Unveiling molecular signatures of preeclampsia and gestational diabetes mellitus with multi-omics and innovative cheminformatics visualization tools

Melanie T. Odenkirk[1], Kelly G. Stratton[2], Marina A. Gritsenko[3], Lisa M. Bramer[2], Bobbie-Jo M. Webb-Robertson[3,4], Kent J. Bloodsworth[3], Karl K. Weitz[3], Anna K. Lipton[3], Matthew E. Monroe[3], Jeremy Ash[1,5,6], Denis Fourches[1,5]*, Brandie D. Taylor[7]*, Kristin E. Burnum-Johnson[8]*, Erin S. Baker[1]*

[1]*Department of Chemistry, North Carolina State University, Raleigh, NC 27695*
[2]*National Security Division, Pacific Northwest National Laboratory, Richland, WA 99354*
[3]*Biological Sciences Division, Pacific Northwest National Laboratory, Richland, WA 99354*
[4]*Department of Biostatistics and Informatics, University of Colorado, Aurora, CO 80045*
[5]*Bioinformatics Research Center, North Carolina State University, Raleigh, NC 27695*
[6]*Department of Statistics, North Carolina State University, Raleigh, NC 27695*
[7]*College of Public Health, Temple University, Philadelphia, PA 19133*
[8]*Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, WA 99354*

*Co-corresponding authors:
Denis Fourches
Correspondence: dfourch@ncsu.edu

Brandie D. Taylor
Correspondence: b.depaoli-taylor@temple.edu

Kristin E. Burnum-Johnson
Correspondence: kristin.burnum-johnson@pnnl.gov

Erin S. Baker
Correspondence: ebaker@ncsu.edu

Supplementary Materials

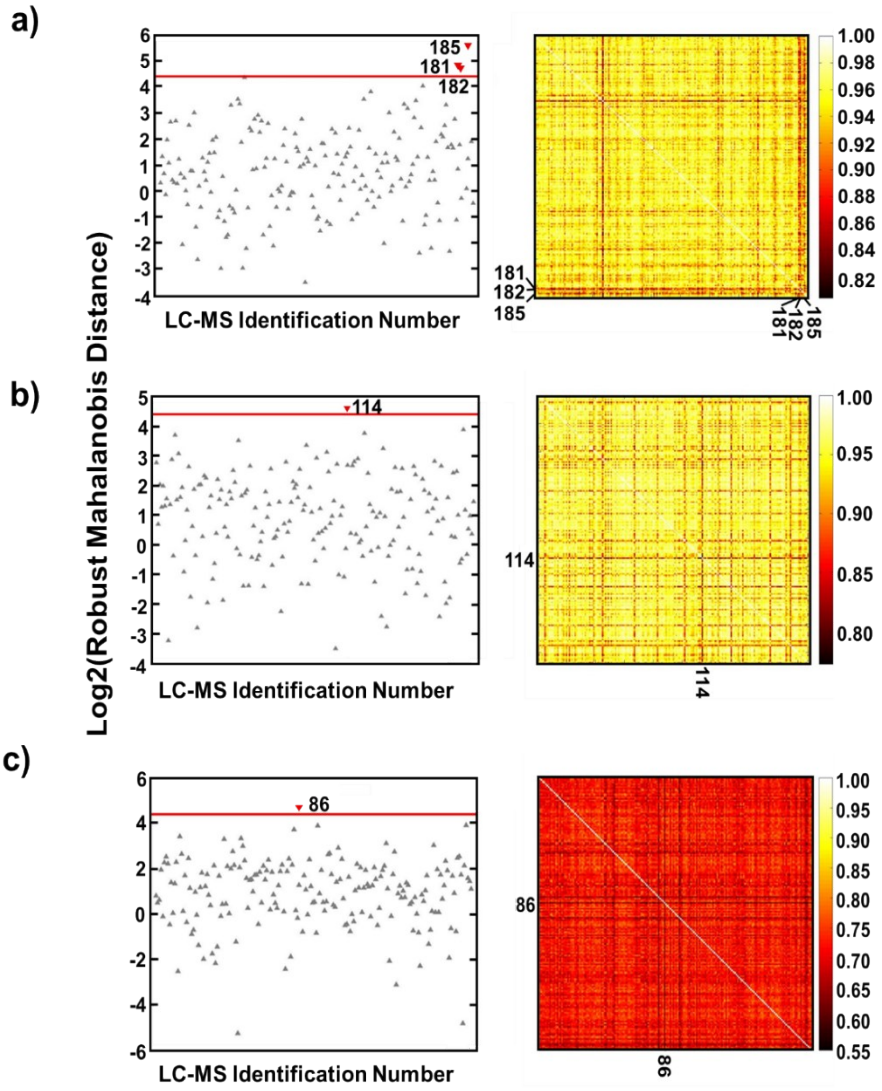

Figure S1. Complication vs. Control outlier assessment. RMD-PAV (left) identifies LC-MS datasets that are extreme deviants from the remaining datasets (above red line) for negative ion lipidomics (a), positive ion lipidomics (b) and proteomics (c) statistical analysis. Heatmap of Pearson correlation (right) confirms one control patient outliers for proteomics, three PRE outliers for negative ion lipidomics and one GDM outlier was removed from positive ion lipidomics.

Table S1- Clinical Information. Expansion of patient cohort information for GDM, PRE and Control patients provided in Table 1. Each row represents clinical information for each of the 186 patients analyzed by LC-IMS-MS analysis, with identification numbers for each patient given in column B. Variable descriptions are given across row 1.

Table S2- HPLC Gradient. HPLC gradient and column wash profiles used for the lipidomic LC analyses.

Elution Gradient

Time (min)	% MPA	% MPB	Flow Rate (mL/min)
0	60	40	0.25

2	50	50	0.25
3	40	60	0.25
12	30	70	0.25
15	25	75	0.25
17	22	78	0.25
19	15	85	0.25
22	8	92	0.25
25	1	99	0.25
34	1	99	0.25

Column Wash

Time (min)	% MPA	% MPB	Flow Rate (mL/min)
34.5	60	40	0.3
35	1	99	0.3
35.5	1	99	0.3
36	60	40	0.35
37	60	40	0.3
38	60	40	0.25

Table S3 - Filtered Proteins. List of all identified proteins following peptide filtering. Column A through C, protein reference number, accession number and name are given for all identified proteins. Column D through E, protein peptide count and number of unique peptides. Columns E-H, results of differential abundance analysis comparing control to GDM (columns F,G) and PRE (columns H,I) samples (F,H = p-values; G,I = $\log_2$ fold changes). Adjusted p-values are from ANOVA (quantitative comparison) with a Dunnett test correction. Column J-K, flags for statistical significance based on ANOVA and q-test analyses with p-value < or = 0.05 for GDM (J) and PRE (K). 0=does not meet criteria for statistical significance by ANOVA or g-test; ++ =meets criteria for statistical significance by Holm-adjust g-test and case is upregulated compared to control; + =meets criteria for statistical significance by ANOVA and case is upregulated compared to control; - =meets criteria for statistical significance by ANOVA and case is downregulated compared to control; -- =meets criteria for statistical significance by Holm-adjust g-test and case is downregulated compared to control (only proteomics). Columns L-GN, patient $\log_2$ normalized protein abundances. Protein sequence coverage (GO) with associated peptides (GP) identified by mass spectrometry.

Table S4 - More Sig Protein. List of all statistically significant proteins for either complication vs. control comparison. Columns B, C, D and E contain protein reference number and uniprot annotations. Column F and G, GDM vs. Control (F) and PRE vs. Control (G) Flag annotations. Columns H and I, peptide coverage of total (H) and unique

(I) peptides. Columns O-U, additional information for significant proteins that met filtering requirements.

Table S5 - Lipidomics. List of all statistically significant lipids for either complication vs. control comparison. Column A, Lipid species identified in plasma. Isomers whose peaks are separated in retention time are annotated with '*' and '**'. Multiple lipid identifications from the same MS/MS spectra are separated by a semi-colon. Note that some of the lipid species identified in positive ionization were also identified in negative ionization. Mode of ionization is specified in column D. Column B, the sum of carbons composing the chains (fatty acid and long base for sphingolipids). Column C, the sum of number of double bonds composing the chains (fatty acid and long base for sphingolipids). Columns E-H, results of differential abundance analysis comparing control to GDM (columns E,F) and PRE (columns G,H) samples (E,G = p-values; F,H = $\log_2$ fold changes). Adjusted p-values are from ANOVA (quantitative comparison) with a Dunnett test correction. Column J-K, flags for statistical significance based on p-value < or = 0.05 for GDM (J) and PRE (K). 0=does not meet criteria for statistical significance by ANOVA or g-test; + =meets criteria for statistical significance by ANOVA and case is upregulated compared to control; - =meets criteria for statistical significance by ANOVA and case is downregulated compared to control. Columns O-GR, normalized, $\log_2$ transformed values (1 column per sample) of control (O-DC) and GDM (DD-EU) and PRE (EV- GR). For negative mode lipidomics, three samples were determined to be outliers (TAM_7194, PE; TAM_721, PE; TAM_7635, GDM) and were not included for statistical analysis of negative ion lipidomics results. For negative mode lipidomics, one sample was determined to be an outlier (TAM_205, GDM) and was not included for statistical analysis for positive ion lipidomics results.

Table S6- GDM vs. PRE test. Statistical output for GDM vs. PRE comparison. Statistical analyses were conducted using an unpaired t-test on proteomics, negative lipidomics and positive lipidomics separately to assess disease differentiation between pregnancy complication PRE and GDM (GDM v. PRE). Species that were significant in original disease versus control comparisons are specified in Column A. Subsequent statistical results are outlined in columns B-E. Column F-H, flags for statistical significance based on p-value < or = 0.05 for GDM v. PRE (F), GDM vs. Ctrl (G) and PRE vs. Ctrl (H). 0=does not meet criteria for statistical significance by ANOVA or g-test; ++ =meets criteria for statistical significance by Holm-adjust g-test and case is upregulated compared to control (only proteomics); + =meets criteria for statistical significance by ANOVA and case is upregulated compared to control; - =meets criteria for statistical significance by ANOVA and case is downregulated compared to control; -- =meets criteria for statistical significance by Holm-adjust g-test and case is downregulated compared to control (only proteomics).

Table S7- STRING Enrichment. Functionally enriched biological processes of significant proteins in Complication versus Control comparisons.

	Go Term	Biological Process	Count in Gene Set	False Discovery Rate
PRE	GO:0042981	regulation of apoptotic process	24 of 1501	0.0119
	GO:0060326	cell chemotaxis	7 of 183	0.0097
	GO:0006957	complement activation, alternative pathway	5 of 13	6.07E-06
	GO:0006956	complement activation	12 of 49	5.06E-12
	GO:0019835	cytolysis	5 of 32	0.00021
	GO:0006887	exocytosis	43 of 774	1.86E-20
	GO:0042730	fibrinolysis	6 of 21	1.71E-06
	GO:0007596	blood coagulation	21 of 288	6.02E-12
	GO:0006954	inflammatory response	26 of 482	4.57E-12
	GO:0045087	innate immune response	26 of 676	3.07E-09
	GO:0006869	lipid transport	9 of 272	0.0057
	GO:0006810	transport	73 of 4130	1.69E-10
	GO:0007155	cell adhesion	20 of 843	0.00039
GDM	GO:0033344	cholesterol efflux	3 of 24	0.0073
	GO:0006954	inflammatory response	8 of 482	0.0144
	GO:0045087	innate immune response	8 of 676	0.0447
	GO:0006810	transport	28 of 4130	0.0188
	GO:0006950	response to stress	27 of 3267	0.0042
	GO:0007155	cell adhesion	12 of 843	0.0063