Documentation for the Metabolomics Datasets
Dr. Rachel Murphy, Health ABC Ancillary Study #AS13-147AM

1. General description ……..2
2. Methods……..2
3. Statistical recommendations……..……..3
4. Dataset structure and contents……...……3
5. References……..4

1. General description

Metabolomics is the study of small molecules participating in cellular metabolism. Advances in metabolic profiling technologies have made it possible to simultaneously assay hundreds of metabolites, providing insight into an organism's metabolic status. Several studies suggest that metabolomics may be novel biomarkers for a diverse range of disease (1-3). Further, metabolites may play important regulatory roles in disease pathways (4, 5) and even serve as effectors of disease processes (6). Metabolomics has only recently been applied to epidemiologic studies, with data predominantly supplied by two providers: the Broad Institute and Metabolon Inc. Each of these providers' supplies identified metabolites, with an estimated overlap of ~150 metabolites. The reliability of metabolite values for both platforms has been validated (7, 8).

a. Broad dataset (metabolomics_broad.sas7bdat)
319 African American men were selected at random from all African American men with stored plasma from the Health ABC Year 2 visit.

b. Broad ICC dataset (metabolomics_broadicc.sas7bdat)
Contains the interclass correlations, an indication of the reliability of the analyses for blind duplicates.

c. Metabolon dataset (metabolomics_metabolon.sas7bdat)
40 African American men were selected at random from the sample of 319 African American men. Metabolites were analyzed.

2. Methods

a. Broad dataset
Metabolites were measured in fasting Health ABC Year 2 EDTA plasma that had not been previously thawed at the Broad Institute of MIT and Harvard, under the direction of Dr. Clary Clish. Metabolites were analyzed using a combination of liquid chromatography tandem mass spectrometry methods that measure complimentary sets of metabolite classes, ranging from polar metabolites such as organic acids to non-polar lipids such as triglycerides. The Broad platform uses three approaches: Method 1: Amines and polar metabolites that ionize in the positive ion mode including amino acids, amino acid metabolites, dipeptides, and other cationic metabolites. Method 2: Central metabolites and polar metabolites that ionize in the negative ion mode. These metabolites include sugars, sugar phosphates, organic acids, purine and pyrimidines. Method 3: Polar and non-polar lipids. Metabolite signals were retained as measured LC/MS/MS peak areas, which are proportional to metabolite concentration. Details of sample extraction and metabolite profiling can be found in Townsend et al. (7).

b. Broad ICC dataset
Duplicates (5% of the total sample, N=17) were chosen at random. A macro developed by Donna Spieglman at Harvard University was used (%icc9) which is publicly available (http://www.hsph.harvard.edu/donna-spiegelman/software/).

c. Metabolon dataset
Metabolon Inc. used a single non-targeted extraction with methanol followed by protein precipitation to recover a diverse set of metabolites. Samples were then analyzed using ultra high performance liquid-phase chromatography and gas chromatography coupled with mass spectrometry and tandem mass spectrometry. The mass spectra peaks were compared to a chemical reference library generated from 2,500 standards to identify individual metabolites. Samples that were unknown are included in the dataset (see description in 4. below). Details on the analytical method have been previously published (8, 9).

3. Statistical recommendations

ICC's rather than CV's are provided because ICCs are more robust to the use of relative concentrations which is the unit of measure for metabolites.

Normalization of metabolites using log-transformation and grand mean centering are recommended. Additionally, for the Metabolon dataset we recommend normalizing for batch (SAMPLE_BOX_NUMBER) and run day (SAMPLE_RUN_DAY) (10).

4. Dataset structure and contents

The datasets contain a single observation per participant (n=319 for Broad; n=40 for Metabolon).

Care should be taken if merging the Broad and Metabolon datasets because a number of metabolites overlap between the analytical platforms and thus, there are duplicate metabolite variable names.

Metabolite variable names/labels can be cross-referenced with the Human Metabolome database to obtained detailed information about the metabolite (http://www.hmdb.ca).

Unknown metabolites in the Metabolon dataset have variable names with the format "X_12346_2".

To download the Metabolomics datasets, please click on "Datasets/Documentation" in the banner on the Health ABC website, then click on "Current Datasets" and select the "Metabolomics" link under the "Biospecimens (all years)" section of downloadable datasets. The 3 datasets are contained in the Metabolomics.exe downloadable file.

To download this documentation, please click on "Datasets/Documentation" in the banner on the Health ABC website, click on "Biospecimens (all years)" in the "Dataset Documentation" section of the webpage, and then select "Metabolomics Documentation."

Key variables: HABCID HABC Enrollment ID without the 2-letter prefix

To use the data, please contact the PI at your site.

5. References

1. Wang TJ, Larson MG, Vasan RS, Cheng S, Rhee EP, McCabe E, et al. Metabolite profiles and the risk of developing diabetes. Nature medicine. 2011;17:448-453.
2. Sabatine MS, Liu E, Morrow DA, Heller E, McCarroll R, Wiegand R, et al. Metabolomic identification of novel biomarkers of myocardial ischemia. Circulation. 2005;112:3868-3875.
3. Hische M, Larhlimi A, Schwarz F, Fischer-Rosinsky A, Bobbert T, Assmann A, et al. A distinct metabolic signature predicts development of fasting plasma glucose. Journal of clinical bioinformatics. 2012;2:3.
4. Sapieha P, Sirinyan M, Hamel D, Zaniolo K, Joyal JS, Cho JH, et al. The succinate receptor GPR91 in neurons has a major role in retinal angiogenesis. Nature medicine. 2008;14:1067-1076.
5. He W, Miao FJ, Lin DC, Schwandner RT, Wang Z, Gao J, et al. Citric acid cycle intermediates as ligands for orphan G-protein-coupled receptors. Nature. 2004;429:188-193.
6. Newgard CB, An J, Bain JR, Muehlbauer MJ, Stevens RD, Lien LF, et al. A branched-chain amino acid-related metabolic signature that differentiates obese and lean humans and contributes to insulin resistance. Cell metabolism. 2009;9:311-326.
7. Townsend MK, Clish CB, Kraft P, Wu C, Souza AL, Deik AA, et al. Reproducibility of metabolomic profiles among men and women in 2 large cohort studies. Clin Chem. 2013;59:1657-1667.
8. Sreekumar A, Poisson LM, Rajendiran TM, Khan AP, Cao Q, Yu J, et al. Metabolomic profiles delineate potential role for sarcosine in prostate cancer progression. Nature. 2009;457:910-914.
9. Evans AM, DeHaven CD, Barrett T, Mitchell M, Milgram E. Integrated, nontargeted ultrahigh performance liquid chromatography/electrospray ionization tandem mass spectrometry platform for the identification and relative quantification of the small-molecule complement of biological systems. Anal Chem. 2009;81:6656-6667.
10. Moore SC, Matthews CE, Sampson JN, Stolzenberg-Solomon RZ, Zheng W, Cai Q, et al. Human metabolic correlates of body mass index. Metabolomics. 2014;10:259-269.