Molecular Omics

Unveiling Molecular Signatures of Preeclampsia and Gestational Diabetes Mellitus with Multi-Omics and Innovative Cheminformatics Visualizations

Journal:	*Molecular Omics*
Manuscript ID	MO-RES-06-2020-000074.R1
Article Type:	Research Article
Date Submitted by the Author:	07-Aug-2020

Complete List of Authors:	Odenkirk, Melanie; North Carolina State University Stratton, Kelly; Pacific Northwest National Laboratory Gritsenko, Marina; Pacific Northwest National Laboratory Bramer, Lisa; Pacific Northwest National Laboratory Webb Robertson, Bobbie-Jo; Pacific Northwest National Laboratory, Computational Biosciences Division Bloodsworth, Kent; Pacific Northwest National Laboratory Weitz, Karl; Pacific Northwest National Laboratory, Lipton, Anna; Pacific Northwest National Laboratory Monroe, Matthew; Pacific Northwest National Laboratory, Biological Sciences Division Ash, Jeremy; North Carolina State University Fourches, Denis; North Carolina State University Taylor, Brandie; Temple University Burnum-Johnson, Kristin; Pacific Northwest National Laboratory, Biological Sciences Division Baker, Erin; North Carolina State University, Chemistry

ARTICLE

Unveiling Molecular Signatures of Preeclampsia and Gestational Diabetes Mellitus with Multi-Omics and Innovative Cheminformatics Visualization Tools

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

Melanie T. Odenkirk[a], Kelly G. Stratton[b], Marina A. Gritsenko[c], Lisa M. Bramer[b], Bobbie-Jo M. Webb-Robertson[c,d], Kent J. Bloodsworth[c], Karl K. Weitz[c], Anna K. Lipton[c], Matthew E. Monroe[c], Jeremy R. Ash[a,e,f], Denis Fourches[a,e*], Brandie D. Taylor[g*], Kristin E. Burnum-Johnson[h*], and Erin S. Baker[a,j*]

[*]Co-corresponding authors

To fully enable the development of diagnostic tools and progressive pharmaceutical drugs, it is imperative to understand the molecular changes occurring before and during disease onset and progression. Systems biology assessments utilizing multi-omic analyses (e.g. the combination of proteomics, lipidomics, genomics, etc.) have shown enormous value in determining molecules prevalent in diseases and their associated mechanisms. Herein, we utilized multi-omic evaluations, multi-dimensional analysis methods, and new cheminformatics-based visualization tools to provide an in depth understanding of the molecular changes taking place in preeclampsia (PRE) and gestational diabetes mellitus (GDM) patients. Since PRE and GDM are two prevalent pregnancy complications that result in adverse health effects for both the mother and fetus during pregnancy and later in life, a better understanding of each is essential. The multi-omic evaluations performed here provide new insight into the end-stage molecular profiles of each disease, thereby supplying information potentially crucial for earlier diagnosis and treatments.

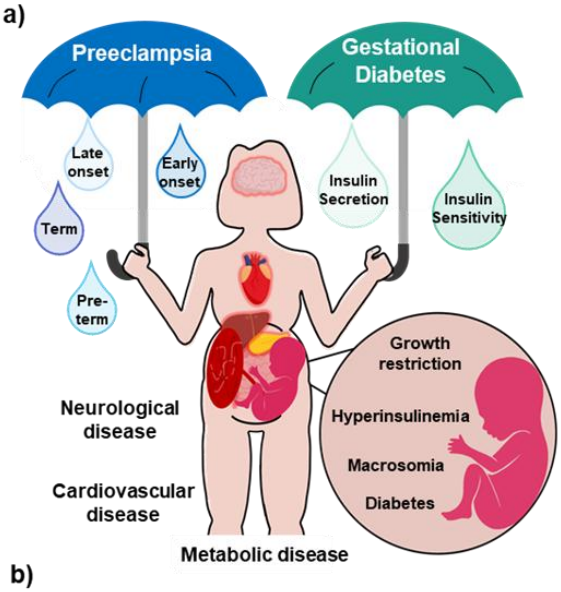
Introduction

Preeclampsia (PRE) and gestational diabetes mellitus (GDM) are two predominant maternal complications that result in an increased likelihood of morbidity and mortality for both the fetus and mother. Additionally, women with PRE or GDM and their offspring have an increased risk of chronic health outcomes such as type II diabetes, hypertension and cardiovascular disease[1, 2]. PRE is a systemic maternal syndrome affecting 3-8% of pregnancies and is a leading cause of maternal mortality and morbidity[3]. Clinically, PRE is defined as the new onset of hypertension and proteinuria or evidence of systemic organ dysfunction after 20 weeks of gestation[3, 4]. The only treatment for PRE is delivery, which often leads to iatrogenic preterm birth and increases the risk of infant morbidity and mortality [3, 4]. Similarly, GDM affects 5-7% of pregnancies[5] and is clinically defined as the onset of glucose intolerance during pregnancy. GDM increases the risk for gestational hypertension, PRE, birth injury, macrosomia, and neonatal hypoglycemia[6, 7], and its' diagnosis is typically evaluated based on risk criteria relating to body mass index (BMI), family history of diabetes, personal history of GDM, and glycosuria[8]. Women who classify as high risk for GDM are administered oral glucose tolerance tests at their first appointment, whereas women who do not meet high risk criteria are not typically examined until 24 to 28 weeks of gestation[8]. The late diagnosis of both GDM and PRE severely limits intervention timelines and allows for substantial disease progression. Rates of both complications are rising throughout the world due partially to increased prevalence of pre-pregnancy obesity and advanced maternal age. Thus, PRE and GDM are serious public health concerns[9, 10].

[a.] *Department of Chemistry, North Carolina State University, Raleigh, NC 27695*
[b.] *National Security Division, Pacific Northwest National Laboratory, Richland, WA 99354*
[c.] *Biological Sciences Division, Pacific Northwest National Laboratory, Richland, WA 99354*
[d.] *Department of Biostatistics and Informatics, University of Colorado, Aurora, CO 80045*
[e.] *Bioinformatics Research Center, North Carolina State University, Raleigh, NC 27695*
[f.] *Department of Statistics, North Carolina State University, Raleigh, NC 27695*
[g.] *College of Public Health, Temple University, Philadelphia, PA 19133*
[h.] *Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, WA 99354*
[i.] *Comparative Medicine Institute, North Carolina State University, Raleigh, NC 27695*

Electronic Supplementary Information (ESI) available: Raw proteomic and lipidomic data are available on MassIVE [https://massive.ucsd.edu]). Normalized, log2 abundances of proteins and lipids are given in supplemental information. The R code for the cheminformatic visualization tools is available on Github: https://github.com/BakerLabNCSU/LipidAnalysis. See DOI: 10.1039/x0xx00000x

For many diseases, early diagnosis is commonly approached using universal biomarkers. Despite the identification of pathway aberrations linked to PRE and GDM, such as inflammation, insulin resistance and mild hyperlipidemia, there are currently no biomarkers for either condition with clinical utility. This is largely attributed to the broad disease definitions of PRE and GDM that encompass multiple novel mechanisms and subtypes (**Figure 1a**). For example, it is well accepted that PRE can be differentiated by both the timeline of disease onset and delivery, while GDM subtypes include mechanistic differences such as insulin



b)

Clinical Data	Control	PRE	GDM
Race	Hispanic 8%, Asian 2%, Black 2%, White 3%, Other 2%, Not specified 85%	Hispanic 8%, Asian 2%, Black 10%, Other 2%, Not specified 80%	Hispanic 2%, Asian 4%, Black 4%, Other 2%, Not specified 89%
BMI	31.9 ± 4.8	34.5 ± 7.0	34.2 ± 6.1
Age	27.5 ± 5.9	27.0 ± 6.7	31.9 ± 6.2
Infant Sex	Female 54%, Male 46%	Female 39%, Male 61%	Female 47%, Male 53%
Gestational age (wks)	38.9 ± 1.6	38.9 ± 1.4	38.5 ± 1.2
Birth weight (g)	3306 ± 576	3331 ± 526	3274 ± 438
Type of Labor	Spontaneous 20%, Augmented 35%, Induced 19%, No labor 15%	Spontaneous 22%, Augmented 31%, Induced 18%, No labor 29%	Spontaneous 11%, Augmented 9%, Induced 27%, No labor 53%

Figure 1. PRE and GDM are prevalent maternal complications leading to serious maternal and fetal complications and morbidities. a) Subtypes of PRE and GDM are shown as raindrops and increased health risks for the mother and fetus later in life are noted around each. b) Synopsis of cohort information for GDM, PRE and Control patients. Race, type of labor and infant sex are shown as patient distributions. BMI, age, gestational age and birth weight are given for the cohort average +/- the cohort's standard deviation.

secretion and insulin sensitivity. However, few studies have focused on molecular differences specific to term PRE (delivery after 37 weeks of gestation)[11, 12] despite this subtype accounting for 60% of PRE cases[11]. GDM not only suffers from mechanistic differentiation of disease subtypes, but lacks comprehensive annotation of the intricate molecular mechanisms underlying the condition as a whole[13]. Moreover, overlapping pathology among common pregnancy disorders further complicates diagnoses, an obstacle best addressed through holistic annotation.

Proteins are essential molecules with a variety of biological functions noted as being dysregulated in PRE and GDM[11, 14]. Lipids and lipid mediators also play integral roles in biological processes including proliferation, apoptosis, migration, metabolism, inflammation and pro-resolving immune responses. Since the placenta relies on fatty acid oxidation for energy and lipotoxicity inhibits trophoblast invasion and placental development, lipids also play important roles in PRE and GDM. However, despite being crucial for fetal development and placental energy processes, the vital roles of these species have been less extensively annotated compared to other omics. Furthermore, since PRE and GDM risk factors such as obesity and maternal age influence protein and lipid processes[15, 16], an examination of the molecular mechanisms occurring and changing in both complications should shed light onto their unique pathophysiology. Thus, a holistic evaluation through sensitive multi-omic analyses and improved data visualization tools is needed to provide key information for a comprehensive understanding of PRE and GDM pathophysiology. In this study, we address these challenges and provide an in-depth profile of term PRE and GDM pregnancy complications through the analysis of proteomic and lipidomic differences in the plasma for 191 pregnant women at admission to labor and delivery (**Figure 1b**).

Experimental

Human Sample Collection. This study utilized stored plasma from Peribank[17], a database and biospecimen repository. Peribank enrolls women at the time of labor and delivery from four facilities in Houston, TX. Trained study personnel approached participants at the time of admission to labor and delivery. After consent was obtained, over 4,700 variables of clinical information were directly extracted from electronic medical records and accompanying prenatal records alongside directed subject questioning. PeriBank was approved by the Institutional Review Board at Texas Children's Hospital and Baylor College of Medicine. For this study, analysis was limited to women with singleton pregnancies and had no current or history of chronic conditions including pre-existing diabetes, heart disease, renal disease or hypertension. A detailed description of cohort profiles for the Control, GDM, and PRE cohorts is presented in Supplemental Table 1 (**Table S1**). The current investigation and Peribank usage of samples was approved by the Institutional Review Board at Temple University (IRB protocol number 26904). All PeriBank samples were strictly collected once consent was properly obtained under rigorous uniform protocols by perinatal and placental pathology-trained personnel. Maternal blood for this study was collected at the time of admission, after admission, and up

to 24 h after delivery. It was then stored at 4˚C until processing.

Proteomic Extraction, Depletion and Digestion. Individual human plasma samples were partitioned and depleted of the 14 highly abundant proteins using a Multiple Affinity Removal Column, 4.6 x 100 mm, Hu-14 (Agilent, Santa Clara, CA). This column was coupled to a 1200 series HPLC (Agilent, Santa Clara, CA) composed of a quaternary pump with degasser and connected to Peltier-cooled autosampler, diode array detector, and refrigerated fraction collector. The unbound, flow-through fraction containing low- and medium- abundance proteins was collected and concentrated. Buffer exchanged to 50 mM Ammonium Bicarbonate (ambic) pH 8.0 was then performed using a 3-kDa molecular mass cutoff Amicon centrifugal filters (Millipore, Burlington, MA) and following the manufacturer's instructions. The protein concentration of the depleted samples was measured by BCA Protein Assay (Thermo Scientific, San Jose, CA) and the volume of each sample was adjusted with 50 mM ambic. All processing steps following depletion were carried out in a 96 1 mL deep-well plate to minimize batch effects associated with processing, utilizing automated protocols on an epMotion 5075 (Eppendorf, Hauppauge, NY).

To perform digestion of the depleted samples, 145 µL aliquots were utilized, giving a starting amount of approximately 100 µg total protein. Aliquots were transferred to a 96-well plate preloaded with urea to achieve a concentration of 8 M urea in the sample solution. 500 mM Dithiothreitol (DTT) (Sigma-Aldrich, St. Louis, MO) was then added to a final concentration of 10 µM and samples were incubated for 1 h at 37°C with constant shaking at 1200 rpm in BioShake iQ Thermal Mixer (Bulldog Bio, Portsmouth, NH) to denature the proteins and reduce disulfide bonds. Reduced cysteine residues were alkylated by adding 1 M iodoacetamide (Sigma-Aldrich, St. Louis, MO) to a final concentration of 40 mM and incubated in the dark at 25°C for 45 min. Samples were then diluted 4-fold with 50 mM ambic, 1 mM $CaCl_2$ prior to the addition of sequence grade TPCK treated trypsin (Trypsin Gold, Mass Spectrometry Grade, Promega) in a 1:50 enzyme to protein ratio. Enzymatic digestion was carried out for 16 h at 25°C with a constant shaking at 1200 rpm in BioShake iQ Thermal Mixer. The reaction was stopped by acidifying the samples with 10% trifluoroacetic acid (Sigma-Aldrich St. Louis, MO) to a final concentration of 0.1%. Peptides were desalted using solid-phase extraction (SPE) via Strata C18-E cartridges formatted in 96-well plate (Phenomenex, Torrance, CA). Peptides were eluted with 80 % ACN, 0.1 % TFA using Preston 100 Positive Pressure Manifold (Phenomenex, Torrance, CA) and concentrated down in a SpeedVac concentrator. After final sample peptide concentrations were evaluated by BCA Assay, they were normalized with 2% acetonitrile, 0.1% formic acid to 0.1 µg/µL prior to LC-IMS-MS analyses.

PRE, GDM and Control depleted sample pools were created by combining equal amounts of digested peptide samples from each group. 300 µg pooled samples were separated on a Waters reversed phase XBridge C18 column (250 mm × 4.6 mm column containing 5-µm particles, and a 4.6 mm × 20 mm guard column) (Waters, Milford, MA) using an Agilent 1200 HPLC System. The sample was loaded on C18 column in solvent A (5 mM ammonium format, pH 10.0) and a 109-min LC gradient with solvent B (5 mM ammonium format, pH 10, 90% acetonitrile) was applied. The LC gradient started with a linear increase of solvent A to 10% B in 3 min, then linearly increased to 30% B in 86 min, 10 min to 42.5% B, 5 min to 55% B and another 5 min to 100% solvent B. The flow rate was 0.5 mL/min. A total of 96 fractions were collected into a 96 deep well plate throughout the LC gradient. These fractions were concatenated into 24 fractions by combining 4 fractions that were each 24 fractions apart (i.e., combining fractions #1, #25, #49, #73; #2, #26, #50, #74; and so on). For proteome analysis, each concatenated fraction was dried down and re-suspended in 2% acetonitrile, 0.1% formic acid to adjust the peptide concentration to 0.1 µg/µL prior to LC-IMS-MS analyses.

Lipid Extraction. Extraction of the unbound lipids was performed using a modified Folch extraction[18, 19]. For the plasma lipid extraction, 25 µL of plasma was transferred into a 2.0 mL tube where 600 µL of -20°C 2:1 chloroform/methanol were added. Each sample was vortexed for 30 sec then transferred into a shaker at 22°C for 60 min at 600 rpm. The samples were vortexed again for 30 sec and 125 µL of water was added to induce a phase separation. The samples were gently inverted several times, placed at room temperature for 5 min and then centrifuged at 10,000 x g for 5 min at 4°C and put on ice to maintain the clear phase separation. Finally, 200 µL of the top polar layer was removed, dried in a speedvac, and stored at -80°C for later analysis of polar metabolites, while 350 µL of the bottom nonpolar layer was removed, dried in a speedvac, and stored at -20°C in 250 µL of 2:1 chloroform/methanol for lipid analyses. Prior to analysis, the total lipid extracts were dried down and then reconstituted in 100 µL of MeOH. To generate pooled case and control samples for LC-IMS-MS analyses, 5 µL aliquots from each reconstituted control plasma sample were removed and combined.

LC-IMS-MS Analyses

Proteomic Analyses. Analysis of the 191 human proteomic extracted plasma samples was performed on an in-house built instrument that couples a 1-m ion mobility separation with an Agilent 6224 TOF MS upgraded to a 1.5 meter flight tube providing resolution of ~25,000 in enhanced dynamic range mode[20]. A fully automated in-house built 2-column HPLC system equipped with in-house packed capillary columns was used with mobile phase A consisting of 0.1% formic acid in water and phase B comprised of 0.1% formic acid in acetonitrile [21]. A 60-min gradient with shorter columns (30 cm long columns with an o.d. of 360 µm, i.d. of 75 µm, and 3-µm C_{18} packing material) was used with the IMS-MS. The gradient linearly increased mobile phase B from 0 to 60% until the final 2-min of the run when B was purged at 95%. 5 µL of each sample was injected for both analyses and the HPLC was operated under a constant flow rate of 0.4 µL/min for the 100-min gradient and 1 µL/min for the 60-min gradient. IMS-MS data were collected from 100-3200 *m/z*.

Lipidomic Analyses. For the 191 human lipidomic extracted plasma samples, the Agilent 6560 IM-QTOF MS platform (Santa Clara, CA) outfitted with the commercial gas kit (Alternate Gas Kit, Agilent) and a precision flow controller (640B, MKS Instruments) was utilized. IMS-MS data were collected in both positive and negative mode from 50-1700 *m/z* with a cycle time of 1 sec/spectra to increase the signal of low abundance species. For the LC analyses, a Waters Aquity UPLC H class system was used. 10 µL of each sample was injected onto a reversed phase Waters CSH column (3.0 mm x 150 mm x 1.7 µm particle size). The lipids in the mixture were separated over a 34 min gradient (mobile phase A: acetonitrile/water (40:60) containing 10 mM ammonium acetate; mobile phase B: acetonitrile/ isopropyl alcohol (10:90) containing 10 mM ammonium acetate) at a flow rate of 250 µL min^{-1}. The gradient and column wash are provided in **Table S2**. The resolution of isomeric lipid species noted herein (A and B pairs) were largely resolved due to retention time using this LC method, while IMS collision cross section values were utilized to increase confidence of the identified lipids.

Quantification and Statistical Analysis

Data Pre-processing. Statistical analysis for both the proteomic and lipidomic studies for the Complication versus Control patients was completed in MATLAB (version 9.1). In the proteomic study, 14,429 unique peptides were observed from the LC-IMS-MS analyses. These peptides correspond to 1,093 proteins using the identification criteria of at least 2 peptides detected per protein where at least one peptide must be unique as outlined in **Tables S3 and S4**. In the lipidomic study, 288 unique lipids were identified as shown in **Table S5**. Potential outliers in the proteomic and lipidomic analyses were identified using the RMD-PAV algorithm [22], and were then confirmed by Pearson correlation and subsequently removed from the dataset (**Figure S1**). Three patients in negative ion lipidomics (3 PRE patients), one patient in positive lipidomics (GDM patient) and one in proteomics (Control patient) were removed following outlier assessment. Lipids and peptides with inadequate data for either qualitative or quantitative statistical tests were also removed [23]. Statistical analysis of proteins and lipids for the complication comparisons (PRE vs. GDM) was completed with MetaboAnalyst (version 4.0) [24] with results shown in **Table S6**.

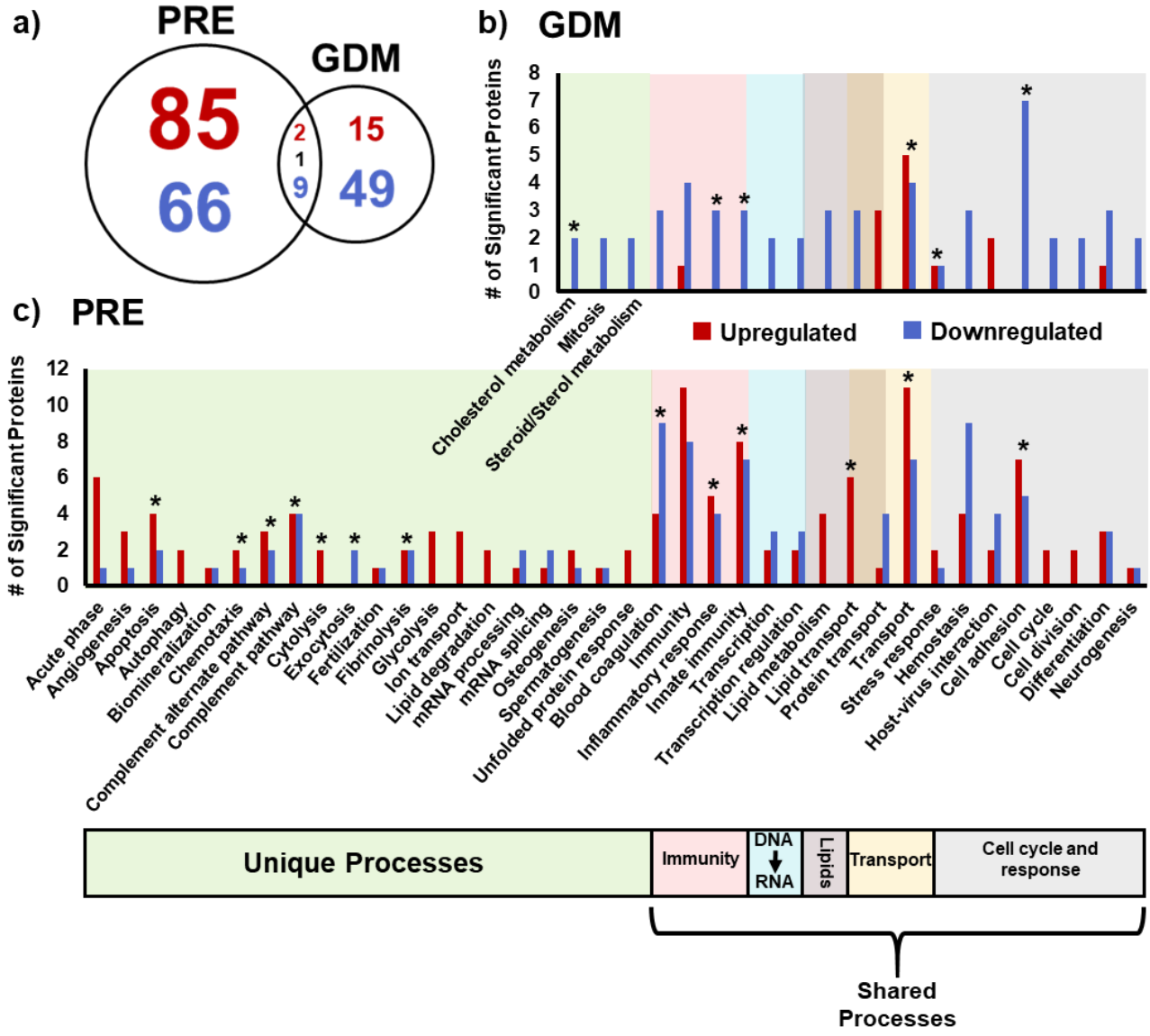
Protein Quantification and Statistics. Peptides and proteins were evaluated using Analysis of Variance (ANOVA) with a Dunnett test correction [25] and a Bonferroni-corrected g-test [26] to compare the GDM and PRE patients to the Control patients (**Supplemental Tables S3 and S4**). Following outlier removal, ANOVA comparisons for Complication vs. Control comparisons consisted of 97 control patients, 48 term PRE and 45 GDM patients to give statistical powers of 78.6% for GDM vs. Control and 80.3% for PRE vs. Control at an effect size of 0.5 (medium). For g-test analyses, power at a large effect size (0.5) was determined to be 60.9% for GDM vs. Control and 61.6% for PRE vs. Control comparisons. To perform signature-based protein quantification, BP-Quant with the background probability of a zero signature set to the default of 0.9 was used [27]. Only significant proteins with greater than 2 peptides/protein and at least 1 unique peptide for identification were utilized in this study (**Table S4**). Proteins with a Dunnett multiple testing correction adjusted ≤ 0.05 were considered significant for discrimination. Of filtered proteins, 227 proteins were significant in at least one Complication versus Control comparison. Enrichment analysis of significant proteins for each Complication versus Control comparison was assessed using STRING (version 11.0)[28]. Biological process associations below a false discovery rate of 0.05 are noted with GO annotation, count in gene set, and false discovery rate expanded upon in **Table S7**.

Lipid Quantitation and Statistics. Positive and negative mode lipidomic data was also evaluated using ANOVA with a Dunnett correction (**Table S5**) [25]. Lipids with adjusted $p \leq 0.05$ were considered relevant for Complication versus Control discrimination. ANOVA for Complication vs. Control comparisons in positive ion lipidomics data consisted of 98 control patients, 48 term PRE and 44 GDM patients following outlier removal. This yielded statistical powers of 78.1% for GDM vs. Control and 80.4% for PRE vs. Control at an effect size of 0.5 (medium). Following outlier removal for negative ion lipidomics, three PRE patients were removed to yield sample sizes of 98, 45 and 45 for control, PRE and GDM groups; respectively. Statistical power for both of these analyses was 78.7% at a medium effect size (0.5). Of the 288 unique lipid identifications, 120 were significant in at least one Complication versus Control comparison.

Pre-processing, Molecular Descriptor, Clustering, and Circular Dendrogram. Proteins with known biological processes and pathways of interest for GDM and PRE were evaluated using STRING[28], UNIPROT[29], and KEGG[30] online tools. Lipid relationships, which are traditionally probed with heatmaps and spreadsheets, were expanded upon in this study to assess biological linkages using a cheminformatics-powered structural clustering of head groups and fatty acyls. Head group clustering was performed based on chemicals initially represented as SMILES strings[31], then converted into 2D standardized structures, and further characterized using an ECFP_6 fingerprint. The hierarchical clustering was done using Euclidean distance and average linkage method using *fingerprint* and *ggtree* packages in R (Version 3.6.0). For lipids with multiple potential identifications, a representative SMILES was chosen to denote all possible identifications. Fatty acyl tail composition was also used to relate lipids. For our analyses, most *sn*-1 and *sn*-2 fatty acyl positions were unknown, so all possible positions were considered to account for potential backbone position effects. For lipids with multiple identities, lipids were partitioned into all possible identifications to visualize fatty acyl effects. Visualization of adjusted p-values was utilized in this clustering technique due to their set cut-off for significance (adjusted p-value ≤ 0.05) using the *pheatmap* package in R. Statistical upregulation is represented by red and downregulation by blue with darker colors indicating a more significant adjusted p-value. Fold changes of significant lipids were also investigated to emphasize analytes with the greatest discrepancy in disease and control groups.

Results and Discussion

In this study, we evaluated the proteomic and lipidomic profiles in plasma of 191 pregnant women from the Peribank[17] bio-repository. All plasma samples were collected at admission to labor and delivery. Samples were obtained from 48 women with term PRE, 45 women with GDM and 98 healthy control women. All women were singleton pregnancies with no current or history of chronic conditions including pre-existing diabetes, heart disease, renal disease or hypertension. A synopsis of cohort profiles for GDM, Control and PRE is presented in **Figure 1b** with full clinical information on patients that participated in this study detailed further in **Table S1**. The majority of patients for all three cohorts were Hispanic with variation across groups such as age, infant sex, and type of labor (**Table S1**). To provide sensitive analyses of this cohort, we utilized a multi-dimensional platform coupling liquid chromatography, ion mobility spectrometry and mass spectrometry (LC-IMS-MS) separations. This platform was capable of molecular speciation due to the three orthogonal techniques, providing simultaneous evaluations of polarity, structure and mass for each species studied and detected. Proteomic and lipidomic specific extractions were performed on the plasma samples and each resulting sample was analyzed with the LC-IMS-MS platform to provide protein and lipid quantification information for statistical analyses comparing the PRE, GDM and the Control groups. Initially, each omic evaluation was performed separately and novel cheminformatics visualization tools were created to assess the structural and biological relationships for the identified molecules. After the individual assessments, the proteomic and lipidomic results were rolled together in a multi-omic



assay illustrating unique molecular mechanisms for each disease.

Protein Associations Specific to PRE and GDM

In the bottom-up proteomic analyses of the 191 pregnant women, 14,429 unique peptides were detected in the plasma samples and rolled into 1,093 proteins using the identification criteria of at least 2 peptides detected per protein where at least one peptide must be unique (**Table S3 and S4**). Of these proteins, analysis of variance (ANOVA) evaluations identified 163 and 76 statistically significant proteins ($\alpha \leq 0.05$) for PRE and GDM when compared to Control patients after a Dunnett multiple comparisons correction[25] (**Figure 2a**). Despite shared placental dysfunction in GDM and PRE, only a slight overlap was observed in the statistically significant proteins as shown in the Venn Diagram in Figure 2a. Furthermore, of the statistically significant proteins found for the two complications (**Figure 2a**), PRE illustrated more upregulation, while GDM was predominantly downregulated compared to the Control group. Specifically, of the 103 significantly upregulated proteins for both conditions, 85 were unique to PRE, 15 were unique to GDM, 2 were upregulated in both, and only 1 was upregulated in GDM but downregulated in PRE. A total of 125 proteins were observed to be significantly downregulated for both conditions with 66 unique to PRE, 49 to GDM, 9 downregulated in both and the 1 protein which was upregulated in GDM but downregulated in PRE.

The statistically significant proteins for each complication were further explored by evaluating both their unique Uniprot biological functions[29], STRING [28] functional enrichment analysis (**Figure 2b and Figure 2c**) and associated KEGG[30] pathways (**Figure 3**). Again, very different functions and mechanisms were observed for each complication. For GDM, a considerable decrease in protein expression was observed as 76% of its' statistically significant species were downregulated (**Figure 2a**) including transport, cellular adhesion, blood coagulation and cholesterol metabolism, all of which were found to be functionally enriched at a false discovery rate less than 0.05 using STRING (**Figure 2b, Table S5**). Notably, this proteome-wide depression affected starch and carbohydrate metabolism enzymes lysosomal alpha-glucosidase (LYAG) and beta-1,4-galactosyltransferase 1 (B4GT1) which were both downregulated in GDM patients (**Figure 3**). In addition to the dysregulation in carbohydrate metabolism we also found a significant change in proteins related to beta cell function such as glutathione synthetase (GSHB) and receptor-type tyrosine-protein phosphatase kappa (PTPRK). These proteins could subsequently trigger elevated maternal glucose, the criteria currently used for GDM diagnosis. The reduced protein synthesis and glucose intolerance observed in GDM may also be related through protein/nucleic acid deglycase DJ-1 (PARK7), an enzyme that has been reported to regulate transcription mechanisms of mRNA and maintain glucose homeostasis and diabetes onset[32]. This protein was observed to be downregulated in GDM with a $\log_2$ fold change of -0.54 and adjusted p-value of 0.0049.

In PRE, both upregulation and downregulation were observed for the statistically significant proteins as 53% of differentially expressed proteins were upregulated (**Figure 2a**). The PRE biological function analyses showed the most dysregulation in processes related to innate immunity, inflammatory response and blood coagulation and complement activation, which were all determined to be functionally enriched at a false discovery rate less than 0.05 (**Figure 2c, Table S5**). Further evaluation illustrated that a majority of the proteins associated with these processes were linked to the blood coagulation and complement cascade (BCC), a pathway that serves as a moderator of innate immunity (complement) and fibrin clot formation (coagulation)[33]. BCC pathway activation through the complement and coagulation cascade was largely observed to be upregulated. Complement activation promotes innate immunity responsible for maintaining host homeostasis, promoting inflammation and enhancing immune system pathogen defense[34]. Dysregulation of species associated with this process is agreeable with instances of immune dysregulation commonly reported in PRE studies[35]. Plasminogen activator inhibitor 1 (PAI1) was among the most differentially upregulated proteins in PRE patients (-1.25 $\log_2$ fold change and adjusted p-value of 0.00277). Since this enzyme serves to downregulate fibrinolytic activity, its upregulation could induce endothelial dysfunction by means of tissue and/or blood vessel damage, thereby activating the coagulation cascade in the BCC pathway. Taken together, BCC activation of complement and coagulation cascades is consistent with reduced placental perfusion and hemodynamic placental dysfunction potentially triggering hypertension while complement activation reflects a state of immune dysregulation commonly reported in PRE cases[33, 36]. Both of these effects could serve as a driving force of inflammation as well as dysregulation of metabolic processes and dyslipidemia. The glycolysis-related enzymes alpha-enolase (ENOA), glyceraldehyde-3-phosphate dehydrogenase (G3P), phosphoglycerate kinase 1 (PGK1) and 3-mercaptopyruvate sulfurtransferase (THTM) were also found

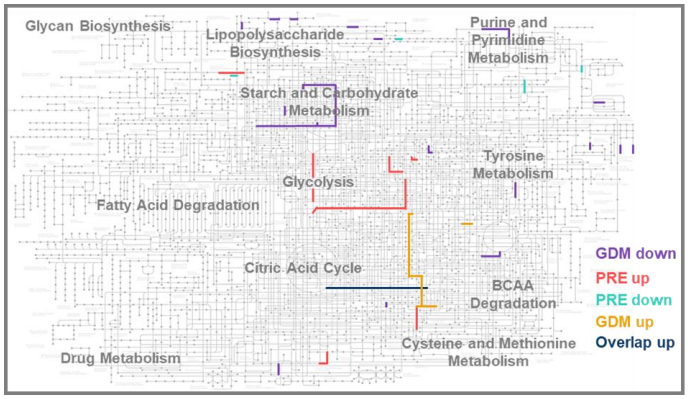


Figure 3. KEGG metabolic pathway analyses illustrate specific areas of up- and downregulation for PRE and GDM. Namely, GDM exhibited downregulation in starch and carbohydrate metabolism and purine and pyrimidine metabolism, while PRE showed upregulation in glycolysis specific enzymes.

Figure 2. Statistically significant proteins detected for PRE and GDM vs. controls. **a**) Th GDM and their slight overlap. **b**) Each statistically significant protein was then associat decrease in functions such as cellular adhesion, immunity, and lipid transport, and b exhibited an overall increase in cellular adhesion, immunity and lipid transport, in immunity. Functional enrichment of biological processes with false discovery rates belo proteins are shown in red while downregulated are in blue.

to be uniquely upregulated in PRE (**Figure 3**), promoting the hypothesis of metabolic abnormality of the placenta, a proposed mechanism unique to the late onset PRE subtype[12]. Interestingly, although transport had both up- and downregulation in both conditions, lipid transport was only downregulated in GDM and only upregulated in PRE (**Figure 2b and 2c**). These findings were further investigated in the following lipidomic evaluations.

Lipid Associations Specific to PRE and GDM

In the lipidomic evaluations of the 191 patients, 288 unique lipids were identified from three main lipid categories: 46 sphingolipids, 72 glycerolipids and 170 glycerophospholipids. Previously, evaluating the biological implications for each lipid species has largely relied on the manual elucidation of lipid relationships. However, with hundreds of lipids detected in lipidomic studies, this can be exceedingly challenging and result in an incomplete picture of disease pathophysiology. Challenges also arise when annotating lipids with existing pathways as broad lipid classifications are often rolled together through shared structural elements, such as head group, even when different directionality of species is observed. For example, many studies will indicate that phospholipids, or even more specifically phosphatidylethanolamine (PE), are only changing in one direction even when different species have conflicting directionalities such as PE(16:0_22:4) being downregulated in PRE while PE(18:0_20:3) is upregulated. Since structural annotations of polar metabolites has previously shown valid association to biological function[37], we created a novel lipidomic analysis workflow to more readily assess how head group and fatty acyl structural elements of lipids trigger specific biological responses (**Figure 4**).

In our lipidomic workflow, first the 288 lipids were identified from the LC-IMS-MS analyses (**Figure 4, Step 1**). All identified lipids were then structurally evaluated by mapping their SMILES[31] strings to an ECFP_6 fingerprint that encodes the presence/absence of key functional groups and 2D sub-structural motifs of each (**Figure 4, Steps 2A and 2B**). ANOVA evaluations were simultaneously performed for the PRE and GDM patients by comparing them to the Control group to acquire p-values and $\log_2$ fold changes for each lipid (**Figure 4, Step 2C**). To begin grouping structurally similar lipids, hierarchical clustering then utilized the ECFP_6 fingerprint, Euclidean distance, and an average linkage method to generate a circular dendrogram for visual evaluation of the changes occurring (**Figure 4, Step 3A**). Fatty acyl variations were also assessed by grouping all lipids with the same fatty acyl components into heatmaps to allow a visual assessment of their statistical significance (**Figure 4, Step 3B**).

Upon assessing the lipidomic changes for PRE and GDM, 81 and 28 unique lipids were found to be statistically significant ($\alpha \leq 0.05$) with a Dunnett correction for multiple comparisons [25] as shown by the Venn diagram in **Figure 5a**. Similar to the proteomic analyses, only a slight overlap was observed for the statistically significant lipids with 11 lipids shared between conditions. The results also showed that PRE illustrated significant lipid dysregulation with 39 unique lipids statistically upregulated and 42 downregulated, while GDM exhibited fewer with 15 unique lipids statistically upregulated and 13 downregulated. The lipids were then evaluated based on head and fatty acyl groups (**Figure 5b**) using the previously described workflow to further assess the role of each component. While the overall comparison of all identified lipids is very important to clearly elucidate molecular effects, in the proceeding analyses only lipids found to be statistically significant in both diseases were assessed for their head group and fatty acyl structural similarity **(Figures 5c and 5d).**

Upon evaluation of the dendrograms for the three categories of lipids detected in GDM and PRE, the most significant variation for both complications occurred in the glycerophospholipid category. Since these lipids are the most abundant membrane lipids and serve integral roles in membrane transport, enzyme activities and extracellular signaling, this finding also relates to the lipid transport dysregulation observed in the proteomic analyses for both complications. The three lipid categories were then further broken down into specific lipid head groups, resulting in three sphingolipid classes (ceramides, galactose/glucose ceramides, sphingomyelins), two glycerolipid classes (diacylglycerols (DG) and triacylglycerols (TG)), and five glycerophospholipid classes (phosphatidylethanolamine (PE), phosphatidylserine (PS), phosphatidylcholine (PC), phosphatidylinositol (PI), and phosphatidylglycerol (PG)). The PE and PC classes were further broken down into their fatty acyl linkages of either the alkenyl ether linkage (plasmalogens, PE P- and PC P-) or the alkyl ether linkage (PE O- and PC O-) subclasses since each have unique biological roles. In PRE, all PIs were downregulated and all DGs upregulated. This result matched a previous study of placental

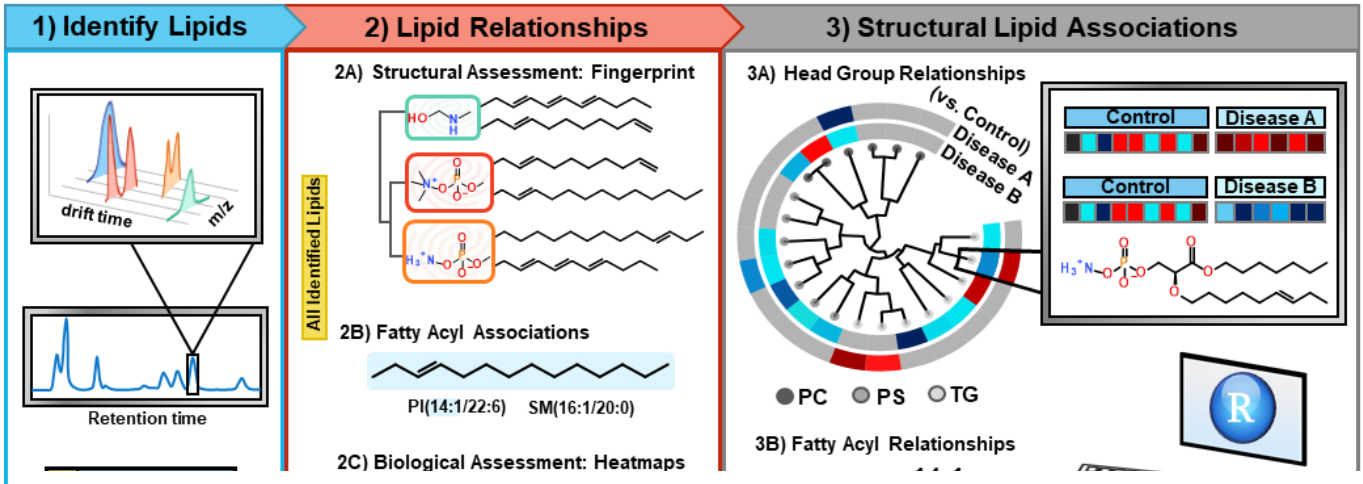


Figure 4. The lipidomics data analysis workflow utilized in this study. **1**) Lipids were first identified from the LC-IMS-MS analyses based on LC retention time, *m/z* and IMS collision cross section. **2A**) Next, the identified lipids were clustered using an ECFP_6 fingerprint for building head group associations. **2B**) Lipid identifications were further related by shared fatty acyl groups. **2C**) Heatmaps and statistical tests were also performed in this step to evaluate the significance of each lipid with regard to the PRE, GDM and Control patients. **3**) Finally, the structural information and statistical significance of each lipid was combined to generate **3A**) a circular dendrogram for head group associations and **3B**) heatmaps of fatty acyl variation allowing concurrent visualization of all conditions and their statistical significance based on either adjusted p-value or $\log_2$ fold change. Upregulated and downregulated lipids are shown in red and blue with the darker colors indicating a more significant adjusted p-value.

lipidomics, linking elevated DGs in PRE to acyl CoA, a placental source of energy[38]. PRE also exhibited changes specific to lipid species in the PC, TG and SM classes. Similar to the DGs, TGs were mainly upregulated except for 3 species. This trend was also noted in serum samples of PRE patients with significant TG dysregulation as early as 10 weeks gestation[39]. On the other hand, PCs were largely downregulated with the exception of lyso species, which were upregulated and may suggest enhanced PC degradation[40]. SMs in PRE were both up and downregulated.

For GDM, most lipid species from various head groups were largely downregulated except for PE and PE plasmalogen (PE P-) species, which were all upregulated. The significant downregulation of lipids in GDM is potentially indicative of free fatty acid accumulation (FFA), a phenomenon that has previously been associated with insulin resistance and hyperinsulinemia [41]. Since the head group trends noted for both PRE and GDM did not always correlate, head group composition was concluded to only be a partial contributor to lipid dysregulation. Furthermore, since previous studies have demonstrated that enzymes have enhanced selectivity for fatty acyl length, double bond position, and sn-positioning on the lipid backbone, fatty acyl composition and corresponding changes were also assessed [42].

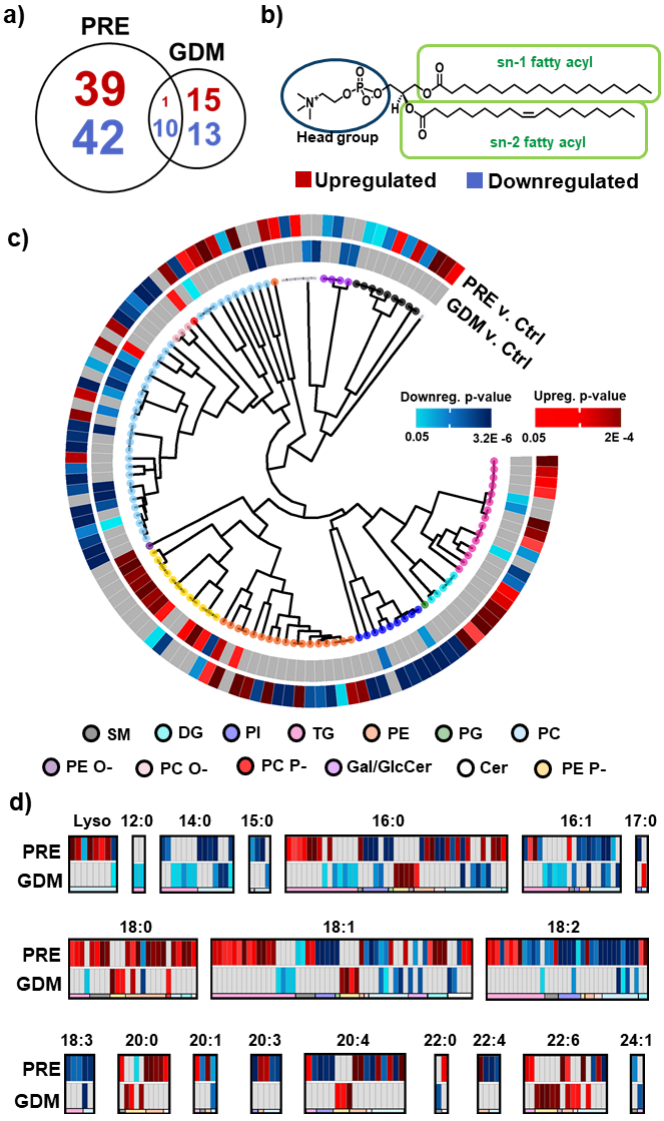


Figure 5. Statistically significant lipids detected for PRE and GDM. **a**) The Venn diagram illustrates 81 and 28 unique statistically significant lipids observed for PRE and GDM with little overlap between conditions (11 lipids). **b**) The lipids were then evaluated based on head groups (circle) and fatty acyl groups (green rectangles). **c**) Circular dendrogram from the hierarchical clustering of the head groups of significant lipids visually illustrates great differences between lipid responses for PRE and GDM. **d**) Evaluation of fatty acyl group presence also showed differences between the two diseases. All identified but insignificant lipids are shown in grey and statistically significant upregulated and downregulated lipids are shown in red and blue with the darker colors indicating

To probe fatty acyl composition in all of the identified lipids, an additional visualization tool that parses out lipids with common fatty acyl groups was utilized (**Figure 4, Step 3B**). As most lipids typically consist of two or more fatty acyl groups with a specific glycerol backbone position, this technique relies on the identification of the specific fatty acyl moieties by MS/MS analysis. In the fatty acyl analyses of PRE and GDM, lipids specifically containing 12:0, 14:0, 15:0, 18:3, 22:4 or 24:1 fatty acyls were found to be downregulated across both disease states and 20:0 and 22:6 were shown to be upregulated (**Figure 5d**). However, the head group/fatty acyl combinations were often different in the two diseases. For example, 22:6 fatty acyls were upregulated in GDM for PE P-, PC and PE O- species, while in PRE it was for TG and PE species. Furthermore, myristic acid or 14:0, which has previously been characterized as being negatively associated with 22:6 [43], was also downregulated in both GDM and PRE but not for the same lipid species. Furthermore, LysoPC(0:0_14:0) was the only exception to the upregulation of all other lyso PC species in PRE. Other fatty acyl trends were also observed within the head groups for the complications, including saturated 18-carbon fatty acyls being upregulated in PRE and as unsaturation increased, downregulation heightened. For example, all 18:0 TG species were upregulated and all TGs with 18:3 species were downregulated. Interestingly, GDM only exhibited a few significant downregulated TGs all consisting of shorter, saturated fatty acyl groups (e.g. 12:0, 14:0, and 16:0). This trend was surprising as type II diabetes is commonly associated with an upregulation of circulating TG species which compensate for glucose variations [44]. This trend however may depend on fatty acyl chain lengths as medium chain fatty acids (MCFAs) or those with > 6 carbons and ≤ 12, do not require enzymes for binding, transport or transmembrane location[45]. MCFA-containing TGs therefore could serve as quickly available energy sources for the fetus and placenta to further evidence the dysregulation of placental energy processes previously reported in GDM and PRE studies[45]. As the only upregulated head groups in GDM, the PE and PE-plasmalogen (PE P-) species overexpression was mainly observed in lipids containing long chain polyunsaturated fatty acids (LC-PUFAs). LC-PUFAs have 18 or more carbons and at least 2 double bonds. Plasmalogen phospholipids are commonly characterized with an *sn-2* located LC-PUFA species as the preferential oxidation of the vinyl ether bond protects LC-PUFA species from oxidation and readily involves these lipids in

signaling processes following phospholipase release[46]. As neither of these fatty acyl groups can be synthesized by the fetus, the entire supply of LC-PUFAs, which are vital for neuronal and visual development are transferred across the placenta. Therefore, upregulation of PE and PE P- species containing LC-PUFAs could suggest these fatty acids are being preferentially shunted to PE P- species. This would alter LC-PUFA transfer and lipid signaling processes in placental and fetal metabolism.

Multi-omic Associations

Currently, associations between lipids and proteins lack in literature due to an abundance of isomeric lipids that are incredibly difficult to differentiate, therefore, attempts to relate lipid and protein findings is often limited by broad classifications. In our analyses, we initially assessed proteins that have established lipid associations. Of our statistically significant proteins, only 16 of the 227 had established associations to lipid binding, lipoprotein complexes and lipid metabolic processes (**Table 1a**) and each was assessed for its significance in GDM and PRE. Several very interesting findings were observed. First, alpha-synuclein (SYUA), which protects insulin signaling from saturated-fat-related insulin resistance through the redirection of free fatty acids for transport[47] was

Table 1. Multi-omic associations for PRE and GDM complications. a) Statistically significant lipoproteins in PRE and GDM with GO annotations related to lipid metabolism and/or transport and b and c) Protein and lipid pathway relationships of interest for (b) PRE and (c) GDM.

a)

Code[a]	Description	Sig.[b,c]	p-value	Lipid association
*LBP	Lipopolysaccharide-binding protein	+	6.9E-6	Binds lipid A moiety of bacterial lipopolysaccharides
*APOC1	Apolipoprotein C-I	+	7.7E-4	Inhibits LOL and VLDL receptor binding, associates with HDL
*APOC2	Apolipoprotein C-II	+	7.5E-4	Component of chylomicrons, VLDL, LDL, and HDL
*APOC3	Apolipoprotein C-III	+	4.7E-3	Component of triglyceride-rich VLDL and HDL
*APOF	Apolipoprotein F	++	0.024	Associates with LDL, regulates cholesterol transport
CHM4A	Charged multivesicular body protein 4a	--	0.036	Multivesicular bodies (MVB) formation
ANXA3	Annexin A3	++	4.7E-3	PhospholipaseA2 inhibitor, Binds phospholipids (unknown specificity)
*ANXA5	Annexin A5	--	0.017	Phospholipid binding (head group specificity unknown)
PEBP1	Phosphatidylethanolamine-binding protein 1	--	2.5E-3	Preferentially binds PE containing lipids (PE, Lyso PE, PE O-, PE P-) with lower affinity for PC and PI species
*IPSP	Plasma serine protease inhibitor	++	8.2E-5	Enhanced inhibition activity when bound to PC containing lipids
*SYUA	Alpha-synuclein	++	0.034	Phospholipid binding (head group specificity unknown)
*AACT	Alpha-1-antichymotrypsin	+	3.8E-6	Regulation of lipid metabolism, specificity of lipids unknown
APOM	Apolipoprotein M	-	0.019	Transport of 14:0, 16:0, 18:0, Vitamin A, all-trans and 9-cis-retinoic acid
PRDX6	Peroxiredoxin-6	-	3.9E-4	Sn-2 fatty acyl reduction of phospholipids
PHLD	Phosphatidylinositol-glycan-specific phospholipase D	-	3.9E-3	Hydrolyzes inositol phosphate linkage in proteins anchored by PI glycans
FACR1	Fatty acyl-CoA reductase 1	+	0.032	Catalyzes reduction of C16 or C18 fatty-acyl CoA to fatty alcohols

b) PRE-Specific Species

	Molecule[a]	Sig.[b,c]	p-value	Biological Process
Upregulated	PE(20:0_22:6)	-	1.6E-3	20:0 LCFA upregulation
	PE(18:1_20:0)	-	5.0E-3	
	*PGK1	-	3.7E-5	Glycolysis
	*ENOA	-	2.9E-3	
	*IC1	-	7.7E-4	Blood coagulation and clotting cascade, innate immunity
	*CRP	-	4.1E-4	
	*C1S	+	7.1E-4	
	*CO5	+	7.3E-5	
Downregulated	*CO4B	-	4.7E-3	
	FA12	-	3.9E-4	
	FA7	--	1.7E-3	
	PHLD	-	3.9E-3	PI downregulation
	PI(18:1_18:2)	-	4.0E-4	
	PI(18:1_20:4)_A	-	3.6E-3	
	PI(18:1_20:4)_B	-	5.0E-4	

c) GDM-Specific Species

	Molecule[a]	Sig.[b,c]	p-value	Biological Process
Upregulated	PE(P-20:0_18:1)	+	8.8E-3	Plasmalogen and 22:6 LC PUFA upregulation
	PE(P-18:0_22:6)	+	1.0E-3	
	PE(P-20:0_22:6)	+	6.8E-3	
	PE(P-18:1_22:6)	+	0.027	
	PC(P-18:0_22:6)	+	0.031	
	PC(17:0_22:6)	+	3.2E-3	
Downregulated	PEBP1	--	2.5E-3	
	TG(12:0_16:0_18:1)	-	0.042	TG downregulation
	TG(14:0_16:0_18:1)	-	0.030	
	TG(14:0_16:0_18:2)	-	0.046	
	PDIA6	--	8.1E-3	Redox homeostasis, protein folding
	PDIA5	--	0.034	
	LYAG	--	0.016	Carbohydrate metabolism and glycosylation
	EXT1	--	0.026	
	B4GA1	--	0.017	

[a] Previously reported lipoprotein associations for PRE and GDM are shown with an asterisk (*).

[b] G-test significance is noted with ++ and - - , whereas ANOVA significance is denoted by + and - for the upregulated and downregulated proteins.

[c] Grey illustrates proteins statistically significant in PRE, while blue denote GDM.

upregulated in GDM. Next, annexin A5 (ANXA5), which reflects a hypercoagulable state that impairs blood coagulation processes and fetal growth, was found to be downregulated for PRE, consistent with other studies[14, 48]. Lipid involvement in inflammation in PRE was evidenced through the upregulation of both the lipopolysaccharide-binding protein (LBP), a marker of fetal inflammation noted in **Table 1a**, and C-reactive protein (CRP), a general marker of inflammation, shown in **Table 1b**[49, 50].

Multi-omic comparisons were also used to identify protein and lipid species that uniquely characterize each complication at admission to labor and delivery (**Table 1b and 1c**). Both PRE and GDM demonstrated dysregulation of enzymes associated with phospholipid variation, shedding light on specific head group and fatty acyl trends. Specifically for GDM, phosphatidylethanolamine-binding protein 1 (PEBP1), an enzyme with preferential binding of PE lipid species, and peroxiredoxin-6 (PRXD6) which has been shown to selectively reduce the *sn-2* position of phospholipids were significantly downregulated and serve as unique candidates for GDM diagnosis (**Table 1c**). Notably, a majority of PE P- upregulation was within 22:6-containing species and this trend was extended to other lipid classes with PC(17:0_22:6) also being differentially expressed solely in GDM. Downregulated TG species were also uniquely observed in GDM patients. Interestingly, these species all contain MCFA groups, which can cross the placenta without carrier proteins. Protein disulfide-isomerase A6 (PDIA6) and protein disulfide-isomerase A5 (PDIA5), which affect disulfide bonds, were also uniquely largely downregulated in GDM and could be partial contributors to the altered proteome expression observed. Other notable proteins uniquely observed for GDM were the glycosylation enzymes (LYAG, EXT1 and B4GA1), which are involved in carbohydrate metabolism.

Unlike GDM, PRE was largely characterized by protein dysregulation in enzymes involved in innate immunity with a majority of uniquely dysregulated species serving roles in the BCC pathway, again affirming this pathway and its relationship to PRE diagnosis (**Table 1b**) [33, 36]. Phosphoglycerate kinase 1 (PGK1) and alpha-enolase (ENOA) suppression further defined dysregulation of placental energy processes through their activity in glycolysis. Lipid dysregulation was also observed as characteristically defining PRE. PE species containing 20:0 were uniquely upregulated, a finding of interest considering the upregulation of 22:6 in GDM patients. The most notable multi-omic connection for PRE was that phosphatidylinositol-glycan-specific phospholipase D (PHLD), an enzyme responsible for the hydrolysis of proteins from PI-glycans was downregulated in PRE, giving context to the dysregulation of the PI species observed. Downregulation of PI(18:1_20:4) isomers resolved by LC and IMS were defining characteristics of PRE. Comparison of these species to a dataset collected with Oz-ID allowed for double bond characterization, revealing both isomers shared the same double bond positions of 18:1 (n-9) and 20:4 (n-6,9,12,15)[51]. Therefore, we hypothesize the PI isomers could have different headgroup arrangements due to the fully resolved LC separation and different IMS sizes but further evaluation is needed to confirm this [52]. Ultimately in this study, by utilizing both the lipid and protein relationships for both GDM and PRE, while limited, we found strong evidence for head group and fatty acyl specificity for certain proteins.

Conclusions

To evaluate PRE and GDM, multi-dimensional analyses were performed on a clinically diverse cohort. The proteomic analyses for GDM and PRE illustrated specific proteins predominantly unique to both complications. Significant proteins in GDM correlated with beta cell productivity and carbohydrate metabolism. Given the well-documented association of insulin dysregulation and GDM, these findings were expected. For PRE, proteins associated with blood coagulation and complement cascade (BCC), were of significance. These results showed unique molecular signatures matching the known dysregulation of GDM and PRE, which despite cohort limitations, verify the results of this study and subsequent associations made for less-comprehensively annotated lipidomic and multi-omic associations.

To facilitate improved lipidomic analysis, tailored structural clustering techniques and cheminformatics-powered visualization tools were created to allow for head group and fatty acyl assessments. While other studies have investigated lipid dysregulation for PRE and GDM, most have focused on specific head groups or summed fatty acyl composition. For example, some studies have called out whole subclasses together (e.g. TG upregulation in GDM[39]) or have focused on multiple species in summed notations such as PI(34:2) upregulation in PRE [53]. Assessing the head and fatty acyl groups independently allowed for the evaluation of lipid changes with greater specificity and showcase significant dependencies for PRE and GDM. However, even in our evaluations, exceptions arose, thus even more extensive annotations including double bond orientations and positions are still needed to fully understand some of the alterations occurring in the lipidomic studies.

Several direct protein-lipid associations in the multi-omic analyses shed light on the correlation of protein and lipid dysregulation and known biological implications. Specifically, for GDM, the PE P- lipids having 22:6 fatty acyl groups were upregulated, while PEBP1, which preferentially binds to PEs (among many other roles in the protein signaling cascade), was downregulated. Since these associated molecules were uniquely significant in GDM, their directionality differences provide important mechanistic information for the condition. In PRE, the PI lipids were consistently downregulated in addition to the PHLD enzyme which is specific for PI species. This finding suggests dysregulation in PI associated molecular mechanisms, which was not observed in GDM. While limited biological information causes protein-lipid associations to be sparse for many diseases, the direct relationships observed here suggest a strong interdependence in disease origin and progression. Furthermore, the detected molecular markers

could serve as biomarker candidates for the early detection of both GDM and PRE. However, since our study evaluated patients upon admission to labor and delivery, it only captured disease pathogenesis after clinical diagnosis. To determine whether the observed molecular candidates serve as early diagnostic markers will require longitudinal assessment of a new patient cohort throughout their entire pregnancy to determine early molecular changes prior to clinical diagnosis.

Conflicts of interest

The authors declare no conflicts of interest.

Author Contributions

M.T.O., E.S.B., K.E.B. and B.D.T. analyzed the data and wrote the manuscript. M.A.G. prepared the samples for analysis and K.K.W. and E.S.B. performed the experiments. K.G.S., L.M.B., B.W. developed software for statistical analysis and executed the different statistical tests. M.T.O and K.J.B. assessed the lipidomic changes and M.T.O, A.K.L., M.E.M and K.B.J. performed the proteomic evaluations. M.T.O, J.A. and D.F. developed the cheminformatics-powered visualization tools. E.S.B., K.E.B., B.D.T. and D.F. designed the experiments and supervised different aspects of the project.

Acknowledgements

Portions of this research were supported by grants from the NIH National Institute of Environmental Health Sciences (P30ES025128, P42 ES027704 and P42ES031009), startup funds from North Carolina State University, the NIH Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD) R21 HD084788, and the Discovery Foundation Grant. This data was collected in the W. R. Wiley Environmental Molecular Sciences Laboratory (EMSL) (grid.436923.9), a DOE Office of Science User Facility sponsored by the Office of Biological and Environmental Research and located at Pacific Northwest National Laboratory (PNNL). PNNL is a multiprogram national laboratory operated by Battelle for the Department of Energy (DOE) under Contract DE-AC05-76RLO 1830.

Notes and references

1. L. Bellamy, J. P. Casas, A. D. Hingorani and D. J. Williams, *BMJ*, 2007, **335**, 974.
2. I. Pastore, E. Chiefari, R. Vero and A. Brunetti, *Endocrine*, 2018, **59**, 481-494.
3. J. A. Hutcheon, S. Lisonkova and K. S. Joseph, *Best Pract Res Clin Obstet Gynaecol*, 2011, **25**, 391-403.
4. G. American College of O, Task Force on Hypertension in P., 2013, **122**, 1122-1131.
5. V. A. Moyer and U. S. P. S. T. Force, *Ann Intern Med*, 2014, **160**, 414-420.
6. B. M. Casey, M. J. Lucas, D. D. McIntire and K. J. Leveno, *Obstet Gynecol*, 1997, **90**, 869-873.
7. Y. Yogev, E. M. Xenakis and O. Langer, *Am J Obstet Gynecol*, 2004, **191**, 1655-1660.
8. C. Spaight, J. Gross, A. Horsch and J. J. Puder, *Endocr Dev*, 2016, **31**, 163-178.
9. C. V. Ananth, K. M. Keyes and R. J. Wapner, *BMJ*, 2013, **347**, f6564.
10. S. Y. Kim, C. Saraiva, M. Curtis, H. G. Wilson, J. Troyan and A. J. Sharma, *Am J Public Health*, 2013, **103**, e65-72.
11. R. Bahado-Singh, L. C. Poon, A. Yilmaz, A. Syngelaki, O. Turkoglu, P. Kumar, J. Kirma, M. Allos, V. Accurti, J. Li, P. Zhao, S. F. Graham, D. R. Cool and K. Nicolaides, *Sci Rep*, 2017, **7**, 16189.
12. O. Erez, R. Romero, E. Maymon, P. Chaemsaithong, B. Done, P. Pacora, B. Panaitescu, T. Chaiworapongsa, S. S. Hassan and A. L. Tarca, *PLoS One*, 2017, **12**, e0181468.
13. J. F. Plows, J. L. Stanley, P. N. Baker, C. M. Reynolds and M. H. Vickers, *Int J Mol Sci*, 2018, **19**.
14. B. Liu, Y. Xu, C. Voss, F. H. Qiu, M. Z. Zhao, Y. D. Liu, J. Nie and Z. L. Wang, *PLoS One*, 2012, **7**, e44701.
15. A. T. Frank, B. Zhao, P. O. Jose, K. M. Azar, S. P. Fortmann and L. P. Palaniappan, *Circulation*, 2014, **129**, 570-579.
16. S. L. Rebholz, K. T. Burke, Q. Yang, P. Tso and L. A. Woollett, *Am J Physiol Endocrinol Metab*, 2011, **301**, E416-425.
17. K. M. Antony, P. Hemarajata, J. Chen, J. Morris, C. Cook, D. Masalas, M. Gedminas, A. Brown, J. Versalovic and K. Aagaard, *J Perinatol*, 2016, **36**, 921-929.
18. J. Folch, M. Lees and G. H. Sloane Stanley, *J Biol Chem*, 1957, **226**, 497-509.
19. E. S. Nakayasu, C. D. Nicora, A. C. Sims, K. E. Burnum-Johnson, Y. M. Kim, J. E. Kyle, M. M. Matzke, A. K. Shukla, R. K. Chu, A. A. Schepmoes, J. M. Jacobs, R. S. Baric, B. J. Webb-Robertson, R. D. Smith and T. O. Metz, *mSystems*, 2016, **1**.
20. E. S. Baker, B. H. Clowers, F. Li, K. Tang, A. V. Tolmachev, D. C. Prior, M. E. Belov and R. D. Smith, *J Am Soc Mass Spectrom*, 2007, **18**, 1176-1187.
21. E. A. Livesay, K. Tang, B. K. Taylor, M. A. Buschbach, D. F. Hopkins, B. L. LaMarche, R. Zhao, Y. Shen, D. J. Orton, R. J. Moore, R. T. Kelly, H. R. Udseth and R. D. Smith, *Anal Chem*, 2008, **80**, 294-302.
22. M. M. Matzke, K. M. Waters, T. O. Metz, J. M. Jacobs, A. C. Sims, R. S. Baric, J. G. Pounds and B. J. Webb-Robertson, *Bioinformatics*, 2011, **27**, 2866-2872.
23. N. Beagley, K. G. Stratton and B. J. Webb-Robertson, *Bioinformatics*, 2010, **26**, 280-282.
24. J. W. Chong, D.; Xia, J., *Curr Protoc Bioinformatics*, 2019, **68**.
25. C. W. Dunnett, *Journal of the American Statistical Association*, 1955, **50**, 1096-1121.
26. B. Zhang, N. C. VerBerkmoes, M. A. Langston, E. Uberbacher, R. L. Hettich and N. F. Samatova, *J Proteome Res*, 2006, **5**, 2909-2918.
27. B. J. Webb-Robertson, M. M. Matzke, S. Datta, S. H. Payne, J. Kang, L. M. Bramer, C. D. Nicora, A. K. Shukla, T. O. Metz, K. D. Rodland, R. D. Smith, M. F. Tardiff, J. E. McDermott, J. G. Pounds and K. M. Waters, *Mol Cell Proteomics*, 2014, **13**, 3639-3646.
28. D. Szklarczyk, A. Franceschini, S. Wyder, K. Forslund, D. Heller, J. Huerta-Cepas, M. Simonovic, A. Roth, A. Santos,

K. P. Tsafou, M. Kuhn, P. Bork, L. J. Jensen and C. von Mering, *Nucleic Acids Res*, 2015, **43**, D447-452.
29. C. UniProt, *Nucleic Acids Res*, 2019, **47**, D506-D515.
30. J. Du, Z. Yuan, Z. Ma, J. Song, X. Xie and Y. Chen, *Mol Biosyst*, 2014, **10**, 2441-2447.
31. D. Weininger, *J Chem Inf Comp Sci*, 1988, **28**, 31-36.
32. F. De Lazzari and M. Bisaglia, *Neural Regen Res*, 2017, **12**, 1797-1798.
33. F. Wong and B. Cox, *Journal of Proteomics and Bioinformatics*, 2014, DOI: 10.4172/0974-276x.S10-001.
34. N. S. Merle, R. Noe, L. Halbwachs-Mecarelli, V. Fremeaux-Bacchi and L. T. Roumenina, *Front Immunol*, 2015, **6**, 257.
35. K. P. Conrad and D. F. Benyo, *Am J Reprod Immunol*, 1997, **37**, 240-249.
36. A. Buurma, D. Cohen, K. Veraar, D. Schonkeren, F. H. Claas, J. A. Bruijn, K. W. Bloemenkamp and H. J. Baelde, *Hypertension*, 2012, **60**, 1332-1337.
37. J. R. Ash, M. A. Kuenemann, D. Rotroff, A. Motsinger-Reif and D. Fourches, *J Cheminform*, 2019, **11**, 43.
38. D. L. Bloxam, B. E. Bullen, B. N. Walters and T. T. Lao, *Am J Obstet Gynecol*, 1987, **157**, 97-101.
39. E. Gratacos, E. Casals, C. Sanllehy, V. Cararach, P. L. Alonso and A. Fortuny, *Acta Obstet Gyn Scan*, 1996, **75**, 896-901.
40. J. Chen, X. Cao, Y. Cui, G. Zeng and G. Zhang, *Mol Med Rep*, 2018, **17**, 4011-4018.
41. T. Hajri, A. Ibrahimi, C. T. Coburn, F. F. Knapp, Jr., T. Kurtz, M. Pravenec and N. A. Abumrad, *J Biol Chem*, 2001, **276**, 23661-23666.
42. O. Uhl, H. Demmelmair, M. T. Segura, J. Florido, R. Rueda, C. Campoy and B. Koletzko, *Diabetes Res Clin Pract*, 2015, **109**, 364-371.
43. H. Pschera, B. Larsson and A. Kjaeldgaard, *Gynecol Obstet Invest*, 1989, **28**, 118-122.
44. R. Bitzur, H. Cohen, Y. Kamari, A. Shaish and D. Harats, *Diabetes Care*, 2009, **32 Suppl 2**, S373-377.
45. P. Schonfeld and L. Wojtczak, *J Lipid Res*, 2016, **57**, 943-954.
46. A. Andre, P. Juaneda, J. L. Sebedio and J. M. Chardigny, *Biochimie*, 2006, **88**, 103-111.
47. G. Rodriguez-Araujo, H. Nakagami, Y. Takami, T. Katsuya, H. Akasaka, S. Saitoh, K. Shimamoto, R. Morishita, H. Rakugi and Y. Kaneda, *Sci Rep*, 2015, **5**, 12081.
48. S. Ornaghi, P. Vergani, G. Urban, V. Giardini, F. Moltrasio and B. E. Leone, *Placenta*, 2011, **32**, 264-268.
49. M. G. Mohaupt, *Hypertension*, 2015, **65**, 285-286.
50. M. Pavcnik-Arnol, M. Lucovnik, L. Kornhauser-Cerar, T. Premru-Srsen, S. Hojker and M. Derganc, *Neonatology*, 2014, **105**, 121-127.
51. B. L. J. Poad, X. Zheng, T. W. Mitchell, R. D. Smith, E. S. Baker and S. J. Blanksby, *Anal Chem*, 2018, **90**, 1292-1300.
52. R. F. Irvine, *J Lipid Res*, 2016, **57**, 1987-1994.
53. R. S. Kelly, D. C. Croteau-Chonka, A. Dahlin, H. Mirzakhani, A. C. Wu, E. S. Wan, M. J. McGeachie, W. Qiu, J. E. Sordillo, A. Al-Garawi, K. J. Gray, T. F. McElrath, V. J. Carey, C. B. Clish, A. A. Litonjua, S. T. Weiss and J. A. Lasky-Su, *Metabolomics*, 2017, **13**.