HHS Public Access

Author manuscript

J Proteome Res. Author manuscript; available in PMC 2020 July 13.

Published in final edited form as:

J Proteome Res. 2020 April 03; 19(4): 1361–1374. doi:10.1021/acs.jproteome.9b00319.

Prepregnant Obesity of Mothers in a Multiethnic Cohort Is Associated with Cord Blood Metabolomic Changes in Offspring

Ryan J. Schlueter,
Department of Obstetrics and Gynecology, University of Hawaii, Honolulu, Hawaii 96826, United States

Fadhl M. Al-Akwaa,
Department of Computational Medicine and Bioinformatics, North Campus Research Complex, University of Michigan, Ann Arbor, Michigan 48105, United States

Paula A. Benny,
Department of Epidemiology, University of Hawaii Cancer Center, Honolulu, Hawaii 96813, United States

Alexandra Gurary,
John A. Burns School of Medicine, Department of Tropical Medicine, Medical Microbiology and Pharmacology, University of Hawaii, Honolulu, Hawaii 96813, United States

Guoxiang Xie,
Metabolomics Shared Resource, University of Hawaii Cancer Center, Honolulu, Hawaii 96813, United States

Wei Jia,
Metabolomics Shared Resource, University of Hawaii Cancer Center, Honolulu, Hawaii 96813, United States

Shaw J. Chun,
Department of Epidemiology, University of Hawaii Cancer Center, Honolulu, Hawaii 96813, United States

Ingrid Chern,

Corresponding Author Lana X. Garmire – Department of Computational Medicine and Bioinformatics, North Campus Research Complex, University of Michigan, Ann Arbor, Michigan 48105, United States; lgarmire@med.umich.edu.

Author Contributions

R.J.S. and F.M.A.-A. contributed equally to this work. L.X.G. envisioned the project, obtained funding, designed, and supervised the project and data analysis. R.J.S., I.C., P.A.B., and S.J.C. collected the samples. A.G. prepared the plasma samples. F.M.A.-A analyzed the data. G.X. performed the metabolomics experiments. R.J.S., F.M.A.-A., P.A.B., A.G., G.X., S.J.C., and L.X.G. wrote the manuscript. All authors have read, revised, and approved the manuscript.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jproteome.9b00319.

Unknown metabolites from batch #1 (XLSX)

Unknown metabolites from batch #2 (XLSX)

Unknown metabolites from batch #3 (XLSX)

Discrimination of obese and normal groups by the PLS method; selection of metabolites using elastic net regularization; accuracies of logistic regression models and important features selected by the clinical model; and power analysis and sample size estimation plot using 230 metabolites and 29 metabolites that were selected by the elastic net model (PDF)

The authors declare no competing financial interest.

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jproteome.9b00319

Department of Obstetrics and Gynecology, University of Hawaii, Honolulu, Hawaii 96826, United States

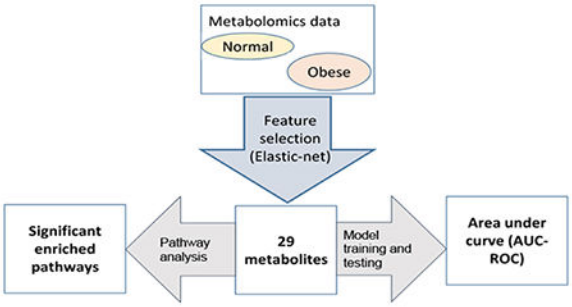
Lana X. Garmire

Department of Computational Medicine and Bioinformatics, North Campus Research Complex, University of Michigan, Ann Arbor, Michigan 48105, United States

Abstract

Maternal obesity has become a growing global health concern that may predispose the offspring to medical conditions later in life. However, the metabolic link between maternal prepregnant obesity and healthy offspring has not yet been fully elucidated. In this study, we conducted a case-control study using a coupled untargeted and targeted metabolomic approach from the newborn cord blood metabolomes associated with a matched maternal prepregnant obesity cohort of 28 cases and 29 controls. The subjects were recruited from multiethnic populations in Hawaii, including rarely reported Native Hawaiian and other Pacific Islanders (NHPI). We found that maternal obesity was the most important factor contributing to differences in cord blood metabolomics. Using an elastic net regularization-based logistic regression model, we identified 29 metabolites as potential early-life biomarkers manifesting intrauterine effect of maternal obesity, with accuracy as high as 0.947 after adjusting for clinical confounding (maternal and paternal age, ethnicity, parity, and gravidity). We validated the model results in a subsequent set of samples ($N = 30$) with an accuracy of 0.822. Among the metabolites, six metabolites (galactonic acid, butenylcarnitine, 2-hydroxy-3-methylbutyric acid, phosphatidylcholine diacyl C40:3, 1,5-anhydrosorbitol, and phosphatidylcholine acyl-alkyl 40:3) were individually and significantly different between the maternal obese and normal-weight groups. Interestingly, hydroxy-3-methylbutyric acid showed significantly higher levels in cord blood from the NHPI group compared to that from Asian and Caucasian groups. In summary, significant associations were observed between maternal prepregnant obesity and offspring metabolomic alternation at birth, revealing the intergenerational impact of maternal obesity.

Graphical Abstract



Keywords

metabolite; metabolomics; obesity; native Hawaiian; Polynesian; bioinformatics; analysis; mother

INTRODUCTION

Obesity is a global health concern. While some countries have a relative paucity of obesity, in the United States, obesity affects 38% of adults.[1,2] As such, maternal obesity has risen to epidemic proportions in recent years and can impose significant risk to both the mother and unborn fetus. By 2015, an estimated 25.6% women were obese before becoming pregnant according to the Centers for Disease Control and Prevention study.[3] Maternal prepregnant obesity can increase the risk for a wide range of health concerns for the baby and the mother during all stages of pregnancy. Moreover, research has recently extended the association of maternal obesity during pregnancy to the subsequent health of offspring such as diabetes or cardiovascular disease.[4] Since the inception of Barker's hypothesis in the 1990s, efforts to connect intrauterine exposures with the development of disease later in life have been the subject of many studies.[5,6] Both obesity and its accompanying morbidities, such as diabetes, cardiovascular diseases, and cancers, are of particular interest because considerable evidence has shown that maternal metabolic irregularities may have a role in genotypic programming in offspring.[7,8] Identifying markers of predisposition to health concerns or diseases would present an opportunity for early identification and potential intervention, thus providing lifelong benefits.[911]

Previous studies have found that infants born to obese mothers consistently demonstrate elevation of adiposity and are at more substantial risk for the development of metabolic disease.[12] While animal models have been used to demonstrate early molecular programming under the effect of obesity, human research to elucidate the underlying mechanisms in origins of childhood disease is lacking.[13] In *Drosophila melanogaster*, the offspring of females given a high-sucrose diet exhibited metabolic aberrations both at the larvae and adult developmental stages.[14–15] Though an invertebrate model, mammalian lipid and carbohydrate systems show high level of conservation in *D. melanogaster*.[16–17] In a mouse model of maternal obesity, progeny demonstrated significant elevations of both leptin and triglycerides when compared with the offspring of control mothers of normal weight.[7] The authors proposed that epigenetic modifications of obesogenic genes during intrauterine fetal growth play a role in adaption to an expected future environment. Recently, Aagaard-Tillery et al. used a primate model to examine the origins of metabolic disturbances and altered gene expression in offspring subjected to maternal obesity.[18] The offspring consistently displayed significant increases in triglyceride level and also fatty liver disease on histologic preparations. However, human studies that explore the fetal metabolic consequences of maternal obesity are still in need of investigation.

Metabolomics is the study of small molecules using high-throughput platforms, such as mass spectroscopy.[19] It is a desirable technology that can detect distinct chemical imprints in tissues and body fluids.[20] The field of metabolomics has shown great promise in various applications including early diagnostic marker identification,[21] where a set of metabolomic biomarkers can differentiate samples of two different states (e.g., disease and normal states). Cord blood metabolites provide information on fetal nutritional and metabolic health[22] and could provide an early window of detection to potential health issues among newborns. Previously, some studies have reported differential metabolite profiles associated with pregnancy outcomes such as intrauterine growth restriction[23] and low birth weight (LBW).[24]

For example, abnormal lipid metabolism and significant differences in relative amounts of amino acids (AA) were found in metabolomic signatures in cord blood from infants with intrauterine growth restriction in comparison with normal-weight infants.[23] In another study, higher phenylalanine and citrulline levels but lower glutamine, choline, alanine, proline, and glucose levels were observed in the cord blood of infants of LBW.[24] However, thus far, no metabolomic studies have been reported to specifically investigate the impact of maternal obesity on metabolomic profiles in fetal cord blood.[23–26]

This study aims to investigate metabolomic changes in fetal cord blood associated with obese (BMI > 30) and normal prepregnant weight (18.5 < BMI < 25) mothers. Uniquely, we recruited mothers from the multiethnic population in Hawaii, including Native Hawaiian and other Pacific Islanders (NHPI). NHPI is a particularly under-represented minority population across most scientific studies.

METHODS

Chemicals and Reagents

Ethanol, pyridine, methoxyamine hydrochloride, C8–C30 fatty acid methyl esters (FAMEs), and ammonium acetate were purchased from Sigma-Aldrich (St. Louis, MO). Liquid chromatography (LC)–mass spectrometry (MS) Optima grade methanol and acetonitrile, formic acid, *N*-methyl-*N*-trimethylsilyltrifluoroacetamide with 1% trimethylchlorosilane (TMCS), and hexane were obtained from Fisher Scientific (Fair Lawn, NJ). The ultrapure water was produced by the Millipore Advantage A10 system with a LC–MS Polisher filter (Billerica, MA). Analytical grade sodium hydroxide, sodium bicarbonate, and anhydrous sodium sulfate were obtained from JT Baker Co. (Phillipsburg, NJ, USA). AA and lipid standards were included in the AbsoluteIDQ p180 Kit (Biocrates Life Sciences, Austria). All other standards were commercially purchased from Sigma-Aldrich and Nu-Chek Prep (Elysian, MN, USA).

Study Population

We performed a multiethnic case-control study at the Kapiolani Medical Center for Women and Children, Honolulu, HI, from June 2015 through June 2017. The study was approved by the Western IRB board (WIRB Protocol 20151223). To avoid confounding of inflammation accompanying labor and natural births,[27] we recruited women scheduled for full-term cesarean section at ≥37 weeks gestation. All subjects fasted for at least 8 h before the scheduled cesarean delivery. Patients meeting inclusion criteria were identified from preadmission medical records with prepregnancy BMI ≥ 30.0 (cases) or 18.5–25.0 (controls). The prepregnancy BMIs were also confirmed during the enrollment. Women with preterm rupture of membranes, labor (being active contractions with cervical dilation), multiple gestations, pregestational diabetes, hypertensive disorders, cigarette smokers, human immunodeficiency virus, hepatitis B virus, and chronic drug users were excluded. Clinical characteristics were recorded, including maternal and paternal age, maternal and paternal ethnicities, mother's prepregnancy BMI, net weight gain, gestational age, parity, gravidity, and ethnicity. Neonate weight was recorded in kilograms, and the weight of the baby was taken directly after the birth in the newborn nursery. For the discovery cohort, a

total of 57 subjects (28 cases and 29 controls) were recruited. Additionally, to confirm the results, we recruited 30 subjects (12 cases and 18 controls) from the same site but different time intervals (July 2017 to June 2018).

Sample Collection, Preparation, and Quality Control

Cord blood was collected under sterile conditions at the time of cesarean section using the Pall Medical cord blood collection kit with 25 mL of citrate phosphate dextrose in the operating room. The umbilical cord was cleansed with chlorhexidine swab before collection to ensure sterility. The volume of the collected blood was measured and recorded before aliquoting to conicals for centrifugation. Conicals were centrifuged at 200*g* for 10 min, and plasma was collected. The plasma was centrifuged at 350*g* for 10 min, aliquoted into polypropylene cryotubes, and stored at −80 °C.

The investigators all took and passed courses where transport, collection, and laboratory use of biologic specimens was tested. During the handling of samples, gloves were used, and documentation for biohazard materials accompanied the transportation of materials. Laboratory-grade cryo-plasticware was used for storage, and all samples were labeled stringently and kept in 100-well boxes with record sheets in an Excel spreadsheet. Upon the use of samples for metabolome profiling, all samples were treated as biohazards. The investigators all received appropriate vaccinations, and personal protective equipment including gloves, lab coats, disposable face mask, and glasses for eye protection were used during sample preparation.

Metabolome Profiling

The plasma samples were thawed and extracted with 3 vol cold organic mixture of ethanol and chloroform and centrifuged at 4 °C at 14,500 rpm for 20 min. The supernatant was split for lipid and AA profiling using Acquity ultraperformance LC coupled to Xevo TQ-S mass spectrometry (UPLC–MS/MS, Waters Corp., Milford, MA). Metabolic profiling of other metabolites including organic acids, carbohydrates, AAs, and nucleotides was done using Agilent 7890A gas chromatography coupled to Leco Pegasus time-of-flight mass spectrometry (Leco Corp., St Joseph, MI). The raw data files generated from LC–MS (targeted) and GC-MS (untargeted) were processed with TargetLynx Application Manager (Waters Corp., Milford, MA) and ChromaTOF software (Leco Corp., St Joseph, MI), respectively. Peak signal, mass spectral data, and retention times were obtained for each metabolite. The detected metabolites from GC–MS were annotated and combined using an automated mass spectral data processing software package.[28] The levels of lipids and AAs detected from LC-MS were measured using the AbsoluteIDQ p180 Kit (Biocrates Life Sciences, Austria) commercially available. The reference standards of these measured lipids and AAs were integrated in the kit.[29] More details of metabolomic experiments and data preprocessing are described in the following subsections.

Sample Preparation for Metabolic Profiling

Plasma samples were prepared as previously described with minor modifications.[30] Each 150 μL of cold organic mixture (ethanol/chloroform = 3:1, v/v) is used to extract small-molecule metabolites from 50 μL of blood sample, spiked with two internal standard

solutions (10 μL of L-2-chlorophenyla-lanine in water, 0.3 mg/mL; 10 μL of heptadecanoic acid in methanol, 1 mg/mL). The sample extracts were centrifuged at 4 °C and 14,500 rpm for 20 min. The supernatant was split for lipid and AA profiling using Acquity ultraperformance liquid chromatography coupled to Xevo TQ-S mass spectrometry (UPLC–MS/MS, Waters Corp., Milford, MA) and for untargeted metabolic profiling using gas chromatography-time-of-flight mass spectrometry (GC–TOFMS).

Quantitation of AAs and Lipids with LC–MS/MS

For targeted metabolomic analyses of plasma samples, the AbsoluteIDQ p180 Kit (Biocrates Life Sciences, Austria) was used, which allows for the simultaneous quantification of metabolites from different compound classes [21 AA, 21 biogenic amines (BA), 40 acylcarnitines (AC), 38 acyl/acyl phosphatidylcholines (PC aa), 38 acyl/alkyl phosphatidylcholines (PC ae), 14 lyso-phosphatidylcholines (lysoPC), 15 sphingomyelins (SM), and the sum of hexoses (H1)]. The lipids, ACs, and the hexoses were determined by FIA-MS/MS, while the AAs and BAs were measured by LC–MS/MS. In brief, each aliquot of the 20 μL supernatant was added to a 96-well Biocrates Kit plate (Biocrates Life Sciences, Austria) for metabolite quantitation. After samples were dried under nitrogen, each 300 μL of extraction solvent (5 mM ammonium acetate in methanol) was added, and the kit plate was gently shaken at room temperature for 30 min. The extracts were derivatized with phenylisothiocyanate for AAs and BA quantification. The data were acquired using MassLynx 4.1 software (Waters) and were analyzed using TargetLynx applications manager version 4.1 (Waters) to obtain calibration equations and the quantitative concentration of each metabolite in the samples. Another aliquot of 20 μL of the extracts was further diluted with 380 μL of methanol with 5 mM ammonium acetate for FIA analysis of lipids. An Acquity ultraperformance liquid chromatography coupled to Xevo TQ-S mass spectrometry (UPLC–MS/MS, Waters Corp., Milford, MA) was used for targeted metabolite analysis of 140 lipids in cell-line samples. Each 10 μL of sample was directly injected into the mass spectrometer with elution solvent (methanol with 5 mM ammonium acetate) at a varied flow rate from 30 to 200 μL/min within 3 min.[31] Concentrations of lipids were directly calculated in MetIDQ (version 4.7.2, Biocrates).

Untargeted Metabolomic Profiling with GC–TOFMS

The protocol for untargeted metabolomic profiling was reported earlier.[32–33] Each aliquot of the 150 μL above supernatant was dried in a vacuum centrifuge concentrator. The dried material is derivatized with 50 μL of methoxyamine (15 mg/mL in pyridine) at 30 °C for 90 min. After the addition of 10 μL of alkynes (retention index standards) and 50 μL of BSTFA (1% TMCS), the mixture is incubated at 70 °C for 60 min. Retention indices of C8–C30 FAMEs were added for retention time correction. Each 1 μL sample was analyzed on an Agilent 7890A gas chromatograph coupled to a Leco Pegasus time-of-flight mass spectrometer (Leco Corp., St Joseph, MI) for global metabolite analysis. The analytes were introduced with a splitless mode to achieve maximum sensitivity and separated on an Rtx-5 MS capillary column (30 m × 0.25 mm i.d., 0.25 μm) (Restek, Bellefonte, PA). The column temperature was initially set to 80 °C for 2 min, increased to 300 °C in 12 min, and maintained at 300 °C for 5 min. The solvent delay was set to 4.4 min. The front inlet temperature, transfer line temperature, and source temperature were set to 260, 270, and 220

°C, respectively. The mass spectrometer was operated on a full scan mode from 50 to 500 at an acquisition rate of 20 spectra/s. To provide a set of data that can be used to assess overall reproducibility and to correct for potential analytical variations, a pooled plasma sample containing aliquots from all study subjects (or representative subjects depending on the number of samples to be tested) was used as a study QC. The QC samples for this project were prepared with the test samples and were injected at regular intervals (after every 10 test samples for GC–TOFMS and after every 12 test samples for UPLC–QTOFMS, respectively) to allow evaluating overall process variability and monitoring platform performance. The acceptance criterion for relative standard deviation is typically set to <20[34] or <30%.[35]

Compound Annotation and Marker Selection

All the features that pass the quality control are subject to compound annotation. This is performed using an in-house library containing ~1000 mammalian metabolites (reference standards). Commercial databases including NIST library 2011, LECO/Fiehn Metabolomics Library, and so forth are also used for compound annotation and verification. For the LC–TQMS data, the data were collected with multiple reaction monitor, and the cone and collision energy for each metabolite used the optimized settings from QuanOptimize application manager (Waters), and the metabolites were all annotated with reference standards. For the GC–TOFMS generated data, identification was processed by comparing the mass fragments and the retention time with our in-house library or the mass fragments with NIST 05 Standard Mass Spectral Databases in NIST MS search 2.0 (NIST, Gaithersburg, MD) software using a similarity of more than 70%.

Metabolomic Data Preprocessing

The raw LC–MS/MS data files were processed with TargetLynx Application Manager (Waters Corp., Milford, MA) to extract peak area and retention time of each metabolite. The raw GC–TOFMS data files were processed with ChromaTOF software (Leco Corp., St Joseph, MI) to extract peak signal and retention times for each metabolite. The detected metabolites were annotated with our internal metabolite database using an automated mass spectral data processing software package, ADAP-GC.[28]

Metabolomic Data Processing

Samples were received in three different batches. Batch # 1 (N = 36), #2 (N = 21), and #3 (N = 30) detected 93, 120, and 106 untargeted metabolites, respectively. A total of 79 untargeted metabolites were common in the discovery cohort (batches 1 and 2) and were combined with the 151 targeted metabolites, yielding 230 metabolites total in the training set. One hundred and fifty-one metabolites were identified from LC, and 79 metabolites were identified from GC. Power analysis was done using the module implemented in MetaboAnalystR[36] on both the whole metabolite list and the selected 29 metabolites. We conducted data preprocessing similar to the previous report.[37] Briefly, we used the K-nearest neighbors (KNN) method to impute missing (~8%) metabolomic data.[38] Using the KNN method, the metabolite missed value was predicted from the mean of the KNNs. To adjust for the offset between high- and low-intensity features and to reduce the heteroscedasticity, the logged value of each metabolite was centered by its mean and autoscaled by its standard deviation.[39] We used quantile normalization to reduce sample-to-sample variation.[40] We

applied partial least squares discriminant analysis (PLS-DA) to visualize how well metabolites could differentiate the obese from normal samples. To explore the contribution of different clinical/physiological factors to metabolomic data, we conducted source of variation analysis. We used ComBat Bioconductor R package[41] to adjust for the batch effects in the metabolomic data.

Classification Modeling and Evaluation

To reduce the dimensionality of our data (230 metabolites vs 57 samples), we selected the unique metabolites associated with separating obese and normal status. To achieve this, we used a penalized logistic regression method called elastic net that was implemented in the glmnet R package.[42] The elastic net method selects metabolites that have nonzero coefficients (*y*-axis, Figure S3C) at different values of λ (*x*-axis, Figure S3C), guided by two penalty parameters α and λ.[42] α sets the degree of mixing between lasso (when α =1) and the ridge regression (when α = 0). λ controls the shrunk rate of coefficients regardless of the value of α. When λ equals zero, no shrinkage is performed, and the algorithm selects all the features. As λ increases, the coefficients are shrunk more strongly, and the algorithm retrieves all features with nonzero coefficients. To find optimal parameters, we performed 10-fold cross-validation for feature selection that yield the smallest prediction minimum square error, similar to previous studies.[43]

We then used the metabolites selected by the elastic net to fit the regularized logistic regression model. Three parameters were tuned: cost, which controls the trade-off between regularization and correct classification, logistic loss, and epsilon, which sets the tolerance of termination criterion for optimization.

To construct and evaluate the model, we conducted crossvalidation (5×), similar to before.[43–45] We trained the model on four folds (80% of data) using leave-one-out cross-validation and measured model performance on the remaining folds (20% of data). We carried out the above training and testing five times on all fold combinations. We plotted the receiver-operating characteristic (ROC) curve for all fold predictions using pROC R package. To adjust confounding other clinical covariants such as ethnicity, gravidity, and parity, we reconstructed the metabolomic model mentioned above by including these factors.

Metabolite Importance Score

To rank the metabolites based on their contribution to the model performance, we used the model-based approach implemented in the Classification and Regression Training (CARET) R package.[46] The importance of each metabolite is evaluated individually using a “filter” approach.[47] The ROC curve analysis is conducted on each metabolite to predict the class using a series of cutoffs. The sensitivity and specificity are computed for each cutoff, and the ROC curve is computed. The trapezoidal rule is used to compute the area under the ROC curve. This area is used as the measure of variable importance. These scores were scaled to have a maximum of 100.

Metabolomic Pathway Analysis

We performed metabolomic pathway analysis on metabolites chosen by the elastic net method using Consensus Pathway DataBase (CPDB). We used r_{corr} function implemented in Hmisc R package to compute the correlations between clinical and metabolomic data.

Source of Variation Analysis

To estimate the relative contribution of the confounder factors such as maternal age and ethnicity to the variability of the metabolomic data, we built the ANOVA model using the metabolomics and cofounder factors; the resulting *F*-value is used to calculate *p*-values.

Data Availability

The metabolomic raw data files as well as processed data generated by this study have been deposited to metabolomics workbench repository (https://www.metabolomicsworkbench.org/., study ID ST001114). The list of unknown metabolites and their characteristics are included in the Supporting Information Tables S1–S3.

RESULTS

Cohort Subject Characteristics

Our discovery cohort consisted of 57 samples (29 normal and 28 obese subjects) from different ethnic groups. It consisted of three ethnic groups: Caucasian, Asian, and NHPI. Women undergoing scheduled cesarean delivery were included based on the previously described inclusion and exclusion criteria (Methods). Demographical and clinical characteristics in obese and control groups are summarized in Table 1. In the Caucasian group (10 mothers), 6 were categorized as nonobese and 4 as obese. In the Asian group (23 mothers), 16 were categorized as nonobese and 7 as obese. In the NHPI group (24 mothers), 7 (24%) were categorized as nonobese and 17 (61%) as obese. The variation in recruitment of cases versus controls in each ethnic background reflects the demographics in Hawaii. Compared to the mothers of normal prepregnant BMI, obese mothers had significantly higher prepregnancy BMI (33.51 ± 4.49 vs 21.89 ± 1.86 kg/m^2, $p = 9.18 \times 10^{-11}$). Mothers had no statistical difference regarding their ages (32.10 ± 4.88 vs 32.48 ± 5.66, $p = 0.7$) or gestational age (39.04 weeks ±0.22 vs 38.93 ± 0.45 $p = 0.38$), excluding the possibility of confounding from these factors. Babies of obese mothers had significantly ($P = 0.03$) higher birth weight averages in comparison with the normal prepregnant weight group, consistent with earlier observations.[48–49].

Preliminary Assessment of Metabolomic Results

Our discovery cohort has a total of 230 metabolites, including 79 untargeted and 151 targeted metabolites. To explore which clinical or physiological covariates were associated with the variations in the metabolomics, we conducted a source of variation analysis using a linear mixed model that includes multiple clinical variables. Figure 1A shows the F value for each factor in the ANOVA model relative to the error. The Y-axis is the *F* value and the *X*-axis is the factors specified in the ANOVA model and the error term. Indeed, maternal obesity was predominantly the most important factor contributing to metabolomic

differences rather than the other factors (Figure 1A). To test if these metabolites allow a clear separation between the obese and normal-weight subjects, we used elastic net regularization-based logistic regression, rather than the PLS-DA model, a routine supervised multivariate method which only yielded modest accuracy AUC = 0.62 (Figure S2B). PLS-DA has high probability to develop models that fit the training data well; however, it produced poor predictive performance on the validation set.[50]

On the other hand, elastic net offers potential improvements over PLS-DA because of the presence of the constraint, which promotes sparse solutions. Moreover, elastic net regularization overcomes the limitation of either ridge and lasso regularization alone and combines their strengths to identify optimized set metabolites.[25] Using the optimized regularization parameters (Figure S3), we identified a total of 29 metabolite features (Table 2), which together yields the highest predictive performance with AUC = 0.97, 95% CI = [0.904-0.986] in 20% hold-out test dataset (Figure 1B). The 29 selected metabolites by the elastic net collectively yields a statistical power of 0.93, much improved from the initial power of 0.67 estimated from the total 230 metabolites (Figure S4), and supports the validity of the statistical modeling approach. Among them, six metabolites have large contributions to the separation between the case and controls, with an importance score of at least 70%, individually (Figure 1C). These are galactonic acid, butenylcarnitine (C4:1), 2-hydroxy-3-methylbutyric acid, phosphatidylcholine diacyl C40:3 (PC aa C40:3), 1,5-anhydrosorbitol, and phosphatidylcholine acyl-alkyl 40:3 (PC ae C40:3). Thus, metabolites selected by the elastic net method indeed improved the prediction power of the model overall.

Calibrated Maternal-Obese Predictive Model with Consideration of Confounding

For statistical rigor, it is important to consider possible clinical confounders (if available), such as maternal and paternal ethnicity and parity (Table 1) that we collected for additional calibration. Toward this, we conducted two investigations. First, we explored the correlations among the demographic factors and metabolomic data. It is evident that several metabolites are correlated with maternal and paternal ethnicity, gravidity, and/or parity (Figure 2A). For example, maternal ethnicity is positively correlated with 2-hydroxy-3-methylbutyric acid. Second, we built a logistic regression model using the abovementioned four covariates alone (parity, gravidity, and maternal and paternal ethnicity). This model yields a modest AUC of 0.701 95% CI = [0.55–0.82] (Figure S4A), again suggesting the existence of confounding. These observations prompted us to recalibrate the 29-metabolite elastic net model by adjusting the metabolomic model using maternal and parental age and ethnicity, gravidity, and parity (Figure 2B).

The resulting modified model remains to have very high accuracy, with AUC = 0.947, 95% CI = [0.88–0.98]. In the new model, besides the original six metabolite features, maternal ethnicity and paternal ethnicity also have importance scores greater than 70% (Figure 2C).

Metabolite Features and Their Pathway Enrichment Analysis

The 29 metabolite features selected by the model belong to AC, glycerophospholipid, AAs, and organic acids classes. Their log-fold changes ranged from −0.66 (2-hydroxy-3-methylbutyric acid) to −0.45 (hydroxyhexadecenoylcarnitine or C16:1-OH) (Figure 3A and

Table 2). Among them, 15 metabolites were higher in obese-associated cord blood samples, including 2-hydroxy-3-methylbutyric acid, galactonic acid, PC ae C40:3, propionylcarnitine (C3), PC aa C40:3, *o*-butanoyl-carnitine (C4:1), hexanoylcarnitine (C6 (C4:1-DC)), phosphatidylcholine diacyl C40:2 (PC aa C40:2), benzoic acid, 1,5-anhydrosorbitol, isovalerylcarnitine (C5), PC ae C40:2, and L-arabitol, octadecenoylcarnitine (C18:1) (Figure 3A, Table 2). The remaining 14 metabolites are lower in obese-associated cord blood samples: malic acid, aspartate, citric acid, PC ae C34:0, isoleucine, PC ae C36:2, oleic acid, PC aa C36:5, PC ae C34:3, PC ae C40:6, C5:1-DC, 2-hydroxybutyric acid, myoinositol, and C16:1-OH (Figure 3A, Table 2). The individual metabolite levels of hexanoylcarnitine (C6(C4:1-DC)), *o*-butanoyl-carnitine (C4:1), PC aa C40:3, propionylcarnitine (C3), PC ae C40:3, galactonic acid, and 2-hydroxy-3-methylbutyric acid increased significantly in obese cases ($p < 0.05$, *t*-test).

To elucidate the biological processes in newborns that may be affected by maternal obesity, we performed pathway enrichment analysis on the 29 metabolite features using the Consensus pathway database (CPDB) tool.[51] We combined multiple pathway databases including KEGG, WikiPathways, Reactome, EHNM, and SMPDB. A list of the enriched pathways is plotted with adjusted *p*-value $q < 0.05$ (Figure 3B). We removed large pathways such as SLC-mediated transmembrane transport for nonspecificity. Among the filtered pathways, alanine and aspartate metabolism is the most significantly enriched pathway ($q = 0.004$). One should note that given the small number of identified metabolites in the dataset, the pathway analysis may have limited reliability.

Influence of Ethinicity on Metabolite Levels

In general, the level of 2-hydroxy-3-methylbutyric acid in obese subjects is higher than that in the normal-weighted subjects (Figure 4A). Our earlier correlational analysis suggested that maternal ethnicity may be correlated with 2-hydroxy-3-methylbutyric acid level (Figure 2A). To confirm this, we conducted two-way ANOVA statistical tests and indeed obtained significant *p*-value ($P = 0.023$, χ^2 test). We thus stratified the levels of 2-hydroxy-3-methylbutyric acid by ethnicity (Figure 4B). There was no significant difference in prepregnant normal-weight subjects across the three ethnic groups (Figure 4B). However, in cord blood samples associated with obese mothers, the concentration of 2-hydroxy-3-methylbutyric acid was much higher in NHPI, as compared to that in Caucasians ($p = 0.05$) or Asians ($p = 0.04$) (Figure 4B). 2-Hydroxy-3-methylbutyric acid originates mainly from ketogenesis through the metabolism of valine, leucine, and isoleucine.[52] Because all subjects fasted 8 h before the C-section, we expect that the confounding from diets is minimized among the three ethnic groups. Thus, the higher 2-hydroxy-3-methylbutyric acid level may indicate the higher efficiency of ketogenesis in babies born from obese NHPI mothers.

Validation on an Independent Cohort

We subsequently collected a new set of 30 patients (18 normal-weight and 12 obese). We decided to treat this set as “validation cohort”, following the convention of machine-learning dataset design, as samples were processed in different times/batches. We aimed to test if the previous model built on the 57 samples is predictive, given the modest size and heterogeneity among samples. We then performed new metabolomic measurements and

processed the data as described earlier. The model built on 57 samples yields an AUC of 0.822 (95% CI = [0.74–0.89], Figure 5A) in the new set of 30 samples, confirming the reproducibility of our findings. Moreover, we observed a similar trend of higher concentration of 2-hydroxy-3-methylbutyric acid in the obese subjects compared to that in the normal-weighted subjects (Figure 5B). Importantly, the levels of 2-hydroxy-3-methylbutyric acid have a similar trend in NHPI compared to Asians and Caucasians (p = 0.001) in the obese group, whereas no statistical difference between ethnicities exists in the control group (Figure 5C). Moreover, within this cohort, four of the six metabolites that had large contributions to the separations between case/control (importance score >70%) in the discovery cohort had consistent trend of changes in the validation cohort.

DISCUSSION

This study aims to distinguish key cord blood metabolites associated with maternal prepregnancy obesity. As maternal obesity is a health condition rather than a disease, we had to set stringent inclusion and exclusion criteria to exclude as many confounding factors as possible to ensure the quality of the metabolomic data. To avoid the sources of confounding from labor and vaginal delivery (diets, multiple operators due to unpredictable delivery time, etc.), we only targeted mothers having elective C-sections. We also excluded obese mothers who had known complications during pregnancy, such as pregestational diabetes, smoking, and hypertension. To minimize confounding due to maternal diet, all subjects fasted 8 h before the cesarean section. These criteria helped to improve the quality of the samples and hence metabolomic data, albeit the size of the study is modest.

Such careful experimental design did yield good data quality, as the source of variation analysis did show that maternal obesity is the only dominate factor contributing to metabolomic difference in the cord blood. Additionally, we conducted rigorous statistical modeling and found that metabolites can distinguish the two maternal groups with accuracy as high as AUC = 0.97 under cross-validation (or 0.947 after adjusting for confounding effects). Among all metabolites and physiological/demographic features selected by the combined model, galactonic acid has the largest impact on the model performance (importance score = 86%). Galactonic acid is a sugar acid and breakdown product of galactose. When present in sufficiently high levels, galactonic acid can act as an acidogen and a metabotoxin, which has multiple adverse effects on many organ systems. Galactonic acid was previously shown to be associated with diabetes in a mouse model because of a proposed mechanism of oxidative stress.[53] On the other hand, maternal ethnicity has the largest impact among physiological factors (importance score = 84%).

Few cord blood metabolomic studies have been carried out to associate with maternal obesity directly or birth weight.[24,54,55] In a recent Hyperglycemia and Adverse Pregnancy Outcome (HAPO) Study, Lowe et al. reported that branched-chain AAs and their metabolites, such as valine, phenylalanine, leucine/isoleucine, and AC C4, AC C3, AC C5 are associated with maternal BMI in a meta-analysis over four large cohorts (400 subjects in each).[55] In another study to associate cord blood metabolomics with LBW, Ivorra et al. found that newborns of LBW (birth weight <10th percentile, n = 20) had higher levels of phenylalanine and citrulline, compared to the control newborns (birth weight between the

75th and 90th percentiles, $n = 30$).[24] They also found lower levels of choline, proline, glutamine, alanine, and glucose in newborns of LBW; however, there were no significant differences between the mothers of the two groups. In our study, isoleucine is also identified as one of the 29 metabolite features related to maternal obesity; although alanine itself is not selected by the model to be a maternal obesity biomarker in cord blood, we did find that alanine and aspartate metabolism are enriched in the cord blood samples associated with the maternal obesity group.

Metabolomics pathway analysis on the metabolite features in the model identified six filtered significant pathways (Figure 3A). Among them, alanine and aspartate metabolism was previously reported to be associated with obesity.[56,57] Aspartate and alanine cycling has known association with insulin resistance and metabolic-related diseases, such as cancer.[45,58] Alanine, a highly gluconeogenic AA, contributes to the development of glucose intolerance in obesity, as circulating alanine levels are elevated in obese mothers. Our study also demonstrates that in infants of obese mothers this pathway is also enriched. Additionally, glycolysis is the metabolic pathway that converts glucose into pyruvate, while gluconeogenesis is the reverse generating glucose from noncarbohydrate carbon substrates. The offspring of obese but not normal-weight mothers in another study demonstrated the downregulation of the glycolysis pathway ($p = 0.049$).[59] Recent research showed that increase in hepatic gluconeo-genesis was a major source of the total maternal glucose used by the fetus.[60] Interestingly, 1,5-anhydrosorbitol, which has been shown to be a maternal marker of short-term glycemic control, was observed in our cord blood study as a marker too, likely from maternal origin. Thus, the changes in glycolysis and gluconeogenesis may suggest that obese mothers have greater glucose metabolism compared to normal controls. Phosphatidylinositol (PI) metabolism is a key regulator for energy metabolism. We found elevated levels of lipids such as PC aa 40:3 and PC ae 40:3 in obese subjects, in concert with this pathway. Altogether, the cord blood in babies of obese mothers demonstrates pathways enriched in metabolic syndrome and obesity, even though the phenotypic differences (obesity) do not exist in the babies but only in mothers.

Notably, our study has identified five metabolites which are previously not reported in the literature in association with obesity or maternal obesity: galactonic acid, L-arabitol, indoxyl sulfate, 2-hydroxy-3-methylbutyric acid, and citric acid. Except citric acid, all the other four metabolites are increased in obese-associated cord blood samples. 2-Hydroxy-3-methylbutyric acid concentrations varied by ethnicity but only in babies born from obese prepregnant mothers. 2-Hydroxy-3-methylbutyric acid is known to accumulate in high levels during ketoacidosis and fatty acid breakdown. Therefore, the higher elevation of 2-hydroxy-3-methylbutyric acid is likely due to increased cellular ketoacidosis and fatty acid breakdown in newborns from obese prepregnant mothers. To the best of our knowledge, this is the first study that shows differences in the 2-hydroxy-3-methylbutyric acid concentration levels among different ethnicities. Additionally, indoxyl sulfate is a metabolite of the AA tryptophan. As tryptophan is commonly found in fatty food, red meat, and cheese, it is possible that high levels of indoxyl sulfate detected in the cord blood associated with obese prepregnant mothers could be due to the maternal high fat diet. Oppositely, citric acid, a compound associated with the citric acid cycle,[61] is decreased in the cord blood associated with obese prepregnant mothers. This could be related to the lower vegetable and fruit

consumptions among obese prepregnant mothers. In all, the data suggest that maternal obesity may impact offspring cord blood metabolites. Further research into the specific mode of action of these metabolites would be beneficial in understanding its association with maternal obesity.

This study may benefit from improvement in the future follow-up. We determined the ethnicity of the subjects by selfreporting rather than genotyping because of the restriction of the currently approved IRB protocol. Additionally, there has been debate on the use of BMI as an indicator of obesity;[62] more direct measures of body fat could be considered, such as skin-fold thickness measurements, bioelectrical impedance, and energy X-ray absorptiometry.[63,64] Moreover, dietary and exposomic data will be very interesting to study in a followup large-scale cohort with IRB approval. Nevertheless, this study has established relationships between cord blood metabolomics and maternal prepregnant obesity, which in turn is associated with socioeconomic disparities.

CONCLUSIONS

In this study, we identified 29 cord blood metabolites that are associated with maternal obesity with high accuracy in a discovery set of 57 samples and a validation set of 30 samples. These metabolites may have the potential to be maternal obesity-related biomarkers in newborns.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

We thank Drs. Joseph Kaholokula and Alika Maunakea from the Native Hawaiian Health Department of University of Hawaii for giving suggestions. We acknowledge the services provided by the Molecular and Cellular Immunology Core which is funded in part by P30GM114737 from the Centers of Biomedical Research Excellence (COBRE) program of the National Institute of General Medical Sciences, a component of the National Institutes of Health. Dr. Lana Garmire's research is supported by grants K01ES025434 awarded by NIEHS through funds provided by the trans-NIH Big Data to Knowledge (BD2K) initiative (http://datascience.nih.gov/bd2k), P20 COBRE GM103457 awarded by NIH/NIGMS, R01 LM012373 awarded by NLM, and R01 HD084633 awarded by NICHD to L.X.G. Funding was also provided in part by the Department of Obstetrics and Gynecology, University of Hawaii. The metabolomic services were provided by the UH Cancer Center Metabolomics Shared Resource.

REFERENCES

(1). Flegal KM; Carroll MD; Kit BK; Ogden CL Prevalence of obesity and trends in the distribution of body mass index among us adults, 1999-2010. J. Am. Med. Assoc 2012, 307, 491–497.

(2). Ogden CL; Carroll MD; Kit BK; Flegal KM Prevalence of obesity in the United States, 2009-2010. NCHS Data Brief, 2012; (82), pp 1–8.

(3). Branum A; Kirmeyer SE; Gregory ECW. Prepregnancy Body Mass Index by Maternal Characteristics and State: Data from the Birth Certificate, 2014, 2016; pp 1–11.

(4). Ryckman K; Smith C Epigenetic and developmental influences on the risk of obesity, diabetes, and metabolic syndrome. Diabetes, Metab. Syndr. Obes.: Targets Ther 2015, 8, 295–302.

(5). Barker DJP Fetal origins of coronary heart disease. Br. Med. J 1995, 311, 171–174. [PubMed: 7613432]

(6). Benny PA; Al-Akwaa FM; Dirkx C; Schlueter RJ; Wolfgruber TK; Chern IY; Hoops S; Knights D; Garmire LX Placenta microbiome diversity is associated with maternal prepregnancy obesity and placenta biogeography. 2019, bioRxiv:659797.

(7). Li CCY; Young PE; Maloney CA; Eaton SA; Cowley MJ; Buckland ME; Preiss T; Henstridge DC; Cooney GJ; Febbraio MA; Martin DIK; Cropley JE; Suter CM Maternal obesity and diabetes induces latent metabolic defects and widespread epigenetic changes in isogenic mice. Epigenetics 2013, 8, 602–611. [PubMed: 23764993]

(8). Pi-Sunyer X The medical risks of obesity. Postgrad. Med 2009, 121, 21–33. [PubMed: 19940414]

(9). Isganaitis E; Rifas-Shiman SL; Oken E; Dreyfuss JM; Gall W; Gillman MW; Patti M-E Associations of cord blood metabolites with early childhood obesity risk. Int. J. Obes 2015, 39, 1041–1048.

(10). Rauschert S; Uhl O; Koletzko B; Hellmuth C Metabolomic biomarkers for obesity in humans: a short review. Ann. Nutr. Metab 2014, 64, 314–324. [PubMed: 25300275]

(11). Hivert MF; Perng W; Watkins SM; Newgard CS; Kenny LC; Kristal BS; Patti ME; Isganaitis E; DeMeo DL; Oken E; Gillman MW Metabolomics in the developmental origins of obesity and its cardiometabolic consequences. J. Dev. Origins Health Dis 2015, 6, 65–78.

(12). Heerwagen MJR; Miller MR; Barbour LA; Friedman JE Maternal obesity and fetal metabolic programming: a fertile epigenetic soil. Am. J. Physiol. Regul Integr. Comp. Physiol 2010, 299, R711–R722. [PubMed: 20631295]

(13). Rauschert S; Kirchberg FF; Marchioro L; Koletzko B; Hellmuth C; Uhl O Early Programming of Obesity Throughout the Life Course: A Metabolomics Perspective. Ann. Nutr. Metab 2017, 70, 201–209. [PubMed: 28301839]

(14). Brookheart RT; Duncan JG Drosophila melanogaster: An emerging model of transgenerational effects of maternal obesity. Mol. Cell. Endocrinol. 2016, 435, 20–28. [PubMed: 26687062]

(15). Brookheart RT; Swearingen AR; Collins CA; Cline LM; Duncan JG High-sucrose-induced maternal obesity disrupts ovarian function and decreases fertility in Drosophila melanogaster. Biochim. Biophys. Acta 2017, 1863, 1255–1263.

(16). Trinh I; Boulianne GL Modeling obesity and its associated disorders in Drosophila. Physiology 2013, 28, 117–124. [PubMed: 23455770]

(17). Diop SB; Bodmer R Drosophila as a model to study the genetic mechanisms of obesity-associated heart dysfunction. J. Cell Mol. Med 2012, 16, 966–971. [PubMed: 22303936]

(18). Aagaard-Tillery KM; Grove K; Bishop J; Ke X; Fu Q; McKnight R; Lane RH Developmental origins of disease and determinants of chromatin structure: maternal diet modifies the primate fetal epigenome. J. Mol. Endocrinol 2008, 41, 91–102. [PubMed: 18515302]

(19). Wishart DS; Tzur D; Knox C; Eisner R; Guo AC; Young N; Cheng D; Jewell K; Arndt D; Sawhney S; Fung C; Nikolai L; Lewis M; Coutouly M-A; Forsythe I; Tang P; Shrivastava S; Jeroncic K; Stothard P; Amegbey G; Block D; Hau DD; Wagner J; Miniaci J; Clements M; Gebremedhin M; Guo N; Zhang Y; Duggan GE; Macinnis GD; Weljie AM; Dowlatabadi R; Bamforth F; Clive D; Greiner R; Li L; Marrie T; Sykes BD; Vogel HJ; Querengesser L HMDB: the Human Metabolome Database. Nucleic Acids Res. 2007, 35, D521–D526. [PubMed: 17202168]

(20). Ciborowski M; Zbucka-Kretowska M; Bomba-Opon D; Wielgos M; Brawura-Biskupski-Samaha R; Pierzynski P; Szmitkowski M; Wolczynski S; Lipinska D; Citko A; Bauer W; Gorska M; Kretowski A Potential first trimester metabolomic biomarkers of abnormal birth weight in healthy pregnancies. Prenat. Diagn 2014, 34, 870–877. [PubMed: 24733416]

(21). Gowda GN; Zhang S; Gu H; Asiago V; Shanaiah N; Raftery D Metabolomics-based methods for early disease diagnostics. Expert Rev. Mol. Diagn 2008, 8, 617–633.

(22). Sanz-Cortés M; Carbajo RJ; Crispi F; Figueras F; Pineda-Lucena A; Gratacos E Metabolomic profile of umbilical cord blood plasma from early and late intrauterine growth restricted (IUGR) neonates with and without signs of brain vasodilation. PLoS One 2013, 8, No. e80121. [PubMed: 24312458]

(23). Favretto D; Cosmi E; Ragazzi E; Visentin S; Tucci M; Fais P; Cecchetto G; Zanardo V; Viel G; Ferrara SD Cord blood metabolomic profiling in intrauterine growth restriction. Anal. Bioanal. Chem 2012, 402, 1109–1121. [PubMed: 22101423]

(24). Ivorra C; Garcia-Vicent C; Chaves FJ; Monleón D; Morales JM; Lurbe E Metabolomic profiling in blood from umbilical cords of low birth weight newborns. J. Transl Med 2012, 10, 142. [PubMed: 22776444]

(25). La Torre D; Seppanen-Laakso T; Larsson HE; Hyotylainen T; Ivarsson SA; Lernmark A; Oresic M Decreased cord-blood phospholipids in young age-at-onset type 1 diabetes. Diabetes 2013, 62, 3951–3956. [PubMed: 23929934]

(26). Zhang A; Sun H; Wang X Power of metabolomics in biomarker discovery and mining mechanisms of obesity. Obes. Rev 2013, 14, 344–349. [PubMed: 23279162]

(27). Gomez-Lopez N; StLouis D; Lehr MA; Sanchez-Rodriguez EN; Arenas-Hernandez M Immune cells in term and preterm labor. Cell. Mol. Immunol 2014, 11, 571–581. [PubMed: 24954221]

(28). Ni Y; Qiu Y; Jiang W; Suttlemyre K; Su M; Zhang W; Jia W; Du X ADAP-GC 2.0: deconvolution of coeluting metabolites from GC/TOF-MS data for metabolomics studies. Anal. Chem 2012, 84, 6619–6629. [PubMed: 22747237]

(29). Sumner LW; Amberg A; Barrett D; Beale MH; Beger R; Daykin CA; Fan TW-M; Fiehn O; Goodacre R; Griffin JL; Hankemeier T; Hardy N; Harnly J; Higashi R; Kopka J; Lane AN; Lindon JC; Marriott P; Nicholls AW; Reily MD; Thaden JJ; Viant MR Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). Metabolomics 2007, 3, 211–221. [PubMed: 24039616]

(30). Qiu Y; Cai G; Su M; Chen T; Zheng X; Xu Y; Ni Y; Zhao A; Xu LX; Cai S; Jia W Serum metabolite profiling of human colorectal cancer using GC-TOFMS and UPLC-QTOFMS. J. Proteome Res 2009, 8, 4844–4850. [PubMed: 19678709]

(31). Qiu Y; Zhou B; Su M; Baxter S; Zheng X; Zhao X; Yen Y; Jia W Mass spectrometry-based quantitative metabolomics revealed a distinct lipid profile in breast cancer patients. Int. J. Mol. Sci 2013, 14, 8047–8061. [PubMed: 23584023]

(32). Qiu Y; Cai G; Zhou B; Li D; Zhao A; Xie G; Li H; Cai S; Xie D; Huang C; Ge W; Zhou Z; Xu LX; Jia W; Zheng S; Yen Y; Jia W A distinct metabolic signature of human colorectal cancer with prognostic potential. Clin. Canc. Res 2014, 20, 2136–2146.

(33). Nepal M; Ma C; Xie G; Jia W; Fei P Fanconi Anemia complementation group C protein in metabolic disorders. Aging 2018, 10, 1506–1522. [PubMed: 29930218]

(34). Dunn WB; Broadhurst D; Begley P; Zelena E; Francis-McIntyre S; Anderson N; Brown M; Knowles JD; Halsall A; Haselden JN; Nicholls AW; Wilson ID; Kell DB; Goodacre R The Human Serum Metabolome, C., Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. Nat. Protoc 2011, 6, 1060–1083. [PubMed: 21720319]

(35). Lewis MR; Pearce JTM; Spagou K; Green M; Dona AC; Yuen AHY; David M; Berry DJ; Chappell K; Horneffer-van der Sluis V; Shaw R; Lovestone S; Elliott P; Shockcor J; Lindon JC; Cloarec O; Takats Z; Holmes E; Nicholson JK Development and Application of Ultra-Performance Liquid Chromatography-TOF MS for Precision Large Scale Urinary Metabolic Phenotyping. Anal. Chem. 2016, 88, 9004–9013. [PubMed: 27479709]

(36). Chong J; Yamamoto M; Xia J MetaboAnalystR 2.0: From Raw Spectra to Biological Insights. Metabolites 2019, 9, 57.

(37). Alakwaa FM; Chaudhary K; Garmire LX Deep Learning Accurately Predicts Estrogen Receptor Status in Breast Cancer Metabolomics Data. J. Proteome Res 2018, 17, 337–347. [PubMed: 29110491]

(38). Beretta L; Santaniello A Nearest neighbor imputation algorithms: a critical evaluation. BMC Med. Inf. Decis. Making 2016, 16, 74.

(39). van den Berg RA; Hoefsloot HC; Westerhuis JA; Smilde AK; van der Werf MJ Centering, scaling, and transformations: improving the biological information content of metabolomics data. BMC Genom. 2006, 7, 142.

(40). Jauhiainen A; Madhu B; Narita M; Narita M; Griffiths J; Tavare S Normalization of metabolomics data with applications to correlation maps. Bioinformatics 2014, 30, 2155–2161. [PubMed: 24711654]

(41). Leek JT; Johnson WE; Parker HS; Jaffe AE; Storey JD The sva package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics 2012, 28, 882–883. [PubMed: 22257669]

(42). Paolino JP Rasch Model Parameter Estimation via the Elastic Net. J. Appl. Meas 2015, 16, 353–364. [PubMed: 26771565]

(43). AlAkwaa FM; Yunits B; Huang S; Alhajaji H; Garmire LX Lilikoi: an R package for personalized pathway-based classification modeling using metabolomics data. GigaScience 2018, 7, giy136.

(44). Alkawaa FM; Chaudhary K; Garmire LX Deep learning accurately predicts estrogen receptor status in breast cancer metabolomics data. J. Proteome Res 2018, 17, 337. [PubMed: 29110491]

(45). Huang S; Chong N; Lewis NE; Jia W; Xie G; Garmire LX Novel personalized pathway-based metabolomics models reveal key metabolic pathways for breast cancer diagnosis. Genome Med. 2016, 8, 34. [PubMed: 27036109]

(46). Kuhn M Building Predictive Models in R Using the caret Package. J. Stat. Software 2008, 28, 26.

(47). Gevrey M; Dimopoulos I; Lek S Review and comparison of methods to study the contribution of variables in artificial neural network models. Ecol Model. 2003, 160, 249–264.

(48). Briese V; Voigt M; Hermanussen M; Wittwer-Backofen U Morbid obesity: pregnancy risks, birth risks and status of the newborn. Homo 2010, 61, 64–72. [PubMed: 20042187]

(49). Shankar K; Harrell A; Liu X; Gilchrist JM; Ronis MJJ; Badger TM Maternal obesity at conception programs obesity in the offspring. Am. J. Physiol. Regul Integr. Comp. Physiol 2008, 294, R528–R538. [PubMed: 18032473]

(50). Bujak R; Daghir-Wojtkowiak E; Kaliszan R; Markuszewski MJ PLS-Based and Regularization-Based Methods for the Selection of Relevant Variables in Non-targeted Metabolomics Data. Front. Mol. Biosci 2016, 3, 35. [PubMed: 27508208]

(51). Kamburov A; Stelzl U; Lehrach H; Herwig R The ConsensusPathDB interaction database: 2013 update. Nucleic Acids Res. 2013, 41, D793–D800. [PubMed: 23143270]

(52). Liebich HM; Forst C Hydroxycarboxylic and oxocarboxylic acids in urine: products from branched-chain amino acid degradation and from ketogenesis. J. Chromatogr. B: Biomed. Sci. Appl 1984, 309, 225–242.

(53). Fahrmann J; Grapov D; Yang J; Hammock B; Fiehn O; Bell GI; Hara M Systemic alterations in the metabolome of diabetic NOD mice delineate increased oxidative stress accompanied by reduced inflammation and hypertriglyceremia. Am. J. Physiol. Endocrinol. Metab 2015, 308, E978–E989. [PubMed: 25852003]

(54). Ferreira DLS; Williams DM; Kangas AJ; Soininen P; Ala-Korpela M; Smith GD; Jarvelin M-R; Lawlor DA Association of pre-pregnancy body mass index with offspring metabolic profile: Analyses of 3 European prospective birth cohorts. PLoS Med. 2017, 14, No. e1002376. [PubMed: 28829768]

(55). Lowe WL Jr.; Bain JR; Nodzenski M; Reisetter AC; Muehlbauer MJ; Stevens RD; Ilkayeva OR; Lowe LP; Metzger BE; Newgard CB; Scholtens DM; Group, H. S. C. R. Maternal BMI and Glycemia Impact the Fetal Metabolome. Diabetes Care 2017, 40, 902–910. [PubMed: 28637888]

(56). Sookoian S; Pirola CJ Alanine and aspartate aminotransferase and glutamine-cycling pathway: their roles in pathogenesis of metabolic syndrome. World J. Gastroenterol 2012, 18, 3775–3781. [PubMed: 22876026]

(57). Cirulli ET; Guo L; Leon Swisher C; Shah N; Huang L; Napier LA; Kirkness EF; Spector TD; Caskey CT; Thorens B; Venter JC; Telenti A Profound Perturbation of the Metabolome in Obesity Is Associated with Health Risk. Cell Metab. 2019, 29, 488–500. [PubMed: 30318341]

(58). Xie G; Zhou B; Zhao A; Qiu Y; Zhao X; Garmire L; Shvetsov YB; Yu H; Yen Y; Jia W Lowered circulating aspartate is a metabolic feature of human breast cancer. Oncotarget 2015, 6, 33369–33381. [PubMed: 26452258]

(59). Baker PR 2nd; Patinkin Z; Shapiro ALB; De La Houssaye BA; Woontner M; Boyle KE; Vanderlinden L; Dabelea D; Friedman JE Maternal obesity and increased neonatal adiposity correspond with altered infant mesenchymal stem cell metabolism. JCI Insight 2017, 2, No. e94200.

(60). Maffeis C; Morandi A Effect of Maternal Obesity on Foetal Growth and Metabolic Health of the Offspring. Obes. Facts 2017, 10, 112–117. [PubMed: 28384625]

(61). Muroyama K; Murosaki S; Yamamotoi Y; Odaka H; Chung HC; Miyoshi M Anti-obesity effects of a mixture of thiamin, arginine, caffeine, and citric acid in non-insulin dependent diabetic KK mice. J. Nutr. Sci. Vitaminol 2003, 49, 56–63. [PubMed: 12882397]

(62). Freedman DS; Horlick M; Berenson GS A comparison of the Slaughter skinfold-thickness equations and BMI in predicting body fatness and cardiovascular disease risk factor levels in children. Am. J. Clin. Nutr 2013, 98, 1417–1424. [PubMed: 24153344]

(63). Wohlfahrt-Veje C; Tinggaard J; Winther K; Mouritsen A; Hagen CP; Mieritz MG; de Renzy-Martin KT; Boas M; Petersen JH; Main KM Body fat throughout childhood in 2647 healthy Danish children: agreement of BMI, waist circumference, skinfolds with dual X-ray absorptiometry. Eur. J. Clin. Nutr 2014, 68, 664–670. [PubMed: 24473457]

(64). Steinberger J; Jacobs DR; Raatz S; Moran A; Hong C-P; Sinaiko AR Comparison of body fatness measurements by BMI and skinfolds vs dual energy X-ray absorptiometry and their relation to cardiovascular risk factors in adolescents. Int. J. Obes 2005, 29, 1346–1352.

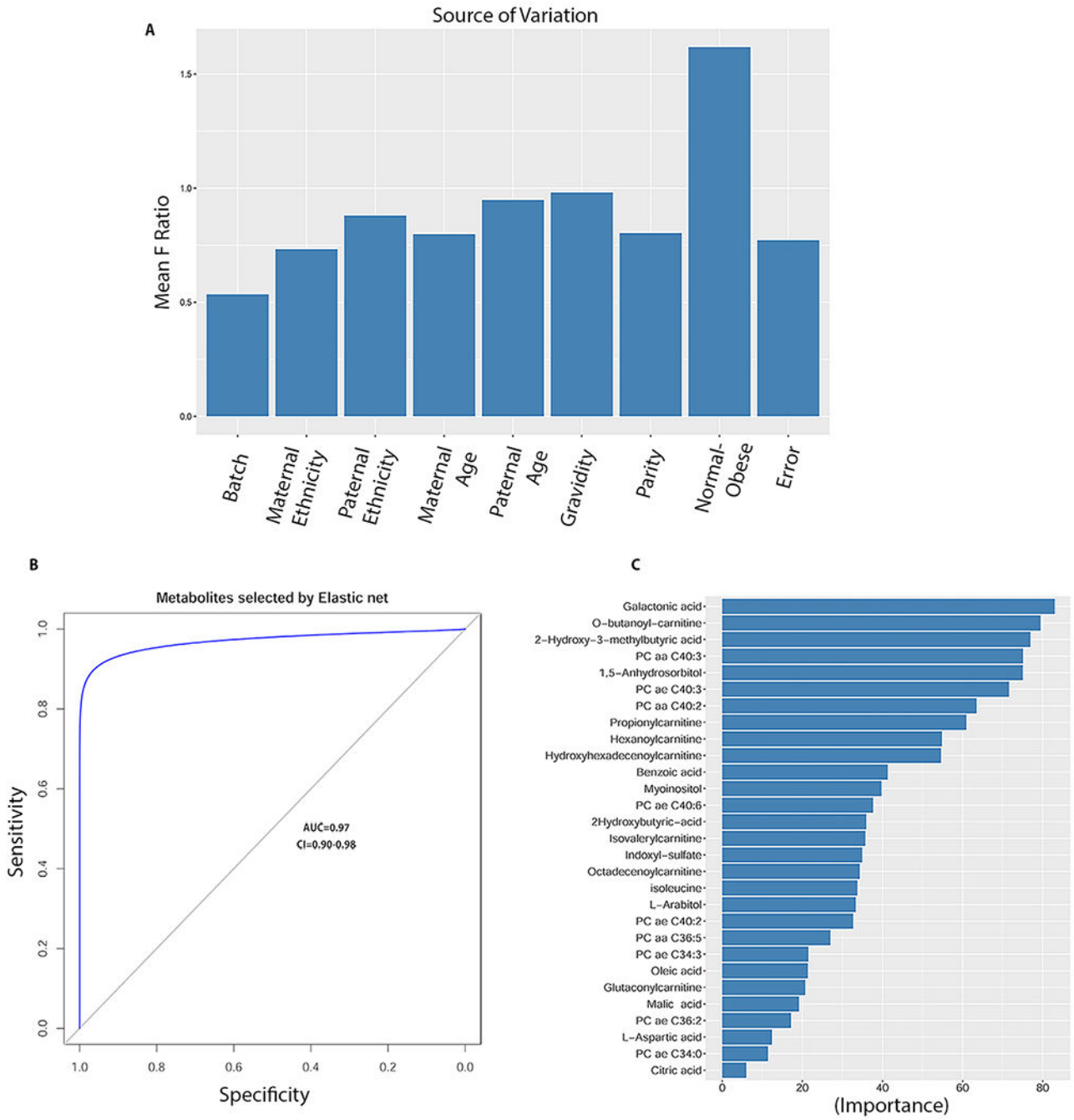


Figure 1.

Source of variation and accuracies of logistic regression models and important features selected by the metabolomic model. (A) ANOVA plot of clinical factors using the metabolite levels in cord blood samples. Averaged ANOVA *F*-statistics are calculated for potential confounding factors, including obesity, gravida, parity, paternal and maternal age, and ethnicity. (B) Model accuracy represented by classification ROCs. (C) Ranking of contributions (percentage) of selected metabolomic features in the model.

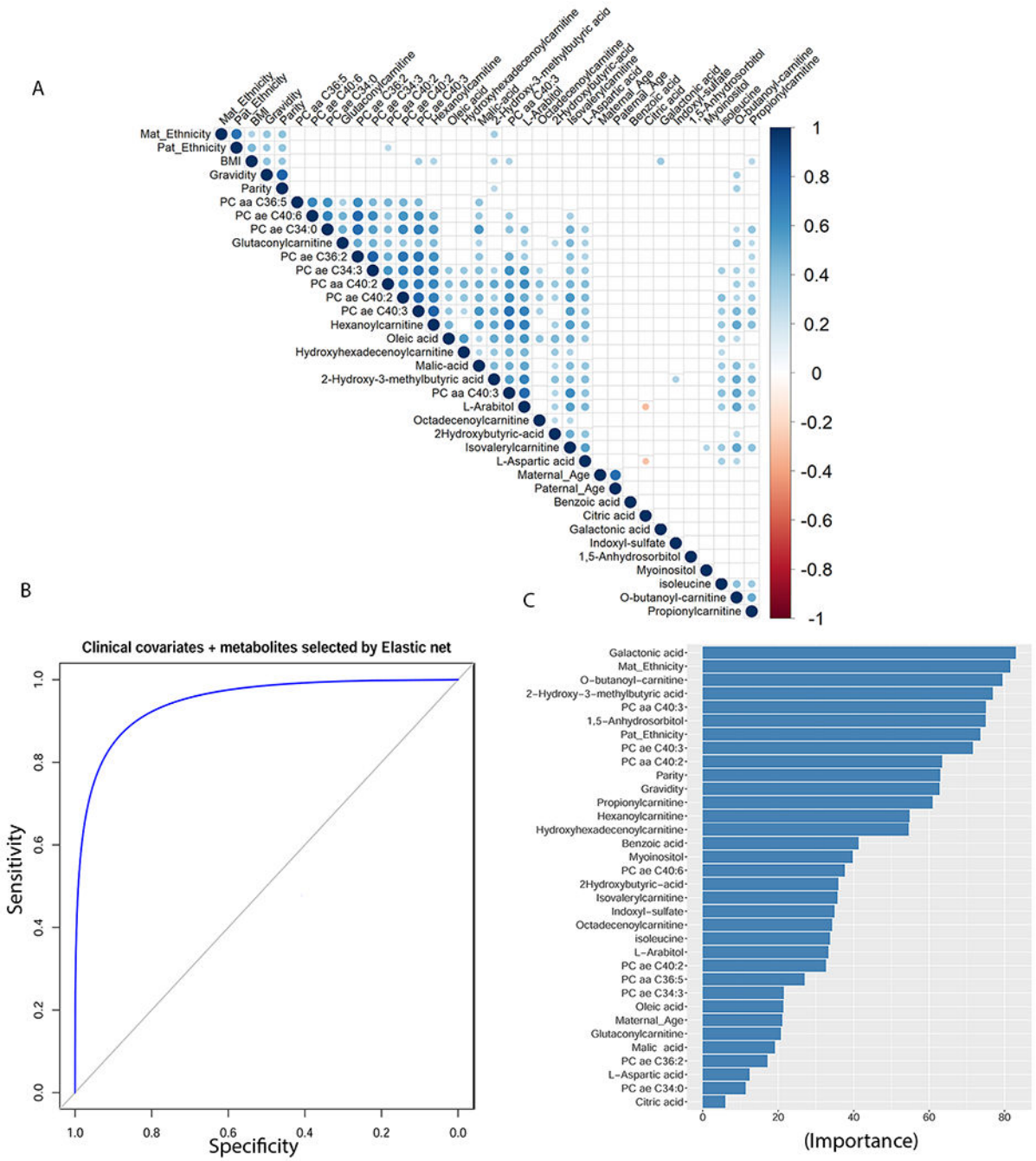


Figure 2.

(A) Correlation coefficients among demographical/physiological factors and the metabolomic data. The blue color indicates positive correlations, and the red color indicates negative correlations. (B) ROCs of the combined model with metabolomic and physiological/demographic data. (C) Ranking of contributions (percentage) of selected features in the model (B).

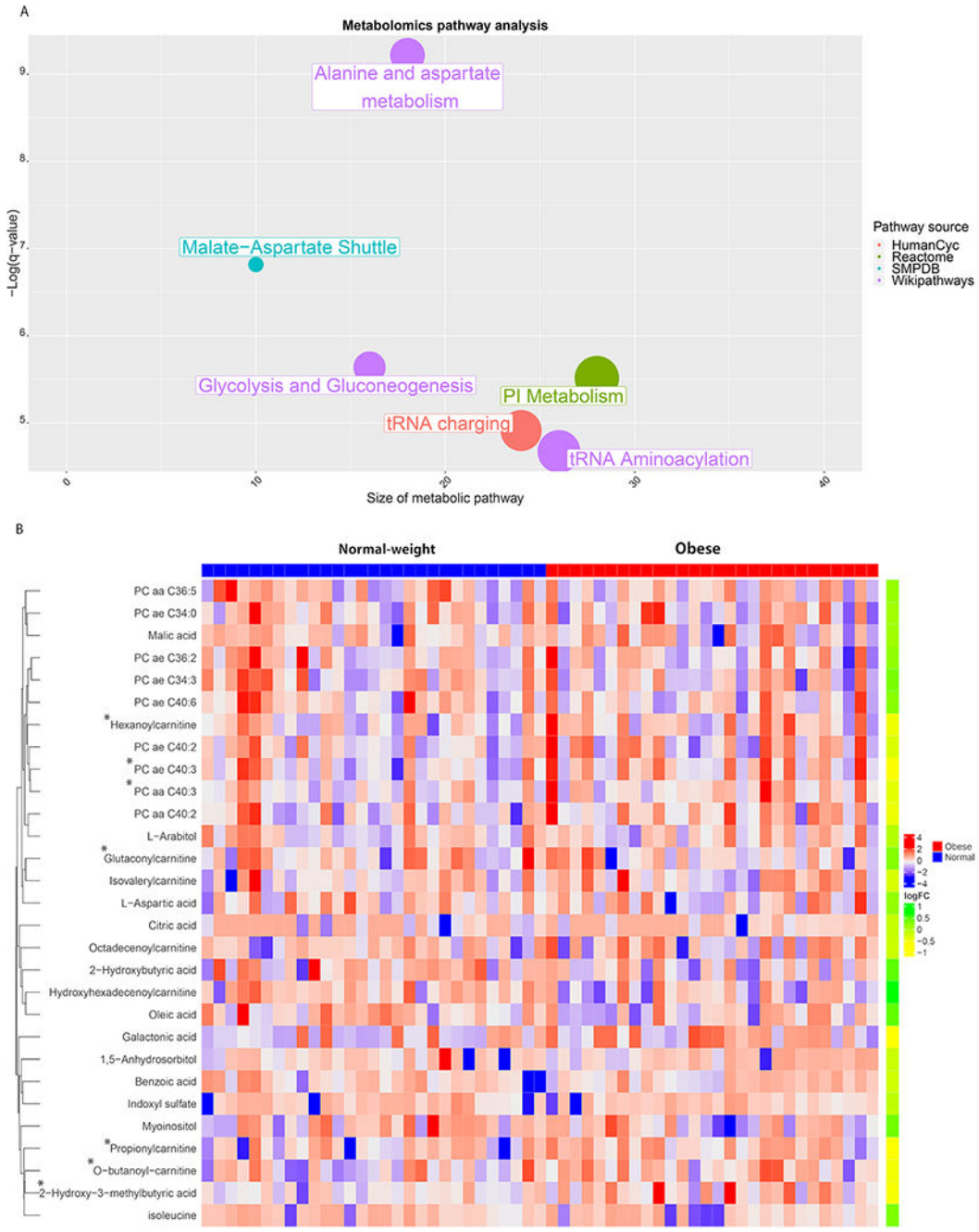


Figure 3.

Analysis of the 29 selected metabolites. (A) Heat map of selected metabolites separated by the maternal group. * indicates metabolites that show significant *p*-values ($P < 0.05$, *t*-test) individually. (B) Pathway analysis of the 29 metabolites. *X*-axis shows the size of the metabolomic pathway. *Y*-axis shows the adjusted *p*-value calculated from the CPDB tool. The size of the nodes represents the size of the metabolomic pathway (number of metabolites involved in each pathway). The color of the nodes represents the source of these pathways.

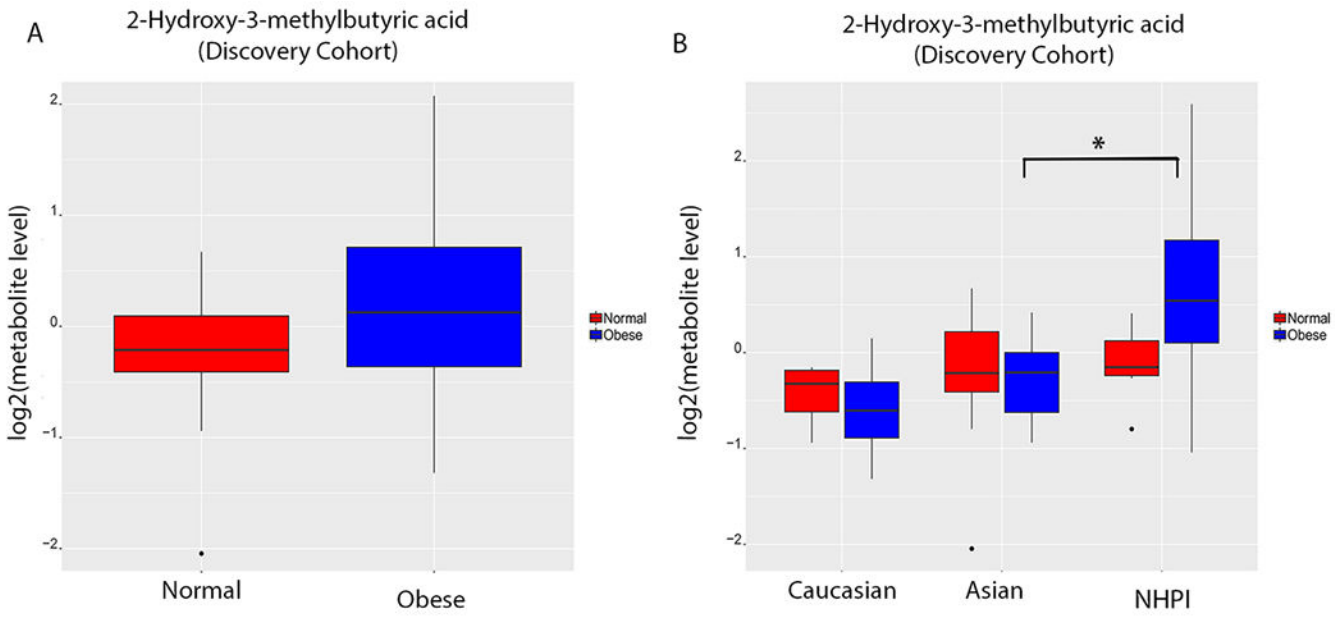


Figure 4.

Box plot of 2-hydroxy-3-methylbutyric acid in the discovery cohort, stratified by (A) normal (n = 29) and obese (n = 28) subjects and further by the (B) three ethnic groups: Asian, Caucasian, and NHPI.

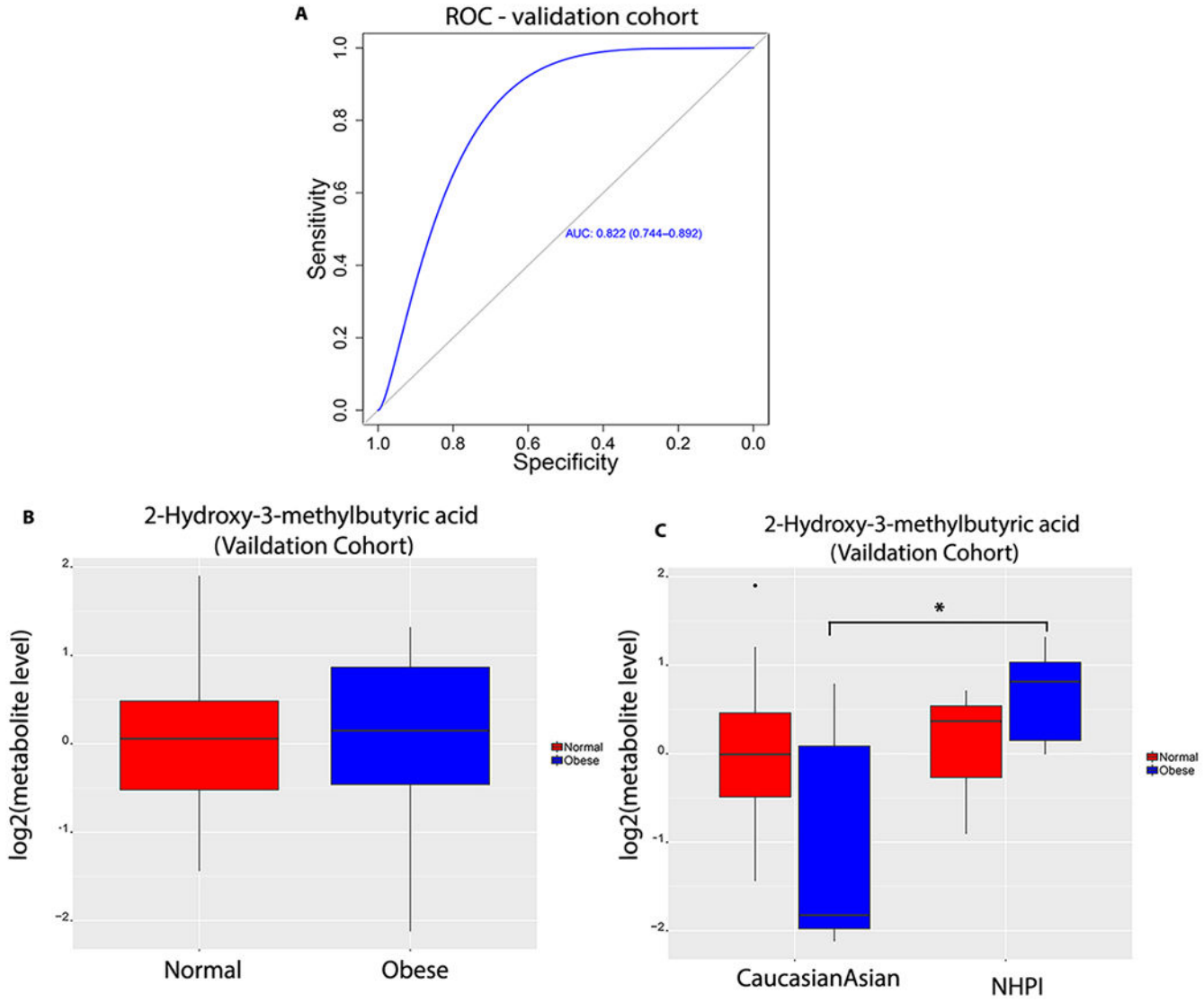


Figure 5.

Validation with a subsequent cohort ($n = 30$). (A) Accuracy on classifying cases vs controls in the validation cohort using the model built on the discovery cohort, as shown in Figure 2B. (B,C) Box plots of 2-hydroxy-3-methylbutyric acid, stratified by (B) normal ($n = 18$) and obese ($n = 12$) subjects and further by (C) ethnic groups of Asians/Caucasians vs NHPI. Asians ($n = 2$) and Caucasians ($n = 3$) were combined, as the number of patients of these ethnicities in the obese group is small in the obese group. *: statistically significant with p-value <0.05 (t-test).

Table 1.

Demographical and Clinical Characteristics in Obese and Control Groups

	mean (SD)				
	control (*n* = 29)	**case (*n* = 28)**	***P*-value[a]**	**confounder[b]**	**SOV[c]**
maternal age, years	32.48 (5.66)	32.10 (4.88)	0.78	yes	0.79
paternal age, years	34.68(7.14)	35.21(6.43)	0.79	yes	0.94
prepregnancy BMI, kg/m^2	21.89(1.86)	33.51(4.49)	1.12×10^{-14}	yes	1.61
gestational age, weeks	39.04(0.218)	38.93(0.45)	0.3812	not included	
net weight gain	30.85(10.92)	29.4(13.55)	0.7335		
baby weight (kg)	3.29(0.32)	3.54(0.5)	0.03		
head circle (cm)	34.89(1.10)	35.55(1.36)	0.05		
baby length (cm)	51.3(1.9)	51.4(2.36)	0.8		
parity			0.03	yes	0.80
0	5	2			
1	16	7			
2	7	10			
3 and above	1	9			
gravidity			0.12	yes	0.98
1	5	1			
2	12	5			
3	7	8			
4 and above	5	14			
maternal ethnicity			0.01	yes	0.73
Caucasian	6	4			
Asian	16	7			
Pacific island	7	17			
paternal ethnicity			0.03	yes	0.88
Caucasian	8	3			
Asian	14	9			
Pacific island	7	16			

[a]Categorical variables were compared using χ^2 test, whereas continuous variables were compared using *t*-test.

J Proteome Res. Author manuscript; available in PMC 2020 July 13.

[b] Yes if the factor is accounted for when building the classifier model.

[c] Source of variation of the factor. It represents the contribution of the factor to the variability of the metabolomics data. It is the F-stats that were taken from a linear regression ANOVA model comprising maternal and paternal age, ethnicity, parity, and gravidity (Figure 1A).

Table 2.

List of Metabolites Associated with Obese-Control Maternal Status, Selected by Elastic Net Regularization-Based Logistic Regression[a]

					fold change[b] (obese—normal-weight)	
ID	**metabolite biochemical name**	**abbreviation**	**class**	**PubChem ID**	**log FC**	***P*_value**
1	2-hydroxy-3-methylbutyric acid	2-hydroxy-3-methylbutyric acid	hydroxy fatty acids	99823	−0.66	0.01
2	galactonic acid	galactonic acid	medium-chain hydroxy acids	128869	−0.63	0.02
3	phosphatidylcholine acyl-alkyl C40:3	PC ae C40:3	glycerophospholipid	53481745	−0.62	0.02
4	propionylcarnitine	C3	AC	NA	−0.56	0.03
5	phosphatidylcholine diacyl C40:3	PC aa C40:3	glycerophospholipid	52922703	−0.56	0.03
6	*O*-butanoyl-carnitine, butenylcarnitine	C4:1	AC	NA	−0.56	0.03
7	hexanoylcarnitine, fumarylcarnitine	C6 (C4:1-DC)	AC	NA	−0.54	0.04
8	phosphatidylcholine diacyl C40:2	PC aa C40:2	glycerophospholipid	24779063	−0.48	0.07
9	benzoic acid	benzoic acid	organic compound (benzoic acids)	243	−0.45	0.08
10	hydroxyhexadecenoylcarnitine	C16:1-OH	AC	NA	0.45	0.09
11	myoinositol	myoinositol	cyclic polyalcohol	NA	0.39	0.14
12	1,5-anhydrosorbitol	1,5-anhydrosorbitol	monosaccharides	64960	−0.37	0.16
13	isovalerylcarnitine, valerylcarnitine, methylbutyrylcarnitine	C5	AC	NA	−0.37	0.16
14	phosphatidylcholine acyl-alkyl C40:2	PC ae C40:2	glycerophospholipid	53481739	−0.32	0.22
15	2-hydroxybutyric acid	2-hydroxybutyric acid	organic compounds (α-hydroxy acids)	11266	0.32	0.22
16	glutaconylcarnitine, mesaconylcarnitine	C5:1-DC	AC	NA	0.27	0.30
17	L-arabitol	L-arabitol	sugar alcohols	439255	−0.27	0.31
18	phosphatidylcholine acyl-alkyl C40:6	PC ae C40:6	glycerophospholipid	24779341	0.26	0.32
19	octadecenoylcarnitine	C18:1	AC	NA	−0.23	0.38
20	phosphatidylcholine acyl-alkyl C34:3	PC ae C34:3	glycerophospholipid	53481709	0.22	0.40
21	phosphatidylcholine diacyl C36:5	PC aa C36:5	glycerophospholipid	24778723	0.22	0.41
22	oleic acid	oleic acid	unsaturated fatty acid	445639	0.20	0.45
23	phosphatidylcholine acyl-alkyl C36:2	PC ae C36:2	glycerophospholipid	6443070	0.19	0.46
24	indoxyl sulfate	indoxyl sulfate	arylsulfates	10258	−0.18	0.49
25	isoleucine	isoleucine	AAs	6306	0.16	0.55

ID	metabolite biochemical name	abbreviation	class	PubChem ID	fold change[b] (obese—normal-weight) log FC	*P*_value
26	phosphatidylcholine acyl-alkyl C34:0	PC ae C34:0	glycerophospholipid	11803170	0.09	0.73
27	citric acid	citric acid	organic compounds (tricarboxylic acids)	311	0.06	0.82
28	L-aspartic acid	L-aspartic acid	organic compounds (aspartic acid)	5960	0.04	0.89
29	malic acid	malic acid	organic compounds (beta hydroxy acids)	525	0.01	0.98

ID	metabolite biochemical name	ChEBI	METLIN	KEGG	InChiKey
1	2-hydroxy-3-methylbutyric acid	60645	5396		NGEWQZIDQIYUNV-UHFFFAOYSA-N
2	galactonic acid	16534	3336	C00880	RGHNJXZEOKUKBD-MGCNEYSASA-N
3	phosphatidylcholine acyl-alkyl C40:3	NA	NA		UMEDQTSZFVUAIR-PBXDXGQFSA-N
4	propionylcarnitine	NA	NA	NA	UFAHZIUFPNSHSL-UHFFFAOYSA-N
5	phosphatidylcholine diacyl C40:3	NA	NA	C00157	LZXZOHSYOCWXFB-DWTYVZKUSA-N
6	*O*-butanoyl-carnitine, butenylcarnitine	NA	NA	NA	
7	hexanoylcarnitine, fumarylcarnitine	NA	NA	NA	VVPRQWTYSNDTEA-UHFFFAOYSA-N
8	phosphatidylcholine diacyl C40:2	NA	NA	C00157	FVYIRCQXEGIKHY-QYQAGSNQSA-N
9	benzoic acid	30746	1297	C00180	WPYMKLBDIGXBTP-UHFFFAOYSA-N
10	hydroxyhexadecenoylcarnitine	NA	NA	NA	WAGYLURELCUJPG-RPZXQYRRSA-N
11	myoinositol	17268	5221	C00137	CDAISMWEOUEBRE-GPIVLXJGSA-N
12	1,5-anhydrosorbitol	16070	3775	C07326	MPCAJMNYNOGXPB-SLPGGIOYSA-N
13	isovalerylcarnitine, valerylcarnitine, methylbutyrylcarnitine	NA	NA	NA	IGQBPDJNUXPEMT-UHFFFAOYSA-N
14	phosphatidylcholine acyl-alkyl C40:2	NA	NA		OQRGDDBNUMCNLP-ABEXPZPBSA-N
15	2-hydroxybutyric acid	1148	3783	C05984	AFENDNXGAFYKQO-UHFFFAOYSA-N
16	glutaconylcarnitine, mesaconylcarnitine	NA	NA	NA	JXVUHLILXGZLFR-DNQSNQRASA-N
17	L-arabitol	18403	141	C00532	HEBKCHPVOIAQTA-IMJSIDKUSA-N

ID	metabolite biochemical name	abbreviation	class	PubChem ID	fold change[b] (obese—normal-weight)	
					log FC	*P*_value
18	phosphatidylcholine acyl-alkyl C40:6	NA	NA		QBZALASVZLFAHF-KYPHJRDXSA-N	
19	octadecenoylcarnitine	NA	NA	NA	DGNPJQDFCXFOEZ-OEPBMOORSA-N	
20	phosphatidylcholine acyl-alkyl C34:3	NA	NA		QLEHHUPUHJPURI-PWYDUFMYSA-N	
21	phosphatidylcholine diacyl C36:5	NA	NA	C00157	KLTHQSWIRFFBRI-CPFPVJFHSA-N	
22	oleic acid	16196	190	C00712	ZQPPMHVWECSIRJ-KTKRTIGZSA-N	
23	phosphatidylcholine acyl-alkyl C36:2	NA	NA		FDNRZRXSENYWER-JTLDGTJHSA-N	
24	indoxyl sulfate	43355	524		BXFFHSIDQOFMLE-UHFFFAOYSA-N	
25	isoleucine	17191	5193	C00407	AGPKZVBTJJNPAG-WHFBIAKZSA-N	
26	phosphatidylcholine acyl-alkyl C34:0	NA	NA		PXPSGTINXJQLBR-VQJSHJPSSA-N	
27	citric acid	30769	124	C00158	KRKNYBCHXYNGOX-UHFFFAOYSA-N	
28	L-aspartic acid	17053	5206	C00049	CKLJMWTZIZZHCS-REOHCLBHSA-N	
29	malic acid	6650	118	C00711	BJEPYKJPYRNKOW-UHFFFAOYSA-N	

[a] Metabolites are sorted by the average log fold change of cases over controls.

[b] Fold change was calculated as mean (log2(obese)) – mean (log2(control)).