Multi-omics Analysis of Umbilical Cord Hematopoietic Stem Cells from a Multi-ethnic Cohort of Hawaii Reveals the Transgenerational Effect of Maternal Pre-Pregnancy Obesity

Yuheng Du[1], Paula A. Benny[2], Ryan J. Schlueter[2], Alexandra Gurary[2], Annette Lum-Jones[3], Cameron B Lassiter[3], Fadhl M. AlAkwaa[4], Maarit Tiirikainen[3], Dena Towner[2], W. Steven Ward[2], Lana X Garmire[1*]

1. Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI

2. Department of Obstetrics and Gynecology, University of Hawaii, Honolulu, HI

3. University of Hawaii Cancer Center, Population Sciences of the Pacific Program-Epidemiology, Honolulu, HI

4. Department of Neurology, University of Michigan, Ann Arbor, MI

*. Corresponding author email: lgarmire@med.umich.edu

Keywords: obesity, Native Hawaiian, Hematopoietic stem cells, Multi-omics, cord blood, methylation, pregnancy

Abbreviation:

BH: Benjamini-Hochberg

BMI: Body mass index

C: Acylcarnitines

DE: Differential expression

DIABLO: Data Integration Analysis for Biomarker discovery using Latent cOmponents

DM: Differential methylation

DMR: Differentially methylated regions

DOHaD: Developmental Origins of Health and Disease

EWAS: Epigenome-wide association studies

FC: Fold Change

FDR: False positive results

KEGG: Kyoto Encyclopedia of Genes and Genomes

LASSO: Least absolute shrinkage and selection operator

LDA: Linear Discriminant Analysis

LOG: Logistic regression

PAM: Partition Around Medoids

PANDA: Preferential Attachment-based common Neighbor Distribution derived Associations

PC aa: Diacyl phosphatidylcholines

PC ae: Acyl-alkylphosphatidylcholines

PCC: Pearson correlation coefficients

PPI: Protein-Protein Interaction

RBF: Radial Basis Function

RF: Random Forest

RPART: Recursive Partitioning and Regression Trees

SOV: Source of variance

SVD: Singular value decomposition

SVM: Supportive Vector Machine

TSS: Transcription start site

uHSCs: Umbilical cord blood hematopoietic stem cells

UMAP: Uniform Manifold Approximation and Projection

VLCAD: Very long-chain acyl-CoA dehydrogenase

WGCNA: Weighted Gene Co-expression Network Analysis

Abstract

Background: Maternal obesity is a health concern that may predispose newborns to a high risk of medical problems later in life. To understand the transgenerational effect of maternal obesity, we conducted a multi-omics study, using DNA methylation and gene expression in the CD34+/CD38-/Lin- umbilical cord blood hematopoietic stem cells (uHSCs) and metabolomics of the cord blood, all from a multi-ethnic cohort (n=72) from Kapiolani Medical Center for Women and Children in Honolulu, Hawaii (collected between 2016 and 2018).

Results: Differential methylation (DM) analysis unveiled a global hypermethylation pattern in the maternal pre-pregnancy obese group (BH adjusted $p<0.05$), after adjusting for major clinical confounders. Comprehensive functional analysis showed hypermethylation in promoters of genes involved in cell cycle, protein synthesis, immune signaling, and lipid metabolism. Utilizing Shannon entropy on uHSCs methylation, we discerned notably higher quiescence of uHSCs impacted by maternal obesity. Additionally, the integration of multi-omics data-including methylation, gene expression, and metabolomics-provided further evidence of dysfunctions in adipogenesis, erythropoietin production, cell differentiation, and DNA repair, aligning with the findings at the epigenetic level.

Conclusions: This study reveals the significant correlation between pre-pregnancy maternal obesity and multi-omics level molecular changes in the uHSCs of offspring, particularly in DNA methylation.

Introduction

Maternal obesity has emerged as a primary health concern during pregnancy, with its prevalence alarmingly increasing. According to a study by the Centers for Disease Control and Prevention, the percentage of women experiencing pre-pregnancy obesity in the United States escalated from 26% to 29% between 2016 and 2019 [1]. Born to mothers with obesity, higher birth weight is associated with a higher incidence of childhood cancers such as leukemia and neuroblastoma [2,3], as well as greater risks of prostate and testicular cancers in men [4–6] and breast cancer in women [7]. Moreover, maternal obesity may have a transgenerational effect and set the stage for increased chronic disease susceptibility later in the adulthood of offspring [8,9]. The hypothesis of the utero origin of diseases proposes that numerous chronic diseases have their origins in the fetal stage, the earliest phase of human development [10,11]. Some researchers have speculated higher stem cell burdens in newborn babies born from obese mothers [12]. In particular, a study showed increases in cord blood CD34+CD38- stem cell and CD34+ progenitor cell concentrations with maternal obesity [13], suggesting that the higher proportions of stem cells in cord blood may make the babies more susceptible to obesity and cancer risks. However, so far little work provides the direct molecular links as to how maternal obesity affects the cellular function and increases the disease risk in offspring.

To seek answers in this area, we conducted an epigenome centered multi-omics study to directly pinpoint the effect of maternal obesity in umbilical cord blood hematopoietic stem cells (uHSCs). Epigenetics is chosen as the center of multi-omics integration, as it is both inheritable and susceptible to modification by diseases. Thus it may serve as a plausible mediator in the transmission of the effects of maternal obesity to offspring. We incorporate gene expression for cord blood stem cells and metabolomics data from the cord blood serum as the downstream

readout of epigenetics changes. By elucidating these molecular connections, we provide a systematic understanding of how maternal obesity during pregnancy can influence the multiple types of molecular profiles of newborns. Such knowledge may ultimately help develop early therapeutic interventions at the molecular level to mitigate these transgenerational health risks due to maternal obesity.

Methods

Overview of the maternal pre-pregnancy cohort with baby cord blood

In this study, baby cord blood samples from 72 pregnant women (34 obese; 38 non-obese) who delivered at Kapiolani Medical Center for Women and Children in Honolulu, Hawaii (2015-2018) were collected. The study was approved by the Western IRB (WIRB Protocol #20151223). Patients meeting the inclusion criteria were identified from pre-admission medical records with pre-pregnancy BMI $\geq$ 30.0 (maternal obesity) or 18.5-25.0 (normal weight). All participants involved in this study provided written informed consent before the collection of cord blood samples. Pregnant women undergoing elected C-sections at $\geq$37 weeks gestation were included, in order to minimize confounding events during the labor. Patient exclusion criteria included pregnant women with preterm rupture of membranes, labor, multiple gestations, pregestational diabetes, hypertensive disorders, cigarette smokers, infection of human immunodeficiency virus or hepatitis B virus, and chronic drug use. Demographic and phenotypic information was recorded, including maternal and paternal age, ethnicity, gestational weight gain, gestational age, parity, and gravidity. For newborns, Apgar scores were documented

at 1 minute and 5 minutes post-delivery. The Apgar score serves as a comprehensive assessment of a newborn's health, with a normal range considered to be between 7 to 10 [14].

Sample preparation and methylation profiling

The baby cord blood sample was collected in the operating room under sterile conditions at the time of the C-section (Pall Medical Cord Blood Collection Kit containing 25ml citrate phosphate dextrose). The umbilical cord was first cleansed with chlorhexidine swabs before cord blood collection. The total volume of collected blood was measured and recorded before aliquoting to conical tubes for centrifugation. The tubes were centrifuged at 200g for 10 min, and plasma was collected. The plasma volume was replaced with 2% FBS/PBS. Negative selection reagents were added to the blood and incubated for 20 min (Miltenyi Biotec, Auburn, CA). The cord blood was diluted with an equal volume of 2% FBS/PBS. A 20ml aliquot of the diluted blood was layered over a density gradient (15ml Lymphoprep) and centrifuged at 1200g for 20 min. The top layer containing an enriched population of stem cells was collected, centrifuged at 300g for 8 min, and then washed in 2% FBS/PBS. Red blood cells were lysed using ammonium chloride (9:1) with incubation on ice for 10 min, washed twice, and then resuspended in 100µl of 2% FBS/PBS. Cells were stained with Lineage FITC and CD34 APC for 45 min on ice, washed twice, and then sorted using the BD FacsAria III. Hematopoietic stem cells (CD34+/CD38-/Lin-) were collected and stored at -80°C until DNA/RNA extraction.

DNA and RNA were extracted simultaneously using the AllPrep DNA/RNA extraction kit (Qiagen). DNA purity and concentration were quantified in Nanodrop 2000 and Picogreen assay. Bisulfite conversion of 500 ng DNA was performed using the EZ DNA Methylation kit (Zymo), followed by sample processing for Infinium HumanMethylation450 bead chips (Illumina)

according to the manufacturer's instructions. Bead chips were analyzed at the Genomics Shared Resource at the University of Hawaii Cancer Center.

Bulk RNA sequencing

A total of 50 RNA samples were prepared for bulk RNA Sequencing. RNA concentration and RIN score were assayed using Nanodrop 2000 and Agilent Bioanalyzer. A total of 200 ng of high-quality RNA (RIN≥7) was subjected to library construction (polyA) and sequenced on HS4000 (2x100) at the Yale Center for Genome Analysis, Connecticut to obtain at least 25M paired reads per sample.

Methylation data pre-processing

R version 3.6.3 was used for all downstream analyses. Raw intensity data (.idat) were extracted using the 'ChAMP' package (version 2.16.2) in R [15–18]. Quality control and processing were performed following the ChAMP pipeline. Background controls were subtracted from the data, and raw data that did not pass detection P-value of 0.05 were removed. For each CpG site, the methylation score was initially calculated as the beta value, a fluorescence intensity ratio between 0 and 1. CpG sites whose probes had known underlying SNPs and association with XY chromosomes were removed from analysis due to potential confounding. After BMIQ normalization [19], the batch effect due to non-biological technical variation caused by experiment handling was removed using the ComBat function in the ChAMP package, confirmed by the singular value decomposition (SVD) plot. The M-values for differential analysis were

transformed from beta-values using lumi (ver 3.1.4) in R [20–23]. A total of 410,765 CpG sites remained for downstream analysis after probe filtering, normalization, and batch removal.

Source of variation analysis and confounding adjustment

To eliminate potential confounding factors of pre-pregnant maternal obesity among the 13 clinical factors, we conducted a source of variation analysis to identify the clinical factors that significantly contribute to the methylation level variation, as done before [24,25]. The variables with F statistics greater than 1 (the error value) were determined as confounders and subjected to confounding adjustment. These factors include the baby's sex, net weight gain during the pregnancy, maternal age, maternal ethnicity, paternal ethnicity, gravidity, and gestational age. To adjust for confounding, a linear regression model is built using the 'limma' package to fit methylation M values of each CpG site, using the confounding factors above. The remaining residuals on the M values were considered to be confounding-adjusted for the subsequent bioinformatics analysis of DNA methylation.

Bioinformatics analysis of differential methylation (DM)

A moderated t-test from the 'limma' R package (version 3.42.2) [26] was used for detecting DM CpG sites between healthy controls and cases with M values. The p-values were adjusted for multiple hypotheses testing using Benjamini-Hochberg FDR. CpG sites with FDR <0.05 were considered statistically significant. To minimize the effect of the gestational age, CpG sites located within the gestational-age-related differentially methylated regions (DMR) were removed. A total of 130 DMRs related to gestational age were identified using linear regression analysis performed with bumphunter [27] across eight public datasets including a total of 248

patients.: GSE31781 [28], GSE36829 [29], GSE59274 [28,30], GSE44667 [31], GSE74738 [32], GSE49343 [33], GSE69502 [34], and GSE98224 [35,36]. The complete list of DMRs was included in **Supplementary Table 1**. Hypermethylation and hypomethylation states were defined by the values of log2 Fold Change (log_2FC) of M values in cases compared to controls: hypermethylation if bigger than 0, and hypomethylation if less than 0. Corresponding genes and feature locations of these differential CpG sites were annotated using IlluminaHumanMethylation450kanno.ilmn12.hg19 (ver 0.6.0) [37] .

KEGG pathway enrichment analysis

'gometh' function from R package "missMethyl" (version 1.26.1) [38–41] was used for KEGG pathway enrichment [42–44] with DNA methylation data. DM sites were used for pathway enrichment within five supergroups from KEGG pathways: Metabolism, Genetic Information Processing, Environmental Information Processing, Cellular Processes, and Organismal Systems. Pathways with adjusted p-values less than 0.05 were considered significant. Pathway scores for protein pathways (KEGG: Transcription, Translation, Folding, sorting and degradation) and immune pathways (KEGG: Immune system) were calculated with averaged beta values from the promoter region CpG sites.

Weighted co-expression network analysis

Firstly, we adjusted all beta values with clinical confounders, then summarized the DM CpG sites at the gene level by averaging the beta values in the promoter regions (those in the TSS200 and TSS1500 promoter regions). Next, we transformed adjusted beta values to adjusted M values for the downstream adjacency matrix construction. We used adjusted M values for the weighted

gene co-expression network analysis (WGCNA) with R package 'WGCNA' (version 1.70-3) [45,46]. The soft threshold for the weighted adjacency matrix with an adjusted R^2>0.8 was 7. The topological overlap matrix was constructed for hierarchical clustering. Modules were identified by the dynamic tree-cut algorithm. The networks were exported to Cytoscape with an edge weight greater than 0.03 in each module. The genes with the highest betweenness and degree in the WGCNA network were identified as the hub genes for different modules.

Protein-protein interaction network analysis

For the protein-protein interaction (PPI) network analysis, DM genes are used as the inputs and were mapped on the STRING database (version 10) [47]. Significantly functionally associated protein pairs were identified using PANDA (Preferential Attachment based common Neighbor Distribution derived Associations) (version 0.9.9) [48]. KEGG pathways associated with these protein pairs (in terms of genes) were found using PANDA. The bipartite network graph was visualized using Cytoscape (version 3.8.1) [49].

Stemness score computation

The stemness score was based on Shannon entropy and scaled plasticity, as proposed previously [50]. Shannon entropy has been widely applied in developmental biology, particularly in stem cell research [51–53]. The formulas are shown below:

$$Entropy = \sum_{i=1}^{N} \frac{-\frac{CpG_i}{\sum_{i=1}^{N} CpG_i} log\left(\frac{CpG_i}{\sum_{i=1}^{N} CpG_i}\right)}{log(N)}$$

$$StemnessScore = \frac{Entropy - min(Entropy)}{max(Entropy) - min(Entropy)}$$

N is the total number of CpG sites. CpG is represented by the beta value on each CpG probe. The stemness score was calculated for all samples using all remaining 410,765 CpG sites after the preprocessing. A Wilcoxon rank test was performed between the stemness scores of the healthy and maternally obese groups.

Bulk RNA-seq data processing

The Illumina universal adapter regions of raw RNA-seq data were first trimmed using BBMap (version 38.91) [54]. All raw sequences passed the quality control using fastqc (version 0.11.8) [55]. The trimmed .fastq files were aligned by STAR (version 2.7.0f) [56] to the human Ensembl genome (Homo_sapiens.GRCh38.dna.primary_assembly.fa) and Ensembl annotation (Homo_sapiens.GRCh38.94.gtf). The gene expression counts were calculated using featureCount [57] from Subread (ver 1.6.4) [58].

Differential expression (DE) of RNA-Seq data

The statistically significant DE genes between healthy controls and maternally obese cases were found adjusting for the same clinical confounders for methylation analysis using the 'DESeq2' (version 1.26.0) [59] and 'limma-voom' function from 'limma' package [26]. The p-values were adjusted for multiple hypotheses testing using BH adjustment. Genes with adjusted p-values less than 0.05 were considered statistically significant.

Correlation analysis between bulk RNA-seq and methylation data

A subset of 47 patients have done both methylation and RNA-seq assyas. Differential expression analysis was done on the bulk RNA-seq data using the 'limma' package, with a threshold of BH

adjusted $p<0.05$ to be differential genes. Pearson correlation coefficients (PCC) were calculated between gene expression and methylation beta values from the promoter regions, among the same patients**.** As mostly a negative correlation between gene expression and DNA methylation in the promoter region is expected, genes with a high negative correlation (PCC<-0.2) were used for pathway enrichment using TOPPFUN [60–62]. Top genes of interest were selected with the absolute value Fold Change>1.5 in gene expression and gene-methyl correlation <-0.3 for hyper- and hypo-methylated CpG sites.

Metabolomics analysis

Metabolomics data were acquired from a previously published study involving 87 patients in the same cohort from three batches (metabolomics workbench study ID ST001114) [63]. Targeted metabolites were generated with LC-MS, and untargeted metabolites were generated with GC-MS. After removal of compounds missing in more than 10% samples, a total of 185 metabolites remained, including 10 amino acids (AA), 40 acylcarnitines (AC), 35 acyl/acyl phosphatidylcholines (PC aa), 38 acyl/alkyl phosphatidylcholines (PC ae) and 62 untargeted metabolites. The raw metabolite data were log transformed, standardized, normalized using variance stabilization normalization (VSN), and batch corrected with ComBat function in sva pacakge[64]. Differential metabolites were identified by limma, with clinical confounders adjustment.

Multi-omics integration on metabolomics, epigenomics, and transcriptomics

A subset of 42 patients have the matched methylation, gene expression, and metabolomics data. We applied multi-omics integration with Data Integration Analysis for Biomarker discovery

using Latent cOmponents (DIABLO) implemented in the mixomics package [65]. DIABLO finds the correlated consensus latent variables among different omics in the supervised manner. Top DIABLO features for each omic were selected based on the loading values. We integrated the pathway level methylation, gene, and metabolite interaction using pathview [66].

Results

Overview of study design and cohort characteristics

This study aims to investigate the transgenerational effect of pre-pregnancy maternal obesity on offspring. A total of 72 patients who elected to deliver full-term babies through C-sections were recruited from Kapiolani Medical Center for Women and Children in Honolulu, Hawaii from 2016 to 2018. Among them, 38 deliveries are in the healthy control group and 34 are cases with pre-pregnancy maternal obesity. We excluded natural virginal births, to avoid its potential confounding effect on multi-omics profiles. We also carried out stringent recruitment selection criteria, including matching the mothers' ages as much as possible, as well as similar net gestational weight gain to minimize its confounding effect over maternal pre-pregnancy maternal obesity. The overall study design is shown in **Figure 1**. Briefly, upon collecting the blood samples, umbilical cord blood hematopoietic stem cells (uHSCs) were enriched by FACS sorting with CD34+CD35-LIN- markers (see **Methods**). We extracted DNA and RNA from these uHSCs for Illumina 450k array based DNA methylation and bulk RNA-Seq sequencing respectively. The plasma from these cord blood samples was subject to untargeted metabolomics assays using GC-MS and targeted metabolomics assays using LC–MS [63]. Given the rationale that DNA methylation could be the mediator for exerting the transgenerational effect of maternal

obesity [67,68], we carried out multi-omics data integration analysis in the DNA methylation-centric manner.

The demographic details and clinical information of these patients are summarized in **Table 1**. The distributions of the most representative variables are shown in **Figure 2**. Among categorical demographic variables, the distribution of baby sex had no statistical difference between obese and health groups, whereas the ethnicity distributions among mothers and fathers, parity and gravidity are statistically different ($P<0.05$) between the two groups (**Figure 2A-2E**). Besides maternal pre-pregnancy BMI, other maternal parameters such as maternal age, gestational week, net weight gain and hemoglobin are also not statistically significantly different between the two groups per study design (**Figure 2F-2I, Table 1**). While mothers of Asian ethnicity are the majority in the control group, NHPIs account for the majority of the maternal-obese group, revealing the health disparity issue known in the state of Hawaii [69]. Moreover, the control group has lower parities and gravidities, compared to the cases. Babies born to obese mothers show significantly higher ($P<0.05$) body weights compared to the control group, as expected [70]. Other parameters including the baby gender, head circumference, body length, and APGAR score at 5 min after birth are not statistically significantly different between case and control groups (**Figure 2J-2M**).

Global hypermethylation pattern revealed by CpG level methylation analysis

For scientific rigor, it is critical to adjust for confounding in DNA methylation association analysis [71]. Thus we performed the source of variance (SOV) analysis on the beta values of the DNA methylation with respect to physiological and phenotypic information, in order to assess

potential confounding factors systematically [24,25,71]. As shown in **Figure 3A**, marginal F-statistics in the SOV analysis show that the dominating contribution to DNA methylation variation is maternal pre-pregnancy obesity status, confirming the quality of the study design which aimed to minimize other confounders' effect. The other minor confounding factors include baby sex, maternal age, maternal ethnicity, net weight gain during pregnancy, paternal ethnicity, gravidity, and gestational age (F-statistics>1). After adjusting these factors by linear regression, all have reduced F-statistics of less than 0.5 (**Figure 3B**) except maternal pre-pregnancy obesity, confirming the success of confounding removal.

Next, we conducted differential methylation (DE) analysis on the confounding adjusted DNA methylation data (**Methods**). We observed a global hypermethylation pattern in pre-pregnancy obese samples, with 10,254 hypermethylated vs. 5394 hypomethylated CpG sites (**Figure 3C**). The top 20 differentially hypermethylated and hypomethylated CpG sites are reported in **Table 2**, respectively. These CpG sites are related to a wide variety of biological functions, including inflammation (CD69, ADAM12), transcription factors (ZNF222, HMGN4, LHX6, TAF3), proliferation and apoptosis (HDAC4, DHRS4, LRCH3, SAFB2, CRADD, EBF3, PRKAR1B). Some top DM CpG sites are associated with obesity directly, including HDAC4 [72] and PLEC1[73].

We further examined the distributions of these differentially methylated sites, relative to the CpG island regions and promoter proximity. (**Figure 3D-E**). A big fraction (38.7%) of the DM sites are located in CpG islands [74,75], significantly higher than that from the Illumina 450K annotation (P<2E-16). CpG islands are more frequent in the hypermethylated sites (41.3%) as compared to these in the hypomethylated sites (33.6%), consistent with the global hypermethylation pattern.

Relative to gene localization, DM sites are most frequent (32.1%) in the promoter regions (including 16.8% and 15.3% in TSS200 and TSS1500 respectively) as expected.

Functional analyses reveal the association between maternal obesity and cell cycle, immune function and metabolic changes in the cord blood of offspring

To investigate the biological functions related to the epigenome alternation, we conducted systematic analysis of DM sites employing multiple methods: KEGG pathway enrichment analysis, Weighted Gene Co-expression Network Analysis (WGCNA), and Protein-Protein Interaction (PPI) network analysis.

KEGG pathway enrichment analysis on hypermethylated CpG sites identified five significant pathways with FDR<0.05 (**Figure 4A)**, including the cell cycle, ribosome, nucleocytoplasmic transport, ribosome biogenesis in eukaryotes, and mTOR signaling pathway. Cell cycle, ribosome, and nucleocytoplasmic transport pathways are essential to normal cell functioning. mTOR signaling pathway coordinates the nutrient-mediated metabolism, immune responses and cell cycle progression, and dysregulation of mTOR could lead to various diseases such as cancer and obesity [76]. There was no significantly enriched pathway emerging from hypomethylated CpG sites. The maternally obese group shows significantly higher methylation levels in KEGG protein synthesis and immune system pathway collections compared to the control group, indicating repression in immune response as well as translation and protein synthesis (**Figure 4B-C**). Similarly, we further explored the differential potential, or stemness, of umbilical cord Hematopoietic Stem Cells (uHSCs). We first confirmed the homogeneity of uHSCs by single-cell RNA sequencing UMAP plot (**Supplementary Figure 2**). We calculated the cell

stemness scores using the DNA methylation beta values similar to others [77]. uHSCs derived from the maternally obese group exhibit significantly elevated stemness scores ($P<0.01$) in comparison to the control group **(Figure 4D)**, confirming the results in KEGG pathway analysis.

Next, we applied WGCNA to cluster co-regulation of gene-level methylation, by averaging CpG sites to affiliated genes (see **Methods**). Five co-expression modules are identified, using the M-values adjusted for clinical confounders (**Supplementary Figure 1A)**, and all modules show positive correlations with maternal obesity except one. The largest turquoise module (**Figure 4E**) is related to cell cycle, protein synthesis, and transport and vesicle trafficking pathways through pathway enrichment analysis. Some hub genes in this module are identified, including INTU, ANAPC7, and AGBL5. These genes were reported essential for maintaining cell polarity (INTU)[78], proliferation (ANAPC7) [79] and glycemic control (AGBL5) [80]. The brown module (**Figure 4F**) is enriched with immune response pathways, in which TLR6 is identified as a hub gene. The other yellow module is related to ion homeostasis, and the gray module is related to the p53 pathway, apoptosis, cell senescence, and ER stress (**Supplementary Figure 1B)**. The only negatively correlated blue module is associated with axon guidance and VEGF signaling pathway (**Supplementary Figure 1B**).

Furthermore, we examined the PPI network, using the gene-level DNA methylation as surrogates (**Figure 4G**). The PPI analysis identifies 14 unique pathways (FDR < 0.05) predominantly associated with hypermethylated CpG sites in the TSS200 and TSS1500 regions. The top five largest pathways included ribosome, proteasome, cell cycle, axon guidance, RNA polymerase, and neuroactive ligand-receptor interaction. Taken all three types of systematic analyses together,

cell cycle, immune function and protein synthesis are ubiquitously highlighted, suggesting that these biological functions in cord blood stem cells are negatively impacted by maternal obesity.

Multi-omics analysis reveals disruptions in cell cycle and metabolic pathways

To systematically investigate the epigenetic, transcriptomic, and metabolomic alterations induced by maternal obesity, we performed multi-omics integration analysis on this cohort. We employed DIABLO, a supervised integration method that extracts features associated with maternal obesity, based on the correlations in the embedding space [65]. **Figure 5A-C** shows that methylation data provide the clearest separation between obese and control groups, confirming the value of the earlier DNA methylation centered analysis.

The top 25 features from each omic with the highest feature weights (loadings) following integrated canonical correlation analysis are demonstrated in **Figure 5D-F**. The methylation features with the highest weights related to maternal obesity include CpG sites involved in cell-cycle control, glucose metabolism, and adipogenesis (FOXO1[81]), DNA repair (LIG3, SMUG1), erythropoietin pathway and differentiation (EPO, CSNK2A1, CSF1), which are hypermethylated in the obese group. Hypomethylation of LEP (encoding leptin) was also observed as a top feature, aligning with prior findings that maternal obesity is associated with elevated maternal leptin levels, a known marker of adipose tissue [82]. These featured CpG sites indicate repression in fat metabolism and DNA repair and reduced differentiation potential. In the transcriptomic space, many genes related to mRNA splicing (SRRM1, IGF2BP1, IGF2BP2,

CNOT4) have increased expression levels due to maternal obesity. Among the metabolite features, essential sugars (glucose, xylose), poly-unsatuated fatty acids (oleic acids, DHA, arachidonic acid), and phosphatidylcholine (PCs) are mostly decreased in the obese group; whereas most acylcarnitines (C) are elevated. The metabolic changes show an overall accumulation of saturated fatty acid, but repression on fat breakdown, glucose and unsaturated fatty acid generation. As poly-unsatuated fatty acids (eg. arachidonic acid) have important anti-inflammatory effects, the results indicate a pro-inflammatory environment in offspring born of pre-pregnant obese mothers.

Discussion

Maternal obesity is one of the most urgent health concerns worldwide. Pre-pregnancy maternal obesity could cause various pregnancy-related complications and predispose offspring to cardiometabolic complications and chronic diseases in the long term [9]. Multiple cross-continental large cohort meta-analyses have shown that maternal obesity is directly associated with offspring's risk of obesity, coronary heart disease, insulin resistance, and adverse neurodevelopmental outcome based on longitudinal observational studies [9,83,84]. To directly ping-point the molecular level changes in offspring by maternal pre-pregnancy obesity, we used cord blood stem cells as the studying material, which serve as a great surrogate revealing newborn's metabolism and immune system changes at the time of birth [85]. The current study expands previous effects and investigates the direct impact of maternal obesity on uHSCs programming, the progenitor of immune cell population, using a multi-omics (epigenetic, gene expression, and metabolite) analysis approach from a unique multi-ethnic cohort.

Centered around methylation changes, three complimentary functional analysis approaches (KEGG, WGCNA, and PPI) consistently demonstrated that maternal obesity impacts multiple bilogical functions including hypermethylation in promoters of genes involved in cell cycle, ribosome biogenesis, and mTOR signaling pathways. Moreover, mTOR signaling pathway also plays a crucial role in metabolism and cell cycle regulation, disruption in this pathway leads to insulin resistance and long-term diseases [86]. We observed a significant increase in stemness scores among uHSCs affected by maternal obesity, aligning with expected downregulation in the cell cycle gene expression due to observed hypermethylation in the promoters of these genes. Higher stembess scores indicate enhanced quiescence, shift the balance between stem cell maintenance and differentiation towards the former. Unlike adult HSCs, fetal/neonatal HSCs typically exhibit higher proliferation and self-renewal capabilities, crucial for blood cell regeneration and innate immune system development [87]. Our findings provide strong epigenetic evidence that maternal obesity compromises the maturation processes in neonatal uHSCs, which may predispose newborns to immunological disorders.

The subsequent multi-omics integration analysis expanded conclusions from methylation analysis to additional metabolomics readouts that are also linked to biological functions eg. cell cycle and inflammatory pathway. We thus propose the conceptual model to illustrate the effect of maternal pre-pregnancy obesity (**Figure 6**). Maternal obesity leads to nutrient deficiency with lower levels of essential amino acids and fatty acids in the newborn blood and disrupts the lipid metabolism homeosis in offspring. These metabolite changes further induce cell membrane instability and repress cell cycle progression, cell proliferation[88], enhancing the dysregulation of these functions preexisting at the methylation level. Lipid dysregulation may also enhance the

pro-inflammatory environment, which in turn induces complications in offspring later in life, such as cardiovascular diseases. Such proposed model is also consistent with and further strengthens previous studies at the metabolomics or epigenome levels. For example, previous metabolomics studies of cord blood showed metabolic derangement predisposes newborns to cardiometabolic and endocrine diseases, and disrupt the normal hormone function and neonatal adiposity [82,89]. Previous epigenome-wide association study (EWAS) with cord blood found a strong association between DNA methylation pattern and postnatal BMI trajectory until adolescent [90].

In summary, this newborn study demonstrates the direct impact of maternal pre-pregnancy obesity and on newborn blood at the multi-omics level, which includes increased cell cycle arrest, impairment in the uHSC differentiation capacity, more inflammation and disruption in lipid metabolism.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the author(s) used GPT-4.0 in order to improve the readability. After using this tool/service, the author(s) reviewed and edited the content thoroughly and take(s) full responsibility for the content of the publication.

Data availability statement

Data generated in this study have been submitted and will be available through the National Institutes of Health Gene Expression Omnibus (accession number pending). Code to produce the analyses in this manuscript are available through GitHub (https://github.com/lanagarmire/)

Acknowledgments

This research was supported by grants R01 LM012373 and LM012907 awarded by NLM, and R01 HD084633 awarded by NICHD to L.X. Garmire, as well as in part by the NCI Cancer Center Support Grant (CCSG) number P30 CA071789 awarded to Genomics and Bioinformatics Shared Resource (RRID:SCR_019085). This research was supported in part by training funding provided by the NIH grant T32 GM141746 and Advanced Proteogenomics of Cancer (T32 CA140044).

Author contributions

LG envisioned this project, obtained the funding, supervised the study and revised the manuscript. YD performed the data analysis, generated the figures, and wrote the initial manuscript. RS consented patients and obtained the samples from the hospital, with coordination from PB. DT and SW coordinated with the patient recruitment and study. PB coordinated all the multi-omics assays. PB, CL, and FA designed the DNA methylation assays. AG performed FACS sorting of cord blood cells. ALJ performed the Illumina Meth 450 assay, MT supervised the Genomics Shared Resource analyses and provided a critical review of the manuscript. All authors have read the manuscript.

Conflicts of interest

None

References

1. Leddy, M.A., Power, M.L., and Schulkin, J. (2008). The impact of maternal obesity on maternal and fetal health. Rev. Obstet. Gynecol. *1*, 170–178.

2. Hjalgrim, L.L., Westergaard, T., Rostgaard, K., Schmiegelow, K., Melbye, M., Hjalgrim, H., and Engels, E.A. (2003). Birth weight as a risk factor for childhood leukemia: a meta-analysis of 18 epidemiologic studies. Am. J. Epidemiol. *158*, 724–735.

3. Harder, T., Plagemann, A., and Harder, A. (2010). Birth weight and risk of neuroblastoma: a meta-analysis. Int. J. Epidemiol. *39*, 746–756.

4. Cnattingius, S., Lundberg, F., Sandin, S., Grönberg, H., and Iliadou, A. (2009). Birth characteristics and risk of prostate cancer: the contribution of genetic factors. Cancer Epidemiol. Biomarkers Prev. *18*, 2422–2426.

5. Eriksson, M., Wedel, H., Wallander, M.-A., Krakau, I., Hugosson, J., Carlsson, S., and Svärdsudd, K. (2007). The impact of birth weight on prostate cancer incidence and mortality in a population-based study of men born in 1913 and followed up from 50 to 85 years of age. Prostate *67*, 1247–1254.

6. Michos, A., Xue, F., and Michels, K.B. (2007). Birth weight and the risk of testicular cancer: a meta-analysis. Int. J. Cancer *121*, 1123–1131.

7. Silva, I. dos S., De Stavola, B., McCormack, V., and Collaborative Group on Pre-Natal Risk Factors and Subsequent Risk of Breast Cancer (2008). Birth size and breast cancer risk: re-analysis of individual participant data from 32 studies. PLoS Med. *5*, e193.

8. Van Cleave, J., Gortmaker, S.L., and Perrin, J.M. (2010). Dynamics of obesity and chronic health conditions among children and youth. JAMA *303*, 623–630.

9. Godfrey, K.M., Reynolds, R.M., Prescott, S.L., Nyirenda, M., Jaddoe, V.W., Eriksson, J.G., and Broekman, B.F. (2017). Influence of maternal obesity on the long-term health of offspring. The lancet. Diabetes & endocrinology *5*. 10.1016/S2213-8587(16)30107-3.

10. Barker, D.J. (2007). The origins of the developmental origins theory. J. Intern. Med. *261*. 10.1111/j.1365-2796.2007.01809.x.

11. Barker, D.J. (2000). In utero programming of cardiovascular disease. Theriogenology *53*. 10.1016/s0093-691x(99)00258-7.

12. Qiu, L., Low, H.P., Chang, C.-I., Strohsnitter, W.C., Anderson, M., Edmiston, K., Adami, H.-O., Ekbom, A., Hall, P., Lagiou, P., et al. (2012). Novel measurements of mammary stem cells in human umbilical cord blood as prospective predictors of breast cancer susceptibility in later life. Ann. Oncol. *23*, 245–250.

13. Strohsnitter, W.C., Savarese, T.M., Low, H.P., Chelmow, D.P., Lagiou, P., Lambe, M., Edmiston, K., Liu, Q., Baik, I., Noller, K.L., et al. (2008). Correlation of umbilical cord blood haematopoietic stem and progenitor cell levels with birth weight: implications for a prenatal influence on cancer risk. Br. J. Cancer *98*, 660–663.

14. Apgar, V. (1953). A Proposal for a New Method of Evaluation of the Newborn Infant. Anesthesia & Analgesia *32*, 260.

15. Morris, T.J., Butcher, L.M., Feber, A., Teschendorff, A.E., Chakravarthy, A.R., Wojdacz, T.K., and Beck, S. (2014). ChAMP: 450k Chip Analysis Methylation Pipeline. Bioinformatics *30*, 428–430.

16. Aryee, M.J., Jaffe, A.E., Corrada-Bravo, H., Ladd-Acosta, C., Feinberg, A.P., Hansen, K.D., and Irizarry, R.A. (2014). Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics *30*, 1363–1369.

17. Fortin, J.-P., Triche, T.J., Jr, and Hansen, K.D. (2017). Preprocessing, normalization and integration of the Illumina HumanMethylationEPIC array with minfi. Bioinformatics *33*, 558–560.

18. Zhou, W., Laird, P.W., and Shen, H. (2017). Comprehensive characterization, annotation and innovative use of Infinium DNA methylation BeadChip probes. Nucleic Acids Res. *45*, e22.

19. Teschendorff, A.E., Marabita, F., Lechner, M., Bartlett, T., Tegner, J., Gomez-Cabrero, D., and Beck, S. (2013). A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. Bioinformatics *29*, 189–196.

20. Du, P., Kibbe, W.A., and Lin, S.M. (2007). nuID: a universal naming scheme of oligonucleotides for illumina, affymetrix, and other microarrays. Biol. Direct *2*, 16.

21. Lin, S.M., Du, P., Huber, W., and Kibbe, W.A. (2008). Model-based variance-stabilizing transformation for Illumina microarray data. Nucleic Acids Res. *36*, e11.

22. Du, P., Kibbe, W.A., and Lin, S.M. (2008). lumi: a pipeline for processing Illumina microarray. Bioinformatics *24*, 1547–1548.

23. Du, P., Zhang, X., Huang, C.-C., Jafari, N., Kibbe, W.A., Hou, L., and Lin, S.M. (2010). Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. BMC Bioinformatics *11*, 587.

24. He, B., Liu, Y., Maurya, M.R., Benny, P., Lassiter, C., Li, H., Subramaniam, S., and Garmire, L.X. (2021). The maternal blood lipidome is indicative of the pathogenesis of severe preeclampsia. J. Lipid Res. *62*. 10.1016/j.jlr.2021.100118.

25. Chen, Y., He, B., Liu, Y., Aung, M.T., Rosario-Pabón, Z., Vélez-Vega, C.M., Alshawabkeh, A., Cordero, J.F., Meeker, J.D., and Garmire, L.X. (2022). Maternal plasma lipids are involved in the pathogenesis of preterm birth. Gigascience *11*. 10.1093/gigascience/giac004.

26. Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. *43*, e47.

27. Jaffe, A.E., Murakami, P., Lee, H., Leek, J.T., Fallin, M.D., Feinberg, A.P., and Irizarry, R.A. (2012). Bump hunting to identify differentially methylated regions in epigenetic epidemiology studies. Int. J. Epidemiol. *41*, 200–209.

28. Novakovic, B., Yuen, R.K., Gordon, L., Penaherrera, M.S., Sharkey, A., Moffett, A., Craig, J.M., Robinson, W.P., and Saffery, R. (2011). Evidence for widespread changes in promoter methylation profile in human placenta in response to increasing gestational age and environmental/stochastic factors. BMC Genomics *12*. 10.1186/1471-2164-12-529.

29. GEO Accession viewer https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE36829.

30. Chu, T., Bunce, K., Shaw, P., Shridhar, V., Althouse, A., Hubel, C., and Peters, D. (2014). Comprehensive analysis of preeclampsia-associated DNA methylation in the placenta. PLoS One *9*. 10.1371/journal.pone.0107318.

31. Blair, J.D., Yuen, R.K., Lim, B.K., McFadden, D.E., von Dadelszen, P., and Robinson, W.P. (2013). Widespread DNA hypomethylation at gene enhancer regions in placentas associated with early-onset pre-eclampsia. Mol. Hum. Reprod. *19*. 10.1093/molehr/gat044.

32. Hanna, C.W., Peñaherrera, M.S., Saadeh, H., Andrews, S., McFadden, D.E., Kelsey, G., and Robinson, W.P. (2016). Pervasive polymorphic imprinted methylation in the human placenta. Genome Res. *26*. 10.1101/gr.196139.115.

33. Blair, J.D., Langlois, S., McFadden, D.E., and Robinson, W.P. (2014). Overlapping DNA methylation profile between placentas with trisomy 16 and early-onset preeclampsia. Placenta *35*. 10.1016/j.placenta.2014.01.001.

34. Price, E.M., Peñaherrera, M.S., Portales-Casamar, E., Pavlidis, P., Van Allen, M.I., McFadden, D.E., and Robinson, W.P. (2016). Profiling placental and fetal DNA methylation in human neural tube defects. Epigenetics Chromatin *9*. 10.1186/s13072-016-0054-8.

35. Leavey, K., Wilson, S.L., Bainbridge, S.A., Robinson, W.P., and Cox, B.J. (2018). Epigenetic regulation of placental gene expression in transcriptional subtypes of preeclampsia. Clin. Epigenetics *10*. 10.1186/s13148-018-0463-6.

36. Wilson, S.L., Leavey, K., Cox, B.J., and Robinson, W.P. (2018). Mining DNA methylation

alterations towards a classification of placental pathologies. Hum. Mol. Genet. *27*. 10.1093/hmg/ddx391.

37. Hansen, K.D. IlluminaHumanMethylation450kanno. ilmn12. hg19: annotation for Illumina's 450k methylation arrays. R package version 0.6. 0.

38. Maksimovic, J., Gordon, L., and Oshlack, A. (2012). SWAN: Subset-quantile within array normalization for illumina infinium HumanMethylation450 BeadChips. Genome Biol. *13*, R44.

39. Phipson, B., and Oshlack, A. (2014). DiffVar: a new method for detecting differential variability with application to methylation in cancer and aging. Genome Biol. *15*, 465.

40. Maksimovic, J., Gagnon-Bartsch, J.A., Speed, T.P., and Oshlack, A. (2015). Removing unwanted variation in a differential methylation analysis of Illumina HumanMethylation450 array data. Nucleic Acids Res. *43*, e106.

41. Phipson, B., Maksimovic, J., and Oshlack, A. (2016). missMethyl: an R package for analyzing data from Illumina's HumanMethylation450 platform. Bioinformatics *32*, 286–288.

42. Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. *28*, 27–30.

43. Kanehisa, M. (2019). Toward understanding the origin and evolution of cellular organisms. Protein Sci. *28*, 1947–1951.

44. Kanehisa, M., Furumichi, M., Sato, Y., Kawashima, M., and Ishiguro-Watanabe, M. (2022). KEGG for taxonomy-based analysis of pathways and genomes. Nucleic Acids Res. 10.1093/nar/gkac963.

45. Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics *9*, 559.

46. Langfelder, P., and Horvath, S. (2012). Fast R Functions for Robust Correlations and Hierarchical Clustering. J. Stat. Softw. *46*. 10.18637/jss.v046.i11.

47. Szklarczyk, D., Gable, A.L., Nastou, K.C., Lyon, D., Kirsch, R., Pyysalo, S., Doncheva, N.T., Legeay, M., Fang, T., Bork, P., et al. (2021). The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. Nucleic Acids Res. *49*, D605–D612.

48. Li, H., Tong, P., Gallegos, J., Dimmer, E., Cai, G., Molldrem, J.J., and Liang, S. (2015). PAND: A Distribution to Identify Functional Linkage from Networks with Preferential Attachment Property. PLoS One *10*, e0127968.

49. Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated

models of biomolecular interaction networks. Genome Res. *13*, 2498–2504.

50. He, B., and Garmire, L.X. (2021). ASGARD: A Single-cell Guided pipeline to Aid Repurposing of Drugs. ArXiv.

51. MacArthur, B.D., and Lemischka, I.R. (2013). Statistical mechanics of pluripotency. Cell *154*, 484–489.

52. Martínez, O., and Reyes-Valdés, M.H. (2008). Defining diversity, specialization, and gene specificity in transcriptomes through information theory. Proc. Natl. Acad. Sci. U. S. A. *105*, 9709–9714.

53. Kannan, S., Farid, M., Lin, B.L., Miyamoto, M., and Kwon, C. (2021). Transcriptomic entropy benchmarks stem cell-derived cardiomyocyte maturation against endogenous tissue at single cell level. PLoS Comput. Biol. *17*, e1009305.

54. Bushnell, B. (2014). BBMap: A Fast, Accurate, Splice-Aware Aligner.

55. Andrews, S. FastQC: a quality control tool for high throughput sequence data. Available online. Retrieved May.

56. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

57. Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923–930.

58. Liao, Y., Smyth, G.K., and Shi, W. (2013). The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. Nucleic Acids Res. *41*, e108.

59. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.

60. Chen, J., Xu, H., Aronow, B.J., and Jegga, A.G. (2007). Improved human disease candidate gene prioritization using mouse phenotype. BMC Bioinformatics *8*, 392.

61. Chen, J., Aronow, B.J., and Jegga, A.G. (2009). Disease candidate gene identification and prioritization using protein interaction networks. BMC Bioinformatics *10*, 73.

62. Chen, J., Bardes, E.E., Aronow, B.J., and Jegga, A.G. (2009). ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. Nucleic Acids Res. *37*, W305–W311.

63. Schlueter, R.J., Al-Akwaa, F.M., Benny, P.A., Gurary, A., Xie, G., Jia, W., Chun, S.J., Chern, I., and Garmire, L.X. (2020). Prepregnant Obesity of Mothers in a Multiethnic Cohort Is Associated with Cord Blood Metabolomic Changes in Offspring. J. Proteome Res. *19*, 1361–1374.

64. Johnson, W.E., Li, C., and Rabinovic, A. (2006). Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics *8*, 118–127.

65. Singh, A., Shannon, C.P., Gautier, B., Rohart, F., Vacher, M., Tebbutt, S.J., and Cao, K.-A.L. (2019). DIABLO: an integrative approach for identifying key molecular drivers from multi-omics assays. Bioinformatics *35*, 3055.

66. Luo, W., and Brouwer, C. (2013). Pathview: an R/Bioconductor package for pathway-based data integration and visualization. Bioinformatics *29*, 1830–1831.

67. Heard, E., and Martienssen, R.A. (2014). Transgenerational epigenetic inheritance: myths and mechanisms. Cell *157*, 95–109.

68. King, S.E., and Skinner, M.K. (2020). Epigenetic Transgenerational Inheritance of Obesity Susceptibility. Trends Endocrinol. Metab. *31*, 478–494.

69. Morisako, A.K., Tauali‘i, M., Ambrose, A.J.H., and Withy, K. (2017). Beyond the Ability to Pay: The Health Status of Native Hawaiians and Other Pacific Islanders in Relationship to Health Insurance. Hawaii J. Med. Public Health *76*, 36.

70. Heslehurst, N., Vieira, R., Akhter, Z., Bailey, H., Slack, E., Ngongalah, L., Pemu, A., and Rankin, J. (2019). The association between maternal body mass index and child obesity: A systematic review and meta-analysis. PLoS Med. *16*, e1002817.

71. Liu, W., Yang, X., Mao, Z., Du, Y., Lassiter, C., AlAkwaa, F.M., Benny, P.A., and Garmire, L.X. Severe preeclampsia is not associated with significant DNA methylation changes but cell proportion changes in the cord blood - caution on the importance of confounding adjustment. medRxiv. 10.1101/2023.08.31.23294898.

72. Abu-Farha, M., Tiss, A., Abubaker, J., Khadir, A., Al-Ghimlas, F., Al-Khairi, I., Baturcam, E., Cherian, P., Elkum, N., Hammad, M., et al. (2013). Proteomics Analysis of Human Obesity Reveals the Epigenetic Factor HDAC4 as a Potential Target for Obesity. PLoS One *8*. 10.1371/journal.pone.0075342.

73. Rönn, T., Volkov, P., Gillberg, L., Kokosar, M., Perfilyev, A., Jacobsen, A.L., Jørgensen, S.W., Brøns, C., Jansson, P.A., Eriksson, K.F., et al. (2015). Impact of age, BMI and HbA1c levels on the genome-wide DNA methylation and mRNA expression patterns in human adipose tissue and identification of epigenetic biomarkers in blood. Hum. Mol. Genet. *24*. 10.1093/hmg/ddv124.

74. Lim, W.-J., Kim, K.H., Kim, J.-Y., Jeong, S., and Kim, N. (2019). Identification of DNA-Methylated CpG Islands Associated With Gene Silencing in the Adult Body Tissues of the Ogye Chicken Using RNA-Seq and Reduced Representation Bisulfite Sequencing. Front. Genet. *10*, 346.

75. Ching, T., Song, M.-A., Tiirikainen, M., Molnar, J., Berry, M., Towner, D., and Garmire, L.X. (2014). Genome-wide hypermethylation coupled with promoter hypomethylation in the chorioamniotic membranes of early onset pre-eclampsia. Mol. Hum. Reprod. *20*,

885–904.

76. Meng, D., Frank, A.R., and Jewell, J.L. (2018). mTOR signaling in stem and progenitor cells. Development *145*. 10.1242/dev.152595.

77. Guo, M., Bao, E.L., Wagner, M., Whitsett, J.A., and Xu, Y. (2017). SLICE: determining cell differentiation and lineage based on single cell entropy. Nucleic Acids Res. *45*, e54.

78. Dai, D., Li, L., Huebner, A., Zeng, H., Guevara, E., Claypool, D.J., Liu, A., and Chen, J. (2012). Planar cell polarity effector gene Intu regulates cell fate-specific differentiation of keratinocytes through the primary cilia. Cell Death Differ. *20*, 130–138.

79. Liu, J., and Fuchs, S.Y. (2006). Cross-talk between APC/C and CBP/p300. Cancer Biol. Ther. *5*. 10.4161/cbt.5.7.3103.

80. Corbi, S.C.T., Bastos, A.S., Nepomuceno, R., Cirelli, T., Dos Santos, R.A., Takahashi, C.S., Rocha, C.S., Orrico, S.R.P., Maurer-Morelli, C.V., and Scarel-Caminaga, R.M. (2017). Expression Profile of Genes Potentially Associated with Adequate Glycemic Control in Patients with Type 2 Diabetes Mellitus. Journal of diabetes research *2017*. 10.1155/2017/2180819.

81. Behl, T., Kaur, I., Sehgal, A., Singh, S., Zengin, G., Negrut, N., Nistor-Cseppento, D.C., Pavel, F.M., Aron, R.A.C., and Bungau, S. (2021). Exploring the Genetic Conception of Obesity via the Dual Role of FoxO. Int. J. Mol. Sci. *22*. 10.3390/ijms22063179.

82. Kadakia, R., Zheng, Y., Zhang, Z., Zhang, W., Hou, L., and Josefson, J.L. (2017). Maternal pre-pregnancy BMI downregulates neonatal cord blood LEP methylation. Pediatr. Obes. *12 Suppl 1*, 57–64.

83. Yu, Z., Han, S., Zhu, J., Sun, X., Ji, C., and Guo, X. (2013). Pre-pregnancy body mass index in relation to infant birth weight and offspring overweight/obesity: a systematic review and meta-analysis. PLoS One *8*, e61627.

84. Sureshchandra, S., Marshall, N.E., and Messaoudi, I. (2019). Impact of pregravid obesity on maternal and fetal immunity: Fertile grounds for reprogramming. J. Leukoc. Biol. *106*, 1035–1050.

85. Levy, O. (2005). Innate immunity of the human newborn: distinct cytokine responses to LPS and other Toll-like receptor agonists. J. Endotoxin Res. *11*, 113–116.

86. Ong, P.S., Wang, L.Z., Dai, X., Tseng, S.H., Loo, S.J., and Sethi, G. (2016). Judicious Toggling of mTOR Activity to Combat Insulin Resistance and Cancer: Current Evidence and Perspectives. Front. Pharmacol. *7*. 10.3389/fphar.2016.00395.

87. Mack, R., Zhang, L., Breslin, P., and Zhang, J. (2021). The fetal-to-adult hematopoietic stem cell transition and its role in childhood hematopoietic malignancies. Stem cell reviews and reports *17*, 2059.

88. Kwok, A.C., and Wong, J.T. (2005). Lipid biosynthesis and its coordination with cell cycle progression. Plant Cell Physiol. *46*. 10.1093/pcp/pci213.

89. Denizli, M., Capitano, M.L., and Kua, K.L. (2022). Maternal obesity and the impact of associated early-life inflammation on long-term health of offspring. Front. Cell. Infect. Microbiol. *12*, 940937.

90. Meir, A.Y., Huang, W., Cao, T., Hong, X., Wang, G., Pearson, C., Adams, W.G., Wang, X., and Liang, L. (2023). Umbilical cord DNA methylation is associated with body mass index trajectories from birth to adolescence. EBioMedicine *91*, 104550.

91. Oesterreich, S., Allredl, D.C., Mohsin, S.K., Zhang, Q., Wong, H., Lee, A.V., Osborne, C.K., and O'Connell, P. (2001). High rates of loss of heterozygosity on chromosome 19p13 in human breast cancer. Br. J. Cancer *84*, 493–498.

92. Hong, E.A., Gautrey, H.L., Elliott, D.J., and Tyson-Capper, A.J. (2012). SAFB1- and SAFB2-mediated transcriptional repression: relevance to cancer. Biochem. Soc. Trans. *40*, 826–830.

93. Hammerich-Hille, S., Kaipparettu, B.A., Tsimelzon, A., Creighton, C.J., Jiang, S., Polo, J.M., Melnick, A., Meyer, R., and Oesterreich, S. (2010). SAFB1 mediates repression of immune regulators and apoptotic genes in breast cancer cells. J. Biol. Chem. *285*, 3608–3616.

Figure legends

Figure 1. Overview of the study design and analysis. In the preparation step, cord blood plasma samples are collected for metabolome profiling and stem cell sorting. DNA and RNA extraction assays are performed on the enriched stem cells for the methylation and RNA-seq analyses. Downstream analyses are mainly focused on the methylation data. Bulk RNA-seq data were used for validations for methylation discoveries. (Created with BioRender.com)

Figure 2. Mother and newborns statistics of the multi-ethnic cohort from Hawaii. (**A-E**) Categorical variables including baby sex, maternal ethnicity, paternal ethnicity, parity and gravidity between control and obese groups are shown in the barplots. P-values using Chi-square test are annotated comparing control and obese groups. **(F-I**) The distributions of maternal age, gestation age, maternal net weight gain during pregnancy, and maternal hemoglobin between control and obese groups are compared. Mean and standard deviation are shown in boxplot. P-values using t-test are annotated. (**J-M**) The distributions of baby weight, baby head circumference, baby length, and APGAR score after 5 minutes of delivery between control and obese groups are compared. Mean and standard deviation are shown in boxplot. P-values using t-test are annotated.

Figure 3. DNA methylation analysis on uHSCs.

(**A-B**) Source of variance plot before and after confounding adjustment. F-statistics are reported for each clinical factor. F statistics greater than 1 are considered to have confounding effects in addition to the case/control difference due to pre-pregnancy maternal obesity. (**C**) Volcano plot of

-log(BH adjusted p-values) against logFC. The cutoff line for adjusted p-value < 0.05 is shown as the red horizontal line. The hyper/hypo threshold is shown as a blue vertical line where logFC=0. Non-significant methylation CpG sites after the differential analysis were shown in gray. Significant CpG sites are colored. After the removal of gestational age-related CpG sites, 10,254 CpG sites are hypermethylated and 5,394 sites are hypomethylated. (**D-E**) Normalized location distribution of differentially methylated CpG sites according to their CpG features in terms of isle regions and gene regions based on the chip annotation. Isle regions include shelf, shore, island, and open-sea. Gene regions include gene body, intergenic region (IGR), TSS200, TSS1500, 5'UTR, 3'UTR, and 1st Exon.

Figure 4. Pathway and network analysis. (**A**) KEGG pathway enrichment for hypermethylated CpG sites from promotor region. Enriched KEGG pathway names, adjusted p-values (-log10 transformed), and the size of enriched gene list are reported for CpG sites from TSS200+TSS1500 regions. The red dotted line shows the threshold cutoff for FDR at -log10(0.05). (**B-C**) Boxplots of averaged beta values for KEGG protein pathway collection and immune pathway collection with Wilcoxon P-values. (**D**) Violin plots of cell entropy scores between control and obese groups with Wilcoxon P-values. (**E-F**) WGCNA network analysis results. WGCNA modules are shown for both the control and the obese group. The top two modules with largest degrees are turquoise and brown modules. Each node represents a gene. Genes co-expressed in each module are annotated. (**G**)Protein-protein interaction (PPI) network. Bipartite graphs represent enriched KEGG pathways and associated genes with significant PPIs. Red nodes represent genes with hypermethylated CpG sites. Blue nodes represent genes with hypomethylated CpG sites. Yellow nodes represented the enriched KEGG pathways. Number of

inter-pathway PPIs are annotated in the rectangular boxes.

Figure 5. Multi-omics integration analysis

(**A-C**) Omics-specific sample plots from DIABLO showing the separation of obese and control samples in methylation data, gene expression data, and metabolomics data respectively. (**D-F**) Importance plot of top 25 features in methylation, gene expression and metabolomics modalities with the highest loadings extracted from the embedding space. The color represents the condition which features contribute the most.

Figure 6. A proposed model of maternal obesity's impact on neonatal development.

Tables

Table 1. Summary statistics of the study cohort.

		Control (n=38)	Case (n=34)
Maternal Age		31.3±5.6	31.6±4.9
Gestational Week		38.9±0.5	39.0±0.3
Net Weight Gain		32.0±11.6	30.9±14.6
Hemoglobin		11.6±1.6	11.0±1.4
Maternal Ethnicity	Asian	21	8
	Caucasian	11	4
	NHPI	6	22
Paternal Ethnicity	Asian	19	11
	Caucasian	11	2
	NHPI	8	21

Baby Sex	Female	17	21
	Male	21	13
Parity	0	7	3
	1	21	7
	2	9	11
	More	1	13
Gravidity	1	6	2
	2	15	5
	3	13	8
	4	3	6
	5	1	4
	More	0	9

Demographic and clinical statistics are reported for the control and maternally obese groups.

Table 2. Top 20 hypermethylated CpG sites and top 20 hypomethylated CpG sites.

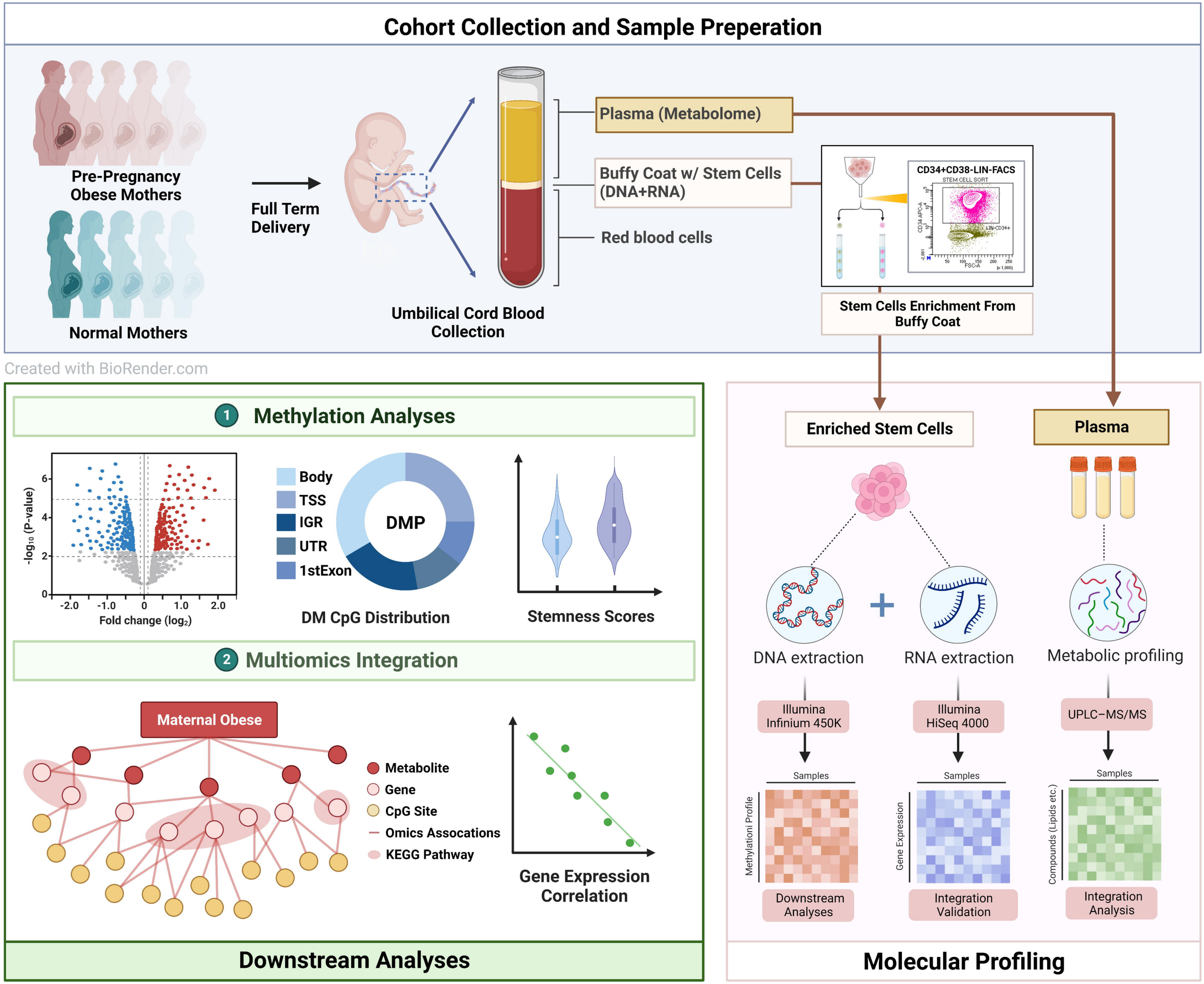
CpG	Gene	Island	Group	logFC	P.Value	adj.P.Val	Type
cg12303247	SYT11	OpenSea	3'UTR	2.188	2.44E-05	6.09E-03	Hyper
cg16818768	PSMG1	Island	TSS1500	1.605	2.15E-05	5.74E-03	Hyper
cg05995465	HDAC4	OpenSea	5'UTR	1.604	1.65E-03	4.65E-02	Hyper
cg01937701	DHRS4	Island	TSS200	1.592	1.96E-10	2.45E-05	Hyper
cg22243583	DLEU1	S_Shore	Body	1.522	2.53E-06	2.01E-03	Hyper
cg16927136	RPL35A	OpenSea	TSS1500	1.507	2.39E-10	2.45E-05	Hyper

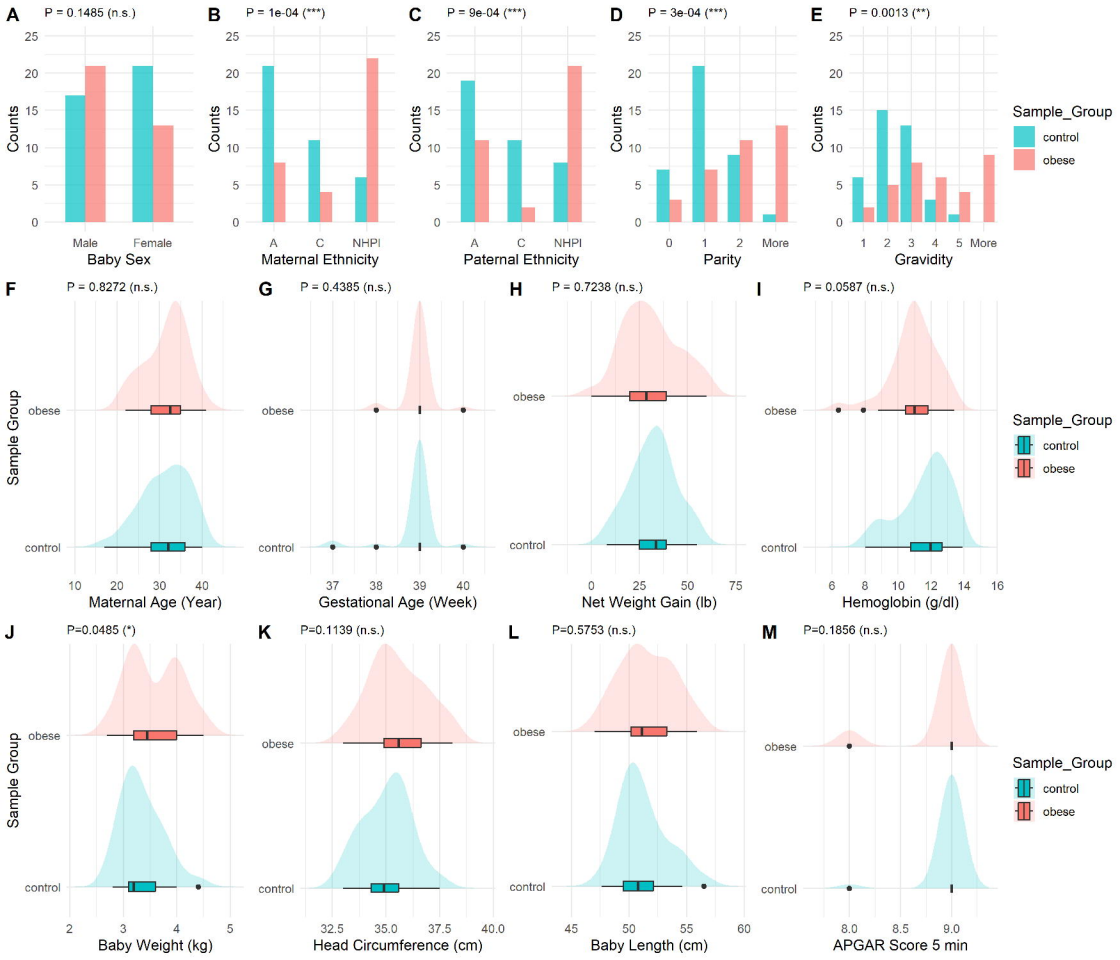
cg08899199	ST7	S_Shore	Body	1.4	7.34E-07	1.13E-03	Hyper
cg05054115	DHRS4	Island	TSS200	1.389	6.66E-08	3.73E-04	Hyper
cg12878710	LRCH3	Island	TSS200	1.387	1.27E-06	1.48E-03	Hyper
cg05130022	HMGN4	N_Shore	TSS200	1.386	1.51E-04	1.45E-02	Hyper
cg05643303	HOXC8	Island	TSS200	1.345	2.70E-05	6.34E-03	Hyper
cg07449543	CHORDC1	S_Shore	TSS200	1.342	6.32E-05	9.53E-03	Hyper
cg25016112	DENND3	OpenSea	Body	1.314	1.22E-03	4.01E-02	Hyper
cg09552166	MSL2	N_Shore	TSS200	1.296	2.30E-05	5.92E-03	Hyper
cg01003902	SAFB2	Island	TSS200	1.269	1.04E-08	1.42E-04	Hyper
cg11028445	FAM96A	N_Shore	TSS1500	1.265	1.97E-04	1.65E-02	Hyper
cg10317138	ADAM12	N_Shore	Body	1.229	5.06E-04	2.60E-02	Hyper
cg09757277	ZNF222	S_Shore	5'UTR	1.229	9.88E-08	4.27E-04	Hyper
cg04117338	CRADD	N_Shore	5'UTR	1.209	1.66E-03	4.67E-02	Hyper
cg07354583	CD69	OpenSea	Body	1.205	5.94E-07	1.02E-03	Hyper
cg04043455	EBF3	S_Shelf	Body	-2.031	6.12E-04	0.029	Hypo
cg20784		N_Shore	Body	-1.812	1.96E-05	0.005	Hypo

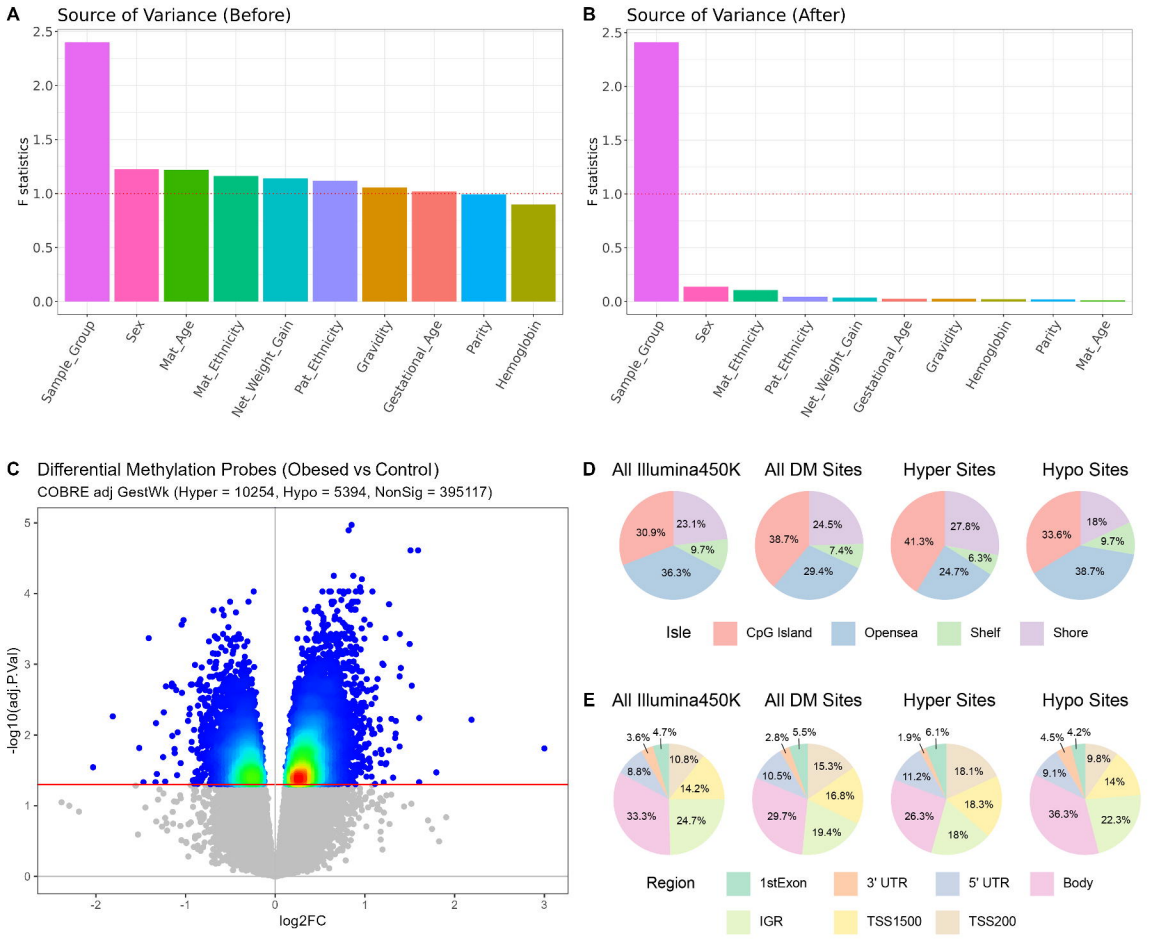
950	PLEC1						
cg09976051	AGA	N_Shore	Body	-1.516	1.67E-04	0.015	Hypo
cg13862711	LHX6	Island	Body	-1.469	1.65E-03	0.047	Hypo
cg16434331	SLC39A11	OpenSea	Body	-1.411	9.52E-08	0	Hypo
cg05636467	EBF3	S_Shelf	Body	-1.335	1.65E-03	0.047	Hypo
cg16858146	TAF3	S_Shelf	Body	-1.33	3.15E-05	0.007	Hypo
cg24796644	MDGA1	Island	Body	-1.242	1.48E-05	0.005	Hypo
cg11064039	PRKAR1B	Island	5'UTR	-1.227	1.58E-03	0.046	Hypo
cg06833656	TBCD	OpenSea	Body	-1.219	2.68E-06	0.002	Hypo
cg25430507	NXPH2	S_Shore	TSS1500	-1.152	2.08E-06	0.002	Hypo
cg03485608	NXPH2	N_Shore	Body	-1.152	2.72E-06	0.002	Hypo
cg00928596	MIR365-1	OpenSea	TSS200	-1.148	7.32E-05	0.01	Hypo
cg12601963	NCRNA00200	Island	Body	-1.132	2.58E-06	0.002	Hypo
cg22772691	SLC12A7	S_Shelf	Body	-1.123	1.79E-04	0.016	Hypo
cg02584267	EBF3	OpenSea	Body	-1.121	2.39E-04	0.018	Hypo

cg19282259	NCRNA00200	Island	TSS200	-1.104	3.49E-06	0.002	Hypo
cg08010094	NXPH2	S_Shore	TSS1500	-1.094	1.04E-03	0.037	Hypo
cg06916001	MIR365-1	OpenSea	TSS200	-1.088	5.74E-05	0.009	Hypo
cg03721387	KRTAP24-1	OpenSea	3'UTR	-1.04	4.29E-06	0.003	Hypo

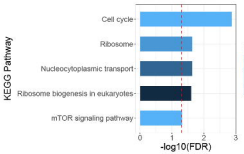
logFC, p-values, BH adjusted p-values, and CpG annotations are reported for the top 20 differentially hypermethylated CpG sites ordered by the adjusted p-values by 'limma' packages. Hypermethylated CpG sites are defined as logFC>0, whereas hypomethylated CpG sites are defined as logFC<0.


Cohort Collection and Sample Preperation
Pre-Pregnancy Obese Mothers
Normal Mothers
Full Term Delivery
Umbilical Cord Blood Collection
Plasma (Metabolome)
Buffy Coat w/ Stem Cells (DNA+RNA)
Red blood cells
CD34+CD38-LIN-FACS
Stem Cells Enrichment From Buffy Coat
Created with BioRender.com
1 Methylation Analyses
-log10 (P-value)
Fold change (log2)
Body
TSS
IGR
UTR
1stExon
DMP
DM CpG Distribution
Stemness Scores
2 Multiomics Integration
Maternal Obese
Metabolite
Gene
CpG Site
Omics Assocations
KEGG Pathway
Gene Expression Correlation
Downstream Analyses
Enriched Stem Cells
Plasma
DNA extraction
RNA extraction
Metabolic profiling
Illumina Infinium 450K
Illumina HiSeq 4000
UPLC–MS/MS
Samples
Methylationi Profile
Gene Expression
Compounds (Lipids etc.)
Downstream Analyses
Integration Validation
Integration Analysis
Molecular Profiling

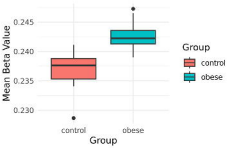

A
P = 0.1485 (n.s.)
Counts
Male
Female
Baby Sex
B
P = 1e-04 (***)
A
C
NHPI
Maternal Ethnicity
C
P = 9e-04 (***)
Paternal Ethnicity
D
P = 3e-04 (***)
0
1
2
More
Parity
E
P = 0.0013 (**)
1
2
3
4
5
More
Gravidity
Sample_Group
control
obese
F
P = 0.8272 (n.s.)
Sample Group
Maternal Age (Year)
G
P = 0.4385 (n.s.)
Gestational Age (Week)
H
P = 0.7238 (n.s.)
Net Weight Gain (lb)
I
P = 0.0587 (n.s.)
Hemoglobin (g/dl)
J
P=0.0485 (*)
Baby Weight (kg)
K
P=0.1139 (n.s.)
Head Circumference (cm)
L
P=0.5753 (n.s.)
Baby Length (cm)
M
P=0.1856 (n.s.)
APGAR Score 5 min


A
Source of Variance (Before)
F statistics
2.5
2.0
1.5
1.0
0.5
0.0
Sample_Group
Sex
Mat_Age
Mat_Ethnicity
Net_Weight_Gain
Pat_Ethnicity
Gravidity
Gestational_Age
Parity
Hemoglobin
B
Source of Variance (After)
F statistics
Sample_Group
Sex
Mat_Ethnicity
Pat_Ethnicity
Net_Weight_Gain
Gestational_Age
Gravidity
Hemoglobin
Parity
Mat_Age
C
Differential Methylation Probes (Obesed vs Control)
COBRE adj GestWk (Hyper = 10254, Hypo = 5394, NonSig = 395117)
-log10(adj.P.Val)
log2FC
D
All Illumina450K
All DM Sites
Hyper Sites
Hypo Sites
30.9%
23.1%
9.7%
36.3%
38.7%
24.5%
7.4%
29.4%
41.3%
27.8%
6.3%
24.7%
33.6%
18%
9.7%
38.7%
Isle
CpG Island
Opensea
Shelf
Shore
E
All Illumina450K
All DM Sites
Hyper Sites
Hypo Sites
3.6%
4.7%
8.8%
10.8%
14.2%
33.3%
24.7%
2.8%
5.5%
10.5%
15.3%
16.8%
29.7%
19.4%
1.9%
6.1%
11.2%
18.1%
18.3%
26.3%
18%
4.5%
4.2%
9.1%
9.8%
14%
36.3%
22.3%
Region
1stExon
3' UTR
5' UTR
Body
IGR
TSS1500
TSS200

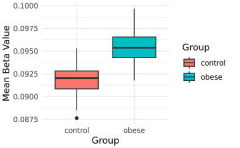
A. Hypermethyltion KEGG Pathways



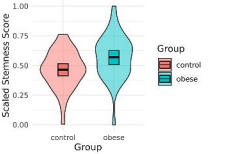
B. Immune Pathway Score (Wilcox p-val<1e-10)



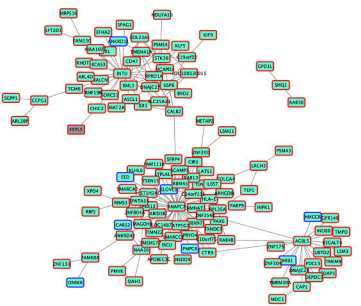
C. Protein Pathway Score (Wilcox p-val<1e-10)



D. Cell Entropy (Wilcox p-val: 0.0085)

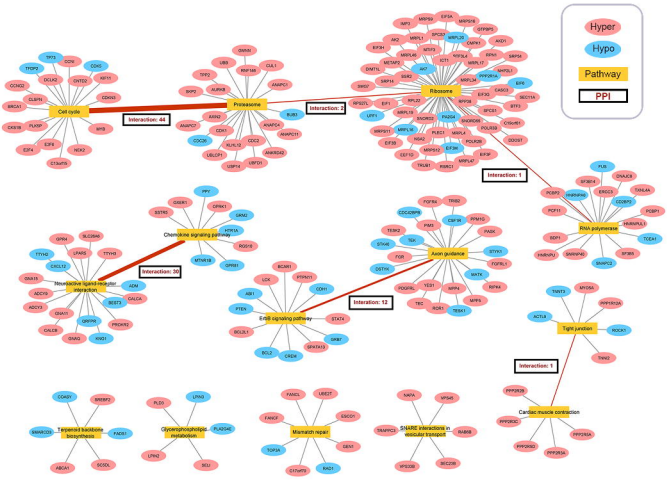


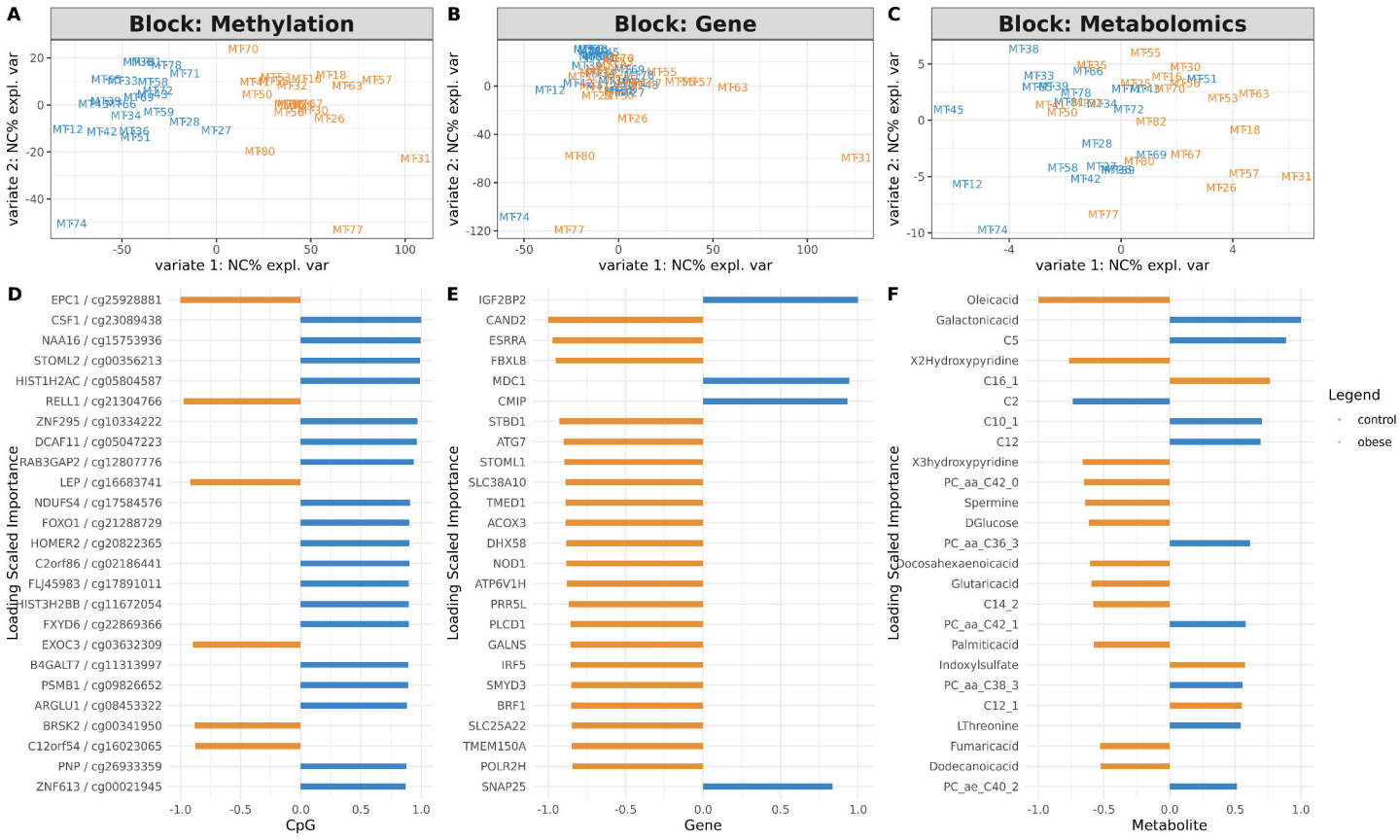
E. Module.turquoise
Correlation: 0.61 (p.value = 1e-08)

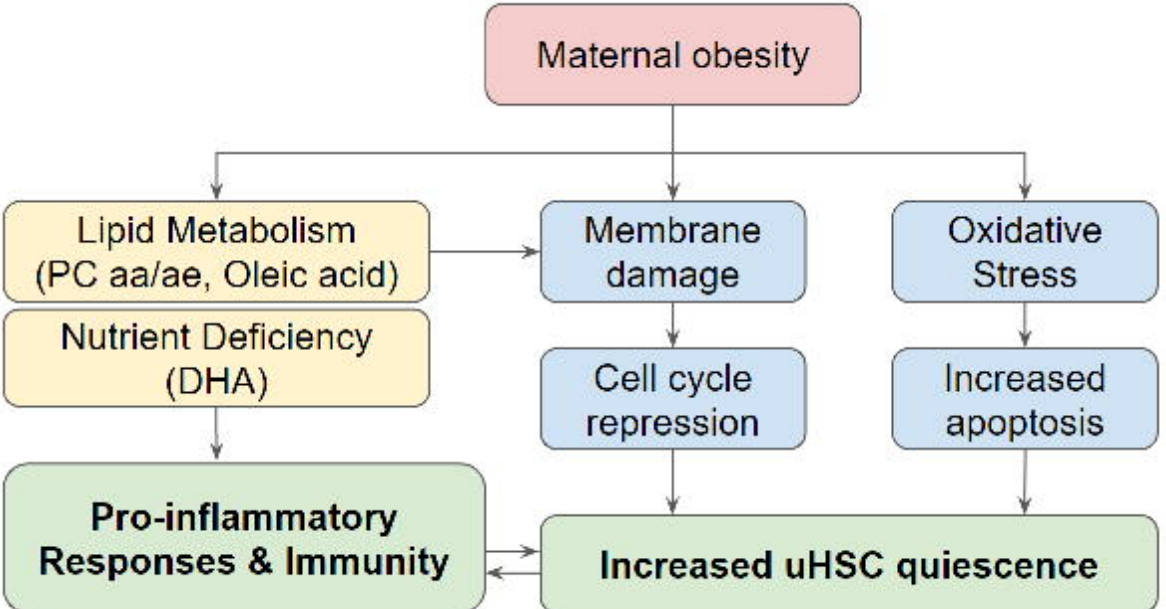


F. Module.brown
Correlation: 0.51 (p.value = 5e-06)

G. Protein-protein interaction networks




A
Block: Methylation
variate 2: NC% expl. var
variate 1: NC% expl. var
B
Block: Gene
variate 2: NC% expl. var
variate 1: NC% expl. var
C
Block: Metabolomics
variate 2: NC% expl. var
variate 1: NC% expl. var
D
Loading Scaled Importance
EPC1 / cg25928881
CSF1 / cg23089438
NAA16 / cg15753936
STOML2 / cg00356213
HIST1H2AC / cg05804587
RELL1 / cg21304766
ZNF295 / cg10334222
DCAF11 / cg05047223
RAB3GAP2 / cg12807776
LEP / cg16683741
NDUFS4 / cg17584576
FOXO1 / cg21288729
HOMER2 / cg20822365
C2orf86 / cg02186441
FLJ45983 / cg17891011
HIST3H2BB / cg11672054
FXYD6 / cg22869366
EXOC3 / cg03632309
B4GALT7 / cg11313997
PSMB1 / cg09826652
ARGLU1 / cg08453322
BRSK2 / cg00341950
C12orf54 / cg16023065
PNP / cg26933359
ZNF613 / cg00021945
CpG
E
Loading Scaled Importance
IGF2BP2
CAND2
ESRRA
FBXL8
MDC1
CMIP
STBD1
ATG7
STOML1
SLC38A10
TMED1
ACOX3
DHX58
NOD1
ATP6V1H
PRR5L
PLCD1
GALNS
IRF5
SMYD3
BRF1
SLC25A22
TMEM150A
POLR2H
SNAP25
Gene
F
Loading Scaled Importance
Oleicacid
Galactonicacid
C5
X2Hydroxypyridine
C16_1
C2
C10_1
C12
X3hydroxypyridine
PC_aa_C42_0
Spermine
DGlucose
PC_aa_C36_3
Docosahexaenoicacid
Glutaricacid
C14_2
PC_aa_C42_1
Palmiticacid
Indoxylsulfate
PC_aa_C38_3
C12_1
LThreonine
Fumaricacid
Dodecanoicacid
PC_ae_C40_2
Metabolite
Legend
control
obese


Maternal obesity
Lipid Metabolism
(PC aa/ae, Oleic acid)
Membrane
damage
Oxidative
Stress
Nutrient Deficiency
(DHA)
Cell cycle
repression
Increased
apoptosis
Pro-inflammatory
Responses & Immunity
Increased uHSC quiescence