A multi-omic analysis of birthweight in newborn cord blood reveals new underlying mechanisms related to cholesterol metabolism

SPONSORED DOCUMENT FROM

METABOLISM

ELSEVIER
FREE Full-Text Article

Metabolism. 2020 Sep; 110: 154292. PMCID: PMC7450273
doi: 10.1016/j.metabol.2020.154292: 10.1016/j.metabol.2020.154292 PMID: 32553738

A multi-omic analysis of birthweight in newborn cord blood reveals new underlying mechanisms related to cholesterol metabolism

Rossella Alfano,[a,b,c,1] Marc Chadeau-Hyam,[a,b,d,1] Akram Ghantous,[e,2] Pekka Keski-Rahkonen,[e,2] Leda Chatzi,[f,g] Almudena Espin Perez,[h] Zdenko Herceg,[e] Manolis Kogevinas,[i,j,k,l] Theo M. de Kok,[m] Tim S. Nawrot,[c,n] Alexei Novoloaca,[e] Chirag J. Patel,[o] Costanza Pizzi,[p] Nivonirina Robinot,[e] Franca Rusconi,[q] Augustin Scalbert,[e] Jordi Sunyer,[j,k,l] Roel Vermeulen,[a,d] Martine Vrijheid,[i,j,l] Paolo Vineis,[a,b,r] Oliver Robinson,[a,3] and Michelle Plusquin[a,b,c,3,*]

[a]Department of Epidemiology and Biostatistics, The School of Public Health, Imperial College London, London, United Kingdom

[b]Medical Research Council-Health Protection Agency Centre for Environment and Health, Imperial College London, London, United Kingdom

[c]Centre for Environmental Sciences, Hasselt University, Diepenbeek, Belgium

[d]Institute for Risk Assessment Sciences (IRAS), Division of Environmental Epidemiology, Utrecht University, Utrecht, the Netherlands

[e]International Agency for Research on Cancer (IARC), 150 Cours Albert Thomas, 69008 Lyon, France

[f]Department of Preventive Medicine, University of Southern California, Los Angeles, CA 90007, United States

[g]Department of Social Medicine, University of Crete, Heraklion, Crete, Greece

[h]Department of Biomedical Informatics Research, Stanford University, CA, United States

[i]Consortium for Biomedical Research in Epidemiology and Public Health (CIBERESP), Madrid, Spain

[j]ISGlobal, Centre for Research in Environmental Epidemiology (CREAL), Barcelona, Spain

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

[l]Universitat Pompeu Fabra (UPF), Barcelona, Catalonia, Spain

[m]Department of Toxicogenomics, Maastricht University, Maastricht, the Netherlands

[n]Environment & Health Unit, Leuven University, Leuven, Belgium

[o]Department of Biomedical Informatics, Harvard Medical School, Boston, MA 02115, United States

[p]Department of Medical Sciences, University of Turin and CPO-Piemonte, Torino, Italy

[q]Unit of Epidemiology, Anna Meyer Children's University Hospital, Florence, Italy

[r]Human Genetic Foundation (HuGeF), Turin, Italy

Michelle Plusquin: michelle.plusquin@uhasselt.be

*Corresponding author at: Centre for Environmental Sciences, Hasselt University, Agoralaan, building D, 3590 Diepenbeek, Belgium. michelle.plusquin@uhasselt.be

[1]Joint first authors.

[2]Joint second authors.

[3]Joint last authors.

Received 2019 Nov 18; Accepted 2020 Jun 11.

Abstract

Background

Birthweight reflects in utero exposures and later health evolution. Despite existing studies employing high-dimensional molecular measurements, the understanding of underlying mechanisms of birthweight remains limited.

Methods

To investigate the systems biology of birthweight, we cross-sectionally integrated the methylome, the transcriptome, the metabolome and a set of inflammatory proteins measured in cord blood samples, collected from four birth-cohorts ($n = 489$). We focused on two sets of 68 metabolites and 903 CpGs previously related to birthweight and investigated the correlation structures existing between these two sets and all other omic features via bipartite Pearson correlations.

Results

This dataset revealed that the set of metabolome and methylome signatures of birthweight have seven signals in common, including three metabolites [PC(34:2), plasmalogen PC(36:4)/PC(O-36:5), and a compound with *m/z* of 781.0545], two CpGs (on the *DHCR24* and *SC4MOL* gene), and two proteins (periostin and CCL22). CCL22, a macrophage-derived chemokine has not been previously identified in relation to birthweight. Since the results of the omics integration indicated the central role of cholesterol metabolism, we explored the association of cholesterol levels in cord blood with birthweight in the ENVIR*ON*AGE cohort ($n = 1097$), finding that higher birthweight was associated with increased high-density lipoprotein cholesterol and that high-density lipoprotein cholesterol was lower in small versus large for gestational age newborns.

Conclusions

Our data suggests that an integration of different omic-layers in addition to single omics studies is a useful approach to generate new hypotheses regarding biological mechanisms. CCL22 and cholesterol metabolism in cord blood play a mechanistic role in birthweight.

Abbreviations: AGA, adequate for gestational age; BMI, body mass index; DOHaD, Developmental Origin of Health and Disease; HDL, high-density lipoprotein; IL, interleukin; IQR, interquartile; LGA, large for gestational age; LDL, low-density lipoprotein; *m/z*, mass-to-charge ratio; ORA, overrepresentation analysis; SGA, small for gestational age; PC, phosphatidylcholine; U, unassigned metabolite; 95CI, 95% confidence interval

Keywords: Birth weight, Cholesterol, DNA methylation, Gene expression, Metabolome, Proteins

1. Introduction

The Developmental Origin of Health and Disease hypothesis (DOHaD) states that later life diseases may be influenced by experiences and conditions in prenatal life [1]. It is hypothesized that the interplay between genotype and in utero environmental factors induces molecular modifications and possibly phenotype differentiation in the fetus [2]. This developmental plasticity is of crucial importance for postnatal life but can induce impairments related to adverse health outcomes [3]. For example, exposure to detrimental environmental factors may induce birthweight changes that in turn may be associated with increased mortality or risk of cardiovascular diseases, mental health problems, and some cancers later in life [[4], [5], [6], [7]]. Among several mechanisms proposed to explain the link between in utero exposures, birthweight and health and diseases in later life, molecular markers identified through “omics” platforms, may play a central role. The theoretical foundation that drives the present study is that birthweight induces molecular modifications that in turn influence later life health, as exemplified in the Fig. 1, however is also possible that birthweight is influenced by molecular changes determined by in utero exposures.

Recently two studies based on high-dimensional molecular measurements in cord blood identified DNA methylation signals and metabolites associated with birthweight: 914 differentially methylated CpG sites were discovered in 8825 neonates from 24 birth-cohorts and 68 metabolites were identified in 499 neonates from four birth-cohorts [8,9]. These studies, together with several metabolomic, gene expression, proteomic, genomic, and epigenomic analyses, have increased our understanding of underlying mechanisms of birthweight [[10], [11], [12], [13], [14], [15], [16], [17], [18], [19], [20]]. However, a study integrating multi-omic levels in cord blood associated with birthweight in the same samples has not yet been performed.

To decipher at several levels the molecular cascades that regulate birthweight, in this paper we propose to integrate DNA-methylation, gene expression data as well as metabolic profiles and a set of inflammatory proteins measured in cord blood samples ($n = 489$) collected from four independent population-based birth-cohorts [21]. We used two birthweight-related sets of molecular signals, metabolites [8] and DNA methylation levels [9], to drive the integrated analyses and translated these signals to the other omic layers in the same samples. Based on results common to the metabolite- and methylation-driven multi-omic integrations, we aimed to identify key molecular associations with birthweight.

2. Material and methods

2.1. Study population and samples collection

Our study population arises from the EXPOsOMICS European project and includes 500 newborns from four population-based cohorts: 200 newborns from ENVIR*ON*AGE, 100 from INMA, 99 from Piccolipiù, and 101 from Rhea [[21], [22], [23], [24], [25]]. Inclusion criteria and protocols are detailed in the respective cohort descriptions and in the Supplementary methods. Before the placenta was delivered, whole blood was withdrawn from cord vessels and immediately frozen at −80 °C. Samples were sent to different laboratories for metabolome, inflammatory proteins and DNA methylome analysis [21]. The transcriptome was measured for the 200 ENVIR*ON*AGE samples participating in the EXPOsOMICS project and cholesterol was measured for the entire ENVIR*ON*AGE cohort.

2.2. Metabolomic profiles

Untargeted metabolomics was performed as previously described [8] and detailed in the Supplementary Methods. Briefly, reversed phase liquid chromatography-quadrupole time-of-flight mass spectrometry (UHPLC-QTOF-MS) system was used in positive ion mode with 499 of the 500 samples successfully analyzed. Raw data preprocessing was performed with Agilent MassHunter software, and metabolic features present in <60% of the samples were excluded, leaving 4712 features for 499 samples available for the subsequent analysis. Data were log-transformed and missing values were imputed using the impute.QRILC function within the "imputeLCMD" R package. Identification of the features of interest was done as previously described by Robinson et al. [8] and level of identification was reported as proposed by Sumner et al. [26].

2.3. DNA methylation profiles

Genomic DNA was extracted from buffy coats according to standard protocol and underwent bisulphite conversion using the Zymo EZ DNA methylationTM kit (Zymo, Irvine, CA, USA), hybridization to Illumina HumanMethylation 450K BeadChip arrays and scanning using an Illumina iScan. As detailed in the Supplementary Methods and elsewhere, we used in-house software to preprocess the data, including removal of probes based on signal intensities and control probes, background subtraction and dye bias correction [27]. After quality control and filtering, 417572 CpG sites for 460 samples were retained for subsequent analysis and methylation levels were expressed as beta values. To account for technically-induced and tissue-related variation in the methylation levels, we ran a preliminary linear model for the methylation beta values (as outcome variable) adjusting for technical variation (array row and position on

the chip), as well as estimated cell type composition using the Bakulski method [28].

2.4. Gene expression

Gene expression levels were measured in cord blood samples ($n = 200$) of the ENVIR*ON*AGE cohort. RNA was extracted using the total RNA miRNeasy mini kit (Qiagen, Venlo, Netherlands) according to the manufacturer's protocol. As detailed in the Supplementary Methods, samples were quality checked and further hybridized onto Agilent Whole Human Genome 8 × 60 K microarrays coupled with Agilent DNA G2505C Microarray Scanner. After preprocessing, quality control, and normalization detailed in the Supplementary methods, 29164 transcripts for 165 samples were left available for further analysis. To account for technical noise, we ran a linear model for the observed gene expression level (as outcome variable) adjusting for technical variation (hybridization date) and white blood cells count.

2.5. Inflammatory proteins

We measured 22 inflammation-related proteins using an R & D Luminex screening assay according to the protocol described by the manufacturer, and c-reactive protein using Solid Phase Sandwich ELISA. Excluding proteins detected in <40% of the study population, we were left with 16 inflammatory proteins (Supplementary Table 1) measured in 493 samples, of which three were excluded. For the remaining 490 samples missing values were imputed following an approach based on likelihood maximization estimation procedure [29]. Protein concentrations were subsequently log-transformed, and to correct for nuisance variation, we employed the same linear model approach described before setting the plate identification as technical covariate.

2.6. Cholesterol

The plasma levels of high-density lipoprotein (HDL), low-density lipoprotein (LDL), and total cholesterol were measured in entire ENVIR*ON*AGE population using Cobas 8000 C702 module analyzer (Roche, Basel, Switzerland). Outliers (>5 standard deviations from mean) were excluded from the analyses. Respectively 1139, 1109 and 1131 samples had valid HDL, LDL, and total cholesterol measurements.

2.7. Anthropometrics and covariates

Birthweight in grams was collected from medical records. In ENVIR*ON*AGE cohort, newborns were classified as small for gestational age (SGA), adequate for gestational age (AGA) or large for gestational age (LGA) if their birthweight for given gestational age, sex, and parity status was respectively below the 10th percentile, between 10th percentile and 90th percentile, or above the 90th percentile calculated for Flanders from the Study Centre for Perinatal Epidemiology (http://www.neonatologie.ugent.be/SPE-standaarden.pdf).

As detailed in the Supplementary Material, covariates were selected based on previous reported associations with birthweight and included: sex of the newborns, parity, gestational age, maternal and paternal age and body mass index (BMI), maternal smoking status during pregnancy, and maternal education.

2.8. Statistical analyses

The study design is schematically represented in Fig. 2.

2.8.1. Exploring correlation structure across omic profiles We adopted an exposome globe approach to investigate the correlation structures across the omic measurements available in our study population and investigated Pearson's correlation coefficients for pairs of omics measurement [30]. We used sets of molecular features from two omic platforms to drive our integrated analyses [31]: (a) a set of 68 metabolites previously associated with birthweight (Supplementary Table 2); and (b) a set of 903 (available

from the total of 914, Supplementary Table 2) CpGs previously associated with birthweight [8,9]. These sets were correlated with all the other untargeted omic features. The statistical significance of all correlation coefficients was assessed by deriving a z-score from Fisher transformation and running a Student's *t*-test test assessing the null hypothesis of no correlation $H_0 : \rho = 0$. We corrected for multiple testing using the stringent Bonferroni correction for the number of tests (i.e. the total number of pairs investigated) and considered significant the correlations with Bonferroni corrected p-values<0·05. The number of samples participating in each analysis is presented in Supplementary Table 3. Results were visualized by means of Circos plots (Circos software version 0.69–6), where only significant correlation coefficients were reported.

In order to assess if our results were biased by heterogeneity between the different cohorts, we used linear models adjusted on the factors differing across cohorts (namely: gestational age, parental ages, weights and heights, parity and maternal education) to test associations between features significantly correlated in our main analyses. Stratification by sex was performed as a sensitivity analysis in order to take into account the birthweight sexual dysmorphism [32].

Finally, we compared the metabolite- and the methylation-driven results to assess if any metabolite-CpG pairs and any omic were in common.

2.8.2. Pathway analysis We performed overrepresentation analyses (ORA) of all transcripts and CpGs significantly correlated in the metabolite-driven analyses and of the CpG significantly correlated with metabolites in the metabolite-driven analysis using ConsensuspathDB online tool (http://consensuspathdb.org/). A pathway was considered significantly enriched if p-values were smaller than 0·05 and included at least 3 genes.

Enriched metabolic pathways within metabolic features correlated with CpGs in the methylation-driven analyses were identified using the *mummichog* program (version 1.1.0) [33] through the MetaboAnalyst platform [34]. We used all mass-to-charge ratio (*m/z*) values and associated p-values of the metabolic features as software input and set *mummichog* parameters to 'positive mode' at ±5 ppm mass tolerance. The p-value cutoff to identify the list of significant *m/z* features was set to false discovery rate adjusted (FDR) p-value equal to 0·05, with the non-FDR significant features used as the reference set. The algorithm searches tentative compound lists from metabolite reference databases against an integrated model of human metabolism to identify functional activity. A pathway was considered significant if gamma adjusted p-values were smaller than 0·05. Visualization of enriched pathways on the KEGGscape network was performed through the MetaboAnalyst platform [34].

2.8.3. Correlation network analysis We selected all significantly correlated omic features in the main analyses and performed network correlation analysis. Correlation networks were plotted using "igraph" package with layout algorithm by Fruchterman and Reingold (version 0.7.1). Only nodes with degree >2 and edges with Bonferroni corrected p-values <0·05 were represented in the networks. Communities were detected using the Louvain algorithm.

2.8.4. In silico identification of methylation quantitative trait loci We searched SNPs associated with the CpG sites significantly correlated to the other omic layers in the publicly available methylation quantitative trait loci (mQTL) database (http://www.mqtldb.org/). We compared the localization on the genome of 233 SNPs associated with own birthweight and 128 SNPs associated with offspring birthweight in the NHGRI-EBI GWAS Catalog with the identified methylation signals (±2 Mb from the genetic variants) and we searched for mQTL associated with these SNPs [35].

2.8.5. Exploring key findings in the ENVIRONAGE cohort Based on omics identified in significant pairs from both the metabolite- and methylation-driven analyses, we generated a hypothesis on mechanisms underlying birthweight. Common omic signals that were significantly correlated with both the metabolite and CpGs birthweight-related sets, were associated with cholesterol metabolism. We ran linear regression models to assess (i) the associations between omics identified in our multi-omic analyses and cholesterol

levels and (ii) the association between birthweight and cholesterol levels. These analyses were restricted to the ENVIR*ON*AGE cohort. All the models were adjusted for gestational age, parity, newborn sex, maternal age, maternal height, maternal BMI, smoking during pregnancy and maternal education, total cholesterol levels (for HDL and LDL analyses), plate (for proteins analyses), cell types composition, array row and position on the chip (for CpGs analyses). Paternal age and anthropometric measurements were not included as adjustment covariates due to missingness. In addition, we used linear models adjusted for the covariates aforementioned to explore if being SGA, AGA and LGA (independent variable) was associated with the levels of omic markers identified and cholesterol (dependent variable). If p-values were smaller than 0·05 results were considered significant. Samples participating in each analysis are reported in the Supplementary Table 3.

3. Results

3.1. Population

Descriptive characteristics of the study population participating in the main multi-omic study are presented by cohort in Table 1 and by sex in Supplementary Table 4 and indicate heterogeneity across cohorts for all covariates, except proportion of girls born and maternal smoking habits during pregnancy.

3.2. Metabolite-driven integration of omics

By correlating the set of (n = 68) metabolites (Supplementary Table 2), previously reported in these same four birth-cohorts to be associated with birthweight by Robinson and colleagues, with the other omic layers we identified 347 significantly correlated omic pairs involving 208 omic features (47 metabolites, 15 inflammatory proteins, 71 transcripts and 75 CpGs) (Fig. 3A and Supplementary Table 5) [8].

Pairs involving transcripts were all positively correlated to metabolites (Supplementary Figs. 1A and 2B), 68% of the metabolite-inflammatory protein pairs were positively correlated (Supplementary Figs. 1C and 2C) while 82% of the significant metabolite-CpG pairs showed negative correlation coefficients (Supplementary Figs. 1B and 2A). The strongest correlation coefficients in absolute value were observed in pairs involving transcripts (absolute range r = 0·42–0·54), followed by pairs involving CpG sites (absolute range r = 0·28–0·39), and inflammatory proteins (absolute range r = 0·18–0·44) (Table 2).

We did not identify one single metabolite related to all three of the other omic layers (Fig. 3B and Supplementary Table 6). Progesterone was the annotated metabolite correlated with most omic features (n = 47), including both proteins and transcripts.

The identified transcriptome signals were involved in 31 significant pathways (p-values<0·05), mainly related to immune response [e.g. Interleukin (IL)12-mediated signaling events and natural killer cell mediated cytotoxicity] (Fig. 3C and Supplementary Table 7). Similarly, the identified CpGs signals were mapped into seven significant pathways (p-values<0·05) including TNFalfa, thermogenesis, and insulin signaling pathways (Fig. 3C and Supplementary Table 7).

To further characterize the 208 omic signals identified in the metabolite-driven analysis, we constructed a correlation network. This network analyses revealed that identified molecules were mainly grouped in distinct communities according to their omic layer (Fig. 3D). In all groups, except group 4, signals of other omics were also present, e.g. progesterone and other four unassigned metabolites (U4, U5, U6 and U46) that were grouped with the transcripts (group 2) (Supplementary Table 8). The network analysis unveiled novel correlations between proteins and CpGs groups but not between transcripts and CpGs or transcripts and proteins.

No mQTL was identified for the identified 75 CpG sites. One to six significant CpG sites were located ±2 Mb from 153 SNPs out of the 233 associated with own birthweight and from 39 SNPs out of the 128 SNPs associated with offspring birthweight (Supplementary Table 9). One SNP, associated with own

birthweight, was located in the same gene (*PDE4B*) as a significant CpG.

Sensitivity analyses agreed mostly with the main analyses, except for transcripts. 29 metabolites were still significantly associated with 32 omic features (17 CpG sites and 15 proteins) after adjustment for gestational age, parental ages, parental weights and heights, parity and maternal education (Supplementary Table 10). Out of the total 347 significant correlations from the main analyses, after stratification by sex, 62 correlations remained significant in boys and 46 in girls (Supplementary Fig. 3), three additional correlations became significant in boys and five in girls (Supplementary Table 11).

3.3. Methylation-driven integration of omics

By correlating a set of 903 CpG sites that have been previously associated with birthweight by Kupers and colleagues (Supplementary Table 2) with the other omic layers, we identified 482 significant pairs involving 241 omic measurements (58 CpGs, 157 transcripts, two proteins, 24 metabolic features) (Fig. 4A, Supplementary Table 12) [9].

As indicated in Table 3, most of the CpG-transcript and CpG-inflammatory protein pairs were negatively correlated (92% and 75% respectively, see Supplementary Figs. 4A,C and 5A–B), and the pairs involving metabolites were predominantly positively correlated (79%, Supplementary Figs. 4B and 5C).

The strongest correlations were observed in the CpG-transcript pairs (absolute range $r = 0{\cdot}45–0{\cdot}57$), followed by CpG-metabolic feature pairs (absolute range $r = 0{\cdot}26–0{\cdot}39$) and CpG-inflammatory protein pairs (absolute range $r = 0{\cdot}22–0{\cdot}24$).

No CpG site was correlated to all three types of omic data (Fig. 4D and Supplementary Table 13). *cg08217545* (located on the *NFIC* gene) was the CpGs involved in the most significant correlations ($n = 86$). Pathways analysis of the identified transcriptome signals resulted in 17 enriched pathways (Fig. 4 B and Supplementary Table 14), mostly involved in signal transduction and immune system (such as G beta:gamma signaling through PI3Kgamma and TNF signaling pathway).

The 24 metabolic features identified were found to represent 14 unique compounds, of which nine could be identified [Unidentifiable phosphatidylcholine (PC)/LysoPC, PC(30:0), PC(34:2), PC(36:4), Plasmalogen PC(38:4) or PC(O-38:5), Plasmalogen PC(36:4) or PC(O-36:5), Plasmalogen PC(36:3) or PC(O-36:4), Cholesterol, Cholestenone]. Full details on retention times masses, and levels of identification are reported in Supplementary Table 15, with the chromatograms and mass spectra in the Additional data. Pathways analysis for the metabolic signals identified ($n = 201$ metabolic features with 0·05 FDR adjusted p-values) revealed three significantly enriched pathways including C21-steroid hormone biosynthesis and metabolism, porphyrin metabolism and omega-3 fatty acid metabolism (Fig. 4C).

Network correlation analysis identified omics grouped in three multi-omic communities of transcripts and CpGs (groups 1, 2 and 5) and three communities mainly or uniquely populated by a single omic type, e.g. groups 3 and 4 are only made of transcripts (Fig. 4E and Supplementary Table 16). The network analysis unveiled novel correlations between metabolites and proteins, and metabolites and transcripts.

Three mQTLs were identified for the 58 significant CpG sites (Supplementary Table 17). None of these three mQTLs has been previously associated with birthweight. We identified one to seven significant CpG sites located ±2 Mb from 188 SNPs out of the 233 associated with own birthweight and from 96 SNPs out of the 128 associated with offspring birthweight (Supplementary Table 18). Only one SNP associated with own birthweight was located in the same gene (*PIM3*) as a significant CpG site.

In sensitivity analyses 13 CpGs were still significantly associated with 24 omics (seven metabolic features and 17 transcripts) after adjustment for gestational age, parental ages, weights and heights, parity and maternal education (Supplementary Table 19). After stratification by sex, of the 482 correlations significant in the main analyses four remained significant in boys and nine in girls (Supplementary Fig. 6), four additional correlations became significant in boys and four in girls (Supplementary Table 20).

3.4. Signals in common between the metabolite- and methylation-driven integration of omics

We identified seven features in common to both the metabolite- and the methylation-driven integration of omics (Fig. 5A and Supplementary Table 21) which are part of 48 unique correlation pairs.

The seven features include three metabolites [PC(34:2), plasmalogen PC(36:4)/PC(O-36:5), and an unidentifiable compound (U)61 of *m/z* 781.0545], two CpG sites (*cg17901584* on the *DHCR24* gene, and *cg05119988* on the *SC4MOL* gene), and two proteins (CCL22 and periostin).

No feature was in common between all omic-layers (Fig. 5B).

Despite that no single transcript was in common, pathways related to immune system and signal transduction were enriched in both metabolite- and the methylation-driven integration of omics (Fig. 5D) and the "chemokine signaling pathway" in particular was significant in both analyses, along with pathways involving IL-2 and JAK-STAT signaling (Supplementary Tables 7 and 14).

Network correlation analysis confirmed the three metabolites were correlated with the two CpGs (Fig. 5C).

No feature was found to be robust to adjustment for factors differing by cohorts (Supplementary Table 22).

As both genes in which the common CpG sites are located are involved in cholesterol biosynthesis we hypothesized that cholesterol metabolism is associated with birthweight and the newly identified omics signals and consequently followed-up our analyses with a verification study in the ENVIR*ON*AGE cohort.

3.5. Cholesterol analysis

To test our hypothesis that cholesterol is related to birthweight and birthweight related molecules, we analyzed the measured levels of each of the seven omics features, common to both the metabolite- and the methylation-driven omic integration, in relation to cord blood measurements of cholesterol available in the ENVIR*ON*AGE cohort (Supplementary Tables 3 and 23). Regressions models were adjusted for gestational age, parity, newborn sex, maternal age, maternal height, maternal BMI, smoking during pregnancy, maternal education, total cholesterol levels (for HDL and LDL analyses), plate (for proteins analyses), cell types composition, array row and position on the chip (for CpGs analyses). We found that an interquartile (IQR) increment of all three metabolic features [PC(34:2), plasmalogen PC(36:4)/PC(O-36:5) and U61], and the two CpG sites, *cg05119988* (on the *SC4MOL* gene) and *cg17901584* (on the *DHCR24* gene), were respectively associated with an increase in total cholesterol levels of 17, 27, 18, 6, 7 mg/dl (p-value<0·01 for all the associations) (Table 4). Additionally, a IQR-increment of plasmalogen PC(36:4)/PC(O-36:5) was associated with an increase of 5 mg/dl of HDL cholesterol (p-value = 0·01) (Table 4).

Analyses stratified by sex showed similar results for total cholesterol in girls and boys (only the p-value of association with CpGs lost statistical significance in girls). Associations involving HDL cholesterol lost statistical significance in the girls' analyses (Supplementary Table 24).

In a larger subset of the ENVIR*ON*AGE cohort (n = 1096) (Supplementary Tables 3 and 23), we found that an IQR-increment in birthweight (equal to 618 g) was associated with an increment 1·14 mg/dl of HDL cholesterol [95% confidence interval (95CI) = 0·43 mg/dl to 1·85 mg/dl, p-value = 1·71e-03] upon adjusted for gestational age, parity, newborn sex, maternal age, maternal height, maternal BMI, smoking during pregnancy and maternal education (Supplementary Table 25). Analyses stratified by sex confirmed this association in girls only (Supplementary Table 25).

Finally, we tested if the levels of omic markers identified and cholesterol differ in SGA, LGA and AGA newborns. Among the seven omic markers identified only methylation of the two CpG sites, *cg05119988* (on the *SC4MOL* gene) and *cg17901584* (on the *DHCR24* gene), was significantly higher in SGA compared to LGA (p-value = 0·03 and 0·01, respectively), and methylation of *cg17901584* was significantly lower in LGA compared to AGA (p-value = 0·03) (Supplementary Table 26). Total cholesterol

was significantly lower (estimate change = −4·68 mg/dl, 95CI = −8·84 mg/dl to –0·50 mg/dl) in SGA compared to AGA newborns (p-value = 0·03) and HDL cholesterol was significantly lower (estimate change = −2·17 mg/dl, 95CI = −4·24 mg/dl to −0·09 mg/dl) in SGA compared to LGA newborns (p-value = 0·04) (Fig. 6). In the analyses stratified by sex, HDL cholesterol levels were lower in SGA girls compared to both AGA (p-value = 0·05) and LGA (p-value = 0·03) (Supplementary Table 25).

4. Discussion

Through an in-depth exploration of birthweight-associated sets of metabolites and methylation sites we have identified commonalities and differences between signals from two different molecular layers [8,9]. From millions of correlations between omic molecules measured in cord blood, our study shows that the set of metabolome and methylome signatures of birthweight have seven signals in common, one of which, the macrophage-derived chemokine CCL22, has not been previously identified in relation to birthweight. CCL22 was negatively correlated to both the metabolite PC(C34:2) and *cg17901584* on the *DHCR24* gene. CCL22 plays a crucial role in the control of T cell immunity [36]. Similarly, we found that the "Chemokine signaling pathway" identified through the gene expression analysis, overlaps between the metabolite- and methylation-driven analyses, supporting a potential link between birthweight and the immune system.

Although a detailed discussion of specific molecules is beyond the purpose of the present study and requires further experimental validation, we highlight here the example of progesterone, which was the annotated metabolite most frequently correlated in metabolite-driven analyses. Higher levels of progesterone in cord blood are observed with lower-weight births [8,37]. Progesterone plays an important role in the suppression of immune responses promoting cord blood T cell differentiation [38]. Furthermore, we observed that progesterone was clustered with transcripts that were most enriched for the IL-12 signaling pathway, which promotes Th1 differentiation and forms a link between innate resistance and adaptive immunity [39]. In the methylation-driven analyses, JAK3 was the most frequently correlated transcript and the JAK-STAT signaling pathway, which plays a critical role orchestrating innate and adaptive immunity, was also significant enriched in the gene expression pathway analysis [40]. No previous study has linked this enzyme to birthweight, yet experimental evidence has associated JAK3 with low grade inflammation, obesity and metabolic syndrome [41]. Progesterone and JAK3 are two signals of many we identified, that illustrate how the cross-omic approach can provide deeper insight in the biological underlying mechanisms of birthweight and highlight avenues for further investigation.

In general, by comparing the metabolite- and methylation-driven analyses we could observe that: i) phosphatidylcholine metabolites, particularly plasmalogens, have been identified in both analyses. Maternal plasmalogens, that are able to cross the placenta, have been recently associated with newborn body composition [42]. However, most previous studies identified lysoPCs rather than PCs as dominant cord blood metabolites related to birthweight [15,18]. ii) In both analyses, metabolites were grouped with two CpGs, albeit different CpGs (*cg05119988* and *cg17901584* in the metabolite-driven analysis and *cg14195992* and *cg15331996* in the methylation-driven analysis). iii) The metabolite-driven network analysis showed relations to groups of features from distinct omic layers while in the methylation-driven analysis three groups contain a mix of different omic types, representing a more profound multi-omics signal. iv) In both analyses we consistently found stronger correlations with gene expression-methylation and gene expression-metabolites than between the other layers. v) Both analyses identified different gene expression signals which suggests that post-transcriptional regulation is specific for metabolites or methylation signals. vi) Metabolite-CpG pairs in the metabolite-driven analyses were mainly negatively correlated, while in the methylation-driven analyses the opposite was observed. The latter difference may arise from the fact that expression levels of genes are positively correlated with the level of methylation within the transcribed region and while only 12% of the CpGs related to metabolites in the metabolite-driven analysis were located in the gene body - this percentage increased to 46% in the methylation-driven analysis. In the methylation-driven analyses the various negative associations may be due to the large

proportion of negative correlations between transcripts and methylation sites (92%), conversely, a single cell study previously described a more complex relation depending on the location of CpG islands [43]. In addition, the analyses showed a general trend for metabolite candidates being positively correlated with gene expression, in agreement with previous literature [44].

CpGs in common between the methylation- and metabolite-driven integrations of omics belong to genes (*DHCR24* and *SC4MOL*) involved in cholesterol biosynthesis. Furthermore, C21-steroid hormone biosynthesis and metabolism was identified as an important metabolic pathway both through our methylation-driven analysis and in direct association with birthweight by Robinson et al. [8]. We therefore performed a verification study in the ENVIR*ON*AGE cohort and showed that plasmalogen PC(36:4)/PC(O-36:5) was positively associated with HDL cholesterol levels that in turn were positively associated with higher birthweight. Phosphatidylcholine metabolism, and in particular plasmalogens, may regulate several important cholesterol biosynthesis processes which improve cholesterol sensing and facilitate interorganelle cholesterol trafficking [45]. During fetal development cholesterol and phospholipids are needed to build membranes, to develop the central nervous system including the brain and they are precursors of bile acids and steroid hormones [46]. While in adults LDLs are the major plasma lipoproteins, at birth cord blood is richer in HDLs because HDLs are produced in blood circulation and are not dependent upon fetal liver production, conversely to LDLs [47]. In our study the positive association between birthweight and HDL cholesterol was further confirmed by the finding of SGA having decreased HDL cholesterol levels compared to LGA newborns. Further, we also found decreased total cholesterol levels in SGA compared to both AGA and LGA newborns. Inconsistent (mostly null) associations of cord total and HDL cholesterol levels with birthweight and between SGA and AGA, that have been previously reported in literature [[48], [49], [50], [51], [52], [53], [54], [55]], may be due to limited sample size or lack of adequate control regarding confounding. Conversely, positive associations between birthweight and HDL cholesterol, in line with our results, have been described by a randomized control trial in 343 obese pregnant women and an observational study in 1522 newborns upon adjustment for main confounders [16,56]. Further, this last study found lower total and HDL cholesterol levels in 105 SGA compared to 1320 AGA newborns. Beyond the traditional association of increased cardiovascular risk with concentrations of total and HDL cholesterol, HDL may have beneficial or detrimental effects on systemic inflammation, obesity and diabetes and aging depending on composition of HDL particles [57].

In the ENVIR*ON*AGE cohort we explored if being SGA was a significant predictor of the omic markers identified in our study. Methylation levels of *cg05119988* and *cg17901584* located on *SC4MOL* and *DHCR24* genes were higher in SGA compared to LGA, and methylation of *cg17901584* was lower in LGA compared to AGA. Cord blood levels of the two CpGs have been previously associated with birthweight [9], but never before to SGA. In our study the methylation of these CpGs was further positively associated with total level of cord blood cholesterol and positively correlated with plasmalogen PC(36:4)/PC(O-36:5) and PC(34:2). Interestingly, previous research has associated methylation of *cg17901584* with waist circumference, PC(36:5) C and with HDL cholesterol in adults [[58], [59], [60]].

Our analysis had a number of weaknesses. Cord blood includes a mixture of cell-types that may demonstrate similar phenotypes but with distinct methylation and gene expression patterns [61,62]. The protein-set was limited to inflammatory proteins ($n = 16$) and the metabolome analysis was limited to a single analytical platform with many metabolites lacking annotation, which is common in metabolomics analyses [63]. Also, the curation of human pathway databases is incomplete and possibly biased towards specific diseases. Although we were not able to analyze genomic data in the same samples, we did not find in silico evidence of genetic variants influencing the DNA methylation sites. While we hypothesized that birthweight influences biomarkers at different omic levels, the cross-sectional study design does not allow assessment of causality and therefore we cannot exclude the possibility that the biomarkers themselves are responsible for birthweight modifications. In this regard, a recent multi-omic study in adults found through mendelian randomization that methylation of one of the two CpGs common to our methylation- and metabolite-driven integrations of omics (*cg17901584*) seems to be a consequence rather than a cause of

obesity [64]. In sensitivity analyses we found differences by sex but we were not able to detect a clear pattern.

The major strengths of our study are the combination of cord blood samples from four different European birth-cohorts and an integrative analysis of different omic levels accompanied by computational and technical challenges. We acknowledge that our results may be affected by differences in birthweight across individual cohorts (Table 1), however the top signals in the two approaches were unaffected by factors differing between cohorts, indicating that study heterogeneity is not the main driver of our findings. Different methods have been described to integrate data obtained from different omic levels [31,65]. Our approach combines multi-omics correlation with pathway and network correlation analysis and aimed at identifying the intermediate biological mechanisms that link birthweight-related omic signatures from the same individuals.

The translational potential of our results lays in the development of an extensive catalog of birthweight associated signals. In observational studies, birthweight has repeatedly been associated with a variety of later-life diseases [[4], [5], [6], [7]]. In the context of the DOHaD, the signals we found may reflect biomolecular changes exhibiting possible health effects later in life. For example, among the key signatures we identified, lower levels of cord blood PCs had been recently associated with higher risk of pulmonary hypertension in infants [66] and cord blood CCL22 has been associated with IgE sensitization in two year-old children [67]. Further, the finding of low levels of total and HDL cholesterol in SGA compared to AGA and LGA respectively, and that methylation of CpGs identified in our multi-omic study differed in SGA compared to LGA suggest a possible route for tailored intervention in SGA newborns that have higher risk of morbidity and mortality both in the perinatal period and in later life [68]. Although multiple molecular layers are linked via complex mechanisms, multi-omics integration, such as in our study, can further clarify relations between the different omics and enable us to study early life dynamics of molecular signals. Before we can translate this knowledge into general applications, the causality of the identified associations should be studied by longitudinal and in vivo experimental studies.

In conclusion, we further substantiated previously identified biomarkers in cord blood linked to birthweight and identified new omic features. Our data suggested that cholesterol and related metabolic pathways are related to birthweight. Our results provide evidence that integration of different omic layers is a useful tool to generate hypotheses on mechanistic pathways. Further studies are required to discover the role of these biomarkers in later life diseases.

Funding information

This work is supported by the Bijzonder Onderzoeksfonds (BOF) Hasselt University through a PhD fellowship [to RA], the “EXPOsOMICS” grant [grant number 308610-FP7 European Commission to PV], and the “STOP” grant [grant number 774548-European Commission H2020 to PV]. The ENVIR*ON*AGE birth-cohort is supported by the EU Program “Ideas” (ERC-2012-StG-310898) and the FWO (G082317N). Piccolipiù cohort has been funded by the CCM grant 2010 and the Italian Ministry of Health.

Disclaimer

Where authors are identified as personnel of the International Agency for Research on Cancer/World Health Organization, the authors alone are responsible for the views expressed in this article and they do not necessarily represent the decisions, policy or views of the International Agency for Research on Cancer/World Health Organization.

Availability of data and materials

EXPOsOMICS data analyzed during the current study are available via NCBI Gene Expression Omnibus (GEO) repository with the Accession No GSE151042 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE151042, for the methylome) and GSE151373 (https://www.ncbi.nlm.nih.gov/geo/query

/acc.cgi?&acc=GSE151373, for the transcriptome), and via the MetaboLights repository with the Accession No MTBLS1684 (https://www.ebi.ac.uk/metabolights/MTBLS1684, for the metabolome). Code relevant to the analyses is available upon request to the corresponding author.

CRediT authorship contribution statement

Rossella Alfano: Conceptualization, Formal analysis, Writing - original draft, Writing - review & editing. **Marc Chadeau-Hyam:** Conceptualization, Writing - review & editing. **Akram Ghantous:** Conceptualization, Data curation, Writing - review & editing. **Pekka Keski-Rahkonen:** Data curation, Writing - review & editing. **Leda Chatzi:** Resources, Writing - review & editing. **Almudena Espin Perez:** Writing - review & editing. **Zdenko Herceg:** Writing - review & editing. **Manolis Kogevinas:** Writing - review & editing. **Theo M. de Kok:** Data curation, Writing - review & editing. **Tim S. Nawrot:** Resources, Writing - review & editing. **Alexei Novoloaca:** Data curation, Writing - review & editing.**Chirag J. Patel:**Writing - review & editing. **Costanza Pizzi:** Resources, Writing - review & editing. **Nivonirina Robinot:** Data curation, Writing - review & editing. **Franca Rusconi:** Writing - review & editing. **Augustin Scalbert:** Writing - review & editing. **Jordi Sunyer:** Writing - review & editing. **Roel Vermeulen:** Data curation, Writing - review & editing. **Martine Vrijheid:** Resources, Writing - review & editing. **Paolo Vineis:** Conceptualization, Supervision, Writing - review & editing. **Oliver Robinson:** Conceptualization, Formal analysis, Writing - review & editing. **Michelle Plusquin:** Conceptualization, Writing - original draft, Writing - review & editing.

Declaration of competing interest

None.

Acknowledgments

None.

Footnotes

Appendix A Supplementary data to this article can be found online at https://doi.org/10.1016/j.metabol.2020.154292.

Appendix A. Supplementary data

Supplementary material including Supplementary Methods, Supplementary Figures 1-6 and Additional data

Supplementary tables

References

1. Gluckman P.D., Hanson M.A., Gluckman P., Hanson M. The developmental origins of health and disease: an overview. In: Hanson M., Gluckman P., editors. Developmental origins of health and disease. Cambridge University Press; Cambridge: 2006. pp. 1–5.

2. Yokoyama Y., Jelenkovic A., Hur Y.M., Sund R., Fagnani C., Stazi M.A. Genetic and environmental factors affecting birth size variation: a pooled individual-based analysis of secular trends and global geographical differences using 26 twin cohorts. Int J Epidemiol. 2018;47:1195–1206. [PMCID: PMC6124623] [PubMed: 29788280]

3. Bianco-Miotto T., Craig J.M., Gasser Y.P., van Dijk S.J., Ozanne S.E. Epigenetics and DOHaD: from

basics to birth and beyond. J Dev Orig Health Dis. 2017;8:513–519. [PubMed: 28889823]

4. Lawlor D.A., Ronalds G., Clark H., Smith G.D., Leon D.A. Birth weight is inversely associated with incident coronary heart disease and stroke among individuals born in the 1950s: findings from the Aberdeen Children of the 1950s prospective cohort study. Circulation. 2005;112:1414–1418. [PubMed: 16129799]

5. O'Donnell K.J., Meaney M.J. Fetal origins of mental health: the developmental origins of health and disease hypothesis. Am J Psychiatry. 2017;174:319–328. [PubMed: 27838934]

6. McCormack V.A., dos Santos Silva I., Koupil I., Leon D.A., Lithell H.O. Birth characteristics and adult cancer incidence: Swedish cohort of over 11,000 men and women. Int J Cancer. 2005;115:611–617. [PubMed: 15700315]

7. Risnes K.R., Vatten L.J., Baker J.L., Jameson K., Sovio U., Kajantie E. Birthweight and mortality in adulthood: a systematic review and meta-analysis. Int J Epidemiol. 2011;40:647–661. [PubMed: 21324938]

8. Robinson O., Keski-Rahkonen P., Chatzi L., Kogevinas M., Nawrot T., Pizzi C. Cord blood metabolic signatures of birth weight: a population-based study. J Proteome Res. 2018;17:1235–1247. [PubMed: 29401400]

9. Kupers L.K., Monnereau C., Sharp G.C., Yousefi P., Salas L.A., Ghantous A. Meta-analysis of epigenome-wide association studies in neonates reveals widespread differential DNA methylation associated with birthweight. Nat Commun. 2019;10:1893. [PMCID: PMC6478731] [PubMed: 31015461]

10. Adkins R.M., Tylavsky F.A., Krushkal J. Newborn umbilical cord blood DNA methylation and gene expression levels exhibit limited association with birth weight. Chem Biodivers. 2012;9:888–899. [PubMed: 22589090]

11. Makikallio K., Kaukola T., Tuimala J., FK S., Hallman M., Ojaniemi M. Umbilical artery chemokine CCL16 is associated with preterm preeclampsia and fetal growth restriction. Cytokine. 2012;60:377–384. [PubMed: 22857868]

12. Engel S.M., Joubert B.R., Wu M.C., Olshan A.F., Haberg S.E., Ueland P.M. Neonatal genome-wide methylation patterns in relation to birth weight in the Norwegian Mother and Child Cohort. Am J Epidemiol. 2014;179:834–842. [PMCID: PMC3969535] [PubMed: 24561991]

13. Gillberg L., Perfilyev A., Brons C., Thomasen M., Grunnet L.G., Volkov P. Adipose tissue transcriptomics and epigenomics in low birthweight men and controls: role of high-fat overfeeding. Diabetologia. 2016;59:799–812. [PubMed: 26750116]

14. Perng W., Rifas-Shiman S.L., McCulloch S., Chatzi L., Mantzoros C., Hivert M.F. Associations of cord blood metabolites with perinatal characteristics, newborn anthropometry, and cord blood hormones in project viva. Metabolism. 2017;76:11–22. [PMCID: PMC5675164] [PubMed: 28987236]

15. Hellmuth C., Uhl O., Standl M., Demmelmair H., Heinrich J., Koletzko B. Cord blood metabolome is highly associated with birth weight, but less predictive for later weight development. Obes Facts. 2017;10:85–100. [PMCID: PMC5644937] [PubMed: 28376503]

16. Patel N., Hellmuth C., Uhl O., Godfrey K., Briley A., Welsh P. Cord metabolic profiles in obese pregnant women: insights into offspring growth and body composition. J Clin Endocrinol Metab. 2018;103:346–355. [PMCID: PMC5761489] [PubMed: 29140440]

17. Kadakia R., Talbot O., Kuang A., Bain J.R., Muehlbauer M.J., Stevens R.D. Cord blood metabolomics: association with newborn anthropometrics and C-peptide across ancestries. J Clin Endocrinol Metab. 2019;104:4459–4472. [PMCID: PMC6735762] [PubMed: 31498869]

18. Lu Y.P., Reichetzeder C., Prehn C., Yin L.H., Yun C., Zeng S. Cord blood lysophosphatidylcholine 16:1 is positively associated with birth weight. Cell Physiol Biochem. 2018;45:614–624. [PubMed: 29402770]

19. Simpkin A.J., Suderman M., Gaunt T.R., Lyttleton O., McArdle W.L., Ring S.M. Longitudinal analysis of DNA methylation associated with birth weight and gestational age. Hum Mol Genet. 2015;24:3752–3763. [PMCID: PMC4459393] [PubMed: 25869828]

20. Agha G., Hajj H., Rifas-Shiman S.L., Just A.C., Hivert M.F., Burris H.H. Birth weight-for-gestational age is associated with DNA methylation at birth and in childhood. Clin Epigenetics. 2016;8:118. [PMCID: PMC5112715] [PubMed: 27891191]

21. Vineis P., Chadeau-Hyam M., Gmuender H., Gulliver J., Herceg Z., Kleinjans J. The exposome in practice: design of the EXPOsOMICS project. Int J Hyg Environ Health. 2017;220:142–151. [PMCID: PMC6192011] [PubMed: 27576363]

22. Janssen B.G., Madhloum N., Gyselaers W., Bijnens E., Clemente D.B., Cox B. Cohort profile: the ENVIRonmental influence ON early AGEing (ENVIRONAGE): a birth cohort study. Int J Epidemiol. 2017;46:1386. [7m] [PubMed: 28089960]

23. Guxens M., Ballester F., Espada M., Fernandez M.F., Grimalt J.O., Ibarluzea J. Cohort profile: the INMA--INfancia y Medio Ambiente--(environment and childhood) project. Int J Epidemiol. 2012;41:930–940. [PubMed: 21471022]

24. Farchi S., Forastiere F., Vecchi Brumatti L., Alviti S., Arnofi A., Bernardini T. Piccolipiu, a multicenter birth cohort in Italy: protocol of the study. BMC Pediatr. 2014;14:36. [PMCID: PMC3926689] [PubMed: 24506846]

25. Chatzi L., Leventakou V., Vafeiadi M., Koutra K., Roumeliotaki T., Chalkiadaki G. Cohort profile: the mother-child cohort in Crete, Greece (Rhea Study) Int J Epidemiol. 2017;46:1392. [3k] [PubMed: 29040580]

26. Sumner L.W., Amberg A., Barrett D., Beale M.H., Beger R., Daykin C.A. Proposed minimum reporting standards for chemical analysis: chemical analysis working group (CAWG) metabolomics standards initiative (MSI) Metabolomics. 2007;3:211–221. [PMCID: PMC3772505] [PubMed: 24039616]

27. Plusquin M., Chadeau-Hyam M., Ghantous A., Alfano R., Bustamante M., Chatzi L. DNA methylome marks of exposure to particulate matter at three time points in early life. Environ Sci Technol. 2018;52:5427–5437. [PubMed: 29597345]

28. Bakulski K.M., Feinberg J.I., Andrews S.V., Yang J., Brown S., LM S. DNA methylation of cord blood cell types: applications for mixed cell birth studies. Epigenetics. 2016;11:354–362. [PMCID: PMC4889293] [PubMed: 27019159]

29. Lubin J.H., Colt J.S., Camann D., Davis S., Cerhan J.R., Severson R.K. Epidemiologic evaluation of measurement data in the presence of detection limits. Environ Health Perspect. 2004;112:1691–1696. [PMCID: PMC1253661] [PubMed: 15579415]

30. Patel C.J., Manrai A.K. Development of exposome correlation globes to map out environment-wide associations. Pac Symp Biocomput. 2015;20:231–242. [PMCID: PMC4299925] [PubMed: 25592584]

31. Pinu F.R., Beale D.J., Paten A.M., Kouremenos K., Swarup S., Schirra H.J. Systems biology and multi-omics integration: viewpoints from the metabolomics research community. Metabolites. 2019;9 [PMCID: PMC6523452] [PubMed: 31003499]

32. Bukowski R., Smith G.C., Malone F.D., Ball R.H., Nyberg D.A., Comstock C.H. Human sexual size dimorphism in early pregnancy. Am J Epidemiol. 2007;165:1216–1218. [PubMed: 17344203]

33. Li S., Park Y., Duraisingham S., Strobel F.H., Khan N., Soltow Q.A. Predicting network activity from

high throughput metabolomics. PLoS Comput Biol. 2013;9

34. Chong J., Soufan O., Li C., Caraus I., Li S., Bourque G. MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis. Nucleic Acids Res. 2018;46 [W486-W94] [PMCID: PMC6030889] [PubMed: 29762782]

35. Buniello A., MacArthur J.A.L., Cerezo M., Harris L.W., Hayhurst J., Malangone C. The NHGRI-EBI GWAS catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. Nucleic Acids Res. 2019;47 [D1005-D12] [PMCID: PMC6323933] [PubMed: 30445434]

36. Rapp M., Wintergerst M.W.M., Kunz W.G., Vetter V.K., Knott M.M.L., Lisowski D. CCL22 controls immunity by promoting regulatory T cell communication with dendritic cells in lymph nodes. J Exp Med. 2019;216:1170–1181. [PMCID: PMC6504218] [PubMed: 30910796]

37. Anandi V.S., Shaila B. Evaluation of factors associated with elevated newborn 17-hydroxyprogesterone levels. J Pediatr Endocrinol Metab. 2017;30:677–681. [PubMed: 28489558]

38. Lee J.H., Ulrich B., Cho J., Park J., Kim C.H. Progesterone promotes differentiation of human cord blood fetal T cells into T regulatory cells but suppresses their differentiation into Th17 cells. J Immunol. 2011;187:1778–1787. [PMCID: PMC3155957] [PubMed: 21768398]

39. Trinchieri G. Interleukin-12 and the regulation of innate resistance and adaptive immunity. Nat Rev Immunol. 2003;3:133–146. [PubMed: 12563297]

40. Safford M.G., Levenstein M., Tsifrina E., Amin S., Hawkins A.L., Griffin C.A. JAK3: expression and mapping to chromosome 19p12-13.1. Exp Hematol. 1997;25:374–386. [PubMed: 9168059]

41. Seif F., Khoshmirsafa M., Aazami H., Mohsenzadegan M., Sedighi G., Bahar M. The role of JAK-STAT signaling pathway and its regulators in the fate of T helper cells. Cell Commun Signal. 2017;15:23. [PMCID: PMC5480189] [PubMed: 28637459]

42. Hellmuth C., Lindsay K.L., Uhl O., Buss C., Wadhwa P.D., Koletzko B. Maternal metabolomic profile and fetal programming of offspring adiposity: identification of potentially protective lipid metabolites. Mol Nutr Food Res. 2019;63 [PMCID: PMC6455915] [PubMed: 29714050]

43. Hu Y., Huang K., An Q., Du G., Hu G., Xue J. Simultaneous profiling of transcriptome and DNA methylome from a single cell. Genome Biol. 2016;17:88. [PMCID: PMC4858893] [PubMed: 27150361]

44. Cuperlovic-Culf M., Ferguson D., Culf A., Morin P., Jr., Touaibia M. 1H NMR metabolomics analysis of glioblastoma subtypes: correlation between metabolomics and gene expression characteristics. J Biol Chem. 2012;287:20164–20175. [PMCID: PMC3370199] [PubMed: 22528487]

45. Honsho M., Abe Y., Fujiki Y. Dysregulation of plasmalogen homeostasis impairs cholesterol biosynthesis. J Biol Chem. 2015;290:28822–28833. [PMCID: PMC4661398] [PubMed: 26463208]

46. Woollett L.A. Review: transport of maternal cholesterol to the fetal circulation. Placenta. 2011;32(Suppl. 2):S218–S221. [PMCID: PMC4699659] [PubMed: 21300403]

47. Nagasaka H., Chiba H., Kikuta H., Akita H., Takahashi Y., Yanai H. Unique character and metabolism of high density lipoprotein (HDL) in fetus. Atherosclerosis. 2002;161:215–223. [PubMed: 11882335]

48. Brittos T., de Souza W.B., Anschau F., Pellanda L. Lipids and leukocytes in newborn umbilical vein blood, birth weight and maternal body mass index. J Dev Orig Health Dis. 2016;7:672–677. [PubMed: 27572697]

49. Aletayeb S.M.H., Dehdashtian M., Aminzadeh M., Moghaddam A.-R.E., Mortazavi M., Malamiri R.A. Correlation between umbilical cord blood lipid profile and neonatal birth weight. Pediatr Pol. 2013;88:521–525.

50. Ghiasi A., Ziaei S., Faghihzadeh S. The relationship between levels of lipids and lipoprotein B-100 in maternal serum and umbilical cord serum and assessing their effects on newborn infants anthropometric indices. Journal of Midwifery and Reproductive Health. 2014;2:227–232.

51. Kenchappa Y., Behera N. Assay of neonatal cord blood lipid levels and its correlation with neonatal gestational age, gender and birth weight: a single center experience. International Journal of Contemporary Pediatrics. 2016;3:718–724.

52. Nayak C.D., Agarwal V., Nayak D.M. Correlation of cord blood lipid heterogeneity in neonates with their anthropometry at birth. Indian J Clin Biochem. 2013;28:152–157. [PMCID: PMC3613497] [PubMed: 24426201]

53. Katragadda T., Mahabala R.S., Shetty S., Baliga S. Comparison of cord blood lipid profile in preterm small for gestational age and appropriate for gestational age newborns. J Clin Diagn Res. 2017;11:SC05–SC7. [PMCID: PMC5324458] [PubMed: 28274013]

54. Hou R.L., Jin W.Y., Chen X.Y., Jin Y., Wang X.M., Shao J. Cord blood C-peptide, insulin, HbA1c, and lipids levels in small- and large-for-gestational-age newborns. Med Sci Monit. 2014;20:2097–2105. [PMCID: PMC4226317] [PubMed: 25357084]

55. Lobo L.L., Kumar H.U., Mishra T., Sundari T., Singh A., Kumar C.V. Small-for-gestational-age versus appropriate-for-gestational-age: comparison of cord blood lipid profile & insulin levels in term newborns (SAGA-ACT study) Indian J Med Res. 2016;144:194–199. [PMCID: PMC5206869] [PubMed: 27934797]

56. Wang J., Shen S., Price M.J., Lu J., Sumilo D., Kuang Y. Glucose, insulin, and lipids in cord blood of neonates and their association with birthweight: differential metabolic risk of large for gestational age and small for gestational age babies. J Pediatr. 2020;220 64-72.e2. [PubMed: 32093929]

57. Hafiane A., Favari E., Daskalopoulou S.S., Vuilleumier N., Frias M.A. High-density lipoprotein cholesterol efflux capacity and cardiovascular risk in autoimmune and non-autoimmune diseases. Metabolism. 2020;104:154141. [PubMed: 31923386]

58. Petersen A.K., Zeilinger S., Kastenmuller G., Romisch-Margl W., Brugger M., Peters A. Epigenetics meets metabolomics: an epigenome-wide association study with blood serum metabolic traits. Hum Mol Genet. 2014;23:534–545. [PMCID: PMC3869358] [PubMed: 24014485]

59. Demerath E.W., Guan W., Grove M.L., Aslibekyan S., Mendelson M., Zhou Y.H. Epigenome-wide association study (EWAS) of BMI, BMI change and waist circumference in African American adults identifies multiple replicated loci. Hum Mol Genet. 2015;24:4464–4479. [PMCID: PMC4492394] [PubMed: 25935004]

60. Braun K.V.E., Dhana K., de Vries P.S., Voortman T., van Meurs J.B.J., Uitterlinden A.G. Epigenome-wide association study (EWAS) on lipids: the Rotterdam study. Clin Epigenetics. 2017;9:15. [PMCID: PMC5297218] [PubMed: 28194238]

61. Adalsteinsson B.T., Gudnason H., Aspelund T., Harris T.B., Launer L.J., Eiriksdottir G. Heterogeneity in white blood cells has potential to confound DNA methylation measurements. PLoS One. 2012;7 [PMCID: PMC3465258] [PubMed: 23071618]

62. Xu Q., Ni S., Wu F., Liu F., Ye X., Mougin B. Investigation of variation in gene expression profiling of human blood by extended principle component analysis. PLoS One. 2011;6 [PMCID: PMC3203156] [PubMed: 22046403]

63. Sun Y.V., Hu Y.J. Integrative analysis of multi-omics data for discovery and functional studies of complex human diseases. Adv Genet. 2016;93:147–190. [PMCID: PMC5742494] [PubMed: 26915271]

64. Liu J., Carnero-Montoro E., van Dongen J., Lent S., Nedeljkovic I., Ligthart S. An integrative

cross-omics analysis of DNA methylation sites of glucose and insulin homeostasis. Nat Commun. 2019;10:2581. [PMCID: PMC6565679] [PubMed: 31197173]

65. Perakakis N., Yazdani A., Karniadakis G.E., Mantzoros C. Omics, big data and machine learning as tools to propel understanding of biological mechanisms and to discover novel diagnostics and therapeutics. Metabolism. 2018;87:A1–A9. [PMCID: PMC6325641] [PubMed: 30098323]

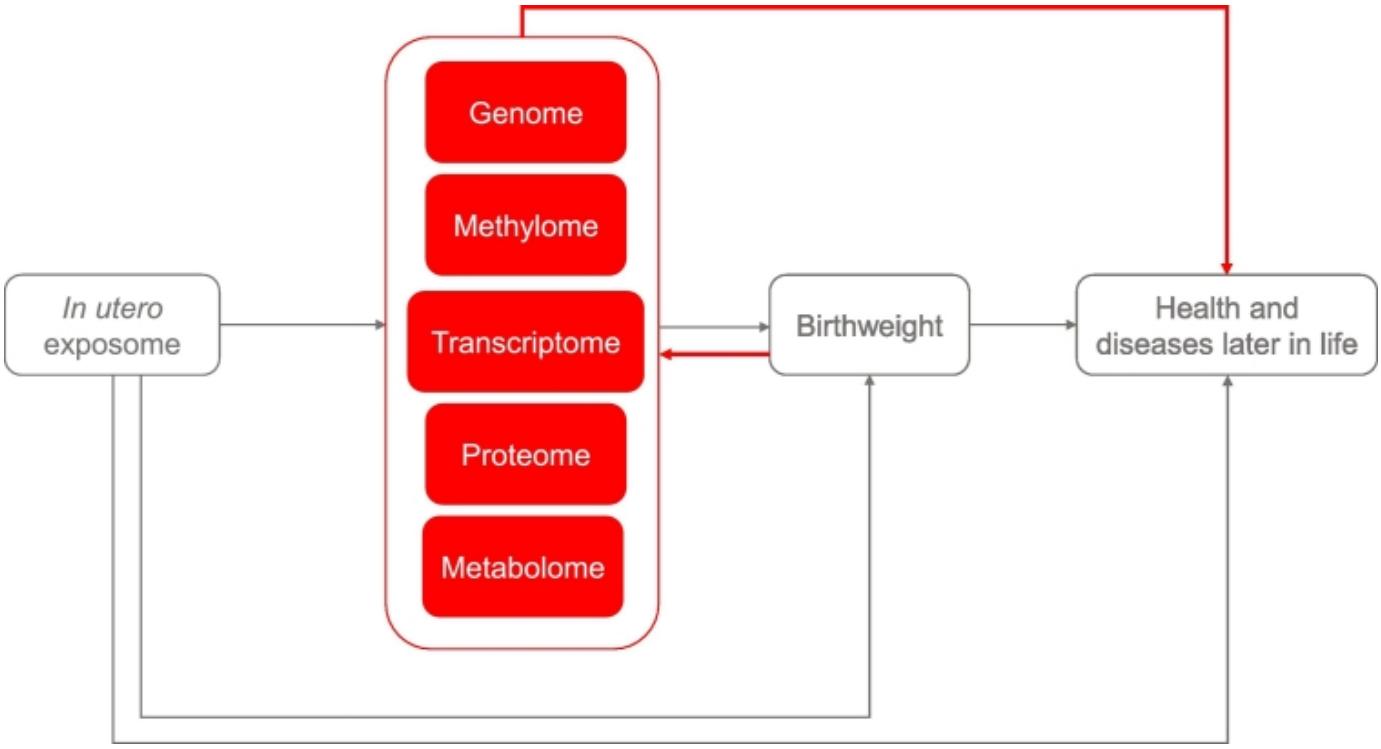
66. La Frano M.R., Fahrmann J.F., Grapov D., Pedersen T.L., Newman J.W., Fiehn O. Umbilical cord blood metabolomics reveal distinct signatures of dyslipidemia prior to bronchopulmonary dysplasia and pulmonary hypertension. Am J Physiol Lung Cell Mol Physiol. 2018;315 [L870-L81] [PMCID: PMC6295510] [PubMed: 30113229]

67. Yeh K.W., Chiu C.Y., Su K.W., Tsai M.H., Hua M.C., Liao S.L. High cord blood CCL22/CXCL10 chemokine ratios precede allergic sensitization in early childhood. Oncotarget. 2017;8:7384–7390. [PMCID: PMC5352329] [PubMed: 27863395]

68. Saenger P., Czernichow P., Hughes I., Reiter E.O. Small for gestational age: short stature and beyond. Endocr Rev. 2007;28:219–251. [PubMed: 17322454]

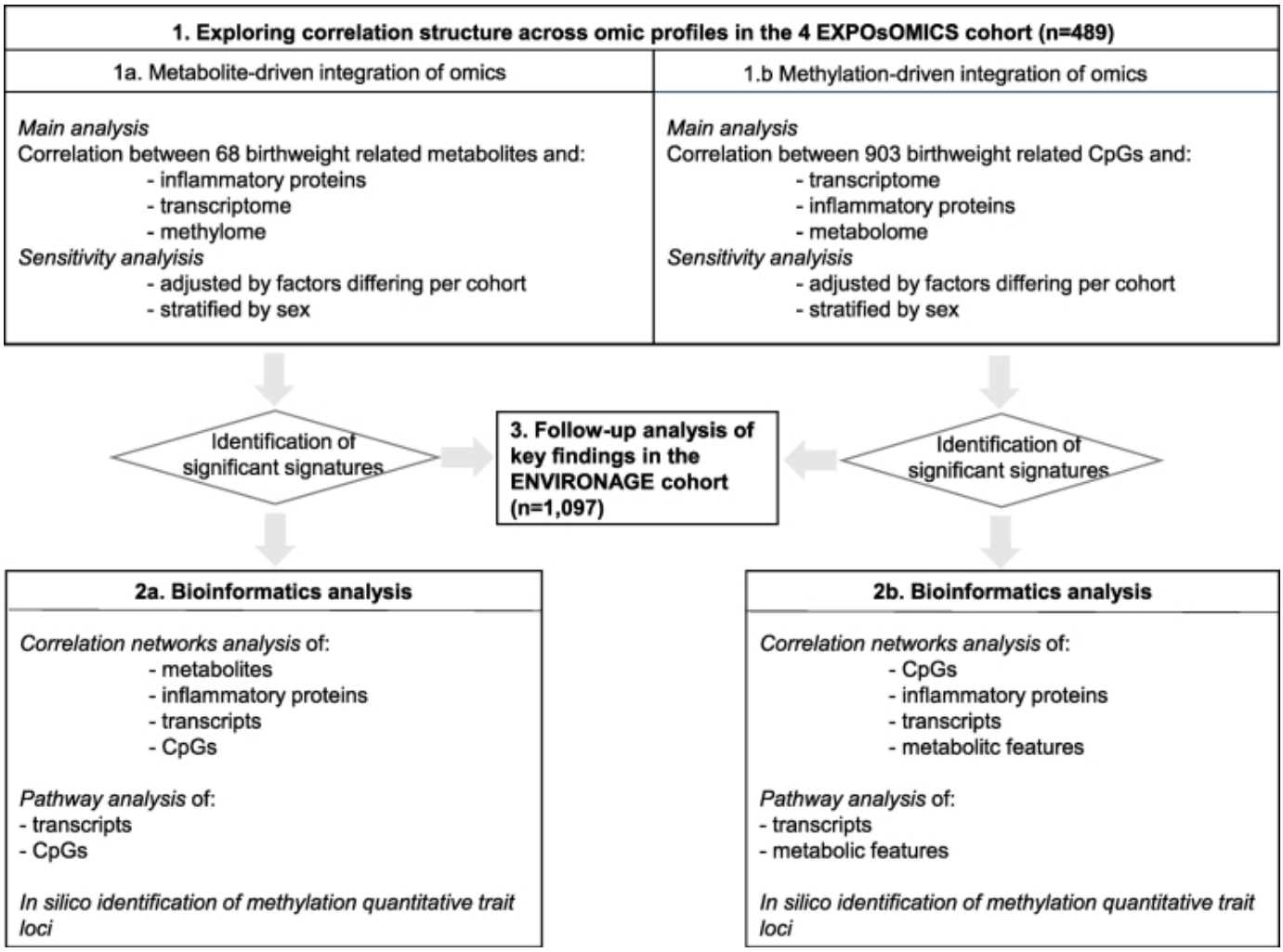
Figures and Tables

Fig. 1



Research hypothesis. The diagram exemplifies the central role of omics in the hypothetical path that drives the present study and according to which birthweight induces molecular modifications that in turn influence the later life health and disease (red arrows); and the alternative pathways where in utero exposures cause the changes in the omic layers leading to variation of birthweight (grey arrows).

Fig. 2



Open in a separate window

Study design. The figure shows the main analysis exploring the structure across omic profiles in the four EXPOsOMICS cohorts and the subsequent pathways analysis, network correlation analysis, in silico identification of methylation quantitative trait loci, and the follow-up analysis exploring key findings in the ENVIR*ON*AGE cohort.

Table 1

Characteristics of the EXPOsOMICS population (by cohort) and the full ENVIRONAGE population.

	EXPOsOMICS population n = 489				P-value[a]	F ENVIR popu n =
	ENVIRONAGE *n* = 195	INMA *n* = 97	Piccolipiù n = 97	RHEA *n* = 100		
Birthweight, g	3389·28 ± 478·29	3305·98 ± 399·33	3217·06 ± 431·42	3258·90 ± 429·69	**8·85e−03**	3413·01
Birthweight,					–	
SGA (<10thPi)	14 (7·2)	–	–	–		76 (6·9)
AGA (≥10thPi& ≤ 90thPi)	155 (79·5)	–	–	–		880 (80
LGA (>90th Pi)	25 (12·8)	–	–	–		141 (12
Gestational age, weeks	39·14 ± 1·53	39·71 ± 1·41	39·57 ± 1·58	38·43 ± 1·32	**2·66e−09**	39·22 ±
Girls	95 (49·0)	50 (51·5)	43 (44·3)	47 (47·0)	0·77	530 (48
Maternal age, years	29·41 ± 4·43	31·48 ± 4·14	33·28 ± 4·46	30·03 ± 4·99	**8·25e−11**	29·41 ±
Maternal BMI, Kg/m^2	23·94 ± 4·06	23·45 ± 3·83	22·63 ± 3·87	25·09 ± 5·37	**7·87e−04**	24·52 ±
Maternal weight, Kg	66·09 ± 11·88	62·52 ± 11·20	60·95 ± 11·16	66·76 ± 15·64	**9·28e−04**	67·82 ±
Maternal height, cm	166·14 ± 6·81	163·15 ± 6·60	164·05 ± 5·70	162·93 ± 5·65	**3·22e−05**	166·189
Maternal smoking	25 (12·9)	23 (24·0)	20 (20·6)	20 (20·2)	0·09	134 (12
Maternal education					**1·57e−03**	
Primary school	27 (14·6)	17 (17·5)	8 (8·2)	8 (8·1)		139 (12
Secondary school	63 (34·1)	46 (47·4)	40 (41·2)	57 (57·6)		393 (35
University of higher	95 (51·4)	34 (35·1)	49 (50·5)	34 (34·3)		565 (51
Multiparity	87 (44·8)	43 (44·8)	51 (52·6)	70 (71·4)	**1·16e−04**	511 (46
Paternal age, years	31·75 ± 5·89	33·46 ± 4·35	36·47 ± 5·57	34·24 ± 5·04	**1·94e−10**	31·95 ±
Paternal BMI, Kg/m^2	25·78 ± 3·46	27·20 ± 3·90	24·97 ± 3·03	25·99 ± 4·49	**1·06e−03**	25·96 ±
Paternal weight, Kg	83·43 ± 15·94	81·11 ± 13·36	78·43 ± 10·65	84·97 ± 14·47	**5·97e−03**	83·65 ±
Paternal height, cm	179·07 ± 7·54	177·08 ± 6·80	177·20 ± 6·30	176·38 ± 7·21	**0·01**	179·49 ±

Open in a separate window

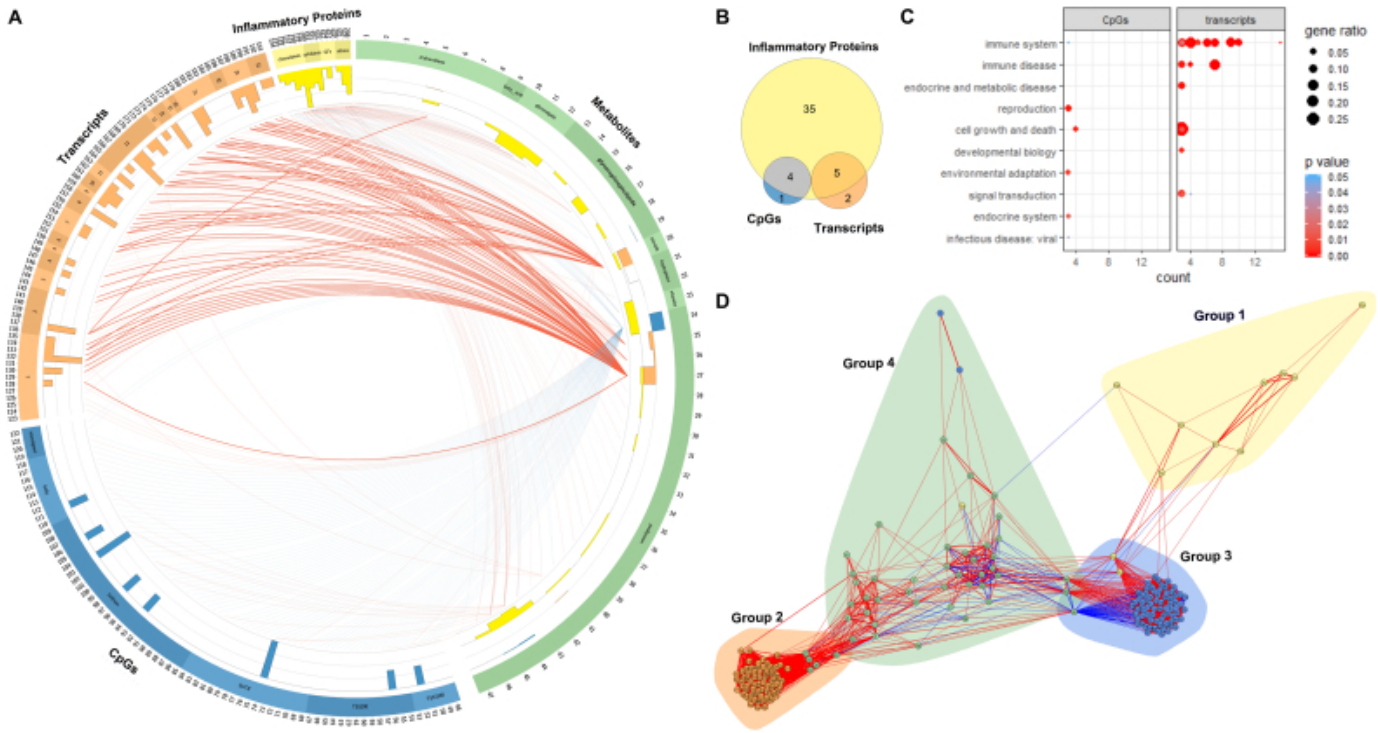
Counts (percentages) and means ± standard deviations are reported for categorical and continuous variables, respectively.

AGA = adequate for gestational age; LGA = large for gestational age; Pi = percentile calculated for Flanders from the Study Centre for Perinatal Epidemiology; SGA = small for gestational age.

[a]P-value for associations between the four EXPOsOMICs birth-cohort. P-values <0·05 are marked in bold.

[b]Between the pooled EXPOsOMICs population and the full ENVIR*ON*AGE population are detected with analysis of variance test (for continuous variables) and chi square test (for categorical variables). P-values <0·05 are marked in bold.

Fig. 3



Significant correlations in the metabolite-driven integration of omics. A The circular plot displays the results of metabolite-driven integration of omics. Tracks from outside to inside are: ideogram, histogram plot and significant links between omic signals. The ideogram shows the omic features significantly correlated grouped in metabolites (green), CpGs (blue), transcripts (orange) and inflammatory proteins (yellow). On the ideogram, features are identified by numbers as reported in the Supplementary Table 27. Alternating bands distinguish classes of metabolites and proteins, genomic regions of genes associated to CpGs and chromosomes on which are located the genes associated to transcripts. The histogram shows for each omic feature (on the x-axis) the scaled percentage of significant correlations per each omic set as identified by colors (on the y-axis in increasing order from outside to inside). In the center of the circular plot each significant correlation coefficient is visualized through a link connecting the two correlated omics. Links are colored according to the sign of the correlation coefficient, where red and blue links mean respectively positive and negative correlations. Thickness of the links grows according to increasing absolute value of correlation coefficients. B The venn diagram shows the count of metabolites significant in the metabolite-inflammatory protein, metabolite-transcriptome and metabolite-methylome analyses represented in yellow, orange and blue circles respectively, and their intersections. C The dot plot shows significant pathways grouped by function, from overrepresentation analysis (ORA) of the transcripts and the CpGs identified in the metabolite-driven integration of omics. Dots size varies according the gene ratio and colors according the p-values. D The network chart shows results from correlation network analysis. The nodes represent the omic features. Nodes are colored according the omic layer they belong to (metabolites in green, CpGs in blue, transcripts in orange and inflammatory proteins in yellow) and are identified by numbers as reported in the Supplementary Table 27. Only nodes with degree>2 and edges with Bonferroni corrected p-values <0·05 are displayed. Hedges are colored according to the sign of the correlation coefficient, where red and blue links mean respectively positive and negative correlations. Communities (groups 1–4), detected using the Louvain algorithm, are marked by circles.

TSS = transcription start site; UTR = untranslated region; GFs = growth factors.

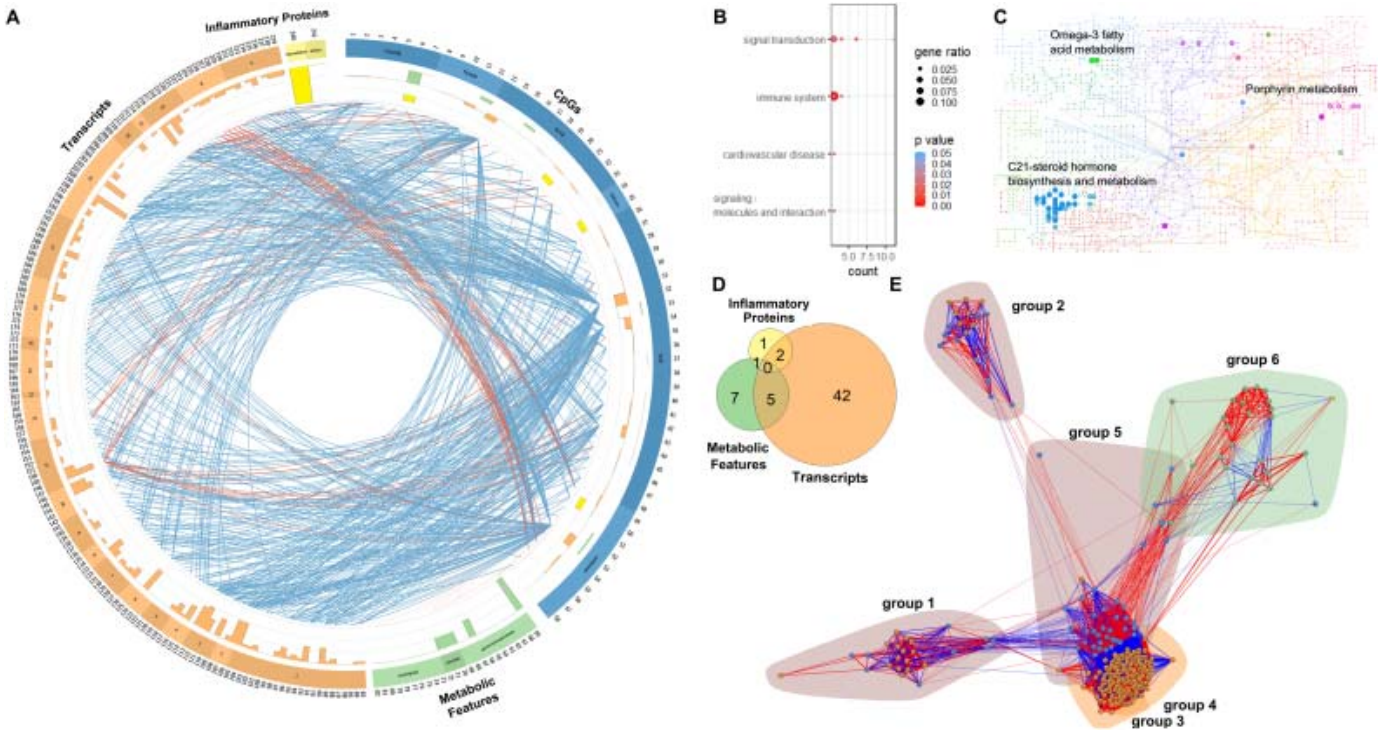
Table 2

Overview of the metabolite-driven integration of omics.

	Metabolite-inflammatory protein pairs	Metabolite-transcript pairs	Metabolite-CpG site pairs
Samples	489	164	460
Correlation pairs	1088	1,983,152	28,394,896
P-value threshold	4·60e−05	2·52e−08	1·76e−09
	Significant correlations		
Correlation pairs	133 (12·2)	129 (0·006)	85 (0·0002)
Omic features	44 metabolites (64·71) 15 proteins (93·75)	7 metabolites (10·29) 71 transcripts (0·24)	5 metabolites (7·35) 75 CpGs (0·02)
r absolute values range	0·18–0·44	0·42–0·54	0·28–0·39
Negative r	43 (32·33)	0 (0)	70 (82·35)

Counts (percentages) are reported. r = correlation coefficient.

Fig. 4



Significant correlations in the methylation-driven integration of omics. A The circular plot depicts the results of the methylation-driven integration of omics. Tracks from outside to inside are: ideogram, histogram plot and significant links between omic signals. The ideogram shows the omic features significantly correlated grouped in CpGs (blue), metabolic features (green), transcripts (orange) and inflammatory proteins (yellow). On the ideogram, features are identified with numbers as reported in the Supplementary Table 28. Alternating bands distinguish classes of metabolites and proteins, genomic regions of genes associated to CpGs and chromosomes on which are located the genes associated to transcripts. The histogram shows for each omic feature (on the x-axis) the scaled percentage of significant correlations per each omic set as identified by colors (on the y-axis in increasing order from outside to inside). In the center of the circular plot each significant correlation coefficient is visualized through a link connecting the two correlated omics. Links are colored according to the sign of the correlation coefficient, where red and blue links mean respectively positive and negative correlations. Thickness of the links grows according to increasing absolute value of correlation coefficients. B The dot plot shows significant pathways, grouped by function, from overrepresentation analysis (ORA) of the transcripts identified in the methylation-driven integration of omics. Dots size varies according the gene ratio and colors according the p-values. C Metabolic network visualization of significantly enriched pathways based on the manually curated KEGG global metabolic network [34]. The metabolites of significantly enriched pathways are represented as nodes on the network. Empty nodes represent compounds identified from the feature list by *mummichog* but not significant, while solid nodes represent significantly enriched features. Note not all metabolites from the KEGG global network are displayed. D The venn diagram shows the count of significant CpGs in the CpG-inflammatory protein, CpG -transcriptome and CpG-metabolic feature analyses represented in yellow, orange and green circles respectively, and their intersections. E The network chart shows results from correlation network analysis. The nodes represent the omic features. Nodes are colored according the omic layer they belong to (metabolites in green, CpGs in blue, transcripts in orange and inflammatory proteins in yellow) and are identified by numbers as reported in the Supplementary Table 28. Only nodes with degree>2 and edges with Bonferroni corrected p-values <0·05 are displayed. Hedges are colored according to the sign of the correlation coefficient, where red and blue links mean respectively positive and negative correlations. Communities (groups 1–6), detected using the Louvain algorithm, are marked by circles.

TSS = transcription start site; UTR = untranslated region.

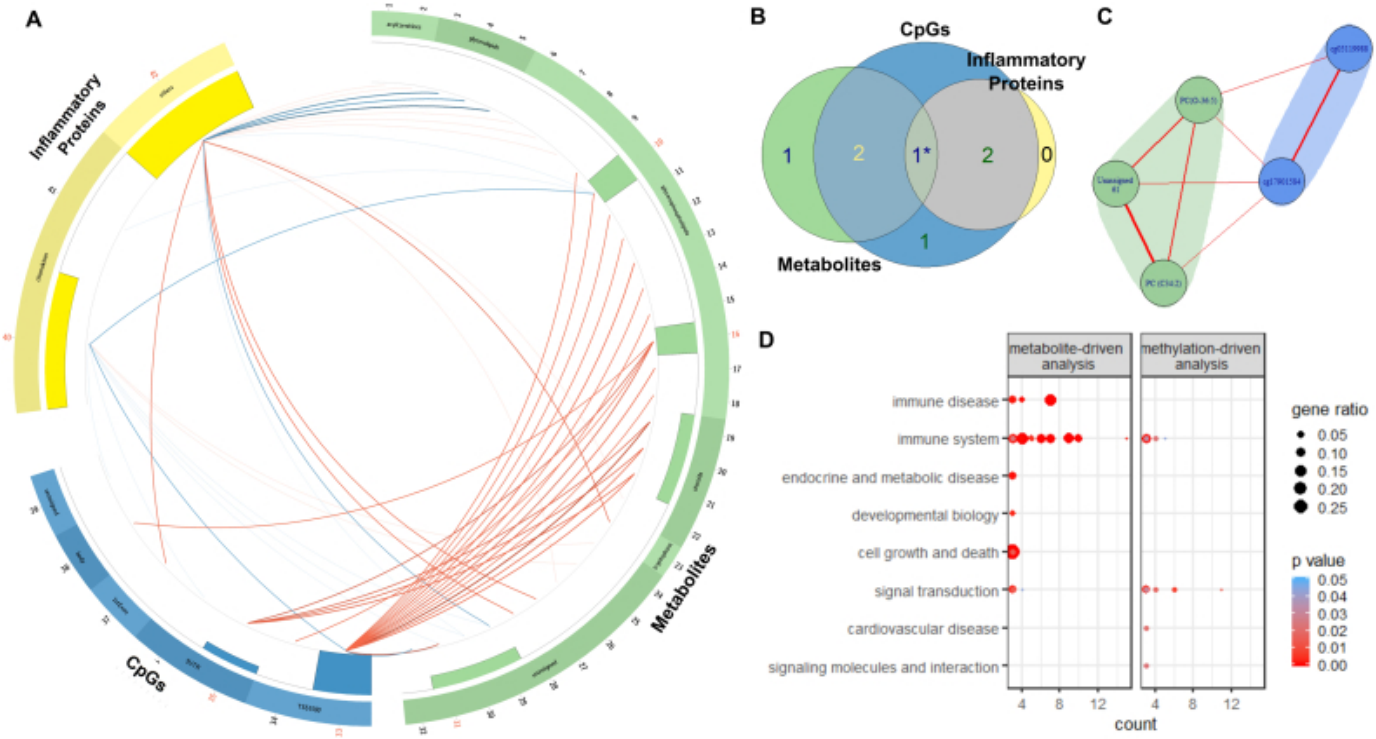
Table 3

Overview of the methylation-driven integration of omics.

	Methylation-inflammatory protein pairs	Methylation-transcript pairs	Methylation-metabolic feature pairs
Samples	450	162	460
Correlation pairs	14,448	26,335,092	4,254,936
P-value threshold	3·46e−06	1·89e−09	1·17−08
		Significant correlations	
Correlation pairs	4 (0·03)	439 (0·002)	39 (0·0009)
Omic features	4 CpGs (0·44) 2 proteins (12·5)	49 CpGs (5·43) 157 transcripts (0·54)	13 CpGs (1·44) 24 metabolic features (0·51)
r absolute values range	0·22–0·24	0·45–0·57	0·26–0·39
Negative r	3 (75)	403 (91·80)	8 (20·51)

Counts (percentages) are reported. r = correlation coefficient.

Fig. 5



Significant correlations common to the metabolite- and the methylation-driven integrations of omics. The figure represents the 48 significant correlations identified by the seven omics commonly significant in the metabolite- and the methylation-driven integrations of omics. A In the circular plot tracks from outside to inside are: ideogram, histogram plot and significant links between omic signals. The ideogram shows the omic features significantly correlated grouped in metabolites (green), CpGs (blue) and inflammatory proteins (yellow). On the ideogram features are identified by numbers as reported in the Supplementary Table 29. Alternating bands distinguish classes of metabolites and proteins, and genomic regions of genes associated to CpGs. The seven common omic features are highlighted in red. The histogram plot shows for each omics feature (on the x-axis) the scaled percentage of significant correlations (on the y-axis in increasing order from outside to inside). In the center of the circular plot each significant correlation coefficient is visualized through a link connecting the two correlated omics. Links are colored according to the sign of the correlation coefficient, where red and blue links mean respectively positive and negative correlations. Thickness of the links grows according to increasing absolute value of correlation coefficients. B The venn diagram shows among the seven common omic features how many are significantly correlated with inflammatory proteins, transcripts and metabolites represented in yellow, orange and green circles respectively, and their intersections. C The network chart shows results from correlation network analysis. The nodes represent the omic features. Nodes are colored according the omic layer they belong to (metabolites in green, and CpGs in blue) and are identified by omic names. Only nodes with degree>2 and edges with Bonferroni corrected p-values <0·05 are displayed. Hedges are colored according to the sign of the correlation coefficient, where red and blue links mean respectively positive and negative correlations. Communities (groups 1–2), detected using the Louvain algorithm, are marked by circles. D The dot plot shows significant pathways grouped by function, from overrepresentation analysis (ORA) of the transcripts identified in the metabolite- and methylation-driven integration of omics. Dots size varies according the gene ratio and colors according the p-values.

TSS = transcription start site; UTR = untranslated region * intersection only between proteins and metabolites.

Table 4

Results from the cholesterol analyses.

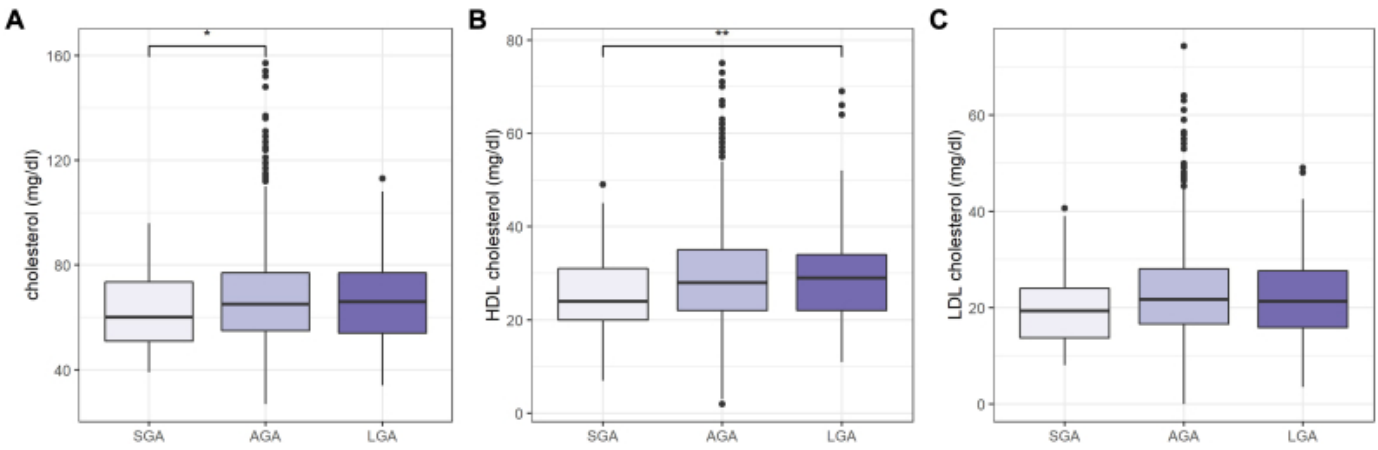
	Total cholesterol		HDL cholesterol		LDL cholesterol	
	Change (mg/dl)	**P-value**	**Change (mg/dl)**	**P-value**	**Change (mg/dl)**	**P-value**
Metabolites[a]	*n* = 182		*n* = 182		*n* = 182	
PC(34:2)	17·08	**7·82e−04**	2·66	0·16	−1·07	0·48
Plasmalogen PC(36:4)/PC(O-36:5)	26·85	**3·05e−07**	5·28	**0·01**	−1·17	0·48
U61	17·70	**3·32e−04**	1·14	0·54	1·72	0·24
CpGs[a]	*n* = 178–179		n = 178–179		*n* = 178–179	
cg05119988	6·02	**7·45e−03**	0·55	0·51	0·17	0·80
cg17901584	6·87	**1·56e−03**	0·86	0·29	0·58	0·38
Proteins[a, b]	n = 176		*n* = 176		*n* = 176	
CCL22	−5·39	0·25	−2·88	0·09	0·90	0·51
Periostin	8·19	0·09	−2·70	0·13	0·89	0·52

P-values <0.05 are marked in bold. CCL22 = macrophage-derived chemokine; change = change in cholesterol levels (in mg/dl) for one interquartile range increment of metabolites, CpGs and proteins; HDL = high-density lipoprotein; LDL = low-density lipoprotein; n = numbers of observation; PC = phosphatidylcholine; U61 = unassigned metabolite 61.

[a]All the analyses are adjusted for gestational age, newborn sex, maternal age, maternal height, maternal BMI, smoking during pregnancy, parity, maternal education, and total cholesterol levels (for HDL and LDL analyses).

[b]Analyses of proteins were additionally adjusted on plate and analyses of CpGs were additionally adjusted for chip, position and cell types composition.

Fig. 6



Cholesterol levels in ENVIR*ON*AGE cohort. A Total, B HDL and C LDL cholesterol levels in cord blood (on the y-axis) in SGA, AGA and LGA newborns (on the x-axis) of the ENVIR*ON*AGE cohort ($n = 1097$) are graphically represented by boxplots.

AGA = adequate for gestational age; LGA = large for gestational age; SGA = small for gestational age.

*p-value between SGA and AGA and ** between SGA and LGA from linear multivariate analysis adjusted for gestational age, newborn sex, maternal age, maternal height, maternal BMI, maternal smoking during pregnancy, parity, maternal education, and cord blood total cholesterol levels (for HDL and LDL analyses) < 0·05.