nature portfolio

https://doi.org/10.1038/s41591-022-01686-6

Supplementary information

Metabolomic and microbiome profiling reveals personalized risk factors for coronary artery disease

In the format provided by the authors and unedited

Supplementary Information

Supplementary Note 1: Metabolomics profiling by Metabolon

Samples were prepared using the automated MicroLab STAR® system from Hamilton Company. To remove protein, dissociate small molecules bound to protein or trapped in the precipitated protein matrix, and to recover chemically diverse metabolites, proteins were precipitated with methanol under vigorous shaking for 2 min (Glen Mills GenoGrinder 2000) followed by centrifugation. The resulting extract was divided into the following fractions: two for analysis by two separate reverse phases (RP)/UPLC-MS/MS methods with electrospray ionization (ESI) in the positive ionization mode, and two for analysis by RP/UPLC-MS/MS ESI and HILIC/UPLC-MS/MS ESI in the negative ionization mode, with additional details as previously described[54]. QC experimental sample served as a technical replicate ("QC-pool") throughout the experiment; extracted water samples served as process blanks; and a mixture of authentic standards that were carefully chosen not to interfere with the measurement of endogenous compounds was spiked into every analyzed sample, which allowed instrument performance monitoring and aided in chromatographic peak alignment. The instrumental variability was then determined by calculating the median relative standard deviation (RSD) of the spiked standards that were added to each sample prior to injection into the mass spectrometers. The overall process variability was determined by calculating the median RSD for all endogenous metabolites (i.e., all non-spiked metabolites) present in 100% of the QC-pools. Experimental samples were randomized across the platform run with all QC samples spaced evenly among the injections, which next allowed for intensity batch corrections using the experimental samples as feature-wise references to account for experimental batch variations and calculate corresponding correction factors. The full details of the normalization procedure, termed MED (Median Run Normalization), are described in [55].

Supplementary Note 2: Treatment-deconfounded analysis

This approach is implemented as an R package, metadeconfoundR. It functions in two steps. In the first, all associations between -omics features and the set of independent variables (disease status, drug treatment status, and risk markers including age, DM status, smoking status, sex, and BMI) are determined under nonparametric statistics (two-sided Mann-Whitney U or Spearman tests, adjusted for multiple -omics features tested using the Benjamini-Hochberg method). For each feature significantly (FDR<10%) associated with disease status (ACS vs control), it is checked whether it has significant associations with any potential confounder. If not, it is considered trivially unconfounded (NC - Not Confounded). If at least one covariate also has a significant association with the feature, then for each such covariate a post-hoc test for confounding is applied. This test takes the form of a nested linear model comparison (likelihood ratio test for P-values), where the dependent variable is the feature (X), and the independent variables are the disease status (A) and the other tested covariate (B) versus a model containing only the covariate (B), thus testing whether disease status (A) adds explanatory power beyond the covariate (B). If this holds (likelihood-ratio test P<0.05) for all covariates (B), then disease status is strictly deconfounded (SD) with regards to its effect on feature X; it cannot be reduced to any confounding factor. For each covariate (B) where significance is lost, a complementary modeling test is performed of the complementary model pairs - predicting X as a function of (A) and (B) versus a model containing (A) alone, thus testing whether the covariate (B) in turn is equally reducible to (A). If for at least one such covariate (B), (B) has an independent effect (likelihood-ratio test P<0.05) on top of (A), then the feature X is considered confounded by (B). However, if in none of the pairwise tests, the original significance holds, then (A) and (B) are considered so correlated that their relative influence cannot be disentangled. We consider these cases laxly deconfounded (LD), in the sense that for these cases clear confounding influence cannot be concluded, but also not ruled out. The R package was applied to the present dataset considering cardiometabolic disease treatment status as binary variables considered at the level of broader pharmacopoeial classes with shared modes of action. The outcome of this analysis is included as **Supplementary Table 7,9**, which lists the resulting status of all significant ACS-associated features, highlighting some as direct features of ACS (whether cause or effects), with others more parsimoniously explained as indirect associations, mediated via the confounding features.

References

54. Bridgewater BR, E. A. High Resolution Mass Spectrometry Improves Data Quantity and Quality as Compared to Unit Mass Resolution Mass Spectrometry in High-Throughput Profiling Metabolomics. *Metabolomics* **04**, (2014).
55. Wulff, J. E. & Mitchell, M. W. A comparison of various normalization methods for LC/MS metabolomics data. *ABB* **09**, 339–351 (2018).