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

Simon Kebede Merid$^a$, Mariona Bustamante$^{b,c,d}$, Marie Standl$^e$, Jordi Sunyer$^{b,c,d,f}$, Joachim Heinrich$^{g,h}$, Nathanaël Lemonnier$^i$, Daniel Aguilar$^j$, Josep Maria Antó$^{h,c,d,f}$, Jean Bousquet$^{h,i,m}$, Loreto Santa-Marina$^{h,o,p}$, Aitana Lertxundi$^{h,o,q}$, Anna Bergström$^{r,s}$, Inger Kull$^h$, Åsa M. Wheelock$^a$, Gerard H. Koppelman$^{v,w}$, Erik Melén$^{a,i,l}$, Olена Gruzieva$^{r,s,t}$

$^a$ Department of Clinical Sciences and Education, Karolinska Institutet, Söderjukhuset, Stockholm, Sweden
$^b$ IgGlobal, Institute for Global Health, Barcelona, Spain
$^c$ Universitat Pompeu Fabra (UPF), Barcelona, Spain
$^d$ Spanish Consortium for Research on Epidemiology and Public Health (CIBERESP), Madrid, Spain
$^e$ Institute of Epidemiology, Helmholtz Zentrum München - German Research Center for Environmental Health, Ingolstädter Landstraße 1, 85764 Neuperb, Germany
$^f$ IMIM (Hospital del Mar Medical Research Institute), Barcelona, Spain
$^g$ Institute and Clinic for Occupational, Social and Environmental Medicine, University Hospital, LMU Munich, Ziemssenstraße 1, 80336 Munich, Germany
$^h$ Allergy and Lung Health Unit, Melbourne School of Population and Global Health, The University of Melbourne, Melbourne, Australia
$^i$ Institute for Advanced Biosciences, UGA-INSERM U1209-CNRS UMR5309, Allée des Alpes, France
$^j$ Biomedical Research Networking Center in Hepatic and Digestive Diseases (CIBEREHD), Instituto de Salud Carlos III, Barcelona, Spain
$^k$ Centre for Occupational and Environmental Medicine, Region Stockholm, Sweden
$^l$ Sächsische Kinderhospital, Berlin, Germany
$^m$ University Hospital, Montpellier, France
$^n$ MACVIA-France, Montpellier, France
$^o$ Health Research Institute-BIODONOSTIA, Basque Country, Spain
$^p$ CIBER Epidemiología y Salud Pública (CIBERESP), Spain
$^q$ Health Department of Basque Government, Sub-directionate of Public Health of Gipuzkoa, 20013 San Sebastian, Spain
$^r$ Preventive Medicine and Public Health Department, University of Basque Country (UPV/EHU), Spain
$^s$ Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden
$^t$ Centre for Occupational and Environmental Medicine, Region Stockholm, Sweden
$^u$ Sachs Children’s Hospital, Stockholm, Sweden
$^v$ Respiratory Medicine Unit, Department of Medicine and Center for Molecular Medicine, Karolinska Institutet, Solna, Stockholm, Sweden
$^w$ University of Groningen, University Medical Center Groningen, Batariëns Children’s Hospital, Department of Pediatric Pulmonology and Pediatric Allergology, Groningen, the Netherlands
$^x$ University of Groningen, University Medical Center Groningen, Groningen Research Institute for Asthma and COPD (GRIAC), Groningen, the Netherlands

Handling Editor: Zorana Jovanovic Andersen

Keywords:
Air pollution
DNA methylation
Gene expression
Integration
Children

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 $< 2.5 \, \mu m$ in diameter (PM$_{2.5}$) at birth, one of them mapping to the NR1I2 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.
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 (Gehringer 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 Frassel, 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 IFN-1 signaling, insulin receptor signaling and NRF2-mediated oxidative stress response pathway in blood (Ruan et al., 2010; Perez et al., 2007), as well as genes associated with bronchial immune responses in bronchoalveolar lavages, protein degradation, and coagulation (e.g. PLA2, F2R, CBL, UBR1) (Pettiti 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 (NO2) 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 have their expression in blood relatively increased when compared with children from a rural less polluted area (van Leeuwen et al., 2006). In other studies, gene-environment interactions with respect to air pollution exposure and expression levels have been reported (Favé et al., 2018; Gref et al., 2017).

To our knowledge, no study has evaluated expression levels across the genome in children and adolescents in relation to individual ambient air pollution exposure. Epigenomic changes related to air pollution are, on the other hand, relatively well studied (Gruzieva et al., 2017; Gruzieva et al., 2019). Although several differentially methylated CpGs and regions have been identified in relation to environmental factors, most studies found rather weak associations with ambient air pollution exposure (Gruzieva et al., 2017; Gruzieva et al., 2019; Lee et al., 2019; Plusquin et al., 2017). Integrating different types of omics data may shed new light on gene modules or molecular pathways that are epigenetically and functionally affected by PM2.5 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 PM2.5 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 PM2.5 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 16.7 years), GINIplus Munich (247 individuals aged 15.2), and INMA (165 individuals aged 4.5) cohorts with the Affymetrix Human Transcription 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, Courtabeuf, 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 Human-Methylation450 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/m²), physical activity at the time of biosampling, 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/m³ increase in PM₂.₅. A Pearson correlation analysis was performed between PM₂.₅ 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 200 bp of the TSS. If no probes mapped to within 200 bp 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 1500 bp 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 (FEM) algorithm in R package (Jiao et al., 2014). The FEM 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.

2.5.3. Functional analyses
Kyoto Encyclopedia 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 Pathway II web server (http://pathway.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 PM₂.₅ 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 PM₂.₅ 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 PM₂.₅ exposure at birth and at current address: 7.6 and 7.3 μg/m³, respectively), and highest for INMA (14.6 and 14.0 μg/m³, 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.

We found genome-wide significant association (FDR P < 0.05) between PM₂.₅ 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

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%)     | 100%              |                |
| Munich                            | 73 (33.9%)     | 52 (21.3%)        |                |
| Jarfalla                          | 79 (35.2%)     | 91 (36.9%)        |                |
| Solna                             | 63 (25.8%)     | 61 (23.6%)        |                |
| Sundbyberg                        | 33 (13.5%)     | 39 (15.2%)        |                |
| Maternal education, 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%)     |
| Autums                            | 68 (27.9%)     | 81 (33.0%)        | 38 (23.0%)     |
| Physical activity, N (%):         |                |                   |                |
| <0.5 h per week                   | 120 (49.2%)    | 137 (55.5%)       | 74 (44.8%)     |
| ≥0.5 h per week                   | 124 (50.8%)    | 110 (44.5%)       | 91 (55.5%)     |
| 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            |

N/A = Not applicable.

* Low (Elementary school or 2-year high school), Medium (3-year High school) and High (University).
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. S2).

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 $|\text{logFC}| > 0.15$. The KEGG pathways analyses based on this selection identified a few enriched pathways related to olfactory transduction, ribosome, complement 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 from the High-quality INTeractomes (HINT) database (http://hint.yulab.org/) using the Functional Epigenetic Module algorithms (FEMs). In these analyses we were able to identify 9 and 6 FEMs in relation to PM$_{2.5}$ exposure both at birth and at the current address, respectively, passing the FDR significance threshold of 0.05 (Table 3).

The size of the significant modules ranged between 10 and 64 to 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).

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).

Six FEMs were identified in association with current PM$_{2.5}$ exposure,

### Table 2

| Probe.Set.ID  | Gene       | Log2FC | SE    | P-value  | FDR  | Direction | Het. P-value |
|--------------|------------|--------|-------|----------|------|-----------|--------------|
| TC10001332.bg.1 | MIR1296    | −0.19  | 0.04  | 1.74E-07 | 0.01 | +++       | 0.196        |
| TC14001976.bg.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.
Environment International 146 (2021) 106248

S.K. Merid et al.

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   |
|                        | SIEGL2| 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   |
|                        | TAF8  | 20                     | 2.17       | <0.0001 |
|                        | TAF5  | 20                     | 2.17       | 0.001   |
|                        | UBE2W | 61                     | 1.35       | 0.012   |
|                        | GGA1  | 14                     | 1.59       | 0.025   |
|                        | TAF8  | 20                     | 2.17       | <0.0001 |
|                        | TAF5  | 20                     | 2.17       | 0.001   |
| Current                | 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.

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).

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).

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, 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. AHR and F2RL3, has been reported as a hallmark of cancer development (Fasannelli 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., 2015), 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 and 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 contrast in levels of air pollution over time (Eeftens et al., 2011; Gulliver et al., 2011). Further, we have explored this aspect in detail in our earlier investigations based on the same data by performing sensitivity analyses using back-extrapolated modeled concentrations of considered pollutants, and this had no impact on the results (Gehring et al., 2015; Gruzieva et al., 2014; Molter et al., 2015). Although misclassification of air pollution exposure might have affected our results, assessments of both exposure and gene expression levels were done independently from each other, making such potential bias likely non-differential. It should also be noted that we considered PM$_{2.5}$ exposure at the birth address that may also be a proxy for prenatal exposure. We, therefore, cannot rule out that at least part of the observed associations with exposure at birth may be attributable to prenatal exposure. Furthermore, the study participants were of European ancestry. Therefore, it is not sure if our findings can be generalized to other ethnic groups. We should also acknowledge age differences between the cohorts included in the present analyses and that our study was not designed to specifically evaluate potential age-related effects in gene expression profiles. In addition, although we adjusted our analyses for a number of potential confounders, still, the possibility of residual confounding cannot be ruled out. Moreover, we used estimated cell counts to account for a potential cell type effect in our analyses, as measured cell counts were not available in all cohorts. Finally, the omics integration analysis assumes interaction at the protein level, whereas the networks probably link to each other at the regulation level. Adding miRNA data in the integration analysis would likely give more power, however the data is not available in the participating cohorts.

In conclusion, this study provides suggestive evidence for associations of PM$_{2.5}$ exposure at birth with differential gene expression in children and adolescents. Importantly, by integrating gene expression and methylation data we could identify several significant interactome hotspots of epigenetic deregulation gene modules in relation to PM$_{2.5}$ exposure both at the 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$_{2.5}$.

**Funding**

The research leading to these results has received funding from the
Declaración de Competiendo Interés

El autores declaran que no tienen conocida conflictos de interés o personalidades que pudieron haber influenciado la redacción de este trabajo.

Aknowledgments

BAMSE: Bamse 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 Ekstrom, Andre Lauber and Niklas Andersson for excellent data management.

GINplus: The authors thank all the families for their participation in the GINplus study. Furthermore, we thank all members of the GINplus Study Group for their excellent work. The GINplus Study Group consists of the following: Institute of Epidemiology, Helmholtz Zentrum München (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 GINplus study was supported by the Commission of the European Communities, the 7th Framework Program: MedDALL 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: MedDALL project and by grants from Instituto de Salud Carlos III (Red INMA G03/176; CB06/02/0041; PI04/1436; PI06/0867; PI08/1151; and PI18/0112 incl. FEDER funds), and Generalitat de Catalunya-CIRIT 1999SGR 00241, Department of Health of the Basque Government (2005111903), Provincial Government of Gipuzkoa (DFG06/002), and annual agreements with the municipalities of the study area (Zumarraga, Urrutxu, Legazpi, Azkoitia y Azpeitia y Beasain).

Appendix A. Supplementary material

Supplementary data can be found at https://doi.org/10.1016/j.envint.2020.106248.

References

Alvarez, A., Woolf, P.J., 2011. RegNetX: predicting relevant regulator-gene relationships in localized prostate tumor samples. BMC Bioinf. 12, 243.
Andersson, C., Lin, H., Liu, C., Levy, D., Mitchell, G.F., Larson, M.G., Vasan, R.S., 2010. An integrated multi-omics approach to identify genetic up-regulations of heart failure and its echocardiographic precursors: the Framingham heart study. Circ. Genom. Precis. Med.
Babh, R., Bowen, B.R., 2003. SDP1 is a peroxisome-proliferator-activated receptor gamma co-activator that binds through its SCAN domain. Biochem. J. 370 (2), 279–287.
Benjamin, Y., Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. 1997;57:589–300.
Bhataranata, M., Orozco, L.D., Hojng, J., Diamante, G., Majid, S., Bennett, B.J., Ahn, I.S., Yang, X., Luzis, A.J., Araujo, J.A., 2019. Diesel exhaust particles dysregulate multiple immunological pathways in murine macrophages: Lessons from microarray and scRNA-seq technologies. Arch. Biochem. Biophys. 687, 108116.
Bousquet, J., Airo, L., Auffray, C., Aklis, M., Cambon-Thomsen, A., Kell, T., Haastela, T., Lambrécht, B.N., Postma, D.S., Sunyer, J., Volonta, R., Aklis, C.A., Ameis, S., Maesano, I., Arno, A., Bachert, C., Ballester, F., Basagana, X., Baumgartner, U., Bindels-Jensen, C., Bruneckef, B., Carlsen, K.H., Chatzi, I., Crameri, R., Eveno, E., Forastiere, F., García-Aymerich, J., Guerra, S., Hammad, H., Heinrich, J., Hirsch, D., Jacquin, B., Kaufmann, F., Merckof, M., Kogevinas, M., Koppelman, G., Kowalski, M.L., Lau, S., Lodrup-Carlsen, K.C., Lopez-Bonet, M., Lotvall, J., Lupinek, C., Maier, D., Makela, M.J., Martinez, F.D., Mestres, J., Monsa, I., Navijn, M.C., Neuhaus, A., Oddie, S., Paklonen, S., Pinc, I., Pison, C., Ranke, F., Reimann, S., Rial-Sebbag, E., Salapatas, M., Siroux, V., Smaghegge, D., Torrment, J., Tostkala, E., van Cauwenberge, P., von Oosterhout, A.J.M., Varrano, R., von Herten, L., Wickman, M., Wijmenga, C., Worm, M., Wright, J., Zuberbier, T., 2011. MedDALL (Mechanisms of the Development of ALLergy): An integrated approach from phenotypes to systems medicine. Mechanisms of the Development of Allergy 66 (5), 596–604.
Brown, C.O., Schiller, J., Fitzgerald, M.P., Singh, N., Salem, K., Zhan, F., Goel, A., 2013. Scavenger receptor class A member 3 (SCARA3) in disease progression and therapy resistance is regulated by multiple mechanisms. Leuk. Res. 37 (B), 963–969.
Cakmak, A.H., Coskunpinar, E., Ikittimur, B., Barman, H.A., Karadag, B., Tiryakioğlu, N. O., Kahraman, K., Vural, V.A., 2015. The prognostic value of circulating microRNAs in heart failure: preliminary results from a genome-wide expression study. J. Cardiovasc. Med. 16 (6), 431–437.
Carmona, J.J., Sofer, T., Hutchinson, J., Cantone, L., Coull, B., Maity, A., Vokonas, P., Lin, X., Schwartz, J., Baccarelli, A.A., 2014. 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. Environ. Health 13, 94.
Chu, J.-H., Hart, J.E., Chhabria, D., Garshick, E., Raby, B.A., Laden, F., 2016. Gene expression network analyses in response to air pollution exposures in the trucking industry. Environ. Health 15 (1), 101.
Das, J., Yu, H., 2012. HINT: High-quality protein interactions and their applications in understanding human disease. BMC Syst. Biol. 6 (1), 92.
Deng, H., Xie, C., Yu, Y., Du, Z., 2020. MicroRNA-1296 expression is associated with prognosis and inhibits cell proliferation and invasion by Wnt signaling in non-small cell lung cancer. Oncol. Lett. 19, 623–630.
Erfens, M., Beelen, R., de Hoogh, K., Bellander, T., Cesaroni, G., Cirach, M., Deleer, C., Debel, A., Donn, E., de Nazelle, A., Dinakopoulos, K., Eriksen, K., Falg, G., Fischer, P., Galassi, C., Grazulieneviicius, R., Heinrich, J., Hoffmann, B., Jerrett, M., Keidel, D., Korek, M., Lanki, T., Lindley, S., Madsen, C., Molter, Á., Naid, G., Nieuwenhuijzen, J., Nonnemacher, M., Pedeli, K., Raasches-Nielsen, O., Paterlaro, E., Quan, U., Ranzini, A., Schindler, C., Stempfle, M., Stedeford, E., Sugiri, D., Tsai, M.-Y., Yi-Tuomii, T., Varro, M.J., Vienema, D., Klot, S.V., Wolf, V., Bruneckef, B., Hoek, G., 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 (20), 11195–11205.
Fave, M.J., Lamaze, F.C., Soave, D., Hodgkinson, A., Gauvin, H., Bruat, V., Grenier, J.-C., Gheba, E., Skend, C., Smargiassi, A., Johnson, M., Idaghdour, Y., Awadalla, P., 2018. Gene-by-environment interactions in urban populations modulate risk phenotypes. Nat. Commun. 9, 1287.
Friedman, J.R., Kaestner, K.H., 2006. The Foxa family of transcription factors in development and metabolism. Genes Dev. Life Sci. 63 (19-20), 2317–2328.
Gehring, U., Wilga, A.H., Hoek, G., Bellander, T., Berdel, D., Brüseke, I., Fuertes, E., Gruzieva, O., Heinrich, J., Hoffmann, B., de Jongste, J.C., Klimper, C.,
