

RESEARCH ARTICLE



## Prominent mediatory role of gut microbiome in the effect of lifestyle on host metabolic phenotypes

Solia Adriouch<sup>a,1</sup>, Eugeni Belda<sup>a,b,1</sup> , Timothy D. Swartz<sup>c</sup>, Sofia Forslund<sup>d,e,f,g,h</sup> , Edi Prifti<sup>a,b</sup> , Judith Aron-Wisniewsky<sup>a,i</sup> , Rima Chakaroun<sup>j,k</sup> , Trine Nielsen<sup>l,m,n</sup> , Christine Poitou<sup>a,i</sup> , Pierre Bel-Lassen<sup>a,i</sup> , Christine Rouault<sup>a</sup>, Tiphaine Le Roy<sup>a</sup>, Petros Andrikopoulos<sup>o</sup> , Kanta Chechi<sup>o</sup> , Francesc Puig-Castellvi<sup>p</sup> , Inés Castro Dionicio<sup>p</sup> , Philippe Froguel<sup>o,p</sup> , Bridget Holmes<sup>q</sup> , Rohia Alili<sup>a</sup>, Fabrizio Andreelli<sup>a,r</sup> , Hedi Soula<sup>a</sup> , Joe-Elie Salem<sup>s</sup> , Gwen Falony<sup>t</sup>, Sara Vieira-Silva<sup>u,v</sup> , Jeroen Raes<sup>w</sup> , Peer Bork<sup>h,x</sup> , Michael Stumvoll<sup>k,y</sup> , Oluf Pedersen<sup>l,z,aa</sup> , S. Dusko Ehrlich<sup>ab,ac</sup> , Marc-Emmanuel Dumas<sup>o,p,ad</sup> , Jean-Michel Oppert<sup>i</sup> , Maria Carlota Dao<sup>ae</sup> , Jean-Daniel Zucker<sup>a,b</sup>  and Karine Clément<sup>a,b,i</sup> 

<sup>a</sup>Sorbonne Université, Inserm, Nutrition and Obesity: Systemic Approaches, NutriOmics, Research Unit, Paris, France; <sup>b</sup>Sorbonne Université, IRD, Unité de Modélisation Mathématique et Informatique des Systèmes Complexes, UMMISCO, Bondy, France; <sup>c</sup>Integrative Phenomics, Paris, France; <sup>d</sup>Charité-Universitätsmedizin Berlin, Freie Universität Berlin and Humboldt-Universität zu Berlin, Berlin, Germany; <sup>e</sup>Max Delbrück Center for Molecular Medicine in the Helmholtz Association (MDC), Berlin, Germany; <sup>f</sup>Experimental and Clinical Research Center, Max Delbrück Center for Molecular Medicine and Charité-Universitätsmedizin Berlin, Berlin, Germany; <sup>g</sup>DZHK (German Centre for Cardiovascular Research), partner site Berlin, Berlin, Germany; <sup>h</sup>Structural and Computational Biology Unit, EMBL, Heidelberg, Germany; <sup>i</sup>Assistance Publique – Hôpitaux de Paris, Department of Nutrition, Pitié-Salpêtrière Hospital, Sorbonne University, Paris, France; <sup>j</sup>Wallenberg Laboratory, Department of Molecular and Clinical Medicine and Sahlgrenska Center for Cardiovascular and Metabolic Research, University of Gothenburg, Gothenburg, Sweden; <sup>k</sup>Medical Department III - Endocrinology, Nephrology, Rheumatology, University of Leipzig Medical Center, Leipzig, Germany; <sup>l</sup>Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Science, University of Copenhagen, Copenhagen; <sup>m</sup>Steno Diabetes Center Copenhagen, Herlev, Denmark; <sup>n</sup>Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark; <sup>o</sup>Division of Systems Medicine, Department of Metabolism, Digestion and Reproduction, Imperial College London, London, UK; <sup>p</sup>EGENODIA Inserm U1283, CNRS UMR8199, Institut Pasteur de Lille, Lille University Hospital, European Genomic Institute for Diabetes (EGID), Université de Lille, Lille, France; <sup>q</sup>FAO, Food and Nutrition Division, Rome, Italy; <sup>r</sup>Assistance Publique – Hôpitaux de Paris, Department of Diabetology, Pitié-Salpêtrière Hospital, Paris, France; <sup>s</sup>Assistance Publique – Hôpitaux de Paris (AP-HP), Clinical Investigation Center (CIC-1901), Department of Pharmacology, Pitié-Salpêtrière Hospital, Sorbonne Université, Institut National de la Santé et de la Recherche Médicale (INSERM), Paris, France; <sup>t</sup>Host Microbe Interactomics Group, Wageningen University & Research, Wageningen, The Netherlands; <sup>u</sup>Institute of Medical Microbiology and Hygiene and Research Center for Immunotherapy (FZI), University Medical Center of the Johannes Gutenberg-University Mainz, Mainz, Germany; <sup>v</sup>Institute of Molecular Biology (IMB), Mainz, Germany; <sup>w</sup>VIB Center for Microbiology, VIB, Leuven, Belgium; <sup>x</sup>EMBL, Heidelberg, Germany; <sup>y</sup>Helmholtz-Institute of Metabolism, Obesity and Vascular Research (HI-MAG) at Leipzig University, Helmholtz Munich, Germany; <sup>z</sup>Center for Clinical Metabolic Research, Herlev-Gentofte University Hospital, Copenhagen, Denmark; <sup>aa</sup>Department of Medicine, University of Copenhagen and Herlev-Gentofte University Hospital, Copenhagen, Denmark; <sup>ab</sup>Department of Clinical and Movement Neurosciences, UCL Queen Square Institute of Neurology, London, UK; <sup>ac</sup>Université Paris-Saclay, INRAE, MGP, Jouy-en-Josas, France; <sup>ad</sup>Victor Phillip Dahdaleh Institute of Genomic Medicine, McGill University, Montréal, QC, Canada; <sup>ae</sup>Department of Agriculture, Nutrition, and Food Systems, University of New Hampshire, Durham, NH, USA

### ABSTRACT

Lifestyle factors influence both gut microbiome composition and host metabolism, yet their combined and mediating effects on host phenotypes remain poorly characterized in cardiometabolic populations. In 1,643 participants from the MetaCardis study, we developed a composite lifestyle score (QASD: dietary quality, physical activity, smoking, and diet diversity) that outperformed individual lifestyle variables in explaining microbial gene richness and exhibited a significant impact on the gut microbiome composition. While bidirectional pathways linking the QASD score, host phenotypes, and microbiome composition were assessed, causal inference-based mediation analyses indicated stronger effects when the microbiome was modeled as the mediator variable, particularly in relation to the insulin resistance-associated profile. Microbiome gene

### ARTICLE HISTORY

Received 30 July 2025  
Revised 28 November 2025  
Accepted 1 December 2025

### KEYWORDS

Gut microbiome; nutrition; lifestyle; mediation

**CONTACT** Eugeni Belda  [eugeni.belda@ird.fr](mailto:eugeni.belda@ird.fr); Karine Clément  [karine.clement@inserm.fr](mailto:karine.clement@inserm.fr)  Sorbonne Université, Inserm, Nutrition and Obesity: Systemic Approaches, NutriOmics, Research Unit, Paris, France

<sup>1</sup>Authors contributed equally.

 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/19490976.2025.2599565>.

© 2025 The Author(s). Published with license by Taylor & Francis Group, LLC.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

richness emerged as a key mediator explaining 27.8% of QASD score's effect on the insulin resistance marker (HOMA-IR), while no significant mediation was observed on BMI. Extended mediation analyses on microbial species and serum metabolomics deconfounded for drug use and clinical profiles identified 47 mediations where microbial taxa mediated more than 20% of the effect of the QASD score on serum metabolites associated with insulin resistance. Notably, several *Faecalibacterium* lineages enriched in individuals with high QASD score played a significant mediatory role in increasing the serum biomarkers of microbiome diversity (as cinnamoylglycine or 3-phenylpropionate). Conversely, elevated levels of secondary bile acids in individuals with low QASD scores were strongly mediated by high levels of *Clostridium bolteae*. These findings highlight distinct and clinically relevant microbiome pathways linking lifestyle behaviors to cardiometabolic risks.

One sentence summary:

The gut microbiome mediates the impact of diet quality and diversity, physical activity and smoking status – combined in a composite lifestyle score – on cardiometabolic phenotypes.

## Introduction

Various medical conditions underpinned by chronic systemic inflammation and insulin resistance, from type 2 diabetes and obesity<sup>1-3</sup> to cardiovascular diseases,<sup>4</sup> have been consistently associated with alterations in the gut microbiota. In these diseases, an imbalance in microbial populations is characterized by reduced gut microbial gene richness (GMGR) and diversity, along with significant changes in microbial composition and function.<sup>5</sup>

Reports regularly describe the importance of a wide array of lifestyle factors in the onset and progression of cardiometabolic diseases, also suggesting a potential mediating role of the gut microbiome in these relationships.<sup>6,7</sup> However, while many studies have reported associations in human populations, only a few have sought to characterize microbiome features that may mechanistically mediate the effects of lifestyle patterns on disease pathophysiology and phenotypes. Among modifiable lifestyle factors, dietary patterns, physical activity, sedentary behavior, and smoking are well-established contributors to cardiometabolic health, that may act either directly or through the modification of gut microbial composition and function.<sup>8,9</sup>

Dietary habits are commonly evaluated based on the intake of specific food items, considering their protective or harmful effects.<sup>10</sup> To capture dietary patterns in relation to health outcomes, several scoring systems have been developed.<sup>10</sup> Among these, the alternative healthy eating index (AHEI) and the dietary approaches to stop hypertension (DASH) score are widely used to assess diet quality and have been linked to reduced cardiovascular risks.<sup>11</sup> On the other hand, the alternative dietary inflammatory index (aDII),<sup>12</sup> capture the inflammatory potential of the diet. Few studies have systematically examined the relationship between these score-based dietary patterns and the gut microbiome composition and potential interactions with cardiometabolic phenotypes.<sup>13-15</sup>

Another component of food consumption is diet diversity (commonly defined by the number of distinct food groups) that is widely recommended to ensure nutrient adequacy.<sup>16</sup> However, dietary diversity is not necessarily indicative of overall diet quality.<sup>17</sup> Commonly used dietary diversity scores rarely account for the distribution of food items within each food groups, referred to here as dietary variety.<sup>17</sup> As such, ecological diversity indices can address these aspects, providing a framework for evaluating the interplay between dietary patterns, the gut microbiome<sup>18</sup> and clinical phenotypes. Previous studies have focused on healthy or elderly populations. For instance, one study reported a positive association between dietary diversity and longitudinal microbial stability in 34 healthy individuals but found no relationship with microbiome richness.<sup>19</sup> The PREDICT 1 study demonstrated that fecal microbiome diversity was associated with habitual dietary intake and diversity in the general population.<sup>7</sup> Another study in 445 elderly subjects employed a diversity score similar to the Simpson index, linking a more diverse diet to a more even distribution of bacterial groups taxa.<sup>20</sup> However, few studies have examined how dietary diversity, independent of dietary quality, may be related to the gut microbiome composition and metabolic health across varying degrees of cardiometabolic disease severity.<sup>7,19,20</sup>

Other lifestyle factors are important for cardiometabolic health. For instance, the Global Burden of Disease Study report (2017) highlights the substantial health risks associated not only with inadequate fruit and vegetable consumption,<sup>21</sup> but also with alcohol intake, smoking, and insufficient physical activity.<sup>21</sup> Alcohol consumption has been shown to differentially affect the gut microbiome composition depending on the presence of underlying disease.<sup>7,22,23</sup> Similarly, physical activity<sup>24</sup> and smoking status<sup>25</sup> have been shown to significantly modify the gut microbiome composition. This suggests that integrative lifestyle indices accounting for these factors should be of interest when investigating the microbiome composition and its links with host phenotypes. However, no study to date has examined the combined effects of physical activity and smoking status alongside dietary indices,<sup>26</sup> in relation to the gut microbiome and metabolic health.<sup>27,28</sup> In particular, it remains to be identified whether a composite lifestyle score combining these lifestyle factors would better capture these associations with the microbiome than individual factors and whether specific microbiome features may partly mediate the effect of lifestyle factors on metabolic phenotypes.

We addressed these aspects in 1,643 participants from the European MetaCardis population encompassing a broad spectrum of cardiometabolic phenotypes and developed a composite lifestyle score, named the QASD, that includes diet quality, physical activity, smoking status, and dietary diversity, to examine its associations with metabolic health phenotypes, notably insulin resistance markers, the serum metabolome, and the fecal microbiome composition in comparison with individual lifestyle features. In addition, we performed bidirectional mediation analyses, providing novel insights into the complex, interrelated pathways linking lifestyle behaviors, the gut microbiome, and cardiometabolic risks.

## Materials and methods

### *Study cohort and sample acquisition*

This study is based on the European MetaCardis cohort, which was originally designed to investigate the relationships between different stages of disease and gut microbiome variation. The cross-sectional MetaCardis cohort included 2,214 participants, covering a wide range of metabolic and cardiac phenotypes. Participants were recruited between 2013 and 2015 from Denmark, Germany, and France. Details regarding the study cohort, including protocol information, exclusion criteria, group definitions, biochemical analyses, and the collection of anthropometric and clinical data, have been extensively described in previous publications.<sup>2,29,30</sup>

For the present study, we focused on a subset of 1,643 individuals, including healthy, normal-weight participants (BMI <25 kg/m<sup>2</sup>), as well as individuals with various metabolic conditions, such as type 2 diabetes (T2D), metabolic syndrome (MS), obesity and cardiovascular disease. This subset is summarized in the flow chart in Supplementary Figure 1 and detailed in Supplementary Table S1. The study protocol was approved by the Ethics Committees of the Medical Faculty at the University of Leipzig, Germany (application number: 047-13-28012013), the Capital Region of Denmark (H-3-2013-145), and the 'Comité de Protection des Personnes' (CPP) Ile-de-France III, France (no. IDRCB2013-A00189-36). The study was registered at <https://clinicaltrials.gov/> (NCT02059538). The cohort design adhered to ethical regulations, including the Declaration of Helsinki and European privacy laws. All participants provided written informed consent.

As a replication cohort for individuals with overweight and obesity, we used data from the GutInside study (see Supplementary Table S2) described in.<sup>31</sup> Participants were recruited in France from September 2018 to January 2020 across various regions and included 433 individuals with overweight or obesity (BMI ≥ 25 kg/m<sup>2</sup>). Although the GutInside participants were involved in a dietary intervention program, only baseline data were analyzed for this study. Similar questionnaires and food records were used to evaluate lifestyle factors.

### *Clinical and anthropometric variables and definition of study groups*

Study groups were defined according to international disease criteria. Overweight and obesity were classified using the WHO criteria, with overweight defined as BMI ≥25 kg/m<sup>2</sup> and obesity defined as

BMI  $\geq 30$  kg/m<sup>2</sup>. MS was determined based on the International Diabetes Federation's 2005 Consensus Worldwide Definition (<http://www.idf.org/metabolic-syndrome>), which requires a waist circumference  $>94$  cm in men and  $>80$  cm in women, along with any two of the following four factors: (i) triglycerides  $\geq 1.7$  mmol/L or treatment for lipid abnormalities (e.g., statins, fibrates, or ezetimibe); (ii) high-density lipoprotein (HDL) cholesterol  $<1.03$  mmol/L in men and  $<1.29$  mmol/L in women, or treatment for lipid abnormalities; (iii) blood pressure (BP)  $\geq 130/85$  mmHg, or treatment with antihypertensive medication; and (iv) fasting plasma glucose  $\geq 5.6$  mmol/L or a diagnosis of T2D. T2D was defined using the American Diabetes Association (ADA) criteria: fasting glucose  $>6.9$  mmol/L, 2-h values during an oral glucose tolerance test  $>11$  mmol/L, HbA1c (glycated hemoglobin)  $\geq 6.5\%$ , or the use of anti-diabetic medication (American Diabetes Association, 2018). Hypertension status was defined according to the American College of Cardiology and American Heart Association guidelines.<sup>32</sup>

For obesity, participants were categorized into two groups: Group 2A: individuals with Grade II obesity (BMI  $\geq 35$  kg/m<sup>2</sup>) without T2D or previous cardiovascular conditions. Group 2B: Predominantly individuals with Grade III obesity (BMI  $\geq 40$  kg/m<sup>2</sup>) eligible for bariatric surgery. T2D was not an exclusion criterion, and participants generally exhibited more severe metabolic impairments compared to Group 2A. For descriptive purposes, Groups 2A and 2B were combined in this study. Participants with heart failure were defined according to guidelines from the American College of Cardiology, American Heart Association, and Heart Failure Society of America.<sup>33</sup> The group with coronary heart disease included patients with first events of acute coronary syndromes, chronic coronary artery disease with or without heart failure (defined as left ventricular ejection fraction  $\geq 45\%$  or  $<45\%$ , respectively). Participants with heart failure due to non-coronary causes were excluded.

Anthropometric measurements, including weight, height and waist circumference, were obtained during clinical visits using standardized procedures and calibrated equipment. Body fat mass and fat-free mass were assessed by bioelectrical impedance analysis. Systolic and diastolic blood pressure were measured using a mercury sphygmomanometer, with three readings taken on each arm; the average of the last two measurements from the right arm was used for analysis. A detailed record of prescribed medications and the patient's medical history was collected, with drug treatments categorized according to their molecular class, as described previously.<sup>29</sup>

## Dietary intake and lifestyle data assessment

### General questionnaire and database

In the MetaCardis EU project, a clearly defined objective was to study the interplay between lifestyle, diet, the fecal metagenome and host metabolic conditions. Dietary data for the MetaCardis population were collected using a web-based, validated food-frequency questionnaire (FFQ) tailored to the cultural habits of each recruitment country. The MetaCardis FFQ was adapted from the validated European Prospective Investigation of Cancer (EPIC)-Norfolk FFQ and incorporated elements from several other European FFQs. The French version of the FFQ version contains 159 food and beverage items, the German FFQ contains 143 items, and the Danish FFQ contains 153 items. The FFQ asked participants to report the frequency of consumption of each food item over the last 12 months, using a 9-level scale, ranging from "never or less than once a month" to "6 times or more per day". For each food item, a generic portion size was specified. Thus, initial responses to the FFQ were frequency choices indicating consumption units per month, week, or day for each food items. These frequencies were converted to daily intake values and then translated into food group intakes (in g/day) and nutrient intakes (in g/day, mg/day, or mcg/day) using data on portion sizes (i.e., grams per portion) and nutrient contents (i.e., nutrients per portion) sourced from nationally relevant food composition databases. A validation study comparing the FFQ against repeated 24-h dietary records among 324 French MetaCardis participants showed good validity for micronutrient intake.<sup>34</sup>

Several compounds and nutrients have been added to the original composition table, allowing for more precise analysis of specific compounds, such as subclasses and totals of polyphenols (using the updated Phenol-Explorer table<sup>35</sup>) and nutrients/metabolites from updated databases, including the USDA (<https://>

fdc.nal.usda.gov/), FooDB (<https://foodb.ca/>), and the Danish National Food Database (DTU Food Database: <https://frida.fooddata.dk/>) for biotin intake<sup>3</sup>. Each nutrient or metabolite of interest was carefully integrated into a consortium internal composition table through dietary matching and detailed recipe decomposition, which was performed by dietitians and experts. The basal metabolic rate (BMR) for each participant was estimated using the Harris and Benedict equation. Participants who significantly under- or overreported energy intake (defined as <0.5 BMR or >3.5 BMR) were excluded from the dietary analysis (less than 10% of the subjects presented nutritional data).

Physical activity and sedentary behavior data were collected using the Recent Physical Activity Questionnaire (RPAQ).<sup>33</sup> The RPAQ assesses habitual physical activity over the past month across four domains: home, work, travel, and leisure time. Physical activity assessment using the RPAQ has been validated against energy expenditure measurements using the doubly-labeled water method.<sup>36</sup> The physical activity data were computed as hours per week and as the metabolic equivalent of task (MET)-hour per week, using a published compendium of physical activities and their corresponding MET values. One MET represents the energy expended at rest, equivalent to 3.5 ml/min/kg of oxygen consumption.<sup>37</sup> Sedentary behavior was assessed through leisure screen time, which represents the sum of time spent watching TV and using the computer during leisure time, expressed in hours/week.

Smoking status data were obtained through multiple questions regarding smoking habits, including the subject's smoking status (non-smoker, former smoker, passive smoker, current smoker), the number of cigarettes smoked, and the date of cessation or resumption of smoking. For the QASD score calculations, smoking status was recoded into two categories (non-smoker/smoker), as shown in Supplementary Table S1. The initial smoking status classification was cross-checked with quantitative data on cigarette consumption, start/stop dates, and passive smoking exposure to confirm the final categorization (non-smoker/smoker).

### **Food groups and nutrient content**

Food items from country-specific FFQ were combined into a list of 160 food items. Food group consumption was calculated by organizing the 160 food items into groups of nutritional interest. The original 22 groups in the MetaCardis study were further divided into 42 subgroups (as shown in Supplementary Table S3, diversity 1 column). The dietary diversity scores were based on these 42 subgroups, following specific criteria. For sensitivity analysis, dietary diversity scores were also constructed by excluding "liquid" foods, such as hot drinks, fruit juices, sugary drinks, alcoholic beverages, and milk, as their quantities could disproportionately affect the diversity scores based on consumption uniformity (Simpson index).

### **Dietary "inflammatory potential" of the diet**

The "inflammatory potential" of the diet was assessed using the alternative dietary inflammatory index (aDII), an adapted version of the original dietary inflammatory index (DII) developed by Cavicchia et al.<sup>12</sup> and based on the methodology outlined by van Woudenberg et al.<sup>38</sup> The aDII includes 34 food items, standardized after adjusting for energy intake using the residual method developed by Willett et al.<sup>39</sup> to minimize variations in dietary intake due to differences in physical activity, body size, and metabolic efficiency. Each food item's value was then multiplied by a literature-based inflammatory weight that reflected the pro/anti-inflammatory potential of the food item,<sup>12</sup> and the resulting values were summed to calculate the aDII score. Higher aDII scores indicate a more proinflammatory diet.

### **Dietary quality scores**

The alternative healthy eating index (AHEI) and dietary approaches to stop hypertension (DASH) scores were computed to evaluate overall dietary quality, and two *a priori* scores, taking into account the existing correlations between the different components of the diet developed in order to overcome the limitations linked to approaches focusing only on food or nutrient groups<sup>40</sup> and based on foods and nutrients

predictive of chronic disease risk.<sup>41,42</sup> As a measure of healthy US-style eating, the AHEI-2010 assigns 0–10 points to each of 11 dietary components based on the portion size.<sup>40,41</sup> The DASH score contains 8 components, each of which receives 1–5 points according to its consumption quintile.<sup>43</sup>

**Dietary diversity and variety scores:** Dietary diversity and variety scores were constructed based on different food group categorizations (42 groups aggregated, with or without “liquid” foods, as shown in Supplementary Table S3) and two different estimation methods: count-based and evenness-based.

**Count-based scores:** We employed count-based measures of diversity, which record the number of distinct food items and groups consumed, without considering the quantities of each consumed food. Simple count-based scores were calculated by assigning 0 points for each food group (for diversity scores) or individual food item (for variety scores) if the participant did not consume the food (consumption = 0 grams) and 1 point if the food was consumed (consumption > 0 grams). Thus, the total diet diversity score ranged from 0 to 42, based on the initial food groups derived from the FFQ (Supplementary Table S3 for the initial 42 groups). While count measures simply tally different food groups and subgroups, they do not account for the distribution of food quantities. Previous research has been limited by these count-based measures and inconsistencies in the number and types of food groups considered.<sup>17</sup> To address these limitations, we conducted sensitivity analyses, systematically varying the number of food groups and excluding liquid foods from the diet.

**Evenness-based scores:** We also used diversity measures that account for both the number and the distribution of different food items, such as the Berry index, also known as the Simpson index,<sup>44</sup> which is commonly applied to assess metagenomic or ecological diversity. This index increases when food items are consumed more equally rather than when they are concentrated in fewer groups. Simpson-based scores are calculated by treating food groups (for diversity) or individual foods within a group (for variety) as different species, similar to how ecological indices measure species diversity in an environment. The Simpson index considers both the number of “species” (food groups) present and their relative abundance:

$$\text{Simpson Index} = 1 - \sum_{i=1}^R P_i^2$$

where  $R$  represents the richness or number of food groups considered, and  $P_i^2$  is the relative abundance of each group (the share of food group  $i$  in the total amount consumed). Subtracting this score from 1 results in an index ranging from 0 to 1, where 0 indicates minimal diversity and 1 represents maximal diversity (equal consumption of all considered items).

To address inconsistencies in the number and types of foods or food groups, we performed sensitivity analyses, varying the groups included in the calculations. Specifically, for total diet diversity, we created scores that excluded liquid foods (see Supplementary Table S3 for additional food groups excluding beverages and milk). We calculated the healthy food diversity (HFD) score developed by Drescher et al.,<sup>45</sup> which serves as an indicator of healthy food diversity, specifically focusing on the diversity of health-promoting products. We also derived a normalized composite dietary diversity and variety score,<sup>46,47</sup> which combines both diversity and variety measures (simple counts), normalized on a scale of 0–100.

These scores capture various aspects of dietary diversity and variety. Some scores reflect the entire diet (referred to as “diversity all diet”), while others focus on specific dietary groups or items (labeled as “variety”). We used diversity and variety scores based on simple counts within consumption groups (labeled as “count”) and scores that account for the evenness of consumption within dietary groups (labeled as “Simpson”).

**Quality activity smoking diversity (QASD) lifestyle score construction:** a new lifestyle score, the QASD, was developed to incorporate diet quality, total physical activity, smoking status, and dietary diversity. Lifestyle components on continuous scales were categorized into three levels to enhance interpretability, harmonize variables measured on different scales (e.g., diet quality-diversity, physical activity), while for smoking status, a value was assigned based on binary classification. This approach ensures robustness across cohorts with heterogeneous distributions, which is consistent with approaches used in other lifestyle indices (e.g., the Healthy Lifestyle Index, Life’s Simple 7, Mediterranean Diet Score<sup>41,48,49</sup>). The values of the four individual scores used to build the QASD are given below.

- **Diet quality (AHEI):** 0 points for the low tertile, 1 point for the medium tertile, and 2 points for the high tertile.

- **Dietary diversity** (Simpson index, excluding beverages): 0 points for the low tertile, 1 point for the medium tertile, and 2 points for the high tertile.
- **Total physical activity** (MET-h.week-1): 0 points for the low tertile, 1 point for the medium tertile, and 2 points for the high tertile.
- **Smoking status**: 0 points for smokers and 1 point for non-smokers or ancient smokers.

The QASD score was computed by taking the unweighted sum of these four scores. The use of the unweighted sum of the subscores is justified because of its simplicity, neutrality, and robustness. In clinical settings, where scores based on variables often represent independent dimensions of patient status (e.g., physiological metrics, diagnostic findings, or treatment responses), the unweighted sum avoids introducing biases from arbitrary or uncertain weightings. This approach ensures equal contribution from each score, maintaining fairness when there is no clear evidence to prioritize one over the others, which is the case here. It also minimizes the risk of errors associated with miscalibrated weightings, which is critical in medical decision-making.<sup>50</sup> Furthermore, the unweighted sum is transparent, easy to interpret, and widely applicable, making it appropriate for clinical contexts where clarity and comparability are essential. This straightforward aggregation method used to construct the QASD score aligns with evidence-based practices and is frequently used in composite medical indices and scoring systems.<sup>51</sup> The QASD score ranges from 0 to 7. To address low numbers in some categories during stratification, we grouped the scores into five categories: “0–1–2,” “3,” “4,” “5,” and “6–7” for Figures 1, 3, and Supplementary Figures 2–5, and into three categories: “0–1–2–3,” “4,” and “5–6–7” for Figures 2, 4, 5, 6, and Supplementary Figures 6–8.

### **Extraction of fecal genomic DNA and gut microbiota sequencing in MetaCardis cohort**

Details of gut microbiota sequencing and taxonomic and functional profiling in the MetaCardis cohort has been extensively described in previous publications.<sup>2,29,30</sup> Briefly, DNA from fecal samples was extracted following the International Human Microbiome Standards (IHMS) guidelines (SOP 07 V2 H). Shotgun sequencing was performed with ion-proton technology (Thermo Fisher Scientific; average  $23.3 \pm 4.0$  million (mean  $\pm$  SD) 150-bp single-end reads per sample). Alien Trimmer (v0.2.4)<sup>52</sup> was used for quality filtering of raw sequencing reads, and non-bacterial contaminants of human and food origin were removed by aligning cleaned reads vs. human genome (RCh37-p10) and the genomes of *Bos taurus* and *Arabidopsis thaliana* (97% identity threshold). Filtered high-quality reads were aligned to the 9.9-million-gene catalogue of the human gut microbiome<sup>53</sup> using Bowtie2 (v.2.3.4; 95% identity threshold),<sup>54</sup> from which a gene abundance table was obtained using METEOR v3.2 software<sup>55</sup> (<https://forgemia.inra.fr/metagenopolis/meteor>). The read counts were downsized to 10 million reads per sample and normalized by the gene size and the number of total mapped reads reported in frequency using the R package MetaOMineR.<sup>56</sup> From the downsized and normalized gene abundance table, functional and phylogenetic quantitative profiles were generated within the MOCAT2 framework.<sup>57</sup> The abundance of 1435 metagenomic species (MGS; co-abundant gene groups with more than 500 genes corresponding to microbial species defined from 1,267 human gut metagenomes used to construct the 9.9-million-gene catalog as previously described<sup>58</sup>) were estimated as the mean abundance of the 50 genes defining a robust centroid of the cluster.<sup>58</sup> The taxonomic annotation of the MGS was carried out by BLASTN of MGS genes against the NCBI database (November 2016 version), with species-level classification assigned when more than 50% of the genes matched the same reference genome with at least 95% sequence identity and 90% coverage. The GMGR was estimated from the average gene count across 10 rarefaction replicates.

### **Extraction of fecal genomic DNA and gut microbiota sequencing in GutInside cohort**

Gut microbiome sequencing of fecal samples from 433 individuals of the GutInside study<sup>31</sup> was performed by GeneWiz (<https://www.genewiz.com/>) using Illumina HiSeq platform, resulting in  $23.17 \pm 5.1$  million read pairs (mean  $\pm$  std. deviation) per sample. Raw reads were processed with NGLess<sup>59</sup> for quality trimming (minimum read quality = 25; minimum read length = 40), host contaminant removal vs. reference human

genome (min. identity = 90%, min. match size = 45), alignment of filtered reads over the 9.9 million gene catalogue (min. identity = 95%, min. match size = 45), and generation of a gene abundance table with the *dist1* metric of NGLess, equivalent to the counting strategy described above for the MetaCardis cohort (first, the genome abundances are computed from unique mapped reads and then are corrected by the multiple mapped reads weighted by the coverage of unique mapped reads). The gene abundance table was processed for rarefaction and FPKM normalization with the MetaOMineR<sup>56</sup> R package. The raw gene abundance table was rarefied to 10 million reads per sample, and the rarefied gene abundance table was normalized according to the FPKM strategy, from which MGS abundances were computed as in the MetaCardis cohort (mean abundance of the 50 genes defining a robust centroid of the cluster if more than 10% of these genes yielded positive signals). From the NGLess-filtered reads, genus-level abundance profiles in the mOTU space were generated with mOTU v2.6.1,<sup>60</sup> from which enterotype classifications was performed after rarefy to 4000 counts per sample using the Dirichlet Multinomial Mixture (DMM) method<sup>61</sup> and implemented in the Dirichlet Multinomial R package (Dirichlet-Multinomial Mixture Model Machine Learning for Microbiome Data. R package version 1.28.0). The minimum Laplace metric was used as criteria to select the best stratification of the cohort,<sup>61</sup> which was at 4 discrete microbiome compositions that corresponded to the Rum, Prev, Bact1 and Bact2 enterotypes described in the Metacardis cohort.<sup>62</sup>

### **Analyses of microbial diversity and community structure**

The estimation of the explanatory power of nutritional features regarding relative microbiome profiles derived from MGS abundance data was performed using univariate and multi-variate stepwise distance-based redundancy analyses (dbRDA) as implemented in the R package *vegan* (Community Ecology Package. R package v.2.2-1).<sup>63</sup> Redundance filtering of collinear lifestyle variables was performed prior to these analyses with the R package FMradio;<sup>64</sup> threshold = 0.9 on the Spearman pairwise correlation matrix). The interindividual variation in the microbiome was visualized by principal coordinate analysis using Bray–Curtis dissimilarity on the basis of the relative abundance matrix derived from MGS abundance data. The environmental fit of nutritional covariates with a significant impact on microbiome composition based on dbRDA analyses over PCoA ordination from the Bray–Curtis inter-sample dissimilarity matrix was computed with the *envfit* function of *vegan* R package. Associations of nutritional variables with enterotype status were tested with logistic regression analyses adjusted for age, gender, center, BMI, and the intake of metformin, statins and PPI.

### **Metabolomic profiling**

We used untargeted UPLC–MS data generated by Metabolon in serum samples and NMR data in urine samples. Full details of sample processing and generation of metabolomic profiles are available in our previous report.<sup>30</sup>

### **Drug deconfounding analysis**

The deconfounding pipeline was employed to determine the extent to which observed differences in microbiome and metabolome feature abundances between pairwise levels of the QASD score (high vs. medium, high vs. low, medium vs. low; low = QASD 1, 2, 3; medium = 4; high = 5, 6, 7) are influenced by confounding factors – such as treatment or risk variables – rather than being inherent characteristics of the specific phenotype itself. We used a post hoc filtering approach implemented in the R package *metadeconfoundR* (version 0.1.8; <https://github.com/TillBirkner/metadeconfoundR> or [Zenodo](#)), developed within the MetaCardis consortium.<sup>29</sup> We followed the procedure described in.<sup>30</sup> Medication status (statin, metformin, and PPI intake, all treated as binary variables) and the MetaCardis clinical group were considered as covariates, and the country of recruitment was included as a random effect. The analysis was conducted on metagenomic (MGS, mOTU, Gut Metabolic Modules (GMM)) and metabolomic features (serum, urine) present in at least 20% of the samples. For metagenomic features, abundances were corrected for bacterial cell count by multiplying by a normalization factor calculated as the bacterial

cell count of the sample divided by the mean bacterial cell count across the entire dataset. Features with strictly deconfounded associations (SD), meaning that the QASD score adds significant explanatory power vs. a model with only the covariates (likelihood ratio test  $p$ -value  $< 0.05$ ) and where the confidence interval does not include zero, were retained for further analyses.

### Statistical analyses

Data management and statistical analyses were performed using SAS version 9.3 and R (R statistical software, Vienna, Austria). The significance levels used were  $p < 0.05$ , FDR  $< 0.01$ , and  $p$  for interaction  $< 0.1$ .

Several adjustment strategies were applied throughout the analyses. For analyses related to gut microbiome gene richness (GMGR), a normally distributed variable, we conducted sensitivity analyses considering potential confounders identified in previous MetaCardis consortium studies.<sup>2,3,29,30</sup> These confounders included age, gender, recruitment center, antibiotic use, metformin and statin intake, and BMI. To isolate the independent effects of nutritional quality, the models were adjusted for energy intake (kcal) and the AHEI score, which is consistent with standard practices for dietary intake analysis. For specific omics features, we used the analysis pipeline developed and published in the MetaCardis study to ensure comparability with other studies (see the previous Drug Deconfounding Analysis section).

### Mediation analysis

Mediation analysis was used to infer causal mediation effects between metagenomic species (MGS), serum metabolites, and pairwise levels of the QASD score (high vs. low, high vs. medium, medium vs. low; low = QASD 1, 2, 3; medium = 4; High = 5, 6, 7). To reduce the number of tests, we focused on MGS and metabolites with strictly deconfounded status from the drug deconfounding pipeline, which exhibited an absolute effect size greater than 0.1 on Cliff's delta between pairwise levels of the QASD score (as shown in Figure 4).

In the first step, pairwise associations between metabolites and MGS were tested individually using linear regression analyses, adjusted for country of origin, age, sex, and intake of metformin, statins, and proton pump inhibitors (PPI). Only significant pairs (FDR  $< 0.05$ ) were retained for further mediation analysis. MGS abundances were normalized using the empirical normal quantile transformation method<sup>65</sup> to ensure normality before analysis as proposed in.<sup>66</sup>

In the second step, we applied a similar regression framework, with the same adjustments, to identify significant mediations under two potential hypotheses regarding the direction of mediation, using the *mediate* function from the mediation R package (version 4.5.0).<sup>67</sup> The first hypothesis examined the role of MGS as mediators in the relationship between the QASD score and the host's metabolomic profile (Direction 1, Figure 5A). The second hypothesis explored whether the relationship between the QASD score and the host metabolome (as a mediator) could subsequently influence the metagenomic landscape (Direction 2 in Figure 5A). Significant mediations were defined as those with an FDR  $< 0.05$  for the average causal mediation effect (ACME), average direct effect (ADE), and total effect. Similar analyses were conducted to infer causal mediation effects between the levels of the QASD score, microbial gene richness, and clinical status of individuals, including measures of body mass (BMI) and metabolic status (glycated hemoglobin, HOMA-IR).

To further integrate insulin resistance profiles in the results of mediation analyses between the QASD, serum metabolites and MGS, we filtered significant mediations by retaining those where both the MGS and the serum metabolite were significantly associated with HOMA-IR in linear regression analyses under the same adjustment framework as in the mediation analyses (FDR  $< 0.05$ ).

## Results

### Characterization of MetaCardis population with lifestyle variables

The 1643 studied individuals of the MetaCardis population include a metabolically healthy group ( $n = 305$ ) and pathology groups defined as follows: individuals with metabolic syndrome ( $n = 220$ ), type 2 diabetes

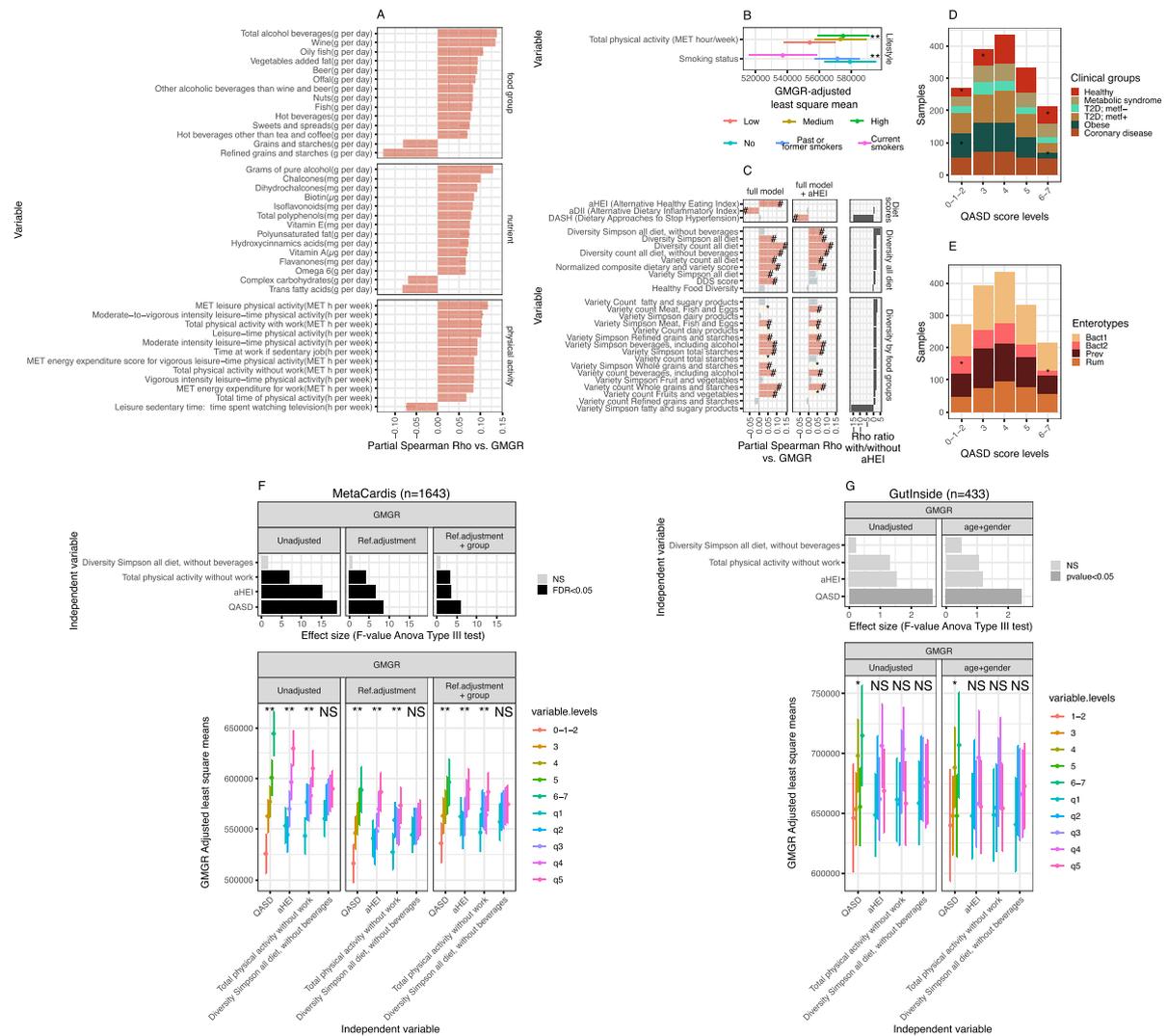
( $n = 477$ ), obesity (BMI  $\geq 30$ ;  $n = 334$ ), and cardiovascular diseases ( $n = 307$ ) (Supplementary Table 1). Pharmacological treatments have been previously identified as factors that often mask disease-related signatures of the gut microbiome.<sup>29,68</sup> Thus, medications such as 5-y retrospective exposure to antibiotics (total number of courses), metformin, statin and proton pump Inhibitors (PPI) usage were included as covariates in the analysis. Analyses of dietary variables were adjusted for total energy intake (in kcal/day), such as the AHEI score, while metagenomic analyses were adjusted for BMI or MetaCardis pathology groups (see Methods section and figure legends for details). To confirm some observations in independent populations, we used data from 433 participants of the GutInside study,<sup>31</sup> which recruited French adults with overweight or obesity across French regions between September 2018 and January 2020. Baseline data were used from this dietary intervention program. Subjects' phenotypes are provided in Supplementary Table 2. Both GutInside and MetaCardis cohorts used the same lifestyle assessment tools: e.g., Food Frequency Questionnaires (FFQ) for habitual dietary intake,<sup>34</sup> the Recent Physical Activity Questionnaire (RPAQ) for physical activity and sedentary behavior<sup>36</sup> and gathered smoking-related information, including past habits and current cigarette consumption. Shotgun metagenomic sequencing was available for both cohorts.<sup>2,29,31</sup>

### ***Dietary components, physical activity traits and smoke associate with gut microbiome gene richness***

We first examined the relationships between the GMGR and individual variables, including food groups ( $n = 65$ ), nutrients ( $n = 72$ ), physical activity ( $n = 36$ ) variables and smoking habits ( $n = 3$ , as qualitative variables). Among the 173 quantitative variables, 40 (23.12%) were significantly associated with the GMGR (FDR  $< 0.05$ ; partial Spearman correlation analyses; Figure 1A). The adjusted effect sizes were modest (Spearman's Rho ranging from  $-0.12$  to  $0.13$ ) suggesting that lifestyle variables, considered independently, modestly impact GMGR variation. As shown in Figure 1A, among the food groups, positive associations with the GMGR were observed for the intake of alcoholic beverages (e.g., wine and beer), tea and coffee (classified as hot beverages), offal, nuts, and fish (including oily fish), as well as sweets and spreads. At the nutrient level, positive associations with the GMGR were found with alcohol, vitamins A/E, D, B7/B8 (e.g., biotin) and poly-unsaturated fatty acids and omega-6 fatty acids, as well as with the intake of total polyphenols and their subclasses, such as chalcones (found in beer), isoflavonoids (found in soja, onions, wine and tea) and hydroxycinnamic acids (found in coffee). These nutrient-based findings are consistent with those at the food group level and are in agreement with reports linking polyphenols to the gut microbiome diversity and health.<sup>69,70</sup>

Our analysis also revealed negative associations with the intake of trans fatty acids, previously implicated in diet-induced inflammation and containing bacteriostatic additives,<sup>71</sup> and with sugars and carbohydrates. At the food group level, we observed a significant negative association between GMGR and "grains and starches" ( $\rho = -0.08$ , FDR =  $7.55e-03$ , Figure 1A), which was driven primarily by the intake of refined grains and simple sugars, which were grouped in this category. This food group shows the strongest negative association with the GMGR (e.g., "Refined grains and starches";  $\rho = -0.12$ , FDR =  $1.15e-05$ ; Figure 1A). This negative association was replicated in stratified analyses across the clinical subgroups (Supplementary Figure 2A). Conversely, a positive association between the intake of fruits and the GMGR was found in the healthy group (Supplementary Figure 2A;  $\rho = 0.22$ , FDR =  $0.016$ ). With respect to physical activity, variables related to both occupational and leisure-time (partly reflecting total physical activity) were positively linked to GMGR (Figure 1A,B), while sedentary behaviors such as screen time in front of television or computer (indicator of sedentary behavior) were negatively associated with the GMGR ( $\rho = -0.072$ , FDR =  $0.018$ ; Figure 1A). Smoking status was significantly associated with the GMGR, with current smokers showing a reduced GMGR (FDR =  $3.72e-04$ ; linear regression; Figure 1B). These associations remained significant after adjustments for clinical status and BMI (Supplementary Figure 2C,D).

Overall, we observed a majority of positive associations between the GMGR and many variables related to dietary intake, including some foods not typically considered healthy and physical activity, whereas sedentary behaviors, smoking, sugars and trans fatty acids were negatively associated with the GMGR.



**Figure 1.** Associations between diet, physical activity factors, and a composite lifestyle score, and gut microbial gene richness (GMGR) are shown. (A) Bar plots showing 40 food groups, nutrients, and physical activity variables significantly associated with the GMGR (FDR < 0.05). Associations are based on partial Spearman's rank correlations adjusted for age, recruitment center, energy intake (kcal), antibiotic treatments (courses in the past 5 y), and the use of metformin, statins, and proton pump inhibitors (PPIs) (MET = metabolic equivalent of task; h per week = hour per week). (B) Associations of lifestyle factors (total physical activity and smoking status) with the GMGR. Physical activity is categorized into tertiles (low/medium/high). The estimated marginal means of the GMGR across these levels are derived from linear regression models adjusted for age, recruitment center, energy intake (kcal), antibiotic treatments, and the use of metformin, statins, and PPIs (FDR < 0.05; ANOVA tests on linear regression models). (C) Associations of dietary scores (AHEI, DASH, and aDII) and 25 dietary diversity and variety scores with the GMGR (\*:  $p < 0.05$ ; #: FDR < 0.05; partial Spearman's rank correlation). The fully adjusted model (full model) included adjustments for age, recruitment center, energy intake (kcal), antibiotic treatments, and the use of metformin, statins, and PPIs. The fully adjusted model + AHEI includes additional adjustment by the alternative healthy eating index to evaluate independence from diet quality (AHEI itself excluded in these analyses). Scores are presented for global dietary scores (top), overall diet (middle) and specific food groups (bottom). The variables on the y-axis are ordered by the ratio of partial Spearman correlation coefficients with and without AHEI adjustment (right-most panel). (D,E) Distribution of individuals in the MetaCardis population across levels of the QASD score coloured by clinical group and metagenome-built enterotypes respectively (\* = FDR < 0.05, post-hoc analysis for Pearson's chi-square test for count data). (F) Top bar plots represent the effect sizes (F-values) derived from the ANOVA Type III test on linear regression models where the GMGR (dependent variable) is regressed against the QASD score and its individual components (y-axis, 1 regression model  $\times$  independent variable) under different adjustment frameworks (Ref. adjustment: age + recruitment center + energy intake (kcal) + antibiotic treatments, and use of metformin + statins + PPIs; group = metacardis clinical groups) in individuals of MetaCardis cohort. The bottom panels represent the estimated marginal means of the GMGR across the QASD score levels and quintiles of the distribution of QASD score individual components derived from the same linear regression models (\*\*: FDR < 0.05, \* =  $P$ -value < 0.05 & FDR > 0.05, NS = non-significant; ANOVA type III test). (G) Similar analyses as panel D were performed in the GutInside cohort ( $n = 433$ ).

These observations are in agreement with prior studies in general populations,<sup>7,72</sup> which also reported modest effect sizes for individual lifestyle variables.

### ***Dietary diversity and variety associate with microbiome gene richness regardless of dietary quality***

To better capture the complexity of dietary behavior, we further explored associations between the GMGR and dietary scores, including those reflecting dietary diversity, variety and quality as well as the inflammatory potential of the diet. Considering commonly used dietary scores combining food items (e.g., AHEI, DASH, and the inflammatory score (DII)), the AHEI score, which reflects diet quality, showed a significant positive association with the GMGR (partial Spearman  $\rho = 0.12$ , FDR =  $1.28e-05$ ; **Figure 1C**, adjusted model including age, recruitment center, energy intake (kcal/day), antibiotic treatments, and the use of metformin, statins, and PPI). In contrast, the aDII was negatively associated with the GMGR (partial Spearman  $\rho = -0.07$ , FDR =  $4.05e-03$ , **Figure 1C**, with the same adjusted model). No significant association was observed between the DASH score and the GMGR (**Figure 1C**). However, further analyses adjusting for diet quality (AHEI) revealed a nuanced picture: the previously significant association of the aDII with the GMGR became non-significant while the DASH score showed a significant negative association (partial Spearman  $\rho = -0.07$ , FDR =  $8.76e-03$ ). These results likely reflect shared variance across dietary scores, as both DASH and aDII were positively or negatively strongly associated with AHEI (Spearman  $\rho = 0.56$ ,  $p$  value =  $4.42e-136$ ; Spearman  $\rho = -0.44$ ,  $p$  -value =  $5.37e-81$  respectively).

Other indices encompass diet diversity (“diversity all diet”), specific food group variety (“variety”), simple counts of different foods consumed within groups (“count”), and the evenness of food consumption within groups (“Simpson”). We therefore calculated 25 scores to assess these dietary patterns and examined their statistical relationships with the GMGR (see Methods and Supplementary Table S3 for detailed descriptions). We notably calculated the normalized composite “dietary and variety score of the diet”<sup>46,47</sup> and the healthy food diversity index (HFD), which combines both diversity and variety dietary information.<sup>45</sup> Fifteen of these scores were positively associated with the GMGR (**Figure 1C**; partial Spearman  $\rho > 0$ , FDR  $< 0.05$ ; fully adjusted model), including 7 of the 9 scores describing dietary diversity. Notably, 11 of these associations remained significant after adjustment for diet quality (assessed by the AHEI). Furthermore, for 2 scores (e.g., “Diversity Simpson all diet, without beverages” and “Variety count Meat, Fish and Eggs”), the associations with the GMGR were even strengthened after adjusting for AHEI (**Figure 1C**; fully adjusted model + AHEI; see Methods), with the “Diversity Simpson all diet, without beverages” score being the one with the highest increase in the strength of the association with the GMGR after adjusting for AHEI (**Figure 1C**; Rho ratio with/without AHEI, left panel).

By decomposing food diversity indices across food groups, we found that the “Variety count of whole grains and starches” had the strongest positive association with the GMGR, followed by the variety of “fruits and vegetables” (measured by the “count” index), “Beverages, including alcohol” (measured by the count and Simpson indices) and the Simpson index for “meat, fish, and eggs” (**Figure 1C**). Conversely, the variety count for “Refined grains and starches” showed a negative, though statistically non-significant, association with the GMGR. This finding aligns with the negative associations observed at the nutrient and food group levels for “sugars and simple starches”, as shown in **Figure 1A**. Importantly, the observed positive associations with the GMGR extended to diverse foods not typically considered healthy, such as a variety of “beverages, including alcohol”. The score “Diversity count all diet” was the score most strongly associated with the GMGR that remained associated after adjusting for dietary quality. In contrast to observations in other cohorts,<sup>7</sup> the HFD score was not associated with GMGR in the MetaCardis population when considering the entire cohort or clinical subgroups (**Figure 1C**, Supplementary Figure 2B). Taken together, dietary diversity and variety, regardless of perceived food healthfulness, were positively associated with GMGR in this population with cardiometabolic disorders.

### ***A new composite lifestyle score is associated with GMGR, regardless of cardiometabolic phenotype severity***

The modest effect sizes of individual nutritional variables and the influence of other lifestyle factors on GMGR variation in the MetaCardis population prompted us to develop an integrative metric. For this

purpose, we built a composite lifestyle score, e.g., QASD for dietary quality, physical activity, smoking and dietary diversity, by assigning tertile points to diet quality (AHEI, 0/1/2), total physical activity (total physical activity in MET-h.week-1, 0/1/2), and the diet diversity score. We selected the “Diversity Simpson all diet, without beverages” score (0/1/2), because it was associated with GMGR that was statistically independent from AHEI (Figure 1C; full model + AHEI). Notably, the AHEI includes alcohol consumption, which was associated with the GMGR in univariate analysis. For smoking status, 0 point for current smokers and 1 for non-smokers were attributed. Using this additive approach by summing the scores for each component, participants received a total score between 0 (lowest QASD) and 7 (highest QASD), with higher QASD values reflecting more favorable lifestyle patterns. The distribution of QASD scores across the MetaCardis clinical groups showed significant differences (Figure 1D; chi-square test,  $p = 6.1e-09$ ). The number of healthy individuals were significantly enriched in the high-QASD score levels (6–7; FDR < 0.05, post-hoc analysis for Pearson’s chi-square test for Count Data), while subjects with obesity were significantly overrepresented in the lowest-QASD score groups (0–1–2; FDR < 0.05, post-hoc analysis for Pearson’s chi-squared test for count data). In agreement, a higher QASD score was significantly associated with improved metabolic parameters, including lower levels of glycosylated hemoglobin, triglycerides and serum CRP, as well as better corpulence profiles (e.g., lower BMI, percent body fat and waist circumference; Supplementary Table S4;  $p$ -values < 0.05, Kruskal-Wallis test).

Interestingly, we also observed significant differences in the distribution of QASD scores across the enterotypes defined in the MetaCardis population<sup>2</sup> (Figure 1E, chi-square test,  $p = 1.3e-02$ ), with the dysbiotic Bact2 enterotype being significantly enriched at low QASD score levels and significantly depleted at high QASD score levels (FDR < 0.05, post-hoc analysis for Pearson’s chi-squared test for count data). In line with this observation, we found a strong association between the composite QASD score and the GMGR in the whole population, which remained significant in several adjustment models that included the MetaCardis clinical groups as covariates (Figure 1F; FDR < 0.05; ANOVA test on linear regression models of the GMGR including the QASD and additional confounders;  $p$ -value < 0.05 on Tukey post-hoc pairwise tests across the QASD levels). The QASD score captured a broader range of GMGR variation than any of its individual components, notably the AHEI score (Figure 1F, Top Panel). This observation was partially replicated at the nominal  $p$ -value level in the GutInside cohort, where the QASD score outperformed its individual components in explaining GMGR variation (Figure 1G;  $p$ -value < 0.05; ANOVA test on linear regression models of the GMGR, including the QASD score and additional cofounders such as age and gender). Extending these analyses to other clinical variables in addition to the GMGR, such as BMI, HbA1c and HOMA-IR (as a proxy of insulin resistance), revealed significant associations between the QASD score with the three tested clinical variables across 8/9 regression frameworks, being particularly strong for HbA1c. Here, it consistently outperformed its individual components (Supplementary Figure 3A). Established scores such as the aHEI and physical activity were significantly associated with BMI and HOMA-IR in all adjustment models, with the aHEI showing the largest effect sizes for BMI and HOMA-IR (Supplementary Figure 3A). However, the QASD remained complementary, as it captured associations extending beyond single-behavior scores, particularly with HbA1c. These findings were not fully replicated in the GutInside cohort comprising only subjects with overweight and obesity, where only physical activity showed a significant association with HbA1C (Supplementary Figure 3B).

### ***Bi-directional mediation analysis reveals the QASD score's dual impact on the gut microbiome and metabolic health***

The individual associations of the QASD score with the GMGR and with the clinical profile in the MetaCardis population led us to explore potential bidirectional mediation relationships between the QASD score, the gut microbiome, and clinical outcomes to better understand how these factors interact. We first evaluated the statistical interaction between the QASD score, GMGR and clinical variables considering routine and easily measurable phenotypes, e.g., body mass index (BMI) and glucose metabolic status (using glycosylated hemoglobin for glucose control and HOMA-IR), by conducting bi-directional mediation analyses under two different hypotheses with the QASD score as the independent variable (Figure 2A). The first hypothesis (direction 1 in Figure 2A) tested whether the QASD score influences the

GMGR through variation in individuals' clinical and biological variables (as mediators). The second hypothesis (direction 2 in [Figure 2A](#)) tested whether the QASD impacts individuals' profile, via GMGR variation (e.g., as a mediator). We performed this analysis considering the upper and lower levels of the QASD score (QASD score 0–1–2–3;  $n = 662$  vs. QASD score 5–6–7;  $n = 546$ ) as well as each of these levels (e.g., “low” QASD or “high” QASD) versus intermediate QASD score level (QASD = 4;  $n = 435$ ). The mediation analyses supported both hypotheses, especially when comparing upper and lower QASD score levels in relation to metabolic variables (high vs. low, [Figure 2B](#),  $P$ -value  $< 0.05$  for the total effect, average direct effect (ADE) and average causal mediation effect (ACME); mediation analyses adjusted by age, gender, recruitment center, and intake of metformin, statin and proton pump inhibitor).

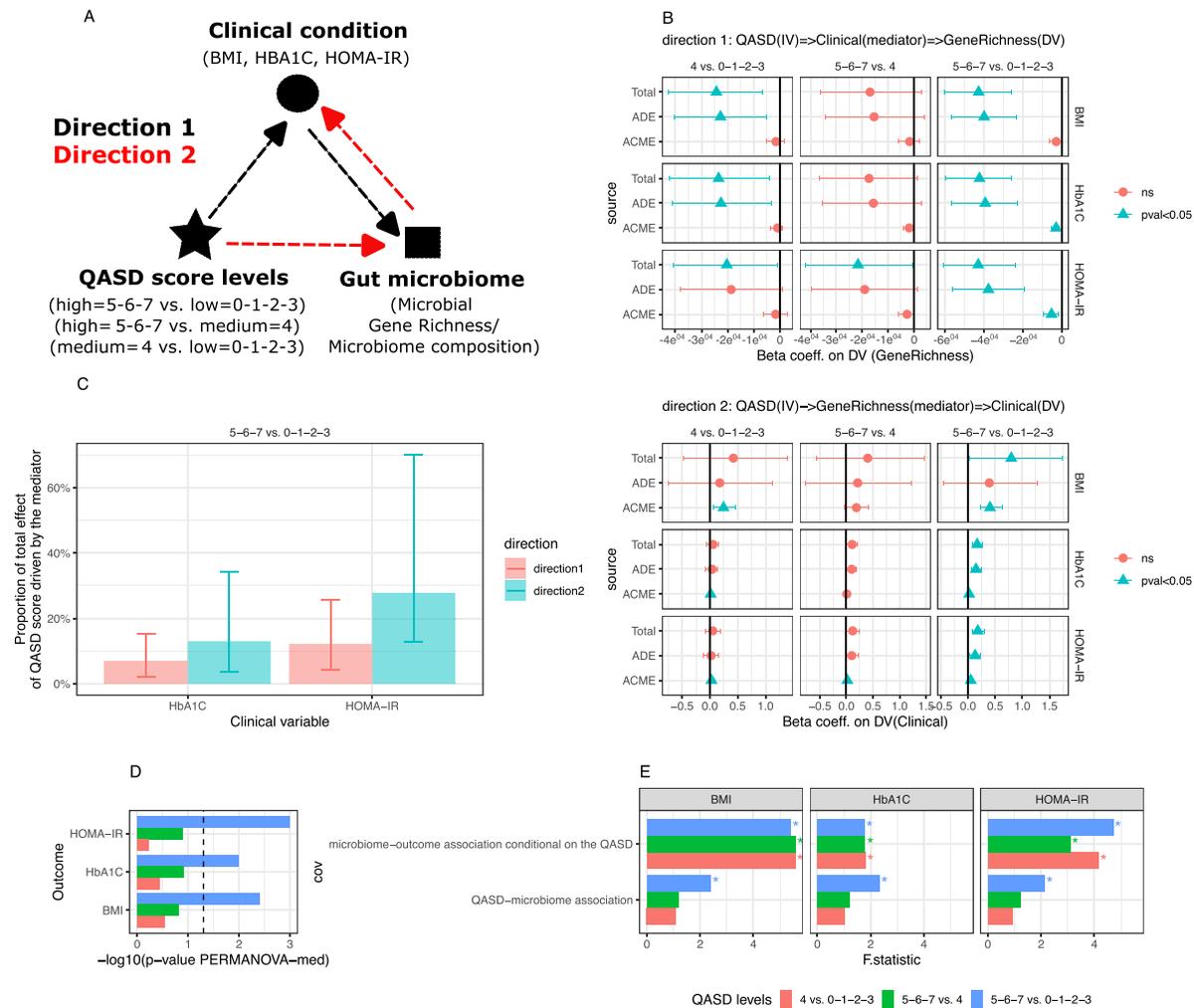
In line with Hypothesis 1, variations in HOMA-IR and glycated hemoglobin explained 12.2% and 6.8%, respectively, of the influence of the QASD score on the GMGR ([Figure 2C](#)). However, our analysis also revealed that the QASD score significantly affects individuals' glucose metabolism status and, that a significant proportion of this effect could be attributed to alterations in the GMGR. As such, in line with Hypothesis 2, 27.8% of the effect of the QASD score on HOMA-IR and 13.0% on glycated hemoglobin could be attributed to changes in the GMGR. Importantly, no significant mediation effect was observed when using BMI as an outcome or mediator. In additional support to these results, PERMANOVA-based mediation analyses<sup>73</sup> using Bray-Curtis distances derived from gut microbiome composition confirmed the significant mediation effect of the gut microbiome on the association between lifestyle (QASD score) and clinical outcomes (BMI, HbA1C and HOMA-IR) when considering the upper and lower levels of the QASD score ([Figure 2D–E](#)).

Overall, we observed a significant bi-directional mediatory effect of GMGR and metabolic phenotypes on reciprocal effects of the QASD score on these variables, although with a stronger effect size when GMGR and gut microbiome composition is explored as mediator of QASD score effect on insulin resistance marker.

### ***The QASD score and specific dietary components explains variations in gut microbiome composition***

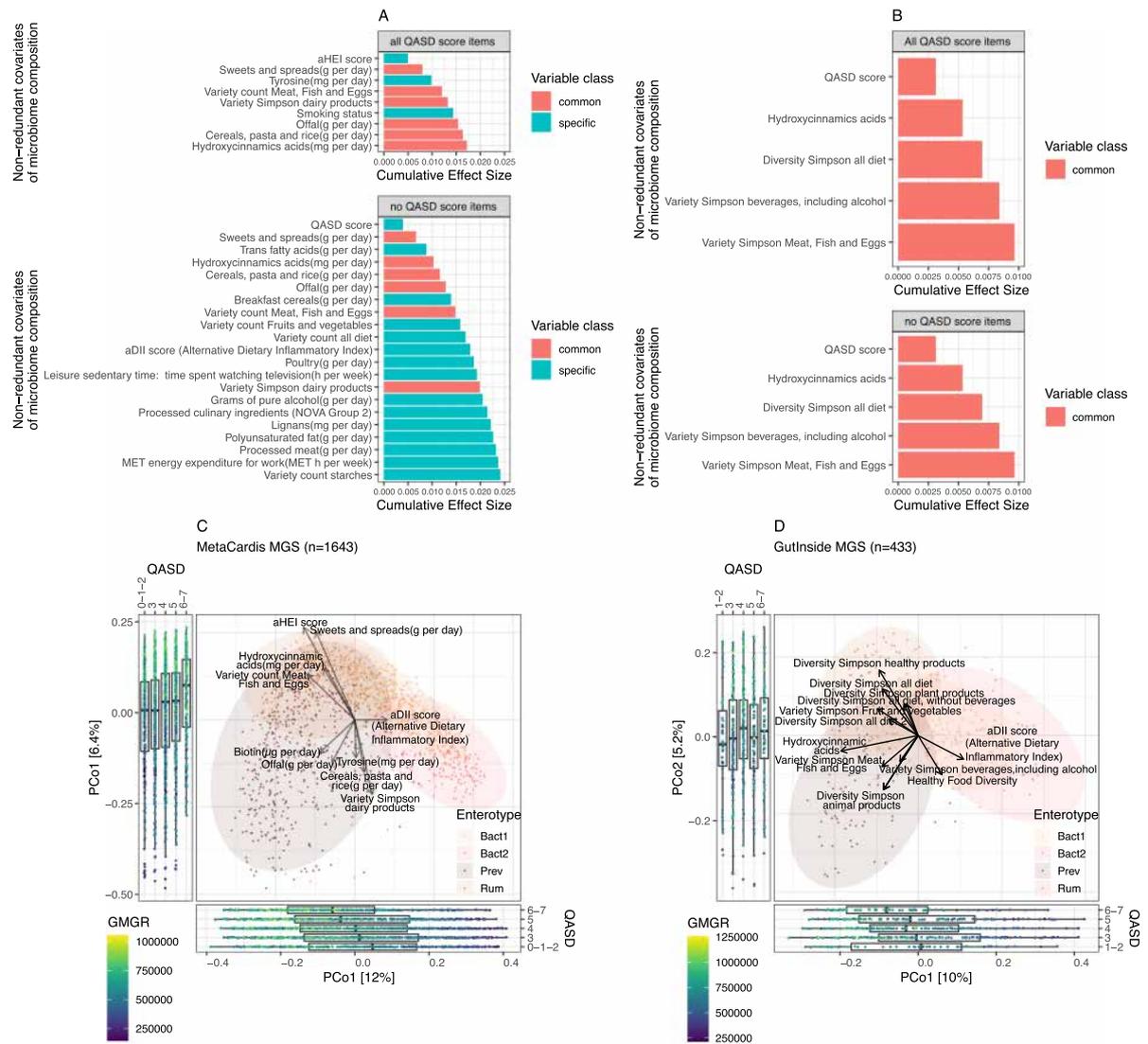
Since GMGR represents only a facet of microbiome composition, we further extended our investigation to deeper evaluate the impact of QASD score together with other nutritional and lifestyle variable on gut microbiome composition both in MetaCardis population and in participants from the GutInside study for confirmation purpose. Distance-based redundancy analyses (dbrDA) identified 47 lifestyle variables significantly associated with microbiome composition, defined as the abundance of metagenomic species (MGS) per individuals (Supplementary Table S5,  $FDR < 0.01$ ). Among these, the AHEI score and the QASD score showed the greatest effect sizes, followed by the intake of sweets and spreads, “hydroxycinnamic acids” (e.g., coffee nutrient), “alcohol” and diet abundance of vitamins such as vitamin D, vitamin A and biotin. A significant association was also found between the inflammatory potential (aDII) of the diet and microbiome composition (Supplementary Table S5). Furthermore, stepwise forward model selection by permutation of these 47 variables provided a model comprising 9 variables with a non-redundant explanatory power for microbiome compositional variation which accounted for 1.72% of the total variance ([Figure 3A](#)). These variables included two components of the QASD score (e.g., “AHEI score” and “Smoking status”), the intake of “Sweets and spreads”, “Tyrosine, Offal”, “Cereals, pasta, rice” and “Hydroxycinnamic acids”, the “Variety count of Meat, Fish Eggs” and the “Variety Simpson dairy products” ([Figure 3A](#), “all QASD score items” panel).

When the 4 individual variables composing the QASD score were excluded from the multivariate approach, the composite QASD score was the lifestyle variable with the highest non-redundant explanatory power in a multivariate model comprising 21 variables that explained 2.41% of the microbiome compositional variability ([Figure 3A](#), “no QASD score items” panel). These findings highlight the added value of a composite such as the QASD score over its individual components in explaining microbiome variability in this population with cardiometabolic disorders. The multivariate model also included variables such as the inflammatory score of the diet and food variety scores and components ([Figure 3A](#), “no QASD score items” panel). These observations were partially replicated in subjects with



**Figure 2.** Mediation analyses of the impact of the QASD score on the gut microbiome and metabolic phenotypes. (A) Schematic representation of the two mediation hypotheses tested to evaluate the impact of the QASD score (independent variable) on the gut microbiome and clinical profiles of 1,643 individuals of MetaCardis study. In Direction 1, the QASD score impacts gut microbiome gene richness (GMGR; dependent variable), which is mediated by changes in clinical profiles (mediator). In Direction 2, the QASD score affects clinical profiles (dependent variable) mediated by alterations in the gut microbial gene richness (GMGR; mediator). The clinical variables tested included BMI, glycated hemoglobin, and HOMA-IR. Pairwise comparisons of the QASD score levels were conducted for low (0–1–2–3) vs. medium (4), low vs. high (5–6–7), and medium vs. high. Mediation analyses were adjusted for recruitment center, age, sex, and the use of metformin, statins, and PPIs (B) Confidence intervals of the beta coefficients representing the total effect (total), average direct effect (ADE), and average causal mediation effect (ACME) of the QASD score on the dependent variable in each direction (GMGR in Direction 1; clinical variables in Direction 2). Color and shape indicate the statistical significance of the mediation effects. Significant mediation in both directions ( $p < 0.05$  for total, ADE, and ACME) was observed only between the upper and lower levels of the QASD score (high vs. low) with glycated hemoglobin and HOMA-IR. (C) Decomposition of the significant mediations observed between high and low levels of the QASD score, presented as a bar plot showing the proportion of the total effect and confidence intervals attributable to the mediator (clinical variable in Direction 1; GMGR in Direction 2). (D) Bar plot representing the  $-\log_{10}$  transformed  $P$ -values ( $x$ -axis) in PERMANOVA-based mediation analyses of the mediation effect of the gut microbiome composition (Bray–Curtis distances derived from MGS abundance profiles), with QASD score as the exposure and clinical variables as outcomes. The dashed line represents the nominal significance level ( $p$ -value = 0.05). (E) Barplots representing the  $F$ -statistic values ( $x$ -axis) for the exposure–microbiome association term and the microbiome–outcome association conditional on the exposure term in PERMANOVA-based mediation analyses ( $* = P$ -value < 0.05).

overweight and obesity from the GutInside study.<sup>31</sup> Here, the QASD score exhibited the strongest association with microbiome compositional variation ( $\text{adj}R^2 = 3.11e-03$ ,  $p = 0.019$ , Supplementary Table S6) and was retained as the variable with the highest non-redundant explanatory power, both with and without its individual components (Figure 3B). These models in the GutInside cohort shared



**Figure 3.** Links between lifestyle factors and microbiome composition in the MetaCardis and GutInside cohorts. (A) Bar plot showing the cumulative effect sizes of multivariate models for non-redundant microbiome compositional variation (adjusted cumulative  $R^2$  values) in individuals from the MetaCardis cohort ( $n = 1,643$ ). The results are presented for models including all individual components of the QASD score (“all QASD score items” panel;  $FDR < 0.01$  in dbRDA analyses and  $p < 0.05$  in stepwise model building,  $n = 9$  variables) and for models excluding the four variables that define the QASD score (“no QASD score items” panel;  $FDR < 0.01$  in dbRDA analyses and  $p < 0.05$  in stepwise model building,  $n = 21$  variables). (B) Bar plot showing cumulative effect sizes of multivariate models for non-redundant microbiome compositional variation (adjusted cumulative  $R^2$  values) in individuals from the GutInside cohort ( $n = 433$ ). The results are shown for models including all individual components of the QASD score (“all QASD score items” panel;  $p < 0.1$  in dbRDA analyses and  $p < 0.05$  in stepwise model building,  $n = 5$  variables) and for models excluding the four QASD score variables (“no QASD score items” panel;  $p < 0.1$  in dbRDA analyses and  $p < 0.05$  in stepwise model building,  $n = 5$  variables). (C) Principal coordinate analysis (PCoA) of inter-individual differences in microbiome profiles (based on Bray–Curtis dissimilarity from MGS abundance data) in the MetaCardis cohort ( $n = 1,643$ ). The arrows in the main panel represent the effect sizes of a post hoc fit of 9 continuous nutritional covariates identified in the multivariate models from panel A. Boxplots represent the distribution of Metacardis individuals at different levels of the QASD score across 1<sup>st</sup> and 2<sup>nd</sup> ordination axis, with points coloured by GMGR. (D) Principal coordinate analysis (PCoA) of inter-individual differences in microbiome profiles (based on Bray–Curtis dissimilarity from MGS abundance data) in the GutInside cohort ( $n = 433$ ). The arrows in the main panel represent effect sizes of a post hoc fit of 12 continuous nutritional covariates identified in the dbRDA analyses ( $p < 0.1$ ). Boxplots represents the distribution of GutInside individuals at different levels of the QASD score across the 1<sup>st</sup> and 2<sup>nd</sup> ordination axis, with points coloured by GMGR. The full results from univariate and multivariate analyses are provided in Supplementary Tables S5 (MetaCardis data) and S6 (GutInside data). The common/specific legends in panels A and B corresponds to variables shared/unshared by models with/without the QASD score items.

variables with those derived from the MetaCardis cohort, including hydroxycinnamic acid intake and consumption of meat, fish, and eggs, captured by different diet diversity metrics (e.g., “Variety count” in MetaCardis, “Simpson diversity” in GutInside; Figure 3A,B). However, among the individual components of the lifestyle score, only the “dietary diversity excluding beverages score” (e.g., “Simpson Diversity Index for all diet, without beverages”) was significantly associated with dbRDA analyses in the GutInside cohort ( $\text{adj}R^2 = 1.14e-03$ ,  $p = 0.048$ ; Supplementary Table S6), although it was not retained in the final multivariate models shown in Figure 3B. Together, these findings underscore that while many individual dietary factors contribute modestly to gut microbiome variation, integrative metrics such as the QASD score may provide a more comprehensive assessment of microbial compositional differences across cardiometabolic populations.

### **The QASD score associates to microbiome enterotype constellations**

We next explored how the QASD score, related variables and other dietary components were associated with enterotypes. In previous studies, 4 distinct microbiome communities (e.g., enterotypes) were identified using Dirichlet Multinomial Mixture Models (DMM).<sup>2,61</sup> These enterotypes correspond to partially overlapping ellipsoids in the PCoA ordination landscape based on Bray-Curtis beta-diversity distances (Figure 3C). Given the prominent dysbiotic character of one enterotype (e.g., Bacteroides 2 or Bact 2) in terms of microbiome diversity and microbial cell density, its strong association with obesity and inflammation<sup>2</sup> and the significant differences in enterotype distribution observed across levels of the QASD score (Figure 1E), we evaluated in depth the association of enterotypes with the QASD score and nutritional/lifestyle variables with a significant impact on the gut microbiome composition. Projecting the 9 lifestyle variables (from Figure 3A) retained in the multivariate model onto this PCoA space revealed that the AHEI score, intake of “sweets and spreads”, “hydroxycinnamic acids” and the “variety count of meat, fish and eggs” were associated with the microbiome compositional space dominated by the Ruminococcus enterotype (Rum, Figure 3C). This enterotype has previously been associated with increased GMGR.<sup>2</sup> A similar pattern was observed for additional variables retained in the multivariate model when the QASD components were excluded, e.g., breakfast cereals, lignans, polyunsaturated fats, dietary variety, starches, fruits and vegetables, and alcohol intake were positively associated with the Rum enterotype. Conversely, the intake of tyrosine, offal, “cereals, pasta and rice” and the variety of dairy products showed associations in the opposite direction in the ordination space (Figure 3C). This observation extended to trans fatty acids, poultry, processed meat and physical activity variables, such as leisure sedentary time and energy expenditure, which were associated with the opposing compositional direction.

Strikingly using our integrative metrics, QASD score levels showed a significant segregation along the 1st and 2nd ordination axes of the PCoA (Figure 3C). The 1st axis, which was strongly anticorrelated with GMGR (Spearman  $Rho = -0.6$ ,  $p < 2.2 \times 10^{-16}$ ), separated individuals with low QASD scores (0–1–2), who clustered at the extreme positive values of the axis corresponding to the Bact2 enterotype region. This enterotype was previously shown to be associated with a lower GMGR, a reduced bacterial load, systemic inflammation and obesity severity.<sup>2,31</sup> In contrast, the second axis, strongly correlated with the GMGR (Spearman  $\rho = 0.7$ ,  $p < 2.2 \times 10^{-16}$ ), was enriched in individuals with high QASD scores (6–7) at its extreme positive values, corresponding to the Ruminococcus enterotype region (Figure 3C). These findings were partially confirmed in the GutInside population, where a high QASD score was associated with the Rum enterotype, while a higher aDII score was associated with the Bact2 enterotype, indicating moderate reproducibility of these associations in individuals with overweight and obesity (Figure 3D).

Logistic regression analyses of enterotype status versus the 47 nutritional covariates retrieved in dbRDA analyses in the MetaCardis population (adjusted for age, sex, BMI, center of recruitment and use of metformin, statins, and PPIs) confirmed the relationships between diet components and enterotype stratification. At the QASD score level, we found a significant decrease in the probability of belonging to the Bact2 group with increased QASD score levels, while no significant association was observed with Ruminococcus status (Supplementary Figure 4A). At the nutritional level, the “variety count of meat, fish and eggs”, the intake of offals, biotin, vitamin D (in food) and “cookies and pastries” were significantly associated with a decrease in the probability of belonging to the Bact2 enterotype ( $p$ -value  $< 0.05$ ;

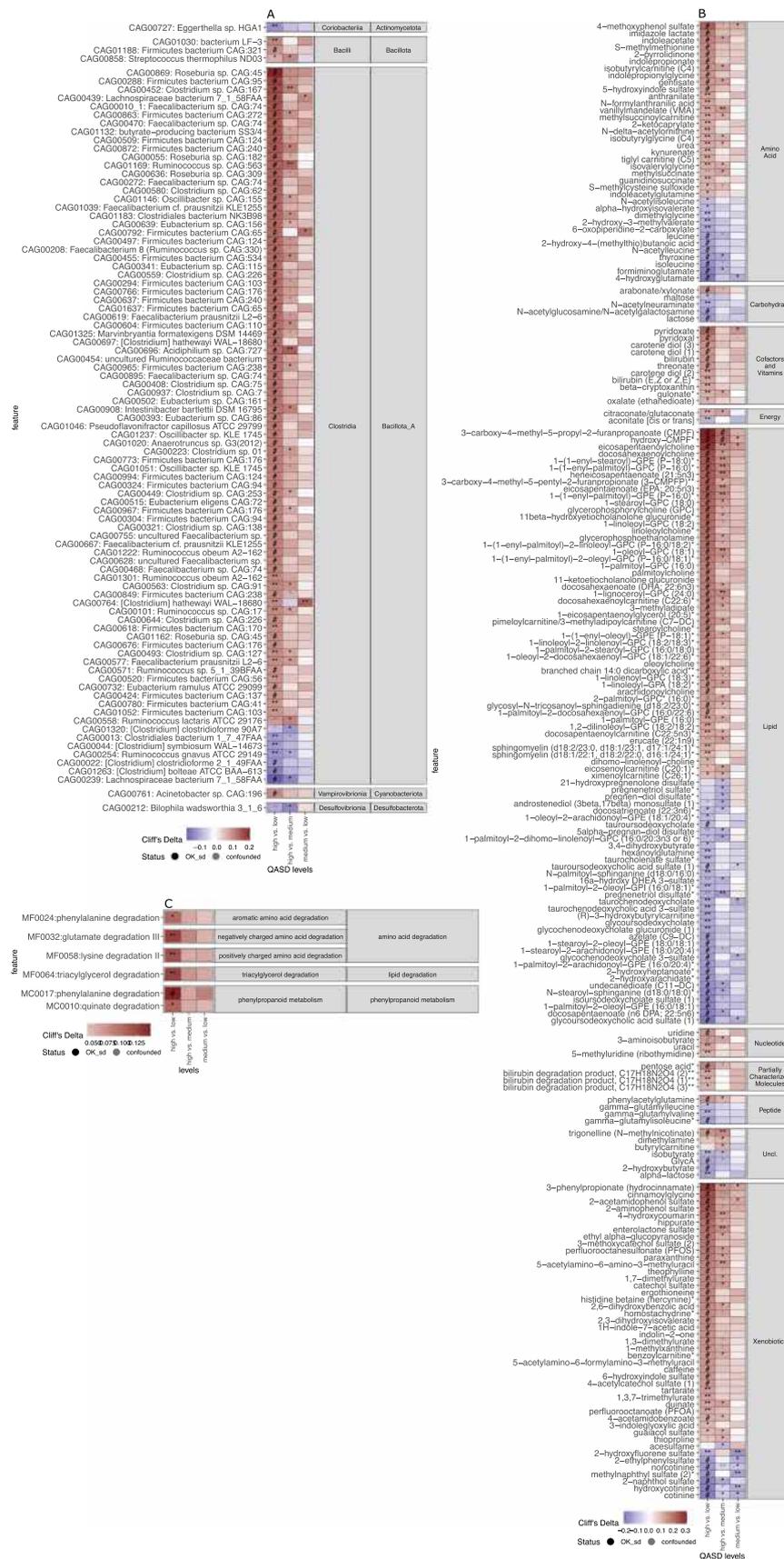
Supplementary Figure 4B–F, Supplementary Results). This again emphasizes on the importance of dietary diversity and variety and not solely on food quality or the perceived healthiness of food. Interestingly, biotin intake in the diet was positively associated with the probability of the Prevotella enterotype in the logistic regression models (Supplementary Figure 4E, Supplementary Results), a community type dominated by bacteria that are auxotrophic for biotin production.<sup>3</sup>

Extended dbRDA analyses incorporating enterotype composition, clinical, biological and lifestyle variables confirmed that, despite the overlaps observed between enterotypes in the PCoA map, enterotypes largely outperformed other covariates in terms of explanatory power of gut microbiome compositional variation (Supplementary Figure 5A). Additionally, this analysis showed that in addition to BMI, fat mass, diabetes status and metformin intake, which explained the greatest variance, the AHEI and the QASD score were the most influential lifestyle variables on microbiome composition variance. The explanatory power of these lifestyle variables surpassed that of medications (e.g., statin intake) and biological variables such as lipid and hepatic blood markers (FDR < 0.01, dbRDA analyses; Supplementary Figure 5A, Supplementary Table S7). Other QASD components (such as smoking status, total physical activity without work (MET h week<sup>-1</sup>)) also emerged as significant contributors in these extended dbRDA analyses (FDR < 0.01; Supplementary Figure 5A). AHEI and smoking status were also retained in a multivariate model comprising 22 nutritional and clinical variables with non-redundant explanatory power, indicating that both the composite lifestyle score and clinical factors exert independent effects on microbiome compositional variation (Supplementary Figure 5B). Finally, associations between the QASD score and the gut microbiome composition were robust when adjusted for key confounders, including age, sex, country, BMI, metformin, statin and PPI intake (Supplementary Figure 5C). These findings highlight the value of considering composite lifestyle metrics, such as the QASD score, to capture relationships between dietary quality and diversity, lifestyle habits and the gut microbiome composition, including considered dysbiotic enterotypes such as Bact2.

### **The QASD score associates with metagenomic species and blood metabolomic signatures**

We next examined more in depth the relationships of the QASD score with metagenomic profiles (taxonomic and functional) and precision phenotypes evaluated by metabolomics. Given the heterogeneous profile of MetaCardis population in terms of clinical and medication profiles, we employed a post hoc filtering approach using the *metadeconfound* R package<sup>29</sup> to account for the potential effects of medications (Metformin, statin, and PPI intake) and clinical status on putative associations of the QASD score with metagenomics and metabolomics measures. Among the examined features, MGS and serum metabolites presented the greatest number of significant and rigorously deconfounded associations, totaling 122 MGS and 319 serum metabolites with an FDR < 0.1. Most of these associations were statistically significant between the upper and lower levels of the QASD score (0–1–2–3, low-QASD, vs. 5–6–7, high-QASD) (Supplementary Figure 6). Only 6 and 11 features were retained in the functional modules (GMM, Figure 4C, discussed in the Supplementary Results) and mOTUs, respectively, while no urine metabolite was retained (Supplementary Figure 6; detailed results in Supplementary Table S8). We thus focused the subsequent analyses on 91 MGS and 272 serum metabolites with absolute effect sizes (Cliff's delta) higher than 0.1.

At the MGS level, higher QASD scores were associated with an enrichment of *Firmicutes* lineages, including *Faecalibacterium prausnitzii* (12 MGS from the *Faecalibacterium* genus and 4 MGS corresponding to different *F. prausnitzii* strains, such as KLE1255 and L2-6), *Eubacterium eligens*, and multiple MGS from the *Oscillibacter* and *Roseburia* genera. In contrast, lower QASD scores were significantly linked to increased levels of *Clostridium bolteae* and *Ruminococcus gnavus* lineages (Figure 4A; Supplementary Results and Supplementary Table S9). We then examined which component of the QASD score captured the largest number of strictly deconfounded significant associations (e.g., using upper and lower levels of tertile decomposition for AHEI, physical activity and Simpson diversity index; smoking status as a binary variable). We found that AHEI and smoking status associate with 97 and 101 MGS, respectively whereas total physical activity and the Simpson diversity index were associated with fewer MGSs, e.g., 17 and 4 MGSs, respectively (Supplementary Figure 7A,B). This finding aligns with the



**Figure 4.** Microbiome and metabolome features associated with the QASD lifestyle score. (A) Heatmap representing the effect sizes (Cliff's delta) of 91 metagenomic species (MGS) showing significant differences between lifestyle score levels (high vs. low; high vs. medium; medium vs. low). Lifestyle score levels are categorized as low (QASD 0, 1, 2, 3), medium (4), and high (5, 6, 7).

(caption on next page)

The effect sizes are strictly deconfounded by metformin, statin, PPI intake, and MetaCardis clinical groups, based on metadeconfoundR results (FDR < 0.1, OK\_sd status meaning strictly deconfounded association, and absolute Cliff's delta > 0.1). Positive values indicate a higher abundance of the MGS at the reference level of the pairwise comparison (e.g., in HvsL, Cliff's delta > 0 denotes higher abundance at the "high" level of the lifestyle score compared to "low", in red; Cliff's delta < 0 denotes lower abundance at the "high" level compared to "low", in blue). The complete list of 122 MGS with significant differences between lifestyle score levels is available in Supplementary Table S8. (B) Same as panel A for 204 annotated serum metabolites with significant differences across lifestyle score levels (FDR < 0.1, OK\_sd status, and absolute Cliff's delta > 0.1). The complete list of 319 metabolites with significant differences across lifestyle score levels is available in Supplementary Table S8. (C) Same as A-B for 6 gut metabolic modules (GMM) with significant differences across lifestyle score levels (FDR < 0.1, OK\_sd status, and absolute Cliff's delta > 0.1). The text labels in heatmaps represent the significance level of the metadeconfoundR results (# = FDR < 0.001, \*\* = FDR < 0.01, \* = FDR < 0.1).



significant, non-redundant impact of AHEI and smoking status on microbiome compositional variation (Figure 3A). Both smoking status and AHEI exhibited similar distributions of MGS changes when Cliff's delta effect sizes were considered. This confirmed that most MGS showing strictly deconfounded associations were enriched in individuals with a high QASD score, high AHEI and non-smokers (Supplementary Figure 7C). Importantly, we also observed a significant positive association between the Cliff's delta effect sizes of MGS changes with the QASD score and with its individual components, meaning that the overall QASD score captures the patterns of MGS variation in their individual components (Supplementary Figure 7D–G).

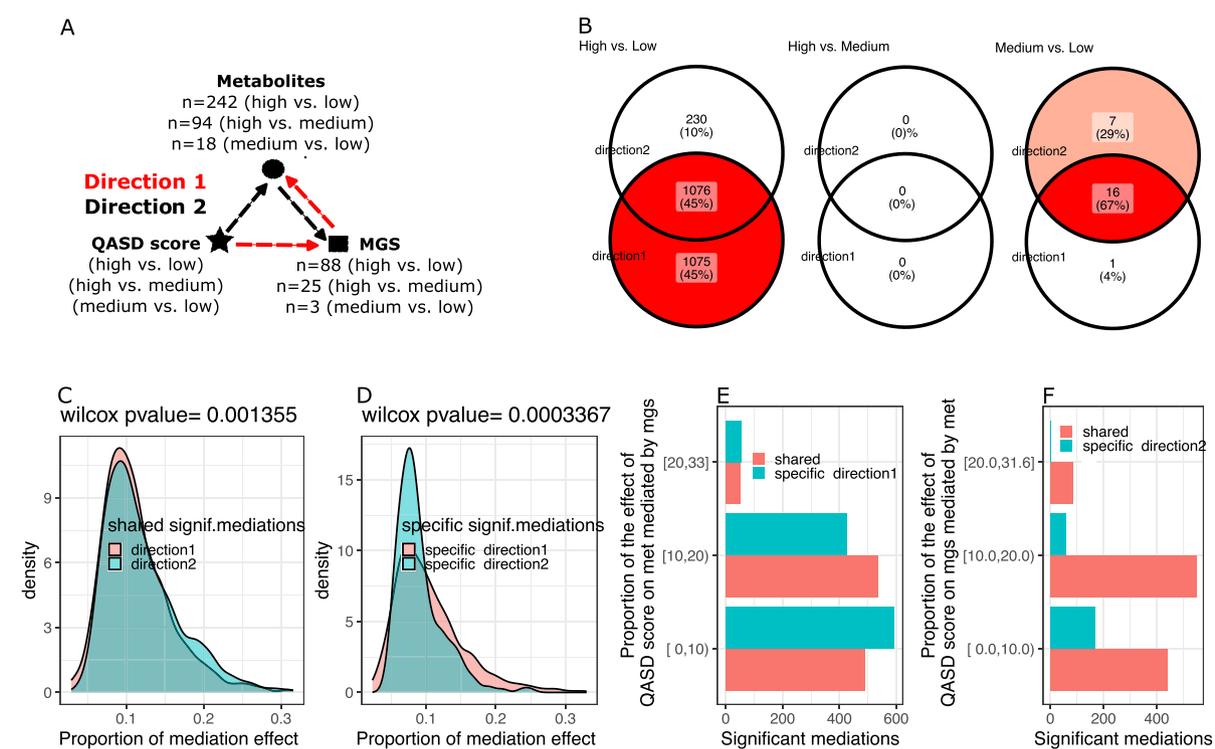
Regarding metabolomic signatures, a high QASD score was positively associated with hippurate, tryptophan-derived compounds, acyl-cholines, coffee-, and pyrimidine-related metabolites. Conversely, lower QASD scores were associated with elevated levels of branched-chain amino acids and their derivatives, bile acid metabolites, markers of tobacco consumption, and dipeptides (Figure 4B, Supplementary Results, and Supplementary Table S10). Deconfounding analyses of serum metabolites for the components of the QASD score showed results consistent with those described for MGS. AHEI and smoking status captured the greatest number of significant associations, with 311 and 377 serum metabolites, respectively, versus 103 and 70 for total physical activity and the Simpson diversity index (Supplementary Figure 7H–I). An increased fraction of specific metabolites was uniquely associated with either AHEI or smoking status (Supplementary Figure 7H–I). AHEI and the QASD score showed similar Cliff's delta effect sizes, whereas smoking status showed a marked negative skew, reflecting enrichment of specific metabolites in smokers (Supplementary Figure 7J). These metabolites (Cliff's delta < -0.5 for non-smokers vs. smokers; FDR < 0.1, strictly deconfounded status) included cotinine, hydroxycotinine and norcotinine, which are indicative of tobacco use. Notably, they were also identified through the QASD score, albeit with smaller effect sizes (Cliff's delta ranging from -0.25 to -0.1; FDR < 0.1; Supplementary Figure 7J,N). Finally, we observed a significant positive association between the Cliff's delta effect sizes of metabolite changes with the QASD score and with its individual components except the Simpson diversity index, indicating that the QASD score overall captures the patterns of metabolite variation in their individual components (Supplementary Figure 7K–N). Overall, this analysis identified distinct MGS and metabolic species associated with high versus low QASD scores, with the QASD score capturing a substantial proportion of metagenomic and serum metabolomic biomarkers of its individual components with comparable effect sizes.

### ***Gut microbiome strongly mediates the effect of the QASD score on serum metabolomics profiles***

Finally, to integrate deconfounded individual MGS and serum metabolome biomarkers associated with the QASD score, we performed extended mediation analyses to assess tripartite relationships, testing whether the metagenome mediates the effect of the QASD score on serum metabolomic profiles, and conversely, whether serum metabolites mediate the impact of the QASD score on MGS abundances. We focused on 91 MGS and 272 metabolomic features that were robustly associated with the QASD score after a strict deconfounding approach (FDR < 0.1), and we demonstrated a substantial effect size (absolute Cliff's delta effect size > 0.1). We tested two primary mediation hypotheses. The first hypothesis proposed that MGS

mediates the relationship between the QASD score and the host's serum metabolome, indicating a microbiome-driven mediation pathway (Direction 1, Figure 5A). The second hypothesis explored an alternative scenario in which the host metabolome mediates the effect of the QASD score on the gut metagenomic landscape (Direction 2, Figure 5A). These hypotheses were tested across all deconfounded MGS-serum metabolite pairs stratified by the QASD score levels (high vs. low, high vs. medium, medium vs. low). importantly, for inclusion in this analysis, each MGS-serum metabolite pair exhibited significant reciprocal associations (FDR < 0.05, linear regression adjusted for center of recruitment, age, sex, and intake of metformin, statins, and PPIs).

Whereas bidirectionality was present, our findings strongly supported the preminence of the first hypothesis. Considering the high and low levels of the QASD score, there were 2,151 significant mediations in Direction 1 and 1,306 in Direction 2. Of these, 1,075 mediations (45.14% of the total) were specific to Direction 1, while 230 (9.65% of the total) were specific to Direction 2, and there were 1,076 significant mediations (45.19% of the total) in both directions (Figure 5B, Supplementary Table S11, Supplementary Figure 8). For intermediate QASD score levels, no significant mediations were observed for high vs.



**Figure 5.** Summary of bidirectional mediation analyses between the QASD score, metabolites, and metagenomic species (MGS). (A) Schematic representation of the two tested directions for causal mediation driven by the QASD score (independent variable). In Direction 1, the QASD score affects serum metabolites (outcome variable) mediated by the abundance of MGS (mediator). In Direction 2, the QASD score impacts the MGS (outcome variable) mediated by serum metabolites (mediator). Mediation tests were conducted for all pairwise combinations of MGS and serum metabolites that showed significant variations between the QASD score levels in the drug deconfounded analyses. All the mediation analyses were adjusted for recruitment center, age, sex, and the use of metformin, statins, and PPIs. (B) Venn diagrams summarizing the number of significant mediations (FDR < 0.05 for average causal mediation effect (ACME), average direct effect (ADE), and total effect) found in each direction across pairwise QASD score levels. (C) Density plots showing the proportion of mediation effects in shared significant mediations for Direction 1 (MGS as a mediator; serum metabolites as outcomes) and Direction 2 (serum metabolites as mediators; MGS as outcomes). (D) Density plots illustrating the proportion of mediation effects in specific significant mediations for Direction 1 and Direction 2. (E) Decomposition of the 2151 significant mediation relationships in Direction 1 (QASD → MGS → metabolite) observed between high and low QASD score levels, based on the proportion of the total effect of the QASD score on serum metabolites mediated by MGS abundances (y-axis). (F) Decomposition of the 1306 significant mediation relationships in Direction 2 (QASD → Metabolite → MGS) observed between high and low QASD score levels based on the proportion of the total effect of the QASD score on MGS abundances mediated by serum metabolite levels (y-axis).

medium levels whereas only 40 significant mediations were found for medium vs. low levels, with 16 (40%) significant mediations in both directions (Figure 5B). This suggests that high and low levels of the QASD score have the most substantial impact on an individual's metagenomic and metabolomic profiles.

Focusing on high and low levels of the QASD score, we compared both directions in terms of the proportion of the mediation effect on the tested outcome variables. Significant mediations ranged from 2.28% to 34.03% of the total effect in Direction 1 (e.g., microbiome as a mediator) and from 3.89% to 31.5% in Direction 2 (metabolome as a mediator), revealing a significant difference between shared mediations despite overlapping distributions ( $p = 1.355e-03$ , Wilcoxon rank-sum test; Figure 5C). More importantly, when direction-specific mediations were examined, the mediation effect was significantly greater in Direction 1 ( $p = 3.367e-04$ , Wilcoxon rank-sum test; Figure 5D), indicating that both the number and strength of specific mediations are higher when the microbiome acts as the mediator.

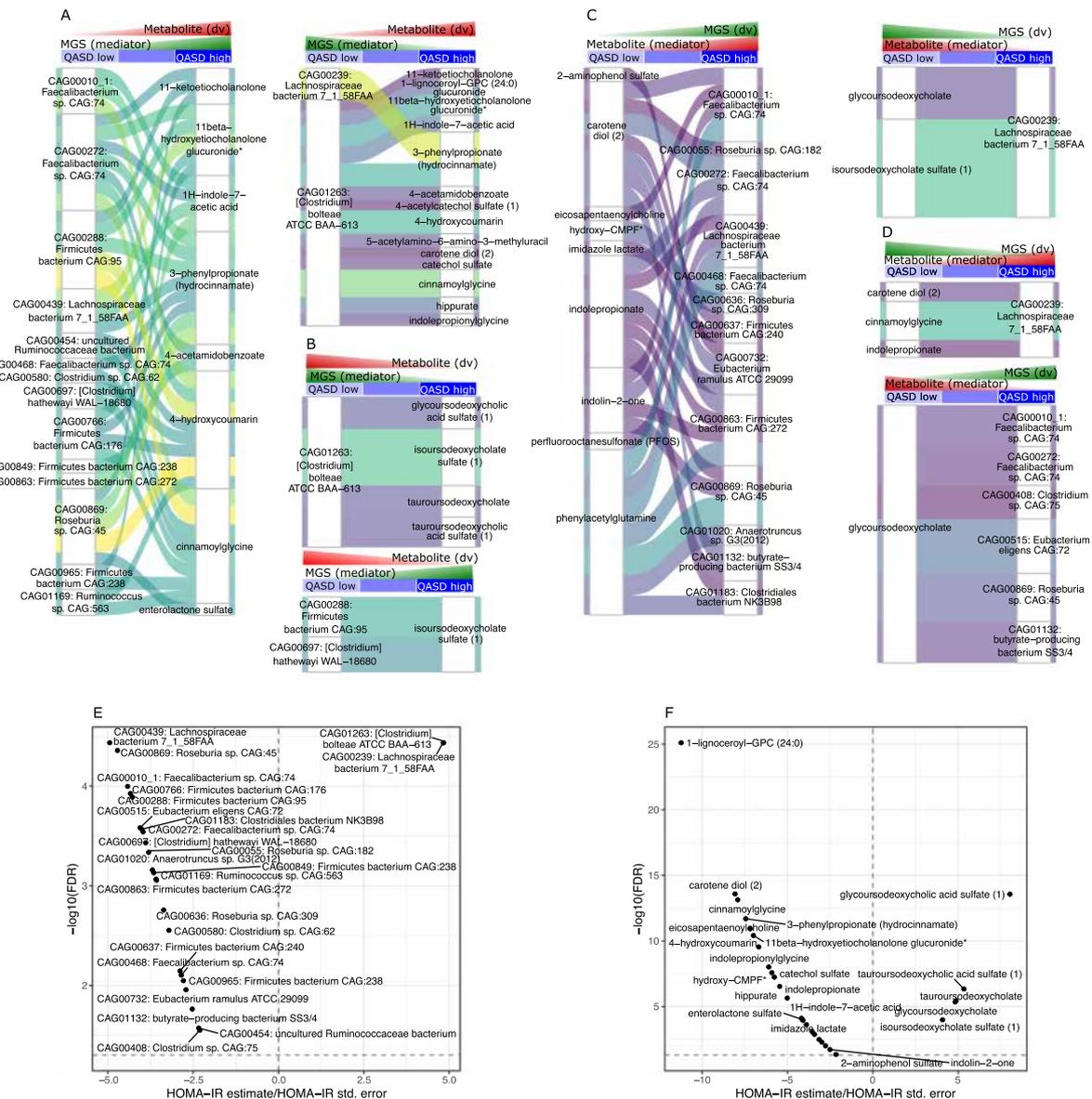
This pattern was also observed when we stratified significant mediations by their strength in terms of the proportion of the mediating effect. We identified 106 and 88 high-impact mediations in each directions 1 and 2, respectively, where the mediator (MGS or serum metabolites, respectively) explained more than 20% of the effect of the QASD score on the exposure variable. Among these, 52.83% of the high-impact mediations in Direction 1 were specific (56/106), compared to 3.4% in Direction 2 (3/88). Similar trends were observed for medium-impact (10%–20%) and low-impact (<10%) mediations (Figure 5E–F).

Given the large number of significant mediations observed and considering the mediating role of the gut microbiome on the QASD-HOMA-IR relationship described earlier (Figure 2), we next focused on the detailed visualization of the strongest significant mediations where the MGS and the serum metabolite were both significantly associated with HOMA-IR (FDR < 0.05, linear regression analyses under the same adjustment framework as mediation analyses). Focusing on the strongest specific mediations in Direction 1 involving annotated serum metabolites (as exposures), we found 47 high-impact mediations (mediation effect > 20%) that involved both positive and negative relationships depending on the effect of the MGS on serum metabolite levels. This causal inference exploration reflects the complex network of interactions between gut microbiome species that may influence the insulin resistance-associated metabolome profile. The strongest mediations were observed in individuals with a high QASD score and involved serum metabolites such as cinnamoylglycine and 3-phenylpropionate, which were previously identified as biomarkers associated with high microbiome diversity.<sup>74,75</sup> Higher levels of these metabolites in individuals with high QASD scores were strongly mediated by increases in several Firmicutes lineages, including *Faecalibacterium* and *Roseburia* (Figure 6A, left panel; Supplementary Table S11). Conversely, these same metabolites were also mediated by decreased levels of *Clostridium bolteae* (27.46% mediation for cinnamoylglycine) and *Lachnospiraceae bacterium 7\_1\_58FAA* (31.37% mediation for 3-phenylpropionate), two MGS enriched in individuals with low QASD scores (Figure 6A, right panel; Supplementary Table S11).

For individuals with low QASD scores, the strongest positive mediations involved increased levels of *Clostridium bolteae*, which significantly mediated the rise of four secondary bile acids, particularly isoursodeoxycholate sulfate (25.2% mediation effect; Figure 6B, top panel). This increase in isoursodeoxycholate sulfate was also mediated by decreased levels of two Firmicutes MGS (*CAG00697: [Clostridium] hathewayi* WAL-18680, 20.12% mediation effect; *CAG00288: Firmicutes bacterium CAG:95*, 22.85% mediation effect; Figure 6B, bottom panel; Supplementary Table S11).

In Direction 2, only 3 high-impact mediations (mediation effect > 20%) were observed. These included phenylacetylglutamine, whose rise in individuals with a high QASD score mediated the rise of *CAG00869: Roseburia sp. CAG:45* (20.99% mediation effect; Figure 6C, left panel), and the *CAG00239: Lachnospiraceae bacterium 7\_1\_58FAA*, whose decrease in individuals with a high QASD score was mediated by decreases in the secondary bile acid isoursodeoxycholate sulfate (24.95% mediation effect; Figure 6C, right panel) and increases in cinnamoylglycine (23.82% mediation effect; Figure 6D, top panel).

Extending the analyses to medium-impact mediations (mediation effect > 10%), mediations associated with high QASD scores included environmental exposures like food-derived carotenoids (carotene diol),<sup>76</sup> which mediated increases in eight MGS ( $12.86\% \pm 1.74\%$  mean  $\pm$  standard deviation mediation effect), or Indolin-2-one, known for its antimicrobial activity,<sup>77</sup> which mediated increases in four MGS ( $11.84\% \pm 1.1\%$  mean  $\pm$  standard deviation mediation effect), while perfluorooctanesulfonate, an environmental pollutant,<sup>78</sup> mediated 12.37% of the increases in *CAG00439: Lachnospiraceae bacterium 7\_1\_58FAA* (Figure 6C, left panel). Additionally, metabolites derived from microbial metabolism, such



**Figure 6.** Overview of the strongest mediations between the QASD score-MGS-Serum metabolites in association with the HOMA-IR phenotype (A, B). Alluvial diagrams representing the 60 strongest significant mediation relationships (MGS-serum metabolites) in Direction 1 between high and low QASD score levels, categorized by the sign of the beta coefficients (increases/decreases in the mediator (MGS) and dependent variable (dv, serum metabolite) between high and low levels of the QASD score). (C, D) Alluvial diagrams representing the 41 strongest significant mediation relationships in Direction 2 between high and low QASD score levels, categorized by the sign of the beta coefficients (increases/decreases in the mediator (serum metabolite) and dependent variable (dv, MGS) between high and low levels of the QASD score). (E) Volcano plot representing the results of the linear regression analyses vs. HOMA-IR (dependent variable) of the 25 MGS included in the mediations represented in panels A–D. Dashed horizontal line represents the threshold for a significant association in multiple testing ( $\text{FDR} < 0.05$  on 91 tested MGS). (F) Similar as E panel for the 28 serum metabolites included in mediations of panels A–D. Full results of mediation analysis and regressions vs. HOMA-IR are available in Supplementary Table S11.

as the aforementioned phenylacetylglutamine (a human metabolite synthesized from phenylalanine degradation),<sup>79</sup> mediated increases in seven MGS ( $17.18\% \pm 2.58\%$  mean  $\pm$  standard deviation mediation effect), and indolepropionate (a tryptophan-derived metabolite)-mediated increases in eight MGS enriched in individuals with high QASD scores ( $11.51\% \pm 1.37\%$  mean  $\pm$  standard deviation mediation effect; Figure 6C, left panel, 6D, top panel). These findings suggest that microbially-derived serum metabolites

may play a mediating role in shaping gut microbial communities across the upper and lower ends of the QASD score.

This mediating role was also observed with other secondary bile acids, such as glyoursodeoxycholate, which, in individuals with low QASD scores, mediated increases in *CAG00239: Lachnospiraceae bacterium 7\_1\_58FAA* (13.33% mediation effect, [Figure 6C](#), right panel) and decreases in six Firmicutes lineages ( $13.07\% \pm 2.21\%$  mean  $\pm$  standard deviation mediation effect; [Figure 6D](#), bottom panel, Supplementary Table S11).

In addition, although the MGS and serum metabolites involved in the mediations described above were all significantly associated with HOMA-IR, we also identified significant mediations driven by environmental exposures such as tobacco consumption. For example, cotinine, enriched in individuals with low QASD scores, significantly mediated decreases in several Firmicutes lineages (e.g., *CAG00055: Roseburia sp. CAG:182* and *CAG00010\_1: Faecalibacterium sp. CAG:74*), despite cotinine itself not being significantly associated with HOMA-IR (mediation effects of 14.08% and 20.51%, respectively; Supplementary Table S11).

Finally, when we examined the linear regression results for the MGS and serum metabolites involved in these mediations, we found a clear pattern: secondary bile acids and MGS that were increased in individuals with low QASD scores (such as *CAG01263: [Clostridium] bolteae ATCC BAA-613* and *CAG00239: Lachnospiraceae bacterium 7\_1\_58FAA*) were positively associated with HOMA-IR. In contrast, other serum metabolites and MGS that were increased in individuals with high QASD scores were negatively associated with HOMA-IR ([Figure 6E,F](#)).

These findings highlight a predominant microbiome-to-metabolome mediation axis, particularly involving Firmicutes lineages, driven by the nutritional and lifestyle factors summarized in the QASD score, that could shape insulin resistance-associated profiles. While bidirectionality exists, the mediating effects were stronger when the microbiome acted as a mediator, and strong mediatory links, such as the rise of secondary bile acids, emerged in individuals with low QASD scores mediated by *CAG01263: [Clostridium] bolteae ATCC BAA-613*. These links underscore the ability of the QASD score to capture nutritional and environmental signatures that impact host phenotypes through alterations in the gut microbiome composition with links to the insulin resistance profiles of the host.

## Discussion

We herein investigated how integrated lifestyle behaviors, captured by a composite lifestyle score (QASD), are related to gut microbiome gene richness and composition (including enterotype stratification) and host metabolic phenotypes in healthy individuals and those with cardiometabolic disorders. Whereas bidirectional interactions are evident, our findings highlight a significant mediating role for the GMGR in the association between QASD and glucose metabolism markers (HbA1c and HOMA-IR). By applying extended causal mediation analyses, we confirmed that the gut microbiome is a significant mediator in the impact of the QASD score on host phenotypes in terms of insulin resistance as well as serum metabolomic profiles. Accordingly, by accounting for different directions in mediation analyses, we identified interplays between insulin resistance-associated bacterial species and serum metabolites, with putative beneficial or detrimental effects on cardiometabolic health.

We first observed that GMGR is not only associated with a large variety of individual food items, macronutrients and micronutrients including polyphenols and vitamins, but also with broader dietary patterns, measures of physical activity and smoking behavior. On the dietary side, we showed that associations between GMGR and dietary diversity (evaluated by different metrics) can be independent of dietary quality. Previous observations in healthy populations such as PREDICT and Flemish Gut cohorts,<sup>6,7</sup> also showed that dietary diversity is associated with the GMGR. Our results are also in agreement with recent findings that omnivorous diets – more diverse than vegetarian or vegan diets – are associated with higher microbial diversity.<sup>15</sup> Moreover, our identified positive associations between physical activity items and the GMGR were in line with observations made in healthier populations recruited in the US Arrivale cohort.<sup>80</sup> Overall, our findings demonstrate that, in individuals with a spectrum of cardiometabolic disorders, greater GMGR is significantly associated with increased dietary

diversity, even when this diversity includes less healthy food items such as alcohol. The polyphenolic compounds present in alcoholic beverages such as red wine could contribute to these associations, as shown in previous studies.<sup>81</sup> We should also consider that patients included in the MetaCardis study reported globally moderate alcohol consumption, with high alcohol consumption being a criteria of exclusion.

Finally, even if high GMGR is generally considered a proxy for gut health, this is not the case in certain contexts, where high microbiome diversity could be associated with non-beneficial exposures.<sup>82,83</sup> Our results underscore the clinical importance of distinguishing between the diversity and quality of food consumption, as greater food variety, if generally recommended, does not inherently reflect healthy eating patterns or more favorable microbiome configurations in populations with pathologies.

Our observations extend beyond the GMGR. Combining lifestyle variables most strongly associated with the GMGR, which include, for example, alcohol as part of the AHEI, demonstrated a greater impact on microbiome composition across the MetaCardis population and pathology subgroups, compared to its individual components. Nevertheless, the effect size of individual dietary or lifestyle factors on gut microbiome compositional variation remains limited, a trend similarly observed in many other populations<sup>84-86</sup>. The four QASD components (e.g., dietary quality and diversity, physical activity levels and smoking status) appear to exert partially non-redundant effects. When we examined the most influential lifestyle variables that had non-redundant associations with microbiome composition, we found that they collectively accounted for less than 1.8% of the microbiome's variance. When clinical covariates were included, the explained variance moderately increased (approximately 2.5%), with AHEI and smoking status identified as key non-redundant contributors. These findings highlight, in individuals with cardiometabolic diseases, the multifactorial nature of microbiome variation shaped by complex interactions among lifestyle behaviors, diet, clinical status, and potentially genetic and other environmental influences not captured in this analysis.<sup>6,87,88</sup>

We described that a low QASD and a high dietary inflammatory index (aDII) were linked to the Bact2 enterotype both in MetaCardis and the GutInside validation cohort. While the concept of enterotypes, which was originally defined based on beta-diversity, has been subject to debate since its introduction,<sup>61,89,90</sup> a growing scientific consensus recognizes that enterotypes represent compositional archetypes within a continuous gradient of gut microbiome community structures.<sup>91</sup> Specific bacterial constellation, such as the Bacteroides-dominated, low-diversity Bact2 enterotype is of particular interest since it has been linked to higher BMI and elevated inflammatory markers in different cohorts, including MetaCardis.<sup>2,31</sup> Here, a combination of diet items, patterns and lifestyle features explained some of these associations in the multivariate models. Interestingly, the association between increased intake of specific foods, such as offals, biotin- and vitamin D-rich items, and even cookies and pastries, and reduced Bact2 prevalence suggests that dietary diversity may help modulate dysbiotic microbiome profiles. In line with this hypothesis, in mouse experiments, in the context of diet-induced obesity, that the combined supplementation of biotin with fructooligosaccharide leads to a significant improvement of metabolic health phenotype and gut microbiome diversity.<sup>3</sup> Our observational findings should nevertheless be interpreted with caution in humans, as most dietary components account for a small fraction of microbiome variance. This underscores the need for tailored, multi-component interventions to evaluate their clinical impact and determine whether putative effects are mediated through microbiome-related mechanisms.

We further explored the metagenomic and host metabolomic signatures and found significant associations between the QASD score and 91 identified MGSs and 272 serum metabolites. Many of the QASD-based associations are consistent with previously reported health-status associations (discussed in depth in the Supplementary Results and Supplementary Tables 9 and 10). We found that MGS enriched in individuals with high QASD scores belonged to bacterial groups such as *F. prausnitzii*, *E. eligens*, *Oscillibacter*, or *Roseburia* lineages previously described with health-promoting effects and found to be depleted in inflammatory diseases,<sup>7</sup> emphasizing the contribution of lifestyle patterns to these bacterial modulations. Conversely, lower QASD scores were associated with MGS previously linked to systemic and postprandial inflammation and deteriorated health, such as *C. bolteae* and *R. gnavus*.<sup>7</sup> The QASD score captured metabolomic signatures of clinical interest, including higher levels of microbiota- and diet-derived metabolites such as hippurate, acyl-cholines, and tryptophan-related compounds, while lower QASD scores were associated with branched-chain amino acids and tobacco-related metabolites. Given the

relevant associations with microbial species and metabolites consistently linked to health-promoting or disease-associated states, these results stimulated explorations of the complex interplay between the QASD score, gut microbiome composition, and metabolic phenotypes.

Interventional studies on personalized diets have indeed demonstrated that gut microbial pathways and species, including subgroups of *Faecalibacterium prausnitzii*, can mediate the effects of various dietary components on metabolic health.<sup>92,93</sup> Research from the LifeLinesDEEP cohort highlighted the mediating role of microbial vitamin B1 and B2 production in the relationship between fruit intake and diabetes risk.<sup>94,95</sup> Findings indicate that the gut microbiome modulates the beneficial association between adherence to a Mediterranean diet and cardiometabolic risk, using prospective data from 307 participants.<sup>96</sup> Moreover, microbiome diversity and composition were also described as mediators of the beneficial effects of a green Mediterranean diet on reducing cardiometabolic risk.<sup>97</sup>

Our mediation analyses, in this cross-sectional design, demonstrated that interactions between the gut microbiome and host metabolism are bidirectional but strongly support a prominent pathway in which lifestyle factors influence the microbiome, which in turn mediates effects on metabolic phenotypes. First, the GMGR significantly mediated the association between the QASD score and clinically relevant glucose metabolism markers, e.g., HbA1C and HOMA1R, while BMI showed no significant mediation relationship. Clinical variables linked to glucose metabolism also mediated the link between lifestyle and the GMGR but with lower mediation effects. These results align with previous evidence that the gut microbiome modulates host glucose homeostasis rather than adiposity.<sup>98</sup>

Our larger-scale mediation analyses expand upon this initial observation linking the GMGR to routine clinical markers by demonstrating, at a systems level, that the gut microbiome significantly mediates the effects of lifestyle on host metabolomic profiles. This was striking in the QASD-associated MGS–metabolite pairs, where the microbiome-to-metabolome direction yielded both a higher number of significant mediations overall (2,151 vs. 1,305 mediations) and direction-specific (1075 vs. 230 mediations) together with larger effect sizes compared to the reverse direction. These results provide strong evidence that specific gut microbial species play pivotal and eventually directional role in translating lifestyle patterns into systemic metabolomic phenotypes. Actually, *Faecalibacterium* and *Roseburia* lineages (e.g., associated with high QASD), were key microbial mediators, found to mediate increases in blood metabolites such as cinnamoylglycine and 3-phenylpropionate, both biomarkers of high microbiome diversity listed in the Microbial Metabolome Database (MiMeDB).<sup>99</sup> Cinnamoylglycine, in particular, has been shown to have a microbial origin in mouse studies, suggesting a need for future mechanistic research.<sup>100</sup> Conversely, *Clostridium bolteae* emerged as a key mediator of increased secondary bile acids, such as isoursodeoxycholate sulfate, in individuals with low QASD scores. This serum biomarker has been linked to postprandial lipemia, inflammatory diseases and impaired liver function.<sup>101</sup> In support of a mechanistic interpretation, genome-scale metabolic models of *C. bolteae* from the AGORA2 repository (including the ATCC\_BAA\_613 strain) contain bile salt hydrolases that deconjugate primary bile acids, producing the isoursodeoxycholate precursor ursodeoxycholate.<sup>102,103</sup> In contrast, we found that serum metabolites reflecting environmental exposures, such as cotinine (a marker of tobacco consumption) or carotene diol (food-derived carotenoids), significantly mediated the impact of the QASD score on gut microbial species. This mediation involved metabolites of microbial origin, including indolepropionate (from microbial degradation of dietary tryptophan), phenylacetylglutamine (from dietary phenylalanine), and secondary bile acids such as glyoursodeoxycholate and isoursodeoxycholate sulfate. These findings reinforce the fact that gut microbiome activity is modulated by metabolic responses captured through the serum metabolome, again supporting the concept of bidirectional mediation. Importantly, MGS and serum metabolites highlighted in these mediations were significantly associated with insulin resistance, revealing a robust enrichment pattern of secondary bile acids in individuals with low QASD scores and high HOMA-IR. This was particularly evident for *CAG01263: [Clostridium] bolteae ATCC BAA-613*, which mediated the abundance of secondary bile acids, and *CAG00239: Lachnospiraceae bacterium 7\_1\_58FAA* also linked to metabolites associated with insulin resistance. Notably, the enrichment of secondary bile acids in individuals with insulin resistance and T2D has recently been reported in two Swedish cohorts,<sup>104</sup> further supporting these findings in the MetaCardis population.

This study has several limitations, including its cross-sectional design and reliance on FFQs to assess long-term dietary patterns, which may overlook short-term intake variability. Future studies should

consider incorporating 24-h recalls to capture short-term dynamic dietary fluctuations. The QASD score, while useful as a composite index reflecting overall lifestyle quality, may mask specific mediation effects driven by individual nutrients – such as tryptophan or phenylalanine – on metabolite levels via distinct microbial species. Previous work has shown both microbiome-mediated effects of diet on metabolites (e.g., *Ruminococcus*-mediated impact of fruit intake on urolithin B) and reverse pathways (e.g., plasma hippurate mediates the effect of coffee on *Methanobrevibacter smithii*).<sup>87</sup>

The statistical strength of the associations of the QASD score with the GMGR and gut microbiome composition observed in the MetaCardis population decreases in our validation cohort (GutInside). This could be explained by differences in cohort characteristics: GutInside is a more homogeneous population (subjects with overweight and moderate obesity from France) compared with MetaCardis, which spans a broader range of geographic origins and cardiometabolic conditions. The smaller sample size of GutInside ( $n = 433$  vs.  $n = 1643$  in MetaCardis) reduces the statistical power to detect associations of similar magnitude. Because the QASD score is derived from population-specific tertile distributions, part of the differences observed between cohorts (e.g., MetaCardis vs. GutInside) may reflect underlying variation in lifestyle component distributions. In this context, MetaCardis could serve as a reference population to calibrate QASD score categories in smaller or less diverse cohorts, but the reduced association in the GutInside population also highlights the importance of cohort context when interpreting lifestyle–microbiome associations and underlines the need for future studies in larger and more diverse populations. Finally, it remains challenging to fully disentangle whether the observed associations between QASD and microbial species or blood metabolites are direct, or instead secondary to gut microbial gene richness (GMGR) or host health status. These relationships are inherently complex, involving multiple potential interaction pathways, which may confound causal inference. In this context, while mediation analysis is a powerful approach for exploring pathways linking lifestyle, microbiota, and metabolism, it inevitably simplifies the complex, bidirectional nature of these interactions and cannot establish definitive causality.

In conclusion, our study highlights the complex and bidirectional influence of lifestyle factors on the gut microbiome and host metabolic phenotype in individuals with cardiometabolic disorders. Through causal mediation analysis, we identified insulin resistance-associated microbial taxa and metabolites, such as *Faecalibacterium prausnitzii*, *Clostridium bolteae*, secondary bile acids, and indolepropionate, as key mediators linking the composite lifestyle score (QASD) to both host metabolomic profiles and the gut microbiome composition. These findings pave the way not only for mechanistic pathway exploration but also for potential biomarkers for identifying at-risk individuals and tailoring interventions. We propose the QASD score as an integrative composite tool to assess lifestyle–microbiome–host interactions and guide personalized nutritional or behavioral strategies. Ultimately, the proposed mediation framework offers a path toward prevention and monitoring of metabolic health in lifestyle intervention.

#### MetaCardis consortium collaborators

The MetaCardis consortium †: Renato Alves, Ehm Astrid Andersson Galijatovic, Olivier Barthelemy, Jean-Philippe Bastard, Jean-Paul Batisse, Randa Bittar, Matthias Blüher, Frederic Bosquet, Olivier Bourron, Mickael Camus, Cecile Ciangura, Arne Dietrich, Morad Djebbar, Angélique Doré, Line Engelbrechtsen, Leopold Fezeu, Sebastien Fromentin, Nathalie Galleron, Marianne Graine, Caroline Grünemann, Agnes Hartemann, Bolette Hartmann, Gerard Helft, Malene Hornbak, Charlotte Kranepuhl, Julien Chilloux, Antonis Myridakis, Lesley Hoyles, Richard Isnard, Sophie Jaqueminet, Niklas Rye Jørgensen, Hanna Julienne, Johanne Justesen, Judith Kammer, Nikolaj Karup, Lars Køber, Michael Kuhn, Véronique Lejard, Ivica Letunic, Florence Levenez, Lajos Marko, Laura Martinez-Gili, Nicolas Maziers, Lucas Moitinho-Silva, Gilles Montalescot, Ana Luisa Neves, Michael Olanipekun, Laetitia Pasero Le Pavin, Luis Pedro Coelho, Nicolas Pons, Benoit Quinquis, Andrea Rodriguez-Martinez, Hugo Roume, Sebastien Schmidt, Tatjana Schütz, Johanne Silvain, Mathilde Svendstrup, Thierry Vanduyvenboden, Stefanie Walther.

#### Disclosure statement

No potential conflict of interest was reported by the author(s).

## Acknowledgments

The authors would like to thank the MetaCardis volunteers for their participation in the study and the patient associations (Alliance du Coeur and CNAO) for their valuable input and support. We are grateful to the nurses, technicians, clinical research assistants, and data managers from the Clinical Investigation Platforms for their efforts in patient investigations, as well as the Clinical Investigation Center (CIC) at Pitié-Salpêtrière Hospital for investigating healthy controls. We acknowledge Quanta Medical for providing regulatory oversight of the clinical study and for their contributions to the processing and management of the electronic data. A pre-print of the manuscript has been deposited on MedRxiv (<https://www.medrxiv.org/content/10.1101/2024.01.12.24301195v2.full.pdf>).

## Funding

This work was supported by the European Union's Seventh Framework Programme for research, technological development, and demonstration under grant agreement HEALTH-F4-2012-305312 (METACARDIS). Additional funding was provided by the Leducq Foundation (17CVD01) and Horizon Europe, EIC-pathfinder "Nutrimune" (Grant agreement ID: 101162457). Part of the work was supported by grants from the Deutsche Forschungsgemeinschaft (DFG): SFB 1052 (project B1), the Fondation pour la Recherche Médicale (FDT201904008276, FDT202106012793), and the French Agency of Research (ANR-NUTRIM\_CHECK, ANR-DeepIntegromics). SKF is supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation; SFB1365 and SFB1470) and the Deutsches Zentrum für Herz-Kreislauf-Forschung (DZHK). K. Chechi acknowledges support from a Medical Research Council Skills Development Fellowship (MR/S020039/1) and a Wellcome Trust-funded Institutional Strategic Support Fellowship (204834/Z/16/Z). PA is funded by a Medical Research Council Career Development Award (MR/Y010051/1). MED acknowledges funding support from the EU IMMEDIATE consortium (101095540) and UKRI Innovate UK (101095556) and by the National Institute for Health Research (NIHR) Imperial Biomedical Research Centre, grants from Guts UK (DG201808), Diabetes UK (19/0006059) and a Medical Research Council grant to MED and PF (MR/X010155/1), by Isite ULNE (R-002-20-TALENT-DUMAS), jointly funded by ANR (ANR-16-IDEX-0004-ULNE), the Hauts-de-France Regional Council (20002845) and by the European Metropolis of Lille (MEL), an ERC Generator Grant "Richness" (R-ERCGEN-23-003-DUMAS) from the University of Lille and an ANR grant (ANR-25-CE14-1815-01). KChe is supported by the NIHR Imperial Biomedical Research Centre (BRC) through a fellowship jointly funded by the Cardiovascular and Multimorbidity Themes. KChe also acknowledges the support of the Medical Research Council Skills Development Fellowship (grant no. MR/S020039/1) and the Wellcome Trust-funded Institutional Strategic Support Springboard Fellowship (grant no. 204834/Z/16/Z). PA is the recipient of a Career Development Award from the Medical Research Council (Grant No. MR/Y010051/1) and acknowledges the the Wellcome Trust-funded Institutional Strategic Support Springboard Fellowship (grant no. 204834/Z/16/Z). This research is also funded by grants from the French National Research Agency (ANR-10-LABX-46 [European Genomics Institute for Diabetes]), from the National Center for Precision Diabetic Medicine – 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). This research was also conducted as part of the CNRS–Imperial–ULille International Research Project in Integrative Metabolism "METABO-LIC". SKF was supported by the DZHK (German Centre for Cardiovascular Research), the DFG (German Research Society) SFB1470 ("HFpEF"), the European Union Horizon Programme ("IMMEDIATE" network) and by the BMBF (German Ministry of Education and Research).

## ORCID

Eugeni Belda  0000-0003-4307-5072  
 Sofia Forslund  0000-0003-4285-6993  
 Edi Pifti  0000-0001-8861-1305  
 Judith Aron-Wisniewsky  0000-0002-3104-9769  
 Rima Chakaroun  0000-0001-9901-1815  
 Trine Nielsen  0000-0002-2066-7895  
 Christine Poitou  0000-0001-7769-6331  
 Pierre Bel-Lassen  0000-0002-8085-2881  
 Petros Andrikopoulos  0000-0002-1555-3498  
 Kanta Chechi  0000-0003-1929-4128  
 Francesc Puig-Castellví  0000-0003-1064-9586  
 Inés Castro Dionicio  0000-0003-4349-3477  
 Philippe Froguel  0000-0003-2972-0784  
 Bridget Holmes  0000-0001-9375-3025

Fabrizio Andreelli  0000-0002-1207-3624  
 Hedi Soula  0000-0001-5306-9712  
 Joe-Elie Salem  0000-0002-0331-3307  
 Sara Vieira-Silva  0000-0002-4616-7602  
 Jeroen Raes  0000-0002-1337-041X  
 Peer Bork  0000-0002-2627-833X  
 Michael Stumvoll  0000-0001-6225-8240  
 Oluf Pedersen  0000-0002-3321-3972  
 S. Dusko Ehrlich  0000-0002-7563-4046  
 Marc-Emmanuel Dumas  0000-0001-9523-7024  
 Jean-Michel Oppert  0000-0003-0324-4820  
 Maria Carlota Dao  0000-0001-6284-7352  
 Jean-Daniel Zucker  0000-0002-5597-7922  
 Karine Clément  0000-0002-2489-3355

## Data availability statement

Raw shotgun sequencing data that support the findings of this study have been deposited in the European Nucleotide Archive with accession codes <https://www.ebi.ac.uk/ena/browser/view/PRJEB37249>, <https://www.ebi.ac.uk/ena/browser/view/PRJEB38742>, <https://www.ebi.ac.uk/ena/browser/view/PRJEB41311>, <https://www.ebi.ac.uk/ena/browser/view/PRJEB46098> and <https://www.ebi.ac.uk/ena/browser/view/PRJEB71898>, with public access. Metabolome data have been uploaded to MassIVE with accession numbers [MSV000088042](https://massive.ucsd.edu/MSV000088042) (serum GCMS) and [MSV000088043](https://massive.ucsd.edu/MSV000088043) (additional isotopically 97 quantified serum metabolites using UPLC–MS/MS). In adherence to EU and national privacy laws, unrestricted access to patient phenotypic data cannot be provided for MetaCardis. Researchers wishing to access individual phenotypic data would need to submit argued applications to the relevant National Data Protection Agencies. These are the Danish Data Protection Agency (<https://www.datatilsynet.dk/english>) for phenotypic data from study participants recruited in Denmark, the Federal Commissioner for Data Protection ([https://www.bfdi.bund.de/EN/Home/home\\_node.html](https://www.bfdi.bund.de/EN/Home/home_node.html)) for phenotypic data from study participants recruited in Germany and the Commission Nationale Informatique & Libertés (<https://www.cnil.fr/en/home>) for phenotypic data of study participants recruited in France. The application procedures are described on the outlined websites. If such permission is granted, phenotypic data will be then made available by the corresponding authors within 5 weeks.

## Author contributions

SA, EB, MCD, and KC conceived and designed the project. FPC, PA, KChe, and MED performed metabolomic analyses, including quality control, data curation, manual re-annotation, and development of an in-house chemical data flow and knowledge base for metabolic and lipid reporting in untargeted studies. SA and EB managed and curated the data. SA conducted the dietary analysis and EB metagenomic analyses and, along with KC, wrote the manuscript. SKF assisted with data management, curation, deployment of confounder-adjusted analysis tools, results interpretation, and provided feedback on the manuscript. TLR, MCD, JDZ, and KC contributed to results interpretation and manuscript feedback. All the authors reviewed and edited the manuscript. KC is the guarantor, accepting full responsibility for the work and/or conduct of the study, having access to all the data, and controlling the decision to publish.

## References

1. Aron-Wisniewsky J, Prifti E, Belda E, Ichou F, Kayser BD, Dao MC, Verger EO, Hedjazi L, Bouillot J, Chevallier J, et al. Major microbiota dysbiosis in severe obesity: fate after bariatric surgery. *Gut*. 2019;68:70–82. doi: [10.1136/gutjnl-2018-316103](https://doi.org/10.1136/gutjnl-2018-316103).
2. Vieira-Silva S, Falony G, Belda E, Nielsen T, Aron-Wisniewsky J, Chakaroun R, Forslund SK, Assmann K, Valles-Colomer M, Nguyen TTD, et al. Statin therapy is associated with lower prevalence of gut microbiota dysbiosis. *Nature*. 2020;581:310–315. doi: [10.1038/s41586-020-2269-x](https://doi.org/10.1038/s41586-020-2269-x).
3. Belda E, Volland L, Tremaroli V, Falony G, Adriouch S, Assmann KE, Prifti E, Aron-Wisniewsky J, Debédát J, Le Roy T, et al. Impairment of gut microbial biotin metabolism and host biotin status in severe obesity: effect of biotin and prebiotic supplementation on improved metabolism. *Gut*. 2022;71:2463–2480. doi: [10.1136/gutjnl-2021-325753](https://doi.org/10.1136/gutjnl-2021-325753).
4. Koren O, Spor A, Felin J, Fåk F, Stombaugh J, Tremaroli V, Behre CJ, Knight R, Fagerberg B, Ley RE, et al. Human oral, gut, and plaque microbiota in patients with atherosclerosis. *Proc Natl Acad Sci*. 2011;108:4592–4598. doi: [10.1073/pnas.1011383107](https://doi.org/10.1073/pnas.1011383107).

5. Le Chatelier E, Nielsen T, Qin J, Prifti E, Hildebrand F, Falony G, Almeida M, Arumugam M, Batto J, Kennedy S, et al. Richness of human gut microbiome correlates with metabolic markers. *Nature*. 2013;500:541–546. doi: [10.1038/nature12506](https://doi.org/10.1038/nature12506).
6. Falony G, Joossens M, Vieira-Silva S, Wang J, Darzi Y, Faust K, Kurilshikov A, Bonder MJ, Valles-Colomer M, Vandeputte D, et al. Population-level analysis of gut microbiome variation. *Science*. 2016;352:560–564. doi: [10.1126/science.aad3503](https://doi.org/10.1126/science.aad3503).
7. Asnicar F. Microbiome connections with host metabolism and habitual diet from 1,098 deeply phenotyped individuals. *Nat Med*. 2021;27:321–332. doi: [10.1038/s41591-020-01183-8](https://doi.org/10.1038/s41591-020-01183-8).
8. Leclercq S, Matamoros S, Cani PD, Neyrinck AM, Jamar F, Stärkel P, Windey K, Tremaroli V, Bäckhed F, Verbeke K, et al. Intestinal permeability, gut-bacterial dysbiosis, and behavioral markers of alcohol-dependence severity. *Proc Natl Acad Sci*. 2014;111:4485. doi: [10.1073/pnas.1415174111](https://doi.org/10.1073/pnas.1415174111).
9. Scheiman J, Luber JM, Chavkin TA, MacDonald T, Tung A, Pham L, Wibowo MC, Wurth RC, Punthambaker S, Tierney BT, et al. Meta-omics analysis of elite athletes identifies a performance-enhancing microbe that functions via lactate metabolism. *Nat Med*. 2019;25:1104–1109. doi: [10.1038/s41591-019-0485-4](https://doi.org/10.1038/s41591-019-0485-4).
10. Katz DL, Dansinger ML, Willett WC. The study of dietary patterns: righting the remedies. *Am J Health Promot*. 2021;35:875–878. doi: [10.1177/08901171211016191b](https://doi.org/10.1177/08901171211016191b).
11. Asghari G, Mirmiran P, Yuzbashian E, Azizi F. A systematic review of diet quality indices in relation to obesity. *Br J Nutr*. 2017;117:1055–1065. doi: [10.1017/S0007114517000915](https://doi.org/10.1017/S0007114517000915).
12. Cavicchia PP, Steck SE, Hurley TG, Hussey JR, Ma Y, Ockene IS, Hébert JR. A new dietary inflammatory index predicts interval changes in serum high-sensitivity c-reactive protein1–3. *J Nutr*. 2009;139:2365–2372. doi: [10.3945/jn.109.114025](https://doi.org/10.3945/jn.109.114025).
13. Bolte LA, Vich Vila A, Imhann F, Collij V, Gacesa R, Peters V, Wijmenga C, Kurilshikov A, Campmans-Kuijpers MJE, Fu J, et al. Long-term dietary patterns are associated with pro-inflammatory and anti-inflammatory features of the gut microbiome. *Gut*. 2021;70:1287–1298. doi: [10.1136/gutjnl-2020-322670](https://doi.org/10.1136/gutjnl-2020-322670).
14. Tindall AM, Petersen KS, Kris-Etherton PM. Dietary patterns affect the gut microbiome—the link to risk of cardiometabolic diseases. *J Nutr*. 2018;148:1402–1407. doi: [10.1093/jn/nxy141](https://doi.org/10.1093/jn/nxy141).
15. Fackelmann G, Manghi P, Carlino N, Heidrich V, Piccinno G, Ricci L, Piperni E, Arrè A, Bakker E, Creedon AC, et al. Gut microbiome signatures of vegan, vegetarian and omnivore diets and associated health outcomes across 21,561 individuals. *Nat Microbiol*. 2025;10:41–52. doi: [10.1038/s41564-024-01870-z](https://doi.org/10.1038/s41564-024-01870-z).
16. Lachat C. Dietary species richness as a measure of food biodiversity and nutritional quality of diets. *PNAS*. 2018;115:127–132. doi: [10.1073/pnas.1709194115](https://doi.org/10.1073/pnas.1709194115).
17. de Oliveira Otto M, Anderson C, Dearborn J, Ferranti E, Mozaffarian D, Rao G. Dietary diversity: implications for obesity prevention in adult populations: a science advisory from the american heart association. *Circulation*. 2018;138.
18. Hanley-Cook GT, Daly AJ, Remans R, Jones AD, Murray KA, Huybrechts I, De Baets B, Lachat C. Food biodiversity: quantifying the unquantifiable in human diets. *Crit Rev Food Sci Nutr*. 2022;63:1–15. doi: [10.1080/10408398.2022.2051163](https://doi.org/10.1080/10408398.2022.2051163).
19. Johnson AJ, Vangay P, Al-Ghalith GA, Hillmann BM, Ward TL, Shields-Cutler RR, Kim AD, Shmagel AK, Syed AN, Walter J, et al. Daily sampling reveals personalized diet-microbiome associations in humans. *Cell Host Microbe*. 2019;25:789–802. doi: [10.1016/j.chom.2019.05.005](https://doi.org/10.1016/j.chom.2019.05.005).
20. Amamoto R, Shimamoto K, Suwa T, Park S, Matsumoto H, Shimizu K, Katto M, Makino H, Matsubara S, Aoyagi Y. Relationships between dietary diversity and gut microbial diversity in the elderly. *Benef Microbes*. 2022;13:453–464. doi: [10.3920/BM2022.0054](https://doi.org/10.3920/BM2022.0054).
21. GBD 2017 Risk Factor Collaborators. Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet London England*. 2018;392:1923–1994. doi: [10.1016/S0140-6736\(18\)32225-6](https://doi.org/10.1016/S0140-6736(18)32225-6).
22. Vassallo G, Mirijello A, Ferrulli A, Antonelli M, Landolfi R, Gasbarrini A, Addolorato G. Review article: Alcohol and gut microbiota – the possible role of gut microbiota modulation in the treatment of alcoholic liver disease. *Aliment Pharmacol Ther*. 2015;41:917–927. doi: [10.1111/apt.13164](https://doi.org/10.1111/apt.13164).
23. Kosnicki KL, Penprase JC, Cintora P, Torres PJ, Harris GL, Brasser SM, Kelley ST. Effects of moderate, voluntary ethanol consumption on the rat and human gut microbiome. *Addict Biol*. 2019;24:617–630. doi: [10.1111/adb.12626](https://doi.org/10.1111/adb.12626).
24. Dohnalová L, Lundgren P, Carty JRE, Goldstein N, Wenski SL, Nanudorn P, Thiengmag S, Huang K, Litichevskiy L, Descamps HC, et al. A microbiome-dependent gut–brain pathway regulates motivation for exercise. *Nature*. 2022;612:739–747. doi: [10.1038/s41586-022-05525-z](https://doi.org/10.1038/s41586-022-05525-z).
25. Antinozzi M, Giffi M, Sini N, Gallè F, Valeriani F, De Vito C, Liguori G, Romano Spica V, Cattaruzza MS. Cigarette smoking and human gut microbiota in healthy adults: a systematic review. *Biomedicines*. 2022;10:510. doi: [10.3390/biomedicines10020510](https://doi.org/10.3390/biomedicines10020510).
26. Forman JP. Diet and lifestyle risk factors associated with incident hypertension in women. *J Am Med Assoc*. 2009;302:401. doi: [10.1001/jama.2009.1060](https://doi.org/10.1001/jama.2009.1060).

27. Rothschild D, Weissbrod O, Barkan E, Kurilshikov A, Korem T, Zeevi D, Costea PI, Godneva A, Kalka IN, Bar N, et al. Environment dominates over host genetics in shaping human gut microbiota. *Nature*. 2018;555:210–215. doi: [10.1038/nature25973](https://doi.org/10.1038/nature25973).
28. Aya V, Flórez A, Perez L, Ramírez JD. Association between physical activity and changes in intestinal microbiota composition: a systematic review. *PLoS One*. 2021;16:e0247039. doi: [10.1371/journal.pone.0247039](https://doi.org/10.1371/journal.pone.0247039).
29. Forslund SK, Chakaroun R, Zimmermann-Kogadeeva M, Markó L, Aron-Wisnewsky J, Nielsen T, Moitinho-Silva L, Schmidt TSB, Falony G, Vieira-Silva S, et al. Combinatorial, additive and dose-dependent drug–microbiome associations. *Nature*. 2021;600:500–505. doi: [10.1038/s41586-021-04177-9](https://doi.org/10.1038/s41586-021-04177-9).
30. Fromentin S, Forslund SK, Chechi K, Aron-Wisnewsky J, Chakaroun R, Nielsen T, Tremaroli V, Ji B, Prifti E, Myridakis A, et al. Microbiome and metabolome features of the cardiometabolic disease spectrum. *Nat Med*. 2022;28:303–314. doi: [10.1038/s41591-022-01688-4](https://doi.org/10.1038/s41591-022-01688-4).
31. Alili R, Belda E, Fabre O, Pelloux V, Giordano N, Legrand R, Bel Lassen P, Swartz TD, Zucker J, Clément K. Characterization of the Gut Microbiota in Individuals with Overweight or Obesity during a Real-World Weight Loss Dietary Program: A Focus on the Bacteroides 2 Enterotype. *Biomedicines*. 2021;10:16. doi: [10.3390/biomedicines10010016](https://doi.org/10.3390/biomedicines10010016).
32. Whelton PK, Carey RM, Aronow WS, Casey DE, Collins KJ, Dennison Himmelfarb C, DePalma SM, Gidding S, Jamerson KA, Jones DW, et al. 2017. ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA guideline for the prevention, detection, evaluation, and management of high blood pressure in adults: executive summary: a report of the American college of cardiology/American heart association task force on clinical practice guidelines. *Hypertension*. 2018;71:1269–1324. doi: [10.1161/HYP.0000000000000666](https://doi.org/10.1161/HYP.0000000000000666).
33. Yancy CW, Jessup M, Bozkurt B, Butler J, Casey DE, Colvin MM, Drazner MH, Filippatos GS, Fonarow GC, Givertz MM, et al. 2017 ACC/AHA/HFSA focused Update of the 2013 ACCF/AHA guideline for the management of heart failure: a report of the American college of cardiology/American heart association task force on clinical practice guidelines and the heart failure society of America. *Circulation*. 2017;136. doi: [10.1161/CIR.0000000000000509](https://doi.org/10.1161/CIR.0000000000000509).
34. Verger EO, Armstrong P, Nielsen T, Chakaroun R, Aron-Wisnewsky J, Gøbel RJ, Schütz T, Delaere F, Gausseres N, Clément K, et al. Dietary assessment in the metacardis study: development and relative validity of an online food frequency questionnaire. *J Acad Nutr Diet*. 2017;117:878–888. doi: [10.1016/j.jand.2016.10.030](https://doi.org/10.1016/j.jand.2016.10.030).
35. Neveu V, Perez-Jimenez J, Vos F, Crespy V, du Chaffaut L, Mennen L, Knox C, Eisner R, Cruz J, Wishart D, et al. Phenol-Explorer: an online comprehensive database on polyphenol contents in foods. *Database*. 2010;2010:bap024. doi: [10.1093/database/bap024](https://doi.org/10.1093/database/bap024).
36. Besson H, Brage S, Jakes RW, Ekelund U, Wareham NJ. Estimating physical activity energy expenditure, sedentary time, and physical activity intensity by self-report in adults. *Am J Clin Nutr*. 2010;91:106–114. doi: [10.3945/ajcn.2009.28432](https://doi.org/10.3945/ajcn.2009.28432).
37. Ainsworth BE, Haskell WL, Herrmann SD, Meckes N, Bassett DR, Tudor-Locke C, Greer JL, Vezina J, Whitt-Glover MC, Leon AS. 2011 Compendium of physical activities: a second update of codes and MET values. *Med Sci Sports Exerc*. 2011;43:1575–1581. doi: [10.1249/MSS.0b013e31821ece12](https://doi.org/10.1249/MSS.0b013e31821ece12).
38. van Woudenberg GJ, Theofylaktopoulos D, Kuijsten A, Ferreira I, van Greevenbroek MM, van der Kallen CJ, Schalkwijk CG, Stehouwer CD, Ocké MC, Nijpels G, et al. Adapted dietary inflammatory index and its association with a summary score for low-grade inflammation and markers of glucose metabolism: the Cohort study on diabetes and atherosclerosis maastricht (CODAM) and the hoorn study. *Am J Clin Nutr*. 2013;98:1533–1542. doi: [10.3945/ajcn.112.056333](https://doi.org/10.3945/ajcn.112.056333).
39. Willett WC, Howe GR, Kushi LH. Adjustment for total energy intake in epidemiologic studies. *Am J Clin Nutr*. 1997;65:1220S–1228S. doi: [10.1093/ajcn/65.4.1220S](https://doi.org/10.1093/ajcn/65.4.1220S).
40. McCullough ML, Feskanich D, Stampfer MJ, Giovannucci EL, Rimm EB, Hu FB, Spiegelman D, Hunter DJ, Colditz GA, Willett WC. Diet quality and major chronic disease risk in men and women: moving toward improved dietary guidance. *Am J Clin Nutr*. 2002;76:1261–1271. doi: [10.1093/ajcn/76.6.1261](https://doi.org/10.1093/ajcn/76.6.1261).
41. Chiuve SE, Fung TT, Rimm EB, Hu FB, McCullough ML, Wang M, Stampfer MJ, Willett WC. Alternative dietary indices both strongly predict risk of chronic disease. *J Nutr*. 2012;142:1009–1018. doi: [10.3945/jn.111.157222](https://doi.org/10.3945/jn.111.157222).
42. Wang P, Song M, Eliassen AH, Fung TT, Clinton SK, Rimm EB, Hu FB, Willett WC, Tabung FK, Giovannucci EL. Optimal dietary patterns for prevention of chronic disease. *Nat Med*. 2023;29:719–728. doi: [10.1038/s41591-023-02235-5](https://doi.org/10.1038/s41591-023-02235-5).
43. Fung TT. Adherence to a DASH-style diet and risk of coronary heart disease and stroke in women. *Arch Intern Med*. 2008;168:713. doi: [10.1001/archinte.168.7.713](https://doi.org/10.1001/archinte.168.7.713).
44. Berry CH. Corporate growth and diversification. *J Law Econ*. 1971;14:371–383. doi: [10.1086/466714](https://doi.org/10.1086/466714).
45. Drescher LS, Thiele S, Mensink GBM. A new index to measure healthy food diversity better reflects a healthy diet than traditional measures. *J Nutr*. 2007;137:647–651. doi: [10.1093/jn/137.3.647](https://doi.org/10.1093/jn/137.3.647).
46. Bianchi CM, Egnell M, Huneau J-F, Mariotti F. Plant protein intake and dietary diversity are independently associated with nutrient adequacy in french adults. *J Nutr*. 2016;146:2351–2360. doi: [10.3945/jn.116.236869](https://doi.org/10.3945/jn.116.236869).
47. Haines PS, Siega-Riz AM, Popkin BM. The diet quality index revised. *J Am Diet Assoc*. 1999;99:697–704. doi: [10.1016/S0002-8223\(99\)00168-6](https://doi.org/10.1016/S0002-8223(99)00168-6).

48. Hasbani NR, Lighthart S, Brown MR, Heath AS, Bebo A, Ashley KE, Boerwinkle E, Morrison AC, Folsom AR, Aguilar D, et al. American heart association's life's simple 7: lifestyle recommendations, polygenic risk, and lifetime risk of coronary heart disease. *Circulation*. 2022;145:808–818. doi: [10.1161/CIRCULATIONAHA.121.053730](https://doi.org/10.1161/CIRCULATIONAHA.121.053730).
49. Trichopoulou A, Costacou T, Bamia C, Trichopoulos D. Adherence to a mediterranean diet and survival in a Greek population. *N Engl J Med*. 2003;348:2599–2608. doi: [10.1056/NEJMoa025039](https://doi.org/10.1056/NEJMoa025039).
50. Kara P, Valentin JB, Mainz J, Johnsen SP. Composite measures of quality of health care: evidence mapping of methodology and reporting. *PLoS One*. 2022;17:e0268320. doi: [10.1371/journal.pone.0268320](https://doi.org/10.1371/journal.pone.0268320).
51. Guthrie B. Measuring the quality of healthcare systems using composites. *Brit Med J*. 2008;337:a639. doi: [10.1136/bmj.a639](https://doi.org/10.1136/bmj.a639).
52. Criscuolo A, Brisse S. AlienTrimmer: a tool to quickly and accurately trim off multiple short contaminant sequences from high-throughput sequencing reads. *Genomics*. 2013;102:500–506. doi: [10.1016/j.ygeno.2013.07.011](https://doi.org/10.1016/j.ygeno.2013.07.011).
53. Li J, Jia H, Cai X, Zhong H, Feng Q, Sunagawa S, Arumugam M, Kultima JR, Prifti E, Nielsen T, et al. An integrated catalog of reference genes in the human gut microbiome. *Nat Biotechnol*. 2014;32(8):834–841.
54. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9:357–359. doi: [10.1038/nmeth.1923](https://doi.org/10.1038/nmeth.1923).
55. Ghozlane A, Thirion F, Plaza Oñate F, et al. Accurate profiling of microbial communities for shotgun metagenomic sequencing with Meteor2. *Microbiome*. 2025;13 227. doi: [10.1186/s40168-025-02249-w](https://doi.org/10.1186/s40168-025-02249-w).
56. Le Chatelier E, Nielsen T, Qin J, Prifti E, Hildebrand F, Falony G, Almeida M, Arumugam M, Batto J, Kennedy S, et al. Richness of human gut microbiome correlates with metabolic markers. *Nature*. 2013;500:541–546. doi: [10.1038/nature12506](https://doi.org/10.1038/nature12506).
57. Kultima JR, Coelho LP, Forslund K, Huerta-Cepas J, Li SS, Driessen M, Voigt AY, Zeller G, Sunagawa S, Bork P. MOCAT2: a metagenomic assembly, annotation and profiling framework. *Bioinform Oxf Engl*. 2016;32:2520–2523. doi: [10.1093/bioinformatics/btw183](https://doi.org/10.1093/bioinformatics/btw183).
58. Nielsen HB, Almeida M, Juncker AS, Rasmussen S, Li J, Sunagawa S, Plichta DR, Gautier L, Pedersen AG, Le Chatelier E, et al. Identification and assembly of genomes and genetic elements in complex metagenomic samples without using reference genomes. *Nat Biotechnol*. 2014;32:822–828.
59. Coelho LP, Alves R, Monteiro P, Huerta-Cepas J, Freitas AT, Bork P. NG-meta-profiler: fast processing of metagenomes using NGLess, a domain-specific language. *Microbiome*. 2019;7:84. doi: [10.1186/s40168-019-0684-8](https://doi.org/10.1186/s40168-019-0684-8).
60. Sunagawa S, Mende DR, Zeller G, Izquierdo-Carrasco F, Berger SA, Kultima JR, Coelho LP, Arumugam M, Tap J, Nielsen HB, et al. Metagenomic species profiling using universal phylogenetic marker genes. *Nat Methods*. 2013;10:1196–1199. doi: [10.1038/nmeth.2693](https://doi.org/10.1038/nmeth.2693).
61. Holmes I, Harris K, Quince C. Dirichlet multinomial mixtures: generative models for microbial metagenomics. *PLoS One*. 2012;7:e30126. doi: [10.1371/journal.pone.0030126](https://doi.org/10.1371/journal.pone.0030126).
62. MetaCardis Consortium, Vieira-Silva S, Falony G, Belda E, Nielsen T, Aron-Wisnewsky J, Chakaroun R, Forslund SK, Assmann K, Valles-Colomer M, et al. Statin therapy is associated with lower prevalence of gut microbiota dysbiosis. *Nature*. 2020;581:310–315. doi: [10.1038/s41586-020-2269-x](https://doi.org/10.1038/s41586-020-2269-x).
63. Oksanen J, et al. 2022. Vegan. Community ecology package.
64. Peeters CFW, Übelhör C, Mes SW, Martens R, Koopman T, de Graaf P, van Velden FH, Boellaard R, Castelijns JA, Beest DET, et al. (2019). Stable prediction with radiomics data. doi: [10.48550/ARXIV.1903.11696](https://doi.org/10.48550/ARXIV.1903.11696).
65. Peng B, Yu RK, DeHoff KL, Amos CI. Normalizing a large number of quantitative traits using empirical normal quantile transformation. *BMC Proc*. 2007;1:S156. doi: [10.1186/1753-6561-1-S1-S156](https://doi.org/10.1186/1753-6561-1-S1-S156).
66. Fan Y, Støving RK, Berreira Ibraim S, Hyötyläinen T, Thirion F, Arora T, Lyu L, Stankevic E, Hansen TH, Déchelotte P, et al. The gut microbiota contributes to the pathogenesis of anorexia nervosa in humans and mice. *Nat Microbiol*. 2023;8:787–802. doi: [10.1038/s41564-023-01355-5](https://doi.org/10.1038/s41564-023-01355-5).
67. Tingley D, Yamamoto T, Hirose K, Keele L, Imai K. mediation: R Package for Causal Mediation Analysis. *J Stat Softw*. 2014;59:1–38. doi: [10.18637/jss.v059.i05](https://doi.org/10.18637/jss.v059.i05).
68. Forslund K, Hildebrand F, Nielsen T, Falony G, Le Chatelier E, Sunagawa S, Prifti E, Vieira-Silva S, Gudmundsdottir V, Krogh Pedersen H, et al. Disentangling type 2 diabetes and metformin treatment signatures in the human gut microbiota. *Nature*. 2015;528:262–266. doi: [10.1038/nature15766](https://doi.org/10.1038/nature15766).
69. Plamada D, Vodnar DC. Polyphenols—gut microbiota interrelationship: a transition to a new generation of prebiotics. *Nutrients*. 2021;14:137. doi: [10.3390/nu14010137](https://doi.org/10.3390/nu14010137).
70. Sanders ME, Merenstein DJ, Reid G, Gibson GR, Rastall RA. Probiotics and prebiotics in intestinal health and disease: from biology to the clinic. *Nat Rev Gastroenterol Hepatol*. 2019;16:605–616. doi: [10.1038/s41575-019-0173-3](https://doi.org/10.1038/s41575-019-0173-3).
71. Abu-Ghazaleh N, Chua WJ, Gopalan V. Intestinal microbiota and its association with colon cancer and red/processed meat consumption. *J Gastroenterol Hepatol*. 2021;36:75–88. doi: [10.1111/jgh.15042](https://doi.org/10.1111/jgh.15042).
72. Singh RK, Chang H, Yan D, Lee KM, Ucmak D, Wong K, Abrouk M, Farahnik B, Nakamura M, Zhu TH, et al. Influence of diet on the gut microbiome and implications for human health. *J Transl Med*. 2017;15:73. doi: [10.1186/s12967-017-1175-y](https://doi.org/10.1186/s12967-017-1175-y).
73. Extension of PERMANOVA to testing the mediation effect of the microbiome. <https://www.mdpi.com/2073-4425/13/6/940>.

74. Menni C, Zhu J, Le Roy CI, Mompeo O, Young K, Rebholz CM, Selvin E, North KE, Mohney RP, Bell JT, et al. Serum metabolites reflecting gut microbiome alpha diversity predict type 2 diabetes. *Gut Microbes*. 2020;11:1632–1642. doi: [10.1080/19490976.2020.1778261](https://doi.org/10.1080/19490976.2020.1778261).
75. Wilmanski T, Rappaport N, Earls JC, Magis AT, Manor O, Lovejoy J, Omenn GS, Hood L, Gibbons SM, Price ND. Blood metabolome predicts gut microbiome  $\alpha$ -diversity in humans. *Nat Biotechnol*. 2019;37:1217–1228. doi: [10.1038/s41587-019-0233-9](https://doi.org/10.1038/s41587-019-0233-9).
76. Meléndez-Martínez AJ, Mandić AI, Bantis F, Böhm V, Borge GIA, Brnčić M, Bysted A, Cano MP, Dias MG, Elgersma A, et al. A comprehensive review on carotenoids in foods and feeds: status quo, applications, patents, and research needs. *Crit Rev Food Sci Nutr*. 2022;62:1999–2049. doi: [10.1080/10408398.2020.1867959](https://doi.org/10.1080/10408398.2020.1867959).
77. Zhou Y, Ju Y, Yang Y, Sang Z, Wang Z, He G, Luo Y. Discovery of hybrids of indolin-2-one and nitroimidazole as potent inhibitors against drug-resistant bacteria. *J Antibiot*. 2018;71:887–897. doi: [10.1038/s41429-018-0076-5](https://doi.org/10.1038/s41429-018-0076-5). Tokyo.
78. Authority (EFSA) E.F.S. Perfluorooctane sulfonate (PFOS), perfluorooctanoic acid (PFOA) and their salts scientific opinion of the panel on contaminants in the food chain. *EFSA J*. 2008;6:653.
79. Zhu Y, Dwidar M, Nemet I, Buffa JA, Sangwan N, Li XS, Anderson JT, Romano KA, Fu X, Funabashi M, et al. Two distinct gut microbial pathways contribute to meta-organismal production of phenylacetylglutamine with links to cardiovascular disease. *Cell Host Microbe*. 2023;31:18–32. doi: [10.1016/j.chom.2022.11.015](https://doi.org/10.1016/j.chom.2022.11.015).
80. Manor O, Dai CL, Kornilov SA, Smith B, Price ND, Lovejoy JC, Gibbons SM, Magis AT. Health and disease markers correlate with gut microbiome composition across thousands of people. *Nat Commun*. 2020;11:5206. doi: [10.1038/s41467-020-18871-1](https://doi.org/10.1038/s41467-020-18871-1).
81. Le Roy CI, Wells PM, Si J, Raes J, Bell JT, Spector TD. Red wine consumption associated with increased gut microbiota  $\alpha$ -diversity in 3 independent cohorts. *Gastroenterology*. 2020;158:270–272. doi: [10.1053/j.gastro.2019.08.024](https://doi.org/10.1053/j.gastro.2019.08.024).
82. Hul MV, Cani PD, Petitfils C, De Vos WM, Tilg H, El-Omar EM. What defines a healthy gut microbiome? *Gut*. 2024. doi: [10.1136/gutjnl-2024-333378](https://doi.org/10.1136/gutjnl-2024-333378).
83. Roager HM, Hansen LBS, Bahl MI, Frandsen HL, Carvalho V, Gøbel RJ, Dalgaard MD, Plichta DR, Sparholt MH, Vestergaard H, et al. Colonic transit time is related to bacterial metabolism and mucosal turnover in the gut. *Nat Microbiol*. 2016;1:16093. doi: [10.1038/nmicrobiol.2016.93](https://doi.org/10.1038/nmicrobiol.2016.93).
84. Hu FB. Dietary pattern analysis: a new direction in nutritional epidemiology. *Curr Opin Lipidol*. 2002;13:3–9. doi: [10.1097/00041433-200202000-00002](https://doi.org/10.1097/00041433-200202000-00002).
85. Cotillard A, Cartier-Meheust A, Litwin NS, Chaumont S, Saccareau M, Lejzerowicz F, Tap J, Koutnikova H, Lopez DG, McDonald D, et al. A posteriori dietary patterns better explain variations of the gut microbiome than individual markers in the American Gut Project. *Am J Clin Nutr*. 2021;332. doi: [10.1093/ajcn/nqab332](https://doi.org/10.1093/ajcn/nqab332).
86. Jacobs DR, Gross MD, Tapsell LC. Food synergy: an operational concept for understanding nutrition. *Am J Clin Nutr*. 2009;89:1543S–1548S. doi: [10.3945/ajcn.2009.26736B](https://doi.org/10.3945/ajcn.2009.26736B).
87. Chen L, Zhernakova DV, Kurilshikov A, Andreu-Sánchez S, Wang D, Augustijn HE, Vich Vila A, Weersma RK, Medema MH, Netea MG, et al. Influence of the microbiome, diet and genetics on inter-individual variation in the human plasma metabolome. *Nat Med*. 2022;28:2333–2343. doi: [10.1038/s41591-022-02014-8](https://doi.org/10.1038/s41591-022-02014-8).
88. Qin Y, Havulinna AS, Liu Y, Jousilahti P, Ritchie SC, Tokolyi A, Sanders JG, Valsta L, Brożynańska M, Zhu Q, et al. Combined effects of host genetics and diet on human gut microbiota and incident disease in a single population cohort. *Nat Genet*. 2022;54:134–142. doi: [10.1038/s41588-021-00991-z](https://doi.org/10.1038/s41588-021-00991-z).
89. Arumugam M, Raes J, Pelletier E, Le Paslier D, Yamada T, Mende DR, Fernandes GR, Tap J, Bruls T, Batto J, et al. Enterotypes of the human gut microbiome. *Nature*. 2011;473:174–180. doi: [10.1038/nature09944](https://doi.org/10.1038/nature09944).
90. Koren O, Knights D, Gonzalez A, Waldron L, Segata N, Knight R, Huttenhower C, Ley RE, Eisen JA. A guide to enterotypes across the human body: meta-analysis of microbial community structures in human microbiome datasets. *PLoS Comput Biol*. 2013;9:e1002863. doi: [10.1371/journal.pcbi.1002863](https://doi.org/10.1371/journal.pcbi.1002863).
91. Costea PI, Hildebrand F, Arumugam M, Bäckhed F, Blaser MJ, Bushman FD, de Vos WM, Ehrlich SD, Fraser CM, Hattori M, et al. Enterotypes in the landscape of gut microbial community composition. *Nat Microbiol*. 2018;3:8–16.
92. Ben-Yacov O, Godneva A, Rein M, Shilo S, Lotan-Pompan M, Weinberger A, Segal E. Gut microbiome modulates the effects of a personalised postprandial-targeting (PPT) diet on cardiometabolic markers: a diet intervention in pre-diabetes. *Gut*. 2023;72:1486–1496. doi: [10.1136/gutjnl-2022-329201](https://doi.org/10.1136/gutjnl-2022-329201).
93. Shoer S, Shilo S, Godneva A, Ben-Yacov O, Rein M, Wolf BC, Lotan-Pompan M, Bar N, Weiss EI, Hourihaddad Y, et al. Impact of dietary interventions on pre-diabetic oral and gut microbiome, metabolites and cytokines. *Nat Commun*. 2023;14:5384. doi: [10.1038/s41467-023-41042-x](https://doi.org/10.1038/s41467-023-41042-x).
94. Wang D, Pham VT, Steinert RE, Zhernakova A, Fu J. Microbial vitamin production mediates dietary effects on diabetic risk. *Gut Microbes*. 2022;14:2154550. doi: [10.1080/19490976.2022.2154550](https://doi.org/10.1080/19490976.2022.2154550).
95. Chen L, Wang D, Garmaeva S, Kurilshikov A, Vich Vila A, Gacesa R, Sinha T, Segal E, Weersma RK, Wijmenga C, et al. The long-term genetic stability and individual specificity of the human gut microbiome. *Cell*. 2021;184:2302–2315. doi: [10.1016/j.cell.2021.03.024](https://doi.org/10.1016/j.cell.2021.03.024).
96. Wang DD, Nguyen LH, Li Y, Yan Y, Ma W, Rinott E, Ivey KL, Shai I, Willett WC, Hu FB, et al. The gut microbiome modulates the protective association between a Mediterranean diet and cardiometabolic disease risk. *Nat Med*. 2021;27:333–343. doi: [10.1038/s41591-020-01223-3](https://doi.org/10.1038/s41591-020-01223-3).

97. Rinott E, Meir AY, Tsaban G, Zelicha H, Kaplan A, Knights D, Tuohy K, Scholz MU, Koren O, Stampfer MJ, et al. The effects of the Green-mediterranean diet on cardiometabolic health are linked to gut microbiome modifications: a randomized controlled trial. *Genome Med.* 2022;14:29. doi: [10.1186/s13073-022-01015-z](https://doi.org/10.1186/s13073-022-01015-z).
98. Zeevi D, Korem T, Zmora N, Israeli D, Rothschild D, Weinberger A, Ben-Yacov O, Lador D, Avnit-Sagi T, Lotan-Pompan M, et al. Personalized nutrition by prediction of glycemic responses. *Cell.* 2015;163:1079–1094. doi: [10.1016/j.cell.2015.11.001](https://doi.org/10.1016/j.cell.2015.11.001).
99. Wishart DS, Oler E, Peters H, Guo A, Girod S, Han S, Saha S, Lui VW, LeVatte M, Gautam V, et al. MiMeDB: the human microbial metabolome database. *Nucleic Acids Res.* 2023;51:D611–D620. doi: [10.1093/nar/gkac868](https://doi.org/10.1093/nar/gkac868).
100. Wikoff WR, Anfora AT, Liu J, Schultz PG, Lesley SA, Peters EC, Siuzdak G. Metabolomics analysis reveals large effects of gut microflora on mammalian blood metabolites. *Proc Natl Acad Sci U S A.* 2009;106:3698–3703. doi: [10.1073/pnas.0812874106](https://doi.org/10.1073/pnas.0812874106).
101. Louca P, Meijnikman AS, Nogal A, Asnicar F, Attaye I, Vijay A, Kouraki A, Visconti A, Wong K, Berry SE, et al. The secondary bile acid isoursodeoxycholate correlates with post-prandial lipemia, inflammation, and appetite and changes post-bariatric surgery. *Cell Rep Med.* 2023;4:100993. doi: [10.1016/j.xcrm.2023.100993](https://doi.org/10.1016/j.xcrm.2023.100993).
102. Heinken A, Hertel J, Acharya G, Ravcheev DA, Nyga M, Okpala OE, Hogan M, Magnúsdóttir S, Martinelli F, Nap B, et al. Genome-scale metabolic reconstruction of 7,302 human microorganisms for personalized medicine. *Nat Biotechnol.* 2023;41:1320–1331. doi: [10.1038/s41587-022-01628-0](https://doi.org/10.1038/s41587-022-01628-0).
103. Heinken A, Ravcheev DA, Baldini F, Heirendt L, Fleming RMT, Thiele I. Systematic assessment of secondary bile acid metabolism in gut microbes reveals distinct metabolic capabilities in inflammatory bowel disease. *Microbiome.* 2019;7:75. doi: [10.1186/s40168-019-0689-3](https://doi.org/10.1186/s40168-019-0689-3).
104. Wu H, Lv B, Zhi L, Shao Y, Liu X, Mitteregger M, Chakaroun R, Tremaroli V, Hazen SL, Wang R, et al. Microbiome–metabolome dynamics associated with impaired glucose control and responses to lifestyle changes. *Nat Med.* 2025;31:1–10. doi: [10.1038/s41591-025-03642-6](https://doi.org/10.1038/s41591-025-03642-6).