This item is the archived peer-reviewed author-version of:

Metabolic targets of endocrine disrupting chemicals assessed by cord blood transcriptome profiling

Reference:

Remy Sylvie, Govarts Eva, Wens Britt, De Boever Patrick, Den Hond Elly, Croes Kim, Sioen Isabelle, Baeyens Willy, van Larebeke Nicolas, Koppe Janna,- Metabolic targets of endocrine disrupting chemicals assessed by cord blood transcriptome profiling

Reproductive toxicology - ISSN 0890-6238 - 65(2016), p. 307-320

Full text (Publisher's DOI): <http://dx.doi.org/doi:10.1016/J.REPROTOX.2016.08.018>

To cite this reference: <http://hdl.handle.net/10067/1363910151162165141>

Metabolic targets of endocrine disrupting chemicals assessed by cord blood transcriptome profiling

Sylvie Remy^{a,b}, Eva Govarts^b, Britt Wens^b, Patrick De Boever^{b,c}, Elly Den Hond^{b,d}, Kim Croes^e, Isabelle Sioen^{f,g}, Willy Baeyens^e, Nicolas van Larebeke^{e,h}, Janna Koppeⁱ, Adrian Covaci^j, Thomas Schettgen^k, Vera Nelen^l, Juliette Legler^m, Greet Schoeters^{b,n,o}

- a. Epidemiology and Social Medicine, University of Antwerp, Universiteitsplein 1, 2610 Wilrijk, Belgium
- b. Environmental Risk and Health, Flemish Institute for Technological Research (VITO), Boeretang 200, 2400 Mol, Belgium
- c. Centre for Environmental Sciences, Hasselt University, Agoralaan gebouw D, 3590 Diepenbeek, Belgium
- d. Directorate of Public Health and Surveillance, Scientific Institute of Public Health, Juliette Wytsmanstraat 14, 1050 Brussels, Belgium
- e. Department of Analytical, Environmental and Geochemistry (AEGC), Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium
- f. Department of Public Health, Ghent University, St. Pietersnieuwstraat 33, 9000 Ghent, Belgium
- g. FWO Research Foundation, Egmontstraat 5, 1000 Brussels, Belgium
- h. Department of Radiotherapy and Experimental cancer research, Ghent University, St. Pietersnieuwstraat 33, 9000 Ghent, Belgium
- i. EKZ-AMC, University of Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands
- j. Toxicological Center, University of Antwerp, Universiteitsplein 1, 2610 Wilrijk, Belgium
- k. Institute for Occupational and Social Medicine, Medical Faculty, RWTH Aachen University, Pauwelsstraße 30, 52074 Aachen, Germany
- l. Department of Health, Provincial Institute for Hygiene, Kronenburgstraat 45, 2000 Antwerp, Belgium
- m. Institute for Environmental Studies, VU University Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands
- n. Biomedical Sciences, University of Antwerp, Universiteitsplein 1, 2610 Wilrijk, Belgium
- o. Environmental Medicine, University of Southern Denmark, Odense, Campusvej 55, 5000 Odense, Denmark

Address correspondence to: Sylvie Remy, Environmental Risk and Health, Flemish Institute for Technological Research (VITO), Boeretang 200, 2400 Mol, Belgium [sylvie.remy@uantwerpen.be], +32 14 33 51 07.

Funding Source: The studies of the Flemish Center of Expertise on Environment and Health were commissioned, financed and steered by the Ministry of the Flemish Community. The research received funding from the European Community's Seventh Framework Programme (FP7/2007–2013) under grant agreement OBELIX 227391. Isabelle Sioen is financially supported by the Research Foundation Flanders.

Financial Disclosure: The authors have no financial relationships relevant to this article to disclose.

Conflict of Interest: The authors have no conflicts of interest to disclose.

Abbreviations:

DDT	Dichlorodiphenyltrichloroethane
EDC	Endocrine disrupting chemicals
FLEHS(I/II/III)	Flemish Environment and Health Survey (I/II/III)
HNF1A	Hepatocyte Nuclear Factor 1 Homeobox A
HPA	Hypothalamic-pituitary-adrenal
ICA1	Islet Cell Autoantigen 1
IL	Interleukin
IPA	Ingenuity Pathway Analysis (Ingenuity Systems®)
IQR	Interquartile Range
IRS2	Insulin Receptor Substrate
NR3C1	Glucocorticoid Receptor
OBELIX	FP7 project: OBesogenic Endocrine disrupting chemicals: Linking prenatal eXposure to the development of obesity later in life
p,p'-DDE	Dichlorodiphenyldichloroethylene
PCB-153	Polychlorinated Biphenyl-153
PFOA	Perfluorooctanoic Acid
PFOS	Perfluorooctane Sulfonate
PGR	Progesterone Receptor
PPARA	Peroxisome proliferator-activated receptor alpha
PRL	Prolactin
TF	Transcription factor
VIF	Variance inflation factor

Authors' contributions:

GS and EDH designed and coordinated the biomonitoring study. JL assisted in the conceptualization the study. VN was responsible for the field work. WB and NVL participated in the design of the study. EG and IS constructed the database. BW performed RNA processing and microarray hybridization. TS carried out the analysis of PFOA and PFOS in plasma. AC carried out the analysis of POPs in plasma. SR performed transcriptome data analysis and wrote the final manuscript. GS helped to draft the manuscript. PDB, JK, KC

and all authors reviewed and revised the manuscript, and approved the final manuscript as submitted.

Abstract

Early life exposure to endocrine disrupting chemicals (EDCs) has been frequently associated with impaired perinatal growth, an important risk factor for later onset of metabolic disorders.

We analyzed whether the cord blood transcriptome showed early indications of alterations in metabolic processes in 195 human samples in relation to cord blood levels of dichlorodiphenyldichloroethylene (p,p'-DDE), polychlorinated biphenyl-153 (PCB-153), perfluorooctanoic acid (PFOA), and perfluorooctane sulfonate (PFOS).

Overall, 39 metabolically relevant transcription factors were significantly enriched (31 by p,p'-DDE, 10 by PCB-153, 8 by PFOA, and 2 by PFOS). These included the glucocorticoid receptor (p,p'-DDE and PCB-153) and the progesterone receptor (PFOA and PFOS). The 'insulin receptor signaling', 'acute phase response signaling', 'Interleukin(IL)-6 signaling', and 'prolactin signaling' pathways were significantly enriched in relation to p,p'-DDE.

Transcriptional changes at birth suggest a role for specific metabolic targets as a link between prenatal EDC exposure and metabolic disorders later in life.

Keywords

Birth cohort; Transcriptomics; Cord blood; Endocrine disrupting chemicals; Metabolic disorders, Glucocorticoid receptor

Introduction

Early development is a critical period of rapid growth and imprinting and of high susceptibility to exogenous influences such as stress and hormonal signals¹. The European FP7 project OBELIX (OBesogenic Endocrine disrupting chemicals: Linking prenatal eXposure to the development of obesity later in life) examined the hypothesis that prenatal exposure to endocrine disrupting chemicals (EDCs) in food plays a role in the development of obesity later in life². Based on a meta-analysis of European birth cohorts, prenatal polychlorinated biphenyl-153 (PCB-153) and dichlorodiphenyldichloroethylene (p,p'-DDE, a common breakdown product of dichlorodiphenyltrichloroethane (DDT)) exposure have been linked to respectively decreased birth weight³ and increased incidence of rapid infant weight gain during the first 24 months after birth⁴. Rapid infant weight gain has been associated to increased risk on childhood obesity at the age of 10⁵. Based on the risk estimates of these associations, Legler et al. calculated that 0.26 % of all cases of childhood obesity can be attributed to prenatal p,p'-DDE exposure in Europe⁶. Changes in birth weight have also been linked with later development of obesity⁷. The association between PCB exposure and birth weight has been mentioned above³. Metabolites of di-2-ethylhexyl phthalate (DEHP) have been related to both high (MEHHP) and low birth weights (MECPP)⁸. Current literature on the associations between perfluorooctanoic acid (PFOA) and perfluorooctane sulfonate (PFOS) and human fetal growth has been recently reviewed⁹. Based on 14 publications, it was concluded that PFOA and PFOS exposures in pregnancy were associated with lower average birth weights in human newborns in most studies, although not all results were statistically significant. Prenatal exposure to persistent organochlorine pollutants (PCBs, DDE, and hexachlorobenzene (HCB)) has been associated with high insulin levels in 5-year-old girls¹⁰. Despite advancements in knowledge on the endocrine and neuronal regulation of energy

balance and food intake, the underlying mode of action of EDCs concerning their impact on obesity and other metabolic disorders is not well understood at the molecular level.

The field of molecular epidemiology to identify targets of prenatal environmental exposure is gaining interest. -Omics wide analyses enable the study of the whole cascade of molecular processes influenced by exposure. In studying the effects of maternal cigarette smoking during pregnancy, positive associations have been reported between cotinine levels in placenta and mRNA levels of CYP1A1 in fetal lung and liver tissue; and with CYP1B1 in lung tissue¹¹. Increases in metabolic enzyme transcripts in fetal liver in relation to maternal smoking has also been demonstrated by O'Shaughnessy et al. (2011) with more marked responses in male fetuses¹². Activation of inflammation/NF-kappaB signaling has been demonstrated in infants born to arsenic-exposed mothers¹³.

Three levels of functional analysis are of specific interest to derive biological processes that link prenatal EDC exposure to increased risk on later development of metabolic diseases. The first level consists of identification of altered transcription factor (TF) regulation, especially of TFs that are known to play a role in metabolic disorders. Many of the TFs involved in metabolic disorders (such as glucose metabolism disorders, insulin resistance, weight gain, and obesity) are activated upon hormone binding. Mimicking endogenous hormone binding is a well-known mechanism of action of EDCs¹⁴. Second, it has been documented that chronic inflammation plays an important role in various metabolic diseases¹⁵. It has been hypothesized that in type 2 diabetes, there is an ongoing cytokine-mediated acute-phase response (part of a wide-ranging activation of innate immunity), and this is closely involved in the pathogenesis of the disease (1,2). Studies providing evidence of this theory have been reviewed and may be summarized as follows: 1) Markers of inflammation are associated with

type 2 diabetes and features of the metabolic syndrome in cross-sectional studies; 2) Markers of inflammation predict the risk for type 2 diabetes; 3) Inflammation is involved in the pathogenesis of atherosclerosis, a common feature of type 2 diabetes; 4) Anti-inflammatory agents decrease the acute-phase response, may reduce the risk of developing type 2 diabetes, and improve control; and 5) Gestational diabetes, a risk factor for type 2 diabetes, is associated with an inflammatory response¹⁶. When profiling immune competent cells, it is hence relevant to include alterations in immunologic pathways as a potential target of EDC exposure. Third, significant influences of EDC exposure on the expression of separate key genes that play a well-known role in disease development/onset should be documented. This is necessary to overcome a limitation of focusing on ‘significance of pathway enrichment’ in –omics studies, which is solely based on ‘number of molecules’ (Fisher Exact Test) and does not take into account the importance of individual molecules in the biological process.

In 195 cord blood samples collected immediately after birth of newborns from the general population of Flanders, we examined global transcriptome responses in cord blood cells in relation to birth parameters (sex, gestational age at birth, season of birth, being delivered by caesarian section, etc.), maternal parameters (age, pre-pregnancy body mass index, etc.), environmental exposure parameters (heavy metals, endocrine disruptors, air pollutants, etc.). Cord blood can be easily obtained. The exposure biomarkers in cord blood reflect exposure at birth, and are considered as a proxy for exposure during the fetal period and as a proxy for maternal concentrations during pregnancy. The relation between maternal levels and cord levels is documented in literature in our study population with respect to metals¹⁸ and in other studies for persistent organic pollutants^{19,20}. Transcriptome wide, the association between cord blood gene expression and tissue expression has not been documented. However, for specific molecules, it has been well described that cord blood gene expression

reflects transcriptional changes in other tissues, such as for the expression of sFLT1 – a key molecule in regulating infant growth – in the placenta^{21,22}.

One of the objectives of the OBELIX project was to relate early life exposure to EDCs with effect biomarkers and health outcome data which are related to risk of obesity later in life². In this context, we focused on molecular signatures that are specific for prenatal exposure to EDCs that can be associated to obesity and metabolic pathways. The statistical approach allowed to adjust the associations for influential factors (covariates) at the gene specific level. Potential influential factors that were considered included co-exposures, maternal body mass index before pregnancy, newborn sex, parity, maternal smoking during pregnancy, maternal alcohol consumption, season of birth, maternal age, and being delivered by caesarean section, gestational age at birth, and cord blood lipid content.

This paper proposes potential targets of EDCs that underlie metabolic disorders later in life by functional analysis of the gene expression changes that correlate with exposure.

Methods

Ethics Statement

A written informed consent was provided by all participating mothers. The study protocol was approved by the ethical committee of the University of Antwerp (Reference UA A08 09).

Study population

The study made use of samples obtained from a birth cohort that was initiated as part of the human biomonitoring program for environmental health surveillance in Flanders, Belgium (2nd Flemish Environment and Health Survey, FLEHSII, 2007-2011)¹⁷. Newborn-mother couples (N = 255) were recruited from the general population of Flanders between August 2008 and July 2009 using a multistage sampling procedure. Ten maternities (two in each of the five Flemish provinces) were randomly selected. Mothers that came to give birth in the maternity were invited to participate if they lived at least 10 years in Flanders and were able to fill in a questionnaire in Dutch (to collect individual health and life style data¹⁷). The number of participants per province was proportional to the number of inhabitants in that province (status at 01/01/2006). A subset of 195 uncomplicated singleton pregnancies was randomly selected for transcriptome profiling.

Exposure assessment

For this paper, we used a subset of chemical exposure markers that were analyzed in the Flemish human biomonitoring program¹⁷. We restricted our analysis to exposure markers that could be detected in more than 50 % of the participants. Lead (Pb) , manganese (Mn) , copper (Cu), thallium (Tl) and arsenic (As) were measured in cord, cadmium (Cd) was measured in maternal whole blood samples by high resolution inductively coupled plasma-mass spectrometry (HR-ICP-MS,Thermo Element II) after micro-wave acid digestion using HNO₃

and H_2O_2 ²³. Persistent chlorinated compounds (polychlorinated biphenyl-153 (PCB-153), dichlorodiphenyldichloroethylene (p,p'-DDE), and hexachlorobenzene (HCB)) were measured in plasma cord samples by using solid-phase extraction and gas chromatography-electron capture negative ionization mass spectrometry^{24,25}. CALUX (Chemical Activated Luciferase gene eXpression) assay is a bioassay used to measure the dioxin like compounds in cord plasma²⁶. Perfluorinated compounds (perfluorooctane sulfonate (PFOS), perfluorooctanoic acid (PFOA)) in cord plasma samples were measured according to Midasch et al. 2007²⁷. Total mercury (Hg) and methylmercury (Hg-MM) were measured in maternal hair by combustion-atomic absorption spectrometry and by headspace injection-gas chromatography-atomic fluorescence spectrometry, respectively²⁸. Hg-MM and not total Hg was used in data analysis. Mono-(2-ethyl-5-carboxypentyl) phthalate (MECPP), a metabolite of di(2-ethylhexyl) phthalate (DEHP), in cord plasma was measured after enzymatic deconjugation²⁹. After addition of the internal standards (¹³C₄-MECPP obtained from Cambridge Isotope Laboratories) used for isotope dilution, a simple protein precipitation step using formic acid was applied. The extracts were analyzed after large volume injection using an on-line trapping column coupled to liquid chromatography and triple quadrupole mass spectrometry. Other metabolites of DEHP were also measured, i.e. mono(2-ethyl-5-hydroxyhexyl)phthalate (MEHHP), and mono(2-ethyl-5-oxohexyl)phthalate (MEOHP). However, these could only be detected in a minority of the samples and are therefore not included in the analysis. In Belgium, several air pollutants, including fine particulate matter (PM 2.5), are continuously measured by a network of automatic monitoring sites (<http://www.irceline.be>). Exposure to PM_{2.5} at the home address was assessed based on interpolation (model RIO) of daily 24 h average concentrations of monitoring data calculated on a 4x4 km² grid to the coordinates of the home addresses of the individual participants³⁰.

Other characteristics of the population

Information on general health and lifestyle parameters (e.g. smoking, consumption of alcohol, length and prenatal weight, maternal age, and parity) were collected from questionnaires. Pre-pregnancy BMI was calculated from self-reported weight and height ($\text{BMI} = \text{weight (kg)}/\text{height}^2 (\text{m}^2)$). Birth parameters (type of delivery, gestational age), were obtained from the clinical records at the maternities. Serum triglycerides (TG) and serum cholesterol (Chol) levels were measured in cord blood samples by spectrophotometry on a Modular analyser (Roche Diagnostics, Belgium). Total lipid (TL) concentration (mg/dL) was calculated by the following formula: $\text{TL} = 1.33 * (\text{TG} + \text{Chol}) + 50.5^{31}$.

Transcriptome profiling

The procedures for cord blood collection, RNA processing, and microarray hybridization (4x44K Agilent Whole Human Genome, design 014850, one-color experimental setup with Cy3-labelling) have been described previously³². The arrays were scanned with an Agilent scanner (G2565BA) and subjected to a primary quality control using the Agilent Feature Extraction Software (Version 10.7). For each feature on the array, the software accurately determines the intensity (gProcessedSignal, signal intensity derived from Cy3 fluorescent dye), which was used as input for further processing and analysis (Matlab, version R2013b). Normalization between arrays was performed by means of quantile normalization upon logarithmic transformation of the data ($\log_2$ -scale). For sequences that are replicated on the array, $\text{mean}(\log_2(\text{gProcessedSignal}))$ was calculated. After filtering for non-control features, 41000 unique sequences remained. As a criterion to remove features with data in the feature noise level, features with median processed signal intensity below 3 were removed. Data obtained from sequences that successfully passed this step of quality control could be confirmed with qPCR as shown in our previous publication for a limited panel of genes³². In

addition, a criterion to filter sequences prior to statistical analysis was based on variation in expression among the dataset. Sequences of which the fold change between the 95th percentile and the 5th percentile of expression was smaller than 2, were not considered for statistical analysis as a data reduction strategy in the context of multiple hypothesis testing. As for all measurements, measurement error exists. Assuming that measurement error is comparable over the whole array, the impact of measurement error on the results is much larger on statistical analysis of transcripts with little variability in the population and may result in unreliable findings. In addition, in general one may assume that transcripts with a very small expression range among the population contribute little to differential biological function although it may not be excluded that small changes in the expression of extremely potent genes may be biologically relevant. Median normalized $\log_2$ transformed signal intensities (over the individual samples) ranged from 1.86 to 19.16. In the remainder of this text, the term ‘expression level’ refers to processed signal intensity ($\log_2$ scale).

Statistical analysis

Statistical analyses were performed with Matlab (version R2013b). The workflow of data analysis is shown schematically by figure 1. For each sequence that successfully passed quality control and that could be annotated to a known gene transcript according to Ingenuity Knowledge Base (Ingenuity ®), a two-tiered approach was followed to indicate which variables are potential determinants of the expression of its corresponding transcript. The first tier consisted of univariate linear regression with the variables listed in **table 1 and 2** as independent variables and the gene expression level ($\log_2$ -scale) as the response variable. Categorical independent variables were treated as nominal variable, continuous independent variables were Ln-transformed prior to analysis. For each sequence, all variables with p-value below 0.2 were selected as input for tier 2. This criterion is often applied in the field of

epidemiology to identify potential important determinants which are subsequently subjected to multiple analysis. In tier 2, stepwise multiple linear regression was performed in which the selection of predictive variables was carried out by an automatic procedure. The outcome of the algorithm is a model that includes the most predictive set of determinants (out of all variables included in the study) as independent variables to explain the expression of a gene. The maximum number of steps allowed by the algorithm was set at 10. For the other criteria default Matlab settings were selected. The outcome was a multiple linear regression model for each of the sequences and their corresponding gene transcripts. Transcripts of which the adjusted coefficient of determination (R^2) of the regression models was smaller than 5% were not considered for further analysis. For these transcripts, we concluded that our dataset did not provide sufficient information, as yet, to explain (at least part of) the variability observed in expression level. Multicollinearity was evaluated by calculating variance inflation factors (VIF) using statistical programming software R³³ package car³⁴. A maximum VIF value of 5 is often applied as a golden standard. In this manuscript we use a conservative criterion of 2 to indicate a possible problem of multicollinearity. Next, we constructed a list of significantly associated genes for each of the determinants (p-value of regression coefficient and/or p-value of interaction term smaller than 0.05).

Sensitivity analysis

A strategy for multiple testing was not included in the workflow, therefore we evaluated the impact of false discovery rate (FDR) control³⁵ at Tier 1. Developing the approach for data analysis (**figure 1**), an arbitrarily chosen cut-off was set at the maximal number of steps in the stepwise regression algorithm. As a sensitivity analysis, we evaluated the impact of lowering this limit from 10 steps to 5.

Functional analysis

IPA (Ingenuity Pathway Analysis, Ingenuity Systems[®]) was used for biological interpretation. Significant gene lists were used as input for 1) upstream regulator analysis (focus on TFs); and 2) canonical pathway analysis. A Fisher-exact test was used to determine whether the gene lists included more genes downstream of a particular TF or more genes associated to a given pathway compared to random chance $P < 0.05$. In addition, we uploaded to IPA the direction of regulation (upregulation for positive estimates, downregulation for negative estimates). Based on this, the activation or inhibition state of the TFs was predicted by calculating the Z-score. Absolute Z-score > 2 was used as criterion for reliable prediction. For genes that included an interaction term between two influential factors, the direction of the expression change was processed as missing to determine z-scores. When e.g. for a given gene there is an interaction between PFOA and sex, the direction of the expression change associated to PFOA may be the opposite in girls versus boys. Upstream regulators for which the number of target molecules in the dataset was two or less were omitted to increase reliability.

For the scope of this manuscript, the EDCs were selected for targeted functional analysis (PFOA, PFOS, PCB-153, p,p'-DDE, HCB, MECP, and dioxin like compounds (CALUX)). The lists of genes significantly associated to either one of the EDCs were screened for significant alterations in transcription factors that are known to be associated to glucose metabolism disorders; insulin resistance; weight gain; and obesity. Second, the significance of canonical pathways categorized as 'cellular immune function' or 'cytokine signaling' (n = 98, listed in **supplementary table 1**) was investigated. Third, we screened for key molecules involved in disease specific pathways, irrespective of the pathway significance level: leptin signaling in obesity; maturity onset diabetes of young; type I diabetes mellitus signaling; type

II diabetes mellitus signaling; insulin receptor signaling, estrogen receptor signaling; androgen signaling; PPAR signaling; PPAR α /RXR α activation; and glucocorticoid receptor signaling. And fourth, a global functional enrichment analysis has been performed at the level of gene ontology (GO), biological processes). The analysis was performed using R³³ and Bioconductor package 'topGO'³⁶. Enrichment has been analyzed by a Fisher Exact Test.

Results

Descriptive statistics

Characteristics of the newborn-mother pairs

An overview of the characteristics of the newborns and their mothers is given in **table 1**. 88 % of the mothers that fulfilled the inclusion criteria were willing to donate a cord sample. The study population was evenly distributed among boys and girls. The majority of the babies was born after 39 or 40 weeks of gestation (61%). 95% of the newborns were delivered vaginally. The age of the mothers was on average 30 years, and the average BMI before pregnancy was 23.15 kg/m². For 39% of the mothers, this was their first child. 13% of the mothers reported smoking during their pregnancy.

Exposure assessment

Summary statistics of the biomarkers of exposure are given in **table 2**. The EDCs were measured in cord plasma. HCB was below level of detection in 49% of the samples and hence treated as a dichotomous variable. The other EDCs were detectable in almost all samples. None of the EDC exposure biomarkers was influenced by maternal BMI (p-value ANOVA > 0.2). A correlation matrix (**Table 3**) shows the Pearson correlation coefficients and corresponding p-values for each exposure pair. Strongly correlated exposures are unlikely to be identified both as associated with the expression of a given gene target based on the statistical procedure that was followed. Highly significant positive correlations ($p < 0.0001$) were observed between As and Hg-MM, Mn and Cu, p,p'-DDE and PCB-153, Hg-MM and PCB-153, PFOS and PCB-153, Hg-MM and PFOS, PFOA and PFOS.

Microarray data

After filtering out sequences within the feature noise level 33,543 unique sequences remained out of 41,000 (non-control) sequences on the array. For 91% of the sequences, the difference between P95 and P5 of normalized signal intensities was at least two-fold; for 28% the difference was at least four-fold; and for 5% the difference was at least eight-fold. This indicates that for the vast majority of the sequences, the expression level of the corresponding transcripts is variable in the population. Among the sequences of which the difference in expression between P95 and P5 was at least two-fold ($n = 30,548$), 24,779 sequences could be annotated to gene transcripts coding for 16,592 different molecules according to the ingenuity database (Ingenuity Systems®, <http://www.ingenuity.com>). This panel of sequences was subjected to further analysis.

Identification of major determinants in neonatal gene expression

For each sequence that successfully passed our filtering criteria and that could be annotated to a known gene transcript according to Ingenuity, a multiple linear regression model was constructed – as depicted in **figure 1** - to identify the best possible set of determinants of its expression level. For 16138 of the sequences (corresponding to 12,219 genes), R^2 of the multiple regression model exceeded 5%. For each of the models, VIFs were calculated. All data together the VIF ranged from 1.000 to 2.442 (99th percentile = 1.612), which indicates that there were no issues of multicollinearity. The VIF exceeded 2 only for the determinant ‘season of birth’ for a limited number of 6 sequences, however the maximum value of 2.442 is far below the standard criterion of 5. For the remainder of the sequences that were analyzed (8,641), our dataset contains too little information to explain, at least part, of the variability in expression observed. Based on this, we generated lists of transcripts that were associated to at least one of the potential determinants in the dataset. In **figure 2**, the determinants are ranked according to the number of genes they are significantly associated with. Gestational age was

the most prominent explanatory variable of the neonatal transcriptome, as it was associated to over 60% of the genes. Based on **figure 2**, other important determinants in the study population were sex of the newborn, season of birth, and cord blood lipid content that were associated with expression of about 20 to 30 percent of the genes. A table with detailed outcome statistics for each gene has been added to the data supplement (**Supplementary table 2**).

Transcriptional changes associated to EDC exposure

The database of multiple regression models allowed to identify 918 genes that were associated to HCB, 845 to p,p'-DDE, 711 to PCB-153, 311 to PFOA, 133 to CALUX, 102 to PFOS, and 65 to MECPD exposure (taking into account relevant influential factors). We did not further consider HCB, CALUX, and MECPD for functional analysis as HCB was below level of detection in 49% of the samples and as such treated as a dichotomous variable in the analysis, and the other two compounds showed little effect on the transcriptome. It is generally not recommended to perform functional enrichment analysis on genelists that are too small. In the context of gene ontology, a rough guide has been suggested at <http://geneontology.org/faq/what-minimum-information-include-functional-analysis-paper> stating that it would be difficult to observe enrichment when analyzing 150 IDs. PFOS was maintained for comparison with structurally related PFOA, although a small number of genes could be associated to this compound. In **figure 3**, examples are given that show the relationship between chemical exposure and gene expression, i.e. PRL (**Figure 3a**) and METNR1 (**Figure 3b**). PRL was selected as it is identified as key molecule in relation to PCB-153 (see section on key molecules), METNR1 was selected as an example of effect modification by sex in the association between PFOA and gene expression. For both genes there was an increase in expression in relation to exposure.

Transcription factors

Upstream regulator analysis revealed that, in total, a panel of 90, 31, 17, and 4 TFs could be identified of which the gene lists, associated with respectively p,p'-DDE, PCB-153, PFOA, and PFOS, were significantly enriched (p-value < 0.05) in downstream target genes. Focusing on TFs that are related to specific conditions of interest (**table 4**), 39 TFs could be identified of which the EDC associated gene lists were significantly enriched (p-value < 0.05) in downstream target genes. The impact of p,p'-DDE (31 significant TFs) was more pronounced compared to other EDCs (10 for PCB-153, 8 for PFOA, and 2 for PFOS). A subset of 11 TFs overlapped between different EDCs, among which the glucocorticoid receptor (NR3C1) between p,p'-DDE and PCB-153 and the progesterone receptor (PGR) between PFOA and PFOS. Detailed information for each of the transcription factors is included in **supplementary table 3**. This table allows extracting the pathways associated to each of the TFs (including non-metabolic related pathways) and the list of matching gene correlates for each of the chemicals. Based on this, one may obtain also obtain an overview of the overlap between different TFs.

Key molecules in disease specific pathways

Focusing on disease specific pathways, the 'insulin receptor signaling' pathway was significantly enriched in the gene list related to p,p'-DDE exposure. This list includes IRS2 (insulin receptor substrate) which is regulated by the insulin receptor (**table 5**). Screening for potential key molecules in disease specific pathways, irrespective of significance of enrichment, we identified HNF1A (hepatocyte nuclear factor 1 homeobox A) in relation to PCB-153; ICA1 (islet cell autoantigen 1) in relation to PFOA; PRL (prolactin) in relation to PCB-153; and PPARA (Peroxisome proliferator-activated receptor alpha) in relation to p,p,'-

DDE. Detailed outcome statistics describing the strength and the significance for each of the exposure-expression relationships are provided by **supplementary table 2**.

Significance of immunologic pathways

An overview of the results of canonical pathway analysis focusing on cellular immune function and cytokine signaling pathways (98 pathways in total) is given in **table 6**. This table included the subset of pathways that were significantly enriched ($p < 0.05$) among the gene lists associated with the EDCs. Detailed outcome statistics describing the strength and the significance for each of the exposure-expression relationships for the individual molecules are provided by **supplementary table 2**. The gene list associated with p,p'-DDE was significantly enriched in genes associated to 17 of the 98 selected pathways. Most of those pathways contained genes that were also significantly associated with the other EDCs, however there was less evidence of pathway enrichment (based on significance). For PFOA, significant enrichment was only demonstrated for 'Natural killer cell signaling', and 'p38 MAPK signaling'. Pathways of interest in the context of diabetes and obesity were 'acute phase response signaling', 'Interleukin(IL)-6 signaling', and 'Prolactin signaling' in response to p,p'-DDE.

Gene ontology enrichment

An overview of the results of the significant GO terms (biological processes) is provided by **supplementary table 4**. Focussing on the TOP-10 of enriched GO terms (based on significance), enriched functions associated with p,p'-DDE were highly related to immune system processes. In addition, GO functions such as 'response to hormone', and 'fat cell differentiation' were significantly enriched ($p < 0.05$) among the list of correlates associated to p,p'-DDE. Considering PCB-153, the TOP-10 of the GO functions was mainly related to

protein modification, apoptotic signaling, and regulation of protein oligomerization. For PFOS, TOP-10 GO functions were related to RNA processing and regulation of binding were the most enriched. The list of TOP-10 enriched functions related to PFOA showed little consistency and was involved in a variety of processes.

Sensitivity analysis

Multiple hypothesis testing

The statistical workflow (**figure 1**) does not include a strategy for multiple hypothesis testing. As a sensitivity analysis, the impact of FDR control at tier 1 has been evaluated. **Supplementary figure 1** shows an overview of the number of probes that are associated with the different determinants based on univariate analysis when considering either p-value below 0.2 or FDR below 0.2. For almost 90% of the sequences for which gestational age was considered as a potential determinant (in Tier 1) the FDR was below 0.2. For Cd, Tl, and Mn, this was true for 55%, 43%, and 31% of the sequences respectively. Other determinants for which more than 20% of the univariate associations are maintained after FDR control are sex (51% remaining) and PM2.5 (22%).

Stepwise regression: 'Number of steps'

The maximal number of steps in the stepwise regression algorithm (tier 2) was arbitrarily set at 10. To evaluate the impact of lowering this value to 5, a sensitivity analysis has been performed. As can be derived from **supplementary table 5**, the influence of changing this parameter may be neglected. The table shows the number of sequences that was influenced by reducing the maximal number of steps from 10 to 5 was below 25 for each of the determinants. Taking into account the intersection between the different determinants, the

multiple regression model (set of explanatory determinants) of a limited set of 177 sequences (among 16138) was influenced.

Discussion

Different from well-controlled *in vitro* and animal studies, the human population is highly dynamic and diverse due to genetic differences but also due to the continuous interaction of the internal environment of chemicals, cytokines, genes, proteins and metabolites inside the body and the variable external environment. Studying associations between (environmental) exposure and effect, multiple regression models allow to control for the impact of covariates (e.g. sex, age, smoking habits, ...). When considering a limited number of endpoints (effects) one may be able to identify the appropriate panel of influential factors based on literature and expert knowledge. However, strategies to *a-priori* defining the appropriate influential variables in -OMICS wide analysis are currently lacking. The method we propose allows to rank the determinants of – in this case – the cord blood transcriptome. For each gene separately, the method enables to identify a panel of determinants that is the most explanatory for the variability in its expression level. Among the determinants that were assessed, gestational age was the most important determinant of the transcriptome as it could be associated with the vast majority of genes. This is in line with previous reports in which it has been demonstrated that gestational age might introduce serious bias in data analyses and needs to be carefully assessed in genome-wide transcription studies³⁷.

The results presented in this manuscript allow to identify targets of EDCs. Out of the panel of EDCs that was assessed in cord blood, PCB-153, p,p'-DDE, PFOA, and PFOS were selected for functional analysis of associated gene transcripts. The level of PFOA in cord blood was comparable to recent reports from studies conducted in the Netherlands²⁹, Norway³⁸, and Germany³⁹, while the levels of PFOS were slightly higher (1.5 to 2.5 fold). The level of HCB in our study population was much lower in comparison to studies conducted in Spain⁴⁰, Italy⁴¹, and Denmark (Faroer Islands)⁴². Establishing reference values for Flanders¹⁷, it must

be noted that the levels of PCB-153 and p,p'-DDE in our study population are rather low in comparison to international literature as we did not focus on specific areas of high exposure⁴⁰⁻⁴³. A study in Belgium (1999), for which an industrial area had been recruited, reported levels that were about 3 to 4 fold higher⁴⁴. To compare the level of PCB-153, the sum of three marker PCBs (138, 153 and 180) was considered to compare with the international literature. In our study population, the geometric mean of the sum of PCB-138,-153, and 180 was 120 ng/L (95% CI: 114-128, data not shown). Two studies that were sampled slightly earlier reported levels that were about 2 fold (Italy⁴¹), and 4 fold higher (Spain⁴³). This was also the case for p,p'-DDE. The levels of EDCs were not measured in mother's blood in this study. However, for PFOA and PFOS we have data from blood levels measured in non-pregnant women within the same age range sampled during the same period in Flanders (data not shown). The geometric mean was about twice as high for PFOA, and 4-fold higher for PFOS compared to the cord blood levels. Phthalate levels (DEHP metabolite MECPP) in cord blood are not widely available. Compared to the LINC study conducted in the Netherlands⁴⁵, MECPP levels were about twice as high.

A list of genes was extracted of which the expression in cord blood cells was significantly associated to cord blood levels of p,p'-DDE, PCB-153, PFOA and PFOS taking into account relevant influential factors. A limitation in the interpretation of the results may be that the EDCs were strongly correlated. Given the statistical approach that was followed, multiple EDCs were unlikely to be identified together as associated with the expression of a given gene target. However, the EDC that has been detected in such cases can be considered as the most important to explain the variability in expression in our dataset.

Most of the alterations in metabolically relevant pathways and genes were attributed to p,p'-DDE exposure. Importantly, population studies in Flanders have already linked prenatal p,p'-DDE exposure levels to increased BMI in 3 year old girls⁴⁶ and to increased abdominal fat at

age 7-9⁴⁷. This observation together with the stronger impact on the transcriptome of cord blood cells suggests that the current levels of p,p'-DDE may be more potent as compared to the other chemicals assessed in our cohort. This may also be concluded from genome wide gene ontology analysis, in which enrichment of fat cell differentiation and functions related to hormone responses was only observed for p,p'-DDE. Prenatal exposures to PCB-153, PFOA, and PFOS also showed links to metabolic disorders but to a lesser extent. There was considerable overlap of differential regulation of TFs between the different EDCs, which may suggest a shared mechanism of prenatal programming. Enrichment of downstream targets of the glucocorticoid receptor (NR3C1) was associated to both p,p'-DDE and PCB-153 exposure. The role of glucocorticoids in fetal programming is gaining interest. As recently reviewed, endogenous glucocorticoids provide a critical developmental trigger⁴⁸. Promoting the production of pulmonary surfactant, glucocorticoids are essential for normal maturation of the fetal lung. In addition, they affect fetal growth and are critical for developmental switching in many other organ systems, including thyroid, kidney, brain, and pituitary⁴⁹. Based on its powerful action, it has been postulated that inappropriate exposure – either excess exposure or exposure at incorrect stages of development – results in altered physiological function throughout life by targeting normal developmental trajectories^{50,51}. The fetal hypothalamic-pituitary-adrenal (HPA) axis is particularly susceptible to long-term programming by glucocorticoids. Aberrant fetal programming of the HPA axis has been associated with - among other - increased susceptibility to metabolic disease⁴⁸. The results of this study suggest that PCB-153 and p,p'-DDE may affect glucocorticoid levels and/or action. This is further supported by the observation that these compounds have already been associated with targets of glucocorticoids, i.e. prenatal and postnatal growth^{3,4}.

The list of genes associated to p,p'-DDE exposure was significantly enriched in pathways including 'acute phase response signaling' and 'IL-6 signaling'. In addition, when performing

global functional analysis on the list of gene correlates associated to p,p'-DDE, gene ontology functions related to immune system processes were the most significantly enriched. Already in 2004, the evidence that supports a role for inflammation and innate immunity in type 2 diabetes has been reviewed¹⁶. Many studies showed that circulating markers of inflammation, acute-phase reactants, or IL-6 (the major cytokine mediator of the acute-phase response) are strong predictors for the development of type 2 diabetes¹⁶. Our results confirm differential regulation of these pathways already at the prenatal stage exposure. The finding that p,p'-DDE is associated with respiratory health symptoms in early childhood⁵² further supports that immunologic pathways may be a target of p,p'-DDE. The canonical pathway 'prolactin signaling' was significantly associated to p,p'-DDE, while the expression of the hormone prolactin itself was related to PCB-153 exposure. Differences in placental PRL gene expression have been demonstrated between 'large for gestational age' versus 'appropriate for gestational age' infants⁵³. A recent animal study provides evidence that defective prolactin signaling impairs pancreatic β -cell development during the critical perinatal period⁵⁴. Our findings are in line with these observations and provide further support for impaired prolactin regulation as a link to adult onset of disease.

In addition, we identified a panel of potential 'key molecules' among the gene lists associated to EDC exposure, i.e. IRS2 (p,p'-DDE), ICA1 (PFOA), and PPARA (p,p'-DDE). IRS2 plays a key role in the hypothalamic regulation of central leptin and peripheral insulin sensitivity, as well as β cell propagation⁵⁵. ICA1 is an autoantigen in insulin-dependent diabetes mellitus. PPAR α plays a central role in fatty acid oxidation and uptake in tissues like liver, heart, skeletal muscle and kidney.

The associations that were tested in our study may suffer from chance findings due to multiple hypothesis testing. Adjustment strategies would lower the false positive detection rate, but would certainly prevent the identification of biologically relevant targets as important sources

of variability are – as yet - unquantified. Such sources of variability include the influence of genetics and variability due to differences in physiological cellular demands. As a strategy to limit chance findings, we focused on existing molecular knowledge of disorders that have been associated to (early) EDC exposure. Our study identified gene expression targets that are functionally relevant in relation to the selected endpoints and exposures which adds to the weight of evidence of causal links between exposure and effect. Confirming our findings in additional cohorts will further strengthen the current evidence. At yet, comparable existing datasets could not be identified to validate our findings. Furthermore, additional *in vitro* experiments may help to further elucidate the proposed metabolic relationships.

Conclusions

In this publication, we propose an approach to identify the most influential set of variables in –OMICS wide epidemiologic analysis. Focusing on EDC exposure, we provide insight of biological processes through which prenatal EDC exposure may contribute to growth changes that may lead to diabetes and obesity later in life. Prenatal p,p'-DDE exposure had more impact at the gene expression level as compared to the other chemicals. Differential regulation of the glucocorticoid receptor, IL-6 signaling, acute phase response signaling, and prolactin signaling may link prenatal p,p'-DDE exposure to increased risk on rapid weight gain, increased childhood BMI, and increased abdominal fat as observed in previous epidemiological studies.

Acknowledgements

The studies of the Flemish Center of Expertise on Environment and Health were commissioned, financed and steered by the Ministry of the Flemish Community. The research received funding from the European Community's Seventh Framework Programme

(FP7/2007–2013) under grant agreement OBELIX 227391. Isabelle Sioen is financially supported by the Research Foundation Flanders.

References

1. Moisiadis VG, Matthews SG. Glucocorticoids and fetal programming part 2: mechanisms. *Nat Rev Endocrinol*. 2014;10:403-411. doi:10.1038/nrendo.2014.74.
2. Legler J, Hamers T, van Eck van der Sluijs-van de Bor M, et al. The OBELIX project: early life exposure to endocrine disruptors and obesity. *Am J Clin Nutr*. 2011;94(6 Suppl):1933S - 1938S. doi:10.3945/ajcn.110.001669.
3. Govarts E, Nieuwenhuijsen M, Schoeters G, et al. Birth weight and prenatal exposure to polychlorinated biphenyls (PCBs) and dichlorodiphenyldichloroethylene (DDE): a meta-analysis within 12 European Birth Cohorts. *Environ Health Perspect*. 2012;120(2):162-170. doi:10.1289/ehp.1103767.
4. Iszatt N, Stigum H, Verner M-A, et al. Prenatal and Postnatal Exposure to Persistent Organic Pollutants and Infant Growth: A Pooled Analysis of Seven European Birth Cohorts. *Environ Health Perspect*. 2015. doi:10.1289/ehp.1308005.
5. Ong KK, Loos RJF. Rapid infancy weight gain and subsequent obesity: systematic reviews and hopeful suggestions. *Acta Paediatr*. 2006;95(8):904-908. doi:10.1080/08035250600719754.
6. Legler J, Fletcher T, Govarts E, et al. Obesity, Diabetes, and Associated Costs of Exposure to Endocrine-Disrupting Chemicals in the European Union. *J Clin Endocrinol Metab*. 2015;100(4):jc.2014-4326. doi:10.1210/jc.2014-4326.
7. Monteiro POA, Victora CG. Rapid growth in infancy and childhood and obesity in later life--a systematic review. *Obes Rev*. 2005;6(2):143-154. doi:10.1111/j.1467-789X.2005.00183.x.
8. de Cock M, De Boer MR, Lamoree M, Legler J, Van De Bor M. Prenatal exposure to endocrine disrupting chemicals and birth weight-A prospective cohort study. *J Environ Sci Health A Tox Hazard Subst Environ Eng*. 2015;1-8. doi:10.1080/10934529.2015.1087753.
9. Bach CC, Bech BH, Brix N, Nohr EA, Bonde JPE, Henriksen TB. Perfluoroalkyl and polyfluoroalkyl substances and human fetal growth: a systematic review. *Crit Rev Toxicol*. 2015;45(1):53-67. doi:10.3109/10408444.2014.952400.
10. Tang-Péronard JL, Heitmann BL, Jensen TK, et al. Prenatal exposure to persistent organochlorine pollutants is associated with high insulin levels in 5-year-old girls. *Environ Res*. 2015;142:407-413. doi:10.1016/j.envres.2015.07.009.
11. Vyhlidal CA, Riffel AK, Haley KJ, et al. Cotinine in human placenta predicts induction of gene expression in fetal tissues. *Drug Metab Dispos*. 2013;41(2):305-311. doi:10.1124/dmd.112.049999.
12. O'Shaughnessy PJ, Monteiro A, Bhattacharya S, Fowler PA. Maternal smoking and fetal sex significantly affect metabolic enzyme expression in the human fetal liver. *J Clin Endocrinol Metab*. 2011;96(9):2851-2860. doi:10.1210/jc.2011-1437.
13. Fry RC, Navasumrit P, Valiathan C, et al. Activation of inflammation/NF-kappaB signaling in infants born to arsenic-exposed mothers. *PLoS Genet*. 2007;3(11):e207. doi:10.1371/journal.pgen.0030207.
14. Zoeller RT, Brown TR, Doan LL, et al. Endocrine-disrupting chemicals and public health protection: a statement of principles from The Endocrine Society. *Endocrinology*. 2012;153(9):4097-4110. doi:10.1210/en.2012-1422.
15. Monteiro R, Azevedo I. Chronic inflammation in obesity and the metabolic syndrome. *Mediators Inflamm*. 2010;2010. doi:10.1155/2010/289645.
16. Pickup JC. Inflammation and activated innate immunity in the pathogenesis of type 2 diabetes. *Diabetes Care*. 2004;27:813-823. doi:10.2337/diacare.27.3.813.

17. Schoeters G, Den Hond E, Colles A, et al. Concept of the Flemish human biomonitoring programme. *Int J Hyg Environ Health*. 2012;215(2):102-108. doi:10.1016/j.ijheh.2011.11.006.
18. Baeyens W, Vrijens J, Gao Y, et al. Trace metals in blood and urine of newborn/mother pairs, adolescents and adults of the Flemish population (2007-2011). *Int J Hyg Environ Health*. 2014;217(8):878-890. doi:10.1016/j.ijheh.2014.06.007.
19. Needham LL, Grandjean P, Heinzow B, et al. Partition of environmental chemicals between maternal and fetal blood and tissues. *Environ Sci Technol*. 2011;45(3):1121-1126. doi:10.1021/es1019614.
20. Aylward LL, Hays SM, Kirman CR, et al. Relationships of chemical concentrations in maternal and cord blood: a review of available data. *J Toxicol Environ Health B Crit Rev*. 2014;17(3):175-203. doi:10.1080/10937404.2014.884956.
21. Rajakumar A, Michael HM, Rajakumar PA, et al. Extra-placental expression of vascular endothelial growth factor receptor-1, (Flt-1) and soluble Flt-1 (sFlt-1), by peripheral blood mononuclear cells (PBMCs) in normotensive and preeclamptic pregnant women. *Placenta*. 2005;26(7):563-573. doi:10.1016/j.placenta.2004.09.001.
22. Makris A, Thornton C, Thompson J, et al. Uteroplacental ischemia results in proteinuric hypertension and elevated sFLT-1. *Kidney Int*. 2007;71(10):977-984. doi:10.1038/sj.ki.5002175.
23. Vrijens J, Leermakers M, Stalpaert M, et al. Trace metal concentrations measured in blood and urine of adolescents in Flanders, Belgium: reference population and case studies Genk-Zuid and Menen. *Int J Hyg Environ Health*. 2014;217(4-5):515-527. doi:10.1016/j.ijheh.2013.10.001.
24. Covaci A, Schepens P. Simplified method for determination of organochlorine pollutants in human serum by solid-phase disk extraction and gas chromatography. *Chemosphere*. 2001;43(4-7):439-447. <http://www.ncbi.nlm.nih.gov/pubmed/11372824>. Accessed February 11, 2015.
25. Covaci A, Voorspoels S. Optimization of the determination of polybrominated diphenyl ethers in human serum using solid-phase extraction and gas chromatography-electron capture negative ionization mass spectrometry. *J Chromatogr B Analyt Technol Biomed Life Sci*. 2005;827(2):216-223. doi:10.1016/j.jchromb.2005.09.020.
26. Koppen G, Den Hond E, Nelen V, et al. Organochlorine and heavy metals in newborns: results from the Flemish Environment and Health Survey (FLEHS 2002-2006). *Environ Int*. 2009;35(7):1015-1022. doi:10.1016/j.envint.2009.05.002.
27. Midasch O, Drexler H, Hart N, Beckmann MW, Angerer J. Transplacental exposure of neonates to perfluorooctanesulfonate and perfluorooctanoate: a pilot study. *Int Arch Occup Environ Health*. 2007;80(7):643-648. doi:10.1007/s00420-006-0165-9.
28. Gao Y, De Galan S, De Brauwere A, Baeyens W, Leermakers M. Mercury speciation in hair by headspace injection-gas chromatography-atomic fluorescence spectrometry (methylmercury) and combustion-atomic absorption spectrometry (total Hg). *Talanta*. 2010;82(5):1919-1923. doi:10.1016/j.talanta.2010.08.012.
29. de Cock M, de Boer MR, Lamoree M, Legler J, van de Bor M. First year growth in relation to prenatal exposure to endocrine disruptors - a Dutch prospective cohort study. *Int J Environ Res Public Health*. 2014;11(7):7001-7021. doi:10.3390/ijerph110707001.
30. Janssen S, Dumont G, Fierens F, Mensink C. Spatial interpolation of air pollution measurements using CORINE land cover data. *Atmos Environ*. 2008;42(20):4884-4903. doi:10.1016/j.atmosenv.2008.02.043.

31. Covaci A, Voorspoels S, Thomsen C, van Bavel B, Neels H. Evaluation of total lipids using enzymatic methods for the normalization of persistent organic pollutant levels in serum. *Sci Total Environ*. 2006;366(1):361-366. doi:10.1016/j.scitotenv.2006.03.006.
32. Remy S, Govarts E, Bruckers L, et al. Expression of the sFLT1 Gene in Cord Blood Cells Is Associated to Maternal Arsenic Exposure and Decreased Birth Weight. *PLoS One*. 2014;9(3):e92677. doi:10.1371/journal.pone.0092677.
33. R Development Core Team R. R: A Language and Environment for Statistical Computing. *R Found Stat Comput*. 2014;1:409. doi:10.1007/978-3-540-74686-7.
34. Fox J, Weisberg S. *An {R} Companion to Applied Regression.*; 2002. doi:10.1177/0049124105277200.
35. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc*. 1995;57:289-300. doi:10.2307/2346101.
36. Alexa A, Rahnenfuhrer J. topGO: topGO: Enrichment analysis for Gene Ontology. 2010.
37. Lian I a, Langaas M, Moses E, Johansson Å. Differential gene expression at the maternal-fetal interface in preeclampsia is influenced by gestational age. *PLoS One*. 2013;8(7):e69848. doi:10.1371/journal.pone.0069848.
38. Gützow KB, Haug LS, Thomsen C, Sabaredzovic A, Becher G, Brunborg G. Placental transfer of perfluorinated compounds is selective--a Norwegian Mother and Child sub-cohort study. *Int J Hyg Environ Health*. 2012;215(2):216-219. doi:10.1016/j.ijheh.2011.08.011.
39. Fromme H, Mosch C, Morovitz M, et al. Pre- and postnatal exposure to perfluorinated compounds (PFCs). *Environ Sci Technol*. 2010;44(18):7123-7129. doi:10.1021/es101184f.
40. Sala M, Ribas-Fitó N, Cardo E, et al. Levels of hexachlorobenzene and other organochlorine compounds in cord blood: exposure across placenta. *Chemosphere*. 43(4-7):895-901. <http://www.ncbi.nlm.nih.gov/pubmed/11372882>. Accessed November 23, 2015.
41. Bergonzi R, Specchia C, Dinolfo M, et al. Distribution of persistent organochlorine pollutants in maternal and foetal tissues: data from an Italian polluted urban area. *Chemosphere*. 2009;76(6):747-754. doi:10.1016/j.chemosphere.2009.05.026.
42. Burse VW, Najam AR, Williams CC, et al. Utilization of umbilical cords to assess in utero exposure to persistent pesticides and polychlorinated biphenyls. *J Expo Anal Environ Epidemiol*. 10(6 Pt 2):776-788. <http://www.ncbi.nlm.nih.gov/pubmed/11138670>. Accessed November 23, 2015.
43. Vizcaino E, Grimalt JO, Glomstad B, Fernández-Somoano A, Tardón A. Gestational weight gain and exposure of newborns to persistent organic pollutants. *Environ Health Perspect*. 2014;122(8):873-879. doi:10.1289/ehp.1306758.
44. Covaci A, Jorens P, Jacquemyn Y, Schepens P. Distribution of PCBs and organochlorine pesticides in umbilical cord and maternal serum. *Sci Total Environ*. 2002;298(1-3):45-53. <http://www.ncbi.nlm.nih.gov/pubmed/12449328>. Accessed November 23, 2015.
45. de Cock M, de Boer MR, Lamoree M, Legler J, van de Bor M. Prenatal exposure to endocrine disrupting chemicals in relation to thyroid hormone levels in infants - a Dutch prospective cohort study. *Environ Health*. 2014;13:106. doi:10.1186/1476-069X-13-106.
46. Verhulst SL, Nelen V, Hond E Den, et al. Intrauterine exposure to environmental pollutants and body mass index during the first 3 years of life. *Environ Health Perspect*. 2009;117(1):122-126. doi:10.1289/ehp.0800003.
47. Delvaux I, Van Cauwenberghe J, Den Hond E, et al. Prenatal exposure to environmental contaminants and body composition at age 7-9 years. *Environ Res*. 2014;132:24-32. doi:10.1016/j.envres.2014.03.019.

48. Moisiadis VG, Matthews SG. Glucocorticoids and fetal programming part 1: outcomes. *Nat Rev Endocrinol*. 2014;1-12. doi:10.1038/nrendo.2014.73.
49. Fowden AL, Li J, Forhead AJ. Glucocorticoids and the preparation for life after birth: are there long-term consequences of the life insurance? *Proc Nutr Soc*. 1998;57(1):113-122. <http://www.ncbi.nlm.nih.gov/pubmed/9571716>. Accessed May 31, 2015.
50. Matthews SG. Antenatal glucocorticoids and programming of the developing CNS. *Pediatr Res*. 2000;47(3):291-300. <http://www.ncbi.nlm.nih.gov/pubmed/10709726>. Accessed May 31, 2015.
51. Kapoor A, Petropoulos S, Matthews SG. Fetal programming of hypothalamic-pituitary-adrenal (HPA) axis function and behavior by synthetic glucocorticoids. *Brain Res Rev*. 2008;57(2):586-595. doi:10.1016/j.brainresrev.2007.06.013.
52. Gascon M, Sunyer J, Casas M, et al. Prenatal exposure to DDE and PCB 153 and respiratory health in early childhood: a meta-analysis. *Epidemiology*. 2014;25(4):544-553. doi:10.1097/EDE.0000000000000097.
53. Ahlsson F, Åkerud H, Schijven D, Olivier J, Sundström-Poromaa I. Gene Expression in Placentas From Nondiabetic Women Giving Birth to Large for Gestational Age Infants. *Reprod Sci*. 2015. doi:10.1177/1933719115578928.
54. Auffret J, Freemark M, Carré N, et al. Defective prolactin signaling impairs pancreatic β -cell development during the perinatal period. *Am J Physiol Endocrinol Metab*. 2013;305(10):E1309-E1318. doi:10.1152/ajpendo.00636.2012.
55. Brady MJ. IRS2 takes center stage in the development of type 2 diabetes. *J Clin Invest*. 2004;114(7):886-888. doi:10.1172/JCI23108.

Table 1 – Descriptive statistics of the study population

Categorical scale							
Determinant	N	Category	Count	Percent			
Maternal body mass index before pregnancy (kg/m ²)	193	<18.5	12	6%			
		[18.5, 25[	139	72%			
		[25, 30[	28	15%			
		>= 30	14	7%			
Newborn Sex	195	Male	98	50%			
		Female	97	50%			
Parity	194	Newborn is first child	75	39%			
		Newborn is second child	65	34%			
		Newborn is third child (or more)	54	28%			
Smoking	190	No smoking during pregnancy	166	87%			
		Smoking during pregnancy	24	13%			
Season of birth	195	Winter	98	50%			
		Spring	33	17%			
		Summer	19	10%			
		Autumn	45	23%			
		<= 25	22	11%			
Maternal age (years)	195	] 25 , 30]	73	37%			
		] 30 , 35]	74	38%			
		> 35	26	13%			
		Vaginal delivery	185	95%			
Type of delivery	195	Caesarean section	10	5%			
		< 1 consumption of alcoholic beverages per month	163	84%			
Alcohol use	194	>= 1 consumption of alcoholic beverages per month	31	16%			
		<= 37 weeks	15	8%			
Gestational age at birth	192	38 weeks	29	15%			
		39 weeks	50	26%			
		40 weeks	67	35%			
		>= 41 weeks	31	16%			
Continuous scale							
Determinant	N	Geomean [95% CI]	Minimum	P25	P50	P75	Maximum
Lipid content in cord blood (mg/dL)	193	206 [199 ; 212]	139	177	200	226	544

Abbreviations: N: Number of participants; P25: 25th percentile; P50: 50th percentile; P75: 75th percentile

Table 2 - Descriptive statistics of exposure biomarkers

Exposure Biomarker	Matrix	N	> LOD / LOQ	Geomean [95% CI]	Minimum	P25	P50	P75	Maximum
Endocrine disrupting chemicals (EDCs)									
Dichlorodiphenyldichloroethylene (ng/L)	Cord plasma	195	195	161 [145 ; 178]	34	91	157	255	1904
Dioxin-like compounds (pg Calux TEQ/L)	Cord plasma	180	170	33,68 [31,38 ; 36,15]	9,04	26,10	35,61	45,45	105,40
Hexachlorobenzene (ng/L)	Cord plasma	195	98	19 [18 ; 21]	10	10	10	21	36
mono-(2-ethyl-5-carboxypentyl) phthalate (ng/mL)	Cord plasma	172	172	0,72 [0,64 ; 0,82]	0,13	0,39	0,65	1,30	4,60
Polychlorinated biphenyl-153 (ng/L)	Cord plasma	195	188	55 [50 ; 60]	10	40	54	85	245
Perfluorooctane sulfonate (µg/L)	Cord plasma	170	170	2,6 [2,4 ; 2,8]	0,8	1,7	2,7	3,7	17,3
Perfluorooctanoic acid (µg/L)	Cord plasma	170	170	1,5 [1,4 ; 1,6]	0,5	1,1	1,6	2,1	4,3
Other environmental exposures									
Arsenic (µg/L)	Cord blood	189	188	0,56 [0,47 ; 0,66]	0,01	0,25	0,53	1,22	14,40
Cadmium (µg/L)	Maternal blood	189	188	0,33 [0,30 ; 0,36]	0,03	0,22	0,30	0,44	2,23
Copper (µg/L)	Cord blood	189	189	604,37 [587,85 ; 621,35]	298,66	535,01	608,16	686,32	993,67
Lead (µg/L)	Cord blood	189	189	8,90 [8,26 ; 9,60]	2,42	6,74	8,57	11,67	67,40
Manganese (µg/L)	Cord blood	189	189	31,85 [30,24 ; 33,54]	7,29	25,84	31,18	39,91	80,12
Methylmercury (µg/G)	Maternal Hair	191	191	0,24 [0,22 ; 0,28]	0,02	0,16	0,27	0,42	1,37
Particulate matter (<= 2.5µM) (µg/m³)	RIO model	192	192	19,81 [19,53 ; 20,08]	13,86	18,60	19,91	21,39	24,36
Thallium (µg/L)	Cord blood	189	189	0,017 [0,016 ; 0,018]	0,008	0,014	0,017	0,021	0,041

* Hexachlorobenzene was included as a dichotomous variable in the gene expression analysis (<LOQ vs. >LOQ).
Abbreviations: N: Number of participants; P25: 25th percentile; P50: 50th percentile; P75: 75th percentile; 95% CI: 95% confidence interval; LOD: Limit of detection; LOQ: Limit of quantification

Table 3 – Biomarkers of exposure: Correlation matrix

	Arsenic	Cadmium	Copper	Manganese	Lead	Thallium	Dioxin-like compounds	Dichlorodiphenylchloroethylene	Methylmercury	Polychlorinated biphenyl-153	Perfluorooctanoic acid	Perfluorooctane sulfonate	Particulate matter (<=2.5µM)
Arsenic													
Cadmium	-0.07 (0.3168)												
Copper	0.07 (0.3599)	-0.02 (0.8098)											
Manganese	0.14 (0.0468)	0.01 (0.9444)	0.38 (<0.0001)										
Lead	0.26 (0.0003)	0.11 (0.1438)	0.15 (0.0402)	0.24 (0.001)									
Thallium	0.03 (0.6345)	0.12 (0.1014)	0.26 (0.0003)	0.12 (0.1104)	0.14 (0.0518)								
Dioxin-like compounds	-0.01 (0.915)	0 (0.9517)	-0.06 (0.4643)	-0.08 (0.3079)	-0.06 (0.458)	-0.11 (0.1464)							
Dichlorodiphenylchloroethylene	-0.01 (0.8849)	-0.08 (0.2645)	0.04 (0.6156)	-0.09 (0.1963)	0.2 (0.0057)	-0.01 (0.8896)	0.12 (0.1134)						
Methylmercury	0.49 (<0.0001)	-0.1 (0.194)	0.09 (0.2364)	0.14 (0.0531)	0.19 (0.0106)	0.02 (0.7807)	0 (0.9638)	0.1 (0.1784)					
Polychlorinated biphenyl-153	0.25 (0.0005)	0.04 (0.6046)	0.08 (0.2584)	0.07 (0.3053)	0.24 (0.0009)	-0.05 (0.4838)	0.19 (0.0106)	0.52 (<0.0001)	0.29 (<0.0001)				
Perfluorooctanoic acid	0.14 (0.0706)	0.02 (0.7711)	-0.07 (0.3445)	-0.15 (0.0539)	0.19 (0.0148)	0.02 (0.8086)	0.2 (0.0104)	0.08 (0.2793)	0.22 (0.005)	0.23 (0.0027)			
Perfluorooctane sulfonate	0.11 (0.1499)	-0.2 (0.0095)	-0.1 (0.2047)	-0.18 (0.0186)	0.14 (0.0819)	-0.01 (0.8555)	0.15 (0.061)	0.23 (0.0028)	0.34 (<0.0001)	0.3 (<0.0001)	0.47 (<0.0001)		
Particulate matter (<=2.5µM)	-0.08 (0.2881)	0.08 (0.2978)	-0.16 (0.0318)	0.03 (0.6506)	0.01 (0.9086)	-0.09 (0.2126)	0.24 (0.0011)	0.03 (0.7258)	-0.07 (0.3286)	0.12 (0.0945)	-0.02 (0.7647)	-0.01 (0.9411)	
mono-(2-ethyl-5-carboxypentyl) phthalate	-0.08 (0.3199)	-0.03 (0.6589)	0.2 (0.0113)	-0.07 (0.3629)	0.01 (0.8545)	0.17 (0.031)	-0.02 (0.8108)	0.11 (0.1701)	-0.01 (0.8473)	0.05 (0.4975)	0.17 (0.0436)	0.15 (0.0773)	-0.07 (0.3587)

Notation: Pearson correlation coefficient (p-value), significant correlations (p < 0.05) are shown in bold font

Table 4 – Overview of transcription factors that mapped to ‘glucose metabolism disorders’, ‘insulin resistance’, ‘obesity’, or ‘weight gain’ and for which the lists of EDC associated genes (p-value < 0.05) were significantly enriched in downstream targets (p-value < 0.05) for at least one of the EDCs. Transcription factors were ordered by the number of EDCs that showed significant enrichment and subsequently alphabetically sorted.

		Glucose Metabolism Disorder	Insulin Resistance	Obesity	Weight Gain				
Symbol	Upstream Regulator					p,p'-DDE	PCB-153	PFOA	PFOS
DDIT3	DNA-damage-inducible transcript 3					N = 7+ *	N = 5+ *	N = 3 *	
CEBPA	CCAAT/enhancer binding protein (C/EBP), alpha					N = 27 * Inh.		N = 15 *	
CLOCK	clock circadian regulator					N = 11+ *	N = 9 *		
FOS	FBJ murine osteosarcoma viral oncogene homolog					N = 34 *		N = 14 *	
FOXO3	forkhead box O3					N = 14 *			N = 3 *
HMGA1	high mobility group AT-hook 1						N = 9 *	N = 6 *	
MYC	v-myc avian myelocytomatosis viral oncogene homolog					N = 49 *	N = 44 *		
NR3C1	nuclear receptor subfamily 3, group C, member 1 (glucocorticoid receptor)					N = 48 *	N = 30 *		
PGR	progesterone receptor							N = 8 *	N = 3 *
TP53	tumor protein p53					N = 68 *	N = 60 *		
TSC22D3	TSC22 domain family, member 3					N = 4 *	N = 3 *		
ATF3	activating transcription factor 3					N = 6 *			
ATF6	activating transcription factor 6					N = 4 *			
BTG2	BTG family, member 2					N = 3+ *			
CEBPB	CCAAT/enhancer binding protein (C/EBP), beta					N = 21+ * Inh.			
CREB1	cAMP responsive element binding protein 1					N = 34 * Inh.			
CREBBP	CREB binding protein					N = 14 *			
CREM	cAMP responsive element modulator					N = 12 *			
CTNNB1	catenin (cadherin-associated protein), beta 1, 88kDa					N = 33 *			
FOXC1	forkhead box C1					N = 3 *			
FOXC2	forkhead box C2					N = 5 *			
FOXO4	forkhead box O4					N = 7 *			
FOXP3	forkhead box P3						N = 6 *		
KLF5	Kruppel-like factor 5 (intestinal)							N = 3 *	
NR0B2	nuclear receptor subfamily 0, group B, member 2						N = 5 *		
NR1H4	nuclear receptor subfamily 1, group H, member 4						N = 12 *		
NR4A1	nuclear receptor subfamily 4, group A, member 1					N = 13 *			
NR4A2	nuclear receptor subfamily 4, group A, member 2					N = 6 *			
PAX6	paired box 6							N = 9 *	
PPARGC1A	peroxisome proliferator-activated receptor gamma, coactivator 1 alpha					N = 12 *			
PYCARD	PYD and CARD domain containing					N = 4 *			
REL	v-rel avian reticuloendotheliosis viral oncogene homolog					N = 9 *			
RUNX1	runt-related transcription factor 1					N = 12 *			
RUNX3	runt-related transcription factor 3					N = 5 *			
SMARCB1	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily b, member 1					N = 13 *			
STAT3	signal transducer and activator of transcription 3 (acute-phase response factor)					N = 29+ * Inh.			
STAT4	signal transducer and activator of transcription 4							N = 7 *	
STAT5B	signal transducer and activator of transcription 5B					N = 12 *			
TFAM	transcription factor A, mitochondrial					N = 3 *			

Symbols: N = ‘Number of target molecules significantly associated with exposure’ + ‘Expression level of transcription factor itself is associated with exposure’ | * ‘p-value transcription factor smaller than 0.05’ | Inh. ‘state of transcription factor is inhibited (Z <= -2)’

Table 5 – Screening for key molecules in disease specific pathways, irrespective of pathway significance. The table includes the list of genes that were significantly associated to the different EDCs, which map to the pathways that were selected. For pathways that were significantly enriched ($p < 0.05$), the p-value is given in front of the gene list.

	p,p'-DDE	PCB-153	PFOA	PFOS
Disease specific pathways				
Leptin Signaling in Obesity	ADCY1 ; STAT3 ; GRB2 ; PRKAR1B ; PIK3R1 ; PIK3CB	MAPK1 ; PRKACB	PIK3R6	PLCL2
Maturity Onset Diabetes of Young (MODY) Signalling	FOXA2	CACNA1S ; HNF1A		
Type I Diabetes Mellitus Signaling	HLA-DRB1 ; SOCS5 ; IL1B ; HLA-C ; ICA1	BCL2 ; MAPK1 ; NFKB2	HLA-G ; ICA1	
Type II Diabetes Mellitus Signaling	IRS2 ; SOCS5 ; SMPD2 ; CEBPB ; KCNJ11 ; PRKCB ; PIK3R1 ; PIK3CB ; SLC27A6	MAPK1 ; NFKB2 ; SLC27A2	PIK3R6 ; ACSL1	
Signaling Pathways				
Androgen Signaling	GNAZ ; GNA11 ; CCND1 ; EP300 ; PRKCB ; PRKAR1B ; GNAI3 ; SHC1	NFKB2 ; SHBG ; GNB2L1 ; MRAS ; GNG7 ; MAPK1 ; PRKACB	GNB3	
Estrogen Receptor Signaling	EP300 ; GRB2 ; MED23 ; SHC1	MAPK1 ; MRAS ; PHB2	MED20	
Glucocorticoid Receptor Signaling	IL1B ; EP300 ; SMARCA2 ; TGFB2 ; PIK3R1 ; PPP3CA ; HSPA6 ; PIK3CB ; STAT3 ; DUSP1 ; CEBPB ; GRB2 ; SHC1	BCL2 ; MAPK1 ; MRAS ; NPPA ; PRKACB ; PRL ; TAT	HSPA8 ; ICAM1 ; IL1R2 ; PIK3R6	FKBP5 ; HSP90AB1
Insulin Receptor Signaling	P = 0.012 PPP1CC ; IRS2 ; VAMP2 ; TSC2 ; GRB2 ; PRKAR1B ; GAB1 ; PIK3R1 ; CBL ; PIK3CB ; STX4 ; SHC1	MAPK1 ; MRAS ; PRKACB	PIK3R6	
PPAR Signaling	IL18 ; IL1B ; EP300 ; PPARA ; GRB2 ; IL36B ; SHC1	MAPK1 ; MRAS ; NFKB2	IL1R2	HSP90AB1
PPAR α /RXR α Activation	ADCY1 ; GNA11 ; IL1B ; EP300 ; PPARA ; GRB2 ; TGFB2 ; PRKCB ; PRKAR1B ; MED23 ; CLOCK ; SHC1	FASN ; MAPK1 ; MRAS ; NFKB2 ; PRKACB	ACVR2B ; IL1R2	PLCL2 ; HSP90AB1

Table 6 - Pathway significance of pathways categorized as ‘immune cell function’ and ‘cytokine signaling’. The table includes the list of genes that are significantly associated to the different EDCs which map to the pathways categorized as ‘immune cell function’ and ‘cytokine signaling’. The subset of pathways that was significantly enriched for at least one of the EDCs (p-value < 0.05, when significant the p-value is given in front of the gene list).

Canonical Pathway	p,p'-DDE	PCB-153	PFOA	PFO5
Acute Phase Response Signaling	P = 0.004 SOCS5 ; IL1B ; ITH3 ; IL6R ; IL36B ; RBP4 ; PIK3R1 ; PIK3CB ; SOD2 ; STAT3 ; IL18 ; CEBPB ; GRB2 ; C9 ; HAMP ; SHC1	HNFI1A ; MAPK1 ; MRAS ; NFKB2 ; ORM2	SOD2	
Dendritic Cell Maturation	P = 0.013 IL1B ; EP300 ; IL36B ; PIK3R1 ; LY75 ; TREM2 ; PIK3CB ; FCGR2A ; IL18 ; IRF8 ; HLA-DRB1 ; HLA-C ; TLR2 ; FCGR3A/FCGR3B	FCGR2B ; FSCN3 ; IL23A ; MAPK1 ; NFKB2	ICAM1 ; PIK3R6	PLCL2
Fcγ3 Receptor-mediated Phagocytosis in Macrophages and Monocytes	P = 0.008 GAB2 ; ACTG1 ; ACTB ; NCF1 ; PRKCB ; PIK3R1 ; ARF6 ; CBL ; FCGR2A ; FCGR3A/FCGR3B	MAPK1 ; PIP5K1A ; TLN2	PLD1 ; RAC2	
FLT3 Signaling in Hematopoietic Progenitor Cells	P = 0.015 STAT3 ; GAB2 ; EP300 ; GRB2 ; PIK3R1 ; CBL ; PIK3CB ; SHC1	MRAS ; MAPK1 ; FLT3LG	PIK3R6	
GMI-CSF Signaling	P = 0.002 STAT3 ; CSF2RA ; CCND1 ; GRB2 ; PRKCB ; PIK3R1 ; PPP3CA ; PIK3CB ; SHC1	GNB2L1 ; MRAS ; MAPK1 ; CISH	PIK3R6	
Graft-versus-Host Disease Signaling	P = 0.028 IL18 ; HLA-DRB1 ; IL1B ; HLA-C ; IL36B		HLA-G	
IL-12 Signaling and Production in Macrophages	P = 0.03 IL18 ; RAB7A ; IRF8 ; CEBPB ; EP300 ; TGFB2 ; PRKCB ; RBP4 ; PIK3R1 ; PIK3CB ; TLR2	IL23A ; MAPK1 ; NFKB2 ; ORM2	LYZ ; PIK3R6	
IL-3 Signaling	P = 0.015 STAT3 ; GAB2 ; GRB2 ; PRKCB ; PIK3R1 ; PPP3CA ; PIK3CB ; SHC1	MRAS ; MAPK1 ; IL3RA	PIK3R6	
IL-6 Signaling	P = 0.016 IL18 ; STAT3 ; IL1B ; MCL1 ; CEBPB ; IL6R ; GRB2 ; IL36B ; PIK3R1 ; PIK3CB ; SHC1	MAPK1 ; MRAS ; NFKB2 ; SRF	PIK3R6 ; CD14 ; IL1R2 ; MAPKAPK2	
Leukocyte Extravasation Signaling	P = 0.013 MLLT4 ; MMP9 ; NCF1 ; PIK3R1 ; ROCK2 ; PIK3CB ; CTNND1 ; RASGRP1 ; GNAI3 ; CDH5 ; ITGA1 ; ACTG1 ; ACTB ; PRKCB ; CYBB ; PECAM1	ITGA2 ; MAPK1 ; MMP2 ; RHOH	PIK3R6 ; ITGB1 ; RAC2 ; ICAM1	
Natural Killer Cell Signaling	P = 0.02 KIR2DS4 (includes others) ; GRB2 ; PRKCB ; KIR2DL4 ; PIK3R1 ; PIK3CB ; HCST ; SHC1 ; FCGR2A ; FCGR3A/FCGR3B	LILRB1 ; MAPK1 ; MRAS	P = 0.029 KLRD1 ; PIK3R6 ; KLRCL1 ; RAC2 ; SH2D1B	
p38 MAPK Signaling	P = 0.013 PLA2G2D ; IL18 ; DUSP1 ; IL1B ; EP300 ; TGFB2 ; DDIT3 ; IL36B ; MAX ; IRAK3 ; CDC25B	DDIT3 ; SRF	P = 0.038 MKNK1 ; PLA2G12B ; IL1R2 ; EEF2K ; MAPKAPK2	
Prolactin Signaling	P = 0.005 STAT3 ; SOCS5 ; CEBPB ; EP300 ; GRB2 ; PRKCB ; PIK3R1 ; PIK3CB ; SHC1	MRAS ; PRL ; MAPK1	PIK3R6	
Role of JAK2 in Hormone-like Cytokine Signaling	P = 0.014 SH2B3 ; STAT3 ; IRS2 ; SOCS5 ; SHC1	PRL	SIRPA	
Role of Pattern Recognition Receptors in Recognition of Bacteria and Viruses	P = 0.006 EIF2S1 ; IL18 ; IL1B ; C5AR1 ; TGFB2 ; PRKCB ; RIPK2 ; PIK3R1 ; PIK3CB ; TLR2 ; CLEC7A ; PTX3	IL17B ; MAPK1 ; NFKB2	PIK3R6	
Toll-like Receptor Signaling	P = 0.042 IL18 ; IL1B ; TNFAIP3 ; PPARA ; IL36B ; TLR2 ; IRAK3	MAPK1 ; NFKB2		
TREM1 Signaling	P = 0.012 IL18 ; STAT3 ; IL1B ; MPO ; GRB2 ; TREM1 ; NLRP6 ; TLR2	NFKB2 ; FCGR2B ; MAPK1 ; NLRP6	ITGB1 ; ICAM1	

Figure 1 – Schematic overview of the workflow of data analysis as illustrated hypothetically for gene X. A two-tiered approach was followed. Initially, all determinants in the dataset were considered, i.e. both exposure biomarkers and characteristics of the newborns and their mothers. In the first tier, associations between the panel of potential determinants and the expression of gene X were investigated in a univariate manner. For associations of which the p-value was smaller than 0.2, data were passed to the second tier. In the second tier, the most predictive set of determinants that passed tier 1 was determined by a stepwise regression algorithm. The output, a multiple linear regression model, was subsequently evaluated based on R^2 (adjusted for the number of determinants in the model). When R^2 was lower than 5%, we concluded that the dataset does not provide sufficient information to explain the expression of gene X. When R^2 was higher than 5% we evaluated whether the determinants were significantly associated with the expression level (p-value coefficients < 0.05). If so, we concluded that the expression level of gene X in cord blood cells was associated to the specified determinants. The same procedure was repeated for all other genes (or sequences) that successfully passed quality control filtering. As such a list of genes associated to each of the determinants was constructed.

Figure 2 – Number of genes associated with the different determinants in the dataset. Functional analysis of the genelists associated with the chemicals highlighted in grey has been performed.

Figure 3 - Added variable plots that show the relationship between chemical exposure and gene expression for two genes. 3a) Increase in expression of METRNL for increasing concentrations of PFOA ($R^2 = 0.49$). The other explanatory variables in the model were cadmium, newborn sex, and gestational age. The association between expression and exposure was different for boys and girls (as shown by effect modification). In girls, a 1.40 fold increase in expression was observed for each ln-unit increase in exposure (p-value 4.59E-

0.8) while a 1.13 fold increase in expression was observed for boys (p-value 0.049). 3b) Association between PRL expression and PCB-153 cord blood levels ($R^2 = 0.06$). A 1.09 fold-increase in expression was observed for each ln-unit increase in PCB-153 levels. The association was adjusted for gestational age. Other detailed outcome statistics of the linear multiple regression models are included in **supplementary table 2**.

Supplementary Material

Supplementary figure 1 – Sensitivity analysis - Multiple hypothesis testing: overview of the number of probes that are associated with the different determinants based on univariate analysis when considering either p-value below 0.2 or FDR below 0.2.

Supplementary table 1 – List of pathways categorized as ‘cellular immune function’ or ‘cytokine signaling pathway’ based on Ingenuity

Supplementary table 2 – Outcome statistics for each of the gene transcript

Supplementary table 3 – List of Associated pathways for each of the transcription factors listed in table 4

Supplementary table 4 – List of enriched gene ontology terms (biological processes) for each of the EDCs.

Supplementary table 5 – Sensitivity analysis - Impact of lowering the maximal number of steps in the stepwise regression algorithm from 10 to 5.

Figure 1

Figure 2

Figure 3a

Added variable plot for PFOA - Boys

Figure 3b

Supplementary Figure 1
[Click here to download Supplementary Material: SupplFigure1.pdf](#)

Supplementary Table 1
[Click here to download Supplementary Material: SupplTable1.xls](#)

Supplementary Table 2
[Click here to download Supplementary Material: SupplTable2.xlsx](#)

Supplementary Table 3
[Click here to download Supplementary Material: SupplTable3.xlsx](#)

Supplementary Table 4
[Click here to download Supplementary Material: SupplTable4.xlsx](#)

Supplementary Table 5
[Click here to download Supplementary Material: SupplTable5.docx](#)

***Conflict of Interest Statement**

[Click here to download Conflict of Interest Statement: 20150717_ConflictOfInterest.pdf](#)