible for these health effects are unclear, several studies have suggested oxidative stress and systemic inflammation as potential intermediate biological responses to air pollutants (Kelly and Fussell 2015). Accumulating evidence suggests that these acute systemic effects of long-term exposure to air pollution can be detected by assessing genome-wide gene expression profiles in peripheral blood cells (Mostafavi et al. 2017).

Exposure to air pollutants has been shown to induce changes in gene expression in animal and *in vitro* experiments (Bhetraratana et al. 2019; Kim et al. 2019; Zhu et al. 2019), but evidence from human studies is scarce. Short-term inhalation studies reported changes in the expression of genes involved in inflammation, tissue growth and host defense against environmental insults, including IGF-1 signaling, insulin receptor signaling and NRF2-mediated oxidative stress response pathway in blood (Huang et al. 2010; Peretz et al. 2007), as well as genes associated with bronchial immune responses in bronchoalveolar lavages, protein degradation, and coagulation (e.g. *PLAU, F2R, CBL, UBR1*) (Pettit et al. 2012) in response to exposure to diesel exhaust. A genome-wide gene expression microarray analysis of 63 non-smoking employees at 10 trucking terminals in the northeastern US identified a set of genes implicated in ischemic heart disease, chronic obstructive pulmonary disease (COPD), lung cancer, and other pollution-related illnesses (Chu et al. 2016). Another study based on 550 healthy subjects participating in cohorts from Italy and Sweden has shown differential gene expression related to long-term exposure to nitrogen oxides (NO_x) similar to tobacco induced changes in the transcriptome (Mostafavi et al. 2017). In a study of school age children inhabiting a severely polluted area from the Czech Republic, numerous genes were found to

153 have their expression in blood relatively increased when compared with children from a rural
154 less polluted area (van Leeuwen et al. 2006). In other studies, gene-environment interactions
155 with respect to air pollution exposure and expression levels have been reported (Fave et al.
156 2018; Gref et al. 2017).
157 To our knowledge, no study has evaluated expression levels across the genome in children
158 and adolescents in relation to individual ambient air pollution exposure. Epigenomic changes
159 related to air pollution are, on the other hand, relatively well studied (Gruzieva et al. 2017;
160 Gruzieva et al. 2019). Although several differentially methylated CpGs and regions have been
161 identified in relation to environmental factors, most studies found rather weak associations
162 with ambient air pollution exposure (Gruzieva et al. 2017; Gruzieva et al. 2019; Lee et al.
163 2019; Plusquin et al. 2017). Integrating different types of omics data may shed new light on
164 gene modules or molecular pathways, which play key roles in cellular responses and
165 subsequent disease (Jiao et al. 2014; Tian et al. 2017). In addition, it has been suggested that
166 omics integration may enhance study power with increased likelihood to identify
167 biologically relevant mechanisms and similarity patterns between groups of subjects (Li et
168 al. 2018; Wang et al. 2014). For example, in a recent integrative study of methylome and
169 transcriptome of human cardiomyocytes, multiple altered methylome and transcriptome
170 signatures in the cardiac disease-specific genes, following exposure to particulate matter
171 less than 2.5 micrometers in diameter ($PM_{2.5}$), have been reported (Yang et al. 2018).
172 Human studies of environmental exposures combining different -omics data are, however,
173 scarce (Vargas et al. 2018).
174 In the present study, we explored the association of gene expression profiles with $PM_{2.5}$
175 exposure at birth and at the time of blood sampling (i.e. pre-school age or adolescence).
176 Moreover, we performed an integrative analysis of gene expression and DNA methylation
177 data to identify molecular pathways that are epigenetically and functionally affected by $PM_{2.5}$

exposure. Within the framework of the European collaborative Mechanisms of the Development of ALLergy (MeDALL) project (Bousquet et al. 2011), we used harmonized data of three European birth cohort studies for which a standardized assessment of air pollution exposure for $PM_{2.5}$ was available.

2. Methods

2.1 Study population

Data from three European birth cohort studies from Sweden (BAMSE (Schultz et al. 2016)); Germany (GINIplus (Heinrich et al. 2017)); and Spain (INMA (Guxens et al. 2011)), participating in the MeDALL project (Bousquet et al. 2011) were included in the present analysis (n total = 656). All cohorts acquired ethics approval and informed consent from participants prior to data collection through local ethics committees.

2.2 Air pollution exposure assessment

In the MeDALL cohorts, annual average concentrations of $PM_{2.5}$ were estimated at home addresses at birth and at the time of bio-sampling (i.e. pre-school age in INMA and adolescence in BAMSE and GINIplus) through land-use regression (LUR) models developed for each study area within the European Study of Cohorts for Air Pollution Effects (ESCAPE) project (Eeftens et al. 2012). Further details about exposure assessment are provided in the online supplement.

2.3 Gene expression measurements

Expression levels were measured in BAMSE (244 individuals of mean age 16.69 years), GINIplus Munich (247 individuals aged 15.2), and INMA (165 individuals aged 4.48) cohorts with the Affymetrix Human Transcriptome Array 2.0 (HTA) (Lemonnier et al. 2020). Whole

blood was collected from study subjects in PAXgene tubes, and RNA was extracted batch-wise (QIAGEN, Courtaboeuf, France). RNA yield and quality were assessed with state-of-the-art spectrophotometry and lab-on-a-chip microfluidic technologies, respectively. RNA of highest quality was selected for amplification, labelling, and hybridization on HTA with WT PLUS kit (Affymetrix Inc.) at the CNRS USR 3010 unit in Lyon. Gene expression levels were normalized with the Robust Multi-array Average (RMA) algorithm including quantile normalization (Irizarry et al. 2003), and version 36 of NetAffx annotation was used to annotate the 67,528 transcript clusters (probes covering a region of the genome reflecting all the exonic transcription evidence known for the region and corresponding to a known or putative gene, including coding and non-coding genes). The empirical Bayes method was applied for batch correction on the main three production phases using ComBat from the sva package in R (Johnson et al. 2007). More detailed information about the gene expression sample assessment can be found elsewhere (Lemonnier et al. 2020). The estimated cell type proportions were calculated from gene expression data using CIBERSORT (http://cibersort.stanford.edu/) (Newman et al. 2015).

2.4 DNA methylation measurements

Epigenome-wide DNA methylation was measured using DNA extracted from whole blood samples collected at the age of 16 years in the BAMSE (n=262) and at 4 years in the INMA (n=201) cohorts (Xu et al. 2018) processed with the Illumina Infinium HumanMethylation450 BeadChip (Illumina Inc., San Diego, USA). More information is provided in the online supplement.

2.5 Statistical analyses

2.5.1 Genome-wide gene expression analyses

We examined the associations between exposure to air pollutants and gene expression levels across the genome by means of linear regression analysis using the limma package in R (Ritchie et al. 2015). Air pollution concentrations were entered as continuous variables without transformation. All analyses were adjusted for sex, age (years), study region (where applicable), maternal education, maternal smoking during pregnancy, second-hand smoking, active smoking status (in adolescence), BMI (kg/m^2), physical activity at the time of bio-sampling, season of blood sampling and cell type. Cohort-specific results of the TWAS were subsequently included in a fixed-effect meta-analysis by combining study-specific weights based on the inverse of the variance. We adjusted for multiple testing using the Benjamini & Hochberg false discovery rate (FDR) correction (Benjamini and Hochberg 1995). The results below FDR P<0.05 threshold were labeled as statistically significant. The results are presented per 5 μg/m^3 increase in $PM_{2.5}$. A pearson correlation analysis was performed between $PM_{2.5}$ at birth and at the time of biosampling for each cohort. All cohort-specific statistical analyses were performed using R (R Development Core Team 2019) and Bioconductor packages (Gentleman et al. 2004), and the meta-analysis was performed using METAL software (Willer et al. 2010).

2.5.2 Integration of DNA methylation and gene expression results

An integration analysis of genome-wide DNA methylation with matched gene expression data available in BAMSE and INMA was performed on 240 and 103 subjects, respectively. First, we combined cohort-specific associations of air pollution exposure with genome-wide DNA methylation in the BAMSE and INMA cohorts (no DNA methylation data were available in GINIplus) after adjustment for potential covariates and cell proportions, in a fixed-effect meta-analysis. In the case of Illumina 450k data, we assigned to a gene the average value of probes mapping to within 200bp of the TSS. If no probes mapped to within 200bp of the TSS,

we used the average of probes mapping to the 1st exon of the gene. If such probes were not present, we used the average of probes mapping to within 1500bp of the TSS. Transcript clusters were mapped to genes as described above. The integration of the genes identified in the TWAS and EWAS meta-analysis, along with protein–protein interaction (PPI) was performed using the Functional Epigenetic Module (FEM) algorithm in R package (Jiao et al. 2014). The FEM algorithm first construct an integrated network with weights on the network edges from the associations between $PM_{2.5}$ and both gene expression and DNA methylation, and then inference of the FEMs as heavy subgraphs on this weighted network (Jiao et al. 2014). The integration of DNA methylation and gene expression profiling was performed after constructing the PPI network for hub gene identification from the High-quality INTeractomes (HINT) database (http://hint.yulab.org/) (Das and Yu 2012). HINT is a database of high-quality PPIs from 8 interactome resources (BioGRID, MINT, iRefWeb, DIP, IntAct, HPRD, MIPS and the PDB) consisting of two types of interactions: binary physical interactions and co-complex associations from different organisms. In total, 150,199 interactions were obtained after removing duplicates and self-linked interactions.

2.5.3 Functional analyses

Kyoto Encyclopaedia of Genes and Genomes (KEGG) (Kanehisa et al. 2019) pathway enrichment analysis was performed using network-based pathway annotation tool BinoX (Ogris et al. 2017) in PathwAX II web server (http://pathwax.sbc.su.se/) (Ogris et al. 2016). The algorithm assesses the statistical significance of 'pathway gene-set' enrichment by evaluating the amount of interactions between genes within a genome wide functional association network. P-values for enrichment were adjusted for multiple testing using the FDR method. Pathways with an estimated FDR P-value below 0.05 were selected as significantly enriched. In addition, to understand the functional involvement of $PM_{2.5}$

exposure in gene expression we performed the KEGG pathway analyses by applying a less stringent cut off P-value < 0.05 and |logFC|> 0.15.

2.5.4 Sensitivity analyses

To address potential age-specific effects, we performed a sensitivity analysis limiting our transcriptome-wide meta-analysis to two adolescent cohorts. Further, to rule out residual confounding effects from active smoking, we performed a sensitivity analysis in the two adolescent cohorts excluding active smokers and compared those with the results based on the whole sample.

3. Results

In total, 656 pre-school children (INMA) and adolescents (GINIplus and BAMSE) were included in the meta-analysis of genome-wide gene expression in relation to birth and current $PM_{2.5}$ exposures. The characteristics of the study subjects are presented in Table 1. Noteworthy, in the subset of children with gene expression measurements in the INMA cohort, the proportion of children exposed to tobacco smoke prenatally, as well as at the time of blood sampling, was higher compared to the other cohorts. Exposure levels, illustrated by box plots in Fig.1, were on average lowest for BAMSE (mean $PM_{2.5}$ exposure at birth and at current address: 7.6 and 7.3 $\mu g/m^3$, respectively), and highest for INMA (14.6 and 14.0 $\mu g/m^3$, respectively). The correlation between birth and current exposure levels was r=0.45 (p=8.46E-14), r=0.32 (p=9.08E-08), r=0.46 (p=2.01E-10) in BAMSE, GINIplus and INMA, respectively.

303 **Table 1. Characteristics of the study participants from the included cohorts.**

Characteristics	BAMSE (n=244)	GINIplus (n=247)	INMA (n=165)
Males, N (%)	136 (55.7%)	120 (48.6%)	87 (52.7%)
Age at the follow-up, yrs mean (SD)	16.7 (0.3)	15.2 (0.2)	4.5 (0.2)
Study region, N (%):	Stockholm 56 (22.9%) Järfälla 92 (37.7%) Solna 63 (25.8%) Sundbyberg 33 (13.5%)	Munich (100%)	Gipuzkoa 57 (34.5%) Sabadell 108 (65.6%)
Maternal education,[a] N (%):			
Low	79 (32.4%)	14 (5.7%)	39 (23.6%)
Medium	73 (29.9%)	87 (35.2%)	61 (36.9%)
High	92 (37.7%)	146 (59.1%)	65 (39.4%)
Season of blood sampling, N (%):			
Winter	48 (19.7%)	32 (12.7%)	28 (17.0%)
Spring	86 (35.3%)	55 (22.3%)	61 (37.0%)
Summer	42 (17.2%)	79 (32.0%)	38 (23.0%)
Autumn	68 (27.9%)	81 (33.0%)	38 (23.0%)
Physical activity, N (%):			
< 5 hours per week	120 (49.2%)	137 (55.5%)	74 (44.85%)
≥ 5 hours per week	124 (50.8%)	110 (44.5%)	91 (55.15%)
Maternal smoking during pregnancy, N (%)	24 (9.8%)	23 (9.3%)	46 (27.9%)
Second-hand smoking at the time of blood sampling, N (%)	31 (12.7%)	26 (10.5%)	84 (50.9%)
Active smoking, N (%)	25 (10.2%)	20 (8.1%)	N/A

304 N/A=Not applicable

305 [a] Low (Elementary school or 2-year high school), Medium (3-year High school) and High
306 (University).

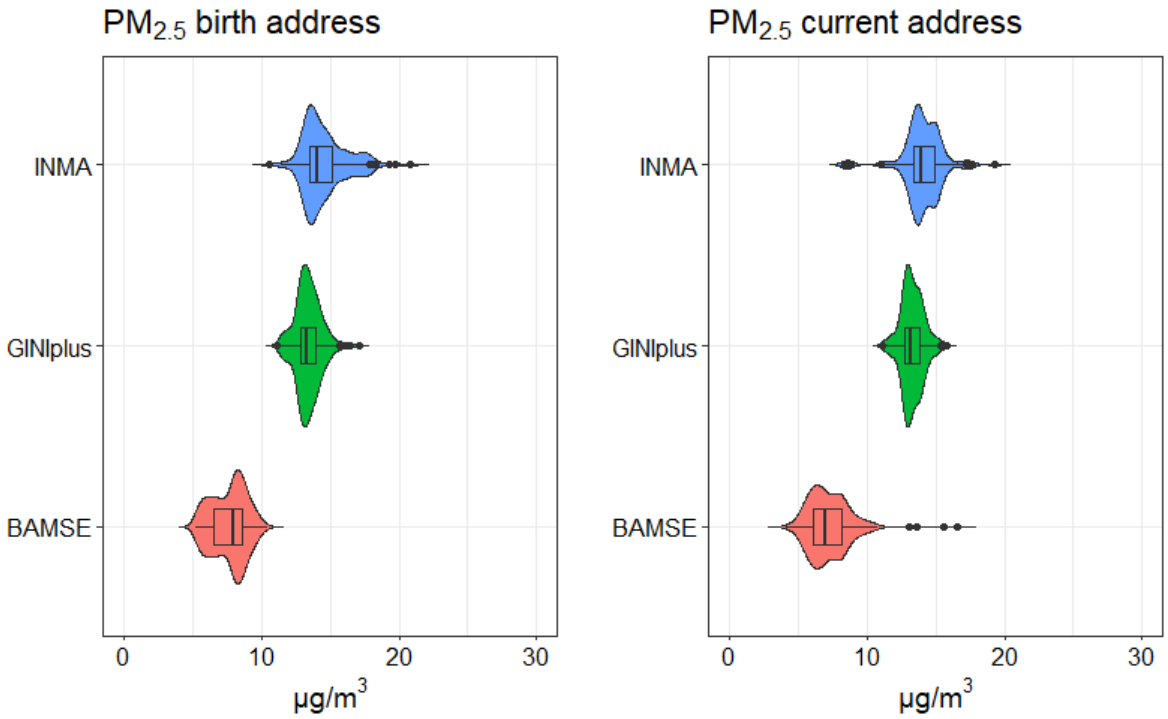

Fig. 1. Exposure for particulate matter ($PM_{2.5}$) at the birth and at the current address at the time of blood sampling, µg/m^3. There are two symmetrical density plots for each box plot. This is an overlay of a violin plot and a box plot.

We found genome-wide significant association (FDR p<0.05) between $PM_{2.5}$ exposure at birth and gene expression for two transcript clusters, namely TC10001332.hg.1 annotated to *MIR1296* gene, and TC14001976.hg.1 that is a long non-coding RNA located near *FOXA1-2* (chr14:38066368-38067552) (Table 2 and forest plot in Fig. S1 and the meta-analysis for all transcript clusters are presented in Supplementary Table 1). No association between current $PM_{2.5}$ exposures and differentially expressed genes (DEGs) at FDR p<0.05 significance level was found (top ten significant transcript clusters are presented in Supplementary Table 2). However, we found that 18 transcript clusters among the top 100 significant DEGs related to birth $PM_{2.5}$ exposure, including *KRBA2*, *NRG1*, *SCAND1* and *ZNF605* genes, were also significantly associated, at the nominal level (p<0.05) and with the same direction, with current $PM_{2.5}$ exposure (Supplementary Table 3). The results of sensitivity analyses limited to two adolescent cohorts demonstrated good agreement with those based on all three cohorts

(genome-wide correlation of beta coefficients = 0.69 and 0.77 in the analysis with $PM_{2.5}$ exposure at birth and at the time of biosampling, respectively (Supplementary Fig. S2).

Further, we found high consistency in the results of analyses including and excluding active smokers, with a very high correlation of beta coefficients, namely 93% and 84% for $PM_{2.5}$ exposure at birth and at the time of biosampling, respectively (Supplementary Fig. S3).

Table 2. FDR-significant DEGs from the meta-analysis of the association between $PM_{2.5}$ exposure at birth and gene expression in children and adolescents (n=656 from the BAMSE, GINIplus, and INMA cohorts).

Probe.Set.ID	Gene	Log2FC	SE	P-value	FDR	Direction*	Het. P-value
TC10001332.hg.1	*MIR1296*	-0.19	0.04	1.74E-07	0.01	---	0.196
TC14001976.hg.1		0.42	0.09	1.02E-06	0.03	+++	0.844

Log2FC = The logarithm fold change (one unit of the logFC translates to a two-fold change in expression), SE = standard error, FDR = false discovery rate and Het.P-value = Heterogeneity P-value based on Cochran's Q-test for heterogeneity. Results are presented per 5 $\mu g/m^3$ increase in $PM_{2.5}$.
* Order of included cohorts in the meta-analysis: BAMSE, GINIplus and INMA

As many as 364 transcript clusters mapping to 102 genes were selected from the analyses of gene expression and $PM_{2.5}$ exposure at birth cut off P-value < 0.05 and |logFC|> 0.15. The KEGG pathways analyses based on this selection identified a few enriched pathways related to olfactory transduction, ribosome, compliment and coagulation cascades and systemic lupus erythematosus (Supplementary Table 4). From the results of analyses based on $PM_{2.5}$ exposure at the time of biosampling, we identified 66 transcripts annotated to 13 genes involved in the ribosome-related pathway (Supplementary Table 5). Thus, the ribosome pathways were common in both $PM_{2.5}$ exposure at birth and at the time of biosampling.
Next, we used two different omics datasets (from BAMSE and INMA) to perform an integration of the genes in the transcriptome-wide analyses (TWAS) and Epigenome-Wide Association Study (EWAS) meta-analysis, along with protein–protein interaction predictions

348 from the High-quality INTeractomes (HINT) database (http://hint.yulab.org/) using the
349 Functional Epigenetic Module algorithms (FEMs). In these analyses we were able to identify
350 9 and 6 FEMs in relation to $PM_{2.5}$ exposure both at birth and at the current address,
351 respectively, passing the FDR significance threshold of 0.05 (Table 3).
352
353

Table 3. Summary of the output of the Functional Epigenetics Modules (FEM) algorithm listing 9 and 6 significant hotspots of differential methylation and expression in relation to birth and current $PM_{2.5}$ exposure.

Exposure to $PM_{2.5}$	Seed	Size (number of genes)	Modularity	FDR *
Birth	*NR1I2*	11	2.12	0.001
	SH3GL2	26	2.00	0.005
	TENT5A	24	1.94	0.006
	MAPK6	64	1.95	0.010
	UBE2W	61	1.35	0.012
	KCTD15	10	1.86	0.014
	MLST8	10	2.11	0.018
	RPP40	12	1.61	0.024
	GGA1	14	1.59	0.025
Current	*TAF8*	20	2.17	<0.0001
	TAF5	20	2.17	0.001
	GNAI3	31	2.07	0.002
	ISLR	22	2.05	0.024
	TRIM69	10	1.84	0.032
	SCARA3	19	1.84	0.040

Columns label the seed gene symbol, the size of the FEM, its modularity (defined as the average of the edge-weights), the associated P-value. *Significant FDR p-value < 0.05.

The size of the significant modules ranged between 10 to 64 and 10 to 31 genes in the analyses with $PM_{2.5}$ exposure at birth and current address, respectively (Supplementary Tables 6 and 7). For the top significant gene module centered around Nuclear Receptor Subfamily 1 Group I Member 2 *(NR1I2)* related to $PM_{2.5}$ exposure at birth, we observed simultaneous hypomethylation and overexpression (Fig. 2). The KEGG pathways involving the 11 genes in *NR1I2* gene module were related to fluid shear stress and atherosclerosis,

protein processing in endoplasmic reticulum, adherens junction and cancer (Supplementary Table 8).

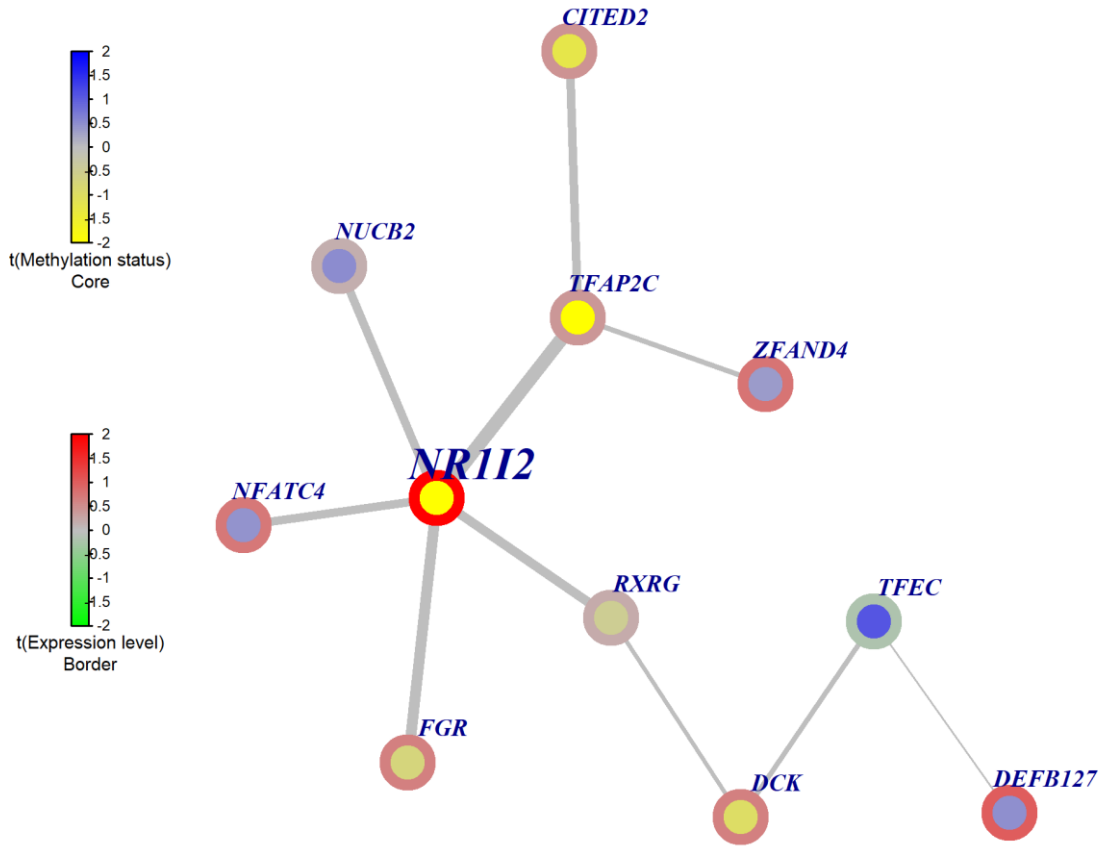

Fig. 2. Top statistically significant functional epigenetic module around seed gene *NR1I2* related to $PM_{2.5}$ exposure at birth address.

The color of node represents the DNA methylation status. Blue represents hypermethylation, whereas yellow represents hypomethylation. The different color of border refers to the expression patterns. The red color of border represents that the genes were upregulated, whereas the green color of border represents that the genes were downregulated. Edge widths are proportional to the average statistic of the gene-gene interaction network.

Another module associated with $PM_{2.5}$ exposure at birth, mitogen-activated protein kinase 6 (*MAPK6*), had the largest size of subnetwork genes (n=64) with modularity of 1.95 (Fig. 3). Subsequent KEGG pathway analysis showed that *MAPK6* module subnetwork genes are implicated in complement and coagulation cascades, as well as fatty acid degradation, oxidative phosphorylation, Huntington, Alzheimer and Parkinson disease (Supplementary Table 9).

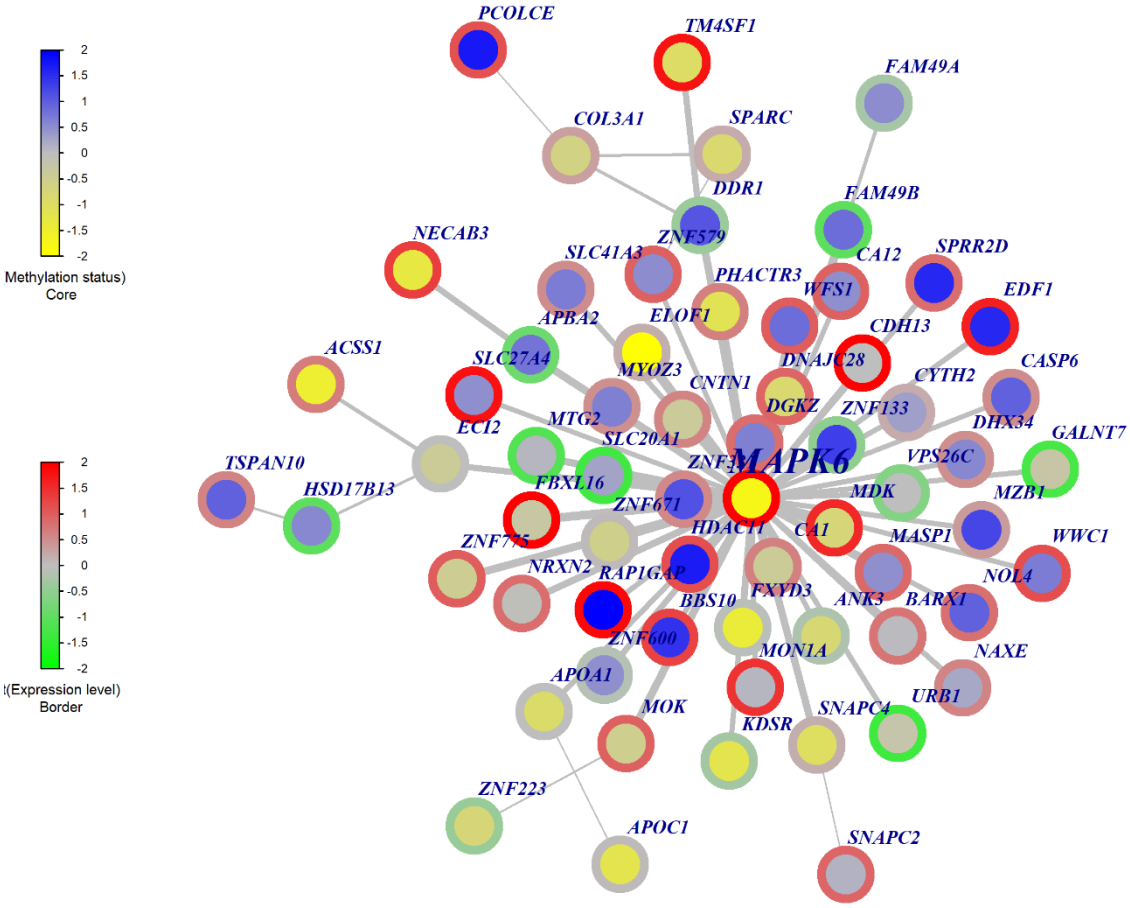

Fig. 3. Top statistically significant functional epigenetic module around seed gene *MAPK6* related to $PM_{2.5}$ exposure at current address.

The color of node represents the DNA methylation status. Blue represents hypermethylation, whereas yellow represents hypomethylation. The different color of border refers to the expression patterns. The red color of border represents that the genes were upregulated, whereas the green color of border represents that the genes were downregulated. Edge widths are proportional to the average statistic of the gene-gene interaction network.

Six FEMs were identified in association with current $PM_{2.5}$ exposure, all distinct from those found with $PM_{2.5}$ exposure at birth. Interestingly, two highly significant gene hubs, TATA-Box Binding Protein Associated Factor 8 (*TAF8*) and TATA-Box Binding Protein Associated Factor 5 (*TAF5*) identified in relation to current $PM_{2.5}$ exposure, contain the same 20 genes in their hubs. The *TAF8* module is shown in Fig. 4. Functional enrichment analyses of the *TAF5* and *TAF8* module subnetwork genes showed that these genes are mainly involved in basal transcription factor, cell cycle, thyroid hormones signaling pathway and notch signaling KEGG pathways (Supplementary Table 10).

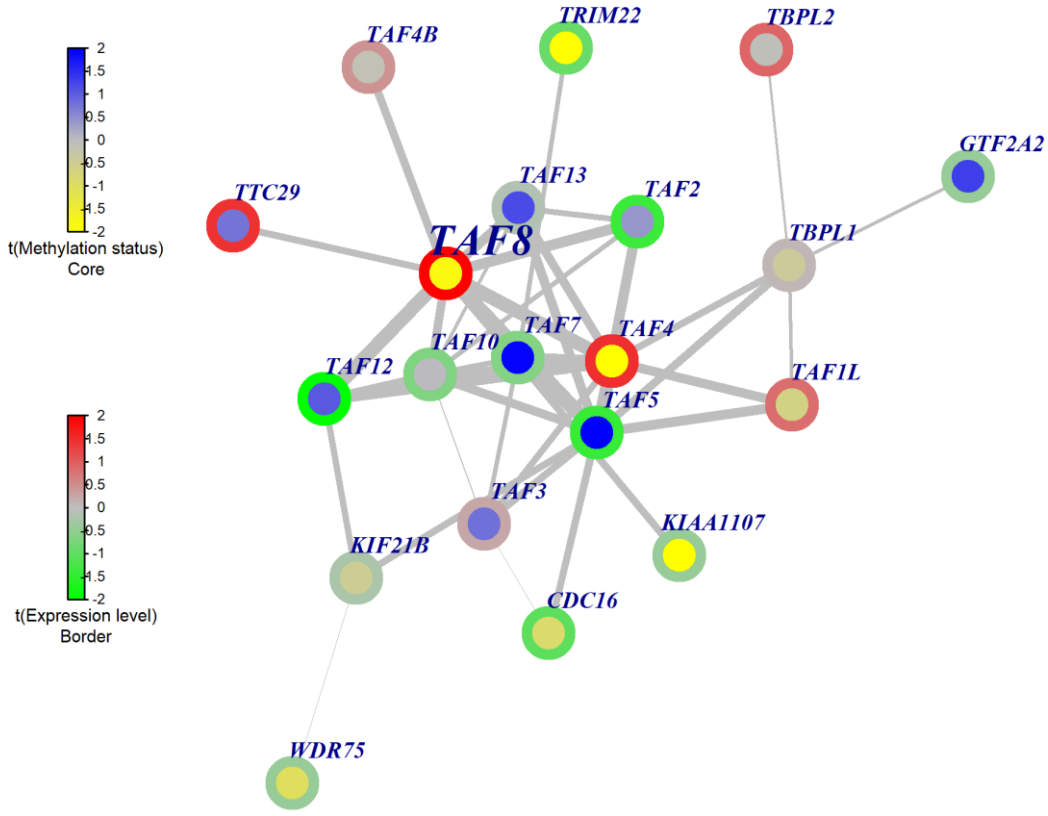

Fig. 4. Top statistically significant functional epigenetic module around seed gene *TAF8* related to $PM_{2.5}$ exposure at current address.

The color of node represents the DNA methylation status. Blue represents hypermethylation, whereas yellow represents hypomethylation. The different color of border refers to the expression patterns. The red color of border represents that the genes were upregulated, whereas the green color of border

represents that the genes were downregulated. Edge widths are proportional to the average statistic of the gene-gene interaction network.

The largest gene module size in relation to current $PM_{2.5}$ exposure was found for *GNAI3* module containing 31 genes with modularity of 2.07 in the subnetwork that appeared to be involved in cancer, prion diseases, rheumatoid arthritis, tight junction, MAPK and other signaling and metabolism KEGG pathways (Supplementary Table 11). Another identified gene module hub associated to the current $PM_{2.5}$ exposure, Scavenger Receptor Class A Member 3 (*SCARA3)*, demonstrated hypomethylation and overexpression (Fig. 5).

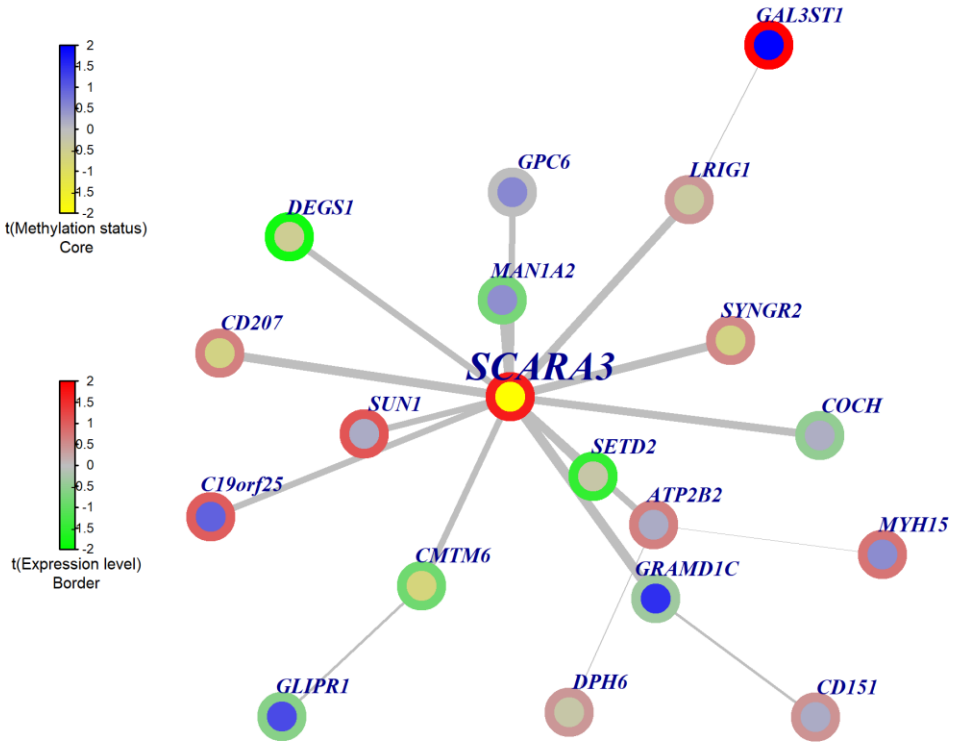

Fig. 5. Statistically significant functional epigenetic module centred around seed gene *SCARA3* related to current $PM_{2.5}$ exposure.

The color of node represents the DNA methylation status. Blue represents hypermethylation, whereas yellow represents hypomethylation. The different color of border refers to the expression patterns. The red color of border represents that the genes were upregulated, whereas the green color of border represents that the genes were downregulated. Edge widths are proportional to the average statistic of the gene-gene interaction network.

4. Discussion

This study represents a large-scale transcriptome-wide meta-analysis evaluating the association between birth and current air pollution exposure and gene expression in children and adolescents. Our meta-analysis results show associations of $PM_{2.5}$ exposure at birth with differential expression of two transcript clusters, TC10001332.hg.1, annotated to *MIR1296* gene and TC14001976.hg.1, long non-coding RNA located close to *FOXA1-2* gene. However, for current $PM_{2.5}$ exposure no FDR significant DEG was found. Further, by integrating gene expression and DNA methylation data into a putative protein interaction network we identified several hubs linked to birth and current $PM_{2.5}$ exposures (e.g. *NR1I2, MAPK6, TAF8* and *SCARA3*).

To date, little is known about air pollution associated transcriptomic signatures in children. Two earlier studies have investigated differential gene expression in the blood of children living in areas of the Czech Republic with differing levels of air pollution. Van Leeuwen and co-authors found increased blood expression of numerous genes among children inhabiting a severely polluted area compared to those residing in a less polluted region (van Leeuwen et al. 2006). In contrast, no clear effect of exposure to air pollutants was found in a study of newborns from the same geographical area (Honkova et al. 2018). These studies did not assess exposure to air pollution at the individual address level, which may have resulted in low statistical power. In the ENVIRONAGE birth cohort, genome-wide gene expression analysis in cord blood identified fifteen transcriptomic pathways altered by prenatal $PM_{2.5}$ exposure (Winckelmans et al. 2017), including “protein processing in endoplasmic reticulum” that was also observed in the present study.

Our transcriptome-wide meta-analysis revealed two DEGs in relation to $PM_{2.5}$ exposure at birth, representing novel associations in the context of air pollution exposure. One of them *MiR-1296*, coding for miRNA that has previously been found to be linked to different types of

cancer, e.g. breast, hepatocellular, colorectal, prostate and lung cancer (Deng et al. 2020; Majid et al. 2010; Phan et al. 2016; Tao et al. 2018; Xu et al. 2017). Also, it has been proposed as a putative circulating prognostic marker of heart failure (Cakmak et al. 2015). *FOXA1/2* (forkhead box a1/2) plays a key role in lung alveolar and respiratory endoderm morphogenesis and differentiation, including α-cells in the endocrine pancreas, liver and prostate luminal ductal epithelia (Friedman and Kaestner 2006; Lee et al. 2005; Wan et al. 2005; Wan et al. 2004). When comparing birth with current $PM_{2.5}$ exposure, we found 18 significant DEGs related to birth $PM_{2.5}$ exposure that were also significant at the nominal level with the same direction of effect in the analysis with current $PM_{2.5}$ exposure that includes *KRBA2, NRG1, SCAND1* and *ZNF605* genes. Molecular lesions of neuregulin 1 (*NRG1)* gene have been proposed as a new molecular signature of invasive mucinous adenocarcinoma of the lung (Trombetta et al. 2017). It has also been repeatedly shown that $PM_{2.5}$ exposure is linked to lung cancer, particularly adenocarcinoma subtype (Raaschou-Nielsen et al. 2013). Another gene, SCAN-domain-containing protein 1 (*SCAND1)* has been suggested to be involved in regulation of lipid metabolism (Babb and Bowen 2003). It has also been shown that the anti-oxidant and anti-inflammatory capacity of high-density lipoproteins can be effected by $PM_{2.5}$ exposures (Ramanathan et al. 2016).

Exploring the functional molecular patterns is critical for understanding the mechanisms of biological responses to air pollution exposure. In this study, we have utilized simultaneously comprehensive multi-omics profiling of biological samples, including genome-wide gene expression and DNA methylation. By integrating gene expression and DNA methylation into a protein interaction network we found 9 significant functional epigenetically deregulated modules associated with $PM_{2.5}$ exposure at birth. The top significant one was centered around seed gene *NR1I2*, also known as steroid and xenobiotic receptor, a transcriptional regulator of a number of important drug metabolizing enzymes and transporters, recently reported among

potential targets for asthma therapy (Wang et al. 2020). The *MAPK6* module associated with $PM_{2.5}$ exposure at birth in the present study, has previously been implicated in air pollution response. *In vivo,* a short-term exposure to black carbon was highly associated with epigenetic changes in the promoter areas of 84 MAPK pathway genes, including *MAPK6* gene (Carmona et al. 2014). Further, in a recent controlled crossover study comparing the whole blood transcriptome profiles from healthy volunteers, obtained pre and post short-term exposure, to high and low levels of air pollution, the authors found that activation of *MAPK6* was strongly related air pollution exposure (Vargas et al. 2018).

Among other identified modules, we found genes also known to be associated with diseases previously linked to air pollution exposure including obesity and type 2 diabetes (i.e., *KCTD15* (Lv et al. 2015; Ng et al. 2010), *MLST8 (Li et al. 2016)*), as well as cardio-vascular diseases (*TRIM69* (Andersson et al. 2019)). Recently, an integrative study of DNA methylation and gene expression reported *SH3GL2* as one of the key genes in pathogenesis of lung adenocarcinoma with hypermethylation and under expression (Jin et al. 2016) opposite to our finding of $PM_{2.5}$ exposure effect of hypomethylation and overexpression. In other studies, hypomethylation in certain genes, e.g. *AHRR* and *F2RL3,* has been reported as a hallmark of cancer development (Fasanelli et al. 2015).

Concurrent exposure to $PM_{2.5}$ at the time of blood sampling was associated with six hotspots of differential expression and methylation in the present study. None of these appeared to overlap with those identified in relation to $PM_{2.5}$ exposure at birth, which may at least partly be attributed to differences in exposure levels as indicated by a moderate correlation between birth and current exposure (0.32-0.46). It is also conceivable that exposure during different time (age) periods of a growing child - from birth to adolescence will result in different molecular responses. Among the six interactome hotspots we also found genes of importance for lung adenocarcinoma, i.e. G protein subunit alpha i3 (*GNAI3)* (Ye et al. 2019), as well as

prostate cancer, *i.e. TAF8* (Alvarez and Woolf 2011). The finding of epigenetically deregulated hotspot centred around the Scavenger receptor class A member 3 (*SCARA3)* gene is of interest, in light of previous reports showing altered expression of this gene induced by oxidative stress (Brown et al. 2013), one of the suggested key mechanisms involved in the adverse health effects of air pollution. Further, differential methylation of this gene has been linked to progression of type 2 diabetes mellitus (Karachanak-Yankova et al. 2016). Another hotspot gene module, Tripartite motif containing 69 (*TRIM69),* has earlier been linked to the incidence of heart failure (Karachanak-Yankova et al. 2016). It remains to be investigated whether FEMs identified in this study can mediate the phenotypic health effects of air pollution exposure on children's and young adults.

The present study is one of the first large-scale studies assessing the association of air pollution exposure, represented by $PM_{2.5}$, during different periods of life on the children´s blood transcriptome. Coordinated gene expression measurements, quality control, normalization procedures, as well as air pollution exposure assessment in all included studies was performed according to a harmonized protocol. Further, integration of genome-wide DNA methylation and gene expression data constitute another major strength of our study. All cohort-specific analyses were conducted according to the same analytical protocol.

This study has some weaknesses. $PM_{2.5}$ exposure assessment was based on the participants' residential address, without taking into consideration time-activity patterns and exposures at nonresidential addresses (e.g. school). Further, the modeled concentrations account only for outdoor air pollution and therefore may not be equivalent to the full range of personal exposures. Another potential limitation is that we utilized purely spatial air pollution models that were based on measurement campaigns carried out between 2008 - 2010 and applied those to the home addresses of study participants over a period beginning from mid-1990s. However, previous validation studies from Europe have demonstrated stability of spatial

522 contrast in levels of air pollution over time (Eeftens et al. 2011; Gulliver et al. 2011). Further,
523 we have explored this aspect in detail in our earlier investigations based on the same data by
524 performing sensitivity analyses using back-extrapolated modeled concentrations of considered
525 pollutants, and this had no impact on the results (Gehring et al. 2015; Gruzieva et al. 2014;
526 Molter et al. 2015).Although misclassification of air pollution exposure might have affected
527 our results, assessments of both exposure and gene expression levels were done independently
528 from each other, making such potential bias likely non-differential. It should also be noted
529 that we considered $PM_{2.5}$ exposure at the birth address that may also be a proxy for prenatal
530 exposure. We, therefore, cannot rule out that at least part of the observed associations with
531 exposure at birth may be attributable to prenatal exposure. Furthermore, the study participants
532 were of European ancestry. Therefore, it is not sure if our findings can be generalized to other
533 ethnic groups. We should also acknowledge age differences between the cohorts included in
534 the present analyses and that our study was not designed to specifically evaluate potential age-
535 related effects in gene expression profiles. In addition, although we adjusted our analyses for
536 a number of potential confounders, still, the possibility of residual confounding cannot be
537 ruled out. Moreover, we used estimated cell counts to account for a potential cell type effect
538 in our analyses, as measured cell counts were not available in all cohorts. Finally, the omics
539 integration analysis assumes interaction at the protein level, whereas the networks probably
540 link to each other at the regulation level. Adding miRNA data in the integration analysis
541 would likely give more power, however the data is not available in the participating cohorts.
542 In conclusion, this study provides suggestive evidence for associations of $PM_{2.5}$ exposure at
543 birth with differential gene expression in children and adolescents. Importantly, by integrating
544 gene expression and methylation data we could identify several significant interactome
545 hotspots of epigenetic deregulation gene modules in relation to $PM_{2.5}$ exposure both at the
546 birth and at the current address. Our study shows the added value of integrating environmental

exposure data with multi-omics information to improve understanding of biologic responses to exposure. Further studies are warranted to get deeper insight into the molecular mechanisms for the harmful health effect of $PM_{2.5}$.

References

Alvarez, A.; Woolf, P.J. RegNetB: predicting relevant regulator-gene relationships in localized prostate tumor samples. BMC bioinformatics 2011;12:243

Andersson, C.; Lin, H.; Liu, C.; Levy, D.; Mitchell, G.F.; Larson, M.G.; Vasan, R.S. An Integrated Multi-Omics Approach to Identify Genetic Underpinnings of Heart Failure and its Echocardiographic Precursors: The Framingham Heart Study. Circ Genom Precis Med 2019;

Babb, R.; Bowen, B.R. SDP1 is a peroxisome-proliferator-activated receptor gamma 2 co-activator that binds through its SCAN domain. The Biochemical journal 2003;370:719-727

Benjamini, Y.; Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. 1995;57:289-300

Bhetraratana, M.; Orozco, L.D.; Hong, J.; Diamante, G.; Majid, S.; Bennett, B.J.; Ahn, I.S.; Yang, X.; Lusis, A.J.; Araujo, J.A. Diesel exhaust particles dysregulate multiple immunological pathways in murine macrophages: Lessons from microarray and scRNA-seq technologies. Archives of biochemistry and biophysics 2019;678:108116

Bousquet, J.; Anto, J.; Auffray, C.; Akdis, M.; Cambon-Thomsen, A.; Keil, T.; Haahtela, T.; Lambrecht, B.N.; Postma, D.S.; Sunyer, J.; Valenta, R.; Akdis, C.A.; Annesi-Maesano, I.; Arno, A.; Bachert, C.; Ballester, F.; Basagana, X.; Baumgartner, U.; Bindslev-Jensen, C.; Brunekreef, B.; Carlsen, K.H.; Chatzi, L.; Crameri, R.; Eveno, E.; Forastiere, F.; Garcia-Aymerich, J.; Guerra, S.; Hammad, H.; Heinrich, J.; Hirsch, D.; Jacquemin, B.; Kauffmann, F.; Kerkhof, M.; Kogevinas, M.; Koppelman, G.H.; Kowalski, M.L.; Lau, S.; Lodrup-Carlsen, K.C.; Lopez-Botet, M.; Lotvall, J.; Lupinek, C.; Maier, D.; Makela, M.J.; Martinez, F.D.; Mestres, J.; Momas, I.; Nawijn, M.C.; Neubauer, A.; Oddie, S.; Palkonen, S.; Pin, I.; Pison, C.; Rance, F.; Reitamo, S.; Rial-Sebbag, E.; Salapatas, M.; Siroux, V.; Smagghe, D.; Torrent, M.; Toskala, E.; van Cauwenberge, P.; van Oosterhout, A.J.; Varraso, R.; von Hertzen, L.; Wickman, M.; Wijmenga, C.; Worm, M.; Wright, J.; Zuberbier, T. MeDALL (Mechanisms of the Development of ALLergy): an integrated approach from phenotypes to systems medicine. Allergy 2011;66:596-604

Brown, C.O.; Schibler, J.; Fitzgerald, M.P.; Singh, N.; Salem, K.; Zhan, F.; Goel, A. Scavenger receptor class A member 3 (SCARA3) in disease progression and therapy resistance in multiple myeloma. Leukemia research 2013;37:963-969

Cakmak, H.A.; Coskunpinar, E.; Ikitimur, B.; Barman, H.A.; Karadag, B.; Tiryakioglu, N.O.; Kahraman, K.; Vural, V.A. The prognostic value of circulating microRNAs in heart failure: preliminary results from a genome-wide expression study. Journal of cardiovascular medicine (Hagerstown, Md) 2015;16:431-437

Carmona, J.J.; Sofer, T.; Hutchinson, J.; Cantone, L.; Coull, B.; Maity, A.; Vokonas, P.; Lin, X.; Schwartz, J.; Baccarelli, A.A. Short-term airborne particulate matter exposure alters the epigenetic landscape of human genes associated with the mitogen-activated protein kinase network: a cross-sectional study. Environmental health : a global access science source 2014;13:94

Chu, J.H.; Hart, J.E.; Chhabra, D.; Garshick, E.; Raby, B.A.; Laden, F. Gene expression network analyses in response to air pollution exposures in the trucking industry. Environmental health : a global access science source 2016;15:101

Das, J.; Yu, H. HINT: High-quality protein interactomes and their applications in understanding human disease. BMC systems biology 2012;6:92

Deng, H.; Xie, C.; Ye, Y.; Du, Z. MicroRNA-1296 expression is associated with prognosis and inhibits cell proliferation and invasion by Wnt signaling in non-small cell lung cancer. Oncology letters 2020;19:623-630

Eeftens, M.; Beelen, R.; de Hoogh, K.; Bellander, T.; Cesaroni, G.; Cirach, M.; Declercq, C.; Dedele, A.; Dons, E.; de Nazelle, A.; Dimakopoulou, K.; Eriksen, K.; Falq, G.; Fischer, P.; Galassi, C.; Grazuleviciene, R.; Heinrich, J.; Hoffmann, B.; Jerrett, M.; Keidel, D.; Korek, M.; Lanki, T.; Lindley, S.; Madsen, C.; Molter, A.; Nador, G.; Nieuwenhuijsen, M.; Nonnemacher, M.; Pedeli, X.; Raaschou-Nielsen, O.; Patelarou, E.; Quass, U.; Ranzi, A.; Schindler, C.; Stempfelet, M.; Stephanou, E.; Sugiri, D.; Tsai, M.Y.; Yli-Tuomi, T.; Varro, M.J.; Vienneau, D.; Klot, S.; Wolf, K.;

Brunekreef, B.; Hoek, G. 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. Environmental science & technology 2012;46:11195-11205
Eeftens, M.; Beelen, R.; Fischer, P.; Brunekreef, B.; Meliefste, K.; Hoek, G. Stability of measured and modelled spatial contrasts in NO2 over time. Occupational and environmental medicine 2011;68:765-770
Fasanelli, F.; Baglietto, L.; Ponzi, E.; Guida, F.; Campanella, G.; Johansson, M.; Grankvist, K.; Johansson, M.; Assumma, M.B.; Naccarati, A.; Chadeau-Hyam, M.; Ala, U.; Faltus, C.; Kaaks, R.; Risch, A.; De Stavola, B.; Hodge, A.; Giles, G.G.; Southey, M.C.; Relton, C.L.; Haycock, P.C.; Lund, E.; Polidoro, S.; Sandanger, T.M.; Severi, G.; Vineis, P. Hypomethylation of smoking-related genes is associated with future lung cancer in four prospective cohorts. Nature communications 2015;6:10192
Fave, M.J.; Lamaze, F.C.; Soave, D.; Hodgkinson, A.; Gauvin, H.; Bruat, V.; Grenier, J.C.; Gbeha, E.; Skead, K.; Smargiassi, A.; Johnson, M.; Idaghdour, Y.; Awadalla, P. Gene-by-environment interactions in urban populations modulate risk phenotypes. Nature communications 2018;9:827
Friedman, J.R.; Kaestner, K.H. The Foxa family of transcription factors in development and metabolism. Cellular and Molecular Life Sciences CMLS 2006;63:2317-2328
Gehring, U.; Wijga, A.H.; Hoek, G.; Bellander, T.; Berdel, D.; Bruske, I.; Fuertes, E.; Gruzieva, O.; Heinrich, J.; Hoffmann, B.; de Jongste, J.C.; Klumper, C.; Koppelman, G.H.; Korek, M.; Kramer, U.; Maier, D.; Melen, E.; Pershagen, G.; Postma, D.S.; Standl, M.; von Berg, A.; Anto, J.M.; Bousquet, J.; Keil, T.; Smit, H.A.; Brunekreef, B. Exposure to air pollution and development of asthma and rhinoconjunctivitis throughout childhood and adolescence: a population-based birth cohort study. Lancet Resp Med 2015;3:933-942
Gentleman, R.C.; Carey, V.J.; Bates, D.M.; Bolstad, B.; Dettling, M.; Dudoit, S.; Ellis, B.; Gautier, L.; Ge, Y.C.; Gentry, J.; Hornik, K.; Hothorn, T.; Huber, W.; Iacus, S.; Irizarry, R.; Leisch, F.; Li, C.; Maechler, M.; Rossini, A.J.; Sawitzki, G.; Smith, C.; Smyth, G.; Tierney, L.; Yang, J.Y.H.; Zhang, J.H. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol 2004;5
Gref, A.; Merid, S.K.; Gruzieva, O.; Ballereau, S.; Becker, A.; Bellander, T.; Bergstrom, A.; Bosse, Y.; Bottai, M.; Chan-Yeung, M.; Fuertes, E.; Ierodiakonou, D.; Jiang, R.; Joly, S.; Jones, M.; Kobor, M.S.; Korek, M.; Kozyrskyj, A.L.; Kumar, A.; Lemonnier, N.; MacIntyre, E.; Menard, C.; Nickle, D.; Obeidat, M.; Pellet, J.; Standl, M.; Saaf, A.; Soderhall, C.; Tiesler, C.M.T.; van den Berge, M.; Vonk, J.M.; Vora, H.; Xu, C.J.; Anto, J.M.; Auffray, C.; Brauer, M.; Bousquet, J.; Brunekreef, B.; Gauderman, W.J.; Heinrich, J.; Kere, J.; Koppelman, G.H.; Postma, D.; Carlsten, C.; Pershagen, G.; Melen, E. Genome-Wide Interaction Analysis of Air Pollution Exposure and Childhood Asthma with Functional Follow-up. American journal of respiratory and critical care medicine 2017;195:1373-1383
Gruzieva, O.; Gehring, U.; Aalberse, R.; Agius, R.; Beelen, R.; Behrendt, H.; Bellander, T.; Birk, M.; de Jongste, J.C.; Fuertes, E.; Heinrich, J.; Hoek, G.; Klumper, C.; Koppelman, G.; Korek, M.; Kramer, U.; Lindley, S.; Molter, A.; Simpson, A.; Standl, M.; van Hage, M.; von Berg, A.; Wijga, A.; Brunekreef, B.; Pershagen, G. Meta-analysis of air pollution exposure association with allergic sensitization in European birth cohorts. J Allergy Clin Immun 2014;133:767-+
Gruzieva, O.; Xu, C.J.; Breton, C.V.; Annesi-Maesano, I.; Anto, J.M.; Auffray, C.; Ballereau, S.; Bellander, T.; Bousquet, J.; Bustamante, M.; Charles, M.A.; de Kluizenaar, Y.; den Dekker, H.T.; Duijts, L.; Felix, J.F.; Gehring, U.; Guxens, M.; Jaddoe, V.V.; Jankipersadsing, S.A.; Merid, S.K.; Kere, J.; Kumar, A.; Lemonnier, N.; Lepeule, J.; Nystad, W.; Page, C.M.; Panasevich, S.; Postma, D.; Slama, R.; Sunyer, J.; Soderhall, C.; Yao, J.; London, S.J.; Pershagen, G.; Koppelman, G.H.; Melen, E. Epigenome-Wide Meta-Analysis of Methylation in Children Related to Prenatal NO2 Air Pollution Exposure. Environmental health perspectives 2017;125:104-110

Gruzieva, O.; Xu, C.J.; Yousefi, P.; Relton, C.; Merid, S.K.; Breton, C.V.; Gao, L.; Volk, H.E.; Feinberg, J.I.; Ladd-Acosta, C.; Bakulski, K.; Auffray, C.; Lemonnier, N.; Plusquin, M.; Ghantous, A.; Herceg, Z.; Nawrot, T.S.; Pizzi, C.; Richiardi, L.; Rusconi, F.; Vineis, P.; Kogevinas, M.; Felix, J.F.; Duijts, L.; den Dekker, H.T.; Jaddoe, V.W.V.; Ruiz, J.L.; Bustamante, M.; Anto, J.M.; Sunyer, J.; Vrijheid, M.; Gutzkow, K.B.; Grazuleviciene, R.; Hernandez-Ferrer, C.; Annesi-Maesano, I.; Lepeule, J.; Bousquet, J.; Bergstrom, A.; Kull, I.; Soderhall, C.; Kere, J.; Gehring, U.; Brunekreef, B.; Just, A.C.; Wright, R.J.; Peng, C.; Gold, D.R.; Kloog, I.; DeMeo, D.L.; Pershagen, G.; Koppelman, G.H.; London, S.J.; Baccarelli, A.A.; Melen, E. Prenatal Particulate Air Pollution and DNA Methylation in Newborns: An Epigenome-Wide Meta-Analysis. Environmental health perspectives 2019;127:57012

Gulliver, J.; Morris, C.; Lee, K.; Vienneau, D.; Briggs, D.; Hansell, A. Land Use Regression Modeling To Estimate Historic (1962-1991) Concentrations of Black Smoke and Sulfur Dioxide for Great Britain. Environmental science & technology 2011;45:3526-3532

Guxens, M.; Ballester, F.; Espada, M.; Fernández, M.F.; Grimalt, J.O.; Ibarluzea, J.; Olea, N.; Rebagliato, M.; Tardón, A.; Torrent, M.; Vioque, J.; Vrijheid, M.; Sunyer, J.; Project, o.b.o.t.I. Cohort Profile: The INMA—INfancia y Medio Ambiente—(Environment and Childhood) Project. International journal of epidemiology 2011;41:930-940

Heinrich, J.; Brüske, I.; Cramer, C.; Hoffmann, U.; Schnappinger, M.; Schaaf, B.; von Berg, A.; Berdel, D.; Krämer, U.; Lehmann, I.; Herbarth, O.; Borte, M.; Grübl, A.; Bauer, C.P.; Beckmann, C.; Behrendt, H.; Ring, J.; Koletzko, S. GINIplus and LISAplus - Design and selected results of two German birth cohorts about natural course of atopic diseases and their determinants. Allergologie select 2017;1:85-95

Honkova, K.; Rossnerova, A.; Pavlikova, J.; Svecova, V.; Klema, J.; Topinka, J.; Milcova, A.; Libalova, H.; Choi, H.; Veleminsky, M.; Sram, R.J.; Rossner, P., Jr. Gene expression profiling in healthy newborns from diverse localities of the Czech Republic. Environmental and molecular mutagenesis 2018;59:401-415

Huang, Y.C.; Schmitt, M.; Yang, Z.; Que, L.G.; Stewart, J.C.; Frampton, M.W.; Devlin, R.B. Gene expression profile in circulating mononuclear cells after exposure to ultrafine carbon particles. Inhalation toxicology 2010;22:835-846

Irizarry, R.A.; Hobbs, B.; Collin, F.; Beazer-Barclay, Y.D.; Antonellis, K.J.; Scherf, U.; Speed, T.P. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics (Oxford, England) 2003;4:249-264

Jiao, Y.; Widschwendter, M.; Teschendorff, A.E. A systems-level integrative framework for genome-wide DNA methylation and gene expression data identifies differential gene expression modules under epigenetic control. Bioinformatics (Oxford, England) 2014;30:2360-2366

Jin, X.; Liu, X.; Li, X.; Guan, Y. Integrated Analysis of DNA Methylation and mRNA Expression Profiles Data to Identify Key Genes in Lung Adenocarcinoma. BioMed research international 2016;2016:4369431-4369431

Johnson, W.E.; Li, C.; Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics 2007;8:118-127

Kanehisa, M.; Sato, Y.; Furumichi, M.; Morishima, K.; Tanabe, M. New approach for understanding genome variations in KEGG. Nucleic acids research 2019;47:D590-d595

Karachanak-Yankova, S.; Dimova, R.; Nikolova, D.; Nesheva, D.; Koprinarova, M.; Maslyankov, S.; Tafradjiska, R.; Gateva, P.; Velizarova, M.; Hammoudeh, Z.; Stoynev, N.; Toncheva, D.; Tankova, T.; Dimova, I. Epigenetic alterations in patients with type 2 diabetes mellitus. Balkan J Med Genet 2016;18:15-24

Kelly, F.J.; Fussell, J.C. Linking ambient particulate matter pollution effects with oxidative biology and immune responses. Annals of the New York Academy of Sciences 2015;1340:84-94

Kim, H.S.; Kim, B.G.; Park, S.; Kim, N.; Jang, A.S.; Seo, Y.R.; Park, M.K. Gene Expression Analysis to Investigate Biological Networks Underlying Nasal Inflammatory Dysfunctions Induced by Diesel Exhaust Particles Using an In Vivo System. The Annals of otology, rhinology, and laryngology 2019:3489419883289

Lee, C.S.; Friedman, J.R.; Fulmer, J.T.; Kaestner, K.H. The initiation of liver development is dependent on Foxa transcription factors. Nature 2005;435:944-947
Lee, M.K.; Xu, C.-J.; Carnes, M.U.; Nichols, C.E.; Ward, J.M.; Kwon, S.O.; Kim, S.-Y.; Kim, W.J.; London, S.J.; The, B.c. Genome-wide DNA methylation and long-term ambient air pollution exposure in Korean adults. Clinical epigenetics 2019;11:37
Lemonnier, N.; Melen, E.; Jiang, Y.; Joly, S.; Menard, C.; Aguilar, D.; Acosta-Perez, E.; Bergstrom, A.; Boutaoui, N.; Bustamante, M.; Canino, G.; Forno, E.; Gonzalez, J.R.; Garcia-Aymerich, J.; Gruzieva, O.; Guerra, S.; Heinrich, J.; Kull, I.; Maurolagoitia, J.I.; Rodriguez, L.S.; Thiering, E.; Wickman, M.; Akdis, C.; Akdis, M.; Chen, W.; Keil, T.; Koppelman, G.H.; Siroux, V.; Xu, C.J.; Hainaut, P.; Standl, M.; Sunyer, J.; Celedon, J.C.; Anto, J.M.; Bousquet, J. A novel whole blood gene expression signature for asthma, dermatitis and rhinitis multimorbidity in children and adolescents. Allergy 2020;
Li, C.X.; Wheelock, C.E.; Skold, C.M.; Wheelock, A.M. Integration of multi-omics datasets enables molecular classification of COPD. The European respiratory journal 2018;51
Li, J.-W.; Lee, H.-M.; Wang, Y.; Tong, A.H.-Y.; Yip, K.Y.; Tsui, S.K.-W.; Lok, S.; Ozaki, R.; Luk, A.O.; Kong, A.P.S.; So, W.-Y.; Ma, R.C.W.; Chan, J.C.N.; Chan, T.-F. Interactome-transcriptome analysis discovers signatures complementary to GWAS Loci of Type 2 Diabetes. Scientific reports 2016;6:35228-35228
Lv, D.; Zhang, D.D.; Wang, H.; Zhang, Y.; Liang, L.; Fu, J.F.; Xiong, F.; Liu, G.L.; Gong, C.X.; Luo, F.H.; Chen, S.K.; Li, Z.L.; Zhu, Y.M. Genetic variations in SEC16B, MC4R, MAP2K5 and KCTD15 were associated with childhood obesity and interacted with dietary behaviors in Chinese school-age population. Gene 2015;560:149-155
Majid, S.; Dar, A.A.; Saini, S.; Chen, Y.; Shahryari, V.; Liu, J.; Zaman, M.S.; Hirata, H.; Yamamura, S.; Ueno, K.; Tanaka, Y.; Dahiya, R. Regulation of Minichromosome Maintenance Gene Family by MicroRNA-1296 and Genistein in Prostate Cancer. 2010;70:2809-2818
Molter, A.; Simpson, A.; Berdel, D.; Brunekreef, B.; Custovic, A.; Cyrys, J.; de Jongste, J.; de Vocht, F.; Fuertes, E.; Gehring, U.; Gruzieva, O.; Heinrich, J.; Hoek, G.; Hoffmann, B.; Klumper, C.; Korek, M.; Kuhlbusch, T.A.; Lindley, S.; Postma, D.; Tischer, C.; Wijga, A.; Pershagen, G.; Agius, R. A multicentre study of air pollution exposure and childhood asthma prevalence: the ESCAPE project. The European respiratory journal 2015;45:610-624
Mostafavi, N.; Vlaanderen, J.; Portengen, L.; Chadeau-Hyam, M.; Modig, L.; Palli, D.; Bergdahl, I.A.; Brunekreef, B.; Vineis, P.; Hebels, D.G.; Kleinjans, J.C.; Krogh, V.; Hoek, G.; Georgiadis, P.; Kyrtopoulos, S.; Vermeulen, R. Associations Between Genome-wide Gene Expression and Ambient Nitrogen Oxides. Epidemiology (Cambridge, Mass) 2017;28:320-328
Newman, A.M.; Liu, C.L.; Green, M.R.; Gentles, A.J.; Feng, W.; Xu, Y.; Hoang, C.D.; Diehn, M.; Alizadeh, A.A. Robust enumeration of cell subsets from tissue expression profiles. Nature methods 2015;12:453-457
Ng, M.C.; Tam, C.H.; So, W.Y.; Ho, J.S.; Chan, A.W.; Lee, H.M.; Wang, Y.; Lam, V.K.; Chan, J.C.; Ma, R.C. Implication of genetic variants near NEGR1, SEC16B, TMEM18, ETV5/DGKG, GNPDA2, LIN7C/BDNF, MTCH2, BCDIN3D/FAIM2, SH2B1, FTO, MC4R, and KCTD15 with obesity and type 2 diabetes in 7705 Chinese. The Journal of clinical endocrinology and metabolism 2010;95:2418-2425
Ogris, C.; Guala, D.; Helleday, T.; Sonnhammer, Erik L L. A novel method for crosstalk analysis of biological networks: improving accuracy of pathway annotation. Nucleic acids research 2017;45:e8-e8
Ogris, C.; Helleday, T.; Sonnhammer, E.L. PathwAX: a web server for network crosstalk based pathway annotation. Nucleic acids research 2016;44:W105-109
Pedersen, M.; Giorgis-Allemand, L.; Bernard, C.; Aguilera, I.; Andersen, A.M.; Ballester, F.; Beelen, R.M.; Chatzi, L.; Cirach, M.; Danileviciute, A.; Dedele, A.; Eijsden, M.; Estarlich, M.; Fernandez-Somoano, A.; Fernandez, M.F.; Forastiere, F.; Gehring, U.; Grazuleviciene, R.; Gruzieva, O.; Heude, B.; Hoek, G.; de Hoogh, K.; van den Hooven, E.H.; Haberg, S.E.; Jaddoe, V.W.; Klumper, C.; Korek, M.; Kramer, U.; Lerchundi, A.; Lepeule, J.; Nafstad, P.; Nystad, W.;

Patelarou, E.; Porta, D.; Postma, D.; Raaschou-Nielsen, O.; Rudnai, P.; Sunyer, J.; Stephanou, E.; Sorensen, M.; Thiering, E.; Tuffnell, D.; Varro, M.J.; Vrijkotte, T.G.; Wijga, A.; Wilhelm, M.; Wright, J.; Nieuwenhuijsen, M.J.; Pershagen, G.; Brunekreef, B.; Kogevinas, M.; Slama, R. Ambient air pollution and low birthweight: a European cohort study (ESCAPE). The Lancet Respiratory medicine 2013;1:695-704

Peretz, A.; Peck, E.C.; Bammler, T.K.; Beyer, R.P.; Sullivan, J.H.; Trenga, C.A.; Srinouanprachnah, S.; Farin, F.M.; Kaufman, J.D. Diesel exhaust inhalation and assessment of peripheral blood mononuclear cell gene transcription effects: an exploratory study of healthy human volunteers. Inhalation toxicology 2007;19:1107-1119

Pettit, A.P.; Brooks, A.; Laumbach, R.; Fiedler, N.; Wang, Q.; Strickland, P.O.; Madura, K.; Zhang, J.; Kipen, H.M. Alteration of peripheral blood monocyte gene expression in humans following diesel exhaust inhalation. Inhalation toxicology 2012;24:172-181

Phan, B.; Majid, S.; Ursu, S.; de Semir, D.; Nosrati, M.; Bezrookove, V.; Kashani-Sabet, M.; Dar, A.A. Tumor suppressor role of microRNA-1296 in triple-negative breast cancer. Oncotarget 2016;7:19519-19530

Plusquin, M.; Guida, F.; Polidoro, S.; Vermeulen, R.; Raaschou-Nielsen, O.; Campanella, G.; Hoek, G.; Kyrtopoulos, S.A.; Georgiadis, P.; Naccarati, A.; Sacerdote, C.; Krogh, V.; Bas Bueno-de-Mesquita, H.; Monique Verschuren, W.M.; Sayols-Baixeras, S.; Panni, T.; Peters, A.; Hebels, D.; Kleinjans, J.; Vineis, P.; Chadeau-Hyam, M. DNA methylation and exposure to ambient air pollution in two prospective cohorts. Environment international 2017;108:127-136

Raaschou-Nielsen, O.; Andersen, Z.J.; Beelen, R.; Samoli, E.; Stafoggia, M.; Weinmayr, G.; Hoffmann, B.; Fischer, P.; Nieuwenhuijsen, M.J.; Brunekreef, B.; Xun, W.W.; Katsouyanni, K.; Dimakopoulou, K.; Sommar, J.; Forsberg, B.; Modig, L.; Oudin, A.; Oftedal, B.; Schwarze, P.E.; Nafstad, P.; De Faire, U.; Pedersen, N.L.; Ostenson, C.G.; Fratiglioni, L.; Penell, J.; Korek, M.; Pershagen, G.; Eriksen, K.T.; Sorensen, M.; Tjonneland, A.; Ellermann, T.; Eeftens, M.; Peeters, P.H.; Meliefste, K.; Wang, M.; Bueno-de-Mesquita, B.; Key, T.J.; de Hoogh, K.; Concin, H.; Nagel, G.; Vilier, A.; Grioni, S.; Krogh, V.; Tsai, M.Y.; Ricceri, F.; Sacerdote, C.; Galassi, C.; Migliore, E.; Ranzi, A.; Cesaroni, G.; Badaloni, C.; Forastiere, F.; Tamayo, I.; Amiano, P.; Dorronsoro, M.; Trichopoulou, A.; Bamia, C.; Vineis, P.; Hoek, G. Air pollution and lung cancer incidence in 17 European cohorts: prospective analyses from the European Study of Cohorts for Air Pollution Effects (ESCAPE). The Lancet Oncology 2013;14:813-822

Ramanathan, G.; Yin, F.; Speck, M.; Tseng, C.-h.; Brook, J.R.; Silverman, F.; Urch, B.; Brook, R.D.; Araujo, J.A. Effects of urban fine particulate matter and ozone on HDL functionality. Particle and Fibre Toxicology 2016;13:26

Ritchie, M.E.; Phipson, B.; Wu, D.; Hu, Y.; Law, C.W.; Shi, W.; Smyth, G.K. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic acids research 2015;43:e47-e47

Schultz, E.S.; Hallberg, J.; Pershagen, G.; Melen, E. Reply: Early-Life Exposure to Traffic-related Air Pollution and Lung Function in Adolescence. American journal of respiratory and critical care medicine 2016;194:385-386

Sram, R.J.; Veleminsky, M., Jr.; Veleminsky, M., Sr.; Stejskalová, J. The impact of air pollution to central nervous system in children and adults. Neuro endocrinology letters 2017;38:389-396

Tao, Y.; Ma, C.; Fan, Q.; Wang, Y.; Han, T.; Sun, C. MicroRNA-1296 Facilitates Proliferation, Migration And Invasion Of Colorectal Cancer Cells By Targeting SFPQ. Journal of Cancer 2018;9:2317-2326

Tian, Z.; Guo, M.; Wang, C.; Xing, L.; Wang, L.; Zhang, Y. Constructing an integrated gene similarity network for the identification of disease genes. J Biomed Semantics 2017;8:32

Trombetta, D.; Rossi, A.; Fabrizio, F.P.; Sparaneo, A.; Graziano, P.; Fazio, V.M.; Muscarella, L.A. NRG1-ErbB Lost in Translation: A New Paradigm for Lung Cancer? Current medicinal chemistry 2017;24:4213-4228

Wan, H.; Dingle, S.; Xu, Y.; Besnard, V.; Kaestner, K.H.; Ang, S.L.; Wert, S.; Stahlman, M.T.; Whitsett, J.A. Compensatory roles of Foxa1 and Foxa2 during lung morphogenesis. The Journal of biological chemistry 2005;280:13809-13816

Wan, H.; Kaestner, K.H.; Ang, S.L.; Ikegami, M.; Finkelman, F.D.; Stahlman, M.T.; Fulkerson, P.C.; Rothenberg, M.E.; Whitsett, J.A. Foxa2 regulates alveolarization and goblet cell hyperplasia. Development (Cambridge, England) 2004;131:953-964

van Leeuwen, D.M.; van Herwijnen, M.H.; Pedersen, M.; Knudsen, L.E.; Kirsch-Volders, M.; Sram, R.J.; Staal, Y.C.; Bajak, E.; van Delft, J.H.; Kleinjans, J.C. Genome-wide differential gene expression in children exposed to air pollution in the Czech Republic. Mutation research 2006;600:12-22

Wang, B.; Mezlini, A.M.; Demir, F.; Fiume, M.; Tu, Z.; Brudno, M.; Haibe-Kains, B.; Goldenberg, A. Similarity network fusion for aggregating data types on a genomic scale. Nature methods 2014;11:333-337

Wang, Y.; Chen, Y.-J.; Xiang, C.; Jiang, G.-W.; Xu, Y.-D.; Yin, L.-M.; Zhou, D.-D.; Liu, Y.-Y.; Yang, Y.-Q. Discovery of potential asthma targets based on the clinical efficacy of Traditional Chinese Medicine formulas. J Ethnopharmacol 2020;252:112635

Vargas, J.E.; Kubesch, N.; Hernandez-Ferrer, C.; Carrasco-Turigas, G.; Bustamante, M.; Nieuwenhuijsen, M.; Gonzalez, J.R. A systemic approach to identify signaling pathways activated during short-term exposure to traffic-related urban air pollution from human blood. Environmental science and pollution research international 2018;25:29572-29583

Willer, C.J.; Li, Y.; Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics 2010;26:2190-2191

Winckelmans, E.; Vrijens, K.; Tsamou, M.; Janssen, B.G.; Saenen, N.D.; Roels, H.A.; Kleinjans, J.; Lefebvre, W.; Vanpoucke, C.; de Kok, T.M.; Nawrot, T.S. Newborn sex-specific transcriptome signatures and gestational exposure to fine particles: findings from the ENVIRONAGE birth cohort. Environ Health-Glob 2017;16

Xu, C.J.; Soderhall, C.; Bustamante, M.; Baiz, N.; Gruzieva, O.; Gehring, U.; Mason, D.; Chatzi, L.; Basterrechea, M.; Llop, S.; Torrent, M.; Forastiere, F.; Fantini, M.P.; Carlsen, K.C.L.; Haahtela, T.; Morin, A.; Kerkhof, M.; Merid, S.K.; van Rijkom, B.; Jankipersadsing, S.A.; Bonder, M.J.; Ballereau, S.; Vermeulen, C.J.; Aguirre-Gamboa, R.; de Jongste, J.C.; Smit, H.A.; Kumar, A.; Pershagen, G.; Guerra, S.; Garcia-Aymerich, J.; Greco, D.; Reinius, L.; McEachan, R.R.C.; Azad, R.; Hovland, V.; Mowinckel, P.; Alenius, H.; Fyhrquist, N.; Lemonnier, N.; Pellet, J.; Auffray, C.; van der Vlies, P.; van Diemen, C.C.; Li, Y.; Wijmenga, C.; Netea, M.G.; Moffatt, M.F.; Cookson, W.; Anto, J.M.; Bousquet, J.; Laatikainen, T.; Laprise, C.; Carlsen, K.H.; Gori, D.; Porta, D.; Iniguez, C.; Bilbao, J.R.; Kogevinas, M.; Wright, J.; Brunekreef, B.; Kere, J.; Nawijn, M.C.; Annesi-Maesano, I.; Sunyer, J.; Melen, E.; Koppelman, G.H. DNA methylation in childhood asthma: an epigenome-wide meta-analysis. The Lancet Respiratory medicine 2018;6:379-388

Xu, Q.; Liu, X.; Liu, Z.; Zhou, Z.; Wang, Y.; Tu, J.; Li, L.; Bao, H.; Yang, L.; Tu, K. MicroRNA-1296 inhibits metastasis and epithelial-mesenchymal transition of hepatocellular carcinoma by targeting SRPK1-mediated PI3K/AKT pathway. Molecular cancer 2017;16:103

Yang, X.; Feng, L.; Zhang, Y.; Shi, Y.; Liang, S.; Zhao, T.; Sun, B.; Duan, J.; Sun, Z. Integrative analysis of methylome and transcriptome variation of identified cardiac disease-specific genes in human cardiomyocytes after PM2.5 exposure. Chemosphere 2018;212:915-926

Ye, J.; Liu, H.; Xu, Z.-L.; Zheng, L.; Liu, R.-Y. Identification of a multidimensional transcriptome prognostic signature for lung adenocarcinoma. 2019;33:e22990

Zhu, J.; Zhao, Y.; Gao, Y.; Li, C.; Zhou, L.; Qi, W.; Zhang, Y.; Ye, L. Effects of Different Components of PM2.5 on the Expression Levels of NF-kappaB Family Gene mRNA and Inflammatory Molecules in Human Macrophage. International journal of environmental research and public health 2019;16