nature medicine

ARTICLES

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

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

Yeela Talmor-Barkan[1,2,3,4,15], Noam Bar[3,4,15], Aviv A. Shaul[1,2], Nir Shahaf[3,4], Anastasia Godneva[3,4], Yuval Bussi[3,4], Maya Lotan-Pompan[3,4], Adina Weinberger[3,4], Alon Shechter[1,2], Chava Chezar-Azerrad[1,2], Ziad Arow[1,2], Yoav Hammer[1,2], Kanta Chechi[5,6,7], Sofia K. Forslund[8,9,10,11], Sebastien Fromentin[12], Marc-Emmanuel Dumas[5,6,13], S. Dusko Ehrlich[12], Oluf Pedersen[14], Ran Kornowski[1,2,16] and Eran Segal[3,4,16] ✉

Complex diseases, such as coronary artery disease (CAD), are often multifactorial, caused by multiple underlying pathological mechanisms. Here, to study the multifactorial nature of CAD, we performed comprehensive clinical and multi-omic profiling, including serum metabolomics and gut microbiome data, for 199 patients with acute coronary syndrome (ACS) recruited from two major Israeli hospitals, and validated these results in a geographically distinct cohort. ACS patients had distinct serum metabolome and gut microbial signatures as compared with control individuals, and were depleted in a previously unknown bacterial species of the *Clostridiaceae* family. This bacterial species was associated with levels of multiple circulating metabolites in control individuals, several of which have previously been linked to an increased risk of CAD. Metabolic deviations in ACS patients were found to be person specific with respect to their potential genetic or environmental origin, and to correlate with clinical parameters and cardiovascular outcomes. Moreover, metabolic aberrations in ACS patients linked to microbiome and diet were also observed to a lesser extent in control individuals with metabolic impairment, suggesting the involvement of these aberrations in earlier dysmetabolic phases preceding clinically overt CAD. Finally, a metabolomics-based model of body mass index (BMI) trained on the non-ACS cohort predicted higher-than-actual BMI when applied to ACS patients, and the excess BMI predictions independently correlated with both diabetes mellitus (DM) and CAD severity, as defined by the number of vessels involved. These results highlight the utility of the serum metabolome in understanding the basis of risk-factor heterogeneity in CAD.

CAD remains a major cause of morbidity and mortality worldwide[1] despite tremendous advances in prevention, diagnosis and treatment[2]. CAD is a complex physiological process that may manifest in multiple, possibly interacting risk factors[3]. Current treatments for CAD are based on traditional and modifiable CAD risk factors and result in only partial success, which is emphasized by the high recurrence rate of cardiovascular disease (CVD) in patients with balanced traditional risk factors[4].

Blood serves as a liquid conveyor for molecules inside the body[5]. Of particular importance are the thousands of circulating small molecules, termed the serum metabolome, which provide unique insights into biological processes, and a valuable source for studying the multifactorial nature of CAD.

The gut microbiome is actively involved in the metabolism of blood metabolites. Several gut-microbiota-derived circulating metabolites are associated with CVD. Trimethylamine N-oxide was established as a marker for CVD in humans[6], with further evidence indicating pro-atherogenicity[7] and prothrombotic[8] in mouse models. Indoxyl sulfate is produced in the liver after degradation of tryptophan by bacterial tryptophanase, and was shown to be associated with arterial stiffness and peripheral vascular disease[9]. P-cresol is a product of colonic bacterial fermentation from phenylalanine and tyrosine and was shown to correlate with increased cardiovascular events[10]. In light of the above, we assume that many yet unknown factors contribute to the mechanism of CAD. With recent technological advances, it is now feasible to study these unknown factors in multi-omic data as well as to identify person-specific omic signatures of CAD, thus promoting a personalized medicine approach.

Here, we recruited a cohort of 199 patients with ACS, for whom we obtained a multi-omic characterization, including metagenomic

[1]Department of Cardiology, Rabin Medical Center, Petah Tikva, Israel. [2]Sackler Faculty of Medicine, Tel-Aviv University, Tel-Aviv, Israel. [3]Department of Computer Science and Applied Mathematics, Weizmann Institute of Science, Rehovot, Israel. [4]Department of Molecular Cell Biology, Weizmann Institute of Science, Rehovot, Israel. [5]Genomic and Environmental Medicine, National Heart and Lung Institute, Faculty of Medicine, Imperial College London, London, UK. [6]Division of Systems Medicine, Department of Metabolism, Digestion and Reproduction, Faculty of Medicine, Imperial College London, London, UK. [7]School of Public Health, Faculty of Medicine, Imperial College London, Medical School Building, St Mary's Hospital, London, UK. [8]Experimental and Clinical Research Center, a cooperation of Charité-Universitätsmedizin and the Max-Delbrück Center, Berlin, Germany. [9]Max Delbrück Center for Molecular Medicine (MDC), Berlin, Germany. [10]MCharité University Hospital, Berlin, Germany. [11]DZHK (German Centre for Cardiovascular Research), Berlin, Germany. [12]University College London, Department of Clinical and Movement Neurosciences, London, UK. [13]European Genomics Institute for Diabetes, UMR1283/8199 INSERM, CNRS, Institut Pasteur de Lille, Lille University Hospital, University of Lille, Lille, France. [14]Novo Nordisk Foundation Center for Basic Metabolic Research, University of Copenhagen, Copenhagen, Denmark. [15]These authors contributed equally: Yeela Talmor-Barkan, Noam Bar. [16]These authors jointly supervised this work: Ran Kornowski, Eran Segal. ✉e-mail: eran.segal@weizmann.ac.il

sequencing of the gut microbiome, and untargeted serum metabolomics. We identified a previously unknown bacterial species of the *Clostridiaceae* family that was depleted in ACS patients, and replicated in a geographically distinct cohort, and showed that it is associated with the levels of multiple circulating metabolites. We further identified a metabolic profile of ACS that intersects with a diverse set of established CAD risk factors, including diabetes, genetics, diet and microbiome. In addition, we found that metabolic deviations of ACS patients are person specific and correlate with clinical parameters and CVD outcomes. We further demonstrated that metabolic aberrations linked with microbiome and diet show a gradual trend in control participants with metabolic impairment, suggesting their involvement in earlier dysmetabolic phases preceding clinically overt CAD. Finally, a metabolomic-based model of BMI, predicted excess BMI levels for patients with ACS that correlated with an impaired glycemic status and disease severity.

Results

Comprehensive multi-omic characterization of ACS. We recruited 199 participants with ACS at the Rabin Medical Center in Israel, for whom we obtained clinical and multi-omic profiling (Methods). These include demographics, anthropometrics, medical parameters, gut microbiome and serum metabolomics using two complementary platforms for a subset of 156 and 191 participants, respectively. For direct case-control comparisons, we leveraged a recently established large cohort[11,12], consisting of 970 non-ACS individuals, for which we measured the above data (Supplementary Table 1). We obtained omics data, using identical methodologies and platforms in subgroups of controls, which we further sampled for individuals matching for baseline characteristics (for the full cohort selection process see Extended Data Fig. 1 and Methods).

To profile the gut microbiome, we used shotgun metagenomics sequencing of stool samples, followed by an in-house computational pipeline (Methods). The serum metabolomics included two complementary platforms: (1) untargeted mass spectrometry measured the levels of 961 metabolites, including lipids, amino acids, xenobiotics, carbohydrates, peptides, nucleotides and approximately 30% un-named compounds (Supplementary Table 2); (2) additional 228 absolute-value-based plasma metabolites and ratios were measured by the proton nuclear magnetic resonance (^{1}H-NMR) platform of Nightingale Health, expanding the detailed lipidomic profiles and adding measurements of clinically validated biomarkers (Supplementary Table 3). All biological samples (stool, serum) were collected near the index event (within 72 hours) to attenuate biological noise.

Broad gut microbiome and serum metabolomics patterns of ACS. Multiple previous studies had linked the meta-organismal pathway with CAD[13,14]. Serum metabolites are known to play a key role in mediating the metabolic and immune interactions between the microbiome and its host, thus providing a fundamental view into the complex dynamics of environmental exposures. To create a microbial and metabolic map under ACS, we compared the gut microbiome and serum metabolome profiles of the ACS patients with non-ACS controls, matched for potential microbiome and metabolome confounders, including age, sex, BMI, smoking status and DM (Supplementary Tables 4 and 5; Methods). As the impact of drug intake on the gut microbiome and serum metabolome was extensively demonstrated[15,16], we applied a recently developed analytical pipeline to de-confound microbiome and metabolome alterations from drug usage[17] (Methods).

We found that the serum profile of patients with ACS exhibits a broad set of perturbations in serum metabolite levels (Fig. 1a and Supplementary Table 6), including 533 significantly altered metabolites (10% false-discovery-rate (FDR) adjusted). Of these, only 29 associations of metabolites were likely to be confounded by medication usage (Supplementary Table 7). Notably, we found that the serum metabolome of ACS followed a major depletion pattern, as 358 metabolites (67%) measured higher on average in control participants (Extended Data Fig. 2a). This trend, however, was inconsistent across major biological pathways (Extended Data Fig. 2b).

Type 2 DM (T2DM) is prevalent among patients with CAD[18]. While we matched the cohorts on DM, we asked whether the diabetic status explained significantly altered metabolic components by comparing the serum metabolome of normoglycemic to diabetic ACS patients. We found that 48 out of the 533 molecules (10% FDR adjusted) significantly differed between the two subgroups. Still, over 90% of the significantly perturbed metabolites were not explained by the glycemic status, suggesting that there are other mechanisms underlying this variation.

Next, we asked which genetic and environmental factors explain the levels of the altered circulating metabolites. To that end, we integrated results from analyses that we recently performed, in which we estimated the explained variance (EV) of individual serum metabolites, based on a comprehensive set of factors, including host genetics, microbiome and dietary data[19]. Here, we further estimated the EV of each metabolite based on traditional risk factors of CAD, which include age, sex, anthropometrics, blood pressure, smoking and diabetic status (Methods). We found that diet and microbiome could better explain ACS-depleted metabolites, while traditional risk factors better explained ACS-enriched metabolites (two-sided Mann–Whitney *U*-test, $P=3\times10^{-5}$ for diet; $P=8\times10^{-4}$ for microbiome, $P=0.01$ for traditional risk factors; Extended Data Fig. 2a).

Comparing the relative abundances of 766 bacterial species between ACS patients and non-ACS controls, we uncovered a distinct signature of the gut microbiome in ACS. We found that ACS patients had higher abundances of Proteobacteria compared to controls matched for age, sex, BMI, smoking status and DM (Fig. 1b; Kruskal–Wallis *H*-test, $P=0.002$) in line with previous reports showing that Proteobacteria flourish in an inflammation state and are a marker of dysbiosis[20].

We identified 20 bacterial genomes significantly enriched in either the ACS or the control individuals (FDR < 10%; Supplementary Table 8), none of which were confounded by differences in clinical parameters or drug usage (Supplementary Table 9). These results replicate previously reported findings, such as a relative depletion in butyrate-producing bacteria[13,14] (*Clostridium*, *Anaerostipes hadrus*, *Streptococcus thermophilus* and *Blautia*), and enrichment in *Odoribacter splanchnicus* and *Escherichia Coli*[21] in patients with ACS.

An ACS-related meta-organismal pathway. Among the 20 significantly enriched genomes, we identified a previously unknown bacterial species of the *Clostridiaceae* family, indexed SGB 4712, which was depleted in the ACS cohort (two-sided Mann–Whitney *U*-test, $P=0.003$; Extended Data Fig. 3a). To validate the robustness of this depletion, we applied our computational quality control (QC) and mapping pipeline to samples from the MetaCardis study (Fromentin et al., unpublished), a geographically distinct cohort of northern European ancestry, consisting of four major groups of participants: ischemic heart disease (IHD; $n=319$; including patients with ACS, IHD and heart failure due to IHD), healthy controls (HC; $n=275$; matched to the IHD group on age and sex), metabolically matched controls (MMC; $n=218$; controls matched with IHD cases on T2DM status and BMI) and untreated metabolically compromised controls (UMCC; $n=211$; individuals with features of the metabolic syndrome and thus at increased risk of IHD but receiving no lipid-lowering or antidiabetic or antihypertensive drugs). Consistent with our findings, the relative abundance of SGB 4712 was lower in IHD compared with HC (two-sided Mann–Whitney *U*-test, $P=6\times10^{-12}$), MMC ($P=0.002$) and UMCC ($P=7\times10^{-5}$), showing a gradual decrease of the relative

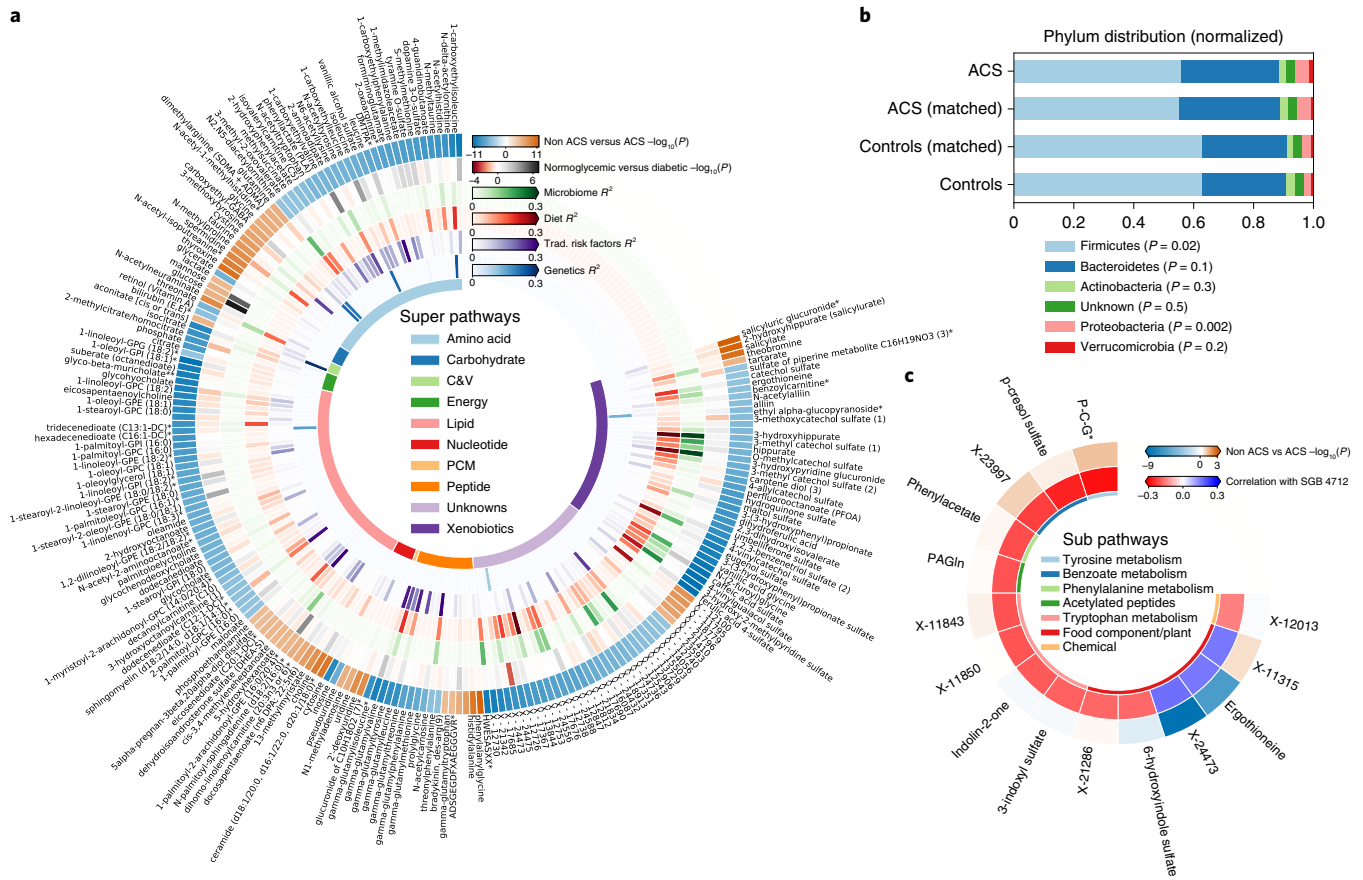


Fig. 1 | Microbiome and serum metabolomics signatures of ACS. a, A circular heatmap showing the top 200 metabolites that differ significantly between ACS and non-ACS control cohorts, matched for age, sex, BMI, smoking status and DM (Methods). Each slice represents a single metabolite, with its name indicated around the outer layer of the chart. The color code is indicated at the top of the panel. The outermost layer indicates the $-\log_{10}(P$ value) (a logistic regression model adjusted for age and sex; Methods) for the enrichment of metabolites between the two cohorts, where orange/blue colors correspond to metabolites enriched/depleted in the ACS cohort. The next layer shows the $-\log_{10}(P$ value) (a logistic regression model adjusted for age, sex and BMI) for each metabolite in diabetic versus normoglycemic ACS patients. Here, black/red colors correspond to metabolites enriched/depleted in diabetic patients. The next four layers show the EV of each metabolite by feature groups, as previously estimated[19]. The metabolites are first sorted by their categories, as indicated in the inner layer, and then by their directional enrichment between the two cohorts. **b**, The distribution of average phylum abundance (normalized to sum to 1.0) among non-ACS and ACS participants (unmatched controls, $n=335$; matched controls, $n=64$; unmatched ACS, $n=199$; matched ACS, $n=64$). *P* values refer to comparisons between the matched cohorts (Kruskal–Wallis). **c**, A circular heatmap showing 15 metabolites that significantly correlate with the relative abundance of SGB 4712 (FDR < 1%, Spearman) in the control cohort. Each slice represents a single metabolite, with its name indicated at the outermost layer of the chart. The color code of each layer is indicated at the top of the panel. The outermost layer indicates the $-\log_{10}(p$ value) (two-sided Mann–Whitney *U*-test) for the enrichment of metabolites between the two cohorts, where orange and blue colors correspond to metabolites enriched and depleted, respectively, in the ACS cohort. The next layer shows the Spearman correlation between each metabolite and the relative abundance of SGB 4712. The metabolites are sorted by their biological pathways, as indicated in the inner layer. Trad., Traditional; C&V, cofactors and vitamins; PCM, partially characterized molecules; P-C-G*, p-cresol-glucuronide*; PAGln, phenylacetylglutamine; DMTPA*, 2,3-dihydroxy-5-methylthio-4-pentenoate (DMTPA)*.

abundance of this species along populations with traditional risk factors for CAD (Extended Data Fig. 3b).

SGB 4712 was significantly associated with the levels of 15 circulating metabolites (FDR < 1%; Spearman correlation *P* value; Supplementary Table 10) in the control cohort (Fig. 1c). Markedly, in the MetaCardis study, the sign of the correlation coefficient for all 15 metabolites with SGB 4712 replicated, with 10 of these associations remaining significant (FDR < 10%; Spearman correlation *P* value; Supplementary Table 11). Several of these metabolites are previously reported to be associated with an increased risk of CAD. These include p-cresol glucuronide and p-cresol sulfate, two major metabolites of p-cresol, a product of colonic bacterial fermentation from phenylalanine and tyrosine, and were associated with CVD in hemodialysis patients[22]; indoxyl sulfate, a protein-bound uremic toxin which was suggested as a CVD risk factor in chronic kidney disease (CKD)[9]; and phenylacetylglutamine, a microbial metabolite associated with overall mortality and CVD in patients with CKD[23], recently shown to contribute to CVD via driving platelet responsiveness and thrombosis through adrenergic receptors[24]. While SGB 4712 was negatively correlated with the above metabolites, it was positively correlated with ergothioneine, a naturally occurring amino acid shown to have antioxidant and cytoprotective capabilities against cellular stressors in vitro[25], and was recently shown to be an independent marker of lower risk of CVD and mortality in humans[26]. Finally, SGB 4712 was associated with seven compounds of unknown chemical structure. These include X-11315 and X-24473, which we predict to be diet derived, and are positively correlated with SGB 4712, and two compounds that we predict to

be tryptophan metabolites, X-11843 and X-11850, which are negatively correlated with SGB 4712.

These results highlight the bacterial genome SGB 4712, as having a potential protective role in CAD development, mediated by an array of circulating blood metabolites, several of which were previously shown to play a central role in the meta-organismal pathway, while others are unknown. Thus, upon further validation in experimental studies, these metabolites may form new targets for attenuation of CAD risk.

The metabolic signature of ACS is person-specific. CAD involves a heterogeneous set of risk factors, and while individuals with CAD share a common endophenotype, they typically exhibit a biologically distinct disease profile[27]. To gain a better understanding of the individual-level variability of ACS, we sought to examine the metabolic deviations from a non-ACS control and ask whether they are person specific. Thus, for each ACS patient, we matched a set of controls for age, sex and BMI (same sex, ∓ 5 years, ∓ 3 BMI points), resulting in a median of 11 controls per ACS individual, and a total of 135 ACS participants for which at least three controls were assigned. We computed their individual deviations, and weighted the top 100 deviating metabolites per individual by their EV as previously estimated based on diet, microbiome, traditional risk factors and genetics (Methods). Finally, for every individual, we averaged these values per determinant factor, to obtain a vector of eight scores, four for either of ACS-enriched and ACS-depleted metabolites (Supplementary Table 12). These scores thus represent a weighted average of the association of the above four factors in the deviation of ACS-enriched and ACS-depleted metabolites.

We found a wide distribution of these scores, indicating that the metabolic deviation of ACS patients from their matched controls is person specific. These deviations exhibit remarkable differences between ACS-enriched and ACS-depleted metabolite sets with respect to their potential determinants (Fig. 2a,b). Notably, both diet and microbiome were far more dominant in associating with deviations of ACS-depleted metabolites compared with ACS-enriched metabolites (two-sided Mann–Whitney *U*-test, $P < 10^{-20}$ for both), suggesting that the microbiome may play a protective role in CAD.

While some patients may present with similar clinical profiles, their underlying physiological states and disease trajectories may differ. To emphasize this intra-CAD patient variability, we selected a homogeneous subgroup of ACS patients with respect to conventional risk factors. These include 17 male patients between ages 60 and 70 years, with low-density lipoprotein (LDL) in range 0.70–1.30 mg ml^{-1} and glycated hemoglobin (HbA1C) below 6%. Despite sharing a similar clinical profile, this subgroup of ACS patients demonstrated heterogeneity in their metabolic deviations (Fig. 2c). As an example, two male patients with nearly identical clinical profiles (ages 63–65 years, LDL 0.77–0.82 mg ml^{-1}, HbA1C 5.5–5.2%), showed a notable difference in the above-computed scores (Fig. 2d). The first patient (blue) had greater metabolic deviations explained by traditional risk factors, compared with the second patient (orange). These patients also differed in their scores for metabolites which were explained by the microbiome, as the first patient had higher scores for ACS-enriched metabolites, while the second patient showed higher scores for ACS-depleted metabolites. This disparity suggests that even though CAD patients may share similar clinical profiles, the metabolic mechanisms underlying their atherosclerotic burden are different.

Next, we asked whether these individual-level scores are associated with known CAD risk factors. Notably, as age is a well-known independent traditional CAD risk factor[28], we found two associations with age, which is considered to be a non-modifiable risk factor (Extended Data Fig. 4a,b). To understand whether these scores may embody a clinically important signature of CAD, we asked whether they are associated with disease state, clinical diagnoses and outcomes. We found that individuals who had a CVD-related outcome within 12 months of recruitment (Methods) showed larger deviations in metabolites explained by traditional risk factors (Fig. 2e; two-sided Mann–Whitney *U*-test, $P = 0.005$). This is despite age, sex, T2DM and smoking status not being predictive of these outcomes ($P > 0.05$ for all). Surprisingly, these individuals had significantly lower scores for genetically related ACS-enriched metabolic deviations (Extended Data Fig. 4c; two-sided Mann–Whitney *U*-test, $P = 0.002$). In addition, we found that ACS patients with T2DM had larger deviations in levels of ACS-enriched metabolites that are microbiome associated, compared with normoglycemic individuals (Fig. 2f; two-sided Mann–Whitney *U*-test, $P = 0.003$).

The microbiome plays a role in the early stages of CAD. Atherosclerosis is a progressive disease that develops over years, where each stage in the formation of an atherosclerotic plaque is characterized by a different pathological process[29]. In the early stages, the growth of an atherosclerotic plaque on the vessel wall is typically associated with impairment of the metabolic state[30].

In an attempt to interpret the involvement of each metabolic component along the chronological timeline of CAD development, we applied our analysis of individual metabolic deviations to metabolically impaired controls (defined as either diagnosis of T2DM, hypertension or dyslipidemia, or BMI > 35), and to a random subset of non-ACS individuals (Fig. 3 and Supplementary Tables 13 and 14; Methods). We view the latter as representing the background variability of such scores in non-ACS participants. By construction, we expect that the ACS patients will have the highest scores on average, as the labeling of a metabolite as depleted/enriched is based on its trend from the comparison of ACS versus non-ACS control. When comparing the scores of these three groups, we found consistent differences in score distributions. The metabolic aberrations linked with microbiome and diet show a gradual trend, with significant metabolite deviations in control participants with metabolic impairment, compared with a random subset of control individuals (diet, Fig. 3a, two-sided Mann–Whitney *U*-test, $P = 0.04$; microbiome, Fig. 3b, $P = 0.007$). In metabolites related to traditional risk factors, we observed no difference when comparing the metabolically impaired and the random control individuals (traditional risk factors, Fig. 3c, $P = 0.7$), while the ACS patients exhibited significantly higher scores.

This suggests that the contribution of the microbiome and diet to ACS might be mediated through impaired metabolic status, as opposed to aberrations in metabolites related to traditional risk factors and genetics that are not yet manifested in metabolically impaired individuals.

Serum metabolomics predict higher BMI for ACS patients. Obesity is a major independent risk factor for CAD, influencing both known risk factors such as dyslipidemia, hypertension, glucose intolerance and inflammatory state, and possibly yet unrecognized mechanisms[31]. BMI measurement is used as a marker for obesity and an indicator of metabolic health[32]. As BMI was shown to be associated with a profound perturbation of the serum metabolome[33], we asked whether and how the BMI–metabolome balance is disrupted in patients with CAD. To that end, we trained a gradient-boosting decision trees[34] (GBDT) algorithm to predict BMI, based only on the serum metabolome profiles of participants in the control cohort, and applied it over a held-out test set, containing both 156 controls and 156 subjects with ACS (Methods). Our metabolome-based model of BMI had superior performance when applied to the control test set (measured-predicted Pearson $R = 0.67$, $P < 10^{-20}$) compared with its performance when applied on the ACS test set ($R = 0.29$, $P = 2 \times 10^{-4}$), suggesting that the metabolome–BMI pattern found in non-ACS subjects is perturbed in patients with ACS (Fig. 4a).

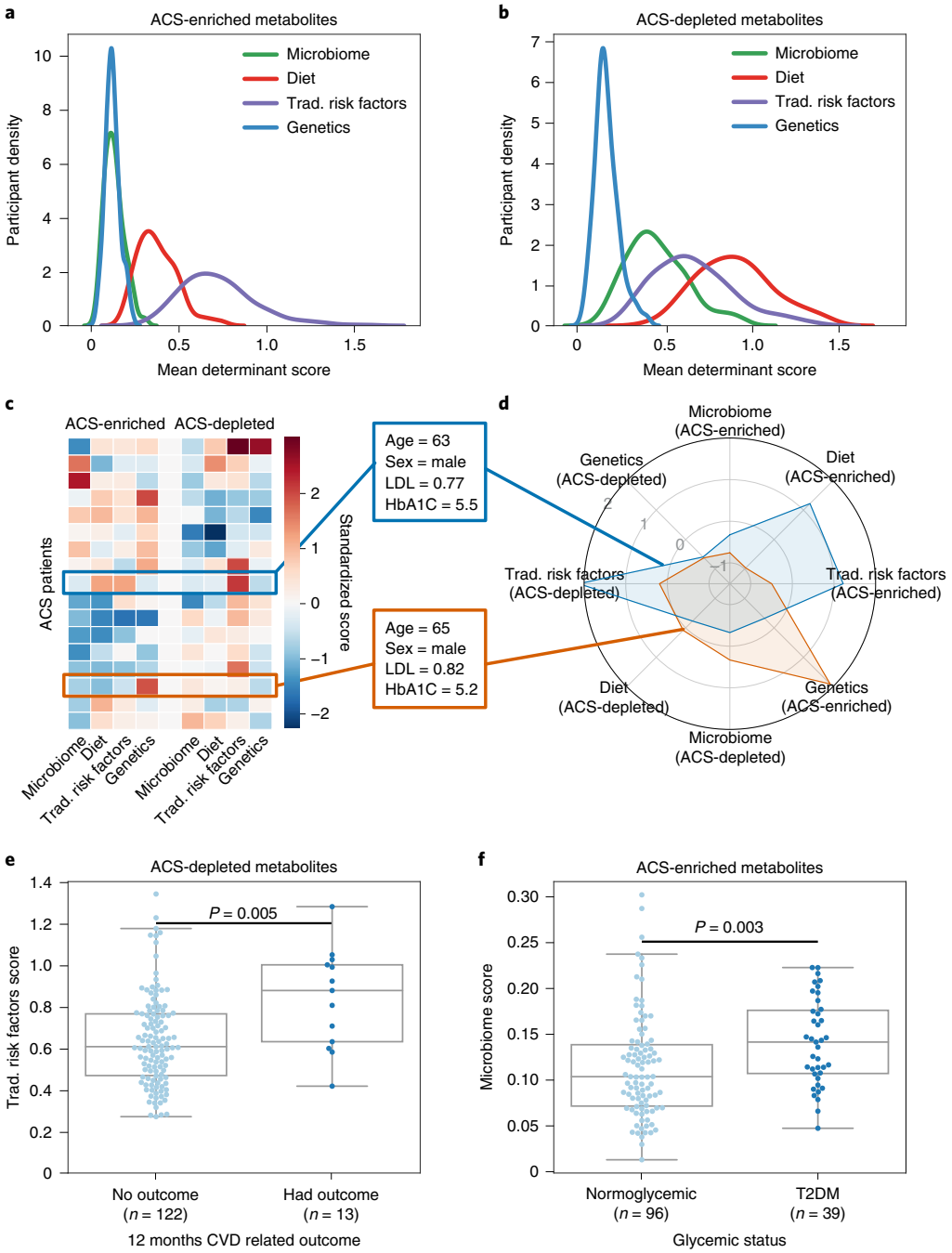


Fig. 2 | Metabolic deviations explained by potential determinants and correlate with clinical parameters. a,**b**, Density plots showing the distribution of ACS participants (*y* axis) versus the mean weighted R^2 of potential determinants (microbiome, diet, traditional risk factors or genetics) for metabolites (*x* axis); metabolites enriched in ACS (**a**); metabolites depleted in ACS (**b**). **c**, Heatmap showing the standardized mean weighted R^2 of potential determinants for ACS-enriched and ACS-depleted metabolites (*x* axis) for a subgroup of 17 ACS patients (*y* axis). **d**, Radar chart showing the standardized mean weighted R^2 of potential determinants for the two ACS subjects who are marked with blue and orange outlines in **c**. **e**,**f**, Box plots (*y* axis: center, median; box, interquartile range (IQR); whiskers, 1.5×IQR) showing the mean weighted R^2 of potential determinants for metabolites in two groups of ACS patients. **e**, The mean weighted R^2 of traditional risk factors for ACS-depleted metabolites compared between ACS patients that had a combined CVD outcome within 12 months ($n=13$) versus not ($n=122$) ($P=0.005$; two-sided Mann–Whitney *U*-test). **f**, the mean weighted R^2 of microbiome for ACS-enriched metabolites compared between ACS patients with T2DM ($n=39$) versus normoglycemic ACS patients ($n=96$) ($P=0.003$; two-sided Mann–Whitney *U*-test).

To investigate these perturbations, we examined the differences between predicted and measured BMI in both the control and ACS test sets, here termed as ΔBMI. We found that our model predicted higher ΔBMI for ACS compared with non-ACS subjects (Fig. 4b).

We next asked whether the metabolic patterns constituting the increased ΔBMI in ACS subjects might translate into clinical manifestation. We found that ACS patients with T2DM had significantly higher ΔBMI compared with normoglycemic ACS patients across all BMI ranges, as a 1.0 s.d. greater ΔBMI was associated

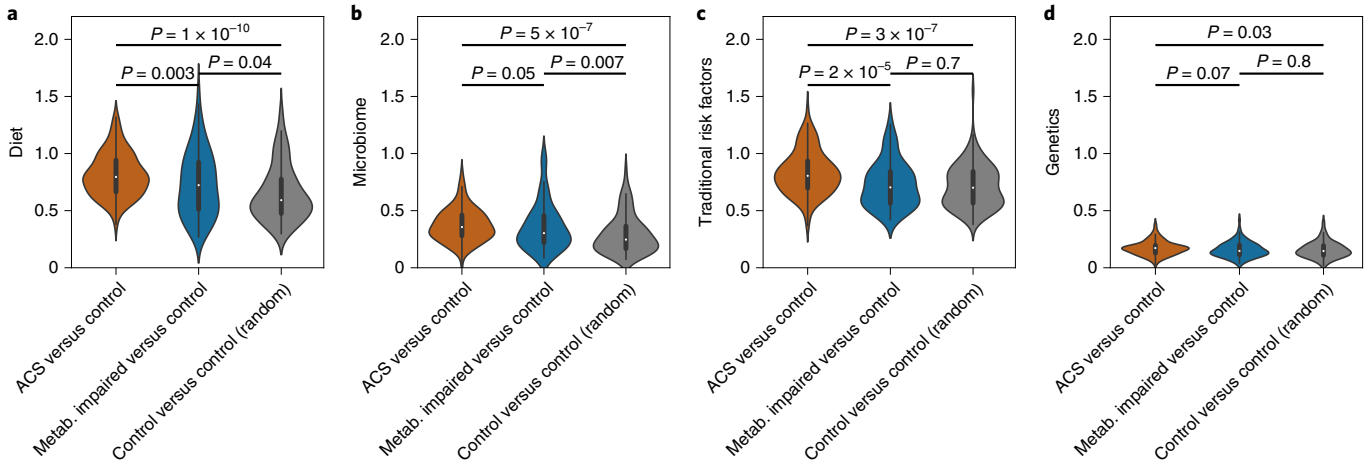


Fig. 3 | Microbiome and diet-related metabolic deviations are present in control participants with metabolic impairment. a–d, Metabolic deviation scores attributed to diet (**a**), microbiome (**b**), traditional risk factors (**c**) and genetics (**d**) computed for three subgroups: (1) ACS individuals ($n=135$) versus non-ACS controls matched for age, sex and BMI (orange); (2) non-ACS controls with metabolic impairment (defined as either: diagnosed with T2DM, hypertension or dyslipidemia, or BMI > 35; $n=102$) versus other non-ACS controls matched for age, sex and BMI (blue); and (3) a random set of non-ACS individuals ($n=132$) versus other non-ACS controls matched for age, sex and BMI (gray). Violin plot-line elements: center line, median; thick limits, upper and lower quartiles; whiskers, 1.5×IQR. *P* values shown are computed using the two-sided Mann–Whitney *U*-test.

with a 1.48-fold higher risk of T2DM (95% CI = 1.09–2.01; $P=0.01$; a logistic regression model adjusted for BMI and age; Fig. 4c). Furthermore, we found that the number of diseased coronary arteries, an index of the arteriographic extent of disease[35], was associated with higher ΔBMI, as a 1.0 s.d. greater ΔBMI was associated with a 1.5-fold higher odds of having three vessels versus one vessel involved (95% CI = 1.03–2.19; $P=0.03$; a logistic regression model adjusted for BMI, age and T2DM; Fig. 4d).

To validate the robustness of these results, we sought to replicate these findings based on other types of metabolomics data and in an independent cohort. To that end, we applied the same prediction procedure to the NMR-based metabolomics data and observed even larger differences in ΔBMI between ACS and controls (Extended Data Fig. 5; Methods). We further applied the same prediction scheme within our companion MetaCardis cohort (Fromentin et al., unpublished) and observed higher BMI predictions for individuals with IHD compared with healthy and metabolically matched controls (Extended Data Fig. 6a,b; Methods). In addition, diabetic individuals with IHD had significantly higher ΔBMI compared with normoglycemic IHD patients across all BMI ranges (Extended Data Fig. 6c).

Finally, we sought to infer which specific metabolites were the main drivers of the high ΔBMI in ACS patients. Using feature attribution analysis (SHapley Additive exPlanations, SHAP[36]), we computed the individual-level contribution of every metabolite (termed SHAP values) to the prediction of BMI in the held-out ACS test set (Supplementary Table 15; Methods). Next, we regressed ΔBMI against each metabolite's SHAP value, adjusting for BMI and T2DM, and found 59 metabolites with significant associations (FDR 1%; Supplementary Table 16). Of these, two lipids, 1-(1-enyl-palmitoyl)-2-oleoyl-GPC (P-16:0/18:1) and 1-(1-enyl-palmitoyl)-2-linoleoyl-GPC (P-16:0/18:2) were negatively correlated with BMI in the control cohort (Spearman $\rho=-0.41$ and -0.37, respectively). The latter was also significantly depleted in patients with more extensive disease (three-versus-one-vessel disease; two-sided Mann–Whitney *U*-test, $P=0.02$). 5-(galactosylhydroxy)-L-lysine was positively correlated with BMI in the control cohort ($\rho=0.33$) and was higher in patients with a three-vessel disease ($P=0.018$).

Discussion

In this study, we obtained a comprehensive multi-omic profiling for 199 patients with ACS to study the multifactorial nature of CAD. By comparing the serum profiles of ACS patients with non-ACS matched controls, we found a unique metabolomics signature of ACS, with hundreds of metabolites significantly perturbed. To date, the majority of studies focused on finding new metabolites that are increased in CAD patients, hence suspected as cardiotoxic[8,24], while our findings emphasize that the metabolomic signature of ACS is characterized by the lack of multiple serum metabolites, many of which are associated with diet and microbiome. These results are in line with the dominance of the microbiome in associating with deviations of ACS-depleted metabolites compared with ACS-enriched metabolites, which further highlights the protective role of the gut microbiome in CAD. A notable example is the previously unknown bacterial species SGB 4712, which we found to be significantly depleted both in ACS patients and in an independent validation cohort. By further linking this bacterium with the levels of both cardiotoxic and cardioprotective metabolites, we demonstrated how the absence of a specific bacterial genome may correspond to an increased risk for CAD, and suggest a concrete target to be evaluated in follow-up intervention studies. Overall, these findings thus direct a new approach in the prediction, and even treatment, of CAD patients.

Most studies to date analyzed CAD patients in bulk, searching for population-level risk factors, instead of focusing on the biological variability at an individual level. In this study, we analyzed the individual-level metabolic deviations of ACS patients, and found that these deviations are person specific, with respect to the genetic and environmental factors underlying their levels. We showed that even a homogeneous subgroup of ACS patients (with respect to conventional risk factors) had heterogeneous metabolic deviation profiles. Thus, our study emphasizes that personalized risk stratification and preventative measures are essential in CAD.

The development of CAD consists of the gradual growth of atherosclerotic plaque on the vessel wall, which is typically associated with an impairment of the metabolic state[30]. To interpret the involvement of each metabolic component along the chronological

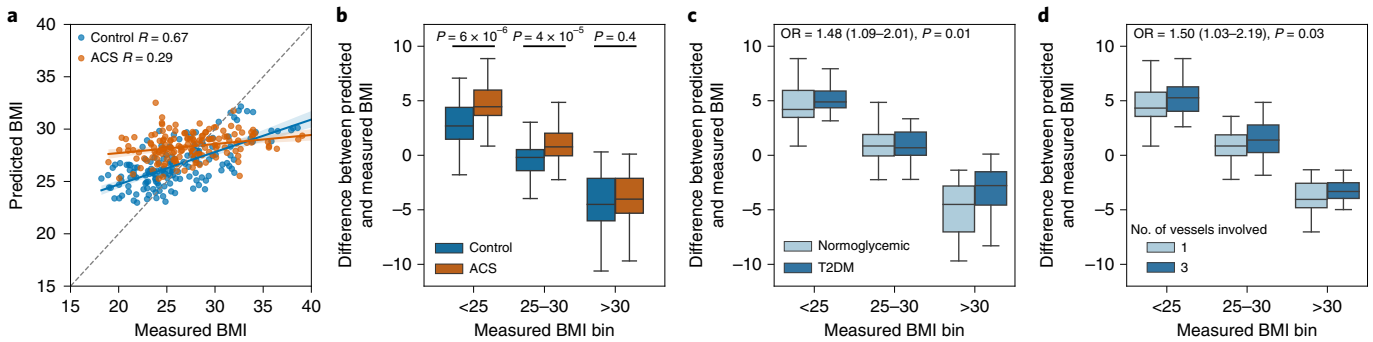


Fig. 4 | A metabolomics-based model of BMI predicts higher BMI in ACS patients and correlates with disease severity. a, Measured (*x* axis) versus predicted (*y* axis) BMI for controls ($n=156$; blue) and ACS ($n=156$; orange) individuals. Line and shaded coloring represent the fitting of a linear model and the 95% confidence interval. **b**, Difference between predicted and measured BMI (*y* axis) of individuals, binned into three BMI groups (<25, 25–30, >30; *x* axis). *P* values shown are computed using the two-sided Man–Whitney *U*-test. **c**, Same as in **b**, only for ACS participants, and each bin is separated into normoglycemic ($n=111$) versus T2DM patients ($n=44$). Higher predicted BMI was associated with an increased incidence of T2DM (odds ratio (OR) = 1.48; 95% confidence interval (CI) = 1.09–2.01; $P=0.01$; a logistic regression model adjusted for BMI and age; Methods). **d**, Same as in **b**, only for ACS participants; each bin is separated into patients with one vessel ($n=71$) versus three vessels involved ($n=36$). Higher predicted BMI was associated with an increased incidence of having three vessels involved (OR = 1.5; 95% CI = 1.03–2.19; $P=0.03$; a logistic regression model adjusted for BMI, age and T2DM; Methods). Box-plot elements: center, median; box, IQR; whiskers, 1.5×IQR.

timeline of CAD development, we applied our analysis of individual metabolic deviations to control participants that are metabolically impaired. We demonstrated that metabolic aberrations linked with microbiome and diet show a gradual trend in control participants with metabolic impairment compared with a random subset of control individuals. In metabolites related to traditional risk factors and genetics, we observed no such difference, while the ACS patients exhibit significantly higher scores. We further showed that alterations in microbial metabolites are more prevalent within ACS patients with T2DM, while traditional risk factors are more common in patients with CVD outcomes. Taken together, these results, are in line with the theory, stating that the microbiome and diet might play a role in an earlier stage of the natural history of CAD (that is, the metabolic syndrome), as opposed to traditional risk factors and genetics which may take place in later shifting into ACS and adverse cardiovascular outcomes.

To investigate obesity as an independent risk factor of CAD, we have devised and thoroughly validated a serum metabolomics-based model of BMI, and showed that higher predicted ΔBMI corresponds to a more extensive atherosclerotic disease, implying that this model might be viewed as a model of metabolic health. Recent studies showed that the lipid 1-linoleoyl-GPC (18:2) is inversely associated with obesity and T2DM[37], and increased levels of the lipid significantly reduced the risk of T2DM[38]. We found 1-linoleoyl-GPC (18:2) and 1-(1-enyl-palmitoyl)-2-linoleoyl-GPC (P-16:0/18:2) to be negatively correlated with BMI in the control cohort and significantly depleted in patients with more extensive CAD, suggesting that these metabolites may serve as potential targets to attenuate the risk of CAD. Furthermore, both metabolites contain one chain of linoleic acid, an essential fatty acid, shown to be inversely related to T2DM risk[39,40]. To infer causality, however, these hypotheses should be further examined in an interventional study.

Our study has several limitations. A valid concern is whether the biomarkers that we identified precede CVD incidence or result from it. Previous studies had shown that a core of dominant bacterial species is stable in the long term[41,42], while the serum profile is far more dynamic[43]. We, therefore, collected all biological samples near the index event, to control for biological noise, and took additional necessary measures to avoid additional biases (Methods). Additionally, the cohorts were not balanced according to T2DM and body fat distribution. However, CAD risk factors such as age, impaired glycemic status, hypertension, hyperlipidemia and smoking were matched in the analyses and our companion manuscript (Fromentin et al., unpublished) thoroughly disentangle the links of some of these factors by systematically comparing matched subgroups of individuals, as well as controlling for covariates using a modeling framework. Nevertheless, some factors may still constitute confounding factors, influencing the levels of the multi-omic data independently of CAD. Another limitation is that the samples were not taken under strict fasting conditions. It may introduce biases to the metabolic signatures that are caused by different postprandial states. The control cohort was recruited as part of a previous study, leading to a potential batch effect. We addressed these caveats by performing a rigorous batch correction procedure, and an additional correction for differences in storage times of samples (Methods). Still, it was shown that no batch correction can fully remove the effect of confounding between batch and outcome[44], and that there is no optimal technique to correct for differences in storage times. Therefore, we further mitigated these concerns by performing an external validation of our results in the MetaCardis datasets. Finally, since this study is based on observational data, the associations found cannot be considered causal.

Using a comprehensive metabolomic and microbiome profiling, we present a deep mapping of the intra-CAD variability. Taken together, our results unravel new paradigms and therapeutic directions, which may form the basis for future mechanistic experiments, preclinical and human interventional studies.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgments, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41591-022-01686-6.

Received: 1 March 2021; Accepted: 6 January 2022;
Published online: 17 February 2022

References

1. Roth, G. A. et al. Global and regional patterns in cardiovascular mortality from 1990 to 2013. *Circulation* **132**, 1667–1678 (2015).
2. Wilkins, J. T. et al. Lifetime risk and years lived free of total cardiovascular disease. *JAMA* **308**, 1795–1801 (2012).

3. Poulter, N. Coronary heart disease is a multifactorial disease. *Am. J. Hypertens.* **12**, 92S–95S (1999).
4. Gaziano, J. M. et al. Use of aspirin to reduce risk of initial vascular events in patients at moderate risk of cardiovascular disease (ARRIVE): a randomised, double-blind, placebo-controlled trial. *Lancet* **392**, 1036–1046 (2018).
5. Psychogios, N. et al. The human serum metabolome. *PLoS ONE* **6**, e16957 (2011).
6. Tang, W. H. W. et al. Intestinal microbial metabolism of phosphatidylcholine and cardiovascular risk. *N. Engl. J. Med.* **368**, 1575–1584 (2013).
7. Brown, J. M. & Hazen, S. L. Metaorganismal nutrient metabolism as a basis of cardiovascular disease. *Curr. Opin. Lipidol.* **25**, 48–53 (2014).
8. Zhu, W. et al. Gut microbial metabolite TMAO enhances platelet hyperreactivity and thrombosis risk. *Cell* **165**, 111–124 (2016).
9. Barreto, F. C. et al. Serum indoxyl sulfate is associated with vascular disease and mortality in chronic kidney disease patients. *Clin. J. Am. Soc. Nephrol.* **4**, 1551–1558 (2009).
10. Meijers, B. K. I. et al. p-Cresol and cardiovascular risk in mild-to-moderate kidney disease. *Clin. J. Am. Soc. Nephrol.* **5**, 1182–1189 (2010).
11. Zeevi, D. et al. Personalized nutrition by prediction of glycemic responses. *Cell* **163**, 1079–1094 (2015).
12. Rothschild, D. et al. Environment dominates over host genetics in shaping human gut microbiota. *Nature* **555**, 210–215 (2018).
13. Jie, Z. et al. The gut microbiome in atherosclerotic cardiovascular disease. *Nat. Commun.* **8**, 845 (2017).
14. Tang, W. H. W., Kitai, T. & Hazen, S. L. Gut microbiota in cardiovascular health and disease. *Circ. Res.* **120**, 1183–1196 (2017).
15. Maier, L. et al. Extensive impact of non-antibiotic drugs on human gut bacteria. *Nature* **555**, 623–628 (2018).
16. Liu, J. et al. Integration of epidemiologic, pharmacologic, genetic and gut microbiome data in a drug-metabolite atlas. *Nat. Med.* **26**, 110–117 (2020).
17. Forslund, S. K. et al. Combinatorial, additive and dose-dependent drug-microbiome associations. *Nature* https://doi.org/10.1038/s41586-021-04177-9 (2021).
18. Kannel, W. B. & McGee, D. L. Diabetes and cardiovascular disease. The Framingham study. *JAMA* **241**, 2035–2038 (1979).
19. Bar, N. et al. A reference map of potential determinants for the human serum metabolome. *Nature* https://doi.org/10.1038/s41586-020-2896-2 (2020).
20. Winter, S. E. & Bäumler, A. J. Dysbiosis in the inflamed intestine: chance favors the prepared microbe. *Gut Microbes* **5**, 71–73 (2014).
21. Carnevale, R. et al. Low-grade endotoxaemia enhances artery thrombus growth via Toll-like receptor 4: implication for myocardial infarction. *Eur. Heart J.* **41**, 3156–3165 (2020).
22. Meijers, B. K. I. et al. Free p-cresol is associated with cardiovascular disease in hemodialysis patients. *Kidney Int.* **73**, 1174–1180 (2008).
23. Poesen, R. et al. Microbiota-derived phenylacetylglutamine associates with overall mortality and cardiovascular disease in patients with CKD. *J. Am. Soc. Nephrol.* **27**, 3479–3487 (2016).
24. Nemet, I. et al. A cardiovascular disease-linked gut microbial metabolite acts via adrenergic receptors. *Cell* **180**, 862–877.e22 (2020).
25. Cheah, I. K. & Halliwell, B. Ergothioneine; antioxidant potential, physiological function and role in disease. *Biochim. Biophys. Acta.* **1822**, 784–793 (2012).
26. Smith, E. et al. Ergothioneine is associated with reduced mortality and decreased risk of cardiovascular disease. *Heart* **106**, 691–697 (2020).
27. Leopold, J. A. & Loscalzo, J. Emerging role of precision medicine in cardiovascular disease. *Circ. Res.* **122**, 1302–1315 (2018).
28. Dhingra, R. & Vasan, R. S. Age as a risk factor. *Med. Clin. North Am.* **96**, 87–91 (2012).
29. Lusis, A. J. Atherosclerosis. *Nature* **407**, 233–241 (2000).
30. Weber, C. & Noels, H. Atherosclerosis: current pathogenesis and therapeutic options. *Nat. Med.* **17**, 1410–1422 (2011).
31. Poirier, P. et al. Obesity and cardiovascular disease: pathophysiology, evaluation, and effect of weight loss: an update of the 1997 American Heart Association Scientific Statement on Obesity and Heart Disease from the Obesity Committee of the Council on Nutrition, Physical Activity, and Metabolism. *Circulation* **113**, 898–918 (2006).
32. Goossens, G. H. The metabolic phenotype in obesity: fat mass, body fat distribution, and adipose tissue function. *Obes. Facts* **10**, 207–215 (2017).
33. Cirulli, E. T. et al. Profound perturbation of the metabolome in obesity is associated with health risk. *Cell Metab.* **29**, 488–500.e2 (2019).
34. Ke, G. et al. LightGBM: a highly efficient gradient boosting decision tree. *NeurIPS Proceedings* https://papers.nips.cc/paper/6907-lightgbm-a-highly-efficient-gradient-boosting-decision-tree.pdf (2017).
35. Ringqvist, I. et al. Prognostic value of angiographic indices of coronary artery disease from the Coronary Artery Surgery Study (CASS). *J. Clin. Invest.* **71**, 1854–1866 (1983).
36. Lundberg, S. M., Erion, G. G. & Lee, S.-I. Consistent individualized feature attribution for tree ensembles. Preprint at arXiv (2018).
37. Pickens, C. A., Vazquez, A. I., Jones, A. D. & Fenton, J. I. Obesity, adipokines, and C-peptide are associated with distinct plasma phospholipid profiles in adult males, an untargeted lipidomic approach. *Sci. Rep.* **7**, 6335 (2017).
38. Vangipurapu, J., Fernandes Silva, L., Kuulasmaa, T., Smith, U. & Laakso, M. Microbiota-related metabolites and the risk of type 2 diabetes. *Diabetes Care* **43**, 1319–1325 (2020).
39. Zong, G. et al. Associations between linoleic acid intake and incident type 2 diabetes among U.S. men and women. *Diabetes Care* **42**, 1406–1413 (2019).
40. Pertiwi, K. et al. Plasma and dietary linoleic acid and 3-year risk of type 2 diabetes after myocardial infarction: a prospective analysis in the alpha omega cohort. *Diabetes Care* **43**, 358–365 (2020).
41. Martínez, I., Muller, C. E. & Walter, J. Long-term temporal analysis of the human fecal microbiota revealed a stable core of dominant bacterial species. *PLoS ONE* **8**, e69621 (2013).
42. Lozupone, C. A., Stombaugh, J. I., Gordon, J. I., Jansson, J. K. & Knight, R. Diversity, stability and resilience of the human gut microbiota. *Nature* **489**, 220–230 (2012).
43. Yousri, N. A. et al. Long term conservation of human metabolic phenotypes and link to heritability. *Metabolomics* **10**, 1005–1017 (2014).
44. Soneson, C., Gerster, S. & Delorenzi, M. Batch effect confounding leads to strong bias in performance estimates obtained by cross-validation. *PLoS ONE* **9**, e100335 (2014).

Methods

Description of cohorts. The ACS cohort was recruited at Rabin Medical Center (Beilinson and Hasharon hospitals) and included 199 participants. The inclusion criteria were ACS patients aged 30–80. 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 a history of organ transplantation or receiving immunosuppressive therapy, or patients with drug or alcohol addiction. The participants answered detailed medical, lifestyle and nutritional questionnaires, and provided stool and serum samples. Both blood and stool samples were not taken under strict fasting conditions. The study was approved by the Ethics Committee of Rabin Medical Center, approval number RMC-622-16. All participants signed written informed consent forms.

The control cohort included 970 previously collected samples of Israeli individuals who were enrolled in previous studies that were approved by Tel Aviv Sourasky Medical Center Institutional Review Board (IRB), approval numbers TLV-0658-12, TLV-0050-13 and TLV-0522-10; Kfar Shaul Hospital IRB, approval number 0-73. Full study designs were described elsewhere[11,12]. In brief, participants were healthy individuals aged 18–70. The participants answered detailed medical, lifestyle and nutritional questionnaires, provided stool and serum samples. Both blood and stool samples were not taken under strict fasting conditions.

Cohort selection. The cohort selection and data acquisition pipeline are shown in Extended Data Fig. 1. All 970 control participants were profiled for microbiome composition as previously described[11]. Matching for stool collection (swab) and DNA library preparation kits (Nextera DNA Flex Library Prep), a subset of 340 samples were considered as valid controls for direct comparison of bacterial species abundances. All 199 participants with ACS were profiled for microbiome composition.

A pilot study of 488 control individuals was previously profiled for serum metabolomics using the Metabolon platform[19]. Of these, we excluded 15 samples of participants that did not have sufficient high-quality clinical data, resulting in a control cohort of 473 individuals. In this study, we profiled a subgroup of 156 participants with ACS for serum metabolomics using the Metabolon platform. Sample selection was on the basis of recruitment date, such that the first 156 available serum samples of individuals with full clinical data were profiled. All available samples of controls (961 of 970) and participants with ACS (191 of 199) were profiled for serum metabolomics using the platform of Nightingale Health.

Clinical data in the ACS cohort. Recruited ACS patients were diagnosed with either unstable angina ($n=30$), non-ST-elevation myocardial infarction ($n=68$), or ST-elevation myocardial infarction ($n=101$). All participants underwent cardiac catheterization in accordance with the standard of care and based upon the decision of the treating cardiologist. Stool and serum samples were stored at the clinical centers at a temperature of −80 °C until delivery. Comprehensive clinical data were collected from medical records, including electrocardiography, vital signs, complete blood count, creatinine and troponin T levels, fasting glucose, HbA1C, lipid profile, and angiographic data, including the number of coronary artery vessels involved.

T2DM, hypertension and dyslipidemia diagnosis. T2DM was defined as fasting plasma glucose ≥7 mmol l^{-1} and/or HbA1C ≥6.5% and/or subjects taking any glucose-lowering agents. Hypertension was defined as systolic blood pressure >130 mmHg and/or diastolic blood pressure >90 mmHg and/or subjects taking antihypertensive drugs. Dyslipidemia was defined as LDL cholesterol >1.60 mg ml^{-1} and/or HDL cholesterol <0.35 mg ml^{-1} and/or subjects taking lipid-lowering drugs. In ACS patients, as the prescription of lipid-lowering drugs was not specific for dyslipidemia, we only considered patients with a previous diagnosis of dyslipidemia. As LDL cholesterol was not directly measured in the non-ACS cohort using standard clinical chemistry, we estimated it based on the Friedewald calculation[45]:

$$(\mathrm{LDL}-C)\left[\frac{\mathrm{mg}}{\mathrm{ml}}\right]=\text{Total cholesterol}-(\mathrm{HDL}-C)-\frac{\text{Triglycerides}}{5}$$

Metabolomics profiling and preprocessing. Metabolite concentrations were measured in serum samples using two different and complementary platforms:

(1) An untargeted LC–MS analysis was performed by Metabolon, Inc., Durham, NC, USA. Full details are available in Supplementary Note 1. A total of 900 serum samples were profiled in two separate runs. In the first run, 540 samples were profiled, 19 of which were control samples (technical replicate) pooled from several individuals, and a total of 1,251 metabolites were identified. In the second run, 360 serum samples were profiled, 10 of which were control samples, identical to those in the first run, and here a total of 1,171 metabolites were identified, 1,011 of which overlapped with the first run. The first run included 457 samples of non-ACS individuals, while the second run included 31 samples of non-ACS individuals and 156 ACS samples. We excluded 15 samples of control individuals for which we did not have sufficient high-quality clinical data, resulting in a control cohort of 473 individuals. To correct for possible batch effects resulting from the separate runs, we first applied the log (base 10) transform over the data and used the control samples within each run to compute the median and standard deviation of every metabolite that was measured in both runs. Then, within each run, for every metabolite that had at least five measurements in control samples (961 metabolites), we subtracted the control's median and divided it by its standard deviation. After merging the data from both runs, for every metabolite, we performed robust standardization (subtracting the median and dividing by the standard deviation) and clipped outlier samples that were farther than 5 s.d. To address the difference in storage times, we first regressed the normalized metabolite intensities against storage times using only samples of non-ACS individuals from the first run and excluded 25 metabolites that were significantly correlated with storage times from downstream analyses ($P<0.05/961$; Spearman correlation P value). To correct the remaining metabolite levels to storage times, for every metabolite with at least 100 non-missing values (all but eight metabolites), using all samples, we regressed its levels against storage times (in days) while adding the participant's age and an indicator variable marking the identity of the cohort as covariates. Then, for all metabolites for which the coefficient of the storage time variable had an estimated P value below 0.05 (143 such metabolites) we applied the correction. Finally, we imputed missing values as the minimum value per metabolite. This resulted in 936 metabolites in both runs which we used in further analyses (Supplementary Table 2).

(2) The proton nuclear magnetic resonance (^{1}H-NMR) platform of Nightingale Health, for which the technical details and relevant epidemiological applications were previously reviewed[46,47]. A total of 1,178 serum samples were profiled in a single run, including 191 samples of ACS patients and 961 samples of non-ACS individuals. This platform provides simultaneous quantification for a total of 228 absolute-value-based (concentrations are estimates based on reference data) plasma metabolites and ratios, mainly expanding the detailed lipidomic profiles and adding measurements of clinically validated biomarkers, including routine lipids, lipoprotein subclass profiling with lipid concentrations within 14 subclasses (lipoprotein-derived variables), fatty-acid composition and various low-molecular-weight metabolites such as amino acids, ketone bodies and glycolysis metabolites (Supplementary Table 3). In cases where the value 'TAG' was provided, we replaced it with zero and added a corresponding binary indicator variable. No further normalization or imputation was used.

Microbiome preprocessing. Sample collection, DNA extraction and sequencing of the samples in this study were previously described[11]. Briefly, we used only samples that were collected using swabs, filtered metagenomic reads containing Illumina adapters, filtered low-quality reads and trimmed low-quality read edges. We detected host DNA by mapping with GEM[48] to the human genome (hg19) with inclusive parameters and removed human reads. We subsampled all samples to have ten million reads.

Bacterial relative abundance estimation was performed by mapping bacterial reads to species-level genome bins (SGB) representative genomes[49]. SGBs were taxonomically labeled with the species label associated with the reference genome(s) present in the bin, considering the most common species label if multiple reference genomes with different assigned species were present (Supplementary Table 17). When no reference genomes were present in the species-level bins, a higher taxonomic level was assigned. We selected all SGB representatives from groups with at least five genomes, and for these representatives' genomes, kept only unique regions as a reference data set. Mapping was performed using bowtie2[50] and abundance was estimated by calculating the mean coverage of unique genomic regions across the 50% most densely covered areas as previously described[51,52].

Identification of differential metabolites. To identify which metabolites significantly differ between the ACS and the control cohort, we applied a logistic regression over the normalized intensities of every metabolite, with age and sex as covariates. This was performed using data following a 1:1 matching for age, sex, BMI, DM and smoking status, resulting in 83 samples of ACS and controls in each cohort (Supplementary Table 4). After matching, there was no significant difference in age ($P=0.36$), sex ($P=1$), BMI ($P=0.24$), smoking ($P=0.31$), DM ($P=0.84$), treatment for DM ($P=0.84$), HbA1C% ($P=0.14$), hypertension ($P=0.21$) and treatment of hypertension ($P=0.21$). Dyslipidemia ($P=0.043$) and treatment for dyslipidemia ($P=0.022$) were significantly unbalanced between the two groups. P values were computed with the Fisher exact test for binary variables, and a two-sided Mann–Whitney U-test for continuous parameters. Standardized differences of the matched parameters are given in Supplementary Table 18. Out of the 936 metabolites, we tested 892, for which we previously obtained estimates of their prediction potential (represented as EV), based on genetic and environmental factors, in a control cohort[19].

Identification of differential microbial species. To identify the microbial genomes that significantly differ between the ACS and the control cohort, we applied a logistic regression over the log-transformed (10 based) relative abundances of every bacterial species, with age and sex as covariates. This was performed using data following a 1:1 matching for age, sex, BMI, DM and smoking status, resulting in 64 samples of non-ACS and ACS in each cohort (Supplementary Table 5). After matching, there was no significant difference in age ($P = 0.25$), sex ($P = 0.44$), BMI ($P = 0.33$), smoking ($P = 1$), DM ($P = 0.8$), treatment for DM ($P = 0.076$), HbA1C% ($P = 0.34$), hypertension ($P = 0.86$), treatment of hypertension ($P = 0.52$), dyslipidemia ($P = 0.37$) and treatment for dyslipidemia ($P = 0.85$). *P* values were computed with the Fisher exact test for binary variables, and a two-sided Mann–Whitney *U*-test for continuous parameters. Standardized differences of the matched parameters are given in Supplementary Table 19. We only tested 766 bacterial species which were present in at least 5% of all samples.

Explained variance of metabolites. We estimated the EV of individual serum metabolites based on a set of environmental and genetic factors, in a healthy cohort, as previously described[19]. Briefly, we used GBDT from the LightGBM (v.2.3.1) package[34], to predict the levels of serum metabolites based on demographics, diet, microbiome, lifestyle, human genetics and clinical data. In this study, we have expanded these analyses to include a set of parameters that represent traditional risk factors of CAD. These include age, sex, waist and hip circumference, waist-to-hip ratio, BMI, sitting systolic and diastolic blood pressure, smoking status and a dummy variable binning HbA1C into normal (<5.7), prediabetes (5.7–6.4) and diabetes (>6.4). To estimate the EV of each metabolite group, we ran a fivefold cross-validation (CV) model, and evaluated the results using the coefficient of determination (R^2).

Treatment de-confounded analysis. To assess the extent to which observed differences between ACS subjects and controls with regards to the microbiome and metabolome features are confounded, in the sense of, being consequences of other (treatment or risk factor) variables different between the groups more so than characteristic of ACS itself, we additionally employed a post-hoc filtering approach introduced in the work by Forslund et al.[17]. Full details are available in Supplementary Note 2.

The individual-level metabolic signature of ACS. For each ACS patient, we matched a non-ACS control subgroup of individuals by age, sex and BMI, such that all controls will have the same sex, an age difference of not more than five years, and not more than three BMI points. This results in a median of 11 controls per ACS sample; 21 ACS patients did not match at least 3 controls and were discarded from further analysis. Next, we separated all metabolites for which explained variance estimates were available into two distinct subsets: ACS-enriched ($n = 338$) and ACS-depleted ($n = 555$) metabolites, and considered these for further analyses. For every ACS patient, we weighted every metabolite by the number of s.d. from its mean, as calculated in the designated control. We only considered deviations of metabolites when they aligned with their general trend as ACS-enriched or depleted. In reverse cases, we assigned the metabolite with a value of zero. Then, for the top 100 deviating metabolites, we multiplied their weight with a binary variable that indicates whether each metabolite had EV >5%, as we previously estimated based on diet, microbiome, traditional risk factors and genetics in the control cohort. Finally, for every individual, we averaged these values per feature group, to obtain a vector of eight scores, four for either ACS-enriched and ACS-depleted metabolites (Supplementary Table 12).

In the analysis shown in Fig. 3 we computed similar scores for: (1) ACS participants ($n = 135$) versus non-ACS controls; (2) non-ACS controls with metabolic impairment (defined as either: diagnosed with T2DM, hypertension or dyslipidemia, or BMI > 35; $n = 102$) versus non-ACS controls; (3) a random set of non-ACS individuals ($n = 132$) versus non-ACS controls. All samples were matched for age, sex and BMI. To compare these scores across the three groups, for every major source, we used the Kruskal–Wallis *H*-test (diet, $P = 5 \times 10^{-9}$; microbiome, $P = 3 \times 10^{-6}$; traditional risk factors, $P = 2 \times 10^{-7}$; genetics, $P = 0.06$). Here, we used the entire set of metabolites, mainly averaging the top 100 deviations out of the 893 metabolites included in this analysis.

Combined cardiovascular outcomes. Cardiovascular outcomes within 12 months of recruitment (exact dates unknown) include: acute myocardial infarction, acute stroke, unplanned percutaneous coronary intervention or cardiovascular-related death.

Serum metabolomics-based prediction models of BMI. We trained a GBDT algorithm (LightGBM) to predict BMI, based on the serum metabolome profiles of 298 control individuals, and applied it over a held-out test set, containing both 156 control and 156 ACS subjects. The serum metabolome profiles we used as input for the model included the set of 936 metabolites measured by the mass spectrometry platform. During training, we ran a random hyperparameter search consisting of five iterations in a threefold CV using the module RandomizedSearchCV from sklearn (v.0.24.2) and chose the best model for prediction.

We applied the same procedure to replicate the prediction of BMI based on metabolomics data from the Nightingale platform. Here, the model was trained, based on the serum metabolome profiles of 763 control individuals, and was applied and evaluated on two held-out test sets, including one of 179 control and another of 179 ACS subjects. In training and prediction based on the NMR data, we excluded the measurements of albumin, glucose and creatinine.

Feature attribution analysis. We used SHAP[53], a framework for interpreting predictions, which assigns each feature an importance value for a particular prediction. Briefly, for a specific prediction, a feature's SHAP value is defined as the change in the expected value of the model's output when this feature is observed compared with when it is missing. It is computed using a sum that represents the impact of each feature being added to the model, averaged over all possible orderings of features being introduced. Individual SHAP values were computed for held-out subjects with ACS using the module TreeExplainer (v.0.35.0)[36]. To identify the metabolites that were the main drivers of the overprediction of BMI, we regressed ΔBMI against each metabolite's SHAP value, adjusting for BMI and T2DM status. We then corrected the regression *p* values for the metabolite's SHAP value using the FDR procedure.

Replications in the independent MetaCardis study. Technical details regarding sample collection, DNA extraction, metagenomic sequencing and initial QC are available in our companion manuscript (Fromentin et al., unpublished). Metagenomics data from the MetaCardis study went through an identical QC and mapping pipeline as the 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 two-sided Mann–Whitney *U*-test over the relative abundances of SGB 4712, comparing individuals with IHD with 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 ten 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., unpublished). 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 exceeded prediction of BMI in subjects with IHD, we applied the same procedure as in the 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 and identified the metabolites, whose SHAP values were significantly associated with ΔBMI (FDR 1%; Supplementary Table 20; adjusted for BMI and T2DM status).

Additional replications are reported in our companion manuscript (Fromentin et al., unpublished). 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 ($p = 2 \bullet 10^{-8}$); (2) effective performances of classifier models (orthogonal projections to latent structures discriminant analysis) built using MetaCardis' ACS-specific biomarkers. Models were trained in the MetaCardis population and tested in our study samples (area under the curve = 0.851).

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The raw metagenomic sequencing data per sample of the 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 the controls are available from the European Genome-phenome Archive (EGA; https://ega-archive.org/): EGAS00001004512. The raw metabolomics data and full clinical phenotypes for the 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.

Code availability

Analysis source code is available at https://github.com/noambar/ACStudy.

References

45. Friedewald, W. T., Levy, R. I. & Fredrickson, D. S. Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. *Clin. Chem.* **18**, 499–502 (1972).
46. Soininen, P., Kangas, A. J., Würtz, P., Suna, T. & Ala-Korpela, M. Quantitative serum nuclear magnetic resonance metabolomics in cardiovascular epidemiology and genetics. *Circ. Cardiovasc. Genet.* **8**, 192–206 (2015).
47. Würtz, P. et al. Quantitative serum nuclear magnetic resonance metabolomics in large-scale epidemiology: a primer on -omic technologies. *Am. J. Epidemiol.* **186**, 1084–1096 (2017).
48. Marco-Sola, S., Sammeth, M., Guigó, R. & Ribeca, P. The GEM mapper: fast, accurate and versatile alignment by filtration. *Nat. Methods* **9**, 1185–1188 (2012).
49. Pasolli, E. et al. Extensive unexplored human microbiome diversity revealed by over 150,000 genomes from metagenomes spanning age, geography, and lifestyle. *Cell* **176**, 649–662.e20 (2019).
50. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
51. Korem, T. et al. Growth dynamics of gut microbiota in health and disease inferred from single metagenomic samples. *Science* **349**, 1101–1106 (2015).
52. Zeevi, D. et al. Structural variation in the gut microbiome associates with host health. *Nature* **568**, 43–48 (2019).
53. Lundberg, S. & Lee, S.-I. A unified approach to interpreting model predictions. Preprint at arXiv (2017).

Acknowledgements

We thank past and present members of the Segal group and the Cardiology Department at Rabin Medical Center for useful discussions. Y.T.-B. received a research grant from the Tel Aviv University Faculty Funds, and from the Gassner Fund for Medical Research. N.B. received a PhD scholarship for Data Science from the Israeli Council for Higher Education (CHE) via the Weizmann Data Science Research Center and is supported by a research grant from the Estate of Tully and Michele Plesser. E.S. is supported by the Crown Human Genome Center, by D. L. Schwarz, J. N. Halpern and L. Steinberg, and by grants funded by the European Research Council and the Israel Science Foundation. M.-E.D. is supported by the NIHR Imperial Biomedical Research Centre, and by grants from the French National Research Agency (ANR-10-LABX-46 [European Genomics Institute for Diabetes]), from the National Center for Precision Diabetic Medici–e – PreciDIAB, which is jointly supported by the French National Agency for Research (ANR-18-IBHU-0001), by the European Union (FEDER), by the Hauts-de-France Regional Council (Agreement 20001891/NP0025517), by the European Metropolis of Lille (MEL, Agreement 2019_ESR_11) and Isite ULNE (R-002-20-TALENT-DUMAS), also jointly funded by ANR (ANR-16-IDEX-0004-ULNE), the Hauts-de-France Regional Council (20002845) and by the European Metropolis of Lille (MEL). K.C. is supported by Medical Research Council (MRC) Skills Development Fellowship (grant number MR/S020039/1) and Wellcome Trust funded Institutional Strategic Support Fellowship (grant number 204834/Z/16/Z).

Author contributions

Y.T.-B. and N.B. conceived the project, designed and conducted all analyses, interpreted the results, wrote the manuscript and are listed in random order. N.R. performed metabolomics analyses and interpreted the results. A.G. conducted microbiome analysis. Y.B. conducted microbiome analysis and provided additional information regarding SGB 4712. A.A.S., A.S., C.C.-A., Z.A. and Y.H. coordinated and designed data collection. M.L.-P. and A.W. developed protocols, performed microbiome sequencing, and processed serum samples. A.W. designed the project and oversaw sample collection and processing. K.C., S.K.F., S.F., M.-E.D., S.D.E. and O.P. performed the replication analysis on the MetaCardis cohort. R.K. and E.S. conceived and directed the project and analyses, designed the analyses, interpreted the results and wrote the manuscript.

Competing interests

The authors declare no competing interests.

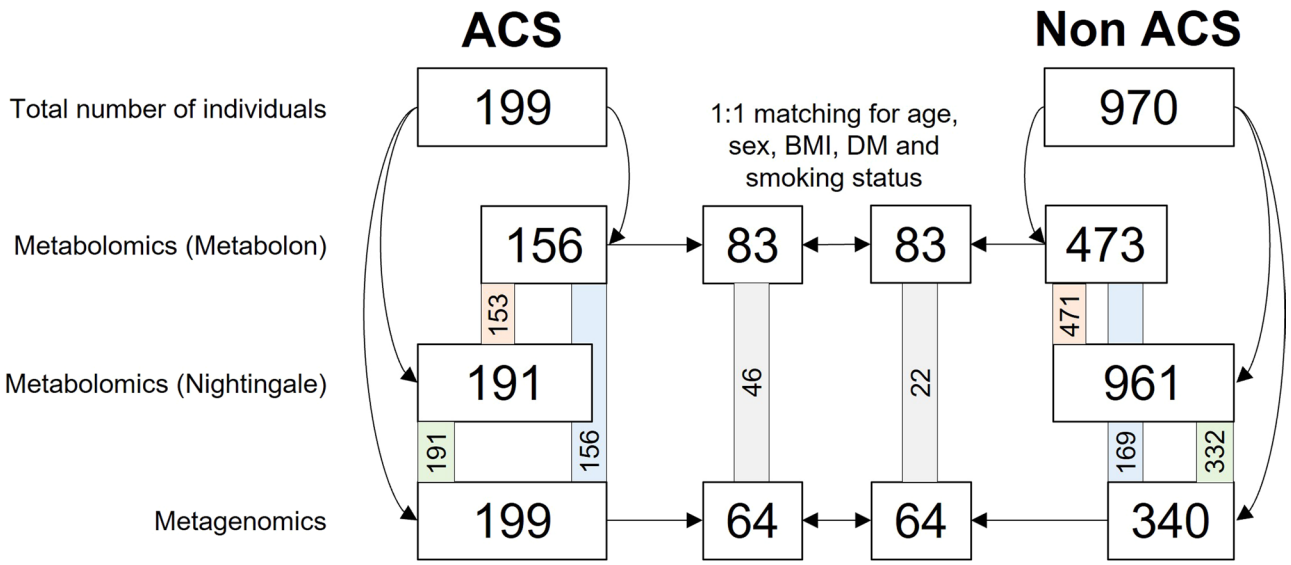
Additional information

Extended data are available for this paper at https://doi.org/10.1038/s41591-022-01686-6.

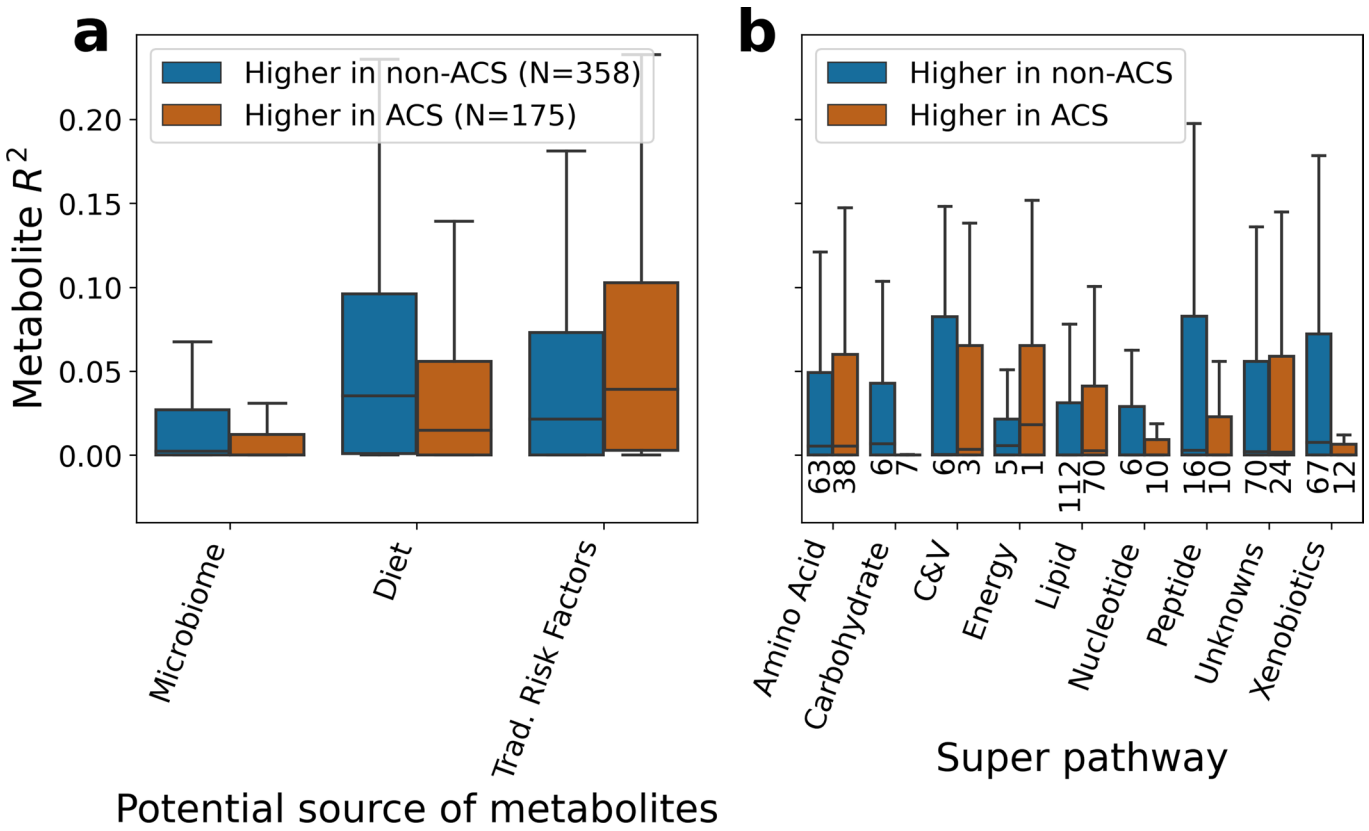
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41591-022-01686-6.

Correspondence and requests for materials should be addressed to Eran Segal.

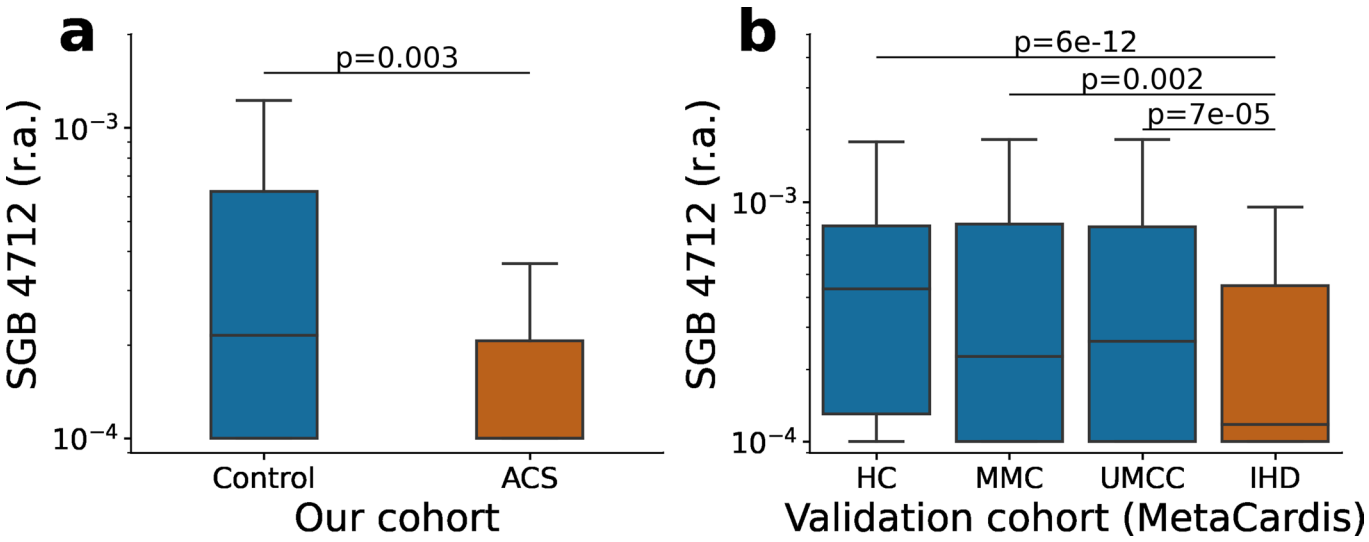
Peer review information *Nature Medicine* thanks Matej Oresic, Manuel Mayr, Ellen Blaak and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Michael Basson was the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.



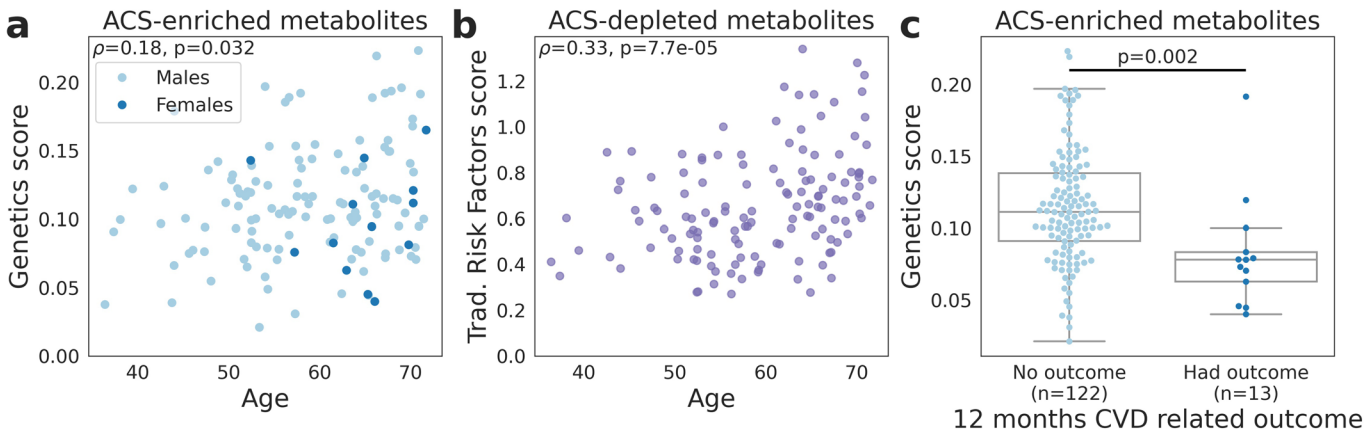
Extended Data Fig. 1 | Cohort selection and data acquisition pipeline. This study includes a total of 199 participants with ACS and 970 non-ACS individuals. Each cell shows the number of individuals who were profiled for the corresponding omic platform indicated on the left. Colored bars connecting cells represent the number of overlapping individuals. For example, there were 169 non-ACS individuals that were profiled both for serum metabolomics using the Metabolon platform and for microbiome composition. The 156 samples of individuals with ACS that were profiled using the Metabolon platform are the first to be enrolled in this study. The 473 samples of non-ACS individuals that were profiled using the Metabolon platform, were profiled as part of our previous study (Bar et al. 2020). All samples of individuals with ACS (n = 191) and of non-ACS individuals (n = 961) for which we had available serum obtained during their recruitment, were profiled using the Nightingale platform. While microbiome data were available for all individuals with and without ACS, we only considered samples for which the collection, DNA extraction and sequencing procedures were identical (n = 199 for ACS; n = 340 for non-ACS). Differential abundance analysis was performed based on subcohorts resulting from 1:1 matching for age, sex, BMI, DM, and smoking status. ACS, Acute Corony Syndrome; BMI Body Mass Index; DM, Diabetes Mellitius.



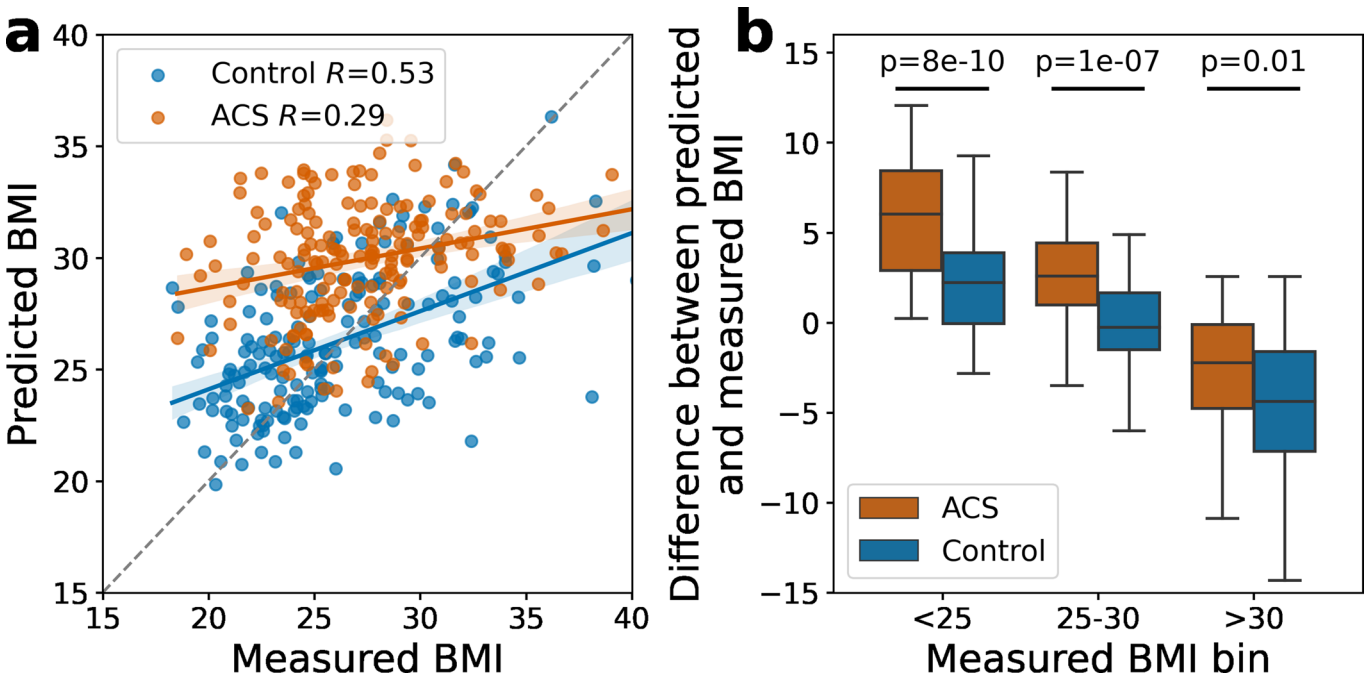
Extended Data Fig. 2 | Breakdown of ACS serum metabolomics pattern by the origin of metabolites and biological pathway. (a) Box plots (y axis: center, median; box, IQR; whiskers, 1.5×IQR) showing the explained variance of metabolites by different feature groups (x-axis) separated to metabolites enriched in ACS (N = 175; orange) and enriched in matched non-ACS controls (N = 358; blue). **(b)** explained variance of metabolites (y axis: center, median; box, IQR; whiskers, 1.5×IQR) by their super pathways (x axis) separated to metabolites enriched in ACS (orange) and enriched in matched non-ACS controls (blue). The number of metabolites per group is shown below each box. Trad., Traditional; C&V, cofactors and vitamins.



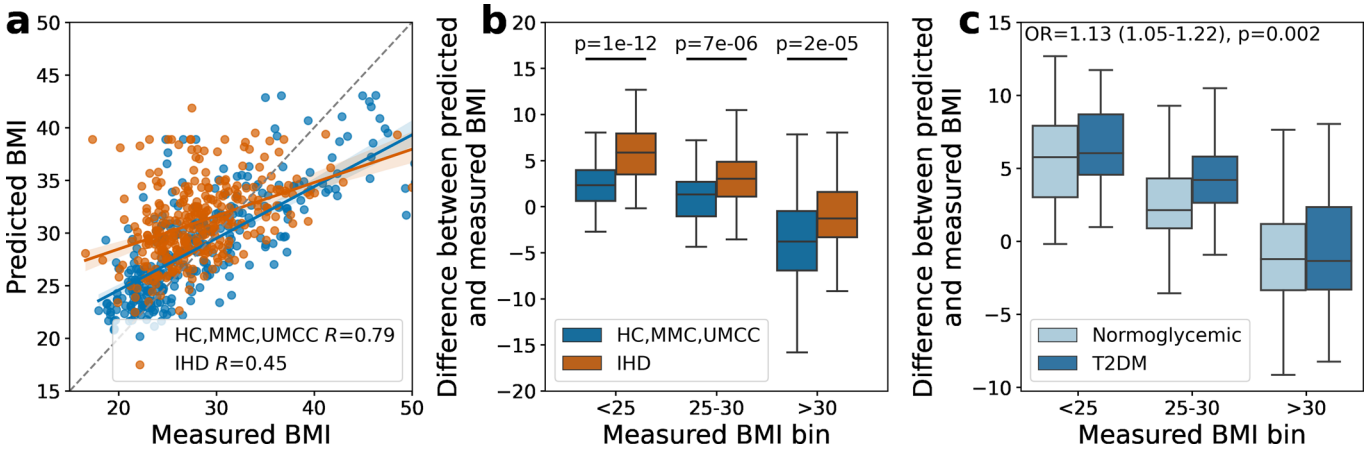
Extended Data Fig. 3 | Depletion of ACS-related bacteria SGB 4712 replicates in an independent validation cohort. (a) Box plots showing the relative abundance of the unknown bacterial species SGB 4712 (y-axis: center, median; box, IQR; whiskers, 1.5×IQR; log scaled) in our ACS and matched controls (x-axis; n = 80 each). The P-value shown is computed using the two-sided Mann–Whitney *U*-test. **(b)** Relative abundance of the unknown bacterial species SGB 4712 (y-axis: center, median; box, IQR; whiskers, 1.5×IQR; log scaled) in four groups from the MetaCardis validation cohort (x-axis; HC, healthy controls, n = 275; MMC, metabolically matched controls, n = 218; UMCC, untreated metabolically compromised controls, n = 211; IHD, ischaemic heart disease, n = 319). The P-value shown is computed using the two-sided Mann–Whitney *U*-test. r.a., relative abundance.



Extended Data Fig. 4 | Clinical data correlates with metabolic deviations. (a) The mean weighted R^2 of genetics for ACS-enriched metabolites (y-axis) versus chronological age (x-axis). Dots are colored by sex. Spearman correlation is computed over all samples (Spearman $\rho = 0.18$; $p = 0.032$). **(b)** The mean weighted R^2 of traditional risk factors for ACS-depleted metabolites (y axis) versus chronological age (*x* axis; Spearman $\rho = 0.33$; $p = 7.7 \times 10^{-5}$). **(c)** The mean weighted R^2 of genetics for ACS-enriched metabolites (y axis: center, median; box, IQR; whiskers, 1.5×IQR) in ACS patients who had a combined CVD outcome (defined as either: acute myocardial infarction, acute stroke, unplanned PCI, or cardiovascular-related death; x axis) versus not (two-sided Mann–Whitney *U-test*, $p = 0.002$).



Extended Data Fig. 5 | Replication of higher predicted BMI in ACS individuals based on NMR metabolomics. Figure panels refer to results of serum metabolomics-based prediction model of BMI trained in a non-ACS control cohort (n = 763) and evaluated on held-out test sets consisting of both controls (n = 179) and individuals with ACS (n = 179; Methods). **(a)** Measured (*x* axis) versus predicted (y-axis) BMI for both controls (blue) and ACS (orange) individuals. Line and shaded coloring represent the fitting of a linear model and the 95% confidence interval. **(b)** Difference between predicted and measured BMI (y axis: center, median; box, IQR; whiskers, 1.5×IQR) of individuals, binned into three BMI groups (<25, 25-30, >30; x-axis). The P-values shown are computed using the two-sided Mann–Whitney *U*-test. BMI, body mass index.



Extended Data Fig. 6 | Replication of higher predicted BMI in IHD individuals in the MetaCardis study. Figure panels refer to results of serum metabolomics-based prediction model of BMI trained in a cohort of individuals without IHD and evaluated on held-out test sets consisting of both 319 IHD and 319 non-IHD individuals (Methods). **(a)** Measured (x axis) versus predicted (y axis) BMI for healthy controls (HC; blue), metabolically matched controls (MMC; blue), and untreated metabolically compromised controls (UMCC; blue), and individuals with ischaemic heart disease (IHD; orange). Line and shaded coloring represent the fitting of a linear model and the 95% confidence interval. **(b)** Difference between predicted and measured BMI (y axis: center, median; box, IQR; whiskers, 1.5×IQR) of individuals, binned into three BMI groups (<25, 25-30, >30; x axis). **(c)** Same as in (b) only for individuals with IHD, and each bin is separated into normoglycemic versus T2DM patients. Higher predicted BMI is associated with an increased incidence of T2DM (OR = 1.13, 95% CI = 1.05-1.22, p = 0.002; a logistic regression model adjusted for BMI and age; Methods). The p values shown are computed using the two-sided Mann–Whitney *U*-test. BMI, body mass index; T2DM, type 2 diabetes mellitus; OR, odds ratio; CI, confidence interval.

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.