Supplementary Methodology

UPLC-MS analysis

Samples were prepared by reconstitution in 90µl HPLC grade water followed by vortex mixing (15 seconds), centrifugation (11 337*g*, 15 minutes) and transfer to vials. Samples were analysed by an Acquity UPLC (Waters Corp. Milford, USA) coupled to a LTQ-Orbitrap mass spectrometry system (Thermo Fisher Scientific, Bremen, Germany) operating in electrospray ionisation mode (4). Samples were analysed consecutively in positive ion mode followed and then consecutively in negative ion mode. Chromatographic separations were performed employing an ACQUITY UPLC BEH 1.7µm-C_{18} column (2.1 x 100mm, Waters Corp. Milford, USA). Solvent A and solvent B were 0.1% formic acid in water and 0.1% formic acid in methanol, respectively. In positive ion mode a flow rate of 0.40ml.min^{-1} was applied with a gradient elution profile (100% A for 1 minute and subsequently ramped to 100% B (curve 5) over 15 minutes, followed by a 4 minute hold at 100% B before a rapid return to 100% A and a hold for 2 minutes). In negative ion mode a flow rate of 0.36ml.min^{-1} was applied with a gradient elution program (100% A for 2 minutes and subsequently ramped to 100% B (curve 4) over 15 minutes, followed by a 5 minute hold at 100% B before a rapid return to 100% A and a hold for 2 minutes). The column and samples were maintained at temperatures of 50°C and 4°C, respectively. A 10µl sample volume was introduced onto the column and 50% of the column effluent was transferred to the mass spectrometer. Centroid MS scans were acquired in the mass range of 50-1000Th using the Orbitrap mass analyser operating with a target mass resolution of 30 000 (FWHM as defined at m/z 400) and a scan time of 0.4s. Mass calibration was performed before each analytical batch using an instrument manufacturer defined calibration mixture (ThermoFisher Scientific, Bremen, Germany).

Data processing of UPLC-MS data

All data was converted to netCDF format using the FileConverter program in the XCalibur software package (ThermoFisher Scientific, Bremen, Germany). Raw data processing. All raw data (in.raw file format) were converted to netCDF file format with the FileConverter program available in XCalibur (ThermoFisher Scientific, Bremen, Germany).

XCMS deconvolution.

XCMS is an open-source deconvolution program available for LC-MS data.(1) Deconvolution using the XCMS program was performed using identical settings to those reported previously(2) with the exception of s/n threshold = 3, step = 0.02, m/z diff = 0.05 and for grouping bandwidth = 10 and mzwidth = 0.05. The esi program (http://msbi.ipb-halle.de/msbi/esi/) available with the XCMS software package was used to write peak output files to an annotated version (as a .csv file) which is more appropriate for these studies. XCMS and esi were run using R version 2.6.0.

Quality Assurance

The performance of analytical instrumentation has to be assessed robustly to ensure that data are of comparable high quality within an analytical run. An approach based on the periodic analysis of a standard biological Quality Control sample (QC sample) together with the patient samples is now accepted as a quality assurance strategy in metabolic profiling. (3, 4) A similar Quality Assurance protocol has been followed in this metabolomic study to assess the repeatability for thousands of endogenous metabolites. A set of pooled QC samples were prepared by mixing equal aliquots from all the samples in a single study. A QC sample is

then injected after every fourth patient sample in each analytical run (a *lead-in* of 10 consecutive QC injections was performed at the start of every analytical run to equilibrise the IPLC column response). At the end of the experimental run, and after XCMS deconvolution each detected peak is normalised to the QC sample using robust Loess signal correction (R-LSC). Here Locally Weighted Scatterplot Smoothing (LOESS) is performed on the QC data with respect to the order of injection. A cubic spline correction curve for the whole analytical run is then interpolated, to which the total data set for that peak is normalized. Using this procedure any attenuation of peak response over an analytical run (i.e. confounding factor due to injection order) is minimised.(3, 4) After R-LSC each peak is required to pass strict Quality Assurance criteria. While there are no generally accepted criteria for the assessment of repeatability in metabolomic data sets, the UK Food and Drug Administration (FDA) suggests a range of criteria that should be applied. In the guidance for bioanalytical method validation in industry (5) the FDA recommends for single analyte tests that tolerance limits are set such that the measured response detected in two-thirds of QC samples is within 15% of the QC mean, except for compounds with concentrations at or near the limit of quantification (LOQ), in these cases a tolerance of 20% is acceptable. In our case, the methods are not specific for one analyte of interest, but instead we aim to detect thousands of analytes, therefore an acceptance tolerance of 20% would seem to be appropriate. Any peak that did not pass the QA criteria was removed from the dataset and thus ignored in any subsequent data analysis.

References:

1. Smith, C.A., Want, E.J., O'Maille, G., Abagyan, R., and Siuzdak, G. 2006. XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal Chem* 78:779-787.
2. Brown, M., Dunn, W.B., Dobson, P., Patel, Y., Winder, C.L., Francis-McIntyre, S., Begley, P., Carroll, K., Broadhurst, D., Tseng, A., et al. 2009. Mass spectrometry tools and metabolite-specific databases for molecular identification in metabolomics. *Analyst* 134:1322-1332.
3. van der Greef, J., Martin, S., Juhasz, P., Adourian, A., Plasterer, T., Verheij, E.R., and McBurney, R.N. 2007. The art and practice of systems biology in medicine: mapping patterns of relationships. *J Proteome Res* 6:1540-1559.
4. Zelena, E., Dunn, W.B., Broadhurst, D., Francis-McIntyre, S., Carroll, K.M., Begley, P., O'Hagan, S., Knowles, J.D., Halsall, A., Wilson, I.D., et al. 2009. Development of a robust and repeatable UPLC-MS method for the long-term metabolomic study of human serum. *Anal Chem* 81:1357-1364.
5. CDER. 2001. Guidance for Industry, Bioanalytical Method Validation. F.a.D.A. Centre for Drug Valuation and Research, editor.

Table SI. The putatively identified metabolites that were significantly different between the cord plasma study (SGA vs. Control) and RUPP plasma study (Normal vs. RUPP).

	Venous Cord Plasma		RUPP model	
Putative Metabolite Indenty based on exact mass	**p-vlaue**	**direction**	**p-value**	**direction**
Cervonyl carnitine **AND**/**OR** 1α,25-dihydroxy-18-oxocholecalciferol	3.09E-06	DOWN	0.004	UP
PC(20:4/0:0) **AND**/**OR** LysoPC(20:4)	6.57E-06	DOWN	0.001	UP
LysoPC(14:0) **OR** PC(O-12:0/2:0)ORlysoPC(14:0) **OR** 1,25-dihydroxy-24-oxo-23-azaergocalciferol	6.63E-06	DOWN	0.008	UP
LysoPC(18:2)	7.81E-06	DOWN	0.012	UP
PC(O-16:1/2:0) **AND**/**OR** PC(16:0/2:0) AND/OR PC(18:1/0:0) **AND**/**OR** LysoPC(18:1) AND/OR LysoPC(20:4)	8.42E-06	DOWN	0.003	UP
1α,25-dihydroxy-24-oxo-23-azaergocalciferol	9.55E-06	DOWN	0.008	UP
LysoPC(16:1) **OR** Cervonyl carnitine	1.10E-05	DOWN	0.005	UP
LysoPC(18:0) **OR** PC	3.80E-05	DOWN	0.0004	UP
Clupanodonyl carnitine **AND**/**OR** Vaccenyl carnitine	3.90E-05	DOWN	0.002	UP
N-propyl-16,16-dimethyl-5Z,8Z,11Z,14Z-docosatetraenoyl amine	5.30E-05	DOWN	0.025	UP
LysoPC(16:0) **OR** PC(O-14:0/2:0) **OR** Docosa-4,7,10,13,16-pentaenoyl carnitine **OR** Clupanodonyl carnitine	5.50E-05	DOWN	0.002	UP
Docosa-4,7,10,13,16-pentaenoyl carnitine **OR** Clupanodonyl carnitine	6.40E-05	DOWN	0.004	UP
12-oxo-c-LTB3 **AND**/**OR** S-(9-deoxy-delta9,12-PGD2)-glutathione	6.90E-05	DOWN	0.014	UP
LysoPC(18:1) **OR** LysoPC(18:1) OR PC(O-16:1/2:0)	7.60E-05	DOWN	0.008	UP
Taurocholate **OR** Taurohyocholate **OR** Tauroursocholic acid **OR** Taurallocholic acid **OR** Tauro-b-muricholic acid	1.32E-04	DOWN	0.010	UP
LysoPC(18:0) OR PC(O-16:0/2:0) **OR** tetracosapentaenoyl carnitine	4.35E-04	DOWN	0.001	UP
tetracosapentaenoyl carnitine	4.74E-04	DOWN	0.002	UP
Phosphatidate	0.001	DOWN	0.015	UP
LysoPC(18:0) **OR** PC	0.001	DOWN	0.002	UP
PC **OR** leukotriene C5	0.002	DOWN	0.004	UP
PC(O-14:0/18:0) **OR** PC(O-16:0/16:0) OR Tricosanamide OR	0.002	DOWN	0.027	UP
Linoelaidyl carnitine **OR** (+/-)N-(1-methyl-2-hydroxy-2-phenyl-ethyl) arachidonyl amine OR N-eicosanoyl-ethanolamine	0.002	DOWN	0.004	UP
1-18:1-lysoPE **OR** PE(18:1(9Z)/0:0) OR PE(18:1(9Z)/0:0)[U] OR N-(5-hydroxy-pentyl) arachidonoyl amine **OR** N-propyl N-(2-hydroxy-ethyl) arachidonoyl amine **OR** (+)N-(2S-hydroxy-propyl) a,a-dimethylarachidonoyl amine **OR** N-(2-isopropyl-5Z,8Z,11Z,14Z-eicosatetraenoyl)-ethanolamine **OR** N-(5Z,8Z,11Z,14Z-tricosatetraenoyl)-ethanolamine **OR** N-(17,17-dimethyl-5Z,8Z,11Z,14Z-heneicosatetraenoyl)-ethanolamine **OR** N-(17-methyl-5Z,8Z,11Z,14Z-docosatetraenoyl)-ethanolamine	0.002	DOWN	0.012	UP
N-(11Z-eicosaenoyl)-ethanolamine **OR** Palmitoylcarnitine **OR** 1-(1Z-hexadecenyl)-sn-glycero-3-phosphoethanolamine **OR** CPA(18:0/0:0)	0.004	DOWN	0.046	UP
PC(O-16:0/18:1) AND/OR PC(O-18:0/16:1) **AND**/**OR** PC(16:0/18:0)	0.004	DOWN	0.001	UP
Tocopherol nicotinate **OR** 11a-(4-dimethylaminophenyl)-1a,25-dihydroxyvitamin D3 **OR** Tocopherol nicotinate OR PC(O-18:0/O-2:1)	0.005	DOWN	0.001	UP
Stearoylglycerone phosphate **OR** 1-Oleoyl-lysophosphatidic acid **OR** LPA(/18:1) **OR** lysoPE(16:0) **OR** lysoPC(13:0) **OR** Aplidiasphingosine OR Sphingofungin A	0.005	DOWN	0.005	UP
1α,25-dihydroxy-11α-phenylcholecalciferol	0.006	DOWN	0.049	DOWN
(15Z)-Tetracosenoic acid **OR** 10,13-Dimethyl-11-docosyne-10,13-diol **OR** trans-selacholeic acid **OR** nervonic acid OR Conessine	0.006	DOWN	0.023	UP
N-Glycoloylganglioside GM2 **OR** PC - more than 20 hits	0.007	DOWN	0.021	UP
PC(16:0/22:5) **AND**/**OR** PC(18:0/20:5) **AND**/**OR** PC(18:1/20:4) **AND**/**OR** PC(16:1/22:4)	0.014	DOWN	0.001	DOWN
PGD2-dihydroxypropanylamine **OR** 15R-PGE2 methyl ester, 15-acetate	0.020	DOWN	0.019	UP

Figure S1

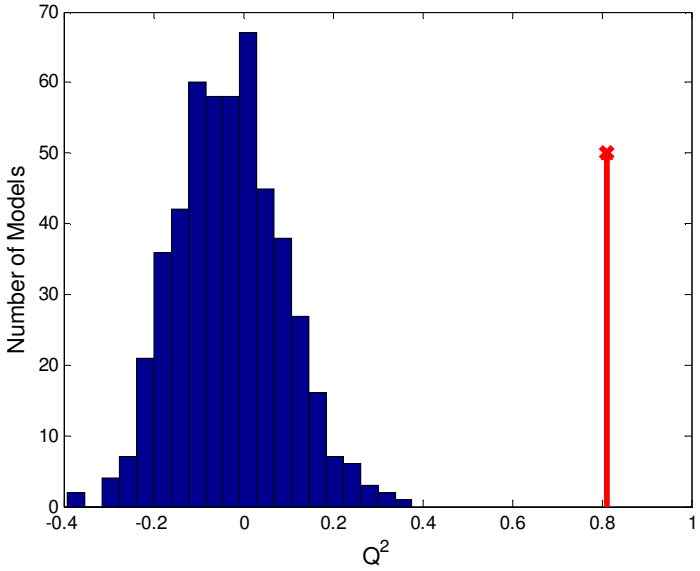

Figure S1. A cross-validated PLS-DA model of all the venous cord plasma metabolite features detected was built using two latent factors with an $R^2 = 0.88$, $Q^2 = 0.81$, and an AUC of 1. Here a reference Q^2 distribution is obtained by calculating all possible PLS-DA models under random reassignment of the case/control labels for each measured metabolic profile. If the correctly labeled model's R2 (red line) value is close to the centre of the reference distribution then the model performs no better than a randomly assigned model and is therefore invalid. A non-parametric test comparing the 'candidate' model (red line) and the permuted H_0 distribution (blue histogram) showed that the probability of a model of this quality randomly occurring was less than 0.001.

Figure S2

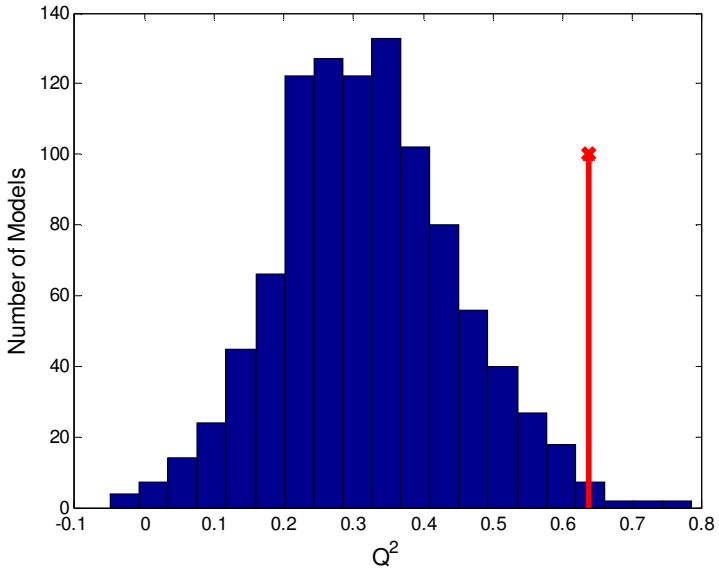

Figure S2. A cross-validated PLS-DA model of all the RUPP plasma metabolite features detected was built using 2 latent factors with an $R^2 = 0.69$, $Q^2 = 0.63$, and an AUC of 0.995. A non-parametric test comparing the 'candidate' model (red line) and the permuted H_0 distribution (blue histogram) showed that the probability of a model of this quality randomly occurring was less than 0.01.

Figure S3(a)

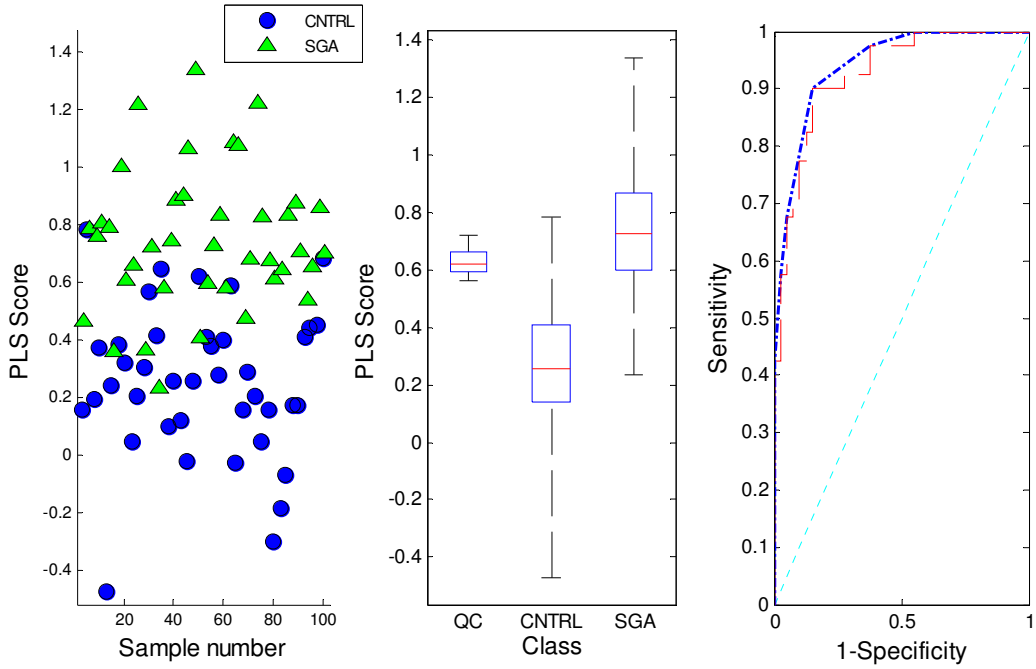

Figure S3(b)

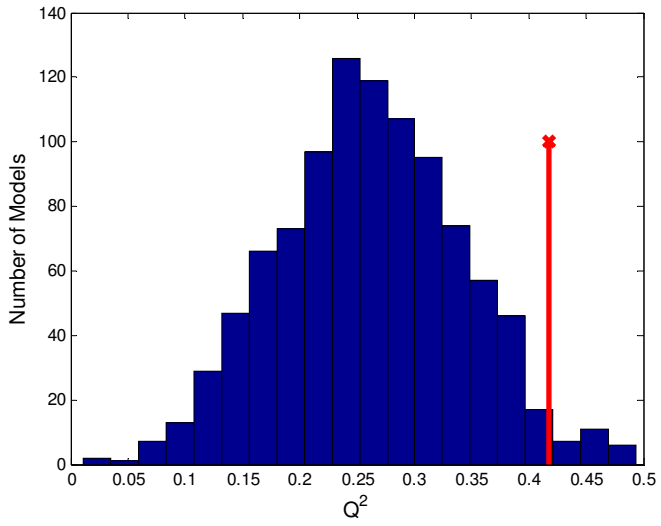

Figure S3. A cross-validated PLS-DA model (3 latent variables) constructed using the week-15 data, using only those metabolites that previously showed significant difference in the VCP study **and** were reproducibly detected in the week-15 study (n=516). This had a Q^2 = 0.48, R^2 = 0.43, AUC of 0.94 and an optimal discriminatory odds ratio of 49 (95% CI 13-184). A non-parametric test comparing the 'candidate' model (red line) and the permuted H_0 distribution (blue histogram) showed that the probability of a model of this quality randomly occurring was less than 0.05.

Figure S4(a)

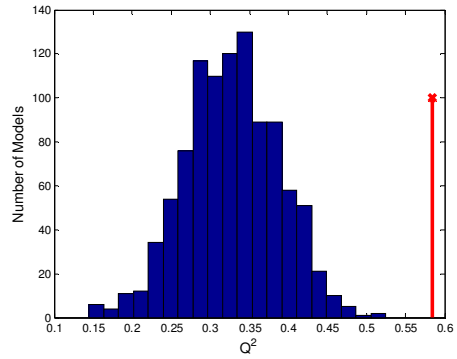

Figure S4(b)

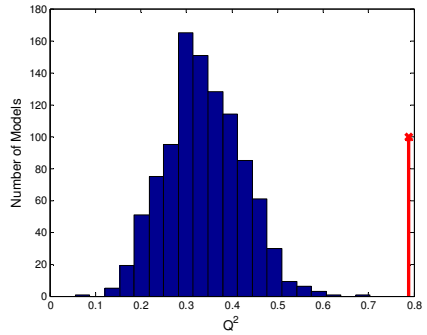

Figure S4(c)

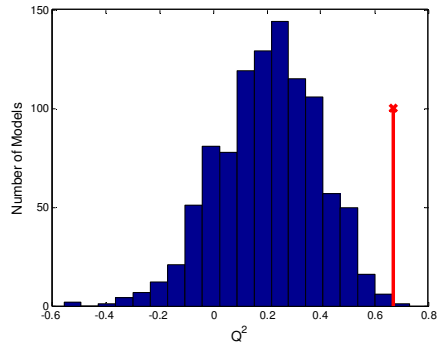

Figure S4. The PLS-DA model predictions for the final 19-metabolite signature found by the Genetic Algorithm Search program (a) Model predictions for the week-15 plasma data. R^2 = 0.61, Q^2 = 0.56, an AUC of 0.90 and an optimal odds ratios of 44 (95% CI 9 – 214). A non-parametric test comparing the 'candidate' model (red line) and the permuted H_0 distribution (blue histogram) showed that the probability of a model of this quality randomly occurring was less than 0.001. (b) Model predictions for the Venous Cord Plasma data. R^2 = 0.83, Q^2 = 0.81, and an AUC of 1. A non-parametric test comparing the 'candidate' model (red line) and the permuted H_0 distribution (blue histogram) showed that the probability of a model of this quality randomly occurring was less than 0.001. (b) Model predictions for the RUPP data. R^2 = 0.66, Q^2 = 0.65, and an AUC of 0.98. A non-parametric test comparing the 'candidate' model (red line) and the permuted H_0 distribution (blue histogram) showed that the probability of a model of this quality randomly occurring was less than 0.001.