Metabolomics profile of umbilical cord blood is associated with maternal pre-pregnant obesity in a prospective multi-ethnic cohort displaying health disparities

Ryan J. Schlueter[1#], Fadhl M. Al-Akwaa[2#], Paula A. Benny[2], Alexandra Gurary[3], Guoxiang Xie[4], Wei Jia[4], Shaw J. Chun[2], Ingrid Chern[1], Lana X. Garmire[1,2*]

[1] University of Hawaii, Department of Obstetrics and Gynecology, 1319 Punahou St Ste 824, Honolulu, HI 96826

[2] University of Hawaii Cancer Center, Department of Epidemiology, 701 Ilalo Street, Honolulu, HI 96813

[3] University of Hawaii, John A. Burns School of Medicine, Department of Tropical Medicine, Medical Microbiology and Pharmacology, 651 Ilalo Street, Bioscience Building 320, Honolulu, Hawaii 96813

[4] University of Hawaii Cancer Center, Metabolomics Shared Resource, 701 Ilalo Street, Honolulu, HI 96813

[#] These authors contributed equally to this work.

[*] Address correspondence to Lana Garmire, PhD . Email address: lgarmire@cc.hawaii.edu

Abstract

Maternal obesity has become a growing global health concern that impacts fetal health and subsequently predisposes the offspring to medical conditions later in life. However, the molecular link between abnormal fetal metabolomic profiles and maternal obesity has not yet been fully elucidated. In this study, we report new discoveries from the newborn cord blood metabolomes associated with a case-control maternal obesity cohort, collected from multi-ethnic populations in Hawaii, including rarely reported Native Hawaiian and other Pacific Islanders (NHPI). This cohort displays significant maternal obesity disparities by subjects' residential area average income and health insurance. An elastic net penalized logistic regression model was constructed to associate cord blood metabolomics and demographic/physiological information to maternal obesity, with accuracy as high as 0.947. We identify 29 metabolites as early-life biomarkers manifesting intrauterine effect of maternal obesity.

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. Recently, research has focused on the association of maternal health during pregnancy and the subsequent effects on the future health of offspring (*3*). Since the inception of Barker's hypothesis in the 1990's, efforts to connect intrauterine exposures with the development of disease later in life has been the subject of many studies (*4*). Both obesity and its accompanying morbidities, such as diabetes, cardiovascular diseases and cancers, are of particular interest as considerable evidence has shown that maternal metabolic irregularities may have a role in genotypic programming in offspring (*5*, *6*). Identifying markers of

predisposition to health concerns or diseases would present an opportunity for early identification and potential intervention, thus providing life-long benefits (*7-9*).

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 (*10*). 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 (*11*). In *Drosophila melanogaster*, offspring of females given a high-sucrose diet exhibited metabolic aberrations both at the larvae and adult developmental stages (*12, 13*). Though an invertebrate model, mammalian lipid and carbohydrate systems show high level of conservation in *Drosophila melanogaster* (*14, 15*). In a mouse model of maternal obesity, progeny demonstrated significant elevations of both leptin and triglycerides when compared with offspring of control mothers of normal weight (*5*). The authors proposed that epigenetic modifications of obesogenic genes during intrauterine fetal growth play a role in adaption to an expected future environment. Recently, Tillery *et al.* used a primate model to examine the origins of metabolic disturbances and altered gene expression in offspring subjected to maternal obesity (*16*). 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 (*17*). It is a desirable technology that can detect distinct chemical imprints in tissues and body fluids (*18*). The field of metabolomics has shown great promise in various applications including early diagnostic marker identification (*19*), where a set of metabolomics biomarkers can differentiate samples of two different states (eg. disease and normal states). Cord blood metabolites provide information on fetal nutritional and metabolic health (*20*), 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 (*21*) and low birth

weight (*22*). For example, abnormal lipid metabolism and significant differences in relative amounts of amino acids were found in metabolomic signatures in cord blood from infants with intrauterine growth restriction in comparison to normal weight infants (*21*). In another study higher phenylalanine and citrulline levels but lower glutamine, choline, alanine, proline and glucose levels were observed in cord blood of infants of low-birth weight (*22*). However, thus far no metabolomics studies have been reported to specifically investigate the impact of maternal obesity on metabolomics profiles in fetal cord blood (*21-24*).

This study aims to investigate metabolomics changes in fetal cord blood associated with obese (BMI>30) and normal pre-pregnant weight (18.5<BMI<25) mothers, in a multi-ethnic population including Native Hawaiian and other Pacific Islanders (NHPI). NHPI is a particularly under-represented minority population across most scientific studies. To ensure the quality of the study, we enrolled the mothers undergoing elected C-sections without any clinically known gestational diseases. In addition to the cord blood samples of their babies at birth, we collected comprehensive EMR records from the subjects, other maternal and paternal parameters such as ethnicities. This study has not only revealed the maternal health disparities on Oahu island, Hawaii, but also discovered the metabolomic links between cord blood and maternal pre pregnant obesity. These metabolites are potential early-life biomarkers affected by maternal obesity.

Materials and methods

Study population

We performed a multi-ethnic case-control study at Kapiolani Medical Center for Women and Children, Honolulu, HI from June 2016 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 (*25*) we recruited women scheduled for full-term cesarean section at ≥ 37 weeks gestation. All subjects fasted for at least 8 hours before the scheduled cesarean delivery. Patients meeting inclusion

criteria were identified from pre-admission medical records with pre-pregnancy BMI ≥30.0 (cases) or 18.5-25.0 (controls). The pre-pregnancy BMIs were also confirmed during the enrollment. Women with preterm rupture of membranes (PROM), labor, multiple gestations, pre-gestational diabetes, hypertensive disorders, cigarette smokers, HIV, HBV, and chronic drug users were excluded. Demographic and clinical characteristics were recorded, including residential zipcode, insurance type, maternal and paternal age, maternal and paternal ethinicities, mother's pre-pregnancy BMI, net weight gain, gestational age, parity, gravidity and ethnicity. A total of 57 samples (28 cases and 29 controls) were collected.

Sample collection, preparation and quality control

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

Metabolome profiling

The plasma samples were thawed and extracted with 3-vol cold organic mixture of ethanol: chloroform and centrifuged at 4 °C at 13500 rpm for 20 min. The supernatant was split for lipid and amino acid profiling with an Acquity ultra performance liquid chromatography coupled to a Xevo TQ-S mass spectrometry (UPLC-MS/MS, Waters Corp., Milford, MA). Metabolic profiling of other metabolites including organic acids, carbohydrates, amino acids, and nucleotides were measured with an Agilent 7890A gas chromatography coupled to a 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 (AMSDP) software package (*26*). The levels of lipids and amino acids detected from LC-MS were calculated with calibration curves established with reference standards.

Metabolomics data processing

We conducted data pre-processing similar to the previous report (*27*). Briefly, we used K-Nearest Neighbors (KNN) method to impute missing metabolomics data (*28*). To adjust for the offset between high and low-intensity features, and to reduce the heteroscedasticity, the logged value of each metabolite was centred by its mean and autoscaled by its standard deviation (*29*). We used quantile normalization to reduce sample-to-sample variation (*30*). We applied partial least squares discriminant analysis (PLS-DA) to visualize how well metabolites could differentiate the obese from normal samples.

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 glment R package (*31*). Elastic net method selects metabolites that have non-zero coefficients as features, guided by two penalty parameters alpha and lambda (*31*). Alpha sets the degree of mixing between lasso (when alpha=1) and the ridge regression (when alpha=0). Lambda controls the shrunk rate of cofficients regardless of the value of alpha. When lambda equals zero, no shrinkage is performed and the algorithm selects all the features. As lambda increases, the coefficients are shrunk more strongly and the algorithm retrives all features with non-zero coefficients. To find optimal parameters, we performed 10-fold cross-validation that yield the smallest prediction minimum square error (MSE). 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 divided samples into 5 folds. We trained the model on four folds (80% of data) using leave one out cross validation (LOOCV) and measured model performance on the remaining fold (20% of data). We carried out the above training and testing five times on all folds combination. We plotted the receiver-operating characteristic (ROC) curve for all folds prediction using pROC R package. To adjust confounding other clinical covariants such as ethnicity, gravidity and parity, we reconstructed the metabolomics model above by including these factors.

Analysis on metabolite features

We used Classification And REgression Training (CARET) R package to rank metabolites based on the model-based approach (*32*). In this approach, each metabolite was assigned a score that estimates its contribution to the model performance (*33*). These scores were scaled to have a maximum of 100. We performed metabolomic pathway analysis on metabolites chosen by the elastic net method using Consensus Pathway DataBase (CPDB). We used rcorr function implemented in Hmisc R package to compute the correlations among clinical metabolomics data.

Data availability

The metabolomics data generated by this study is deposited to Metabolomics workbench (ID1312).

Results

Cohort subjects characteristics

Our cohort consisted of three ethnic groups: Caucasian, Asian and Native Hawaiian and other Pacific Islander (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 non-obese and 4 as obese. In the Asian group (23 mothers), 16 were categorized as non-obese and 7 as obese. In the NHPI group (24 mothers), 7 (24%) were categorized as non-obese and 17 (61%) as obese.

The variation in recruitment of cases versus controls in each ethnic background reflects the demographics in Hawaii. Compared to mothers of normal pre-pregnant BMI, obese mothers have significantly higher pre-pregnancy BMI (33.51+/- 4.49 vs 21.89 +/- 1.86 kg/m2, p=9.18e-11). Mothers have 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 confouding from these factors. Babies of obese mothers have significantly (P=0.03) higher birth weight compared to the normal pre pregnant weight group, consistent with earlier observations (*34*, *35*).

The cohort displays disparities of maternal obesity

In an interest to seek possible relationship between maternal obesity and social economics status, we retrieved the residential zip code and patient insurance from the subjects. To estimate each subject's income, we used the surrogate of annual averaged personal income in that person's zip code, based on the IRS data of year 2016. A 2-tail t-test analysis shows that the difference of annual income between obese and control group's residential areas is statistically significant (p=0.04, Figure 1A). Further, we looked into the differences of insurance carriers of the subjects. In the state of Hawaii, the lowest tier of insurance is Quest care, which covers medicaid and less wealthy population. The mid-tier coverage are provided by HMSA HMO, HMA, TriCare, and United Health Care. The highest and most expensive health coverage is under HMSA PPO, in which patients pay a premium to have more choices in picking healthcare providers. Figure 1B shows the number of obese and non-obese subjects across 11 different working insurance companies in Hawaii. Around 57% of obese group are insured by Quest, compared to 28% of control group with this insurance; whereas 38% of control group are insured by HMSA PPO, compared to 17% of the obese group with this insurance; A two-way ANOVA analysis between obesity vs. high and low insurance tier has almost significant p-value (p=0.06), indicating signs of social economial disparities among mothers.

Preliminary assessment of metabolomics results

We detected a total of 230 metabolites, including 79 untargeted and 151 targeted metabolites (11 amino acids, and 140 lipids). To test if these metabolites allow clear separation between the obese and normal-weight subjects, we used elastic net regularization based logistic regression, rather than the partial least squares (PLS) model, a routine supervised multivariate method which only yielded modest accuracy AUC=0.62 (Fig 1S). Elastic net regularization overcomes the limitation of either ridge and lasso regularization alone, and combines their strengths to identify an optimized set metabolites [25]. Using the optimized regularization parameters (Fig. 2S), we identified a total of 29 metabolite features, which together yields the highest predictive performance with AUC=0.97, 95 % CI=[0.904-0.986] in 20% hold-out test dataset (Figure 2A). Among them, six metabolites have large contributions to the separations between case/control, with an importance score of at least 70% individually (Figure 2A). 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.

Combined predictive model for maternal obesity with consideration of confounding

Some demographic and physiological factors, such as maternal/paternal ethnicity and parity (Table 1) may confound the metabolites signatures above. To check this, we conducted two analyses. First, we explored the correlations among the demographic factors and metabolomics data. It is evident that several metaboloties are correlated with maternal and paternal ethnicity, gravidity, and/or parity (Figure 3). For example, maternal ethnicity is positively correlated with 2-hydroxy-3-methylbutyric acid. Next, we built a logistic regression model using the above-mentioned four covariates (parity, gravidity, maternal and paternal ethnicity). This model yields an modest AUC of 0.701 95% CI=[0.55-0.82] (Figure 2B), again supporting their confounding effect to associate with maternal obesity. These observations prompted us to recalibrate the 29-metabolite elastic net model, by adjusting the metabolomics model using all collected clinical covariants (Figure 2C). The resulting model remains to have very high accuracy, with AUC=

0.947, 95% CI= [0.87-0.97]. In the new model, besides the original 6 metabolite features, maternal ethnicity and paternal ethnicity also have importance scores greater than 70% (Figure 2C).

Metabolomite features and their pathway and enrichment analysis

The 29 metabolite features selected by the model belong to acylcarnitine, glycerophospholipid, amino acids and organic acids classes. Their log fold changes ranged from -0.45 (Hydroxyhexadecenoylcarnitine, or C16:1-OH) to 0.66 (2-hydroxy-3-methylbutyric acid) (Figure 4A). Among them, 15 metabolites are 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, L-arabitol, Octadecenoylcarnitine (C18:1) (Figure 4A, Table 2) . The remaining 14 metabolites are lower in obese associated cord blood samples: malic acid, L-aspartic acid, 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 4A, 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 effected by maternal obesity, we performed pathway enrichment analysis on the 29 metabolite features, using Consensus pathway database (CPDB) tool (*36*). To gain most comphrehsenive pathway list, we combined multiple pathway databases including KEGG, Wikipathways, Reactome, EHNM and SMPDB. A list of 10 pathways are enriched with adjusted p-value $q<0.05$ (Figure 4B). Among them, alanine and aspartate metabolism is the most significantly enriched pathway ($q=0.004$). Transmembrane transport of small molecules and SLC-mediated transmembrane transport are also significantly enriched ($q=0.004$ and $q=0.01$ respectively).

Ethinicity association of metabolites

Our earlier correlational analysis suggested that maternal ethnicity may be correlated with 2-hydroxy-3-methylbutyric acid level. To confirm this, we conducted 2-way ANOVA statistical tests and indeed obtained significant p-value (P=0.023, chi-square test). We thus stratified the levels of 2-hydroxy-3-methylbutyric acid by ethnicity (Figure 5). There is no significant difference in normal pre pregnant-weight subjects across the three ethnic groups (Figure 5A). However, in cord blood samples associated with obese mothers, the concentration of 2-hydroxy-3-methylbutyric acid is much higher in NHPI, as compared to Caucasians (p=0.05) or Asians (p=0.04) (Figure 5B). 2-hydroxy-3-methylbutyric acid originates mainly from ketogenesis through the metabolism of valine, leucine and isoleucine (*37*). Since all subjects have fasted 8 hours before the C-section, we expect the confounding from diets is minimized among the three ethnical groups. Thus the higher 2-hydroxy-3-methylbutyric acid level may indicate the higher efficiency of ketogenesis in babies born from obese NHPI mothers.

Discussion

This study aims to distinguish key cord blood metabolites associated with maternal pre-pregnancy obesity. The novelty of the study is manifested in several folds. First, we have collected a unique multi-ethnic population in Hawaii, which includes Asian, NHPI and Caucasians. Secondly, this is the first human metabolomics study that is also connected to maternal obesity disparities, demonstrated by geographical and health insurance analyses. Thirdly, we utilize state of the art metabolomics technology platform coupling GC-MS and LC-MS platforms, which allows us to detect hundreds of metabolites simultaneously. Lastly, we use the state of art method called elastic net based logistic regression that drastically improves the classification accuracy on cord blood metabolomics data.

We conducted rigorous statistical modeling and found that metabolites can distinguish the two maternal groups with accuracy as high as AUC=0.97 (or 0.947 after adjusting for confounding effects). Metabolomics pathway analysis on the metabolite features in the model identified 10 significant

pathways. Among them, alanine and aspartate metabolism was previously reported to be associated with obesity (*38*). Transmembrane transport was identified as another significant pathway. The transmembrane transport pathway corresponds to the acylcarnitine metabolites in the features. Acylcarnitines are known transmembrane transporters of fatty acids across the mitochondrial membrane (*39*). 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, was previously shown to be associated with diabetes in a mouse model, due to a proposed mechanism of oxidative stress (*40*). On the other hand, maternal ethnicity has the largest impact among physiological factors (importance score =84%).

A very few cord blood metabolomics studies have been carried out to associate with maternal obesity directly, or birth weight (*22, 41, 42*). In a recent Hyperglycemia and Adverse Pregnancy Outcome (HAPO) Study, Lowe et al. reported that branched-chain amino acids such as valine, phenylalanie, leucine/isoleucine and AC C4, AC C3, AC C5 are associated with maternal BMI in a meta-analysis over 4 large cohorts (400 subjects in each) (*42*). In another study to associate cord blood metabolomics with low birth weight (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-90th percentiles, n = 30) (*22*). They also found lower levels of choline, proline, glutamine, alanine and glucose in new borns of LBW, however, there was no significant differences between the mothers of the two groups. In our study, isoleucine is also identified as one of the 29 metablite features related to maternal obesity; although alanine iteself 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 maternal obesity group.

Notably, our study has identified 5 metabolites which are previously not reported in the literature with association to 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 pre-pregnant 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 new borns from obese pre-pregnant 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 amino acid 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 pre-pregnant mothers could be due to the maternal high fat diet. Oppositely, citric acid, a compound associated with the citric acid cycle (*43*), is decreased in the cord blood associated with obese pre-pregnant mothers. This could be related to the lower vegitable and fruit consumptions among obese pre-pregnant 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.

One limitation of the study was the modest sample size, given the stringent inclusion and exclusion crtieria. To avoid the confounding from labor and vaginal delivery, we only targeted mothers having elective C-sections. We also excluded obese mothers who had known complications during pregnancy, such as pre-gestational diabetes, smoking, and hypertension. These criteria helped to improve the qualtiy of the metabolomics data, at the tradeoff of the sample size. Awaring of this potential issue, we assessed the regression model using cross-validation and hold-out testing dataset, rather than using another validation cohort. The second caveat is that the extent of confounding due to maternal diet is unknown, although all subjects fasted 8 hours before the Cesarean section. Fasting states are commonly employed in metabolomics studies. Thirdly, we determined the subjects' ethnicity by self-reporting rather than genotyping, due to the restriction of the currently approved IRB protocol. Additionally, there has been

debates on the use of BMI as an indicator of obesity (*44*), and more direct measures of body fat could be considered such as skin-fold thickness measurements, bioelectrical impedance and energy x-ray absorptiometry (*45, 46*). Lastly, a longitudinal follow-up study on the developmental trajectory of offspring of obese mothers would provide further insights. We plan to conduct a larger-scale maternal-offspring obesity study by addressing all the issues above. Neverthless, this study has established relationships between cord blood metabolomics with maternal pre-pregnant obesity, which in turn is associated with socialeconomical disparties.

Conclusion

In this study, we identified 29 metabolites that are associated with maternal obesity, 5 of which are previously unreported in the literature. These metabolites have the potential to be maternal obesity-related bio-markers in newborns that warranty dietary interventions in early-life.

Author Contributions

LXG envisioned the project, obtained funding, designed and supervised the project and data analysis. RJS, IC, PAB and SJC collected the samples. AG prepared the plasma samples. FMA analyzed the data. GX performed the metabolomics experiments. RJS, FMA, PAB, AG, GX, SJC and LXG wrote the manuscript. All authors have read, revised, and approved the manuscript.

Competing financial interests

The authors declare no competing financial interests.

Acknowledgements

We thank Drs. Joseph Kaholokula and Alika Maunakea from the Native Hawaiian Health Department of University of Hawaii for giving suggestions. The authors 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 LX Garmire. Funding was also provided in part by the Department of Obstetrics and Gynecology, University of Hawaii. The metabolomics services were provided by the UH Cancer Center Metabolomics Shared Resource.

References

1. K. M. Flegal, M. D. Carroll, B. K. Kit, C. L. Ogden, Prevalence of obesity and trends in the distribution of body mass index among us adults, 1999-2010. *JAMA* **307**, 491-497 (2012).
2. C. L. Ogden, M. D. Carroll, B. K. Kit, K. M. Flegal, Prevalence of obesity in the United States, 2009-2010. *NCHS Data Brief*, 1-8 (2012).
3. C. J. Smith, K. K. Ryckman, Epigenetic and developmental influences on the risk of obesity, diabetes, and metabolic syndrome. *Diabetes Metab Syndr Obes* **8**, 295-302 (2015).
4. D. J. Barker, Fetal origins of coronary heart disease. *BMJ* **311**, 171-174 (1995).
5. C. C. Li *et al.*, Maternal obesity and diabetes induces latent metabolic defects and widespread epigenetic changes in isogenic mice. *Epigenetics* **8**, 602-611 (2013).
6. X. Pi-Sunyer, The medical risks of obesity. *Postgrad Med* **121**, 21-33 (2009).
7. E. Isganaitis *et al.*, Associations of cord blood metabolites with early childhood obesity risk. *Int J Obes (Lond)* **39**, 1041-1048 (2015).
8. S. Rauschert, O. Uhl, B. Koletzko, C. Hellmuth, Metabolomic biomarkers for obesity in humans: a short review. *Ann Nutr Metab* **64**, 314-324 (2014).
9. M. F. Hivert *et al.*, Metabolomics in the developmental origins of obesity and its cardiometabolic consequences. *J Dev Orig Health Dis* **6**, 65-78 (2015).
10. M. J. Heerwagen, M. R. Miller, L. A. Barbour, J. E. Friedman, Maternal obesity and fetal metabolic programming: a fertile epigenetic soil. *Am J Physiol Regul Integr Comp Physiol* **299**, R711-722 (2010).
11. S. Rauschert *et al.*, Early Programming of Obesity Throughout the Life Course: A Metabolomics Perspective. *Ann Nutr Metab* **70**, 201-209 (2017).
12. R. T. Brookheart, J. G. Duncan, Drosophila melanogaster: An emerging model of transgenerational effects of maternal obesity. *Mol Cell Endocrinol* **435**, 20-28 (2016).

13. R. T. Brookheart, A. R. Swearingen, C. A. Collins, L. M. Cline, J. G. Duncan, High-sucrose-induced maternal obesity disrupts ovarian function and decreases fertility in Drosophila melanogaster. *Biochim Biophys Acta* **1863**, 1255-1263 (2017).
14. I. Trinh, G. L. Boulianne, Modeling obesity and its associated disorders in Drosophila. *Physiology (Bethesda)* **28**, 117-124 (2013).
15. S. B. Diop, R. Bodmer, Drosophila as a model to study the genetic mechanisms of obesity-associated heart dysfunction. *J Cell Mol Med* **16**, 966-971 (2012).
16. K. M. Aagaard-Tillery *et al.*, Developmental origins of disease and determinants of chromatin structure: maternal diet modifies the primate fetal epigenome. *J Mol Endocrinol* **41**, 91-102 (2008).
17. D. S. Wishart *et al.*, HMDB: the Human Metabolome Database. *Nucleic Acids Res* **35**, D521-526 (2007).
18. M. Ciborowski *et al.*, Potential first trimester metabolomic biomarkers of abnormal birth weight in healthy pregnancies. *Prenat Diagn* **34**, 870-877 (2014).
19. G. A. Gowda *et al.*, Metabolomics-based methods for early disease diagnostics. *Expert Rev Mol Diagn* **8**, 617-633 (2008).
20. M. Sanz-Cortes *et al.*, 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* **8**, e80121 (2013).
21. D. Favretto *et al.*, Cord blood metabolomic profiling in intrauterine growth restriction. *Anal Bioanal Chem* **402**, 1109-1121 (2012).
22. C. Ivorra *et al.*, Metabolomic profiling in blood from umbilical cords of low birth weight newborns. *J Transl Med* **10**, 142 (2012).
23. D. La Torre *et al.*, Decreased cord-blood phospholipids in young age-at-onset type 1 diabetes. *Diabetes* **62**, 3951-3956 (2013).
24. A. Zhang, H. Sun, X. Wang, Power of metabolomics in biomarker discovery and mining mechanisms of obesity. *Obes Rev* **14**, 344-349 (2013).
25. N. Gomez-Lopez, D. StLouis, M. A. Lehr, E. N. Sanchez-Rodriguez, M. Arenas-Hernandez, Immune cells in term and preterm labor. *Cell Mol Immunol* **11**, 571-581 (2014).
26. Y. Ni *et al.*, ADAP-GC 2.0: deconvolution of coeluting metabolites from GC/TOF-MS data for metabolomics studies. *Anal Chem* **84**, 6619-6629 (2012).
27. F. M. Alkawaa, K. Chaudhary, L. X. Garmire, Deep learning accurately predicts estrogen receptor status in breast cancer metabolomics data. *J Proteome Res*, (2017).
28. L. Beretta, A. Santaniello, Nearest neighbor imputation algorithms: a critical evaluation. *BMC Med Inform Decis Mak* **16 Suppl 3**, 74 (2016).
29. R. A. van den Berg, H. C. Hoefsloot, J. A. Westerhuis, A. K. Smilde, M. J. van der Werf, Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC Genomics* **7**, 142 (2006).
30. A. Jauhiainen *et al.*, Normalization of metabolomics data with applications to correlation maps. *Bioinformatics* **30**, 2155-2161 (2014).
31. J. P. Paolino, Rasch Model Parameter Estimation via the Elastic Net. *J Appl Meas* **16**, 353-364 (2015).
32. M. Kuhn, Building Predictive Models in R Using the caret Package. *2008* **28**, 26 (2008).
33. S. G. Arunajadai, Stepwise logistic regression. *Anesth Analg* **109**, 285; author reply 285-286 (2009).
34. V. Briese, M. Voigt, M. Hermanussen, U. Wittwer-Backofen, Morbid obesity: pregnancy risks, birth risks and status of the newborn. *Homo* **61**, 64-72 (2010).
35. K. Shankar *et al.*, Maternal obesity at conception programs obesity in the offspring. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology* **294**, R528-R538 (2008).

36. A. Kamburov, U. Stelzl, H. Lehrach, R. Herwig, The ConsensusPathDB interaction database: 2013 update. *Nucleic Acids Res* **41**, D793-800 (2013).
37. H. M. Liebich, C. Forst, Hydroxycarboxylic and oxocarboxylic acids in urine: products from branched-chain amino acid degradation and from ketogenesis. *J Chromatogr* **309**, 225-242 (1984).
38. S. Sookoian, C. J. Pirola, Alanine and aspartate aminotransferase and glutamine-cycling pathway: their roles in pathogenesis of metabolic syndrome. *World J Gastroenterol* **18**, 3775-3781 (2012).
39. N. Longo, M. Frigeni, M. Pasquali, Carnitine transport and fatty acid oxidation. *Biochim Biophys Acta* **1863**, 2422-2435 (2016).
40. J. Fahrmann *et al.*, Systemic alterations in the metabolome of diabetic NOD mice delineate increased oxidative stress accompanied by reduced inflammation and hypertriglyceremia. *Am J Physiol Endocrinol Metab* **308**, E978-989 (2015).
41. D. L. Santos Ferreira *et al.*, Association of pre-pregnancy body mass index with offspring metabolic profile: Analyses of 3 European prospective birth cohorts. *PLoS Med* **14**, e1002376 (2017).
42. W. L. Lowe, Jr. *et al.*, Maternal BMI and Glycemia Impact the Fetal Metabolome. *Diabetes Care* **40**, 902-910 (2017).
43. K. Muroyama *et al.*, Anti-obesity effects of a mixture of thiamin, arginine, caffeine, and citric acid in non-insulin dependent diabetic KK mice. *J Nutr Sci Vitaminol (Tokyo)* **49**, 56-63 (2003).
44. D. S. Freedman, M. Horlick, G. S. Berenson, 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* **98**, 1417-1424 (2013).
45. C. Wohlfahrt-Veje *et al.*, Body fat throughout childhood in 2647 healthy Danish children: agreement of BMI, waist circumference, skinfolds with dual X-ray absorptiometry. *Eur J Clin Nutr* **68**, 664-670 (2014).
46. J. Steinberger *et al.*, 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 (Lond)* **29**, 1346-1352 (2005).

Figures

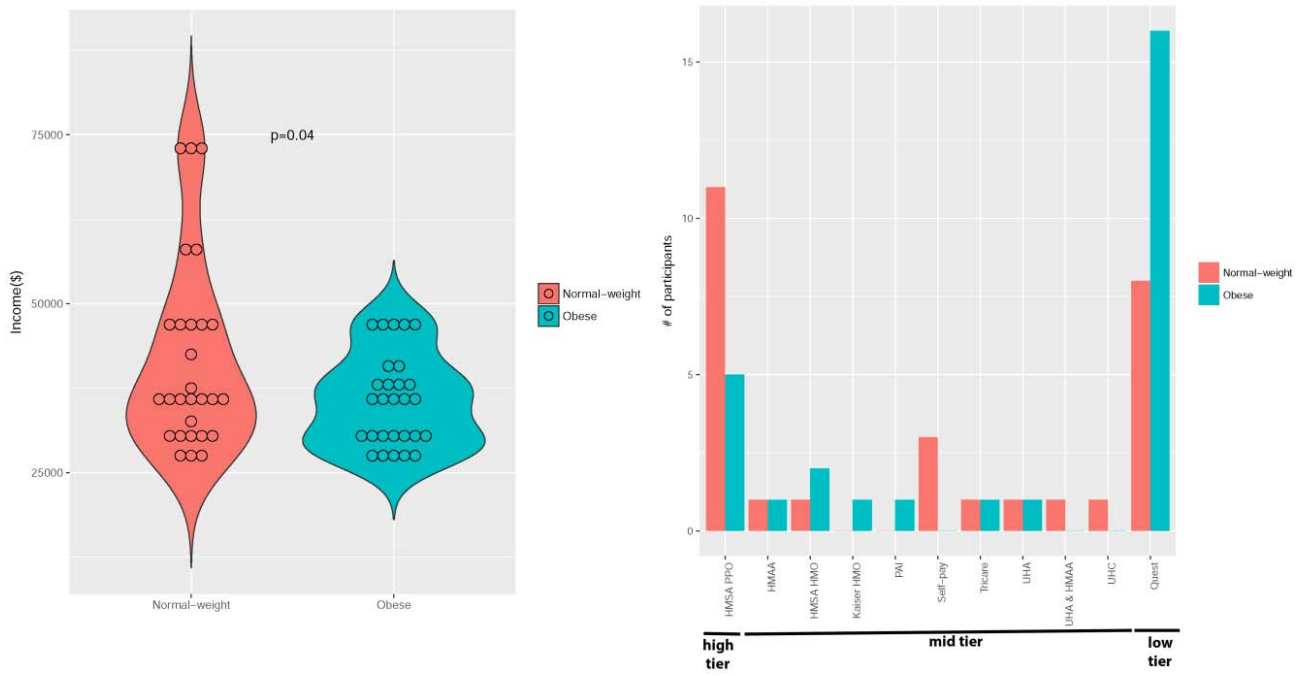
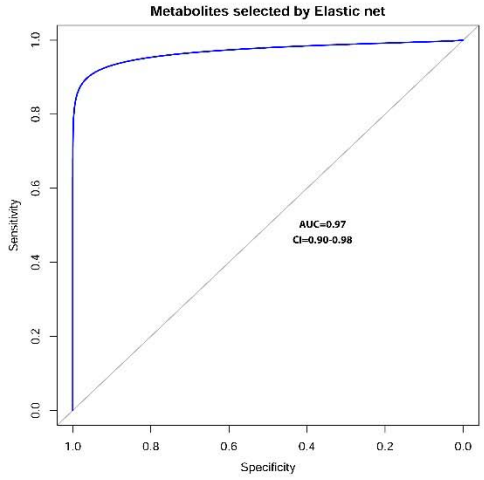
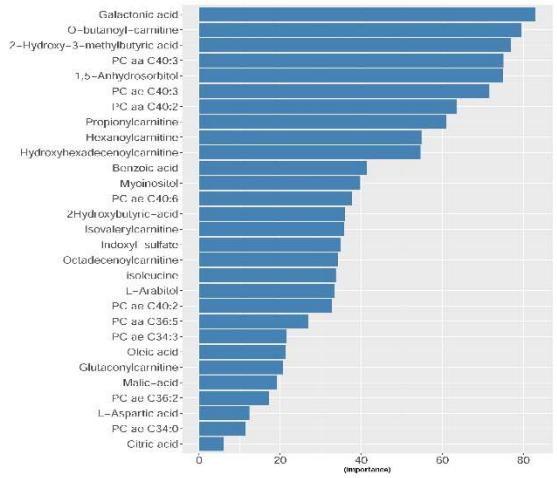


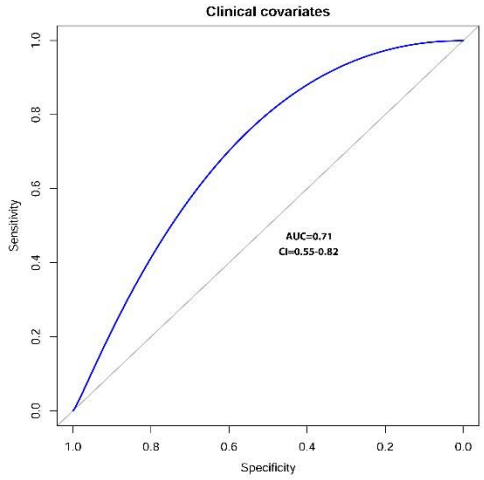
Figure 1: Maternal obesity disparity among subjects. (A) Violin plot of the average annual income from the zip codes where obese and normal pre pregnant weight subjects live. B) Bar plot of insurance companies among subjects. The highest tire of health insurance is HMO PPO; the lowest tire is Quest; and the remaining health insurances are the middle tier (Hawaii Medical Assurance Association (HMAA), Hawaii Medical Service Association (HMSA HMO), Kaiser HMO, Pacific Administrators Inc (PAI), Self-pay, TRICARE, UHA Health Insurance (UHA), United Health Care (UHC).

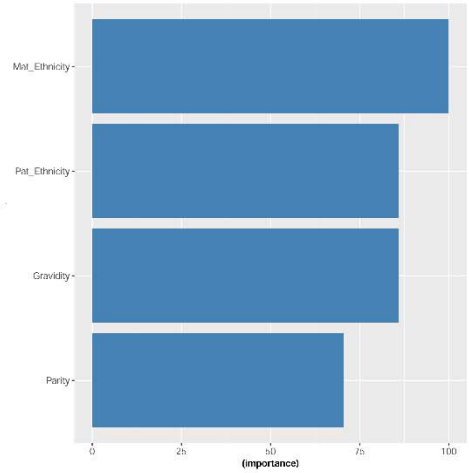
A


Metabolites selected by Elastic net
AUC=0.97
CI=0.90-0.98
Sensitivity
Specificity
1.0
0.8
0.6
0.4
0.2
0.0

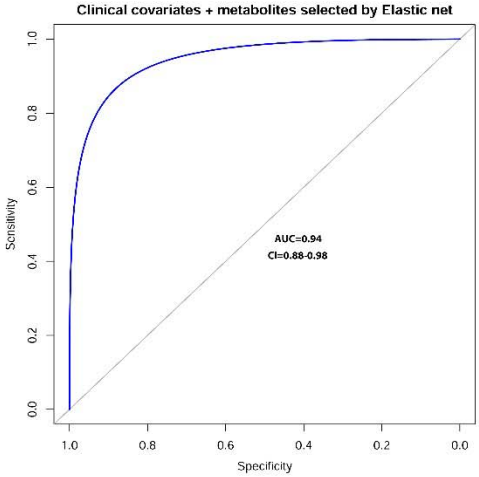

Galactonic acid
O-butanoyl-carnitine
2-Hydroxy-3-methylbutyric acid
PC aa C40:3
1,5-Anhydrosorbitol
PC ae C40:3
PC aa C40:2
Propionylcarnitine
Hexanoylcarnitine
Hydroxyhexadecenoylcarnitine
Benzoic acid
Myoinositol
PC ae C40:6
2Hydroxybutyric-acid
Isovalerylcarnitine
Indoxyl sulfate
Octadecenoylcarnitine
Isoleucine
L-Arabitol
PC ae C40:2
PC aa C36:5
PC ae C34:3
Oleic acid
Glutaconylcarnitine
Malic-acid
PC ae C36:2
L-Aspartic acid
PC ae C34:0
Citric acid
0
20
40
60
80
(importance)

B


Clinical covariates
AUC=0.71
CI=0.55-0.82
Sensitivity
Specificity
1.0
0.8
0.6
0.4
0.2
0.0


Mat_Ethnicity
Pat_Ethnicity
Gravidity
Parity
0
25
50
75
100
(importance)

C


Clinical covariates + metabolites selected by Elastic net
AUC=0.94
CI=0.88-0.98
Sensitivity
Specificity
1.0
0.8
0.6
0.4
0.2
0.0

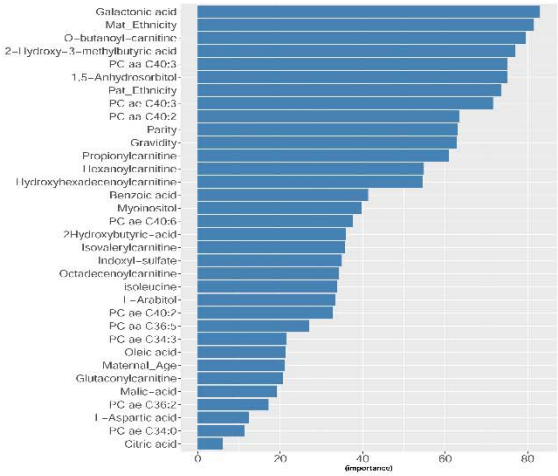

Galactonic acid
Mat_Ethnicity
O-butanoyl-carnitine
2-Hydroxy-3-methylbutyric acid
PC aa C40:3
1,5-Anhydrosorbitol
Pat_Ethnicity
PC ae C40:3
PC aa C40:2
Parity
Gravidity
Propionylcarnitine
Hexanoylcarnitine
Hydroxyhexadecenoylcarnitine
Benzoic acid
Myoinositol
PC ae C40:6
2Hydroxybutyric-acid
Isovalerylcarnitine
Indoxyl-sulfate
Octadecenoylcarnitine
Isoleucine
L-Arabitol
PC ae C40:2
PC aa C36:5
PC ae C34:3
Oleic acid
Maternal_Age
Glutaconylcarnitine
Malic-acid
PC ae C36:2
L-Aspartic acid
PC ae C34:0
Citric acid
0
20
40
60
80
(importance)

Figure 2: Accuracies of logistic regression models and important features selected by the models. Model accuracy is represented by classification Receiver Operator Curves (ROCs). The contributions (percentage) of selected features in each model are ranked from high to low. (A) Metabolomics data based model. (B) Physiological/demographic data based model. (C) Combined model with metabolomics and physiological/demographic data.

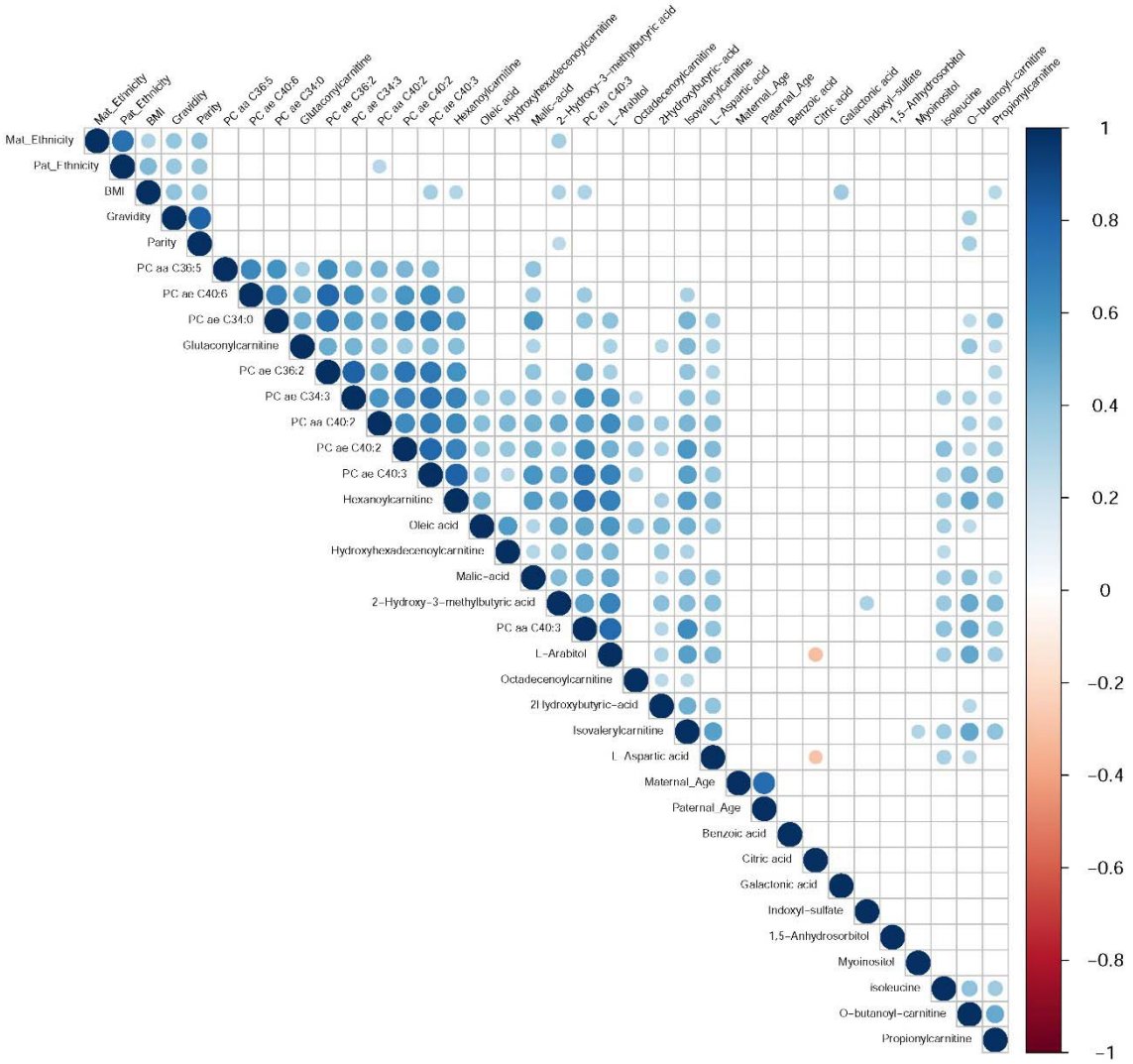


Figure 3: Correlation coefficients among demographical/physiological factors and the metabolomics data. Blue colors indicates positive correlations and red indicated negative correlations.

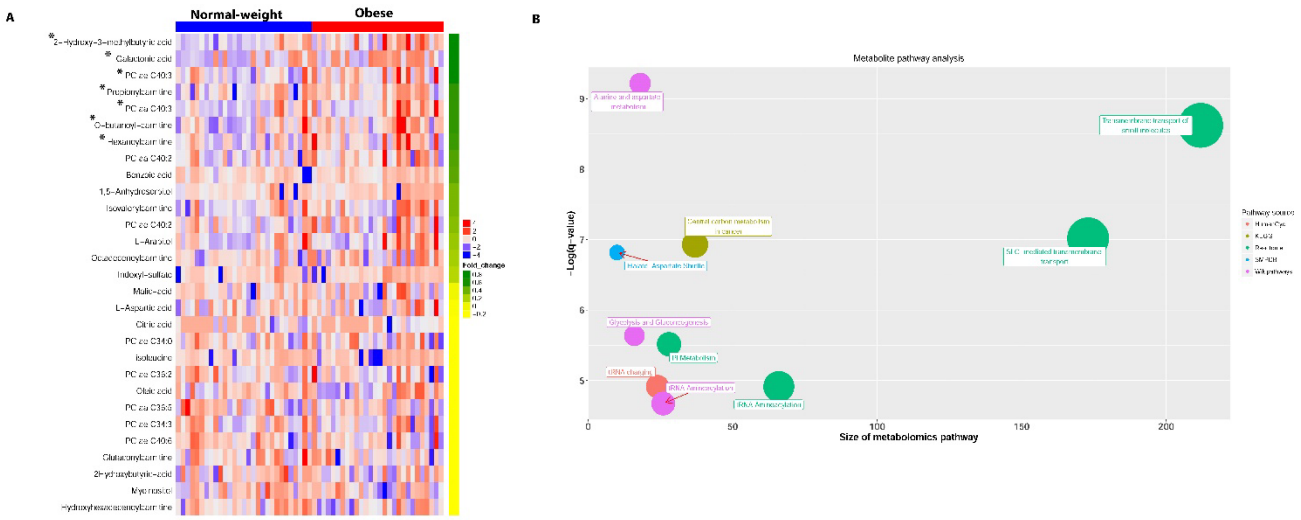


Figure 4: Analysis of the 29 selected metabolites. (A) Heatmap of selected metabolites separated by maternal group. * indicates metabolites that shows significant p-values (P<0.05, t-test) individually. (B) Pathway analysis of the 29 metabolites. X-axis shows size of metabolomic pathway. Y-axis shows the adjusted p-value calculated from CPDB tool. The size of the nodes represents the size of metabolomic pathway (number of metabolites involved in each pathway). The color of the nodes represents the source of these pathways.

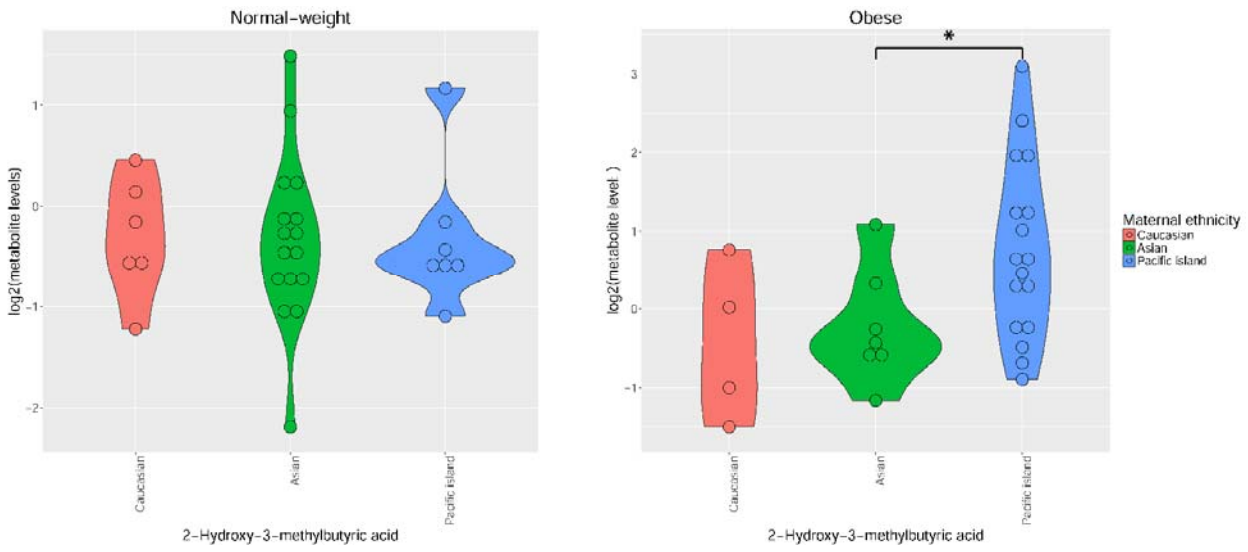
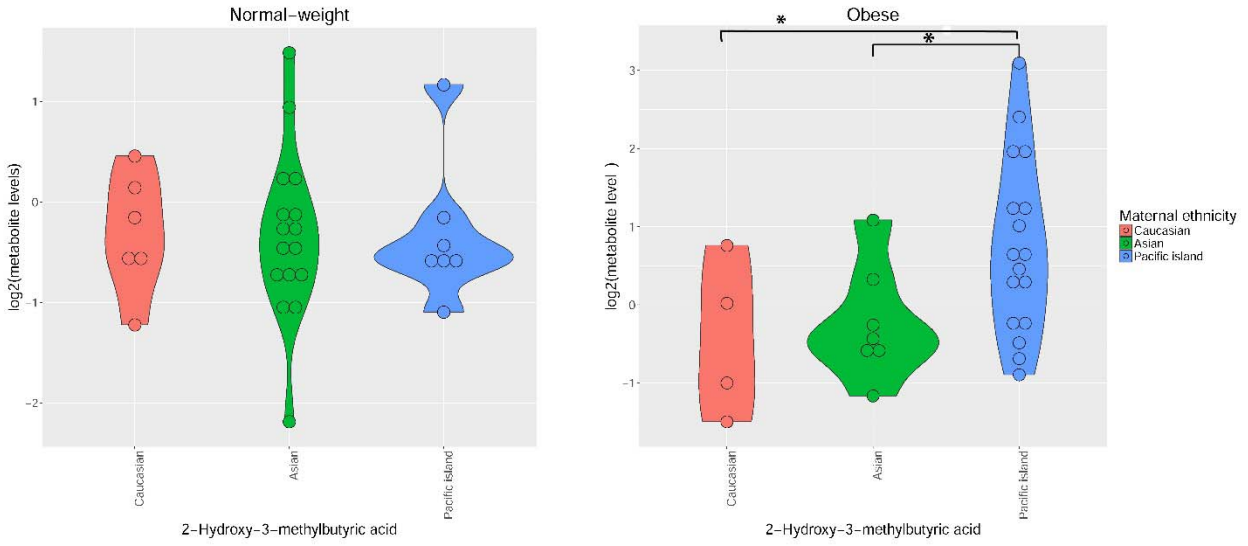


Figure 5: Violin plot of 2-hydroxy-3-methylbutyric acid among 3 ethnic groups. Association between 2-hydroxy-3-methylbutyric acid and the ethnicity in (**A)** normal (n=29) and (**B)** obese (n=28) subjects.



Tables

Table 1: Demographical and clinical characteristics in obese and control groups

	Control(n=29)	**Case(n=28)**	**P-value***
	Mean (SD)		
Maternal age, years	32.48 (5.66)	32.10 (4.88)	0.78
Paternal age, years	34.68(7.14)	35.21(6.43)	0.79
Pre-pregnancy BMI, kg/m2	21.89(1.86)	33.51(4.49)	1.12 e-14
Gestational Age, Weeks	39.04(0.218)	38.93(0.45)	0.3812
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
0	5	2	
1	16	7	
2	7	10	
3 and above	1	9	
Gravidity			0.12
1	5	1	
2	12	5	
3	7	8	
4 and above	5	14	
Maternal Ethnicity			0.01
Caucasian	6	4	
Asian	16	7	
Pacific island	7	17	
Paternal Ethnicity			0.03
Caucasian	8	3	
Asian	14	9	
Pacific island	7	16	

*Categorical variables were compared using chi-square test, whereas continuous variables were compared using *t* test.

Table 2: A list of metabolites associated with obese-control maternal status and selected by elastic net regularization based logistic regression. The metabolites are sorted by the average log fold change of cases over controls.

Metabolites	Chemical name	Fold change [a] (case-control)	Univariate Analysis [b]

		logFC	P_value	Cofficient	P_value
2-hydroxy-3-methylbutyric acid	2-hydroxy-3-methylbutyric acid	0.6609	0.0119	0.65592	0.062950865
Galactonic acid	Galactonic acid	0.6337	0.0158	0.640515	0.06565148
PC ae C40:3	Phosphatidylcholine acyl-alkyl C40:3	0.6249	0.0173	0.762691	0.035189439
C3	Propionylcarnitine	0.5598	0.033	-0.1467	0.648143485
PC aa C40:3	Phosphatidylcholine diacyl C40:3	0.5561	0.0342	-0.33489	0.318665241
C4:1	O-butanoyl-carnitine, butenylcarnitine	0.556	0.0342	-0.44274	0.168989046
C6 (C4:1 -DC)	Hexanoylcarnitine, Fumarylcarnitine	0.5355	0.0414	-0.28551	0.337718
PC aa C40:2	Phosphatidylcholine diacyl C40:2	0.4793	0.0679	0.532796	0.113517583
Benzoic acid	Benzoic acid	0.4549	0.0831	0.279734	0.350259256
1,5-Anhydrosorbitol	1,5-Anhydrosorbitol	0.3664	0.1628	0.636374	0.24536415
C5	Isovalerylcarnitine, Valerylcarnitine, Methylbutyrylcarnitine	0.3654	0.1638	-0.38664	0.196793118
PC ae C40:2	Phosphatidylcholine acyl-alkyl C40:2	0.3242	0.2168	-0.71475	0.042908449
L-Arabitol	L-Arabitol	0.2685	0.3062	0.360549	0.266082992
C18:1	Octadecenoylcarnitine	0.228	0.385	0.253734	0.427416515
Indoxyl sulfate	Indoxyl sulfate	0.1792	0.4948	-0.06239	0.827985019
Malic acid	Malic acid	-0.006	0.9811	0.010217	0.977502972
L-Aspartic acid	L-Aspartic acid	-0.036	0.8899	-0.18507	0.549849292
Citric acid	Citric acid	-0.058	0.8242	-0.08235	0.790831897
PC ae C34:0	Phosphatidylcholine acyl-alkyl C34:0	-0.091	0.7295	0.712	0.058228623
Isoleucine	Isoleucine	-0.158	0.5473	-0.56607	0.089720981
PC ae C36:2	Phosphatidylcholine acyl-alkyl C36:2	-0.193	0.4629	-0.1802	0.553764206
Oleic acid	Oleic acid	-0.2	0.4465	0.183252	0.536574067
PC aa C36:5	Phosphatidylcholine diacyl C36:5	-0.218	0.4059	-0.4694	0.174139565

PC ae C34:3	Phosphatidylcholine acyl-alkyl C34:3	-0.22	0.4008	0.319963	0.31966488
PC ae C40:6	Phosphatidylcholine acyl-alkyl C40:6	-0.261	0.3193	0.741937	0.01875932
C5:1-DC	Glutaconylcarnitine, Mesaconylcarnitine	-0.271	0.3021	-0.26351	0.409158971
2-Hydroxybutyric acid	2-Hydroxybutyric acid	-0.323	0.219	0.250888	0.404894782
Myoinositol	Myoinositol	-0.386	0.1416	0.47233	0.144462991
C16:1 -OH	Hydroxyhexadecenoylcarnitine	-0.447	0.0884	0.809254	0.093896414

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

[b]Univariate logistic regression of each Elanet-selected metabolite adjusted for maternal age, ethnicity, parity, and gravidity.

Supplementary Materials

Supplementary Figure 1: Discrimination of obese and normal groups by Partial Least Squares (PLS) method. **(A)** Discriminant analysis score plot for obese cases (Green) and normal (Red). (**B)** The accuracy of the 10 fold cross-validation of the PLS-DA model. R2 is the sum of squares captured by the model; Q2 is the cross-validation of R2.

Supplementary Figure 2: Selection of metabolites using elastic net regularization. **(A)** Tuning alpha parameter, the parameter representing the degree of mixing between lasso and the ridge regularization. Y-axis is the root mean square error of the 10-fold cross-validation. X-axis is the range of alpha values, with the optimal alpha =0.22. **(B)** Tuning lambda, the parameter controlling the shrunk rate of coefficients in the linear model. Y-axis is the misclassification error of the 10 fold cross validation. X-axis is the range of lambda, with the optimal lambda=0.008. **(C)** The shrinkage coefficients of the metabolites using tuned alpha and lambda.