Data analysis

Untargeted LC-MS

Raw data processing was done using Agilent software (MassHunter Qual and ProFinder). Data analysis was performed separately for positive and negative mode data with Agilent MassProfiler Pro (MPP) package using recursive analysis workflow.

Recursive analysis was performed initially on every sample batch. Molecular feature extractor (MFE) was run first and data were imported into MPP. The following sample groups were defined:

1. mothers at first trimester
2. mothers at term
3. female children at term
4. male children at term
5. pooled sample
6. CHEAR control sample

A subset of features present in at least 80% of the samples of at least one of sample groups was combined with the authentic standards library and used to run the recursive search using Find-by-Formula (FBF) algorithm. The data were imported into MPP and again a subset of features present in at least 80% of the samples of at least one of sample groups was exported as a consistent set (FBF-library) for the given batch.

FBF-libraries for all batches were aligned using MPP, a subset of features present in at least two out of three libraries was created and used to run FBF recursive search for all the samples. This tree-step procedure allows to ensure good coverage and consistent feature naming across the whole experiment.

Data from the second recursive search were imported into MPP and then exported to in-house software to remove duplicates (features with same mass and retention). Cleaned data were then normalized in order to reduce the influence of instrument drift and batch effect on the final analysis. Instrument drift was corrected separately within each sample batch using pooled samples as reference. Correction was done compound by compound using LOESS fitting of peak

intensity vs time for pooled samples and then interpolation for all samples in the batch. Pooled samples were run at the start and end of the batch and at equal intervals within a batch.

Between-batch normalization was then performed compound by compound using the ratio of median values for pooled samples within each batch:

- Median values for pooled samples were computed for each compound within each batch ($M_{i,j}$ where *i* is batch and *j* is feature);
- One batch was chosen as reference and normalization coefficient was computed: $C_{i,j} = M_{i,j} / M_{ref,j}$
- When the feature was absent in particular batch ratio couldn't be computed and was replaced by the median of all ratios
- Values were normalized for non-reference batches by dividing feature area by the respective normalization coefficient.

Lipidomics

The lipids were identified using Lipid Blast software by matching MSMS spectra to the library. Identification results obtained were used to create quantitation method and the data files were processed using MultiQuant 1.1.0.26 (AB-SCIEX, Concord, Canada). Quantitative data were normalized against internal standards. The quality control (QC) samples were mainly used to remove technical outliers and lipid species that were detected below the lipid class-based lower limit of quantification. In total, 8 QC samples evenly distributed along analytical runs of the study were analyzed. The average coefficient of variation of all the lipids detected in the study samples was below 35%.

Male, female and pooled samples were tentatively identified using steroid panel and acyl carnitine data. A list of lipids differentiating between male and female subjects was generated using T-test. PCA analysis performed using this list demonstrated reasonable separation between the groups.