



# Benzo[a]pyrene is associated with dysregulated myelo-lymphoid hematopoiesis in asthmatic children

Hyunok Choi<sup>a,\*,1</sup>, Won-min Song<sup>b,1</sup>, Minghui Wang<sup>b</sup>, Radim J. Sram<sup>c,d</sup>, Bin Zhang<sup>b,\*,\*</sup>

<sup>a</sup> Departments of Environmental Health Sciences, Epidemiology, and Biostatistics, State University of New York at Albany School of Public Health, Rensselaer, NY, USA

<sup>b</sup> Icahn Institute of Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, NY, USA

<sup>c</sup> Department of Genetic Toxicology and Nanotoxicology, Institute of Experimental Medicine, Academy of Sciences of the Czech Republic, v.v.i., Vídeňská 1083, 142 20 Prague 4, Czech Republic

<sup>d</sup> University of Chemistry and Technology, Prague, Faculty of Food and Biochemical Technology, Department of Food Analysis and Nutrition, Technická 3, 166 28 Prague, Czech Republic

## ARTICLE INFO

Handling Editor: Xavier Querol

### Keywords:

Polycyclic aromatic hydrocarbon  
Asthma  
Air pollution  
Gene co-expression network analysis  
Multiscale clustering analysis  
Environmental systems biology

## ABSTRACT

**Background:** The extent to which ambient benzo[a]pyrene (B[a]P) contributes to mechanistically distinct de novo asthma remains unknown.

**Objectives:** To identify molecular signatures and regulatory networks underlying childhood exposure to ambient B[a]P and asthma, using robust and unbiased systems biology approaches.

**Methods:** Clinically confirmed asthmatic ( $n = 191$ ) vs. control ( $n = 194$ ) children (aged, 7–15) were enrolled from a polluted urban center and semi-rural region in Czech Republic. Contemporaneous B[a]P concentration, gene expressions, DNA methylation data were analyzed against asthma diagnosis, as well as a modified prognostic index of asthma, using integrative multiscale co-expression network analysis. Sample-wise cell type compositions were inferred by a machine learning approach (i.e. CIBERSORT) with reference gene expressions of purified 38 distinct hematopoietic cell states from umbilical cord (i.e. stem cell/progenitors) or peripheral blood (i.e. lymphocytes).

**Results:** The median outdoor B[a]P was increased near the homes of the urban children with ‘moderate’ or ‘severe’ prognostic markers of asthma, but not in the urban controls. An elevated B[a]P induced epigenetic suppression of NF- $\kappa$ B inflammation, decreased Natural Killer T (NKT) cells and activated anti-inflammatory IL10-secreting CD8<sup>+</sup> T effective memory cells. B[a]P was positively correlated with an increased expression of a heme biosynthesis gene, *ALAS2*, which in turn, appears to promote concurrent increase of neutrophilic meta-myelocyte and mature CD71<sup>low</sup> erythroid cells. Furthermore, erythroid-specific master transcription regulator gene (*GATA1*), glutathione transferase genes (*GSTM1* and *GSTM3*) and Eosinophil marker (*IL5RA*) were simultaneously activated in the urban asthma cases.

**Conclusions:** B[a]P might contribute to concurrent suppression of pro-inflammatory (e.g. NF- $\kappa$ B mediated NKT cells), and activation of anti-inflammatory pathways (e.g. IL10-secreting CD8<sup>+</sup> T cells) in the urban asthmatic children. In addition, B[a]P appears to elevate heme biosynthesis, which in turn, promotes neutrophilic meta-myelocyte expansion and reduction of CD71<sup>+</sup> erythroids.

**Abbreviations:** ALAS2, 5-aminolevulinic acid synthase-2; B[a]P, Benzo[a]pyrene; *BTN3A2*, Butyrophilin Subfamily 3 Member A2; CAP, Czech Asthma Pathway; CREED, CRowd Extracted Expression of Differential Signatures; CYP450s, cytochromes P450; DEG, Differentially expressed genes; FET, Fisher's Exact Test; ICAC, inner city children asthma/healthy cohort; ICD, International Classification of Diseases; IPAH, idiopathic pulmonary arterial hypertension; JT, Jonkheere-Terpestra; KS, Komologorov-Smirnov; meSCG, correlation analysis of methylation profiles and gene expression; NF- $\kappa$ B, Nuclear factor kappa-B; NK, Nature Killer; AMOD, Atopic March Onset & Diversity; ONOO-, Peroxynitrite; PAH, Polycyclic aromatic hydrocarbon; ROS, reactive oxygen species; SCG, Significantly correlated genes; PM<sub>2.5</sub>, particulate matter < 2.5  $\mu$ m in aerodynamic diameter; FCRL2, Fc receptor-like 2

\* Correspondence to: H. Choi, One University Place, Room 153, Rensselaer, NY 12144-3456, USA.

\*\* Correspondence to: B. Zhang, Leon and Norma Hess Center for Science and Medicine, Room S8-111, Box 1498, 1470 Madison Avenue, New York, NY 10029, USA.

E-mail addresses: [hyunokc@gmail.com](mailto:hyunokc@gmail.com) (H. Choi), [zhangb@hotmail.com](mailto:zhangb@hotmail.com) (B. Zhang).

<sup>1</sup> Equally contributed.

<https://doi.org/10.1016/j.envint.2019.04.052>

Received 2 August 2018; Received in revised form 21 April 2019; Accepted 22 April 2019

0160-4120/ © 2019 Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## 1. Introduction

Asthma is a heterogeneous collection of diseases, which poses a heavy burden on children worldwide (Nunes et al., 2017). Within US, an estimated 40 million people suffer from asthma during lifetime (Centers for Disease Control and Prevention, 2014). However, the existing pharmacotherapies are ineffective in 38% of children and 50% of adults with asthma (Centers for Disease Control and Prevention, 2014). Therefore, a mismatch between etiologic processes and available treatments likely contribute toward current economic burden of ~\$56 billion/year, as of 2011 (Nunes et al., 2017). Asthma develops from interplay between inherited factors of a host and environmental stressors. As the genetic susceptibilities of hosts are unlikely to change within a few decades, vast alterations in the modern living environments, lifestyles, and diet since World War II are speculated to drive the current trend (Guarnieri and Balmes, 2014a; Perera, 2017).

Perinatal environmental exposures, in particular, air pollutants, play critical roles in the children's immune system development. However, it remains unclear to what extent air pollutants *causally* contribute to physiologic alterations (i.e. airway inflammation, remodeling, and airway hyper-responsiveness) prior to/during an induction of new asthma. Childhood exposures to ambient air pollutants, such as particulate matter with a diameter < 2.5 µm (PM<sub>2.5</sub>), or its biologically active components, namely, polycyclic aromatic hydrocarbons (PAHs), could impair lung function (Stern et al., 2013), and increase the risk for early-onset wheezing symptoms (Jedrychowski et al., 2010; Perzanowski et al., 2013), or existing asthma exacerbation (Samoli et al., 2011). Within Czech Republic, prenatal and early-postnatal exposures to benzo[a]pyrene (B[a]P), a representative PAH, predict markedly higher co-morbid conditions of atopic dermatitis, wheezing, and allergic rhinitis during the first five years of life (Dostal et al., 2013). However, to date, it remains unknown whether ambient PAHs contribute to mechanistically distinct alterations in the immune system during the first few years, subsequently inducing *de novo* asthma (Dixon and Poynter, 2016). The present cross-sectional investigation aims to identify PAH-associated molecular signatures and networks of asthma through an agnostic analysis of high-dimensional transcriptome, DNA methylation, and clinical outcome data. In addition to the asthma outcome, we examine whether children's combined index (including, atopic dermatitis, atopic sensitization, wheezing, and/or upper respiratory infection diagnoses) at ≤ 2 years of age, collectively termed Atopic March Onset and Diversity (AMOD) index), are also associated with ambient B[a]P concentration. We test the following hypotheses: 1) high childhood exposures to B[a]P predict concentration-dependent alterations in epigenome and transcriptome; 2) B[a]P-associated asthma can be characterized by modular structures underlying gene co-expression networks (Song and Zhang, 2014; Zhang et al., 2013; Zhang and Horvath, 2005); and 3) such dysregulated gene expression and subnetworks (i.e. modules) are correlated with AMOD index as well as the asthma diagnosis.

## 2. Materials and methods

The subject enrollment and laboratory analysis procedures have been described (Rossnerova et al., 2011, 2013; Sram et al., 2013a, 2013b). A historically industrial city of Ostrava (i.e. defined as P1) and three Southern Bohemia villages (i.e. defined as P0) were chosen as our study sites, based on our prior investigations of early-childhood allergy and respiratory outcomes (Dostal et al., 2013; Sram et al., 2013a, 2013b). Benzo[a]pyrene (B[a]P) was chosen as the representative compound for PAHs as well as particulate matter < 2.5 µm in aerodynamic diameter (PM<sub>2.5</sub>) due to its demonstrated biological activities (Binkova et al., 2003; Sram et al., 2006). Briefly, a case-control study was conducted during November 2008. As all children in Czech Republic are assigned a pediatrician, we enlisted all pediatricians and the nurses within the target districts, who in turn, invited the parents of the

eligible children to answer the questionnaires (Dostal et al., 2013). A similar number of the asthma cases ( $n = 191$ ) and the controls ( $n = 194$ ) were enrolled from a polluted urban center within historically industrial city (i.e., Ostrava, termed **P1**) and a rural background region (i.e. Southern Bohemia, termed **P0**), respectively (Sram et al., 2013b). The asthma and allergy diagnoses were made by pediatricians, and subsequently confirmed by allergologists/pneumologists (Velická et al., 2015). A child was diagnosed as a case of 'current' asthma, primarily using on a combination of the symptoms (starting around 3rd month of age), and additionally considering lung function deficit (using spirometry at age five or higher), and positive response to bronchodilation (also determined at age five or higher) on older children. Following major and minor criteria were considered for infants and toddlers. Major criteria include: i) a history of hospitalization for bronchiolitis or heavy wheezing dyspnea; ii) a minimum of three episodes of wheezing dyspnea during the last 6 months prior to the current clinical visit; iii) positive asthma diagnosis in either parent; and iv) history of atopic dermatitis diagnosis. Minor criteria include: i) allergic rhinitis without a cold; ii) wheezing symptoms without any indications of infection; iii) Eosinophilophilia; and iv) male sex. If the infant/toddler met at least one of the major criteria or two of the minor criteria, he/she was considered to be at a high risk of asthma, and was closely monitored. In older children, asthma was suspected if the child presented with a minimum of three of the following events, including parental report of ≥ 2 episodes of cough (any time during the day), difficulty with breathing, wheezing and/or whistling sound during breathing, behavioral limitations in home and/or school activities, and hospitalization due to symptom exacerbation. Allergy status was determined as positive response (wheal size ≥ 3 mm diameter) based on a skin prick test (PhadiatopH, Pharmacia & Upjohn Diagnostics, Uppsala, Sweden) for the local airborne allergens as described in Online Supporting Material (Heinzerling et al., 2013). Urinary cotinine (ng/mg Creatinine) level was analyzed by radioimmunoassay (Rossnerova et al., 2011). Children whose cotinine concentration ranged between 20 and 449 ng/mg Creatinine were considered secondhand smokers (Dostal et al., 2008). The unexposed children were defined as those with cotinine < 20 ng/mg Creatinine. Actively smoking case and control children (i.e. > 450 ng/mg of Creatinine) were removed from the analysis (Dostal et al., 2008). The control was defined as those who were free from all of the above conditions. The ethics committee of the Institute of Experimental Medicine, Academy of Sciences, approved the study. The parents signed an informed consent.

### 2.1. Polycyclic aromatic hydrocarbon monitoring

Routinely monitored ambient B[a]P data from four monitoring sites within Ostrava and four sites in Southern Bohemia were obtained from The Czech Hydrometeorological Institute (CHMI) monitors airborne eight PAHs and PM<sub>2.5</sub> using a Versatile Air Pollution Sampler (VAPS) as described for November 2008 period (Czech Hydrometeorological Institute, 2008).

### 2.2. Gene expression analysis

Full summary of the procedure has been described (Sram et al., 2013b). Whole blood samples were processed using PAXgene Blood RNA Kit (PreAnalytiX). HumanHT-12 v3 BeadChips (Illumina) with probes for > 48 K transcripts (total available samples = 368). The isolated RNA (200 ng) was reverse transcribed, amplified and biotinylated by in vitro transcription (Illumina TotalPrep RNA Amplification Kit, Ambion, Foster City, CA, USA). The microarrays were processed by BeadStation 500 (Illumina) and interpreted using BeadStudio Software (version 3.3.7, Illumina). Quality control and normalization of the data were performed in R statistical environment (<http://www.r-project.org>) using the Limma package.

### 2.3. Infinium methylation assay

Genomic DNA was isolated from whole venous blood collected in EDTA tubes (Rossnerova et al., 2013). Bisulfite-converted DNA was prepared through enzymatic fragmentation, precipitation, resuspension, and hybridization followed by washing and BeadChip staining. A batch of 12 samples were processed with Infinium Human Methylation 27 K BeadChips (Illumina, catalog number WG-311-2201).

### 2.4. Atopic March Onset and Diversity (AMOD) index

Early-life morbidities (i.e. during the first two years of life) were combined to create an ordinal scale variable (see Online Supporting Material 1). Jonckheere–Terpestra (JT) test was used to non-parametrically assess a linear trend, without assuming an underlying normal distribution at  $\alpha = 0.05$ .

### 2.5. Power analysis

After exclusion of active smokers and others with a poor quality control score, the power analysis was conducted on 174 (P1) and 156 (P0) subjects (Cohen, 1988), while adjusting inflated type I errors from multiple testing (So and Sham, 2011). For the two-tailed correlation test, approximately 2.03E8 pairwise correlation tests were performed for 20,130 genes within the present data. Such number of tests required significance  $< 2.47\text{E-}10$  for individual tests, using Bonferroni corrected  $p\text{-value} < 0.05$  (So et al. 2018). At such significance level, a minimum power of 0.90 corresponds to absolute correlation coefficient (i.e.  $|\rho|$ ) of 0.50 with 174 children in P1. Within CAP data, 2.48E6 pairs ( $\sim 1.2\%$  of all tested gene pairs) well exceed  $|\rho| > 0.50$ . As Bonferroni correction is very conservative and can lead to over-correction due to the independence assumption for individual tests (So and Sham, 2011), we employed permutation-based False Discovery Rate (FDR) correction in detecting significantly correlated gene pairs for MEGENA (Song and Zhang, 2015), leading to 51,612 significant gene pairs with  $\text{FDR} < 0.05$ . Furthermore, our two-tailed  $t$ -test of the differential expression analysis of P1 (174 children) and P0 (156 children) requires a significance level smaller than  $2.48\text{E-}6$ , based on a Bonferroni corrected  $p\text{-value} < 0.05$ , at a minimal power of 0.9. This corresponds to effect size of  $\sim 0.67$ . Similarly, differential methylation analysis of P1 (92 children) and P0 (96 children) requires significance  $< 1.83\text{E-}6$  (Bonferroni corrected  $p\text{-value} < 0.05$ ) across 27,579 individual CpG sites and yields effect size of 0.91 at a minimal power of 0.9. In all differential analyses, type I errors were strictly controlled by correcting  $p$ -values from individual tests by Benjamini–Hochberg FDR correction with  $\text{FDR} < 0.05$  cut-off for the significance (Benjamini and Hochberg, 1995).

### 2.6. Differential gene expression and significantly correlated genes

An integrative analysis of ambient B[a]P, transcriptome, epigenome, asthma or AMOD index was conducted on the case and control groups in the urban and the rural areas, respectively (Fig. 1). Altered gene pathways and functions were identified through enrichment test of Differentially Expressed Genes (DEGs) and Significantly Correlated Genes (SCGs) for the known pathways and gene signatures. DEGs were identified by comparing the following groups: 1) all the children in the urban area (i.e. P1) vs. those in the rural area (i.e. P0); 2) all the asthmatic children (A1) vs. those non-asthmatic (A0); and 3) all the urban asthmatic children (A1P1) vs. the urban control children (A0P1). SCGs were identified by correlating the gene expression profiles with traits of interest (e.g. B[a]P and AMOD).

### 2.7. Differential methylation analysis

Differentially Methylated Regions (DMRs) were detected using the limma package with BH  $\text{FDR} < 0.05$  and  $\text{FC} > 1.2$  by comparing the

P1 and P0 children. We also leveraged the mapping between CpG sites and respective proximal genes annotated by Infinium Human Methylation 27 K BeadChips to identify potential cis-regulated gene expressions by DMRs. By classifying cis-/trans-CpG sites, methylation correlated gene signatures (cis-/trans-meSCGs) were identified by Spearman's correlation analysis between all pairs of methylation profiles (at cis-/trans-CpG sites) and gene expression profiles with BH  $\text{FDR} < 0.05$ . In contrast to DMR, meSCGs capture epigenome-transcriptome interactions within P1 or P0. The resulted DEG, DMR, SCG, and meSCG signatures were then functionally annotated using the Molecular Signature Database (MSigDB) (see Fig. S2F for the gene signatures used in the analysis). All statistical and bioinformatics analyses were performed using the R software environment version 3.3.1 ([www.r-project.org](http://www.r-project.org)).

### 2.8. Co-expression network analysis

Gene co-expression networks were constructed, using Multiscale Embedded Gene Co-expression Network Analysis (MEGENA) method (Song and Zhang, 2014, 2015; Song et al., 2012). Briefly, we selected and embedded gene pairs with significant correlations ( $\text{FDR} < 0.05$ ) onto a 3-dimensional topological sphere (Tumminello et al., 2005). Topology of the resulting “planar filtered networks (PFNs)” was analyzed to identify network clusters (i.e. gene modules), using multiple resolutions (Song and Zhang, 2014, 2015; Song et al., 2012). MEGENA identified a set of co-expressed gene modules (i.e. subnetworks) and their key regulators which were further evaluated for their biological relevance by testing the enrichment of the DEGs and SCGs with Fisher's Exact Test (FET)).

### 2.9. Module ranking

Modules with too few or too many genes are often less biologically meaningful due to over-/under-splitting in the clustering analysis, therefore we focus on the modules with 50 to 2000 genes. Relevance of a gene module to asthmatic outcomes was calculated by a summary score from its enrichment for the previously described DEG and SCG signatures:

$$MS_m = \sum_{t=1}^T -\log_{10}(p_t),$$

where,  $m = 1, \dots, M$ ,  $M$  = the total number of modules,  $T$  = the total number of DEG and SCG signatures, and  $p$  = corrected FET  $p$ -value. DEG and SCG signatures associated with asthma diagnosis and ambient B[a]P concentration were used to calculate the module ranking score (see the Online Supporting Material 2).

### 2.10. Module Differential Connectivity (MDC) analysis

MDC for a given module was defined as a ratio of the mean connectivity among the genes in the module in one network to that of the same set of genes in another network (Zhang et al., 2013).

### 2.11. Comparison of methylation-gene expression correlations between P1 and P0

For each module, the significantly correlated transcript and CpG probe pairs (Benjamini–Hochberg corrected Spearman correlation  $P < 0.05$ ) were identified. The distributions of the correlation coefficients across different groups were compared using the Kolmogorov–Smirnov (KS) test.

### 2.12. In-silico validation of gene-gene interactions

We validated our putative network structures using the established

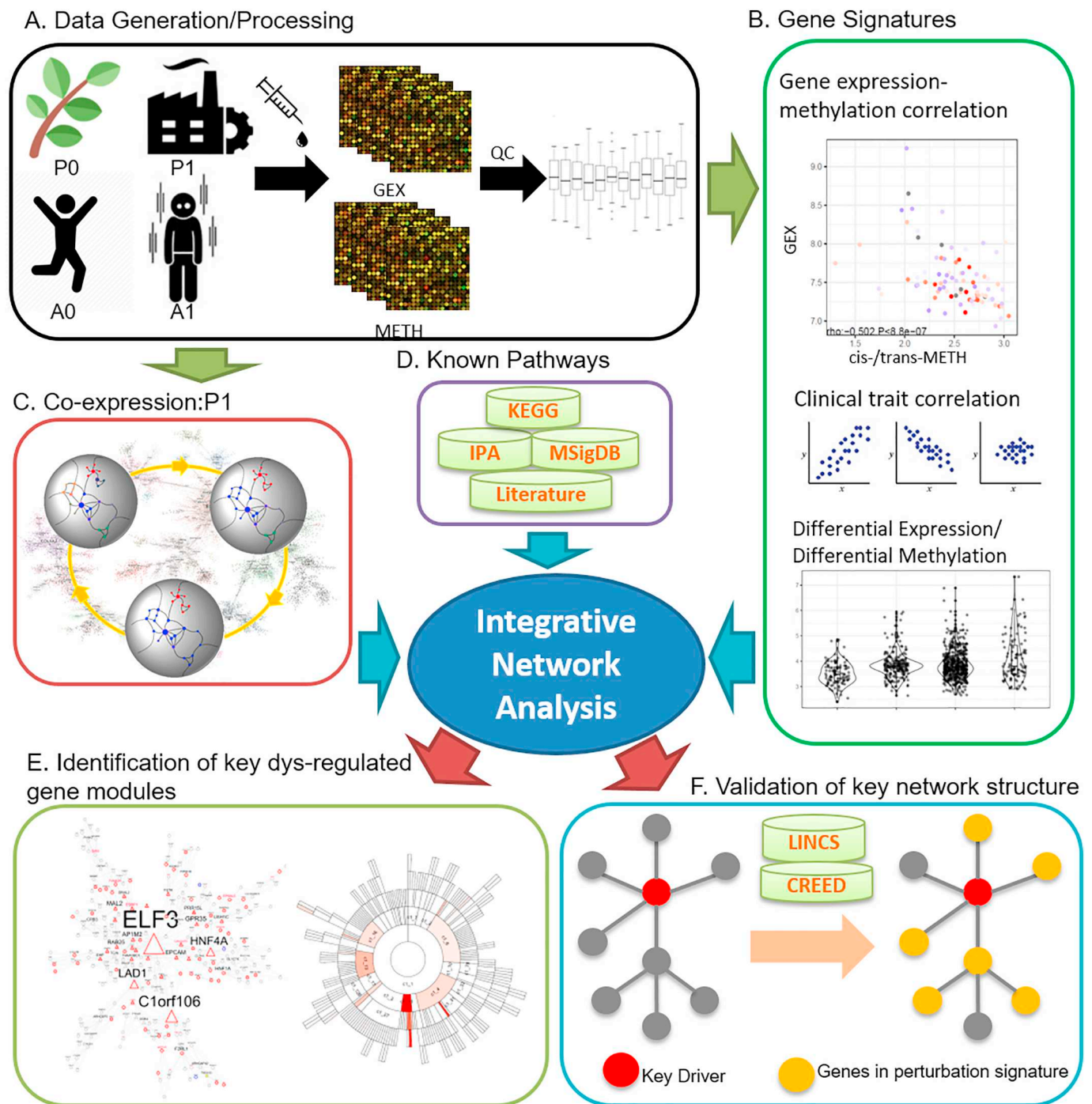


**Fig. 1.** Overview of integrative network analysis of Czech Asthma Project (CAP). **A.** Data Processing: Gene expression (GEX)/methylation (METH) profiles from plasma samples from asthmatic (A1) and controls (A0) in polluted urban (P1) vs. rural (P0) regions are generated. Quality Controls (QC) procedures removed outliers and implausible values. Putative confounding variables are adjusted within the model. **B.** Gene Signatures: gene signatures are identified by correlating gene expression data against methylation profiles, B[a]P, asthma/AMOD outcomes, respectively. **C.** Co-expression: MEGENA is applied to identify co-expression network from the gene expression data for P1 children, relative to that of P0 children. **D.** Known pathways: collections of known pathways and functions (e.g. MSigDB) are compared against our CAP co-expression modules to identify enriched pathways. **E.** Identification of dys-regulated gene modules: Compendium of gene signatures for altered gene expressions/methylations are tested for enrichments in the gene modules. **F.** Validation of predicted network structure: Predicted hub genes from the gene modules are functionally validated by projecting gene perturbations signatures (from LINCS and CREED) onto the neighborhood network structure.

gene perturbation signatures derived from the Library of Network-Based Cellular Signatures (LINCS) and CRowd Extracted Expression of Differential Signatures (CREEDS) (Wang et al., 2016). For the LINCS, we selectively analyzed the signatures generated from lung tissue specific cell-lines, namely HCC515 and A549. The overall gene perturbation signatures in the CREEDS were intersected with the network

neighborhoods of their corresponding perturbed genes in each PFN for enrichment test. The network structure around a gene was deemed successfully validated if the BH corrected FET P is  $< 0.05$ . For example, we validated the subnetwork of ZFAND6 by its knock-down signatures in A549 cells from the LINCS (Fig. S3D-E). We utilized the knock-down signatures from GSE22366 for primary erythroid progenitor cells in the

CREEDS, and the knock-down signatures from lung tissue cell lines (A549 and HCC515) in the LINCS.

### 2.13. Calculation of Asthma/B[a]P exposure association scores in Fig. S2D

In order to identify most dys-regulated known pathways by B[a]P exposure in P1 children, we interrogated enrichment of B[a]P signatures for the MSigDB signatures. For a pair of signatures with an enrichment significance level of FET  $p$ , we defined an enrichment score  $ES$  as  $-\log_{10}$  (corrected FET  $p$ ). B[a]P association score for each MSigDB signature was calculated as follows. First, the positive and negative SCG signatures for  $\ln(B[a]P)$  in P1 were tested for enrichment in each MSigDB signature, leading to two enrichment scores  $ES(B[a]P^+)$  and  $ES(B[a]P^-)$ , respectively. The two enrichment scores were then combined into an aggregate score,  $S(B[a]P^{all}) = ES(B[a]P^+) - ES(B[a]P^-)$ .  $S(B[a]P^{all}) > 0$  denotes a stronger enrichment for the positive SCGs while  $S(B[a]P^{all}) < 0$  shows a better enrichment for the negative SCGs. Similarly for asthma association scores, up-regulated scores for asthma was defined by  $S(\text{asthma}^+) = \sum_i -\log_{10}(\text{FET FDR}_i)$  where  $t$  denotes up-regulated signatures in the comparisons: A1P1 vs A0P1, mild asthma vs healthy control in P1, and severe asthma vs healthy control in P1 as shown in the legend of Fig. S2A. Note that there were no significantly down-regulated genes at  $\text{FDR} < 0.05$  in these comparisons, therefore  $S(\text{asthma}) = S(\text{asthma}^+)$ .

### 2.14. Analysis of CIBERSORT inferred cell population with clinical traits

The cell type compositions in individual samples were inferred by the reference gene expression profiles (RGEs) of 38 distinct cell states in the hematopoietic lineage, purified by multi-parameter flow cytometry (see Table S1 in (Novershtern et al., 2011) for immuno-phenotypes and respective cell populations). The respective classifiers were trained by the CIBERSORT framework. Stem cells and progenitor populations were purified from umbilical cord blood, and terminally differentiated lymphocytes were purified from peripheral blood (Novershtern et al., 2011). We performed Spearman's correlation analysis between the CIBERSORT inferred relative cell populations and the continuous variables (e.g. B[a]P concentration). We performed a non-parametric Wilcoxon rank sum test of the ordinal variables (e.g. AMOD index) with the relative cell populations. The inferred cell population was then associated with asthma or B[a]P concentration by summary scores calculated as follows. In particular, negative and positive correlations were summarized separately so as to provide directions for the final association scores (see the Subsection Calculation of Asthma/B[a]P exposure association scores in Fig. S2D for details). For the asthma association score,  $-\log_{10}$  of Wilcoxon rank sum  $p$ -values were pooled for the following comparisons: asthmatic against controls in P1, severe/moderate/mild AMOD against controls in P1. For B[a]P exposure association score, we calculated  $-\log_{10}(p\text{-value})$  from Spearman correlation analysis with B[a]P exposure.

## 3. Results

The case and control children's exposures as well as their clinical and demographic information are shown (Table 1). All asthmatic children were also positive for their history of mild, moderate, or severe AMOD index (Fig. S1). Higher proportion of the P1 asthma cases met the moderate (21%) or the severe (11%) AMOD case definition, compared to the P0 asthmatics (12% moderate; and 4% severe). While most (88%) P0 asthma cases ( $n = 80$ ) are sensitized to multiple aeroallergens (compared to 16% of P0 sensitized controls ( $n = 15$ )), a much smaller proportion (54%) of P1 asthmatic children ( $n = 47$ ) are sensitized (compared to the P1 controls). The median B[a]P is positively correlated with elevated AMOD index only among the P1 asthma cases (JT test for linear trend  $p < 0.001$ ), but not the P0 asthma cases (JT test  $p > 0.05$ ), compared to their respective controls (Fig. S1).

Furthermore, only the non-atopic P1 asthmatic children (but not the atopic P1 asthmatic children) are associated with significantly positive increase in the median B[a]P concentration (Fig. S1).

### 3.1. Ambient B[a]P concentration drives alterations of key pathways among mild P1 AMOD cases

Ambient B[a]P concentration is significantly associated with altered functional pathways including HEME metabolism, immune response, neurotrophin signaling, and cytoplasmic gene expression (Fig. S2). In particular, the genes positively correlated with the B[a]P level are most robustly enriched for the HEME metabolism pathway (FET  $p = 6.36\text{E-}11$ , 14.4-fold enrichment (FE)), which is also enriched in the up-regulated DEG signatures in the asthmatics (FET  $p = 1.54\text{E-}2$ ; 13.3 FE) as well as the mild AMOD cases (FET  $p = 1.25\text{E-}2$ ; 27.3 FE) compared with the respective controls (Fig. S2; Table S1; Online Supporting Material 2). However, there are no significantly down-regulated genes in the P1 asthmatic children compared to the controls at  $\text{FDR} < 0.05$ . Overall, the HEME metabolism pathway is the most elevated pathway in relation to an increased B[a]P level among the P1 asthmatic kids.

### 3.2. Hematopoietic cellular lineage is altered in relation to B[a]P in the P1 asthmatic children

Not only the B[a]P concentration, but also an elevated AMOD severity are significantly associated with altered hematopoietic cellular lineages in the asthmatic P1 children (Fig. 2A and B). Based on the CIBERSORT algorithm (Newman et al., 2015) and the purified hematopoietic lineage transcriptome (Novershtern et al., 2011), neutrophilic metamyelocytes (termed GRAN2 with markers  $\text{CD34}^- \text{SSC}^{\text{hi}} \text{CD45}^+ \text{CD11b}^+ \text{CD16}^-$ , see Table S1 in (Novershtern et al., 2011)) is the most positively correlated cell population with the B[a]P concentration among the P1 asthma cases (see Fig. S2G for a summary of the associated cell populations in the asthma and B[a]P exposure). Furthermore, the P1 asthmatics also exhibit B[a]P-correlated increase of the following inferred cell types:  $\text{CD8}^+$  T cell populations (TCELLA1:  $\text{CD8}^+ \text{CD62L}^- \text{CD45RA}^+$ , TCELLA2:  $\text{CD8}^+ \text{CD62L}^+ \text{CD45RA}^+$ ); Mature B Cell Class Switched (BCELLA4:  $\text{CD19}^+ \text{IgD}^- \text{CD27}^+$ ); and  $\text{CD71}^-$  low erythroid cells (ERY4:  $\text{CD34}^- \text{CD71}^{\text{lo}} \text{GlyA}^+$ ) (Fig. S2G).

Furthermore, an elevated B[a]P is associated with a decrease in the following inferred cell populations (Fig. S2G): Megakaryocyte / Erythroid Progenitor (MEP:  $\text{CD34}^+ \text{CD38}^+ \text{IL-3R}^- \text{CD45RA}^-$ ),  $\text{CD71}^+$  erythroids (ERY3:  $\text{CD34}^- \text{CD71}^+ \text{GlyA}^+$ ), eosinophils (EOS2:  $\text{FSC}^{\text{hi}} \text{SSC}^{\text{lo}} \text{IL-3R}^+ \text{CD33}^{\text{dim}+}$ ), Naïve  $\text{CD4}^+$  T Cells (TCELLA6:  $\text{CD4}^+ \text{CD62L}^+ \text{CD45RA}^+$ ), and monocytes (MONO2:  $\text{FSC}^{\text{hi}} \text{SSC}^{\text{lo}} \text{CD14}^+ \text{CD45}^{\text{dim}}$ ) among the P1 asthma cases, compared to the P1 controls (Table S3C). Overall, an altered distribution of the cell populations is significantly correlated with the module eigen-genes of several top ranked modules (Spearman  $\text{FDR} < 0.05$ , Fig. 2B).

The ambient B[a]P concentration is significantly associated with an increase in the relative abundance of  $\text{CD71}^-$  erythroid subpopulations among the P1 asthmatic children. While an elevated B[a]P is correlated with an increased  $\text{CD71}^{\text{low}}$  (ERY4), the same exposure is associated with a suppressed  $\text{CD71}^+$  (ERY3) and erythroid progenitor (MEP) cell abundance in the P1 asthma cases (highlighted in red box-Fig. 2B, Fig. S2G).

### 3.3. Top module, mediating the B[a]P-asthma association, is regulated by ZFAND6

Top molecular pathways, which are most robustly associated with the ambient B[a]P concentration and AMOD severity, at the same time, in the P1 children are summarized in Table S2 (see *Co-expression network analysis* in Materials and Methods). The gene modules are rank ordered by their enrichment of differentially expressed/methylated genes (DEGs/DMRs) and significantly correlated genes (SCGs) with

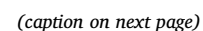


**Fig. 2.** Overview of CIBERSORT inferred hematopoietic lineage cell populations in CAP P1. **A.** Inferred cell populations of hematopoietic lineage. Main: X-axis is the blood samples from polluted region (P1), and y-axis is the relative abundance of individual cell types as -inferred by CIBERSORT algorithm. Distinct cell populations are labeled by different colors as shown in the legend on the left. The stacking order of bar graphs is identical to the legend order on the left. The top dendrogram orders samples by clustering the relative cell abundance matrix, and the bottom heatmap annotates clinical traits of individual samples. The clinical traits include (from top to bottom): asthma status within P1 (red: asthmatic kid, green: healthy kid), AEDS index, AMOD index,  $\ln(\text{cotinine}[\text{ng}/\text{mg}] \text{ in blood})$ , and  $\ln(\text{ambient B[a]P exposure})$ . The colour scales for the clinical traits are shown in the legends below. **B.** Distinct cell populations associate to gene modules, asthma and B[a]P exposure simultaneously. Main: X-axis is different cell populations, and y-axis is top 50 modules from P1 network, ranked by associations to asthma and B[a]P exposure. Heatmap colors show  $-\log_{10}(\text{FDR})$  of Spearman correlations between module eigen-genes and inferred cell populations, signed by correlation coefficient (red: positive, blue: negative; see legend on bottom right). Top: shows associations between cell populations (x-axis) and clinical traits (y-axis) in CAP P1. The heatmap colors show  $-\log_{10}(\text{FDR})$  of Spearman correlation, signed by correlation coefficient (pink: positive, green: negative) for continuous traits, and  $-\log_{10}(\text{FDR})$  of Wilcoxon rank sum test, signed by difference between asthmatic kids and healthy control in P1 (pink/green: greater/smaller cell population in asthmatic kids than healthy kids). Right: shows enrichments of DEGs and SCGs associated asthma and B[a]P exposure. Heatmap colors show  $-\log_{10}(\text{FET FDR})$  of signatures. Bottom: summarizes positive/negative associations between cell populations and clinical traits in upper/lower panel. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

respect to AMOD severity (Fig. 3B; Table 2; the section of *Module Ranking* in Materials and Methods).

The top ranked module, M14, is associated with B[a]P-driven alterations in transcriptome as well as epigenome. M14 is enriched for the negative SCGs of  $\ln(\text{B[a]P})$  (FET  $p = 2.24\text{E-}52$ , 15.0 FE) and hyper-methylated DMRs in the P1, compared to the P0 children (FET  $p = 2.95\text{E-}2$ , 2.4-fold enrichment). M14 is associated with the following inferred cell populations (Fig. 2B; Table S3C): elevated Granulocyte-Monocyte Progenitor (GMP:  $\text{CD}34^+ \text{CD}38^+ \text{IL-3R}\alpha^{\text{lo}} \text{CD}45\text{RA}^+$ ; Spearman  $\rho = 0.34$ , FDR  $p = 1.39\text{E-}4$ ) and increased Natural Killer T cells (NKA4, also known as NKT:  $\text{CD}14^- \text{CD}19^- \text{CD}3^+ \text{CD}1d^+$ ; Spearman  $\rho = 0.26$ , FDR  $p = 6.05\text{E-}3$ ).

The top regulators of M14 (Table 2), including *ZFAND6* (zinc finger AN1 type-6, also known as *AWP1*), are enriched for the genes down-regulated in the severe AMOD cases versus the controls in P1 ( $t$ -test nominal  $p = 7.90\text{E-}02$ ; 1.4-fold down-regulation; Fig. S3B). The *ZFAND6* centered network is validated *in-silico* by its significant enrichment of *ZFAND6* shRNA knockdown signature in A549 lung adenocarcinoma cells (FET  $p = 9.89\text{E-}3$ , 2.2 FE; Fig. S3D and E).

### 3.4. Second top-ranked module suggests B[a]P-driven HEME biosynthesis in P1 asthmatics

The second top ranked module, M2, is associated with mitochondrial HEME metabolism pathway (FET  $p = 2.39\text{E-}16$ ; 10.4 FE). Among the P1 children, M2 is further enriched for the genes positively correlated with B[a]P (FET  $p = 7.46\text{E-}27$ , 18.0 FE), and the DEGs up-regulated in the severe AMOD cases (FET  $p = 12.85\text{E-}13$ , 34.0 FE) with respect to the healthy controls. M2 contains two daughter modules, M25 and M26, which respectively have distinct features (Fig. 4A). M2 and M25 are positively correlated with  $\text{CD}71^{\text{low}}$  erythroid population (Spearman  $\rho$  (M2, ERY4) = 0.42 with FDR = 6.17E-7; Spearman  $\rho$  (M25, ERY4) = 0.40 with FDR = 2.16E-6), and negatively correlated with  $\text{CD}71^+$  erythroid population (Spearman  $\rho$  (M2, ERY3) =  $-0.31$  with FDR = 7.34E-4; Spearman  $\rho$  (M25, ERY3) =  $-0.31$  with FDR = 6.14E-4). In contrast, M26 is strongly correlated with neutrophilic metamyelocyte (Spearman  $\rho$  (M26, GRAN2) = 0.43 with FDR = 3.30E-7), in addition to  $\text{CD}71^{\text{low}}$  erythrocytes (Spearman  $\rho$  (M26, ERY4) = 0.34 with FDR = 1.14E-4).

M25 includes an established regulator of HEME metabolism, *ALAS2* (5-aminolevulinic acid synthase-2), as its hub gene. Increased *ALAS2* expression is not only positively correlated with B[a]P (Spearman's  $\rho = 0.29$ , nominal  $p < 1.74\text{E-}4$ ), but also with an elevated AMOD index (Spearman's  $\rho = 0.22$ , nominal  $p < 5.06\text{E-}03$ ; Fig. 4B). In addition, *ERAF* (also known as alpha hemoglobin stabilizing protein, *AHSP*) is positively correlated with B[a]P (Spearman's  $\rho = 0.39$ , nominal  $P < 1.74\text{E-}4$ ), as well as the AMOD index (Spearman's  $\rho = 0.27$ , nominal  $P < 3.70\text{E-}4$ ). *ALAS2* is validated *in silico* as a key driver of the module M2. Gene signature of *ALAS2* over-expression in bone marrow derived  $\text{CD}34^+$  hematopoietic stem cells from the CREED database (Wang et al., 2016) is enriched in the 3-layer neighborhood of

*ALAS2* (FET  $p = 3.35\text{E-}6$ , 8.3 FE). We replicated the over-expression of *ALAS2* in another independent, US asthma cohort (Yang et al., 2015). A group of US inner-city children with severe asthma demonstrated a significant over-expression of *ALAS2* ( $t$ -test  $p = 7.29\text{E-}02$ ; 1.4-fold increase; Fig. 4C). Collectively, B[a]P-driven activation of HEME synthesis in asthmatic children are correlated with increased mature erythroid subpopulations (i.e. high  $\text{CD}71^{\text{low}}$  & low  $\text{CD}71^+$  erythroid subpopulations) as well as an upsurge of the immature neutrophils (i.e. neutrophilic metamyelocyte).

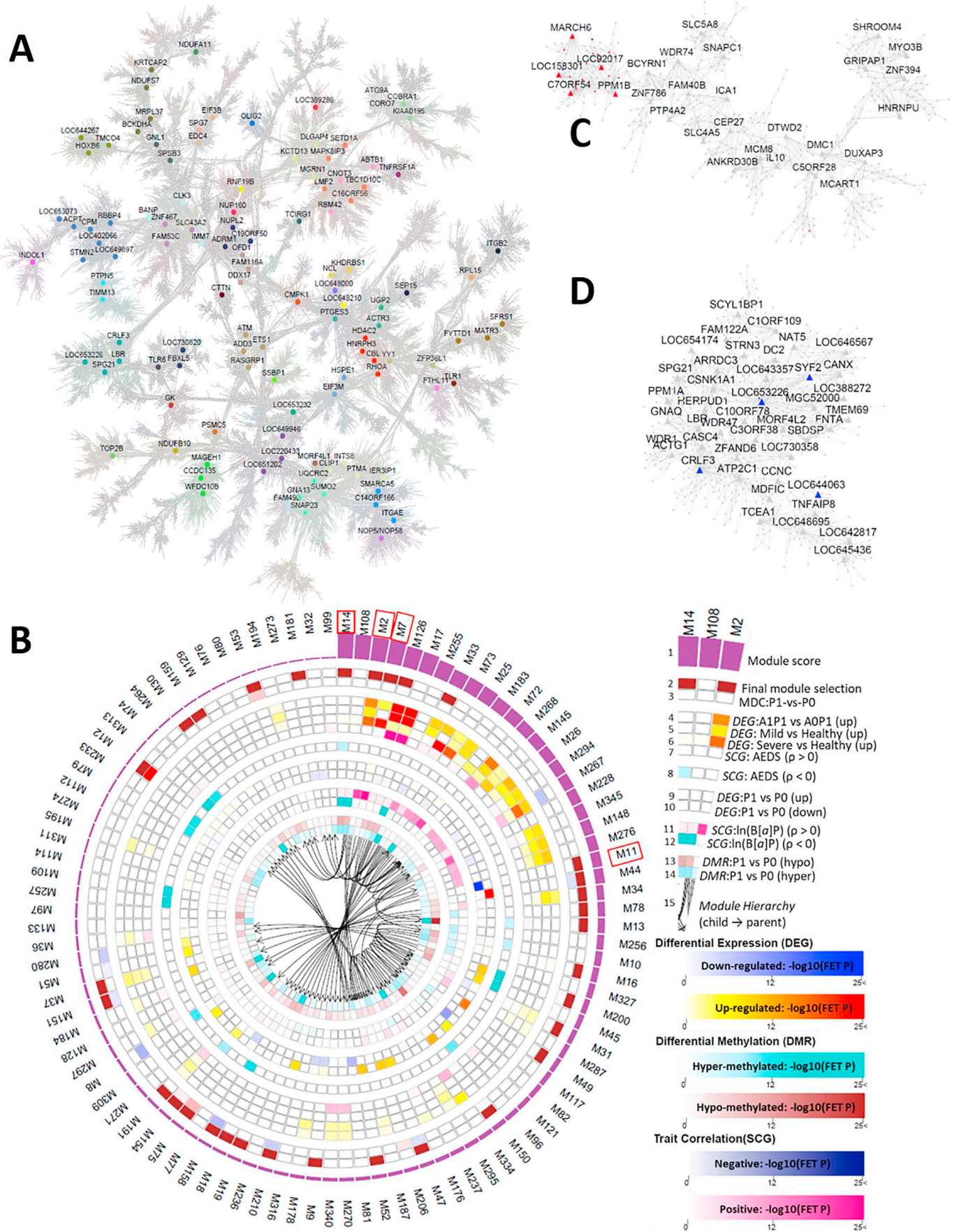
### 3.5. B[a]P negatively regulates NF- $\kappa$ B pathway, resulting in greater relative abundance of $\text{CD}8^+$ TEMRA population

Ambient B[a]P concentration is associated with an up-regulation of the third ranked module, M7, in the asthmatic children (Fig. 3C). Functionally, M7 is associated with an increased zinc finger protein transcription (nominal FET  $p = 4.22\text{E-}4$ ; 2.4 FE). The up-regulated DEGs in the severe P1 AMOD cases, compared to those in the controls, are enriched in M7 (FET  $p = 9.49\text{E-}33$ , 13.6 FE). In addition, the genes positively correlated with ambient B[a]P are enriched in M7 (FET  $p = 1.59\text{E-}17$ , 25.6 FE). M7 is associated with several cell subpopulations including, effective memory RA  $\text{CD}8^+$  T Cell (TCELLA1, also known as  $\text{CD}8^+$  TEMRA with markers  $\text{CD}8^+ \text{CD}62\text{L}^- \text{CD}45\text{RA}^+$ ; Spearman  $\rho = 0.25$ , FDR = 8.53E-3) and early B-cells (PRE\_BCELL2 with markers  $\text{CD}34^+ \text{CD}10^+ \text{CD}19^+$ ; Spearman  $\rho = 0.23$ , FDR  $p = 1.93\text{E-}2$ ). These data suggest that an exposure to high B[a]P concentration promotes an increased relative abundance of  $\text{CD}8^+$  TEMRA and early B-cell populations via M7 in the severe AMOD cases.

The hub gene of M7, *PPM1B* (Protein Phosphatase Magnesium Dependent 1B), is up-regulated in the P1 severe AMOD cases, compared to the P1 controls ( $t$ -test  $p = 1.79\text{E-}2$ ; 1.88-fold). *PPM1B* is known as a negative regulator of NF- $\kappa$ B pathway (Yien and Bieker, 2012). In addition, the other M7 hub genes are strongly correlated with the established NF- $\kappa$ B pathway genes, as curated in the BIOCARTE. Specifically, *IL10*, another hub genes in M7, is negatively correlated with known mediators of NF- $\kappa$ B activations, including *TRAF6* (Pearson's  $\rho = -0.61$ ,  $p = 3.83\text{E-}19$ ), *NFKB1* (Pearson's  $\rho = -0.47$ ,  $p = 3.64\text{E-}11$ ) and *MYD88* (Pearson's  $\rho = -0.45$ ,  $p = 6.38\text{E-}10$ ). Altogether, both B[a]P and the severe AMOD cases are concurrently associated with an up-regulation of *PPM1B* and *IL10*, the regulators of the NF- $\kappa$ B pathway.

### 3.6. B[a]P alters DNA methylation-mediated innate cell development in asthmatic children

Among the top 50 modules, M11 is subject to strong epigenetic regulation. M11 is enriched for the cis-meSCGs (FET  $p = 3.65\text{E-}3$ ; 19.3 FE; Fig. 5A-B) underlying pro-inflammatory innate cell response. This module is significantly correlated with greater relative abundance of the effective memory RA  $\text{CD}8^+$  T Cells (Spearman  $\rho = 0.43$ , FDR = 4.58E-7), the effective Memory  $\text{CD}8^+$  T Cell (Spearman  $\rho = 0.37$ , FDR = 2.48E-5), and a decreased relative abundance in



(caption on next page)

**Fig. 3.** MEGENA derived co-expression network of the P1 children. **A.** The global MEGENA network. The top predicted modules are labeled and circled. **B.** Circos representation of molecular features associated to top 100 gene modules (legends on the right). The heatmap scales are specific on the bottom right. The molecular features include - Track 1: Overall module score. Track 2: Final selection of modules by avoiding module overlaps. Track 3: Module Differential Connectivity (MDC) between P1 and P0. Tracks 4–7: up-regulated DEGs/SCGs in asthmatic and/or allergic kids compared to healthy kids in P1 (see track legend for details). Track 8: Negatively correlated SCGs with AEDS index. Tracks 9/10: up-regulated/down-regulated DEGs in P1, compared to P0. Tracks 11/12: positively/negatively correlated SCGs with ambient B[a]P exposure (i.e.  $\ln(B[a]P)$ ). Tracks 13/14: Hypo-/Hyper-methylated regions in P1, compared to P0. Track 15: The arrows depict module hierarchy from child modules to parent modules. **C/D.** Top up-/down-regulated modules in asthmatic children compared to the control children in P1 (C: M7, D: M108). Red/blue colors represent up-/down-regulated genes in asthmatic children. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

**Table 1**

Demographic characteristics and clinical outcomes of the case and the control children.

|                                      | Control (n = 183) <sup>a</sup> |                | Asthma (n = 178) <sup>a</sup> |                | p-Value <sup>b</sup> |
|--------------------------------------|--------------------------------|----------------|-------------------------------|----------------|----------------------|
|                                      | Median                         | Geometric Mean | Median                        | Geometric Mean |                      |
| benzo[a]pyrene (ng/m <sup>3</sup> )  | 3.83                           | 3.50           | 5.00                          | 5.11           | 0.001                |
| body mass index (kg/m <sup>2</sup> ) | 17.43                          | 17.92          | 18.27                         | 18.61          | 0.027                |
| Child's age (years)                  | 11                             | 10             | 11                            | 10             | 0.840                |
| Cotinine (μg/g creatinine)           |                                |                |                               |                | 0.678                |
| Non-smoker <sup>a</sup>              | 7.32                           | 7.26           | 6.45                          | 6.64           |                      |
| Passive smoker <sup>a</sup>          | 48.22                          | 51.71          | 46.80                         | 53.89          |                      |
|                                      | N                              |                | N                             |                |                      |
| Enrollment region                    |                                |                |                               |                | 0.954                |
| Rural                                | 93                             | (51%)          | 91                            | (51%)          |                      |
| Urban                                | 90                             | (49%)          | 87                            | (49%)          |                      |
| Gender (male)                        | 95                             | (52%)          | 111                           | (62%)          | 0.045                |
| mother smoker (yes)                  | 42                             | (23%)          | 48                            | (27%)          | 0.378                |
| father smoker (yes)                  | 78                             | (43%)          | 72                            | (40%)          | 0.557                |
| Sensitized to allergen (yes)         | 16                             | (8.2%)         | 138                           | (72.3%)        | < 0.001              |
| Atopic march onset score             |                                |                |                               |                | < 0.001              |
| Asymptomatic (0)                     | 175                            | (91%)          | 4                             | (2%)           |                      |
| Mild (1–2)                           | 16                             | (8%)           | 141                           | (74%)          |                      |
| Moderate (3)                         | 1                              | (1%)           | 32                            | (17%)          |                      |
| Severe (4–6)                         | 1                              | (1%)           | 14                            | (7%)           |                      |
| Age group                            |                                |                |                               |                | 0.956                |
| Pre-adolescent                       | 95                             | (49%)          | 93                            | (49%)          |                      |
| Adolescent                           | 99                             | (51%)          | 98                            | (51%)          |                      |
| Other diagnosed allergies (yes)      | 20                             | (10%)          | 191                           | (100%)         | < 0.001              |
| Atopic eczema ≤3 yr of age (yes)     | 5                              | (3%)           | 73                            | (38%)          | < 0.001              |
| Wheezing ≤3 yrs. of age (yes)        | 3                              | (2%)           | 127                           | (67%)          | < 0.001              |

<sup>a</sup> The children whose cotinine levels > 450 ng/mg are removed from the analysis, to preclude confounding. Sample size is reduced because 11 control and 13 asthmatic children are removed based on their active cigarette smoking status. Among the controls, 135 were non-smokers and 48 were passive-smokers. Among the asthmatics, 126 were non-smokers and 52 were passive-smokers.

<sup>b</sup> Jonckheere-Terpstra test was used on the continuous variable. Chi-square test used on the categorical variable.

CD56<sup>+</sup> mature Nature Killer (NK) cells (Spearman  $\rho = -0.44$ , FDR = 2.20E-7) (Table S3C). The activities of M11 genes include NK cell mediated cytotoxicity (FET  $p = 2.85E-18$ , 13.8 FE), and antigen processing and presentation (FET  $p = 9.39E-14$ , 14.6 FE).

Several cis-meSCGs within M11 are associated with the asthma outcome among the P1 children (Fig. 5A). For example, a significant hypo-methylation of *GZMH* (granzyme H), the hub of M11, and *GZMA* (granzyme A) are observed in the mild AMOD cases (t-test  $p = 5.43E-2$ , 0.90-fold decreased methylation). Granzymes (GZMs) represent a group of cytotoxic serine proteases released by NK T-cells to combat infective agents (Vahedi et al., 2014). The result suggests an activation of cytotoxic inflammatory response.

Compared to the P1 children, the P0 children are also under a significant, albeit weaker, epigenetic regulation of the genes in M11

(Kolmogorov-Smirnov (KS) test  $p = 5.60E-2$ ; Fig. 5E). For example, the expression of *CTSW* (Cathepsin W), an NK-cell specific marker of cytotoxic protease, is significantly correlated with *cis*-methylation at the CpG site in both P1 ( $p = 6.72E-05$ ,  $r = -0.415$ ) and P0 ( $p = 3.36E-03$ ,  $r = -0.311$ ) children, although the epigenetic suppression is stronger in the P1 children (the upper panels in Fig. 5D).

In addition, even though another module, M30, is not highly ranked in the P1 network, it is under robust epigenetic regulation. M30 is involved in the glutathione transferase activity (FET  $p = 4.27E-3$ , 49.8 FE). An elevated *cis*-methylation at the CpG site of *GSTM1*, *IL5RA*, *GATA1* and *GSTM3* is respectively associated with significant reduction of transcription of each of above genes (FET  $p = 7.60E-3$ , 28.2 FE). In particular, *GSTM1* is under significant epigenetic regulation in both the P1 children (Spearman  $\rho = -0.698$ , FDR < 1E-308) as well as the P0 children (Spearman  $\rho = -0.721$ , FDR < 1E-308; Table S4). The hub gene of M30, *IL5RA*, is a known biomarker for eosinophils. M30 is positively correlated with an elevated relative abundance of eosinophil population (Spearman  $\rho = 0.41$ , FDR = 1.56E-6). While *IL5RA* is *cis*-regulated in both P1 and P0 children (the lower panels in Fig. 5D), the epigenetic regulation is stronger in the P1 children (Spearman  $\rho$  (*IL5RA*, cg10159529) =  $-0.502$  with FDR of 8.8E-07; Spearman  $\rho$  (*IL5RA*, cg08404225) =  $-0.185$  with FDR of 8.39E-02), than in the P0 children (Spearman  $\rho$  (*IL5RA*, cg10159529) =  $-0.337$  with FDR of 1.41E-03; Spearman  $\rho$  (*IL5RA*, cg08404225) =  $-0.06$  with FDR of 0.529).

### 3.7. Asthma classification in the urban children using network drivers

We further examined the utility of B[a]P-associated hub genes as a predictor of asthma outcome. Using 10-fold cross-validation, the prediction accuracy of the B[a]P-associated asthma outcome is 74% (the expected accuracy by chance is 50%) for the P1 children, and 53% (the expected accuracy is 50%) for the P0 children. That is, the B[a]P-associated mechanisms among the P1 asthmatic children are distinct from those for the P0 asthmatic children.

## 4. Discussion

To date, the extent to which ambient air pollution exposure during childhood contributes to mechanistically distinguishable asthma has not been clarified (Dixon and Poynter, 2016). To our knowledge, the present data suggest for the first time that ambient B[a]P may: i) suppress NF- $\kappa$ B mediated inflammation by reducing NKT cells (M14), while promoting greater relative abundance of the IL10-secreting CD8<sup>+</sup> T cells (TEMRA: M7/M11) in the severe P1 AMOD cases; ii) promote mitochondrial heme biosynthesis, thereby driving the maturation of erythroid cells (i.e. elevated CD71<sup>low</sup> erythroids and reduced CD71<sup>+</sup> subpopulation) (Chen et al., 2009), as well as an increase in the relative abundance of neutrophilic metamyelocytes, again, most acutely in the asthmatic children with severe AMOD history; and iii) regulate glutathione transferase activity for eosinophil (M30) through epigenetic modification of transcription. Collectively, our data suggest that environmentally activated asthma via B[a]P exposure is mediated by a complex molecular network of multiple pathways, which are, in turn, regulated by several driver genes. As such, an effective asthma treatment strategy may require coordinated targeting of hub genes within multiple pathways.

**Table 2**  
Summary of key findings from gene modules associated to asthma and B[a]P exposure.

| Module/pathway                                                  | Altered cell populations                                                                                                                                                     | Top 15 key regulators (bold genes highlighted in main text)                                                                                                                                                                                                                                                 | Notes                                                                                                                                                            |
|-----------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Down-regulated by AMOD/B[a]P (M14)                              | $\rho > 0$ : GMP, NKT                                                                                                                                                        | <i>LOC653226</i> , <i>CRLF3</i> , <i>LBR</i> , <i>SPG21</i> , <i>ATP2C1</i> , <i>SBDSP</i> , <i>LOC648695</i> , <i>TCEA1</i> , <i>CASC4</i> , <i>LOC642817</i> , <i>CANX</i> , <i>TNFAIP8</i> , <i>LOC645436</i> , <b><i>ZFAND6</i></b> , <i>CCNC</i>                                                       | - Validated ZFAND6 as driver, Knock-down of ZFAND6 may inhibit NF- $\kappa$ B.                                                                                   |
| Up-regulated according to AMOD and/or B[a]P (M7)                | $\rho > 0$ : CD8 <sup>+</sup> TEMRA, early B-cells.                                                                                                                          | <i>PPM1B</i> , <b><i>IL10</i></b> , <i>ZNF394</i> , <i>LOC158301</i> , <i>CEP27</i> , <i>C7ORF54</i> , <i>LOC92017</i> , <i>MCART1</i> , <i>BCYRN1</i> , <i>SLC5A8</i> , <i>MARCH6</i> , <i>ICA1</i> , <i>HNRNP</i> , <i>SLC4A5</i> , <i>ANKRD30B</i>                                                       | - Negative correlation with NF- $\kappa$ B activators                                                                                                            |
| Elevated HEME according to AMOD and/or B[a]P (M2 with M25, M26) | $\rho > 0$ : CD71 <sup>low</sup> erythroids (M25), neutrophilic metamyelocyte (M26)<br>$\rho < 0$ : CD71 <sup>+</sup> erythroids (M25)                                       | <i>CTTN</i> , <i>ERAF</i> , <i>SH3BGR2</i> , <i>TSPAN9</i> , <i>C5ORF4</i> , <i>EPB42</i> , <i>HBD</i> , <i>ACRBP</i> , <i>ITGB5</i> , <i>CALD1</i> , <i>MUC6</i> , <i>MMD</i> , <i>DPYSL5</i> , <b><i>ALAS2</i></b> , <i>F13A</i>                                                                          | - More mature erythroid populations (CD71 <sup>low</sup> ) correlate to HEME synthesis,<br>- expansion of acute inflammation marker, neutrophilic metamyelocyte. |
| Epigenetic regulation (M11/M30)                                 | $\rho > 0$ : CD8 <sup>+</sup> TEMRA, CD8 <sup>+</sup> T-cells (M11), and eosinophils (M30)<br>$\rho < 0$ : CD56 <sup>+</sup> mature NK cells (M11), and early B-cells (M30). | - M11: <i>EDG8</i> , <i>GPR56</i> , <i>TGFBR3</i> , <i>FGFBP2</i> , <i>GZMH</i> , <i>KLRD1</i> , <i>PPP2R2B</i> , <i>GZMA</i> , <i>PRF1</i> , <b><i>NGK7</i></b> , <i>STAT4</i><br>- M30: <i>OLIG2</i> , <i>CCL23</i> , <i>SLC29A1</i> , <i>GPR44</i> , <b><i>IL5RA</i></b> , <i>CEBPE</i> , <i>SIGLEC8</i> | - NK-cell mediated cytotoxicity, up-regulation in mild/moderate AMOD (M11),<br>- Epigenetic regulation of eosinophil activity (M30) associates to GSTMs.         |

High ambient B[a]P concentration is associated with a coordinated decrease in the relative abundance of NKT through the NF- $\kappa$ B pathway (via M14), as well as the activation of anti-inflammatory IL10-secreting CD8<sup>+</sup> T cells among the asthmatic children. A decrease in the relative abundance of NKT may reflect anti-inflammatory mechanism within the severe P1 AMOD subset of the asthma cases. Earlier studies have shown that activated NKT cells could exacerbate airway hyper-reactivity and inflammation (Iwamura and Nakayama, 2010). Furthermore, high B[a]P exposure predicts higher relative abundance of mature CD71<sup>+</sup> erythroids as well as the CD8<sup>+</sup> T effective memory cells (TCELLA1/2) among the asthmatic children, compared to the controls. CD8<sup>+</sup> T effective memory cells have been noted as markers for severe asthma (Tsitsiou et al., 2012), and lung function decline in the asthma cases (den Otter et al., 2016).

In addition, B[a]P may contribute toward an elevated heme biosynthesis, which in turn, promotes concurrent increase of neutrophilic metamyelocyte and mature CD71<sup>low</sup> erythroid population among all AMOD cases. Key mechanisms of B[a]P and other polycyclic aromatic hydrocarbons (PAHs), may involve cytochromes P450 (CYP450s) induction (Li et al., 2003), destruction of mature erythrocytes, and an associated hemolytic anemia, as observed in oiled sea birds, mice, and humans, following PAH exposure (Booker and White, 2005; Kamal et al., 2014; Troisi et al., 2007). In our data, not only the mild AMOD, but also the moderate and severe AMOD cases within P1 are associated with increased HEME biosynthesis which is manifested through top asthma-related module M2 and its two primary hub genes, *ALAS2* and *ERAF*. *ALAS2* encodes an erythroid-specific mitochondrial enzyme, which catalyzes an initial step in the HEME biosynthetic pathway (Ferreira and Gong, 1995; Hunter and Ferreira, 2009). Therefore, increased heme biosynthesis might reflect host compensatory mechanism to generate more erythrocytes following B[a]P and other PAH exposure. In fact, anemia is a common comorbid condition in women with high indoor PAH exposure (Kamal et al., 2014, 2015), the asthmatics (Bener et al., n.d.; Ramakrishnan and Borade, 2010), or the pulmonary arterial hypertension patients (Lode et al., 2007). In particular, anemia could be induced by oxidative hemolysis of erythrocytes by circulating PAH metabolites, hemoglobin leakage, as well as an accumulation of Heinz bodies (Couillard and Leighton, 1993; Troisi et al., 2007). Such destruction might, in turn promote the HEME biosynthesis through an elevated *ALAS2* expression (Ferreira and Gong, 1995; Hunter and Ferreira, 2009).

Our observations support a growing body of evidence regarding putative B[a]P-associated asthma subtype and idiopathic pulmonary arterial hypertension (IPAH) as a commonly regulated disease (Achouh et al., 2008; Hayes Jr., 2007; Said et al., 2010). The gene expressions of *ALAS2* and *ERAF/AHSP* (also the key hub genes of M2 in this study) are commonly up-regulated in both the IPAH cases and our putative B[a]P-

associated asthma (observed in the present study), followed by subsequent red blood cell recruitment (Cheadle et al., 2012). Our data warrant a deeper investigation on whether B[a]P-associated asthma and IPAH both induce an injury in the airway as well as the vascular tissues.

Within the erythroid cells and megakaryocytes, heme biosynthesis is regulated by a master transcription regulator *GATA1* (An et al., 2014; Katsumura and Bresnick, 2017; Lee et al., 2017; Moriguchi and Yamamoto, 2014; Tanimura et al., 2016). In the children who are exposed to higher B[a]P, epigenetic dysregulation of the genes within the oxidative stress pathway (*GSTM1* and *GSTM3*) may contribute toward recruitment of eosinophils (suggested by an upregulation of *IL5RA*) by an altered expression of *GATA1*. Such scenario is supported in an experimental model data of pulmonary fibrosis (Soumyakrishnan et al., 2014). Cellular oxidative stress from reactive oxygen species (ROS) buildup or exhaustion of antioxidants could induce fibrogenesis by attracting inflammatory cells (e.g. eosinophils) or initiate signal transduction, leading to fibrogenesis (Soumyakrishnan et al., 2014). This is supported by another key hub gene of the module M7, *IL10*, and IL10-secreting CD8<sup>+</sup> T Cell. *IL10* released by CD8<sup>+</sup> T regulatory cells confer protective roles by reducing NF- $\kappa$ B activity in allergen-induced airway inflammation in asthmatics (Conti et al., 2003). In M7, such concerted upregulation of *PPM1B* and *IL10* suggests suppression of NF- $\kappa$ B mediated inflammation in severe AMOD cases. In a case-control study of Japanese children, *PPM1B* was associated with asthma exacerbation within the NF- $\kappa$ B pathway (Aoki et al., 2009). Earlier investigations of *PPM1B* and *IL10* proposed the two genes as respective negative regulators of NF- $\kappa$ B pathway (Yien and Bieker, 2012). The ambient B[a]P concentration is significantly correlated not only with AMOD outcome severity, but also with the degree of gene activations (e.g. *PPM1B*, *ZFAND6*, and *ALAS2*) across the top three molecular networks. Table 2 summarizes the key molecular processes within the top asthma-related gene subnetworks (modules). *ZFAND6* has been shown to activate NF- $\kappa$ B through TNF- $\alpha$ , by binding to *TRAF2* (Tumor necrosis factor receptor-associated factor 2) (Chang et al., 2011).

A robust body of evidence, including our own, have demonstrated B[a]P as an etiologically pertinent PAH compound (Baldacci et al., 2015; Guarnieri and Balmes, 2014b; Hew et al., 2015; Manners et al., 2014; Perzanowski et al., 2013). Both the laboratory and epidemiologic evidence has shown that B[a]P and other PAHs could induce or enhance allergic sensitization, exacerbate pre-existing asthma, and enhance the risk of de novo asthma development (Binkova et al., 2007; Li et al., 2016; Manners et al., 2014). In particular, B[a]P has been shown to directly target hematopoietic stem cells by binding to aryl hydrocarbon receptors, and subsequently impart a wide array of adverse effects, including mitochondrial functional deficit (van Grevenynghe et al., 2005).

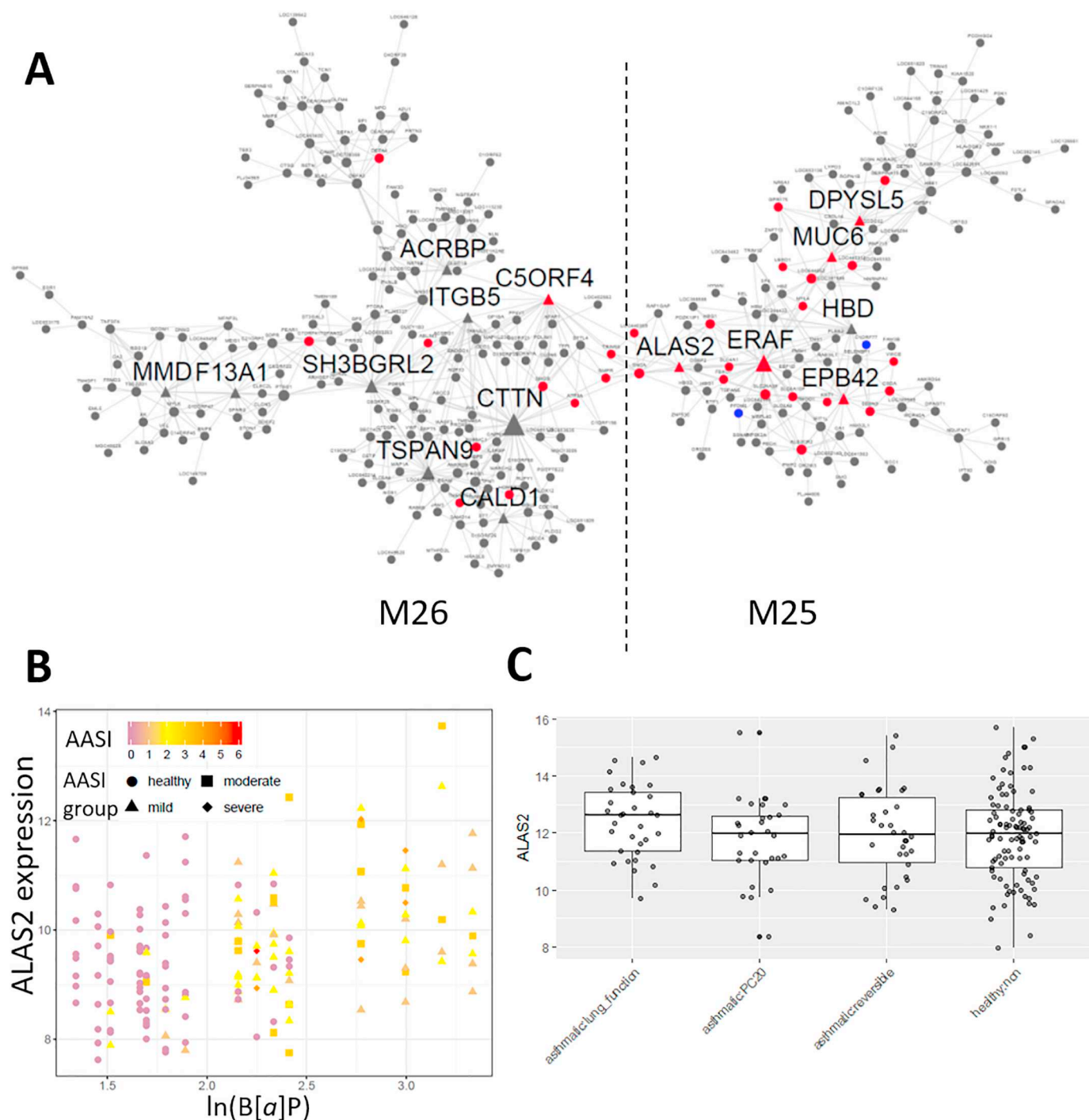


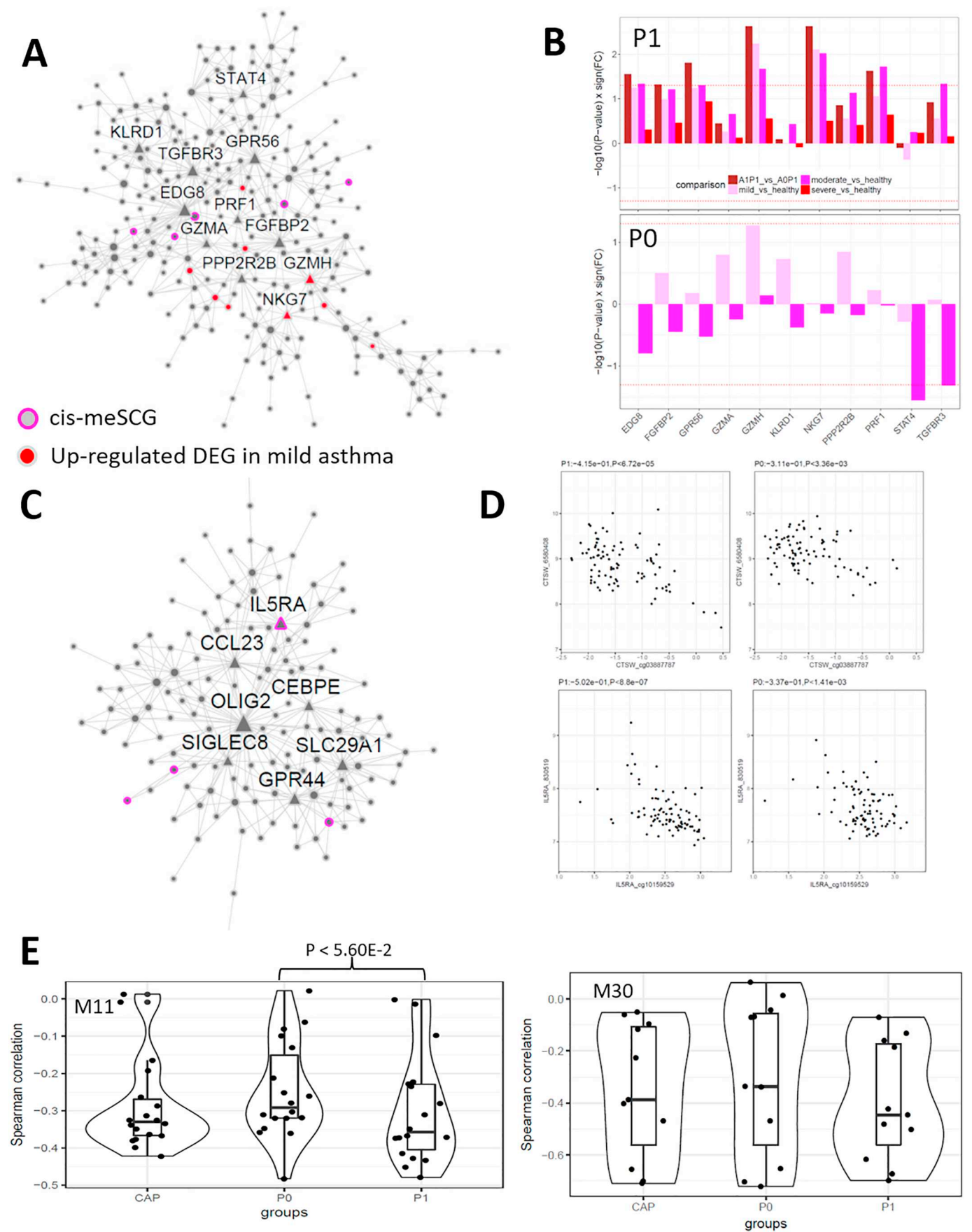
**Fig. 4.** Overview of HEME pathway module (M2). **A.** Network visualization of the module, M2. The subsets of M2 (i.e. M25 and M26) are labeled by a border. Red/blue dots represent up-/down-regulated genes in severe AMOD cases compared to the control children within P1. **B.** Scatter plot between *ALAS2* expression and  $\ln(B[a]P)$ . Colors and shapes represent AMOD index. **C.** Differential expression of *ALAS2* in asthmatic kids from an independent cohort ICAC cohort. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

5. Strengths and limitations

Strengths of this study include integration of unique regional B[a]P monitoring data with large multi-Omic data (i.e. DNA methylation and whole genome transcriptome) to identify B[a]P-associated molecular signatures and networks underlying asthma. Our unbiased network biology approach systematically predicts comprehensive sets of gene co-expression networks with varying degree of sensitivity. Our MEGENA approach is agnostic, and as such, is free of a priori bias. Therefore, the subnetworks and their driver genes warrant further

validation as host- and exposure-specific “meet-in-the-middle” bio-markers.

At the same time, there are several limitations with our study. First, our study is cross-sectional in design. Thus, the threat of reverse causality exists, because the health outcome was measured at the same time as the exposure assessment. To preclude reverse causality, we examined the B[a]P exposure and its risk on AMOD index, an alternate marker of pre-clinical asthma. Individual clinical diagnosis underlying the AMOD index was made at  $\leq 24$  months in age, prior to the current asthma diagnosis (except for lung function test and bronchodilation test).



(caption on next page)

**Fig. 5.** Overview of MEGENA modules enriched for cis-regulated gene expressions by methylation (namely, cis-meSCG signatures). **A/C.** Network visualization of highlighted modules M11 (A) and M30 (C). Nodes surrounded by magenta border are negatively correlated cis-meSCG gene signature, triangles denote MEGENA hub genes, and blue/red filled nodes indicated down-/up-regulated genes in polluted urban compared to background. **B.** Differential expression of hub genes of M11 across asthma severity groups in polluted urban (upper panel) and the background rural (lower panel). **D.** Examples of differential correlations between CpG probes and cis-meSCG signatures in polluted urban (left) and background rural (right). CTSW gene (upper panels) represents an example from M11, and *IL5RA* (lower panels) represents an example from M30. **E.** Correlations between gene expressions and respective cis-acting methylation probes (i.e. cis-meSCGs) within different groups: the urban, rural and the whole CAP cohort. The pairs of gene expression and cis-acting methylations are gathered by enforcing Bonferroni corrected p-value < 0.05 within the urban, rural and the whole CAP cohort separately. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Therefore, we believe that robust association, not only between B[a]P and positive AMOD index determined at  $\leq 24$  months, but also between B[a]P and current asthma diagnosis support primary causal contribution of B[a]P on asthma inception. Proof-of-principle studies is currently under way in both a prospective birth cohort and an experimental model system. Second, we did not directly measure the underlying cell population sizes between P1 and P0 children. Rather, we conducted cell-type deconvolution analysis, in which epigenetically dysregulated gene expression is postulated to alter the relative abundance of cellular subpopulations. Our analysis suggests that dysregulated molecular networks (and the altered distribution of the cell type subpopulations) are more likely to mediate the causal pathway, rather than merely confound the association. This is because the families, whose children have severe AMOD outcome, are unlikely to have chosen to live in districts with high ambient air pollution, following their AMOD diagnosis. Third, we used contemporaneous concentration of ambient B[a]P, rather than that during a suspected window of vulnerability (i.e. prenatal and first two years of life) due to the lack of available data in Ostrava. Fourth, other unmeasured, and correlated air pollutants (e.g. traffic-related air pollutants) may also pose causal effects, and thus pose a threat of residual confounding. The question of residual confounding by other correlated air pollutants is currently being addressed through another on-going prospective birth cohort study, which measures a mixture of pollutants. Fifth, ambient monitoring as primary marker of chronic exposure might have underestimated true exposure from other routes (e.g. oral and dermal). Our pilot analysis of another mother-newborn cohort ( $n = 40$ ) have shown that majority of personal exposure occurs through inhalation during the heating season (Pulkrabova et al., 2016). However, during the summer, major exposure (60–90%) occurs through dietary intake in both the polluted urban and the rural population (Pulkrabova et al., 2016). In newborns, maternal breast milk contribute by about 20–50% of the total PAH exposure (Pulkrabova et al., 2016). Finally, our use of an outdated technique (i.e. 27 K BeadChips) may have limited the sensitivity of full epigenetic regulation of the gene expressions.

## 6. Conclusion

In children with severe AMOD outcome, B[a]P appears to contribute to the suppression of NF- $\kappa$ B mediated inflammation by reduction of NKT cells and induction of IL10-secreting CD8<sup>+</sup> T effective memory cells. Higher heme biosynthesis within mitochondria by B[a]P promotes expansion of neutrophilic metamyelocyte and reduction of CD71<sup>+</sup> erythroid subpopulation. Such enhanced heme biosynthesis might reflect host compensatory mechanism to generate more erythrocytes following B[a]P and other PAH exposures. Furthermore, co-activation of erythroid cell specific *GATA1*, glutathione genes (*GSTM1* and *GSTM3*), and the Eosinophil marker genes (e.g. *IL5RA*) suggest that B[a]P may promote oxidative stress and Eosinophil recruitment. Our comprehensive multiscale network analysis of a large cohort of transcriptomic and epigenetic data in asthma uncovers the global landscape of molecular interactions in asthma as well as the detailed local regulatory circuits in B[a]P-induced asthma. The future work will seek to experimentally validate these predicted novel pathways and key regulators.

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.envint.2019.04.052>.

## Acknowledgements

The study is made possible by the support of The Czech Ministry of Education, Youth and Sport (InterAction programme LTAUSA 17182), The Czech Ministry of the Environment (SP1b35007), Ministry of Education(2B08005), Academy of Sciences Czech Republic (AVOZ50390703), and NIH/National Institute of Allergy and Infectious Diseases (U01AI111598). We thank Dr. Eva Schallerova, Ostrava-Radvanice, and the district of Prachatic for facilitating the collection of blood samples from the children. We thank Dr. David Lawrence for careful review of the manuscript. Drs. Helena Libalova and Elena Tulupova performed transcriptomic analysis. Andrea Rossnerova oversaw the processing and interpretation of the epigenetic data. Pavel Rossner, Jr. coordinated laboratory analysis and participated at gene expression and methylation data interpretation.

## Capsule summary

Median B[a]P concentration is correlated with an elevated asthma severity index only among the urban children. Molecular networks and predicted key molecular regulators underlying asthma are systematically uncovered. Severe cases in the urban region appear to experience higher energy catabolic demand, as indicated by the upregulation of heme metabolism. Furthermore, within the severe urban asthma cases epigenetically altered genes are involved in inflammation, oxidative stress, innate and adaptive (B- and T-cell) differentiation.

## Statement of author contributions

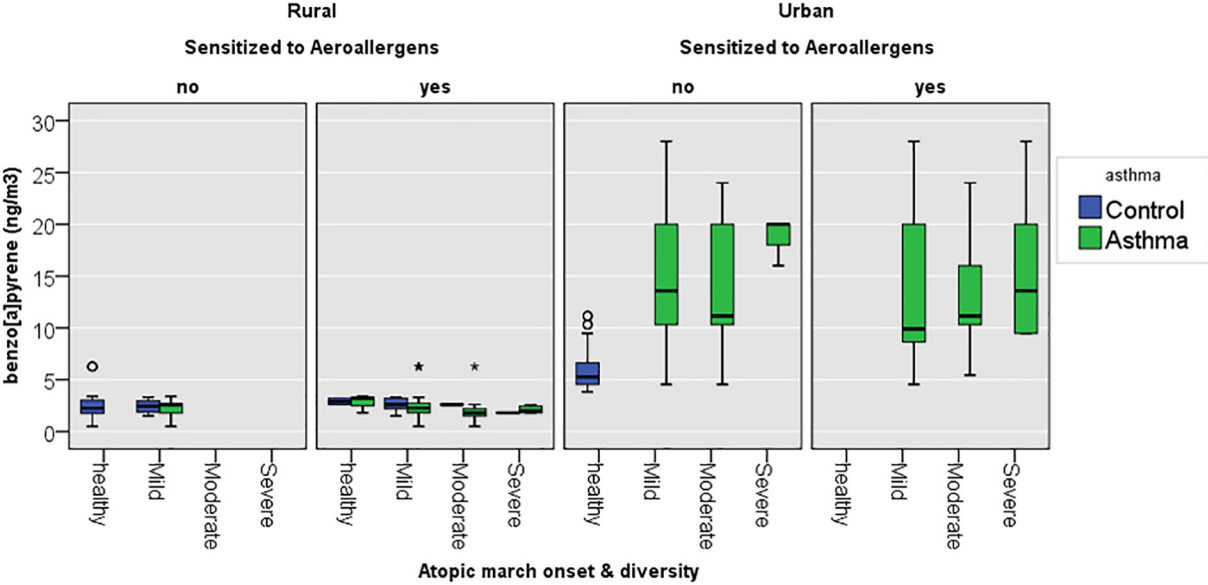
H. Choi and B. Zhang conceived the study. H. Choi conducted statistical analysis and wrote the manuscript. B. Zhang designed the framework of the integrative network analysis and wrote the manuscript. W. Song conducted the network analysis and wrote the manuscript. Miroslav Dostal and Anna Pastorkova organized sampling from pediatricians. Hans Gmuender analyzed statistical significance of genes. Radim J. Sram coordinated the study. All authors reviewed and approved the final manuscript.

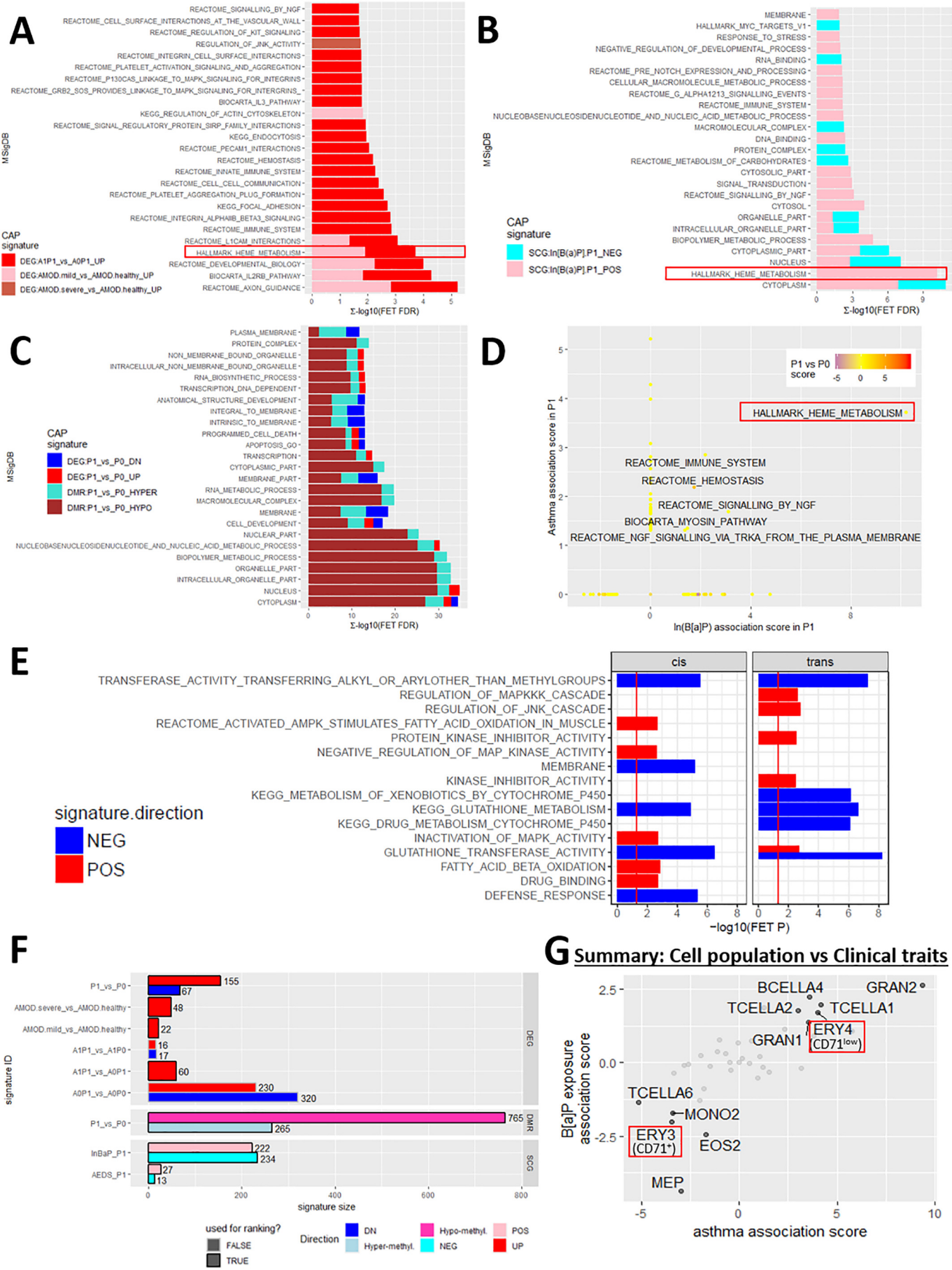
## References

- Achouh, L., Montani, D., Garcia, G., Jais, X., Hamid, A.M., Mercier, O., Simonneau, G., Humbert, M., 2008. Pulmonary arterial hypertension masquerading as severe refractory asthma. *Eur. Respir. J. Off. J. Eur. Soc. Clin. Respir. Physiol.* 32, 513–516. <https://doi.org/10.1183/09031936.00005408>.
- An, X., Schulz, V.P., Li, J., Wu, K., Liu, J., Xue, F., ... Gallagher, P.G., 2014. Global transcriptome analyses of human and murine terminal erythroid differentiation. *Blood* 123 (22), 3466–3477.
- Aoki, T., Matsumoto, Y., Hirata, K., Ochiai, K., Okada, M., Ichikawa, K., Shibasaki, M., Arinami, T., Sumazaki, R., Noguchi, E., 2009. Expression profiling of genes related to asthma exacerbations. *Clin. Exp. Allergy* 39, 213–221. <https://doi.org/10.1111/j.1365-2222.2008.03186.x>.
- Baldacci, S., Maio, S., Cerrai, S., Sarno, G., Baiz, N., Simoni, M., Annesi-Maesano, I., Viegi, G., Study, H., 2015. Allergy and asthma: effects of the exposure to particulate matter and biological allergens. *Respir. Med.* 109, 1089–1104.
- Bener, A., Ehlayel, M., Hamid, Q., n.d. The impact of anemia and hemoglobin level as a risk factor for asthma and allergic diseases YR - 2015/7/1. *Indian J. Allergy, Asthma*

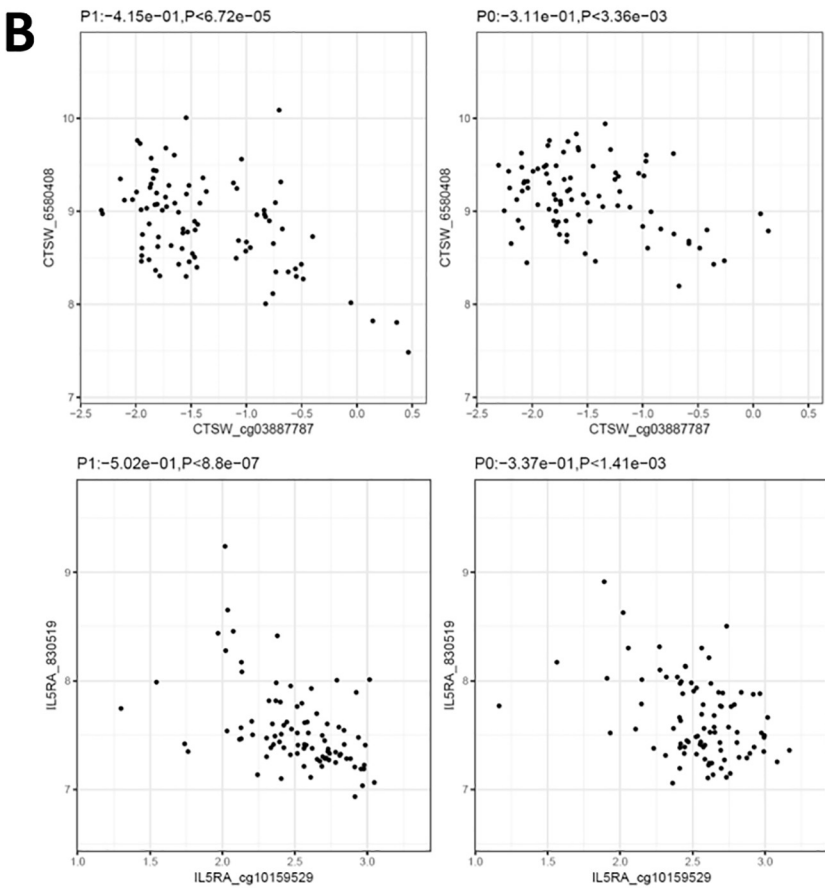
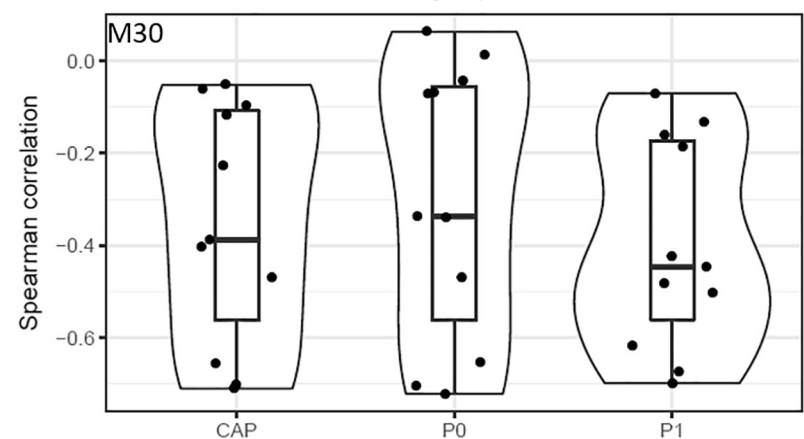
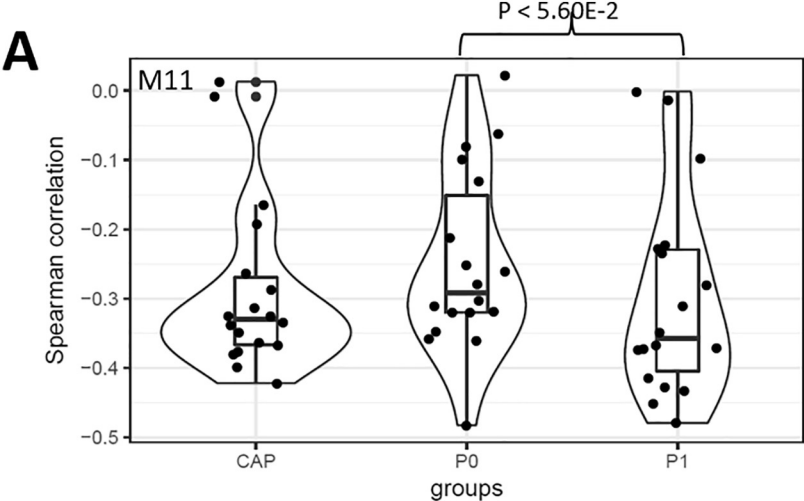
- Immunol. 72 OP-78 VO-29. doi:<https://doi.org/10.4103/0972-6691.178271>.
- Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate - a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Methodol.* 57, 289–300.
- Binkova, B., Cerna, M., Pastorkova, A., Jelinek, R., Benes, I., Novak, J., Sram, R.J., 2003. Biological activities of organic compounds adsorbed onto ambient air particles: comparison between the cities of Teplice and Prague during the summer and winter seasons 2000–2001. *Mutat. Res.* 525, 43–59.
- Binkova, B., Topinka, J., Sram, R.J., Sevastyanova, O., Novakova, Z., Schmuczerova, J., Kalina, I., Popov, T., Farmer, P.B., 2007. In vitro genotoxicity of PAH mixtures and organic extract from urban air particles part I: acellular assay. *Mutat. Res.* 620, 114–122.
- Booker, C.D., White, K.L., 2005. Benzo(a)pyrene-induced anemia and splenomegaly in NZB/WF1 mice. *Food Chem. Toxicol.* 43, 1423–1431. <https://doi.org/10.1016/j.fct.2005.03.018>.
- Centers for Disease Control and Prevention, 2014. Uncontrolled asthma among persons with current asthma [WWW document]. In: *Behav. Risk Factors Surveill. Syst. (BRFSS)—Child Adult Asthma Call-Back Surv. Data, 2006–2010*, URL: [https://www.cdc.gov/asthma/asthma\\_stats/uncontrolled\\_asthma.htm](https://www.cdc.gov/asthma/asthma_stats/uncontrolled_asthma.htm).
- Chang, E.J., Ha, J., Kang, S.S., Lee, Z.H., Kim, H.H., 2011. AWP1 binds to tumor necrosis factor receptor-associated factor 2 (TRAF2) and is involved in TRAF2-mediated nuclear factor-kappaB signaling. *Int. J. Biochem. Cell Biol.* 43, 1612–1620. <https://doi.org/10.1016/j.biocel.2011.07.010>.
- Cheadle, C., Berger, A.E., Mathai, S.C., Grigoryev, D.N., Watkins, T.N., Sugawara, Y., Barkataki, S., Fan, J., Boorgula, M., Hummers, L., Zaiman, A.L., Girgis, R., McDevitt, M.A., Johns, R.A., Wigley, F., Barnes, K.C., Hassoun, P.M., 2012. Erythroid-specific transcriptional changes in PBMCs from pulmonary hypertension patients. *PLoS One* 7, e34951. <https://doi.org/10.1371/journal.pone.0034951>.
- Chen, K., Liu, J., Heck, S., Chasis, J.A., An, X., Mohandas, N., 2009. Resolving the distinct stages in erythroid differentiation based on dynamic changes in membrane protein expression during erythropoiesis. *Proc. Natl. Acad. Sci.* 106 (17413 LP-17418).
- Cohen, J., 1988. *Statistical Power Analysis for the Behavioral Sciences*. Lawrence Erlbaum Associates, Hillsdale, NJ.
- Conti, P., Kempuraj, D., Kandere, K., Di Gioacchino, M., Barbacane, R.C., Castellani, M.L., Felaco, M., Boucher, W., Letourneau, R., Theoharides, T.C., 2003. IL-10, an inflammatory/inhibitory cytokine, but not always. *Immunol. Lett.* 86, 123–129.
- Couillard, C.M., Leighton, F.A., 1993. In vitro red blood cell assay for oxidant toxicity of petroleum oil. *Environ. Toxicol. Chem.* 12, 839–845.
- Czech Hydrometeorological Institute, 2008. B[a]P - Benzo[a]Pyrene Monthly and Annual Air Pollution Characteristics.
- den Otter, I., Willems, L.N.A., van Schadewijk, A., van Wijngaarden, S., Janssen, K., de Jeu, R.C., Sont, J.K., Sterk, P.J., Hiemstra, P.S., 2016. Lung function decline in asthma patients with elevated bronchial CD8, CD4 and CD3 cells. *Eur. Respir. J.* 48, 393–402.
- Dixon, A.E., Poynter, M.E., 2016. Mechanisms of asthma in obesity. Pleiotropic aspects of obesity produce distinct asthma phenotypes. *Am. J. Respir. Cell Mol. Biol.* 54, 601–608.
- Dostál, M., Milcová, A., Binková, B., Kotěšovec, F., Nožička, J., Topinka, J., Šrám, R.J., 2008. Environmental tobacco smoke exposure in children in two districts of the Czech Republic. *Int. J. Hyg. Environ. Health* 211, 318–325.
- Dostal, M., Pastorkova, A., Rychlik, S., Rychlikova, E., Svecova, V., Schallerova, E., Sram, R.J., 2013. Comparison of child morbidity in regions of Ostrava, Czech Republic, with different degrees of pollution: a retrospective cohort study. *Environ. Health* 12, 74.
- Ferreira, G.C., Gong, J., 1995. 5-Aminolevulinic synthase and the first step of heme biosynthesis. *J. Bioenerg. Biomembr.* 27, 151–159.
- Guarnieri, M., Balmes, J.R., 2014a. Outdoor air pollution and asthma. *Lancet* 383, 1581–1592. [https://doi.org/10.1016/S0140-6736\(14\)60617-6](https://doi.org/10.1016/S0140-6736(14)60617-6).
- Guarnieri, M., Balmes, J.R., 2014b. Outdoor air pollution and asthma. *Lancet* 383. [https://doi.org/10.1016/S0140-6736\(14\)60617-6](https://doi.org/10.1016/S0140-6736(14)60617-6).
- Hayes Jr., D., 2007. Idiopathic pulmonary arterial hypertension misdiagnosed as asthma. *J. Asthma* 44, 19–22. <https://doi.org/10.1080/02770900601125243>.
- Heinzel, L., Mari, A., Bergmann, K.-C., Bresciani, M., Burbach, G., Darsow, U., Durham, S., Fokkens, W., Gjomarkaj, M., Haahela, T., 2013. The skin prick test—European standards. *Clin. Transl. Allergy* 3, 3.
- Hew, K.M., Walker, A.I., Kohli, A., Garcia, M., Syed, A., McDonald-Hyman, C., Noth, E.M., Mann, J.K., Pratt, B., Balmes, J., Hammond, S.K., Eisen, E.A., Nadeau, K.C., 2015. Childhood exposure to ambient polycyclic aromatic hydrocarbons is linked to epigenetic modifications and impaired systemic immunity in T cells. *Clin. Exp. Allergy* 45, 238–248. <https://doi.org/10.1111/cea.12377>.
- Hunter, G.A., Ferreira, G.C., 2009. 5-aminolevulinic synthase: catalysis of the first step of heme biosynthesis. *Cell. Mol. Biol.* 55, 102–110.
- Iwamura, C., Nakayama, T., 2010. Role of NKT cells in allergic asthma. *Curr. Opin. Immunol.* 22, 807–813.
- Jedrychowski, W.A., Perera, F.P., Mauger, U., Mrozek-Budzyn, D., Mroz, E., Klimaszewska-Rembiesz, M., Flak, E., Edwards, S., Spengler, J., Jacek, R., Sowa, A., 2010. Intrauterine exposure to polycyclic aromatic hydrocarbons, fine particulate matter and early wheeze. Prospective birth cohort study in 4-year olds. *Pediatr. Allergy Immunol.* 21, e723–e732. <https://doi.org/10.1111/j.1399-3038.2010.01034.x>.
- Kamal, A., Malik, R.N., Martellini, T., Cincinelli, A., 2014. PAH exposure biomarkers are associated with clinico-chemical changes in the brick kiln workers in Pakistan. *Sci. Total Environ.* 490, 521–527. <https://doi.org/10.1016/j.scitotenv.2014.05.033>.
- Kamal, A., Cincinelli, A., Martellini, T., Malik, R.N., 2015. A review of PAH exposure from the combustion of biomass fuel and their less surveyed effect on the blood parameters. *Environ. Sci. Pollut. Res.* 22, 4076–4098.
- Katsumura, K.R., Bresnick, E.H., 2017. The GATA factor revolution in hematology. *Blood* blood-2016.
- Lee, W.Y., Weinberg, O.K., Pinkus, G.S., 2017. GATA1 is a sensitive and specific nuclear marker for erythroid and megakaryocytic lineages. *Am. J. Clin. Pathol.* 147, 420–426.
- Li, N., Sioutas, C., Cho, A., Schmitz, D., Misra, C., Sempf, J., Wang, M., Oberley, T., Froines, J., Nel, A., 2003. Ultrafine particulate pollutants induce oxidative stress and mitochondrial damage. *Environ. Health Perspect.* 111, 455–460.
- Li, N., Georas, S., Alexis, N., Fritz, P., Xia, T., Williams, M.A., Horner, E., Nel, A., 2016. A work group report on ultrafine particles (American Academy of Allergy, Asthma & Immunology): why ambient ultrafine and engineered nanoparticles should receive special attention for possible adverse health outcomes in human subjects. *J. Allergy Clin. Immunol.* 138, 386–396.
- Lode, H.N., Krings, G., Schulze-Neick, I., Dählöw, S., Schroeder, U., Bonnet, R., DaPalma, J., Luck, W., Strau, G., Berger, F., 2007. Pulmonary hypertension in a case of Hb-Mainz hemolytic anemia. *J. Pediatr. Hematol. Oncol.* 29, 173–177.
- Manners, S., Alam, R., Schwartz, D.A., Gorska, M.M., 2014. A mouse model links asthma susceptibility to prenatal exposure to diesel exhaust. *J. Allergy Clin. Immunol.* 134, 63–72 (e7).
- Moriguchi, T., Yamamoto, M., 2014. A regulatory network governing Gata1 and Gata2 gene transcription orchestrates erythroid lineage differentiation. *Int. J. Hematol.* 100, 417–424.
- Newman, A.M., Liu, C.L., Green, M.R., Gentles, A.J., Feng, W., Xu, Y., Hoang, C.D., Diehn, M., Alizadeh, A.A., 2015. Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* 12, 453.
- Novershtern, N., Subramanian, A., Lawton, L.N., Mak, R.H., Haining, W.N., McConkey, M.E., Habib, N., Yosef, N., Chang, C.Y., Shay, T., 2011. Densely interconnected transcriptional circuits control cell states in human hematopoiesis. *Cell* 144, 296–309.
- Nunes, C., Pereira, A.M., Morais-Almeida, M., 2017. Asthma costs and social impact. *Asthma Res. Pract.* 3 (1).
- Perera, F.P., 2017. Multiple threats to child health from fossil fuel combustion: impacts of air pollution and climate change. *Environ. Health Perspect.* 125, 141–148. <https://doi.org/10.1289/EHP299>.
- Perzanowski, M.S., Chew, G.L., Divjan, A., Jung, K.H., Ridder, R., Tang, D., 2013. Early-life cockroach allergen and polycyclic aromatic hydrocarbon exposures predict cockroach sensitization among inner-city children. *J. Allergy Clin. Immunol.* 131. <https://doi.org/10.1016/j.jaci.2012.12.666>.
- Pulkrabova, J., Lankova, D., Urbancova, K., Svarcova, A., Stupak, M., Hajslova, J., 2016. Assessment of body burden of Czech population to polycyclic aromatic hydrocarbons. In: *Environmental Technology and Innovations: Proceedings of the 1st International Conference on Environmental Technology and Innovations (Ho Chi Minh City, Vietnam, 23–25 November 2016)*. CRC Press, pp. 251.
- Ramakrishnan, K., Borade, A., 2010. Anemia as a risk factor for childhood asthma. *Lung India* 27, 51–53. <https://doi.org/10.4103/0970-2113.63605>.
- Rossnerova, A., Spatova, M., Rossner, P., Novakova, Z., Solansky, I., Sram, R.J., 2011. Factors affecting the frequency of micronuclei in asthmatic and healthy children from Ostrava. *Mutat. Res. Mol. Mech. Mutagen.* 708, 44–49.
- Rossnerova, A., Tulupova, E., Tabashidze, N., Schmuczerova, J., Dostal, M., Gmuender, H., Sram, R.J., 2013. Factors affecting the 27K DNA methylation pattern in asthmatic and healthy children from locations with various environments. *Mutat. Res. Mol. Mech. Mutagen.* 741, 18–26.
- Said, S.I., Hamidi, S.A., Gonzalez Bosc, L., 2010. Asthma and pulmonary arterial hypertension: do they share a key mechanism of pathogenesis? *Eur. Respir. J. Off. J. Eur. Soc. Clin. Respir. Physiol.* 35, 730–734. <https://doi.org/10.1183/09031936.00097109>.
- Samoli, E., Nastos, P.T., Paliatatos, A.G., Katsouyanni, K., Priftis, K.N., 2011. Acute effects of air pollution on pediatric asthma exacerbation: evidence of association and effect modification. *Environ. Res.* 111, 418–424.
- So, H.-C., Sham, P.C., 2011. Multiple testing and power calculations in genetic association studies. *Cold Spring Harb. Protoc.* (pdb-top95).
- Song, W.-M., Zhang, B., 2014. Multiscale Embedded Gene Co-expression Network Analysis. (Submitted).
- Song, W.M., Zhang, B., 2015. Multiscale embedded gene co-expression network analysis. *PLoS Comput. Biol.* 11, e1004574. <https://doi.org/10.1371/journal.pcbi.1004574>.
- Song, W.M., Di Matteo, T., Aste, T., 2012. Hierarchical information clustering by means of topologically embedded graphs. *PLoS One* 7, e31929. <https://doi.org/10.1371/journal.pone.0031929>.
- Soumyakrishnan, S., Divya, T., Kalayarasan, S., Sriram, N., Sudhandiran, G., 2014. Daidzein exhibits anti-fibrotic effect by reducing the expressions of proteinase activated receptor 2 and TGFβ1/smad mediated inflammation and apoptosis in bleomycin-induced experimental pulmonary fibrosis. *Biochimie* 103, 23–36. <https://doi.org/10.1016/j.biochi.2014.04.005>.
- Sram, R.J., Binkova, B., Demek, J., Chvatalova, I., Solansky, I., Topinka, J., 2006. Association of DNA adducts and genotypes with birth weight. *Mutat. Res.* 608, 121–128.
- Sram, R.J., Binkova, B., Dostal, M., Merkerova-Dostalova, M., Libalova, H., Milcova, A., Rossner Jr., P., Rossnerova, A., Schmuczerova, J., Svecova, V., 2013a. Health impact of air pollution to children. *Int. J. Hyg. Environ. Health* 216, 533–540.
- Sram, R.J., Dostal, M., Libalova, H., Rossner, P., Rossnerova, A., Svecova, V., Topinka, J., Bartonova, A., 2013b. The European hot spot of B[a]P and PM 2.5 exposure—the Ostrava region, Czech Republic: health research results. *ISRN Public Health* 2013.
- Stern, G., Latzin, P., Rössli, M., Fuchs, O., Proietti, E., Kuehni, C., Frey, U., 2013. A prospective study of the impact of air pollution on respiratory symptoms and infections in infants. *Am. J. Respir. Crit. Care Med.* 187, 1341–1348.
- Tanimura, N., Miller, E., Igarashi, K., Yang, D., Burstyn, J.N., Dewey, C.N., Bresnick, E.H., 2016. Mechanism governing heme synthesis reveals a GATA factor/heme circuit that controls differentiation. *EMBO Rep.* 17, 249–265. <https://doi.org/10.15252/embr.20151465>.

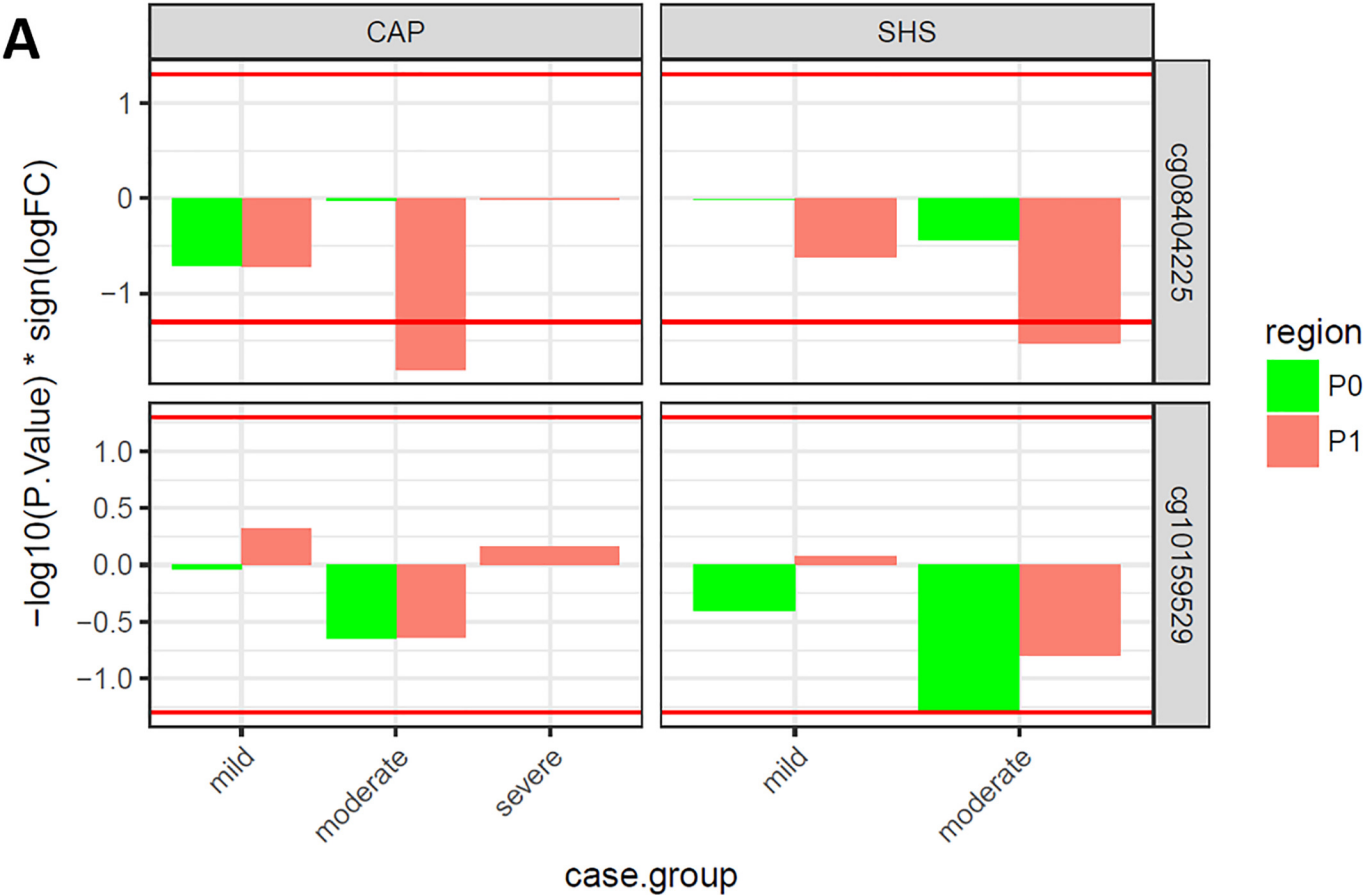
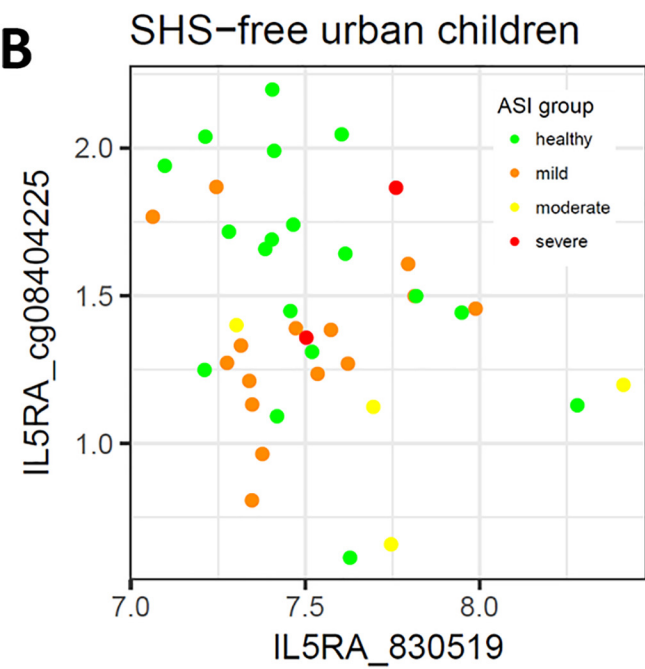
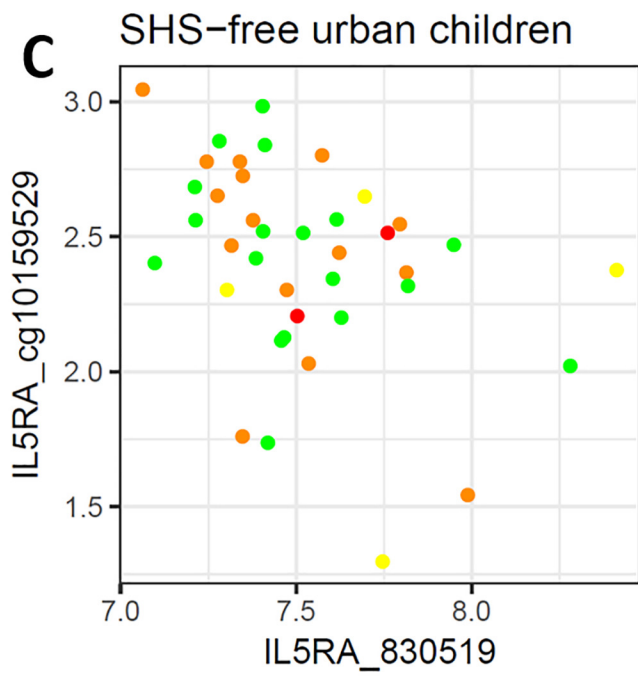
- Troisi, G., Borjesson, L., Bexton, S., Robinson, I., 2007. Biomarkers of polycyclic aromatic hydrocarbon (PAH)-associated hemolytic anemia in oiled wildlife. *Environ. Res.* 105, 324–329. <https://doi.org/10.1016/j.envres.2007.06.007>.
- Tsitsiou, E., Williams, A.E., Moschos, S.A., Patel, K., Rossios, C., Jiang, X., Adams, O.-D., Macedo, P., Booton, R., Gibeon, D., 2012. Transcriptome analysis shows activation of circulating CD8 + T cells in patients with severe asthma. *J. Allergy Clin. Immunol.* 129, 95–103.
- Tumminello, M., Aste, T., Di Matteo, T., Mantegna, R.N., 2005. A tool for filtering information in complex systems. *Proc. Natl. Acad. Sci. U. S. A.* 102, 10421–10426. <https://doi.org/10.1073/pnas.0500298102>.
- Vahedi, F., Fraleigh, N., Vlasschaert, C., McElhaney, J., Hanifi-Moghaddam, P., 2014. Human granzymes: related but far apart. *Med. Hypotheses* 83, 688–693.
- van Grevenynghe, J., Bernard, M., Langouet, S., Le Berre, C., Fest, T., Fardel, O., 2005. Human CD34-positive hematopoietic stem cells constitute targets for carcinogenic polycyclic aromatic hydrocarbons. *J. Pharmacol. Exp. Ther.* 314, 693–702 (<https://doi.org/jpet.105.084780> [pii]10.1124/jpet.105.084780).
- Velická, H., Puklová, V., Keder, J., Brabec, M., Malý, M., Bobák, M., Kotlík, B., Jirík, V., Janout, V., Kazmarová, H., 2015. Asthma exacerbations and symptom variability in children due to short-term ambient air pollution changes in Ostrava, Czech Republic. *Cent. Eur. J. Public Health* 23, 292.
- Wang, Z., Monteiro, C.D., Jagodnik, K.M., Fernandez, N.F., Gundersen, G.W., Rouillard, A.D., Jenkins, S.L., Feldmann, A.S., Hu, K.S., McDermott, M.G., Duan, Q., Clark, N.R., Jones, M.R., Kou, Y., Goff, T., Woodland, H., Amaral, F.M., Szeto, G.L., Fuchs, O., Schussler-Fiorenza Rose, S.M., Sharma, S., Schwartz, U., Bausela, X.B., Szymkiewicz, M., Maroulis, V., Salykin, A., Barra, C.M., Kruth, C.D., Bongio, N.J., Mathur, V., Todoric, R.D., Rubin, U.E., Malatras, A., Fulp, C.T., Galindo, J.A., Motiejunaite, R., Juschke, C., Dishuck, P.C., Lahl, K., Jafari, M., Aibar, S., Zaravinos, A., Steenhuizen, L.H., Allison, L.R., Gamallo, P., de Andres Segura, F., Dae Devlin, T., Perez-Garcia, V., Ma'ayan, A., 2016. Extraction and analysis of signatures from the gene expression omnibus by the crowd. *Nat. Commun.* 7, 12846. doi:<https://doi.org/10.1038/ncomms12846>.
- Yang, I. V. Pedersen, B.S., Liu, A., O'Connor, G.T., Teach, S.J., Kattan, M., 2015. DNA methylation and childhood asthma in the inner city. *J. Allergy Clin. Immunol.* 136. doi:<https://doi.org/10.1016/j.jaci.2015.01.025>.
- Yien, Y.Y., Bieker, J.J., 2012. Functional interactions between erythroid Kruppel-like factor (EKL/KLF1) and protein phosphatase PPM1B/PP2Cβ. *J. Biol. Chem.* 287, 15193–15204. <https://doi.org/10.1074/jbc.M112.350496>.
- Zhang, B., Horvath, S., 2005. A general framework for weighted gene co-expression network analysis. *Stat. Appl. Genet. Mol. Biol.* 4 <https://doi.org/10.2202/1544-6115.1128>. Article17.
- Zhang, B., Gaiteri, C., Bodea, L.-G., Wang, Z., McElwee, J., Podtelezchnikov, A.A., Zhang, C., Xie, T., Tran, L., Dobrin, R., Fluder, E., Clurman, B., Melquist, S., Narayanan, M., Suver, C., Shah, H., Mahajan, M., Gillis, T., Mysore, J., MacDonald, M.E., Lamb, J.R., Bennett, D.A., Molony, C., Stone, D.J., Gudnason, V., Myers, A.J., Schadt, E.E., Neumann, H., Zhu, J., Emilsson, V., 2013. Integrated systems approach identifies genetic nodes and networks in late-onset Alzheimer's disease. *Cell* 153, 707–720.









**A****B****C**

## Materials and Methods

Subject Enrollments and Study Design: The subject enrollment and laboratory analyses procedures have been described (Rossnerova et al., 2013, 2011, Sram et al., 2013a, 2013b). A historically industrial city of Ostrava (*i.e.* defined as P1) and three Southern Bohemia villages (*i.e.* defined as P0) were chosen as our study sites, based on our prior investigations of early-childhood allergy and respiratory outcomes (Dostal et al., 2013; Sram et al., 2013b, 2013a). Benzo[*a*]pyrene (B[*a*]P) was chosen as a representative compound for PAHs as well as particulate matter < 2.5  $\mu\text{m}$  in aerodynamic diameter (PM<sub>2.5</sub>) due to its demonstrated biological activities (Binkova et al., 2003; Sram et al., 2006). Within our target district (Radvanice/Bartovice), mean ambient level of B[*a*]P is more than four-times higher (11.4 ng/m<sup>3</sup>) than that for the P0 province (2.5 ng/m<sup>3</sup>) during November 2008 (*i.e.* transitional season of interest) (Sram et al., 2013b). The transitional season in Czech Republic has been associated with not only a large spatial, but also temporal, variability in ambient PAHs (Sram et al., 2013b). Major sources of air pollution in Ostrava include the power plants and coke ovens for various industries, residential heating, and automobile exhaust (Sram et al., 2013b). In contrast, southern Bohemia does not have any industrial power plants, but is characterized by residential heating and mobile exhaust emissions (Rossnerova et al., 2009). Briefly, a case-control study was conducted during November 2008. As all children in Czech Republic are assigned a pediatrician, we enlisted all pediatricians and the nurses within the target districts, who in turn, invited the parents of the eligible children to answer the questionnaires (Dostal et al., 2013). Historical participation rate by pediatricians to our studies has been > 97% (Hertz-Picciotto et al., 2007). An equal number of case (n=191) and control (n=194) children were enrolled from the polluted urban center within historically industrial city (*i.e.* Ostrava, defined as P1) and rural background region (*i.e.* Southern Bohemia, defined as P0), respectively, Czech Republic (Sram et al., 2013b). The asthma and allergy diagnoses were made by pediatricians, and subsequently confirmed by allergists. The pediatricians also provided International Classification of Diseases, Tenth Revision (ICD-10)-based clinical diagnoses data for each child. We defined the child as an asthma case if he/she met all the following conditions: 1) positive clinical diagnosis of ‘current’ asthma using the ICD-10 codes; 2) currently receive asthma medication; 3) positive history of bronchodilatation test result during the last 12 months; and 4) positive response (wheal size  $\geq 3$  mm diameter) based on a skin tests (PhadiatopH, Pharmacia & Upjohn Diagnostics, Uppsala,

Sweden) for airborne allergens (including hazel (*Corylus avellana*), alder (*Alnus incana*), birch (*Betula alba*), plane (*Platanus vulgaris*), cypress (*Cupressus sempervirens*), grass mix (*Poa pratensis*, *Dactylis glomerata*, *Lolium perenne*, *Phleum pratense*, *Festuca pratensis*, *Helictotrichon pratense*), Olive (*Olea europaea*), mugwort (*Artemisia vulgaris*), ragweed (*Ambrosia artemisiifolia*), *Alternaria alternata* (*tenuis*), *Cladosporium herbarum*, *Aspergillus fumigatus*, *Parietaria*, cat, and dog, *Dermatophagoides pteronyssinus*, *Dermatophagoides farinae*) (Heinzerling et al., 2013). The duration (in months) of corticoid medication was recorded. Diagnostic reliability of ICD-10 diagnoses in physician visits or hospitalizations in our target regions have been demonstrated (Hertz-Picciotto et al., 2007). Parents of the children also answered a questionnaire on the children's medical history, exposures, and lifestyles. Urinary cotinine (ng/mg Creatinine) level is analyzed by radioimmunoassay (Rossnerova et al., 2011). Children whose cotinine concentration ranged between 20 and 449 ng/mg Creatinine are considered secondhand smokers (Dostál et al., 2008). The unexposed children are defined as those with cotinine < 20 ng/mg Creatinine. Actively smoking case and control children (i.e. >450 ng/mg of Creatinine) were removed from the analysis (Dostál et al., 2008). The control is defined as a child who is free from all of the above conditions. The ethics committee of the Institute of Experimental Medicine, Academy of Sciences, approved the study. The parents signed an informed consent.

Polycyclic Aromatic Hydrocarbon Monitoring: Routinely monitored ambient B[a]P data from four monitoring sites within Ostrava and four sites in Southern Bohemia were obtained from The Czech Hydrometeorological Institute (CHMI) monitors airborne eight PAHs and PM<sub>2.5</sub> using a Versatile Air Pollution Sampler (VAPS) as described for November 2008 period ("Air Pollution in Czech Republic: Czech Hydrometeorological Institute," n.d.). Per CHMI protocol, the mean daily level is taken once every three days for total of 10 days/month in southern Bohemia; once every six days for total of 5 days/month in Ostrava. Data quality assurance and control are described (Ghosh et al., 2016; Hertz-Picciotto et al., 2007; Rossner et al., 2012).

Gene expression analysis: Full summary of the procedure can be found on the online supporting material. Whole blood samples are processed using PAXgene Blood RNA Kit (PreAnalytiX). HumanHT-12 v3 BeadChips (Illumina) with probes for more than 48K transcripts (total available samples = 368). The isolated RNA (200 ng) is reverse transcribed, amplified and

biotinylated by in vitro transcription (Illumina TotalPrep RNA Amplification Kit, Ambion, Foster City, CA, USA). The microarrays are processed by BeadStation 500 (Illumina) and interpreted using BeadStudio Software (version 3.3.7, Illumina). Quality control and normalization of the data are performed in R statistical environment (<http://www.r-project.org>) using Limma package.

Infinium Methylation Assay: Genomic DNA was isolated from whole venous blood collected in EDTA tubes (Rossnerova et al., 2013). Bisulfite-converted DNA is prepared through enzymatic fragmentation, precipitation, resuspension, and hybridization followed by washing and BeadChip staining. A batch of 12 samples are processed with Infinium Human Methylation 27K BeadChips (Illumina, catalog number WG-311-2201).

Atopic March Onset and Diversity (AMOD) index: We developed a modified prognostic index of asthma, based on the children's medical history during the first two years of life. Each respective diagnosis of atopic dermatitis, allergic sensitization, wheezing symptoms, at  $\leq 24$  months, as well as the positive result of the spirometry test were given a score of 1. The composite AMOD index for each child is coded as: 0 for the asymptomatic children. A child with allergic sensitization only (category 1) or doctor-diagnosed atopic dermatitis at  $\leq 24$  months, in addition to the sensitization (category 2) were defined as a 'mild' case. A child with allergic sensitization, atopic dermatitis diagnosis, as well as wheezing at  $\leq 24$  months were deemed 'moderate' (category 3). Finally, a child was deemed as a 'severe' case if he/she was allergen sensitized, had an early-onset atopic dermatitis diagnosis, and also had a positive bronchodilation result (category 4); or also had an additional diagnosis of wheezing at  $< 24$  months (category 5) or had an additional diagnosis of allergic rhinitis (category 6). Jonckheere–Terpestra (JT) test is used to non-parametrically assess a linear trend, without assuming an underlying normal distribution at  $\alpha = 0.05$ .

Power Analysis: After exclusion of active smokers and others with a poor quality control score, the power analysis was conducted on 174 (P1) and 156 (P0) subjects, based on Jacob Cohen's work (Cohen, 1988). For two-tailed correlation test, a minimum power of 0.90 corresponds to correlation coefficient of 0.242 at a significance level  $< 0.05$  with 174 children in P1. As our minimum observed correlation coefficient in P1 co-expression network is 0.270 with FDR  $< 0.05$ ,

our power well exceeds 0.90 for detection of significant interactions among the genes. Furthermore, when we consider an ordinal outcome (i.e. AMOD index), our power is even higher, thereby reducing the minimum number of required children within each cell.

For example, let us consider the case of T-test. Under the normality assumption, the effect size is analogous to z-score (i.e. mean difference divided by pooled standard deviation) (Kim, 2015). Noting that the conventional p-value  $< 0.1$  corresponds to z-score  $> 1.65$  in two-tailed T-test, we evaluated the statistical power for an effect size of 1.5. Among the P1 children with the significance level of 0.05 and preset power of 0.9, it requires 4 children in the sparsest cell (relative to 83 healthy controls). Our observed number of children in the sparsest cell were 7 with 'severe' AMOD outcome; 18 for the moderate; 56 for the mild; and 83 for the healthy children, respectively, within P1. Furthermore, in a combined analysis of P1 (174 children) and P0 (156 children), the power associated with an effect size of 1.5 and a significance level at 0.05 for two-tailed T-test is  $> 0.99$ . Similarly, our power to detect a true association for methylation data analysis is  $> 0.98$ , based on an effect size of 1.5, significance level at 0.05 for two-tailed T-test. Such power level is based on 92 children in P1 and 96 children in P0, respectively.

Differentially Expressed Gene and Differentially Methylated Gene Analysis: An integrative analysis of ambient B[a]P, transcriptome, epigenome, asthma or AMOD index was conducted on the entire cases and the controls; and separately for the urban and the rural children. First, a series of quality control operations are performed on the gene expression and methylation data. Altered gene pathways and functions are identified through enrichment tests of *Differentially Expressed Genes (DEGs)* and *Significantly Correlated Genes (SCGs)*. DEGs are identified by comparing the following groups: 1) all urban (i.e. **P1**) vs. all rural (i.e. **P0**) children; 2) all asthmatic (**A1**) vs. non-asthmatic (**A0**) children; and 3) all urban asthmatic (**A1P1**) vs. the urban control children (**A0P1**). SCGs are identified by correlating the gene expression against continuous variables (e.g. B[a]P and AMOD). DEGs are identified by R package limma version 3.28.17, with Benjamini-Hochberg False Discovery Rate (BH FDR)  $< 0.05$  and expression fold change (FC)  $> 1.2$ .

Differentially Methylated Region (DMR) are detected using the limma package with BH FDR  $< 0.05$  and FC  $> 1.2$  by comparing P1 and P0 children. We also leveraged the mapping between

CpG sites and respective proximal genes annotated by Infinium Human Methylation 27K BeadChips to identify potential cis-regulated gene expressions by DMR. Methylation correlated gene signatures (meSCGs) are identified by Spearman's correlation analysis between all pairs of methylation profiles (at cis-/trans-CpG sites) and gene expressions with BH FDR < 0.05. In contrast to DMR, meSCGs captures epigenome-transcriptome interactions within P1 or P0. Resulting DEG, DMR, SCG, and meSCG signatures are functionally annotated using Molecular Signature Database (MSigDB). All statistical and bioinformatics analyses are performed using the R statistical package version 3.3.1 ([www.r-project.org](http://www.r-project.org)) (see **Figure S2F** for the gene signature used in the analysis).

Co-expression network analysis: Gene co-expression networks are constructed using Multiscale Embedded Gene Co-expression Network Analysis (MEGENA) method (Song and Zhang, 2014, 2015; Song et al., 2012). Briefly, gene pairs with significant correlations (FDR < 0.05) were embedded as geodesic paths onto a 3-dimensional topological sphere (Tumminello et al., 2005). Topology of the resulting “planar filtered networks (PFNs)” is analyzed in an unsupervised fashion to identify network clusters (*i.e.* gene modules) at multiple resolutions (Song and Zhang, 2014, 2015; Song et al., 2012). The output of MEGENA is a set of co-expressed gene modules (*i.e.* subnetworks) and their key regulators. Relevance of these gene modules to asthma etiology is evaluated by testing the enrichment of the DEGs and SCGs using Fisher's Exact Test (FET). We projected a compendium of DEGs, SCGs and other gene signatures to the co-expression network to identify significantly enriched gene modules. Significantly enriched gene modules suggest altered according to the respective clinical variable and/or the ambient B[a]P concentration.

Conservation of co-expression modules using independent data sets: Overlap between two modules  $m_i$  and  $m_j$ , denoted as  $m_i \cap m_j$ , is calculated by Jaccard index,  $J_{ij} = |m_i \cap m_j| / |m_i \cup m_j|$  (Handl et al., 2005). Two modules are conserved if Bonferroni adjusted FET  $P < 0.05$  and  $J_{ij} > 0.2$  (*i.e.*, overlap is above 20% of the union of the genes in two modules).

Module ranking: In order to avoid over-/under-splitting the genes into modules, we restricted the module ranking to those within the range of 50 to 2000 genes. Relative importance of a gene module is defined as the module's association with asthmatic outcomes and calculated as below:

$$MS_m = \sum_{t=1}^T -\log_{10}(p_t),$$

where,  $m = 1, \dots, M$ ,  $M$  = total number of modules,  $T$  = the total number of DEG and SCG signatures. We chose gene signatures associated with asthma diagnosis and ambient B[a]P concentration to calculate the module ranking score. Specifically, the signatures include: DEGs between A1P1 and A0P1 (DEG: A1P1 vs A0P1), DEGs between mild asthma and healthy children in P1 (DEG: AMOD-mild vs AMOD-healthy), DEGs between severe asthma and healthy children in P1 (DEG: AMOD-severe vs AMOD-healthy), DEGs between P1 and P0 (DEG: P1 vs P0), DMRs between P1 and P0, and SCGs correlated to natural-log of ambient B[a]P exposure (SCG:  $\ln(B[a]P)$ ). Upon ranking the modules, we further refined the final set of modules by removing overlapping modules. This resulted in the final selection of modules in track 2, **Figure 2A**.

Module Differential Connectivity (MDC) analysis: MDC for a given module is the ratio of the mean connectivity of the module's member genes in one network to that of the same set of genes in another network (Zhang et al., 2013).

Comparison of methylation-gene expression correlations between P1 and P0: For each module, the significantly correlated transcript and CpG probe pairs (Benjamini-Hochberg corrected Spearman correlation  $P < 0.05$ ) are identified. The distributions of the correlation coefficients across different groups are compared using the Kolmogorov-Smirnov (KS) test.

In-silico validation of gene-gene interactions: We validated our putative network structures against the established gene perturbation signatures derived from The Library of Network-Based Cellular Signatures (LINCS) and CRowd Extracted Expression of Differential Signatures (CREEDS) (Wang et al., 2016). For LINCS, we selectively analyzed the signatures generated from lung tissue specific cell-lines, namely HCC515 and A549. For CREEDS, we used overall gene perturbation signatures, which intersected with MEGENA hub genes. These gene perturbation signatures are projected onto the neighborhoods of the corresponding genes in each PFN for enrichment test. The network structure around a gene is successfully validated if the BH corrected FET  $P < 0.05$ . For example, we validated the network interactions shown in **Figure S3E** by projecting *ZFAND6* knock-down signatures in A549 cells from LINCS databases (see

**Figure S3D** for enrichments). For CREED, the knock-down signatures obtained from GSE22366 for primary erythroid progenitor cells were utilized, and for LINCS, the knock-down signatures from lung tissue cell lines (A549 and HCC515).

Calculation of Asthma/B[a]P exposure association scores in Figure S2D: B[a]P association score per MSigDB signature was calculated by  $-\log_{10}(\text{FET FDR } p)$  of positive and negative SCGs of  $\ln(\text{B[a]P})$  in P1 (denoted as  $S(\text{B[a]P}^+)$  and  $S(\text{B[a]P}^-)$ ), and summarizing as an aggregate score,  $S(\text{B[a]P}^{\text{all}}) = S(\text{B[a]P}^+) - S(\text{B[a]P}^-)$ .  $S(\text{B[a]P}^{\text{all}}) > 0$  denotes enrichment of the positive SCGs, and  $S(\text{B[a]P}^{\text{all}}) < 0$  denotes enrichment of negative SCGs. Similarly for asthma association scores, up-regulated scores for asthma is defined by  $S(\text{asthma}^+) = \sum_t -\log_{10}(\text{FET FDR}_t)$  where  $t$  denotes up-regulated signatures in the comparisons: A1P1 vs A0P1, mild asthma vs healthy control in P1, and severe asthma vs healthy control in P1 as shown in the legend of **Figure S2A**. Note that there were not significantly down-regulated genes with  $\text{FDR} < 0.05$  in these comparisons, therefore  $S(\text{asthma}) = S(\text{asthma}^+)$ .

Analysis of CIBERSORT inferred cell population with clinical traits: We performed Spearman's correlation analysis between CIBERSORT inferred relative cell populations and the continuous variables (e.g. B[a]P concentration). We performed non-parametric Wilcoxon rank sum test of the ordinal variables (e.g. AMOD index) with the relative cell populations. As shown in the top panel of Figure 2A, the pink heatmap colors represent greater (i.e. proliferation) cell abundance, whereas the green heatmap color represents smaller (i.e. suppression) proportion of respective cell populations, with respect to asthma outcome and/or B[a]P concentration. We calculated the inferred cell population association against asthma or B[a]P concentration, respectively, by pooling  $-\log_{10}$  of significance p-values, and summing  $-\log_{10}(\text{p-values})$ . In particular, we pooled negative and positive associations separately to provide directions for the final association scores (see Calculation of Asthma/B[a]P exposure association scores in Figure S2D for details). For asthma association score,  $-\log_{10}$  of Wilcoxon rank sum p-values were pooled for the following comparisons: asthmatic kids against healthy kids in P1, severe/moderate/mild AMOD against healthy kids in P1. For B[a]P exposure association score, we calculated  $-\log_{10}(\text{p-value})$  from Spearman correlation analysis with B[a]P exposure. The resulting cell scores are summarized in the top right subfigure of **Figure 2B**.

Eigen-gene analysis to clinical traits: The first principal component of each module is calculated by Principal Component Analysis (PCA), and Spearman correlation analysis is performed on the asthma AMOD index (Langfelder and Horvath, 2007).

Validation of Asthma Outcome Classification: Support Vector Machine (SVM) based classifiers are developed to distinguish the P1 asthmatic children from the asymptomatic controls using the hub genes from the gene modules in the P1 network that are significantly enriched for the AMOD SCGs or the DEGs from the asthma -vs- control comparison. A 10-fold cross validation strategy is adopted for performance evaluation. The asthmatic and the control children are equally distributed in each fold. The same procedure is applied within P1 (174 children: asthma = 87, control = 87) and P0 (156 children: asthma = 83, control = 73) separately to evaluate prediction power for asthma. Given that the features used for the classifiers are based on co-expression networks from P1, we posit that the difference in prediction performances between these two regions may be due to the differences in B[a]P exposure.

## **References**

Kim, H.Y. (2015). Statistical notes for clinical researchers: effect size. *Restor Dent Endod* 40, 328-331.

## MATERIALS AND METHODS

Rationale for B[a]P as representative PAH: Robust body of evidence, including our own, have demonstrated B[a]P as a etiologically pertinent and representative PAH compound (Baldacci et al., 2015; Guarnieri and Balmes, 2014; Hew et al., 2015; Manners et al., 2014; Perzanowski et al., 2013). First, not only PAHs, but also a complex mixtures (e.g. diesel-exhaust particles), increase oxidative injury within the respiratory system, thereby contributing to respiratory system inflammation and alteration in lung structure and function (Guarnieri and Balmes, 2014; Hew et al., 2015; Jedrychowski et al., 2013; Li et al., 2008, 2003; Perzanowski et al., 2013). Both the laboratory and epidemiologic evidence have shown that PAHs could induce or enhance allergic sensitization, exacerbate pre-existing asthma, and enhance the risk of *de novo* asthma development (Binkova et al., 2007; Li et al., 2016; Manners et al., 2014). In particular, B[a]P has been shown to directly target hematopoietic stem cells by binding to aryl hydrocarbon receptors, and subsequently impart a wide array of adverse effects, including mitochondrial functional deficit (van Grevenynghe et al., 2005). Earlier investigation have shown that organic constituents of diesel exhaust particles (DEP), such as B[a]P, critically contribute to an allergic airway inflammation, based on both murine (Diaz-Sanchez et al., 1997) and *in vitro* (Singh et al., 2004) model system. Airborne B[a]P is highly mutually correlated with PM<sub>2.5</sub> (Choi et al., 2015; Sram et al., 2013). Ambient B[a]P concentration > 1 ng/m<sup>3</sup> has been shown to increase the risk of genomic material translocations (Sram et al., 2004), formation of micronuclei of the lymphocytes (Rossnerova et al., 2011), sperm DNA fragmentation (Rubes et al., 2009), intrauterine growth restriction, preterm delivery, and an elevated Cephalization Index (Choi et al., 2008, 2006). Furthermore, high ambient B[a]P concentration is associated with an increased PAH-DNA adduct level, urinary 1-hydroxypyrene, DNA strand breaks, and oxidative damage (WHO, 2010).

Target Sites: The urban target district (Radvanice/Bartovice within Ostrava) is historically associated with high ambient pollution level compared to our second, semi-rural target site. For example, mean ambient level of B[a]P in Radvanice/Bartovice is more than four-times higher (11.4 ng/m<sup>3</sup>) than that for our background, semirural districts (2.5 ng/m<sup>3</sup>) during November 2008 (*i.e.* transitional season of interest) (Sram et al., 2013). The transitional season in Czech Republic is associated with not only a large spatial, but also temporal, variability in ambient

PAHs (Sram et al., 2013). Major sources of air pollution in P1 (i.e. Ostrava) include the power plants and coke ovens for various industries, residential heating, and automobile exhaust (Sram et al., 2013). In contrast, southern Bohemia does not have any industrial power plants, but is characterized by residential heating and mobile exhaust emissions (Rossnerova et al., 2009).

Asthma diagnosis: Based on our earlier investigation, historical participation rate by pediatricians to our studies is > 97% (Hertz-Picciotto et al., 2007). The duration (in months) of corticoid medication was recorded for each child. A diagnosis was made using Tenth Revision (ICD-10). Diagnostic reliability of ICD-10 diagnoses in physician visits or hospitalizations in our target regions have been demonstrated (Hertz-Picciotto et al., 2007). Parents of the children also answered a questionnaire on the children's medical history, exposures, and lifestyles.

Allergy test: The local airborne allergens in the allergen test include hazel (*Corylus avellana*), alder (*Alnus incana*), birch (*Betula alba*), plane (*Platanus vulgaris*), cypress (*Cupressus sempervirens*), grass mix (*Poa pratensis*, *Dactylis glomerata*, *Lolium perenne*, *Phleum pratense*, *Festuca pratensis*, *Helictotrichon pratense*), Olive (*Olea europaea*), mugwort (*Artemisia vulgaris*), ragweed (*Ambrosia artemisiifolia*), *Alternaria alternata* (*tenuis*), *Cladosporium herbarum*, *Aspergillus fumigatus*, *Parietaria*, cat, and dog, *Dermatophagoides pteronyssinus*, *Dermatophagoides farinae*).

Polycyclic Aromatic Hydrocarbon Monitoring: Per CHMI protocol, the mean daily level was taken once every three days for total of 10 days/month in southern Bohemia; once every six days for total of 5 days/month in Ostrava. Data quality assurance and control are described (Ghosh et al., 2016; Hertz-Picciotto et al., 2007; Rossner et al., 2012).

Atopic March Onset and Diversity (AMOD) index: Each respective diagnosis of atopic dermatitis, allergic sensitization, wheezing symptoms, at  $\leq 24$  months, as well as the positive result of the spirometry test (taken at age five or higher) were given a score of 1. The composite AMOD index for each child was coded as: 0 for the asymptomatic children. A child with allergic sensitization only (category 1) or doctor-diagnosed atopic dermatitis at  $\leq 24$  months, in addition to the sensitization (category 2) was defined as a 'mild' case. A child with allergic sensitization, atopic dermatitis diagnosis, as well as wheezing at  $\leq 24$  months was deemed 'moderate' (category 3). Finally, a child was deemed as a 'severe' case if he/she was allergen

sensitized, had an early-onset atopic dermatitis diagnosis as well as a positive bronchodilation result (category 4); or also had an additional diagnosis of wheezing at < 24 months (category 5) or had an additional diagnosis of allergic rhinitis (category 6).

Conservation of co-expression modules using independent data sets: Overlap between two modules  $m_i$  and  $m_j$ , denoted as  $m_i \cap m_j$ , is calculated by Jaccard index,  $J_{ij} = \frac{|m_i \cap m_j|}{|m_i \cup m_j|}$  (Handl et al., 2005). Two modules are conserved if Bonferroni adjusted FET  $P < 0.05$  and  $J_{ij} > 0.2$  (i.e., overlap is above 20% of the union of the genes in two modules).

Eigen-gene analysis to clinical traits: The first principal component of each module is calculated by Principal Component Analysis (PCA), and Spearman correlation analysis is performed on the asthma AMOD index (Langfelder and Horvath, 2007).

Validation of Asthma Outcome Classification: Support Vector Machine (SVM) based classifiers are developed to distinguish the P1 asthmatic children from the asymptomatic controls using the hub genes from the gene modules in the P1 network that are significantly enriched for the AMOD SCGs or the DEGs from the asthma -vs- control comparison. A 10-fold cross validation strategy is adopted for performance evaluation. The asthmatic and the control children are equally distributed in each fold. The same procedure is applied within P1 (174 children: asthma = 87, control = 87) and P0 (156 children: asthma = 83, control = 73) separately to evaluate prediction power for asthma. Given that the features used for the classifiers are based on co-expression networks from P1, we posit that the difference in prediction performances between these two regions may be due to the differences in B[a]P exposure.

Adjustment of Tobacco Smoke Exposure in our Network Models: We first validated the children cotinine concentration against their self-or parent- reported exposure to tobacco smoke exposure at home (Choi et al., 2017). The children's median Cotinine concentration is linearly associated with their self-reported number of smokers at home (Choi et al., 2017). Subsequently, we removed all children who were deemed an active smoker. The non-smoking children who are passively exposed were adjusted by including their Cotinine concentration in all models of the analysis. Specifically, we tested correlations between module eigen-genes (Langfelder and Horvath, 2007) and cotinine measurements from the blood children to identify associations between smoking exposure and gene modules. Eigen-genes represent the first principal

components of gene expressions belonging to the respective modules, and serve as representative features of the gene modules in general (Langfelder and Horvath, 2007). In particular, correlation analysis between eigen-gene and clinical trait is an established approach to test the clinical associations at module level (Langfelder and Horvath, 2007; Liu et al., 2015; Ponsuksili et al., 2015; Schieffer et al., 2017; Shen et al., 2006). With FDR < 0.05, we did not find any significantly correlated module to cotinine levels (see Table S4B). Therefore, gene modules are not confounded by the children's passive cigarette smoke exposure. Furthermore, DEG/SCG/DMR signatures were derived by excluding kids with reported self/familial smoking exposure to minimize the cigarette smoking effect. This ensures the module ranking scheme by enrichments of these signatures are minimally affected by cigarette smoking exposure. Overall, our observed exposure-outcome associations are not confounded by cigarette smoke exposure.

## REFERENCES

- Baldacci, S., Maio, S., Cerrai, S., Sarno, G., Baiz, N., Simoni, M., Annesi-Maesano, I., Viegi, G., Study, H., 2015. Allergy and asthma: Effects of the exposure to particulate matter and biological allergens. *Respir. Med.* 109, 1089–1104.
- Binkova, B., Topinka, J., Sram, R.J., Sevastyanova, O., Novakova, Z., Schmuczerova, J., Kalina, I., Popov, T., Farmer, P.B., 2007. In vitro genotoxicity of PAH mixtures and organic extract from urban air particles part I: acellular assay. *Mutat Res* 620, 114–122.
- Choi, H., Jedrychowski, W., Spengler, J., Camann, D.E., Whyatt, R.M., Rauh, V., Tsai, W.-Y., Perera, F.P., 2006. International studies of prenatal exposure to polycyclic aromatic hydrocarbons and fetal growth. *Environ. Health Perspect.* 114, 1744–1750. <https://doi.org/10.1289/ehp.8982>
- Choi, H., Rauh, V., Garfinkel, R., Tu, Y., Perera, F.P., 2008. Prenatal exposure to airborne polycyclic aromatic hydrocarbons and risk of intrauterine growth restriction. *Environ. Health Perspect.* 116, 658–665. <https://doi.org/10.1289/ehp.10958>
- Choi, H., Tabashidze, N., Rossner, P., Dostal, M., Pastorkova, A., Kong, S.W., Gmuender, H., Sram, R.J., 2017. Altered vulnerability to asthma at various levels of ambient Benzo[a]Pyrene by CTLA4, STAT4 and CYP2E1 polymorphisms. *Environ. Pollut.* 231. <https://doi.org/10.1016/j.envpol.2017.07.057>
- Choi, H., Zdeb, M., Perera, F., Spengler, J., 2015. Estimation of chronic personal exposure to airborne polycyclic aromatic hydrocarbons. *Sci. Total Environ.* 527–528, 252–261. <https://doi.org/10.1016/j.scitotenv.2015.04.085>

- Diaz-Sanchez, D., Tsien, A., Fleming, J., Saxon, A., 1997. Combined diesel exhaust particulate and ragweed allergen challenge markedly enhances human in vivo nasal ragweed-specific IgE and skews cytokine production to a T helper cell 2-type pattern. *J Immunol* 158.
- Ghosh, R., Rossner, P., Honkova, K., Dostal, M., Sram, R.J., Hertz-Picciotto, I., 2016. Air pollution and childhood bronchitis: Interaction with xenobiotic, immune regulatory and DNA repair genes. *Environ. Int.* 87, 94–100.
- Guarnieri, M., Balmes, J.R., 2014. Outdoor air pollution and asthma. *Lancet* 383. [https://doi.org/10.1016/s0140-6736\(14\)60617-6](https://doi.org/10.1016/s0140-6736(14)60617-6)
- Handl, J., Knowles, J., Kell, D.B., 2005. Computational cluster validation in post-genomic data analysis. *Bioinformatics* 21, 3201–3212. [https://doi.org/DOI 10.1093/bioinformatics/bti517](https://doi.org/DOI%2010.1093/bioinformatics/bti517)
- Hertz-Picciotto, I., Baker, R.J., Yap, P.-S., Dostál, M., Joad, J.P., Lipsett, M., Greenfield, T., Herr, C.E.W., Beneš, I., Shumway, R.H., 2007. Early childhood lower respiratory illness and air pollution. *Environ. Health Perspect.* 115, 1510.
- Hew, K.M., Walker, A.I., Kohli, A., Garcia, M., Syed, A., McDonald-Hyman, C., Noth, E.M., Mann, J.K., Pratt, B., Balmes, J., Hammond, S.K., Eisen, E.A., Nadeau, K.C., 2015. Childhood exposure to ambient polycyclic aromatic hydrocarbons is linked to epigenetic modifications and impaired systemic immunity in T cells. *Clin Exp Allergy* 45, 238–248. <https://doi.org/10.1111/cea.12377>
- Jedrychowski, W.A., Perera, F.P., Spengler, J.D., Mroz, E., Stigter, L., Flak, E., Majewska, R., Klimaszewska-Rembiesz, M., Jacek, R., 2013. Intrauterine exposure to fine particulate matter as a risk factor for increased susceptibility to acute broncho-pulmonary infections in early childhood. *Int J Hyg Env. Heal.* 216, 395–401. <https://doi.org/http://dx.doi.org/10.1016/j.ijheh.2012.12.014>
- Langfelder, P., Horvath, S., 2007. Eigengene networks for studying the relationships between co-expression modules. *BMC Syst Biol* 1, 54. <https://doi.org/10.1186/1752-0509-1-54>
- Li, N., Georas, S., Alexis, N., Fritz, P., Xia, T., Williams, M.A., Horner, E., Nel, A., 2016. A work group report on ultrafine particles (American Academy of Allergy, Asthma & Immunology): Why ambient ultrafine and engineered nanoparticles should receive special attention for possible adverse health outcomes in human subjects. *J. Allergy Clin. Immunol.* 138, 386–396.
- Li, N., Sioutas, C., Cho, A., Schmitz, D., Misra, C., Sempf, J., Wang, M., Oberley, T., Froines, J., Nel, A., 2003. Ultrafine particulate pollutants induce oxidative stress and mitochondrial damage. *Env. Heal. Perspect* 111, 455–460.
- Li, N., Xia, T., Nel, A.E., 2008. The role of oxidative stress in ambient particulate matter-induced lung diseases and its implications in the toxicity of engineered nanoparticles. *Free Radic. Biol. Med.* 44, 1689–1699.
- Liu, R., Guo, C.-X., Zhou, H.-H., 2015. Network-based approach to identify prognostic

- biomarkers for estrogen receptor–positive breast cancer treatment with tamoxifen. *Cancer Biol. Ther.* 16, 317–324.
- Manners, S., Alam, R., Schwartz, D.A., Gorska, M.M., 2014. A mouse model links asthma susceptibility to prenatal exposure to diesel exhaust. *J. Allergy Clin. Immunol.* 134, 63–72. e7.
- Perzanowski, M.S., Chew, G.L., Divjan, A., Jung, K.H., Ridder, R., Tang, D., Diaz, D., Goldstein, I.F., Kinney, P.L., Rundle, A.G., 2013. Early-life cockroach allergen and polycyclic aromatic hydrocarbon exposures predict cockroach sensitization among inner-city children. *J. Allergy Clin. Immunol.* 131, 886–893. e6.
- Ponsuksili, S., Siengdee, P., Du, Y., Trakooljul, N., Murani, E., Schwerin, M., Wimmers, K., 2015. Identification of common regulators of genes in co-expression networks affecting muscle and meat properties. *PLoS One* 10, e0123678.
- Rossner, P., Svecova, V., Schmuczerova, J., Milcova, A., Tabashidze, N., Topinka, J., Pastorkova, A., Sram, R.J., 2012. Analysis of biomarkers in a Czech population exposed to heavy air pollution. Part I: bulky DNA adducts. *Mutagenesis* ges057.
- Rossnerova, A., Spatova, M., Rossner, P., Novakova, Z., Solansky, I., Sram, R.J., 2011. Factors affecting the frequency of micronuclei in asthmatic and healthy children from Ostrava. *Mutat. Res. Mol. Mech. Mutagen.* 708, 44–49.
- Rossnerova, A., Spatova, M., Rossner, P., Solansky, I., Sram, R.J., 2009. The impact of air pollution on the levels of micronuclei measured by automated image analysis. *Mutat Res* 669, 42–47. [https://doi.org/S0027-5107\(09\)00146-8](https://doi.org/S0027-5107(09)00146-8) [pii]10.1016/j.mrfmmm.2009.04.008
- Rubes, J., Rybar, R., Prinosilova, P., Veznik, Z., Chvatalova, I., Solansky, I., Sram, R.J., 2009. Genetic polymorphisms influence the susceptibility of men to sperm DNA damage associated with exposure to air pollution. *Mutat. Res.* In press.
- Schieffer, K.M., Choi, C.S., Emrich, S., Harris, L., Deiling, S., Karamchandani, D.M., Salzberg, A., Kawasawa, Y.I., Yochum, G.S., Koltun, W.A., 2017. RNA-seq implicates deregulation of the immune system in the pathogenesis of diverticulitis. *Am. J. Physiol. Liver Physiol.* 313, G277–G284.
- Shen, R., Ghosh, D., Chinnaiyan, A., Meng, Z., 2006. Eigengene-based linear discriminant model for tumor classification using gene expression microarray data. *Bioinformatics* 22, 2635–2642.
- Singh, P., DeMarini, D.M., Dick, C.A.J., Tabor, D.G., Ryan, J. V, Linak, W.P., Kobayashi, T., Gilmour, M.I., 2004. Sample characterization of automobile and forklift diesel exhaust particles and comparative pulmonary toxicity in mice. *Environ. Health Perspect.* 112, 820.
- Sram, R.J., Beskid, O., Binkova, B., Rossner, P., Smerhovsky, Z., 2004. Cytogenetic analysis using fluorescence in situ hybridization (FISH) to evaluate occupational exposure to carcinogens. *Toxicol. Lett.* 149, 335–344.

Sram, R.J., Dostal, M., Libalova, H., Rossner, P., Rossnerova, A., Svecova, V., Topinka, J., Bartonova, A., 2013. The European Hot Spot of B [a] P and PM 2.5 Exposure—The Ostrava Region, Czech Republic: Health Research Results. ISRN Public Health 2013.

van Grevenynghe, J., Bernard, M., Langouet, S., Le Berre, C., Fest, T., Fardel, O., 2005. Human CD34-positive hematopoietic stem cells constitute targets for carcinogenic polycyclic aromatic hydrocarbons. *J Pharmacol Exp Ther* 314, 693–702.  
<https://doi.org/jpet.105.084780> [pii]10.1124/jpet.105.084780

WHO, 2010. WHO guidelines for indoor air quality: selected pollutants.