Study Design

Prior to any nutritional metabolomic analyses on human subjects the study must be carefully designed and implemented. Fasting plasma/serum samples are collected before and after the nutritional intervention study. Food consumption is monitored by 4- or 7-day food diaries in predefined days or food frequency questionnaires which are revised by trained personnel together with subjects during visits to research center. Dietary intakes of nutrients are calculated based on nutritional composition of consumed food items by software specified for dietary calculations (e.g. NUTRICA® or AivoDiet, Aivo Finland).

When the study focuses on causal effects of e.g. bilberries, fatty fish and whole grains, randomized controlled trial (RCT) design is applied. Subjects are randomized to treatment and control groups. Responses to dietary food items are then analyzed by comparing treatment group to control group (1). RCT generally provides evidence of changes on presumably homogenous study population.

Metabolomics has been successfully applied to population-based cohorts. Noerman et al (2)(2018) identified several metabolites which may link higher egg intake with lower risk of development of type 2 diabetes. The samples used by Noerman et al. originated from the Kuopio Ischaemic Heart Disease Risk Factor Study (KIHD) in Eastern Finland(2). Studies like this are characterized by a long follow-up time and a high number of participants from a selected population that usually originates from or lives in certain geographic area and fits to pre-set criteria. In the KIHD study the emphasis was to investigate risk factors for cardiovascular diseases and atherosclerosis in middle-aged men from eastern Finland. This ongoing prospective cohort started in the 1984-1989 and consisted of 2682 participants with eligibility rate of 83%. The baseline dietary intake was calculated as the daily mean of a 4-day food record. Venous blood sample collection took place during the examination visit at baseline after an overnight fast and restriction of alcohol consumption 3 days prior to the visit. After the sample collection, the serum samples were stored at -80 °C until metabolomics analysis was performed. A subcohort based on 2x2 grouping according to different levels of dietary intake (e.g. eggs) and the incidence of diseases (e.g. type 2 diabetes), as performed by Noerman and colleagues, could be useful to determine the potential molecular link between diet and disease using non-targeted metabolomic approach.

A third type of study design commonly used to describe metabolic response of the diet can be classified as postprandial. In this kind of study set-up an immediate response of nutrients is investigated by comparing samples between the fasting stage to those taken after test food. Postprandial study setup combined with metabolomics has been able to separate the metabolic response of breakfast meals modelling vegan, lacto-ovo vegetarian, and omnivore diets in serum(3). Differences of the response of different metabotypes can also be studied (4).

Tutorial on Pathway Analyses Tools

a. MetaboAnalyst (https://www.metaboanalyst.ca/)
b. Enrichment Analysis
    i. Paste the list of the HMDB or KEGG ID to the list of a compound names for the over-representation analysis in the Enrichment Analysis module of MetaboAnalyst 4.0
    ii. Specify the ID type, then click "Submit"
    iii. Choose "pathway-associated metabolite sets (SMPDB)" and use only metabolites sets containing at least 2 compounds
    iv. Upload a reference metabolome based on your analytical platform or use all compounds in the selected metabolite set library, then click "Submit"
    v. Note that the results would be available both in network and bar view, then click "Submit" to be redirected to the download links.
    vi. Download download.zip to download all results files, or select only certain files you need.
c. Pathway Analysis
    i. Paste the list of the HMDB or KEGG ID in the Pathway Analysis module in MetaboAnalyst 4.0
    ii. Choose *Homo sapiens* KEGG as the reference library for human-based samples, or other libraries depending on sample origin
    iii. Depending on what you want, choose the test for over-representation analysis and pathway topology. We use Fischer exact test for and relative-betweenness centrality, respectively

iv. Upload your reference metabolome, or use all compounds in the selected pathway
v. Import the results

d. Network Explorer
i. Paste the list of the HMDB or KEGG ID in the Network Explorer module in MetaboAnalyst 4.0
ii. Check the list of the metabolites, delete the unrecognized ones
iii. Choose "metabolite-metabolite interaction network" mode if you only upload the list of metabolites.
iv. Click "Proceed" to view the network
v. Click "Download" to import the results

2. MetScape (http://metscape.ncibi.org/).

a. Build a network:

a. Open the Cytoscape software.
b. Choose the "Apps" menu on the tool bar.
c. Choose the first option "App Manager" from the dropdown menu.
d. In App Manager window, in the search box, enter "MetScape". MetScape should appear in the second column.
e. Click on "MetScape" and then Click on "Install" at the bottom of the window. Once the app is installed, it will appear in the "Apps" menu. One-time free registration is required the first time, the MatScape is opened.
f. From "Apps" menu, "MetScape", click the "Built Network" and then "Pathway-based".
g. In "Input" section, choose your Organism (Human, Plant, or Mouse).
h. In "Import" section, "Select" button upload the experimental data (*.CSV). The MetScape main window has three tabs that provide users with the following options Load a list of compounds, one compound per line (compound names, KEGG IDs, HMDB IDs, BIGG IDs, EHMN IDs), or load a file containing normalized experimental metabolite data with metabolite KEGG IDs and corresponding values at given time points or under specific experimental conditions.
i. Under "Options" section, "Network Type", Choose pathway-specific networks by choosing a pathway from the drop-down list.
j. click "Build Network" at the bottom of Control Panel.

b. Correlation Calculator (optional step)

Correlations are measures between pairs of metabolites. Correlation Calculator is a standalone Java application that provides methods of

calculation pairwise correlations among repeatedly measured entities. It is designed for use with quantitative metabolite measurements, such as Mass Spectrometry data, on a set of samples. The workflow allows inspection and/or saving of results at various stage, and the final correlation results can be dynamically imported into version 3.1 or higher of MerScape as a correlation network.

a) The Correlation Calculator can be downloaded from MetScape website (http://metscape.ncibi.org/CorrelationCalculator-1.0.1.jar).
b) The input data file is a CSV file that contains a table of measurements across multiple samples. Although metabolites must be labeled, sample labels are optional.
c) Samples may be in rows or columns.
d) After launching the calculator, click the browse button.
e) Select the appropriate data file and click Open (make sure to specify the file format.
f) Under, "Data Normalization", Select "Log2-Transform Data" AND "Autoscale Data" and click "Run".
g) Click "View Normalized Data" to view the results. To save the data click the "Save" button. if the data are already normalized before loading it into the calculator, this normalized step can be skipped.
h) Pearson's Correlations is performed to filter out metabolites; this step is optional. To use Pearson's Correlations, click "Run" under "Data Alalysis". Histogram and heatmap view are available for this analysis.
i) The last step is to use a Partial Correlation Method, either Debiased Sparse Partial Correlation (DSPC) or Basic Partial Correlation, and then click "Run". The Correlation Calculator calculates the partial correlation values, *p*-values, and *q*-values for each compound pair.
j) To view the result in the MetScape, click "View in MetScape" where interactive visualization and exploration can be performed.

c. Correlation network:

To build a correlation network in MetScape, appropriate data file formatting is required. Two types of data file formats are accepted. The first data file format is column-based (recommended format). The first row of column-based file must have column heading of the user's choosing. The first two columns must contain metabolite names or IDs. Additional columns contain values such as *p*-values, *q*-values, and correlation values. The second data file format is a matrix format, where the first row and the first column contain metabolite names, and the rest of the rows and columns contain correlation values.

a) Open the Cytoscape

b) Go to the “Apps” menu and click on “MetScape”.
c) Select “Build Network” and then “Correlation-based”. Now, a MetScape tap displays on the left side of the screen, in the Control Panel.
d) Under the “Input” section, click the “Select” button. Select the location of the correlation file and click open.
e) A new window will appear showing potential matches found in the MetScape database, for each compound in the input file.
f) Use the dropdown arrows for each compound to choose the best match. If the compound is not found in the system, it will say “Not Found”. Your mapping selection will be saves, so that it will appear as the default option in the future.
g) Select “OK”.
h) Under “Edge Mapping” in Control Panel, use the dropdown menu next to “Base Edges on” and select the appropriate column from your data file (e.g. correlation values column.
i) Under “Edge Mapping” in Control Panel, use the dropdown menu next to “Tooltip Labels” and select the appropriate column from your data file (e.g. *p*-values column).
j) Click “Build Network”.

References

1. Hanhineva K, Lankinen MA, Pedret A, Schwab U, Kolehmainen M, Paananen J, et al. Nontargeted Metabolite Profiling Discriminates Diet-Specific Biomarkers for Consumption of Whole Grains, Fatty Fish, and Bilberries in a Randomized Controlled Trial. J Nutr [Internet]. 2015 Jan 1 [cited 2019 Dec 12];145(1):7–17. Available from: https://academic.oup.com/jn/article/145/1/7/4644407

2. Noerman S, Kärkkäinen O, Mattsson A, Paananen J, Lehtonen M, Nurmi T, et al. Metabolic Profiling of High Egg Consumption and the Associated Lower Risk of Type 2 Diabetes in Middle-Aged Finnish Men. Mol Nutr Food Res [Internet]. 2018 Dec 12 [cited 2019 Dec 12];1800605. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/mnfr.201800605

3. Rådjursöga M, Lindqvist H, Pedersen A, Karlsson B, Malmodin D, Ellegård L, et al. Nutritional Metabolomics: Postprandial Response of Meals Relating to Vegan, Lacto-Ovo Vegetarian, and Omnivore Diets. Nutrients [Internet]. 2018 Aug 10 [cited 2020 Jan 22];10(8):1063. Available from: http://www.mdpi.com/2072-6643/10/8/1063

4. Fiamoncini J, Rundle M, Gibbons H, Thomas EL, Geillinger-Kästle K, Bunzel D, et al. Plasma metabolome analysis identifies distinct human metabotypes in the postprandial state with different susceptibility to weight loss–mediated metabolic improvements. FASEB J [Internet]. 2018 Oct [cited 2020 Jan 22];32(10):5447–58. Available from: https://www.fasebj.org/doi/10.1096/fj.201800330R