nature research

Corresponding author(s): Eran Segal

Last updated by author(s): Dec 30, 2021

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No software was used for data collection
Data analysis	Python 3.7.4, with packages: pandas 1.2.5, numpy 1.21.0, scikit-learn 0.24.2, scipy 1.7.0, shap 0.35.0, LightGBM 2.3.1, statsmodels 0.12.2. Bowtie 2 version 2.3.4.2. hg19 human reference genome. GEM version 1.376.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The raw metagenomic sequencing data per sample of our controls are available from the European Nucleotide Archive (ENA; https://www.ebi.ac.uk/ena): PRJEB11532. The raw metabolomics data and phenotypes per sample of our controls are available from the European Genome-phenome Archive (EGA; https://ega-archive.org/): EGAS00001004512. The raw metabolomics data and full clinical phenotypes for our cohort of individuals with ACS are available from the EGA: EGAS00001005342.
Additional data regarding SGB 4712, including the genome sequence, gene annotation and closest references are available at https://github.com/noambar/ACStudy/tree/master/SGB_4712.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	This is an exploratory multi-omic study, in which thousands of variables were compared between two cohorts. Given the limited prior knowledge of the size of the effect of the variables examined in this study, sample size calculation was infeasible. The number of cases included in this study was determined by the volume of patients meeting the inclusion criteria who arrived at the recruiting medical centers during the recruitment period. Throughout this study, results presented are those that reached statistical significance following multiple hypothesis correction. Thus, it is likely that with greater sample size, additional findings of lower effect sizes may be detected.
Data exclusions	No data was excluded.
Replication	Technical details regarding sample collection, DNA extraction, and metagenomic sequencing, and initial QC are available in our companion manuscript (Fromentin et al.). Metagenomics data from the MetaCardis study went through an identical QC and mapping pipeline as our control and ACS cohorts to produce estimates of the relative abundances of bacterial representative genomes. To test whether the CAD-related depletion of SGB 4712 replicates in the MetaCardis study, we applied the Mann-Whitney U test over the relative abundances of SGB 4712, comparing individuals with IHD to each of the three other sub cohorts (HC, MMC, UMCC). To replicate the associations of the 15 serum metabolites mentioned in Supplementary Table 10 with SGB 4712, we computed the Spearman correlation between their levels and the relative abundance of SGB 4712 within the healthy cohort (n=273; Supplementary Table 11). The sign of the correlation coefficient for all 15 metabolites with SGB 4712 replicated, with 10 of these associations remaining significant (FDR<10%). Full technical details regarding serum sample collection, preparation, and metabolomics analyses in the MetaCardis study are available in our companion manuscript (Fromentin et al.). The metabolomics data considered for replication in this study included normalized intensities of 859 metabolites measured by the untargeted LC-MS platform of Metabolon that overlapped with the measured metabolites in our study. To replicate the excessed prediction of BMI in subjects with IHD, we applied the exact same procedure as in our cohort. Here, the model was trained based on the normalized serum samples of a random subset of 383 out of 702 non-IHD individuals (from the HC, MMC, and UMCC subgroups). We then applied and evaluated the model on two held-out test sets, including one of 319 subjects with IHD and another of the remaining 319 non-IHD controls. In addition, we replicated the feature attribution analysis of the main metabolite contributors to the overprediction of BMI in individuals with IHD. To that end, we applied the same procedure as in our cohort and identified the metabolites, whose SHAP values were significantly associated with ?BMI (FDR 1%; Supplementary Table 20; adjusted for BMI and T2DM status). Additional replications are reported in our companion manuscript (Fromentin et al.). These include validation of MetaCardis' ACS-specific markers in our study subjects using two approaches: (1) Significant correlation between the ACS-specific markers identified in the MetaCardis study exhibiting comparative effect sizes in our study samples (p=4.4·10-11); (2) Effective performances of classifier models (OPLS-DA) built using MetaCardis' ACS-specific biomarkers. Models were trained in the MetaCardis population and tested in our study samples (AUC=0.92).
Randomization	Being a case-control study, this study did not include any randomization. When comparing the groups, covariates were controlled by a performing a 1:1 matching for age, sex, BMI, DM, and smoking status.
Blinding	There was no group allocation in this study.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Human research participants

Policy information about studies involving human research participants

Population characteristics	Full population characteristics are described in supplementary table 1. Population characteristics for the matched cohorts used for metabolomics and microbiome comparisons are described in supplementary tables 4, 5.
Recruitment	The participants were recruited during hospitalization in the Cardiology department at Rabin Medical Center as close as possible to the index event (within 72 hours) and only after signing an informed consent form (ICF). Therefore, critically ill patients, intubated patients, hemodynamically unstable patients or physical conditions that did not allow patients to sign ICF were not recruited to the study. The inclusion criteria were ACS patients aged between 30 and 80. The exclusion criteria were antibiotic usage in the past 3 months, bariatric surgery or intestinal resection except for appendectomy, inflammatory bowel disease, active cancer, infectious diseases (including hepatitis B or C and human immunodeficiency viruses), autoimmune disease, patients with past history of organ transplantation or receiving immunosuppressive therapy, or patients with drug or alcohol addiction. It may cause a bias toward recruitment of "healthier" participants to the study. The participants answered detailed medical, lifestyle, and nutritional questionnaires, provided stool and serum samples for metagenomic sequencing and metabolomics. Both blood and stool samples were not taken under strict fasting conditions (similar as for the control cohort). This may introduce biases to the metabolic signatures that are caused by different postprandial states.
Ethics oversight	The study was approved by the Ethics Committee at Rabin Medical Center (Petah-Tikva, Israel) and the ethical comity at Weizmann Institute of Science (Rehovot, Israel), approval number RMC-622-16. All participants signed written informed consent forms.

Note that full information on the approval of the study protocol must also be provided in the manuscript.