S1 Appendix. Supplemental Methods Section

Targeted Metabolomic Analysis

Both mass spectrometry (MS) and NMR (nuclear magnetic resonance) platforms were used for targeted metabolomic analysis. We have previously published details regarding the NMR-based approach [1, 2]. Briefly, serum samples were filtered through 3-kDa cut-off centrifuge filter units (Amicon Micoron YM-3; Sigma-Aldrich, St. Louis, MO) to remove blood proteins. Aliquots of each serum sample were transferred into the centrifuge filter devices and spun (10,000 rpm for 20 minutes) to remove macromolecules (primarily protein and lipoproteins) from the sample. The filtrates were checked visually for any evidence that the membrane was compromised and for these samples the filtration process was repeated with a different filter and the filtrate inspected again. The subsequent filtrates were collected and the volumes were recorded. If the total volume of the sample was under 300 µL an appropriate amount from a 50 mM NaH_2PO_4 buffer (pH 7) was added until the total volume of the sample was 300 µL. Any sample that had to have buffer added to bring the solution volume to 300 uL, was annotated with the dilution factor and metabolite concentrations were corrected in the subsequent analysis. After this, 35 □L of D_2O and 15 □L of buffer solution containing 11.667 mM DSS (disodium-2, 2-dimethyl-2-silceptentane-5-sulphonate), 730 mmol imidazole and 0.47% NaN_3 in H_2O) was added to the sample.

Three hundred and fifty (350) □L of serum was then transferred to a micro cell NMR tube (Shigemi, Inc., Allison Park, PA). ^{1}H-NMR spectra were collected on a 500-MHz Inova (Varian Inc, Palo Alto, CA) spectrometer with a 5-mm ITCN Z-gradient PFG room-temperature probe. The singlet produced by the DSS methyl groups was used as an internal standard for chemical shift referencing (set to 0 ppm) and for quantification. All ^{1}H-NMR spectra were

processed and analyzed using the Chenomx NMR Suite Professional Software package version 7.6 (Chenomx Inc, Edmonton, AB, Canada) and were manually fitted to an internal database. Each spectrum was evaluated by at least 2 NMR spectroscopists to minimize errors.

For the MS-based approach, targeted quantitative metabolomic analysis of the serum was performed by combining direct injection mass spectrometry with reverse-phase LC-MS/MS using the commercially available AbsoluteIDQ™ kit, which is available from BIOCRATES Life Sciences AG (Austria). A detailed description of this approach was also previously published by us [3]. Targeted identification and quantification of up to 180 different endogenous metabolites including amino acids, acylcarnitines, biogenic amines, glycerophospholipids, sphingolipids and sugars was performed using this technique.

Untargeted Metabolomic Analysis

Serum Metabolite Extraction

Metabolites were extracted from serum samples via protein precipitation with methanol. Three volumes of ice-cold methanol were added to 20 μL of serum, vortexed and incubated on ice for 15 min. This was followed by centrifugation at 14,000 rpm for 15 min and then the supernatants were dried using a SpeedVac.

Dansylation Labeling

Serum samples were re-suspended in 25 μL of 5:1 v/v H_2O:ACN. The 25 μL sample solutions were mixed with sodium carbonate/sodium bicarbonate buffer and ACN. The solutions were vortexed, spun down and mixed with 25 μL of freshly prepared ^{12}C-dansyl chloride solution (18 mg/mL) (for light isotope labeling) or ^{13}C-dansyl chloride solution (18 mg/mL) (for heavy isotope labeling) [4]. The dansylation reaction was allowed to proceed for 60 min at 40°C.

After 60 min, NaOH was added to the reaction mixture to quench the excess dansyl chloride. The solution was then incubated at 40°C for another 10 min. Finally, formic acid in 50/50 acetonitrile/H_2O was added to consume excess NaOH and to make the solution acidic.

LC-UV Quantification

For sample normalization, a Waters ACQUITY UPLC system (Milford, MA, USA) with a binary solvent manager, a sampler manager, and a photodiode array (PDA) detector was used for the quantification of the labeled metabolites [5]. First, 5 μL of the labeled serum solutions were injected onto a Waters ACQUITY BEH C18 column (2.1 mm × 5 cm, 1.7 μm particle size, 130 Å pore size) for a fast step-gradient run. Solvent A was 0.1% (v/v) formic acid in 5% (v/v) acetonitrile; Solvent B was 0.1% (v/v) formic acid in acetonitrile. The gradient started with 0% B for 1 min and was stepwise increased to 95% within 0.01 min and held at 95% B for 1 min to ensure complete elution of all labeled metabolites. The gradient was restored to 0% B in 0.5 min and held in this condition for 3.5 min to re-equilibrate the column. The flow rate was 0.45 mL/min. The detection wavelength was set at 338 nm.

LC-MS

The labeled metabolites were analyzed using a Bruker Maxis Impact QTOF mass spectrometer (Bruker, Billerica, MA) linked to an Agilent 1100 series binary HPLC system (Agilent, Palo Alto, CA). The labeled serum samples were injected onto an Agilent reversed phase Eclipse Plus C18 column (2.1 mm × 10 cm, 1.8 μm particle size, 95 Å pore size) for separation. Solvent A was 0.1% (v/v) formic acid in 5% (v/v) acetonitrile, and solvent B was 0.1% (v/v) formic acid in acetonitrile. The gradient started at 20% B and stepwise increased to 35% B within 3.5 min, followed by 14.5-min ramp to 65% B, then raised to 99% B within 6 min and finally held at 99% B for 10 min to ensure complete elution of all labeled metabolites. The

flow rate was 180 μL/min. The MS conditions were as follows: nebulizer gas pressure 1.8 bar; drying gas temperature 230°C; drying gas flow 8 L/min; capillary voltage 4500 V; and scan range 150–1000. All MS spectra were obtained in the positive ion mode.

Data Analysis (untargeted metabolomics)

The resulting MS data were internally calibrated and processed using our in-house peak-pair picking software, IsoMS, written in the R language [6]. This program eliminates false positive peaks, such as isotopic peaks, common adduct ions, and multiple charged ions. Only the protonated ion pairs with S/N greater than 10 were exported for further analysis. The extracted peak-pair data from the LC-MS spectra were aligned by retention time matching (within 30 seconds) and accurate mass matching (within 8 ppm). After alignment, missing values in the aligned files were filled by the Zerofill program [7]. Metabolite identification was performed based on accurate mass and retention time match to a dansyl standard library [8]. Putative identification was done based on accurate mass match to metabolites in the human metabolome database (HMDB) (www.hmdb.ca) and the predicted human metabolite library in MyCompoundID (MCID) (www.mycompoundid.org) [9].

Overall Statistical Analysis

Our statistical approaches have been described in detail in prior publications [10, 11, 12]. Standard metabolomics statistical analyses were utilized [10, 11]. Log-transformation and Pareto-scaling were performed to normalize the metabolite concentration data for following multivariate analysis. Principal component analysis (PCA) and Partial Least Squares Discriminant Analysis (PLS-DA) were performed to identify distinct metabolite patterns. To assess the significance of the separations achieved by PLS-DA, permutation testing was performed (using 2000 data resampling steps) to determine the corresponding p-values.

Permutation testing allows one to determine the likelihood that the observed separation between cases and controls on a PLS-DA plot is due to chance [3]. In addition, Variable Importance in Projection (VIP) plots were generated from the PLS-DA data. VIP plot scores estimate the importance of each variable in the projection used in a PLS-DA model which is often used for variable selection. MetaboAnalyst was used to perform univariate analysis as well as PCA and PLS-DA analyses [11]. Custom programs written using the R statistical software package (http://www.r-project.org) and STATA 12.0 (http://www.stata.com) were used to perform all other statistical analyses.

Mean (SD) metabolite concentrations in cases and controls were compared using a two-tailed t-test. The Mann Whitney U test was performed for non-normal distribution of each metabolite. Logistic regression analysis was used to generate the optimal predictive models for stillbirth prediction. Independent variables and potential confounders considered in each of the prediction models. These included clinical or demographic factors such as maternal age, ethnicity, parity (multi- versus nulliparous), tobacco use, prior history of chronic hypertension, other medical disorders, conception method (spontaneous versus others), BMI, fetal CRL (crown rump length) measurement, delta NT (observed – expected based on CRL), UtPI, maternal serum PAPP-A (MoM) and β-hCG (MoM) concentrations.

Two groups of models were developed, one was based on metabolites only, and the other was based on metabolites plus the other potential predictors listed above. We evaluated the diagnostic accuracy of these prediction models for overall stillbirth as well as several stillbirth subgroups: i) early occurring < 28 weeks, ii) < 32 weeks, iii) stillbirth related to fetal growth restriction/ placental abnormality and iv) unexplained stillbirth. Metabolite data were log-transformed, and metabolites with a p-value < 0.3 (using univariate analysis) were selected for developing logistic

regression models. We employed a k-fold cross-validation (CV) technique to ensure that the logistic regression models were robust [12]. LASSO (Least Absolute Shrinkage and Selection Operator) was used for variable/predictor selection [13] and stepwise variable selection was utilized to optimize all the model components [14] via 10-fold CV. The threshold used for inclusion of a metabolite or other clinical or other variables required that the particular variable be selected > 8 times of the 10 cross-validations performed. From the selected features, robust logistic equations for stillbirth prediction were developed. Areas under the Receiver Operating Characteristic curve were (AUROC or AUC) [12] along with sensitivity and specificity values were calculated.

REFERENCES

1. Bahado-Singh RO, Akolekar R, Mandal R, Dong E, Xia J, Kruger M, Wishart DS, Nicolaides K. First-trimester metabolomic detection of late-onset preeclampsia. American journal of obstetrics and gynecology. 2013;208:58 e1-7.
2. Bahado-Singh RO, Syngelaki A, Akolekar R, Mandal R, Bjondahl TC, Han B, Dong E, Bauer S, Alpay-Savasan Z, Graham S, Turkoglu O, Wishart DS, Nicolaides KH. Validation of metabolomic models for prediction of early-onset preeclampsia. American journal of obstetrics and gynecology. 2015;213:530 e1- e10.
3. Bahado-Singh RO, Ertl R, Mandal R, Bjorndahl TC, Syngelaki A, Han B, Dong E, Liu PB, Alpay-Savasan Z, Wishart DS, Nicolaides KH. Metabolomic prediction of fetal congenital heart defect in the first trimester. American journal of obstetrics and gynecology. 2014;211:240 e1- e14.
4. Guo K, Li L. Differential 12C-/13C-isotope dansylation labeling and fast liquid chromatography/mass spectrometry for absolute and relative quantification of the metabolome. Analytical chemistry. 2009;81:3919-32.
5. Wu Y, Li L. Determination of total concentration of chemically labeled metabolites as a means of metabolome sample normalization and sample loading optimization in mass spectrometry-based metabolomics. Analytical chemistry. 2012;84:10723-31.
6. Zhou R, Tseng CL, Huan T, Li L. IsoMS: automated processing of LC-MS data generated by a chemical isotope labeling metabolomics platform. Analytical chemistry. 2014;86:4675-9.
7. Huan T, Li L. Counting missing values in a metabolite-intensity data set for measuring the analytical performance of a metabolomics platform. Analytical chemistry. 2015;87:1306-13.
8. Huan T, Wu Y, Tang C, Lin G, Li L. DnsID in MyCompoundID for rapid identification of dansylated amine- and phenol-containing metabolites in LC-MS-based metabolomics. Analytical chemistry. 2015;87:9838-45.
9. Li L, Li R, Zhou J, Zuniga A, Stanislaus AE, Wu Y, Huan T, Zheng J, Shi Y, Wishart DS, Lin G. MyCompoundID: using an evidence-based metabolome library for metabolite identification. Analytical chemistry. 2013;85:3401-8.
10. Wishart DS. Computational approaches to metabolomics. Methods Mol Biol. 2010;593:283-313.

11. Xia J, Mandal R, Sinelnikov IV, Broadhurst D, Wishart DS. MetaboAnalyst 2.0--a comprehensive server for metabolomic data analysis. Nucleic Acids Res. 2012;40:W127-33.
12. Xia J, Broadhurst DI, Wilson M, Wishart DS. Translational biomarker discovery in clinical metabolomics: an introductory tutorial. Metabolomics. 2013;9:280-99.
13. Tibshirani R. Regression Shrinkage and Selection via the Lasso. Journal of the Royal Statistical Society Series B (Methodological). 1996;58:267-88.
14. Hastie T, Tibshirani R, Friedman JH. The Elements of Statistical Learning: Data Mining, Inference, and Prediction: Springer; 2001.