Corresponding Author: Cristina Menni

Date: 12/03/2018

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work we publish. This form is published with all life science papers and is intended to promote consistency and transparency in reporting. All life sciences submissions use this form; while some list items might not apply to an individual manuscript, all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Experimental design

1. Sample size

   Describe how sample size was determined.

   This is the first study to examine fecal metabolites, therefore there is no prior knowledge on the effect sizes we are to expect.
   However, having included 786 individuals with fecal metabolites, anthropometric, microbial and GWAS data our study has 80% power to detect a difference of 0.175 SD for the anthropometric measures adjusting for multiple testing (FDR<0.05) and 80% power to detect a difference of 0.40 for GWAS metabolites ratio adjusting for 31226 tested ratios. The largest effect size found by us in the GWAS is -0.17 for a metabolite ratio with mean=0 SD=0.30 (i.e. an effect of 0.56 SD) which is larger than the effect needed to achieve 99% power for a p< 2 x 10[-12] needed for the GWAS of metabolite ratios (adjusting for 31226 tested ratios)

2. Data exclusions

   Describe any data exclusions.

   No data were excluded from the analysis

3. Replication

   Describe whether the experimental findings were reliably reproduced.

   We replicated our GWAS-metabolomics results in an independent sample of 230 individuals

4. Randomization

   Describe how samples/organisms/participants were allocated into experimental groups.

   This is not a clinical trial and so randomization is not relevant

5. Blinding

   Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

   This is not a clinical trial and so blindness is not relevant

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or the Methods section if additional space is needed).

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☐	☒	A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly.
☐	☒	A statement indicating how many times each experiment was replicated
☐	☒	The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
☐	☒	A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☐	☒	The test results (e.g. *p* values) given as exact values whenever possible and with confidence intervals noted
☐	☒	A summary of the descriptive statistics, including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
☐	☒	Clearly defined error bars

See the web collection on statistics for biologists for further resources and guidance.

Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.

R version 3.3.3
R package mets (version 1.1.1) to estimate heritability
GEMMA for the GWAS

For all studies, we encourage code deposition in a community repository (e.g. GitHub). Authors must make computer code available to editors and reviewers upon request. The *Nature Methods* guidance for providing algorithms and software for publication may be useful for any submission.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.

N/A

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

N/A

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.

N/A

b. Describe the method of cell line authentication used.

N/A

c. Report whether the cell lines were tested for mycoplasma contamination.

N/A

d. If any of the cell lines used in the paper are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.

N/A

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.

N/A

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.

We included 786 individuals, predominantly females (93.4%), with an average age of 65.2 (±7.6) and an average of BMI of 26.1 (±4.7) from the TwinsUK cohort. Results of the genome-wide association study were replicated in an independent set of 230 individuals (98.3% female) from the TwinsUK study, aged 66.9 (±8.6) and an average BMI of 27.2 (±5.2). Metabolite concentrations were measured from fecal samples by Metabolon Inc., Durham, USA, using an untargeted LC/MS platform. 16S rRNA microbial sequencing was extracted from fecal samples using Illumina MiSeq.
Genotypes were determined using whole-genome sequencing Ethical approval by St Thomas' Hospital ethics committee; all participants provided informed written consent.