## Study population

In 1989–1990, 32,826 NHS participants provided blood samples and completed a short questionnaire [1]. Briefly, women arranged to have their blood drawn (two 15mL sodium heparin tubes) and shipped with an ice pack, via overnight courier, to our laboratory, where it was processed and separated into plasma, red blood cell, and white blood cell components and frozen in gasketed cryovials in the vapor phase of liquid nitrogen freezers. Between 1996 and 1999, 29,611 NHSII participants provided blood samples (three 10mL sodium heparin tubes) and completed a short questionnaire [2]. Premenopausal women (n=18,521) who had not taken hormones, been pregnant, or lactated within the past 6 months provided blood samples drawn 7–9 days before the anticipated start of their next menstrual cycle (luteal phase). Other women (n = 11,090) provided a single 30-mL untimed blood sample. Samples were shipped and processed identically to the NHS samples. Briefly, whole blood samples were centrifuged for 20 minutes at 2500 RPM and 4 degrees C. Each blood sample was aliquoted into eight smaller cryovials: two 4.5mL and three 1.8mL plasma tubes, two 1.8mL white blood cell (buffy coat) tubes, and one 4.5mL red blood cell tube. Large aliquots were thawed and aliquoted to 125uL aliquots to send to the laboratory for analysis.

## Metabolite profiling

Plasma metabolites were profiled at the Broad Institute of MIT and Harvard (Cambridge, MA) using three complimentary liquid chromatography tandem mass spectrometry (LC-MS/MS) methods designed to measure polar metabolites and lipids as well as free fatty acids as described previously [3-6]. For each method, pooled plasma reference samples were included every 20 samples and results were standardized using the ratio of the value of the sample to the value of the nearest pooled reference multiplied by the median of all reference values for the metabolite. Samples from the two cohorts were run together, with matched case-control pairs (as sets) distributed randomly within the batch, and the order of the case and controls within each pair randomly assigned. Therefore, the case and its control were always directly adjacent to each other in the analytic run, thereby limiting variability in platform performance across matched case-control pairs. In addition, 64 quality control (QC) samples, to which the laboratory was blinded, were also profiled. These were randomly distributed among the participants’ samples.

Hydrophilic interaction liquid chromatography (HILIC) analyses of water soluble metabolites in the positive ionization mode were conducted using an LC-MS system comprised of a Shimadzu Nexera X2 U-HPLC (Shimadzu Corp.; Marlborough, MA) coupled to a Q Exactive mass spectrometer (Thermo Fisher Scientific; Waltham, MA). Metabolites were extracted from plasma (10 µL) using 90 µL of acetonitrile/methanol/formic acid (74.9:24.9:0.2 v/v/v) containing stable isotope-labeled internal standards (valine-d8, Sigma-Aldrich; St. Louis, MO; and phenylalanine-d8, Cambridge Isotope Laboratories; Andover, MA). The samples were centrifuged (10 min, 9,000 x g, 4°C), and the supernatants were injected directly onto a 150 x 2 mm, 3 µm Atlantis HILIC column (Waters; Milford, MA). The column was eluted isocratically at a flow rate of 250 µL/min with 5% mobile phase A (10 mM ammonium formate and 0.1% formic acid in water) for 0.5 minute followed by a linear gradient to 40% mobile phase B (acetonitrile with 0.1% formic acid) over 10 minutes. MS analyses were carried out using electrospray ionization in the positive ion mode using full scan analysis over 70-800 m/z at 70,000 resolution and 3 Hz data acquisition rate. Other MS settings were: sheath gas 40, sweep gas 2, spray voltage 3.5 kV, capillary temperature 350°C, S-lens RF 40, heater temperature 300°C, microscans 1, automatic gain control target 1e6, and maximum ion time 250 ms.

Plasma lipids were profiled using a Shimadzu Nexera X2 U-HPLC (Shimadzu Corp.; Marlborough, MA). Lipids were extracted from plasma (10 µL) using 190 µL of isopropanol containing 1,2-didodecanoyl-sn-glycero-3-phosphocholine (Avanti Polar Lipids; Alabaster, AL). After centrifugation, supernatants were injected directly onto a 100 x 2.1 mm, 1.7 µm ACQUITY BEH C8 column (Waters; Milford, MA). The column was eluted isocratically with 80% mobile phase A (95:5:0.1 vol/vol/vol 10mM ammonium acetate/methanol/formic acid) for 1 minute followed by a linear gradient to 80% mobile-phase B (99.9:0.1 vol/vol methanol/formic acid) over 2 minutes, a linear gradient to 100% mobile phase B over 7 minutes, then 3 minutes at 100% mobile-phase B. MS analyses were carried out using electrospray ionization in the positive ion mode using full scan analysis over 200–1100 m/z at 70,000 resolution and 3 Hz data acquisition rate. Other MS settings were: sheath gas 50, in source CID 5 eV, sweep gas 5, spray voltage 3 kV, capillary temperature 300°C, S-lens RF 60, heater temperature 300°C, microscans 1, automatic gain control target 1e6, and maximum ion time 100 ms. Lipid identities were denoted by total acyl carbon number and total double bond number.

Metabolites of intermediate polarity, including free fatty acids and bile acids, were profiled using a Nexera X2 U-HPLC (Shimadzu Corp.; Marlborough, MA) coupled to a Q Exactive (Thermo Fisher Scientific; Waltham, MA). Plasma samples (30 µL) were extracted using 90 µL of methanol containing PGE2-d4 as an internal standard (Cayman Chemical Co.; Ann Arbor, MI) and centrifuged (10 min, 9,000 x g, 4°C). The supernatants (10 µL) were injected onto a 150 x 2.1 mm ACQUITY BEH C18 column (Waters; Milford, MA). The column was eluted isocratically at a flow rate of 450 µL/min with 20% mobile phase A (0.01% formic acid in water) for 3 minutes followed by a linear gradient to 100% mobile phase B (0.01% acetic acid in acetonitril) over 12 minutes. MS analyses were carried out using electrospray ionization in the negative ion mode using full scan analysis over m/z 70-850. Additional MS settings are: ion spray voltage, -3.5 kV; capillary temperature, 320°C; probe heater temperature, 300 °C; sheath gas, 45; auxiliary gas, 10; and S-lens RF level 60.

Raw data from orbitrap mass spectrometers were processed using TraceFinder 3.3 software (Thermo Fisher Scientific; Waltham, MA) and Progenesis QI (Nonlinear Dynamics; Newcastle upon Tyne, UK) and targeted data from the QTRAP 5500 system were processed using MultiQuant (version 2.1, SCIEX; Framingham, MA). For each method, metabolite identities were confirmed using authentic reference standards or reference samples.

## Permutation Testing and Metabolite Grouping Analyses

Case-control status was permuted within a matched case-control pair for conditional logistic regression analyses. In the subtype analyses using unconditional logistic regression, each control was matched to a case within that analysis, preserving the initial matching criteria as much as possible. The smallest p-value across all tested metabolites in each permutation run was recorded. The permutation p-value for test of the overall null (no metabolite is associated with ovarian cancer) was estimated as k/(5,001), where k is the number of permutations where the smallest p-value (across all metabolites) was smaller than the smallest observed p-value. We estimated the permutation adjusted p-value for each metabolite by using the stepdown min P approach by Westfall and Young [7]implemented in the R package ***NPC***, which is based on the previously computed permutation p-values, all tested metabolites and their correlation structure. Metabolite Set Enrichment Analysis (MSEA) [8], implemented in the R package ***FGSEA*** [9], ranks the metabolites by the estimated beta coefficient of the association with risk and uses this metric to identify enriched metabolite groups at the two extremes of the distribution of beta estimates (positive/inverse associations). Weighted Gene Co-expression Network Analysis (WGCNA) [10], implemented in the R package ***WGCNA*** uses hierarchical clustering to identify groups of correlated metabolites, called metabolite modules, which reflect a scale-free network topology of the measured metabolites [11]. Modules were derived based on control samples only. Each module was summarized by its first principal component (PC) among all analyzed samples. A score was derived for each metabolite module based on the linear combination of measured metabolite values weighted by their corresponding loadings on the first PC summarizing the module. The score was subsequently used in conditional/unconditional logistic regressions to assess associations with risk of ovarian cancer overall and by histotypes.

1. Hankinson, S.E., et al., *Plasma prolactin levels and subsequent risk of breast cancer in postmenopausal women.* Journal of the National Cancer Institute, 1999. **91**(7): p. 629-634.

2. Tworoger, S.S., P. Sluss, and S.E. Hankinson, *Association between plasma prolactin concentrations and risk of breast cancer among predominately premenopausal women.* Cancer research, 2006. **66**(4): p. 2476-2482.

3. Mascanfroni, I.D., et al., *Metabolic control of type 1 regulatory T cell differentiation by AHR and HIF1-α.* Nature medicine, 2015. **21**(6): p. 638.

4. O’sullivan, J.F., et al., *Dimethylguanidino valeric acid is a marker of liver fat and predicts diabetes.* The Journal of clinical investigation, 2017. **127**(12): p. 4394-4402.

5. Paynter, N.P., et al., *Metabolic predictors of incident coronary heart disease in women.* Circulation, 2018. **137**(8): p. 841-853.

6. Townsend, M.K., et al., *Reproducibility of metabolomic profiles among men and women in 2 large cohort studies.* Clinical chemistry, 2013. **59**(11): p. 1657-1667.

7. Westfall, P.H., S.S. Young, and S.P. Wright, *On adjusting P-values for multiplicity.* Biometrics, 1993. **49**(3): p. 941-945.

8. Subramanian, A., et al., *Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles.* Proceedings of the National Academy of Sciences, 2005. **102**(43): p. 15545-15550.

9. Sergushichev, A., *An algorithm for fast preranked gene set enrichment analysis using cumulative statistic calculation.* BioRxiv, 2016: p. 060012.

10. Langfelder, P. and S. Horvath, *WGCNA: an R package for weighted correlation network analysis.* BMC bioinformatics, 2008. **9**(1): p. 559.

11. Barabasi, A.-L. and Z.N. Oltvai, *Network biology: understanding the cell's functional organization.* Nature reviews genetics, 2004. **5**(2): p. 101-113.