Article

Statin therapy is associated with lower prevalence of gut microbiota dysbiosis

https://doi.org/10.1038/s41586-020-2269-x

Received: 5 August 2019

Accepted: 3 April 2020

Published online: 6 May 2020

Check for updates

Sara Vieira-Silva[1,2,36], Gwen Falony[1,2,36], Eugeni Belda[3,4,36], Trine Nielsen[5], Judith Aron-Wisnewsky[3,6], Rima Chakaroun[7], Sofia K. Forslund[8,9,10,11], Karen Assmann[3], Mireia Valles-Colomer[1,2], Thi Thuy Duyen Nguyen[1,2], Sebastian Proost[1,2], Edi Prifti[3,4,12], Valentina Tremaroli[13], Nicolas Pons[14], Emmanuelle Le Chatelier[14], Fabrizio Andreelli[3,15], Jean-Phillippe Bastard[16,17], Luis Pedro Coelho[9,18], Nathalie Galleron[14], Tue H. Hansen[5], Jean-Sébastien Hulot[19,20,21], Christian Lewinter[22], Helle K. Pedersen[5], Benoit Quinquis[14], Christine Rouault[3], Hugo Roume[14], Joe-Elie Salem[19], Nadja B. Søndertoft[5], Sothea Touch[3], MetaCardis Consortium*, Marc-Emmanuel Dumas[23,24], Stanislav Dusko Ehrlich[14], Pilar Galan[25], Jens P. Gøtze[26], Torben Hansen[5], Jens J. Holst[5], Lars Køber[22], Ivica Letunic[27], Jens Nielsen[28], Jean-Michel Oppert[6], Michael Stumvoll[7,29], Henrik Vestergaard[5], Jean-Daniel Zucker[3,4,12], Peer Bork[9,30,31], Oluf Pedersen[5], Fredrik Bäckhed[5,13], Karine Clément[3,6,37] ✉ & Jeroen Raes[1,2,37] ✉

Microbiome community typing analyses have recently identified the *Bacteroides*2 (Bact2) enterotype, an intestinal microbiota configuration that is associated with systemic inflammation and has a high prevalence in loose stools in humans[1,2]. Bact2 is characterized by a high proportion of *Bacteroides*, a low proportion of *Faecalibacterium* and low microbial cell densities[1,2], and its prevalence varies from 13% in a general population cohort to as high as 78% in patients with inflammatory bowel disease[2]. Reported changes in stool consistency[3] and inflammation status[4] during the progression towards obesity and metabolic comorbidities led us to propose that these developments might similarly correlate with an increased prevalence of the potentially dysbiotic Bact2 enterotype. Here, by exploring obesity-associated microbiota alterations in the quantitative faecal metagenomes of the cross-sectional MetaCardis Body Mass Index Spectrum cohort (n = 888), we identify statin therapy as a key covariate of microbiome diversification. By focusing on a subcohort of participants that are not medicated with statins, we find that the prevalence of Bact2 correlates with body mass index, increasing from 3.90% in lean or overweight participants to 17.73% in obese participants. Systemic inflammation levels in Bact2-enterotyped individuals are higher than predicted on the basis of their obesity status, indicative of Bact2 as a dysbiotic microbiome constellation. We also observe that obesity-associated microbiota dysbiosis is negatively associated with statin treatment, resulting in a lower Bact2 prevalence of 5.88% in statin-medicated obese participants. This finding is validated in both the accompanying MetaCardis cardiovascular disease dataset (n = 282) and the independent Flemish Gut Flora Project population cohort (n = 2,345). The potential benefits of statins in this context will require further evaluation in a prospective clinical trial to ascertain whether the effect is reproducible in a randomized population and before considering their application as microbiota-modulating therapeutics.

Indications that alterations in the faecal microbiome are linked to the development of obesity[5] have resulted in intense research efforts since the early days of metagenomics. However, developing a comprehensive blueprint of an obesity-associated microbiota constellation has proved challenging[6]. Although compositional observations still remain inconclusive[7], obesity and obesity-related comorbidities have clearly been associated with alterations in the intestinal microbiota, including lowered faecal-community richness and reduced proportional abundances of butyrate producing bacteria[7–9].

Cross-sectional microbiome-association studies are inherently limited regarding the inference of causality, and are potentially biased by unaccounted confounders. However, they remain highly suitable for

A list of affiliations appears at the end of the paper.

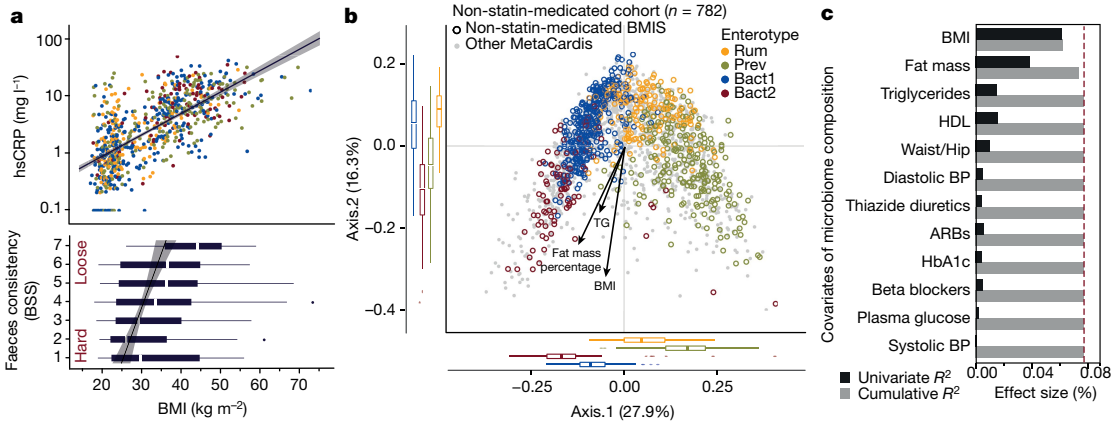


Fig. 1 | Microbiome variation in the non-statin-medicated BMIS cohort. **a**, Correlations between BMI and inflammation levels (top; serum hsCRP, $n = 763$ biologically independent samples, Spearman's $\rho = 0.70$, $P_{adj} = 1.60 \times 10^{-113}$) and faeces consistency (bottom; Bristol Stool Scale (BSS), $n = 772$ biologically independent samples, Spearman's $\rho = 0.16$, $P_{adj} = 9.13 \times 10^{-6}$). Adjustment for multiple testing (P_{adj}) was performed using the Benjamini–Hochberg method. **b**, Principal coordinates analysis of inter-individual differences (genus level Bray–Curtis dissimilarity) in the microbiome profiles of the non-statin-medicated BMIS cohort (open circles, coloured by enterotype; Extended Data Fig. 4), with the rest of the MetaCardis dataset in the background ($n = 1{,}240$ biologically independent samples, grey dots). Arrows represent the effect sizes of a post hoc fit of significant microbiome covariates identified in the multivariate model in **c**. **c**, Variables correlating most to microbiome compositional variation in the non-statin-medicated BMIS cohort (dbRDA, genus-level Bray–Curtis dissimilarity), either independently (univariate effect sizes in black) or in a multivariate model (cumulative effect sizes in grey). The cut-off for significant non-redundant contribution to the multivariate model is represented by the red dashed line. In **a**, **b**, the body of the box plot represents the first and third quartiles of the distribution, the line represents the median, and the whiskers extend from the quartiles to the last data point within 1.5× IQR, with outliers beyond.

explorative analyses, as they enable the scale requirements imposed by the moderate effect-sizes[10] to be met with relative ease. As part of the European Union MetaCardis project, a large-scale observational cohort study was set up to investigate the role of gut microorganisms in the progression of cardio-metabolic diseases through a combination of metagenomic, metabolomic and clinical approaches (http://www.metacardis.net). Recruitment efforts resulted in the enrolment of more than 2,000 participants (Supplementary Fig. 1) and involved, amongst others, the assembly of a transnational $n = 888$ Body Mass Index Spectrum cohort (BMIS; median BMI = 31.5 kg m^{-2}, range = 18.0–73.3; Supplementary Tables 1, 2). Faecal samples were analysed using a quantitative microbiome profiling pipeline[1] adapted for shotgun metagenomics data and were subsequently annotated with customized metabolic modules[11] (Supplementary Table 3). Because more than 42% of BMIS participants reported taking at least one type of medication at the time of sampling, we assessed the potential confounding effect of the most frequently disclosed therapeutics (those consumed by more than 10% of participants; Extended Data Fig. 1a, Supplementary Table 1) on the association between microbiota and obesity; this was achieved by evaluating their explanatory power on relative genus-level microbiome variation as compared with the effect-sizes of obesity parameters and host variables constituting the International Diabetes Federation consensus definition of metabolic syndrome[12] (Supplementary Table 4). Statins were identified as the drugs with largest explanatory power, contributing to genus-level microbiome variation beyond the effect of obesity-related parameters and metabolic syndrome features ($n = 869$, stepwise distance-based redundancy analysis (dbRDA), $R^2 = 0.24\%$, adjusted P value (P_{adj}) = 0.032; Extended Data Fig. 1b, c). Statin-medicated participants ($n = 106$) were most commonly prescribed simvastatin (48%; 31% atorvastatin, 21% other statins), which had an effect on microbiome variation similar to that of general statin intake (Extended Data Fig. 1d, Supplementary Table 4). To enable an—in terms of medication—least-confounded evaluation of BMI–microbiome associations, statin-medicated participants were excluded from the explorative analyses presented below.

In accordance with the premise of the analysis, within the $n = 782$ non-statin-medicated BMIS cohort (Supplementary Table 1), we found that BMI correlated both with changes in stool consistency (higher BMI values were associated with looser stools, as assessed using the Bristol Stool Scale; $n = 772$, Spearman's $\rho = 0.16$, $P_{adj} = 9.13 \times 10^{-6}$) and with host inflammation markers (for example, serum levels of highly sensitive C-reactive protein (hsCRP), $n = 763$, Spearman's $\rho = 0.70$, $P_{adj} = 1.60 \times 10^{-113}$; Fig. 1a, Supplementary Table 5). Regarding metadata variables that define obesity or metabolic syndrome, only BMI, fat mass percentage and serum fasting triglycerides were found to explain a both significant and non-redundant fraction of compositional microbiome variation ($n = 764$, stepwise dbRDA, $R^2 = 6.22\%$ ($P_{adj} = 1.0 \times 10^{-4}$), 1.15% ($1.0 \times 10^{-4}$) and 0.39% (0.009), respectively; Fig. 1b, c, Supplementary Table 6). All three covariates correlated with microbiome gene richness ($n = 771$, Spearman's $\rho = -0.45$ to -0.26, $P_{adj} = 4.0 \times 10^{-39}$ to 1.6×10^{-13}), a proxy for microbial biodiversity proposed as a marker of metabolic health in obese individuals[8], and with faecal microbial load ($n = 771$, Spearman's $\rho = -0.17$ to -0.13, $P_{adj} = 4.1 \times 10^{-6}$ to 3.1×10^{-4}; Extended Data Fig. 2, Supplementary Table 7, Supplementary Fig. 2). Additionally, BMI, fat mass percentage and triglycerides could all be linked to quantitative variation in specific microbiome features, in terms of composition as well as metabolic potential (Supplementary Table 8). Notable associations included the decrease in *Akkermansia*[13]—which is associated with metabolic health—with increasing BMI ($n = 432$, Spearman's $\rho = -0.23$, $P_{adj} = 6.8 \times 10^{-9}$), alongside an increase in, for example, *Acidaminococcus* spp. ($n = 163$, Spearman's $\rho = 0.23$, $P_{adj} = 5.8 \times 10^{-9}$), a genus that has previously been linked to body mass in a large Korean cohort[14]. The abundance of *Faecalibacterium*—a genus with potential anti-inflammatory properties[15]—was negatively correlated with all three parameters assessed, but was most closely associated with serum triglyceride levels ($n = 753$, Spearman's $\rho = -0.16$, $P_{adj} = 2.5 \times 10^{-4}$). Covariation patterns between BMI, fat mass percentage or triglyceride levels and gut-microbial metabolic modules consisted nearly exclusively of negative correlations (Supplementary Table 8), reflecting the accompanying overall decrease in total microbial load (Supplementary Table 7). Among the features that decrease with all three variables, we highlight that the variation in the butyryl-CoA–acetate CoA-transferase pathway[16]—the most common butyrate production pathway in colon bacteria ($n = 771$, Spearman's $\rho = -0.27$ to

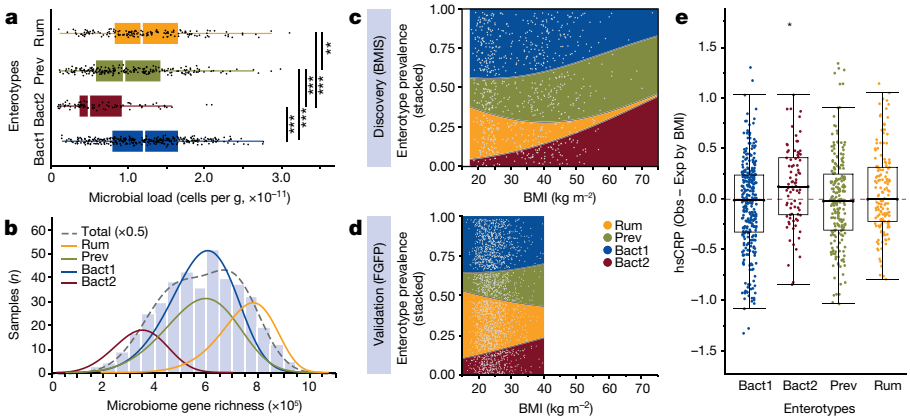


Fig. 2 | Characterization of enterotypes and variation in prevalence with BMI in the non-statin-medicated BMIS cohort. a, Distribution of faecal microbial loads across enterotypes, showing decreased microbial density in the Bact2 enterotype ($n = 771$ biologically independent samples, Kruskal–Wallis with post hoc Dunn test, $^{***}P_{adj} < 0.001$; $^{**}P_{adj} < 0.01$; Supplementary Table 9). **b**, Distribution of gene richness between enterotypes, with low richness samples corresponding to the Bact2 community constellation ($n = 782$ biologically independent samples). **c**, Variation in the prevalence of enterotypes with the BMI of individuals, showing the significant increase in Bact2 prevalence with obesity ($n = 782$ biologically independent samples, binomial logistic regression, Bact2 relative risk = 1.05, $P = 1.2 \times 10^{-7}$). Coloured areas represent the stacked enterotype prevalence along the BMI gradient, with lines provided by multivariate logistic regression of enterotypes by BMI, and data points (light grey) jittered at the corresponding BMI. **d**, Validation of the association between BMI and Bact2 prevalence in the independent FGFP dataset ($n = 2{,}051$ participants, excluding statin-medicated individuals; binomial logistic regression, relative risk = 1.03, $P = 9.4 \times 10^{-3}$). **e**, Inflammatory levels are higher in Bact2 carriers than would be expected on the basis of BMI, as shown by the distribution of residuals of the linear regression between serum CRP and BMI ($n = 763$ biologically independent samples, one-sample location test (dotted line, null hypothesis; mean = 0), Cohen's $d = 0.27$, $^{*}P_{adj} = 0.018$). In **a**, **e**, the body of the box plot represents the first and third quartiles of the distribution, the line represents the median, and the whiskers extend from the quartiles to the last data point within 1.5× IQR, with outliers beyond.

−0.20, $P_{adj} = 3.1 \times 10^{-13}$ to 6.0×10^{-8}; Extended Data Fig. 3a–c)—is in line with previous reports linking this pathway with metabolic health[8]. The metabolism of microbiota-derived butyrate by colonocytes is essential for the maintenance of hypoxic conditions within the colon environment[17], and the disruption of microbial butyrate production has been suggested to induce low-diversity gut microbiota dysbiosis[18].

To investigate a potential association between BMI and the prevalence of faecal microbiome community constellations, we enterotyped the BMIS cohort using Dirichlet multinomial mixtures on genus-level molecular operational taxonomic unit (mOTU) profiles. By applying probabilistic models to group samples that potentially originate from the same community, stratification based on Dirichlet multinomial mixtures reproducibly identifies microbiome constellations across datasets without making any claims regarding the putative discrete nature of the strata detected. Our analyses confirmed previous reports of microbiome variation centred around four enterotypes[1,2,19] (Fig. 1b, Extended Data Fig. 4a, b), hereafter termed Ruminococcaceae (Rum), *Bacteroides*1 (Bact1), *Bacteroides*2 (Bact2) and *Prevotella* (Prev) on the basis of their respective genus-level proportional abundance profiles (Extended Data Fig. 4c). Cell counts differed between enterotypes[1], with the low-richness Bact2 samples ($n = 782$, Kruskal–Wallis, $\chi^2 = 325.65$, $P_{adj} = 5.5 \times 10^{-70}$) also exhibiting the lowest microbial loads ($n = 771$, Kruskal–Wallis, $\chi^2 = 80.14$, $P_{adj} = 2.9 \times 10^{-17}$; Fig. 2a, Supplementary Table 9).

A quantitative compositional and functional analysis of the differences between enterotypes aligned with previous reports[11] (Supplementary Table 10). Further to the findings highlighted above, we found that Bact2 communities displayed the lowest abundances of *Akkermansia* ($n = 771$, Kruskal–Wallis, $\chi^2 = 141.12$, $P_{adj} = 2.0 \times 10^{-29}$) and of *Faecalibacterium* ($n = 771$, Kruskal–Wallis, $\chi^2 = 112.73$, $P_{adj} = 1.7 \times 10^{-23}$), as well as a decreased butyrate production potential ($n = 771$, Kruskal–Wallis, $\chi^2 = 167.12$, $P_{adj} = 4.7 \times 10^{-35}$; Extended Data Fig. 3d). Whereas no significant differences in *Acidaminococcus* levels could be noted between enterotypes ($n = 771$, Kruskal–Wallis, $\chi^2 = 6.47$, $P_{adj} = 0.12$), taxa such as *Eggerthella*—a genus that is considered part of a normal microbiota but is also linked to gastrointestinal infections as well as bacteraemia[20]—was found to occur in higher absolute numbers against the background of overall reduced microbial load, as observed in Bact2 communities ($n = 771$, Kruskal–Wallis, $\chi^2 = 224.95$, $P_{adj} = 4.1 \times 10^{-47}$; Extended Data Fig. 5a, b, Supplementary Table 10). Co-abundance gene group analyses additionally indicated enterotype differentiation at the species level (Supplementary Table 11). For example, in Bact2-enterotyped communities, the *Bacteroides* fraction was observed to be proportionally depleted in *Bacteroides caccae* ($n = 768$, Kruskal–Wallis, $\chi^2 = 78.40$, $P_{adj} = 1.3 \times 10^{-15}$) and *Bacteroides cellulosilyticus* ($n = 768$, Kruskal–Wallis, $\chi^2 = 64.79$, $P_{adj} = 5.3 \times 10^{-13}$) when compared with Rum, Prev and Bact1 samples. By contrast, it seemed to be enriched in *Bacteroides fragilis* ($n = 768$, Kruskal–Wallis, $\chi^2 = 65.26$, $P_{adj} = 3.5 \times 10^{-11}$; Extended Data Fig. 6, Supplementary Table 11), which is considered to be among the most pathogenic and immunomodulatory of the *Bacteroides* species[21,22].

The prevalence of enterotypes along a gene-richness axis in the non-statin-medicated cohort confirmed previous observations of a bimodal distribution[8]; however, the Bact2 community type enabled further refinement of richness stratifications through the deconvolution of overlapping peaks (Fig. 2b). The prevalence of Bact2 was found to increase with BMI, from 3.90% among lean or overweight participants (BMI < 30) to 17.73% among obese participants (BMI ≥ 30) ($n = 782$, binomial logistic regression, relative risk = 1.05, $P = 1.2 \times 10^{-7}$, where relative risk can be interpreted as the scale factor necessary to obtain the prevalence of the Bact2 enterotype after a unit increase in BMI; Fig. 2c; Supplementary Table 12). Notwithstanding methodological differences, this finding was validated in the independent, amplicon-sequenced Flemish Gut Flora Project[10] dataset (FGFP, $n = 2{,}051$; excluding statin-medicated participants; binomial logistic regression, relative risk = 1.03, $P = 9.3 \times 10^{-3}$; Fig. 2d). In line with previous findings from the FGFP[2], Bact2 hosts from the BMIS cohort displayed more pronounced systemic inflammation levels when compared to non-Bact2 participants, here assessed through serum hsCRP concentrations (Kruskal–Wallis, $\chi^2 = 48.61$, $P = 1.37 \times 10^{-10}$; Extended Data Fig. 7a; Supplementary Table 13). Notably, the inflammatory tone of Bact2 hosts exceeded the levels anticipated on the basis of their obesity status ($n = 86$, one-sample location test on residuals of non-statin-medicated BMIS linear regression

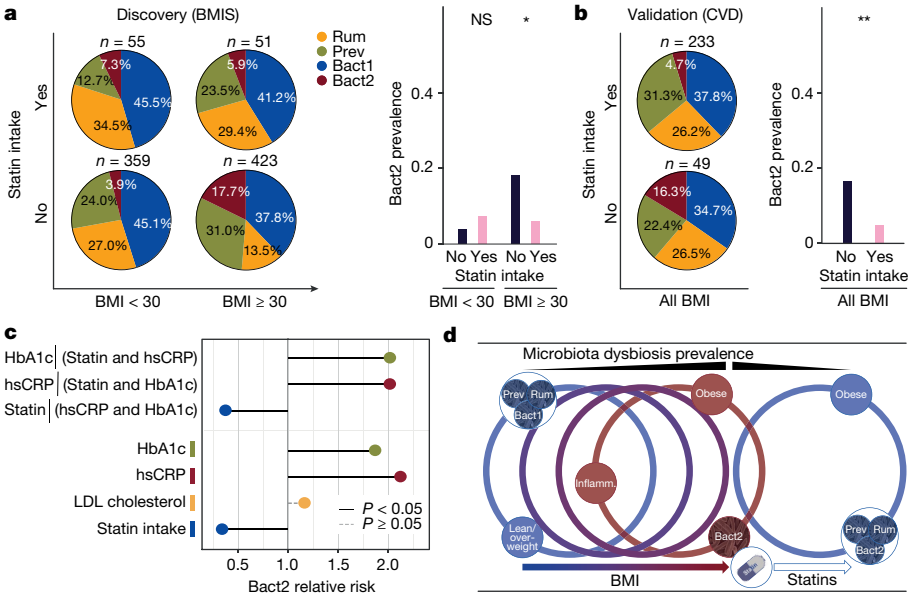


Fig. 3 | Association between the prevalence of the Bact2 enterotype, obesity and statin intake. a, Bact2 prevalence in obese (BMI ≥ 30) compared with lean and overweight (BMI < 30) individuals in the BMIS cohort ($n = 888$), stratified according to statin medication status. Statin-medicated obese individuals display a significantly lower prevalence of Bact2 as compared with non-statin-medicated individuals (bar plots; statin-medicated versus non-statin-medicated, 5.88% versus 17.73%, $n = 888$ biologically independent samples, Fisher's two-tail exact test, log likelihood = −2.88, $^{*}P = 0.028$). **b**, The lower Bact2 prevalence among statin-medicated compared with non-statin-medicated individuals is validated in the MetaCardis cardiovascular disease cohort, comprising $n = 282$ non-diabetic patients with cardiovascular disease (CVD; statin-medicated versus non-statin-medicated, 4.72% versus 16.33%, $n = 282$ biologically independent samples, Fisher's two-tail exact test, log likelihood = −3.47, $^{**}P = 0.008$). **c**, Relative risk of obese BMIS individuals ($n = 474$ participants) harbouring a Bact2 enterotype as a function of statin intake and serum biomarkers for potential (side) effects of statins (lipidemic control (LDL-cholesterol), inflammatory modulation (hsCRP) and glucose regulation (HbA1c)). Variables were modelled independently or together in univariate or multivariate models, respectively (Supplementary Table 19). The latter suggests that statin intake remains associated with a reduction in dysbiosis risk after partialing-out hsCRP and HbA1c ($n = 462$ biologically independent samples, multivariate binomial logistic regression, Statin | (hsCRP and HbA1c) relative risk = 0.36, $P_{adj} = 0.039$). Adjustment for multiple testing (P_{adj}) was performed on univariate tests using the Benjamini–Hochberg method (represented by black lines when significant ($P_{adj} < 0.05$), or otherwise a dashed grey line ($P_{adj} = 0.15$)). **d**, Graphical summary of the main results regarding the prevalence of the Bact2 enterotype, BMI and statin intake. In the present BMIS cohort, we identify Bact2 as an inflammation-associated microbiome community constellation, with increasing prevalence along a BMI gradient in non-statin-medicated individuals. Statin therapy is associated with attenuated inflammation and a Bact2 prevalence comparable to that observed among lean and overweight subjects. Circles represent individual host configurations in terms of body mass, microbiota community type, and inflammation status.

between hsCRP and BMI, Cohen's $d = 0.27$, $P_{adj} = 0.018$; Fig. 2e, Extended Data Fig. 7b, Supplementary Table 14). In a multivariate model, the BMI and the Bact2 carrier status of the participants both provided a non-redundant contribution to increased systemic inflammation levels, corresponding to a 1.04 ($n = 763$, linear multivariate model, $P_{adj} = 2.2 \times 10^{-16}$) and a 1.16 ($P_{adj} = 0.004$) unit increase risk in serum hsCRP levels, respectively (Supplementary Table 15). These observations support the qualification of the Bact2 microbiota configuration as an (low-grade) inflammation-associated, potentially dysbiotic enterotype[1,2].

Whether initiating or sustaining pro-inflammatory processes and metabolic derailment, countering dysbiosis of the gut ecosystem has been suggested to contribute to the maintenance of host health and the containment of obesity-related comorbidities. However, no effective microbiome modulation strategy has yet been established. Here, within the limitations of the cross-sectional study design, we identify statin treatment as a potential lever in the management of dysbiosis. In contrast to the findings from the non-statin-medicated participants, we observed that Bact2 prevalence no longer significantly increased with BMI in statin-medicated individuals ($n = 106$, binomial logistic regression, relative risk = 1.03, $P = 0.60$). Among obese individuals, only 5.88% of statin-medicated individuals were enterotyped as Bact2, compared with 17.73% of non-statin-medicated participants (Fisher's two-tail exact test, log likelihood = −2.88, $P = 0.028$; Fig. 3a, Supplementary Table 16). When exploring whether accounted clinical parameters, anticipated treatment responses, co-medication or key microbiome covariates[10] could be associated with the observed differences in Bact2 prevalence, we noted that statin-medicated obese participants displayed ameliorated lipid metabolism (low-density lipoprotein (LDL)-cholesterol, $n = 473$, Mann–Whitney U-test, $r = -0.17$, $P_{adj} = 0.002$) and inflammation status (hsCRP, $n = 462$, Mann–Whitney U-test, $r = -0.23$, $P_{adj} = 8.4 \times 10^{-6}$; Supplementary Table 17)—both expected outcomes of statin therapy[23]. Besides minor differences in the incidence of concomitant drug intake (notably aspirin intake being more prevalent among statin-medicated participants; $n = 474$, Fisher's two-tailed exact test, log likelihood = −17.36, $P_{adj} = 2.2 \times 10^{-7}$) and glucose metabolism (lower HbA1c levels among non-statin-medicated participants, $n = 474$, Mann–Whitney U-test, $r = 0.17$, $P_{adj} = 0.001$)—the latter being a known side effect of statin treatment[24]—the statin-medicated subcohort was characterized as older (median age statin-medicated versus non-statin-medicated, 61 versus 47; $n = 474$, Mann–Whitney U-test, $r = 0.34$, $P_{adj} = 1.4 \times 10^{-11}$) and less obese (BMI 33.5 versus 40.8; $n = 474$, Mann–Whitney U-test, $r = -0.25$, $P_{adj} = 2.1 \times 10^{-6}$). However, among these significant covariates, and excluding variables that reflect pleiotropic effects of statins—that is, levels of LDL-cholesterol and inflammation markers—only statin intake and blood HbA1c levels were shown to have a significant, non-redundant explanatory power for Bact2 prevalence (Supplementary Table 18), with the latter being associated with an increased probability of Bact2 carrier status ($n = 472$, multivariate binomial logistic regression, statin intake relative risk = 0.31, $P_{adj} = 0.013$; HbA1c relative risk = 2.00, $P_{adj} = 0.009$). Although 41% of

BMIS participants reported taking non-statin drugs, (co-)medication status did not affect the outcome of Bact2 prevalence analyses in obese participants (Extended Data Fig. 8). Low prevalence of the Bact2 enterotype among statin-medicated individuals was validated in the accompanying MetaCardis cardiovascular disease dataset (non-diabetic patients with cardiovascular disease (CVD); Bact2 prevalence among statin-medicated versus non-statin-medicated participants, 4.72% versus 16.33%; $n = 282$, Fisher's two-tailed exact test, log likelihood = −3.47, $P = 0.008$; Fig. 3b, Supplementary Table 16). Here—and in accordance with observations in non-CVD disease cohorts[1,2]—increased Bact2 prevalence was not limited to obese non-statin-medicated patients with CVD, but could also be noted within the non-statin-medicated lean and overweight subgroup. In the independent FGFP dataset—which targets an average representation of a Western population, and therefore covers a narrower BMI spectrum ($n = 2{,}345$; median BMI = 24.2, range = 16–40)—we confirmed lowered Bact2 prevalence among statin-medicated individuals given their BMI ($n = 2{,}345$, multivariate binomial logistic regression, Statin | BMI, relative risk = 0.72, $P_{adj} = 0.045$; Extended Data Fig. 9, Supplementary Table 16). Additional evidence—which is indicative of causality in statin-associated microbiota variation—is provided by a recent small-scale intervention study in a rat model, which demonstrates reversion of microbiota alterations induced by a high-fat diet and hypercholesterolemia upon treatment with atorvastatin, resulting in an increased microbiome richness[25]. Although caution should be applied when extrapolating findings from the rodent microbiome to a human context, these results do demonstrate directionality in statin–microbiota associations, although the effect of atorvastatin (31% of statin-medicated participants) in the present BMIS cohort did not reach statistical significance (Extended Data Fig. 1, Supplementary Table 4).

The cross-sectional nature of the MetaCardis dataset did not enable us to establish a causal chain of events that lead to a lower prevalence of the Bact2 enterotype among statin-medicated individuals. Given the putatively independent effects of statin therapy on levels of serum hsCRP and LDL-cholesterol[23], we modelled the association of both variables with Bact2 prevalence for obese participants in the BMIS cohort. Although no significant effect of LDL-cholesterol concentrations was found ($n = 473$, univariate binomial logistic regression, LDL-cholesterol relative risk = 1.16, $P_{adj} = 0.15$), lower hsCRP levels were associated with a lower prevalence of the Bact2 enterotype ($n = 462$, univariate binomial logistic regression, hsCRP relative risk = 2.11, $P_{adj} = 0.003$; Supplementary Table 19). A multivariate model for the prediction of Bact2 prevalence—which covers treatment (statin intake), treatment outcome (hsCRP levels), as well as side effects of treatment (HbA1c concentrations)—indicated a significant additive contribution of statin therapy to the reduction of dysbiosis risk ($n = 462$, multivariate binomial logistic regression, Statin | (hsCRP and HbA1c) relative risk = 0.36, $P_{adj} = 0.039$; Fig. 3c, Extended Data Fig. 10, Supplementary Table 19); this suggests that the effect of statins is greater than solely the attenuating effect on the inflammation status of the host. Nevertheless, the pleiotropic effect of statins on microbiome community constellations seemed to be closely associated with a concomitant effect on host inflammation levels. At this point, at least two mechanistic interpretations of our observations—or a combination of both—remain possible (Fig. 3d). On one hand, aligning with the microbiota–inflammation hypothesis, statins could counteract the microbial contribution to inflammatory and metabolic obesity comorbidities through (in)direct modulation of the microbiota. Consistent with this, in vitro studies have demonstrated that statins affect the growth of several gut microorganisms[26]. Conversely, the demonstrated anti-inflammatory effects of statins could alleviate gut host–microbial interactions and enable the subsequent development of enterotypes that are not associated with inflammation. However, it should be stressed that the cross-sectional design of our study does not allow us to rule out potential confounding by indication (lower Bact2 prevalence resulting from the specific condition that prompted statin prescription) or by unaccounted diagnosis-associated diet or lifestyle alterations (participants adopting health-promoting and/or microbiota-modulating activities complementary to statin therapy).

For many years, strategies for the modulation of microbiota have revolved around (next-generation) probiotics and prebiotics—introducing or promoting the growth of beneficial bacteria or bacterial consortia. It is only recently that a revived interest in the effect of small molecules and drugs on the colon ecosystem, as well as individual faecal isolates, has been noted[26,27]. Although we cannot rule out a potential effect of unaccounted confounders, nor can we infer causality from the associations observed, our analyses reveal that statin therapy is linked with a lowered prevalence of a pro-inflammatory microbial community type in obese individuals. Our results align well with previous, sparse reports of a beneficial effect of statins in pathologies in which a role of the gut microbiota has been postulated[28]—including interventional[29] as well as epidemiological[30] evidence in Crohn's disease, a condition that has previously been linked to a high prevalence of Bact2[1,2]. Within the limitations of the cross-sectional nature of the cohorts analysed—and emphasizing the need for interventional follow-up research using a randomized, double-blind, placebo-control study design to exclude potential confounding by indication—our findings suggest statins as a possible target for the development of future drug-based strategies for the modulation of the intestinal microbiota.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-020-2269-x.

1. Vandeputte, D. et al. Quantitative microbiome profiling links gut community variation to microbial load. *Nature* **551**, 507–511 (2017).
2. Vieira-Silva, S. et al. Quantitative microbiome profiling disentangles inflammation- and bile duct obstruction-associated microbiota alterations across PSC/IBD diagnoses. *Nat. Microbiol.* **4**, 1826–1831 (2019).
3. Probert, C. S., Emmett, P. M. & Heaton, K. W. Some determinants of whole-gut transit time: a population-based study. *QJM* **88**, 311–315 (1995).
4. Ford, E. S. Body mass index, diabetes, and C-reactive protein among U.S. adults. *Diabetes Care* **22**, 1971–1977 (1999).
5. Turnbaugh, P. J. et al. An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444**, 1027–1031 (2006).
6. Sze, M. A. & Schloss, P. D. Looking for a signal in the noise: revisiting obesity and the microbiome. *MBio* **7**, e01018-16 (2016).
7. Walters, W. A., Xu, Z. & Knight, R. Meta-analyses of human gut microbes associated with obesity and IBD. *FEBS Lett.* **588**, 4223–4233 (2014).
8. Le Chatelier, E. et al. Richness of human gut microbiome correlates with metabolic markers. *Nature* **500**, 541–546 (2013).
9. Karlsson, F. H. et al. Symptomatic atherosclerosis is associated with an altered gut metagenome. *Nat. Commun.* **3**, 1245 (2012).
10. Falony, G. et al. Population-level analysis of gut microbiome variation. *Science* **352**, 560–564 (2016).
11. Vieira-Silva, S. et al. Species-function relationships shape ecological properties of the human gut microbiome. *Nat. Microbiol.* **1**, 16088 (2016).
12. Alberti, K. G. M. M., Zimmet, P. & Shaw, J. Metabolic syndrome—a new world-wide definition. A consensus statement from the International Diabetes Federation. *Diabet. Med.* **23**, 469–480 (2006).
13. Depommier, C. et al. Supplementation with *Akkermansia muciniphila* in overweight and obese human volunteers: a proof-of-concept exploratory study. *Nat. Med.* **25**, 1096–1103 (2019).
14. Yun, Y. et al. Comparative analysis of gut microbiota associated with body mass index in a large Korean cohort. *BMC Microbiol.* **17**, 151 (2017).
15. Quévrain, E. et al. Identification of an anti-inflammatory protein from *Faecalibacterium prausnitzii*, a commensal bacterium deficient in Crohn's disease. *Gut* **65**, 415–425 (2016).
16. Louis, P. et al. Restricted distribution of the butyrate kinase pathway among butyrate-producing bacteria from the human colon. *J. Bacteriol.* **186**, 2099–2106 (2004).
17. Litvak, Y., Byndloss, M. X. & Bäumler, A. J. Colonocyte metabolism shapes the gut microbiota. *Science* **362**, eaat9076 (2018).
18. Kriss, M., Hazleton, K. Z., Nusbacher, N. M., Martin, C. G. & Lozupone, C. A. Low diversity gut microbiota dysbiosis: drivers, functional implications and recovery. *Curr. Opin. Microbiol.* **44**, 34–40 (2018).
19. Ding, T. & Schloss, P. D. Dynamics and associations of microbial community types across the human body. *Nature* **509**, 357–360 (2014).

20. Gardiner, B. J. et al. Clinical and microbiological characteristics of *Eggerthella lenta* bacteremia. *J. Clin. Microbiol.* **53**, 626–635 (2015).
21. Mazmanian, S. K., Liu, C. H., Tzianabos, A. O. & Kasper, D. L. An immunomodulatory molecule of symbiotic bacteria directs maturation of the host immune system. *Cell* **122**, 107–118 (2005).
22. Wexler, H. M. Bacteroides: the good, the bad, and the nitty-gritty. *Clin. Microbiol. Rev.* **20**, 593–621 (2007).
23. Ridker, P. M. et al. Reduction in C-reactive protein and LDL cholesterol and cardiovascular event rates after initiation of rosuvastatin: a prospective study of the JUPITER trial. *Lancet* **373**, 1175–1182 (2009).
24. Muscogiuri, G. et al. The good and bad effects of statins on insulin sensitivity and secretion. *Endocr. Res.* **39**, 137–143 (2014).
25. Khan, T. J. et al. Effect of atorvastatin on the gut microbiota of high fat diet-induced hypercholesterolemic rats. *Sci. Rep.* **8**, 662 (2018).
26. Maier, L. et al. Extensive impact of non-antibiotic drugs on human gut bacteria. *Nature* **555**, 623–628 (2018).
27. Forslund, K. et al. Disentangling type 2 diabetes and metformin treatment signatures in the human gut microbiota. *Nature* **528**, 262–266 (2015).
28. Zeiser, R. Immune modulatory effects of statins. *Immunology* **154**, 69–75 (2018).
29. Grip, O., Janciauskiene, S. & Bredberg, A. Use of atorvastatin as an anti-inflammatory treatment in Crohn's disease. *Br. J. Pharmacol.* **155**, 1085–1092 (2008).
30. Ungaro, R. et al. Statins associated with decreased risk of new onset inflammatory bowel disease. *Am. J. Gastroenterol.* **111**, 1416–1423 (2016).

[1]Laboratory of Molecular Bacteriology, Department of Microbiology and Immunology, Rega Institute, KU Leuven, Leuven, Belgium. [2]Center for Microbiology, VIB, Leuven, Belgium. [3]Nutrition and Obesities: Systemic Approaches Research Unit (NutriOmics), INSERM, Sorbonne Université, Paris, France. [4]Institute of Cardiometabolism and Nutrition, Integromics Unit, Paris, France. [5]Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. [6]Nutrition Department, Pitie-Salpêtrière Hospital, Assistance Publique Hôpitaux de Paris, Paris, France. [7]Medical Department III – Endocrinology, Nephrology, Rheumatology, University of Leipzig Medical Center, Leipzig, Germany. [8]Experimental and Clinical Research Center, Charité-Universitätsmedizin and Max-Delbrück Center, Berlin, Germany. [9]Structural and Computational Biology, European Molecular Biology Laboratory, Heidelberg, Germany. [10]Max Delbruck Center for Molecular Medicine in the Helmholtz Association (MDC), Berlin, Germany. [11]DZHK (German Centre for Cardiovascular Research), Partner Site Berlin, Berlin, Germany. [12]Unité de Modélisation Mathématique et Informatique des Systèmes Complexes, UMMISCO, Sorbonne Université, IRD, Bondy, France. [13]Wallenberg Laboratory, Department of Molecular and Clinical Medicine and Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden. [14]Université Paris-Saclay, INRAE, Metagenopolis, Jouy en Josas, France. [15]Diabetes Department, Pitie-Salpêtrière Hospital, Assistance Publique Hôpitaux de Paris, Paris, France. [16]UF Biomarqueurs Inflammatoires et Métaboliques, Biochemistry and Hormonology Department, Tenon Hospital, Assistance Publique Hôpitaux de Paris, Paris, France. [17]Centre de Recherche Saint-Antoine, Sorbonne Université-INSERM UMR-S 938, IHU ICAN, Paris, France. [18]Institute of Science and Technology for Brain-Inspired Intelligence, Fudan University, Shanghai, China. [19]NICO Cardio-oncology Program, CIC-1421, Department of Pharmacology, Pitié-Salpêtrière Hospital, Assistance Publique Hôpitaux de Paris, INSERM, Sorbonne Université, Paris, France. [20]Université de Paris, PARCC, INSERM, Paris, France. [21]CIC1418 and DMU CARTE, Assistance Publique–Hôpitaux de Paris, Hôpital Européen Georges-Pompidou, Paris, France. [22]Department of Cardiology, Rigshospitalet, University of Copenhagen, Copenhagen, Denmark. [23]Computational and Systems Medicine, Department of Metabolism, Digestion and Reproduction, Faculty of Medicine, Imperial College London, London, UK. [24]Genomic and Environmental Medicine, National Heart & Lung Institute, Faculty of Medicine, Imperial College London, London, UK. [25]Sorbonne Paris Cité Epidemiology and Statistics Research Centre (CRESS), U1153 INSERM, U1125, INRA, CNAM, University of Paris, Nutritional Epidemiology Research Team (EREN), Bobigny, France. [26]Department of Clinical Biochemistry, Rigshospitalet, University of Copenhagen, Copenhagen, Denmark. [27]Biobyte Solutions, Heidelberg, Germany. [28]Department of Biology and Biological Engineering, Chalmers University of Technology, Gothenburg, Sweden. [29]Helmholtz Institute for Metabolic, Obesity and Vascular Research (HI-MAG) of the Helmholtz Zentrum München at the University of Leipzig, Leipzig, Germany. [30]Molecular Medicine Partnership Unit, University of Heidelberg and European Molecular Biology Laboratory, Heidelberg, Germany. [31]Department of Bioinformatics, Biocenter, University of Würzburg, Würzburg, Germany. [36]These authors contributed equally: Sara Vieira-Silva, Gwen Falony, Eugeni Belda. [37]These authors jointly supervised this work: Karine Clément, Jeroen Raes. *A list of participants and their affiliations appears at the end of the paper. [✉]e-mail: karine.clement@inserm.fr; jeroen.raes@kuleuven.be

MetaCardis Consortium

Renato Alves[9], Chloe Amouyal[3,15], Ehm Astrid Andersson Galijatovic[5], Olivier Barthelemy[32], Jean-Paul Batisse[32], Magalie Berland[14], Randa Bittar[33], Hervé Blottière[14], Frederic Bosquet[15], Rachid Boubrit[32], Olivier Bourron[15], Mickael Camus[14], Dominique Cassuto[6], Cecile Ciangura[6,15], Jean-Philippe Collet[32], Maria-Carlota Dao[3], Jean Debedat[3], Morad Djebbar[32], Angélique Doré[14], Line Engelbrechtsen[5], Soraya Fellahi[16,17], Sebastien Fromentin[14], Philippe Giral[34], Marianne Graine[4], Agnes Hartemann[15], Bolette Hartmann[5], Gerard Helft[32], Serge Hercberg[25], Malene Hornbak[5], Richard Isnard[32], Sophie Jaqueminet[15], Niklas Rye Jørgensen[35], Hanna Julienne[14], Johanne Justesen[5], Judith Kammer[7], Mathieu Kerneis[32], Jean Khemis[6], Nikolaj Krarup[5], Michael Kuhn[9], Aurélie Lampuré[3], Véronique Lejard[14], Florence Levenez[14], Lea Lucas-Martini[6], Robin Massey[14], Nicolas Maziers[14], Jonathan Medina-Stamminger[6], Lucas Moitinho-Silva[9], Gilles Montalescot[32], Sandrine Moutel[6], Laetitia Pasero Le Pavin[14], Christine Poitou-Bernert[3,6], Francoise Pousset[32], Laurence Pouzoulet[34], Sebastian Schmidt[9], Johanne Silvain[32], Mathilde Svendstrup[5], Timothy Swartz[3,4], Thierry Vanduyvenboden[14], Camille Vatier[6], Eric Verger[3] & Stefanie Walther[7]

[32]Cardiology Department, Pitie-Salpêtrière Hospital, Assistance Publique Hôpitaux de Paris, Paris, France. [33]Biochemistry Department, Pitie-Salpêtrière Hospital, Assistance Publique Hôpitaux de Paris, Paris, France. [34]Endocrinology Department, Pitie-Salpêtrière Hospital, Assistance Publique Hôpitaux de Paris, Paris, France. [35]Department of Clinical Biochemistry, Rigshospitalet, Glostrup University of Copenhagen, Copenhagen, Denmark.

Methods

Data reporting

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Sample collection

Ethical compliance. Ethical approval was obtained from the Ethics Committee CPP Ile-de France, Ethics Committee at the Medical Faculty at the University of Leipzig, and the Ethical Committees of the Capital Region of Denmark. The study protocol (also comprising an interventional arm which is not part of the analysis presented) was registered at https://clinicaltrials.gov (study number NCT02059538). The study design (observational cohort study) complied with all relevant ethical regulations, aligning with the Helsinki Declaration and in accordance with European privacy legislation. All participants provided written informed consent.

Study cohort. The $n = 888$ transnational Body Mass Index spectrum (BMIS) cohort was assembled as part of the overall MetaCardis recruitment efforts (Supplementary Fig. 1). Participants were recruited between 2013 and 2015 in the clinical departments of the Pitié-Salpêtrière Hospital (Paris, France), the Integrated Research and Treatment Center for Adiposity Diseases (Leipzig, Germany), and in the Novo Nordisk Foundation Center for Basic Metabolic Research (Copenhagen, Denmark). Potential participants were evaluated for suitability according to standardized inclusion and exclusion criteria across the three recruitment centres. Exclusion criteria included history of abdominal cancer/radiation therapy on the abdomen, history of intestinal resection (except for appendectomy), acute or chronic inflammatory or infectious diseases (including hepatitis C virus, hepatitis B virus and HIV), history of organ transplantation or receiving immunosuppressive therapy, severe kidney failure (MDRD glomerular filtration rate < 50 ml (min $1.73m^2)^{-1}$), or drug or alcohol addiction. All study participants had to be free of any antibiotic use in the three months before inclusion. The BMIS ($n = 888$) cohort consisted of a MetaCardis sub-cohort, defined by exclusion of cardiovascular patients (defined in the MetaCardis consortium study protocol as patient groups 4, 5, 6 and 7) and any individual with type-2 diabetes. Diagnosis of type-2 diabetes was defined using the American Diabetes Association definition: fasting glycemia >6.9 mmol l^{-1} and/or 2 h values in the oral glucose tolerance test >11 mmol l^{-1} and/or haemoglobin A1c (HbA1c, glycated haemoglobin) ≥ 6.5% and/or use of any antidiabetic treatment. The MetaCardis project sample size calculation was focused on the objectives of multi-omics integration and metagenomic-wide metabolome-wide association study (MW2AS) across groups of patients ranging different cardiometabolic phenotypes. On the basis of unpublished data from consortium partners, a sample size of 2,000 individuals was deemed required to detect a significant association (with and without concomitant risk factors). No specific sample size calculation was performed before BMIS sub-cohort recruitment. On the basis of the baseline prevalence of Bact2 enterotype (with baseline defined as lean/overweight individuals $P(\text{Bact2}) = 14\%$) in the amplicon-sequenced FGFP cohort, the present study cohort size enabled us to identify a minimum difference of 7.4% in Bact2 prevalence between the two groups: lean or overweight ($n = 414$) versus obese ($n = 474$) as significant (power = 80%, alpha = 0.05).

Validation cohorts. MetaCardis cardiovascular disease (CVD, $n = 282$). The CVD cohort was recruited as described above as part of the MetaCardis cohort, and corresponds to patients with cardiovascular disease and without diabetes, defined in the MetaCardis consortium study protocol as patient groups 4, 5, 6 and 7. Flemish Gut Flora Project (FGFP, $n = 2,345$). The FGFP cohort is part of a population-level cross-sectional sampling of the Flemish population described in ref. [10] and re-sequenced with dual-indexed HiSeq amplicon sequencing as analysed in ref. [31]. Ethical approval for the FGFP sampling was granted by the Commissie Medische Ethiek UZ-VUB (B.U.N.143201215505) and the Ethische Commissie Onderzoek UZ/KU Leuven (S58125). The inclusion and exclusion criteria defined for recruitment of the MetaCardis cohort and, more specifically, the BMIS subset, were applied to the FGFP: inclusion age between 18 and 75 years old, exclusion of acute or chronic inflammatory or infectious diseases (notably diagnosis of inflammatory bowel disease and recent gastroenteritis), and exclusion of patients with diabetes—defined as having a diagnosis of diabetes or increased glycated haemoglobin A1c levels (HbA1c ≥ 6.5%), or the use of any antidiabetic treatment. The disease diagnoses used for exclusion were reported by the general practitioners of the participants. The medical questionnaire and blood sampling for analysis (including HbA1c) were performed within one week of faecal sampling.

Sample collection. Faeces were collected according to International Human Microbiome Standards (IHMS) guidelines (modified SOP 04 V1 (collection without anaerobic bag)). In brief, participants were handed a collection kit, collected samples at home, and stored them temporarily (less than 48 h) at −20 °C until they were transported frozen (on dry ice) to the collection centre (Pitié-Salpêtrière Hospital (France), University Hospital of Leipzig (Germany) or Frederiksberg Hospital (Denmark)). Blood samples were collected during the clinical examination visit after overnight fasting.

Metadata collection. Participant phenotyping was performed according to standardized operational procedures and included the acquisition of biological samples and the assessment of clinical parameters and anthropometrics including age, gender, smoking status, weight, height, BMI, blood pressure, body composition, and waist and hip circumference measurements. Body fat mass and fat-free mass were determined through bioelectrical impedance analysis. Systolic and diastolic blood pressure were measured using a mercury sphygmomanometer (measures were taken three times on each arm; the mean of the last two measurements on the right arm was used for analyses). During the interview at the clinical visit, a detailed list of prescribed medications (based on direct recall or medication list when provided) as well as the medical history of the patient was compiled. Subjects were questioned on adherence to their medication plan. Five-year antibiotic intake was assessed by recall in France and Denmark, whereas participants in Germany were requested to provide medication anamnesis from their general practitioners or physicians (drugs prescribed over the past five years). All medication data was curated jointly by the study physicians at each centre so as to harmonize presentation. The metadata necessary for reproducing the results presented in the article are available in Supplementary Table 2.

Sample analyses

Blood analyses. Blood metabolic markers were assessed in local routine laboratories. Analyses of adipokines, measures of glycaemia, inflammatory markers, and free fatty acids were centralized; plasma and serum samples were stored at the respective clinical centres at −80 °C until shipment to a central measuring facility. Blood cell counts (leukocytes, monocytes, neutrophils and immune cells) were measured using flow cytometry as described previously[32]. Fasting glucose, total cholesterol, high-density-lipoprotein cholesterol, triglycerides and HbA1c were measured using enzymatic methods. LDL-cholesterol concentrations were measured enzymatically for German participants; values for French and Danish subjects were calculated using the Friedwald equation. Kinetic assays based on coupled enzyme systems were used to measure alanine aminotransferase, aspartate aminotransferase and γ-glutamyltransferase levels. Free fatty acid concentrations were assessed by photometrics (Diasys Diagnostic Systems). A chemiluminescence assay (Insulin Architect, Abbott) was used to measure

serum insulin and C-peptide levels in a fasting state and at 30 and 120 min during an oral glucose tolerance test. Serum leptin was determined using the Human Leptin Quantikine ELISA Kit (R&D Systems); adiponectin was measured using an ELISA sandwich assay (HMW & Total Adiponectin ELISA Kit, ALPCO). Levels of hsCRP were determined by an IMMAGE automatic immunoassay system (Beckman-Coulter). Blood concentrations of high-sensitivity interleukin 6 (hsIL6) and CD14 were measured using the Human IL-6 Quantikine HS and the Human Quantikine ELISA Kit (R&D Systems), respectively. A Luminex assay (ProcartaPlex Mix&Match Human 13-plex, eBioscience) was set up to measure the following cytokines: interferon gamma-induced protein 10 (IP-10), C-X-C motif chemokine ligand 5 (CXCL5), CC-Chemokin ligand 2 (CCL2), Eotaxine, Interleukine 7 (IL-7), macrophage migration inhibitory factor (MIF), macrophage inflammatory protein 1β (MIP 1β), stromal cell-derived factor 1 (SDF1) and vascular endothelial growth factor A (VEGFA).

Metagenomic analyses of faecal samples. Total faecal DNA was extracted following the International Human Microbiome Standards (IHMS) guidelines (SOP 07 V2 H) and sequenced using an Ion proton system (Thermo Fisher Scientific) resulting in 23.3 ± 4.0 million (mean ± s.d.) 150-bp single-end reads per sample on average. Reads were cleaned using AlienTrimmer (v0.2.4)[33] to remove resilient sequencing adapters and to trim low quality nucleotides at the 3′ side (quality and length cut-off of 20 and 45 bp, respectively). Cleaned reads were subsequently filtered from human and potential food contaminant DNA (using human genome RCh37-p10, *Bos taurus* and *Arabidopsis thaliana* with an identity score threshold of 97%). Gene abundance profiling was performed using the 9.9-million-gene integrated reference catalogue of the human microbiome[34]. Filtered high-quality reads were mapped with an identity threshold of 95% to the 9.9-million-gene catalogue using BowTie (v.2.2.6) included in the METEOR software[35]. A gene abundance table was generated by means of a two-step procedure using METEOR. First, the uniquely mapping reads (reads mapping to a single gene in the catalogue) were attributed to their corresponding genes. Second, shared reads (reads that mapped with the same alignment score to multiple genes) were attributed according to the ratio of their unique mapping counts. The gene abundance table was processed for rarefaction and normalization and further analysis using the R package MetaOMineR[36]. To decrease technical bias due to different sequencing depth and avoid any artefacts of sample size on low-abundance genes, read counts were rarefied. The gene abundance table was rarefied to 10 million reads per sample by random sampling of 10 million mapped reads without replacement. The resulting rarefied gene abundance table was normalized according to the FPKM (fragments per kilobase of transcript per million mapped reads) strategy (normalization by the gene size and the number of total mapped reads reported in frequency) to give the gene abundance profile table and binned by functional and phylogenetic categories as carried out within the MOCAT2 framework[37]. 1,436 metagenomic species (MGS; co-abundant gene groups with more than 500 genes corresponding to microbial species) were clustered from 1,267 human gut metagenomes used to construct the 9.9-million-gene catalogue[34], as described previously[38]. MGS abundances were estimated as the mean abundance of the 50 genes defining a robust centroid of the cluster (if more than 10% these genes gave positive signals). MGS taxonomical annotation was performed using all genes by sequence similarity using NCBI blast N; a species-level assignment was given if more than 50% of the genes matched the same reference genome of the NCBI database (November 2016 version) at a threshold of 95% of identity and 90% of gene length coverage. The remaining MGS were assigned to a given taxonomical level from genus to superkingdom if more than 50% of their genes had the same level of assignment. Microbial gene richness (gene count) was calculated by counting the number of genes that were detected at least once in a given sample, using the average number of genes counted in ten independent rarefaction experiments.

Determination of faecal microbial load. Microbial loads of faecal samples of were determined as described previously[1,2]. In brief, 0.2 g frozen (−80 °C) aliquots were dissolved in physiological solution (9 g l^{-1} NaCl; Baxter S.A.) to a total volume of 100 ml. Subsequently, the slurry was diluted 1,000 times. Samples were filtered using a sterile syringe filter (pore size 5 µm; Sartorius Stedim Biotech). Next, 1 ml of the microbial cell suspension obtained was stained with 1 µl SYBR Green I (1:100 dilution in DMSO; shaded 15 min incubation at 37 °C; 10,000 concentrate, Thermo Fisher Scientific). The flow cytometry analysis was performed using a C6 Accuri flow cytometer (BD Biosciences)[39]. Fluorescence events were monitored using the FL1 533/30 nm and FL3 > 670 nm optical detectors. In addition, forward and sideward-scattered light was also collected. The BD Accuri CFlow (v.1.0.264.21) software was used to gate and separate the microbial fluorescence events on the FL1/FL3 density plot from the faecal sample background. A threshold value of 2,000 was applied on the FL1 channel. The gated fluorescence events were evaluated on the forward/sideward density plot, so as to exclude remaining background events. Instrument and gating settings were kept identical for all samples (fixed staining/gating strategy[39]; Supplementary Fig. 2). On the basis of the exact weight of the aliquots analysed, cell counts were converted to microbial loads per gram of faecal material.

Analyses of faecal metagenomes

Quantitative microbiome profiling. Phylogenetic quantitative microbiome profiles were built using a modified version of the pipeline described in ref. [1]. In short, sample abundance profiles were downsized to even sampling depth, defined as the ratio between sampling size (average mOTU marker genes coverage[40]) and microbial load (average total cell count per gram of frozen faecal material). The sequencing depth of each sample was rarefied to the level necessary to equate the minimum observed sampling depth in the cohort. The rarefied mOTU abundance matrix was converted into numbers of cells per gram and quantitative microbiome profiling matrices created for phylum to species levels. Functional quantitative microbiome profiles and quantitative co-abundance gene groups[38] profiles were constructed by multiplication of relative proportions to an indexing factor proportional to the microbial cell densities of the samples (load), defined as the sample load divided by the median load over the entire MetaCardis cohort. The processed microbiome profiles can be downloaded at http://raeslab.org/software/BMIS/.

Customized module analyses. Customized module sets included previously described gut metabolic modules[11] covering bacterial and archaeal metabolism specific to the human gut environment with a focus on anaerobic fermentation processes, expanded with a specific set of six modules focusing on bacterial trimethylamine metabolism[41]. Additionally, following a previously published strategy to build manually curated gut-specific metabolic modules[11,31], we constructed a new set of modules to describe and map microbial phenylpropanoid metabolism (phenylpropanoid metabolism modules, PPM) from shotgun metagenomic data. This set of 20 modules, following KEGG syntax, is provided in the Supplementary Information, including references to the original publications in which the pathways were described (Supplementary Table 3). Abundances of customized modules were derived from the orthologue abundance tables using Omixer-RPM v1.0 (https://github.com/raeslab/omixer-rpm)[11,42]. The coverage of each metabolic variant encoded in a module was calculated as the number of steps for which at least one of the orthologous groups was found in a metagenome, divided by the total number of steps constituting the variant. The presence or absence of a module was identified with a detection threshold of more than 66% coverage to provide tolerance to

misannotations and missing data in metagenomes. Module abundance was calculated as the median of orthologue abundances in the pathway with maximum coverage.

Statistical analyses

Statistical analyses were performed in R using the following packages: vegan[43] v.2.5-3, phyloseq[44] v.1.26.0, FSA[45] v.0.8.24, coin[46] v.1.2-2, DirichletMultinomial[47] v.1.24.0, Hmisc[48] v.4.1-1, car[49] v.3.0-2, sjstats[50] v.0.17.5, and nnet[51] v.7.3-12. All statistical tests used were two-sided. All *P* values were corrected for multiple testing when appropriate using the Benjamini–Hochberg method (P_{adj}), only $P_{adj} < 0.05$ were reported as significant.

Faecal microbiome derived features and visualization. Observed richness was calculated using phyloseq[44]. Microbiome inter-individual variation was visualized by principal coordinates analysis using Bray–Curtis dissimilarity on the genus-level relative abundance matrix with Hellinger transformation.

Partitioning of microbiome variation across clinical explanatory variables. The estimation of the explanatory power of clinical features regarding relative, genus-level, microbiome profiles variation was performed using univariate or multivariate stepwise distance-based redundancy analysis as implemented in the R package vegan[43].

Microbiome community typing. Enterotyping (or community typing) of the genus-level abundance microbial profiles with Hellinger transformation was performed on the basis of the Dirichlet multinomial mixtures (DMM) approach implemented in the R package DirichletMultinomial, as described in ref. [52] on the whole of the $n = 2{,}022$ MetaCardis cohort. Although the dissimilarity/distance-based approaches were applied to screen for covariate-associated microbiome trends throughout the whole of the BMIS cohort, DMM-based stratification allows identification of covariates not only associated with the strata, but also linked to fluctuations in the prevalence of one (or more) particular microbiota constellation(s). This makes enterotyping a valuable strategy when assessing microbiome variation in pathologies that are not expected to be characterized by generalized dysbiosis with varying severity according to diagnosis[53], but—by contrast—by the increased occurrence of a single dysbiotic community type with prevalence depending on the condition studied[1,2,31], as proposed here for obesity.

Microbiome features and clinical features associations. Taxa unclassified at the genus level or present in fewer than 20% of samples were excluded from the statistical analyses. Pearson or Spearman correlations were used, respectively, for linear or rank-order correlations between continuous variables, including genera abundances and metadata. The Mann–Whitney *U*-test was used to test median differences of continuous variables between two different groups. For more than two groups, the Kruskal–Wallis test with post-hoc Dunn test were used. Statistical differences in the prevalence of enterotypes between groups were evaluated using pairwise Fisher's exact tests. Modelling the association between the prevalence of enterotypes (Bact1, Bact2, Prev, Rum) or Bact2 prevalence (Bact2 = Yes/No) and single (univariate) or multiple (multivariate) dependent variables (clinical metadata features) was performed using generalized linear models, namely multinomial or binomial logistic regression (for enterotypes or Bact2 prevalences, respectively) with significance evaluated by likelihood ratio tests using the R package car. Risk ratio estimates (and their confidence intervals) were retrieved using the R package sjstats, by conversion of the odds ratios of the generalized linear models[54], the latter corresponding to exponential transformation of the model coefficients.

Reporting summary

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

Data availability

Raw amplicon sequencing data used in this study have been deposited in the EMBL-EBI European Nucleotide Archive (ENA) under accession number PRJEB37249. The metadata and processed microbiome data required for the reanalysis of results presented in the manuscript are respectively provided as Supplementary Table 2 and available for download at http://raeslab.org/software/BMIS/. For clinical cohort-related questions, contact K.C.

31. Valles-Colomer, M. et al. The neuroactive potential of the human gut microbiota in quality of life and depression. *Nat. Microbiol.* **4**, 623–632 (2019).
32. Touch, S. et al. Mucosal-associated invariant T (MAIT) cells are depleted and prone to apoptosis in cardiometabolic disorders. *FASEB J.* **32**, 5078–5089 (2018).
33. Criscuolo, A. & Brisse, S. AlienTrimmer: a tool to quickly and accurately trim off multiple short contaminant sequences from high-throughput sequencing reads. *Genomics* **102**, 500–506 (2013).
34. Li, J. et al. An integrated catalog of reference genes in the human gut microbiome. *Nat. Biotechnol.* **32**, 834–841 (2014).
35. Cotillard, A. et al. Dietary intervention impact on gut microbial gene richness. *Nature* **500**, 585–588 (2013).
36. Prifti, E. & Le Chatelier, E. MetaOMineR: a quantitative metagenomics data analyses pipeline. R package v.1.1 (2015).
37. Kultima, J. R. et al. MOCAT2: a metagenomic assembly, annotation and profiling framework. *Bioinformatics* **32**, 2520–2523 (2016).
38. Nielsen, H. B. et al. Identification and assembly of genomes and genetic elements in complex metagenomic samples without using reference genomes. *Nat. Biotechnol.* **32**, 822–828 (2014).
39. Prest, E. I., Hammes, F., Kötzsch, S., van Loosdrecht, M. C. M. & Vrouwenvelder, J. S. Monitoring microbiological changes in drinking water systems using a fast and reproducible flow cytometric method. *Water Res.* **47**, 7131–7142 (2013).
40. Kultima, J. R. et al. MOCAT: a metagenomics assembly and gene prediction toolkit. *PLoS ONE* **7**, e47656 (2012).
41. Falony, G., Vieira-Silva, S. & Raes, J. Microbiology meets big data: the case of gut microbiota-derived trimethylamine. *Annu. Rev. Microbiol.* **69**, 305–321 (2015).
42. Darzi, Y., Falony, G., Vieira-Silva, S. & Raes, J. Towards biome-specific analysis of meta-omics data. *ISME J.* **10**, 1025–1028 (2016).
43. Oksanen, J. et al. vegan: Community Ecology Package. R package v.2.2-1 (2015).
44. McMurdie, P. J. & Holmes, S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE* **8**, e61217 (2013).
45. Ogle, D. H. FSA: Fisheries Stock Analysis. R package v.0.8.13. (2017).
46. Hothorn, T., Hornik, K., van de Wiel, M. A. & Zeileis, A. A Lego system for conditional inference. *Am. Stat.* **60**, 257–263 (2006).
47. Morgan, M. DirichletMultinomial: Dirichlet-multinomial mixture model machine learning for microbiome data. R package v.1.18.0 (2017).
48. Harrell, F. E. Hmisc: Harrell Miscellaneous. R package v.4.1-1 (2018).
49. Fox, J. & Weisberg, S. *An R Companion to Applied Regression*. (Sage, 2011).
50. Lüdecke, D. sjstats: Statistical Functions for Regression Models v.0.17.5 (2019).
51. Venables, W. N. & Ripley, B. D. *Modern Applied Statistics with S*. (Springer, 2002).
52. Holmes, I., Harris, K. & Quince, C. Dirichlet multinomial mixtures: generative models for microbial metagenomics. *PLoS ONE* **7**, e30126 (2012).
53. Duvallet, C., Gibbons, S. M., Gurry, T., Irizarry, R. A. & Alm, E. J. Meta-analysis of gut microbiome studies identifies disease-specific and shared responses. *Nat. Commun.* **8**, 1784 (2017).
54. Grant, R. L. Converting an odds ratio to a range of plausible relative risks for better communication of research findings. *Br. Med. J.* **348**, f7450 (2014).

Acknowledgements We thank the study participants and nurses for their contributions to the project. MetaCardis was funded by European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement HEALTH-F4-2012-305312 (MetaCardis project) and the French National Agency of Research (ANR; 'Investissement d'Avenir' FORCE, Metagenopolis grant ANR-11-DPBS-0001 and ICAN ANR-10-IAHU-05). The promotor of the clinical study was the Assistance Publique Hôpitaux de Paris (APHP). S.V.-S. was supported by a post-doctoral fellowship from the Research Foundation Flanders (FWO Vlaanderen). The Raes laboratory is supported by the VIB Grand Challenges programme, KU Leuven, the Rega Institute for Medical Research, and the FWO EOS program (30770923). The Novo Nordisk Foundation Center for Basic Metabolic Research is an independent research institution at the University of Copenhagen partially funded by an unrestricted donation from the Novo Nordisk Foundation. M.-E.D. was funded by the NIHR Imperial Biomedical Research Centre.

Author contributions M.-E.D., S.D.E., P.G., J.P.G., T.H., J.J.H., L.K., I.L., J.N., J.-M.O., M.S., H.V., J.-D.Z., P.B., O.P., F.B., K.C. (the MetaCardis Consortium coordinator) and J.R. conceived the MetaCardis study protocol, including clinical standard operating procedures, study objectives and study design. T.N., J.A.-W. and R.C. coordinated recruitment and sample collection efforts over the different cohorts. T.N., J.A.-W., R.C. and K.A. curated and harmonized the clinical metadata. S.V.-S., G.F., E.B., T.N., J.A.-W., S.K.F., K.A., R.C., M.V.-C., S.P., E.P., V.T., N.P., E.L.C., F.A.,

J.-P.B., L.P.C., N.G., T.H.H., J.-S.H., C.L., H.K.P., B.Q., C.R., H.R., J.-E.S., N.B.S., S.T. and the MetaCardis Consortium assisted in sample collection, analyses, and/or data pre-processing and exploration. Faecal microbial DNA extraction and shotgun sequencing was performed by N.P., E.L.C. and S.F. Flow-cytometry-based faecal microbial load estimations were performed by T.T.D.N. Statistical analyses were designed and executed by S.V.-S., G.F., E.B., K.A., S.K.F. and M.V.-C. The manuscript was drafted by G.F., S.V.-S., K.C. and J.R. All authors revised the article and approved the final version for publication.

Competing interests J.R., S.V.-S., G.F. and M.V.-C. are listed as inventors on patent application PCT/EP2018/084920, in the name of VIB VZW, Katholieke Universiteit Leuven, KU Leuven R&D and Vrije Universiteit Brussel, covering the features of the microbiome associated with inflammation described in ref. [2].

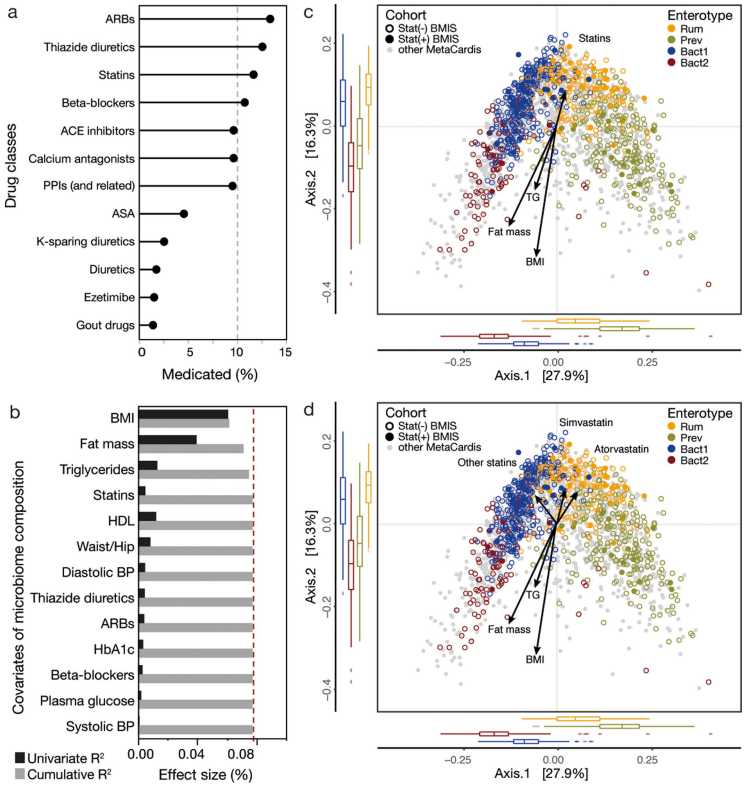
Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41586-020-2269-x.

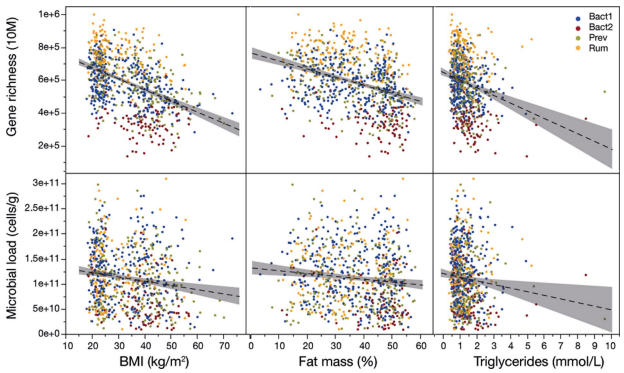
Correspondence and requests for materials should be addressed to K.C. or J.R.

Peer review information *Nature* thanks Peter Turnbaugh and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

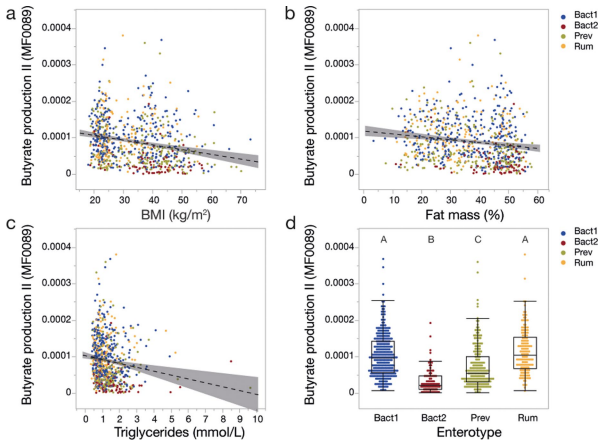
Reprints and permissions information is available at http://www.nature.com/reprints.



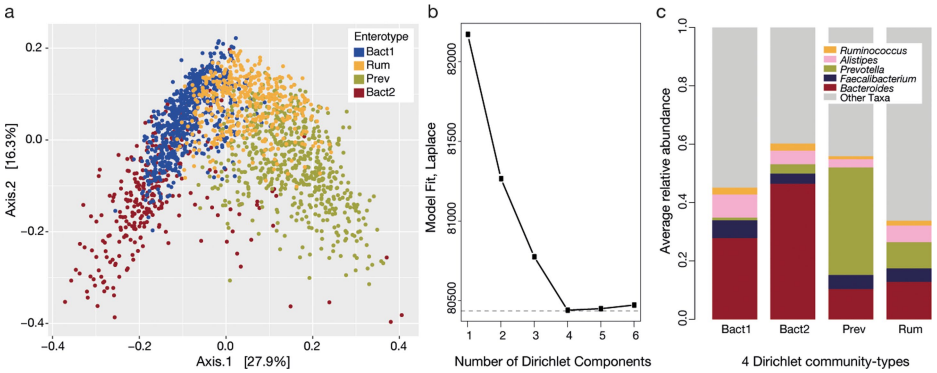
Extended Data Fig. 1 | Microbiome variation in the BMIS cohort ($n = 888$ participants). a, Percentage of subjects in the BMIS cohort taking medication of the stated drug classes. ACE inhibitors, angiotensin converting enzyme inhibitors; ARB, angiotensin II receptor blockers; ASA, acetylsalicylic acid; PPI, proton-pump inhibitors. **b**, Best model explaining inter-individual microbiome variation based on obesity-defining and metabolic-syndrome-defining variables as well as on most frequently disclosed therapeutics (taken by more than 10% of participants; Supplementary Table 4). Explanatory power of the included variables are reported for the variables taken individually (black bars; $n = 888$ biologically independent samples, univariate dbRDA)) or in a multivariate model (grey bars; $n = 888$ biologically independent samples, multivariate dbRDA). **c**, Principal coordinates analysis of inter-individual differences (genus level Bray–Curtis dissimilarity) in the microbiome profiles of the BMIS cohort ($n = 888$ biologically independent samples, data points coloured by enterotypes (Extended Data Fig. 4)) with the rest of the MetaCardis dataset in the background ($n = 1{,}134$, grey dots). Full and open circles corresponding to statin-medicated (Stat(+)) and non-statin-medicated participants (Stat(−)), respectively. Arrows represent the effect sizes of a post hoc fit of significant microbiome covariates identified in the multivariate model in **b**. **d**, Same principal coordinates analysis as in **c**, with the statin intake variable split into the separate statin classes ($n = 888$ biologically independent samples, simvastatin ($n = 51$), atorvastatin ($n = 33$) and other statins ($n = 22$); Supplementary Table 4). In **c**, **d**, the body of the box plot represents the first and third quartiles of the distribution, the line represents the median, and the whiskers extend from the quartiles to the last data point within 1.5× the interquartile range (IQR), with outliers beyond.



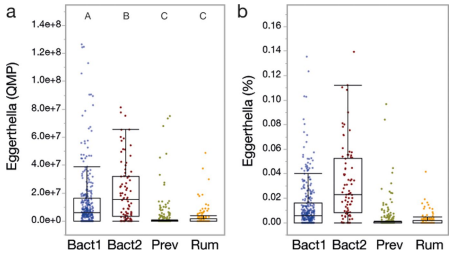
Extended Data Fig. 2 | The association of BMI, fat mass percentage and serum fasting triglyceride levels with faecal microbial gene richness and faecal microbial load in the non-statin-medicated BMIS cohort (*n* = 782 participants). All three covariates were found to be associated with both microbiome gene richness ($n = 711$ biologically independent samples, Spearman's $\rho = -0.45$ to -0.26, $P_{adj} = 4.0 \times 10^{-39}$ to 1.6×10^{-13}), a proxy for microbial biodiversity previously suggested as a marker of metabolic health in obese individuals[8], and faecal microbial load ($n = 711$ biologically independent samples, Spearman's $\rho = -0.17$ to -0.13, $P_{adj} = 4.1 \times 10^{-6}$ to 3.1×10^{-4}; Supplementary Table 7). Adjustment for multiple testing (P_{adj}) was performed using the Benjamini–Hochberg method. Least square linear regression lines (dashed line) with 95% confidence interval (grey shading) are provided for visual representation of the non-parametric testing provided in Supplementary Table 7. Data points are coloured by enterotype classification.



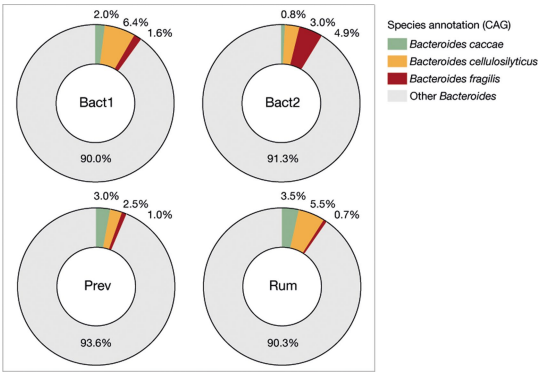
Extended Data Fig. 3 | Association between the variation in quantitative butyrate production potential and the BMI, fat mass percentage and triglycerides levels of participants, or the enterotype classification of the samples, in the non-statin-medicated BMIS cohort (*n* = 782 participants). Quantitative functional microbiome profiles were constructed by multiplication of relative proportions to an indexing factor proportional to the microbial load of the samples. The module 'butyrate production II' describes butyrate production from the butyryl-CoA–acetate CoA-transferase pathway—the most common among colon bacteria. **a**–**d**, The abundance of the butyrate production II module was negatively correlated with BMI ($n = 771$ biologically independent samples, Spearman's $\rho = -0.27$, $P_{adj} = 3.1 \times 10^{-13}$) (**a**), fat mass percentage ($n = 771$ biologically independent samples, Spearman's $\rho = -0.21$, $P_{adj} = 6.0 \times 10^{-8}$) (**b**) and tryglyceride levels ($n = 771$ biologically independent samples, Spearman's $\rho = -0.20$, $P_{adj} = 6.4 \times 10^{-8}$) (**c**), and significantly decreased in the Bact2 enterotype compared with the others (Bact2 < Prev < Bact1 = Rum; $n = 771$ biologically independent samples, Kruskal–Wallis $P_{adj} = 4.71 \times 10^{-35}$; different letters denote enterotypes with a significant pairwise difference (post hoc Dunn tests provided in Supplementary Table 10) (**d**). The body of the box plot represents the first and third quartiles of the distribution, the line represents the median, and the whiskers extend from the quartiles to the last data point within 1.5× IQR, with outliers beyond. In **a**–**d**, adjustment for multiple testing (P_{adj}) was performed using the Benjamini–Hochberg method.



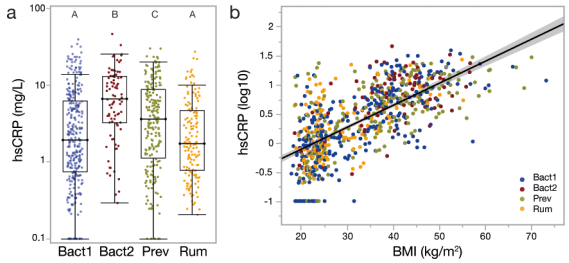
Extended Data Fig. 4 | Enterotyping of the MetaCardis dataset (n = 2,022 biologically independent samples). a, Principal coordinates visualization of the four enterotypes resulting from community typing was performed using DMM[52] on genus-level faecal microbiome profiles. **b**, Information criteria (minimum Laplace) used to determine the optimal number of clusters (enterotypes) for the MetaCardis dataset (n = 2,022 biologically independent samples) DMM-based community typing. **c**, Average relative composition of the enterotypes for key genera, used to label the enterotypes *Bacteroides*1 (Bact1; high percentages of *Bacteroides* and *Faecalibacterium*), *Bacteroides*2 (Bact2; high percentages of *Bacteroides* and low percentages of *Faecalibacterium*), *Prevotella* (Prev; high percentages of *Prevotella*) and Ruminococcaceae (Rum; low percentages of *Bacteroides* and *Prevotella*), on the basis of their respective genus-level proportional abundance profiles.



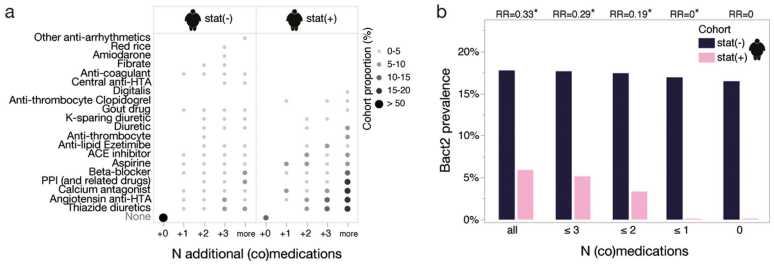
Extended Data Fig. 5 | Increased quantitative abundance of Eggerthella in the Bact2 enterotype of the non-statin-medicated BMIS cohort. **a**, Difference in quantitative *Eggerthella* abundances between enterotypes (Prev = Rum < Bact1 < Bact2; $n = 771$ biologically independent samples, Kruskal–Wallis $P_{adj} = 4.10 \times 10^{-47}$; different letters denote enterotypes with a significant pairwise difference (post hoc Dunn tests provided in Supplementary Table 10)). Adjustment for multiple testing (P_{adj}) was performed using the Benjamini–Hochberg method. **b**, Difference in the proportion of *Eggerthella* (normalized by the sample total microbial load) between enterotypes, showing a comparable trend to that seen in **a** ($n = 771$ biologically independent samples). The body of the box plot represents the first and third quartiles of the distribution, the line represents the median, and the whiskers extend from the quartiles to the last data point within 1.5× IQR, with outliers beyond.



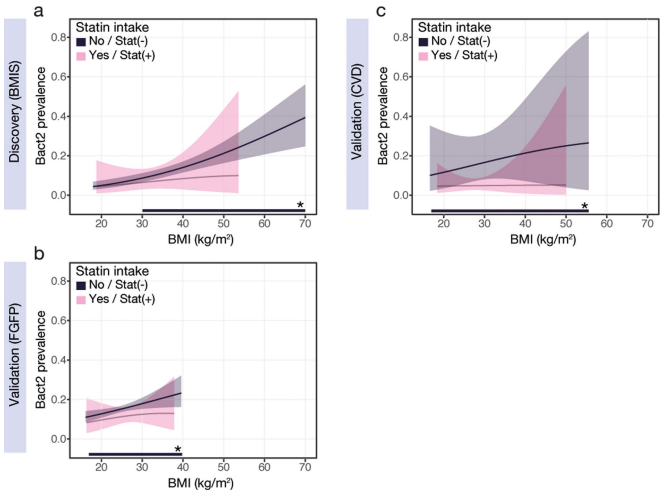
Extended Data Fig. 6 | Species dominating the *Bacteroides* fraction in the different enterotypes of the non-statin-medicated BMIS cohort. The top associations with the Bact2 enterotype—with the proportions they contribute to the total fraction shown in the ring chart—were the depletion in *B. caccae* ($n = 768$ biologically independent samples, Kruskal–Wallis, $P_{adj} = 1.3 \times 10^{-15}$) and *B. cellulosilyticus* ($n = 768$ biologically independent samples, Kruskal–Wallis, $P_{adj} = 5.3 \times 10^{-13}$) when compared with the Rum, Prev and Bact1 enterotypes, and the enrichment in *B. fragilis* ($n = 768$ biologically independent samples, Kruskal–Wallis, $P_{adj} = 3.5 \times 10^{-11}$; Supplementary Table 11). Species were defined by species-level annotation of metagenomic species, and their proportional abundances were defined relative to the genus abundance. Samples for which the genus had a low total abundance (below the 20% quantile for all species belonging to the top 10 genera) were excluded from the analysis ($n = 768$ biologically independent samples were included). Adjustment for multiple testing (P_{adj}) was performed using the Benjamini–Hochberg method.



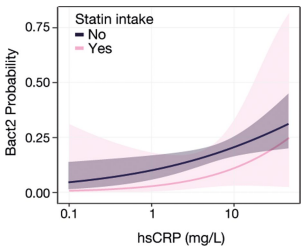
Extended Data Fig. 7 | Systemic inflammation and its relation to enterotypes and to BMI in the non-statin-medicated BMIS cohort. **a**, Individuals with faecal samples enterotyped as Bact2 displayed more pronounced systemic inflammation levels as assessed through fasting serum hsCRP concentrations when compared with participants classified as Rum, Prev and Bact1 ($n = 763$ biologically independent samples, Kruskal–Wallis $P = 1.37 \times 10^{-10}$; Rum = Bact1 < Prev < Bact2; different letters denote enterotypes with a significant pairwise difference (post hoc Dunn tests provided in Supplementary Table 13)). The body of the box plot represents the first and third quartiles of the distribution, the line represents the median, and the whiskers extend from the quartiles to the last data point within 1.5× IQR, with outliers beyond. **b**, Linear model of the correlation between host systemic inflammation (hsCRP concentration, $\log_{10}$-transformed) and BMI, fitted by least squares regression ($n = 763$ biologically independent samples; estimated intercept = −0.8681, estimated slope = 0.0379, $R^2 = 0.47$, $P = 1.5 \times 10^{-108}$).



Extended Data Fig. 8 | Control for the effect of additional medication taken by obese statin-medicated or non-statin-medicated individuals of the BMIS cohort (*n* = 888 participants) on the association between reduced Bact2 prevalence and statin intake. a, List of drugs taken by non-statin-medicated and statin-medicated obese BMIS participants separated into 5 groups: those reporting no (co-)medication (beyond statin intake) (+0), and those reporting one (+1), two (+2), three (+3) and more than three (more) (co-)medications. The size and colour of the dots represent the fraction of the non-statin-medicated or statin-medicated obese BMIS participants falling within that group. **b**, Difference in prevalence of the Bact2 enterotype in statin-medicated compared with non-statin-medicated obese BMIS participants, with decreasing co-medication threshold for inclusion of participants. For 'all', the total number of statin-medicated and non-statin-medicated obese BMIS participants were included ($n = 474$ biologically independent samples); then only subjects reporting three or fewer (≤ 3; $n = 419$), two or fewer (≤ 2; $n = 369$), one or fewer (≤ 1; $n = 296$) or no (0; $n = 226$) (co-)medications were included. The relative risk and respective significance level associated with the prevalence of the Bact2 enterotype given statin intake is provided above the bar plots (Fisher's exact test, two-sided, $^*P < 0.05$, relative risk = P(Bact2|Statin = Yes)/P(Bact2|Statin = No)).



Extended Data Fig. 9 | Variation in prevalence of the Bact2 enterotype with BMI and statin intake in the BMIS discovery cohort, and in the FGFP and CVD validation cohorts. a–**c**, Variation in the prevalence of the Bact2 enterotype with BMI for statin-medicated and non-statin-medicated individuals, showing the significant effect (represented by the range bar with an asterisk; Supplementary Table 16) of statin intake given individuals' BMI, in the BMIS obese participants (n = 474 biologically independent samples, multivariate binomial logistic regression, Statin | BMI, relative risk = 0.34, *P_{adj} = 0.025) (**a**); the FGFP cohort, a population-level recruitment with a much narrower BMI range than the BMIS cohort (n = 2,345 biologically independent samples, multivariate binomial logistic regression, Statin | BMI, relative risk = 0.72, *P_{adj} = 0.045) (**b**) and the MetaCardis CVD cohort (n = 271 biologically independent samples, excluding 11 individuals for which BMI was not known, multivariate binomial logistic regression, Statin | BMI, relative risk = 0.29, *P_{adj} = 0.021) (**c**). In **a**–**c**, the fit lines were obtained by multinomial logistic regression of enterotypes as predicted by BMI, for statin-medicated and non-statin-medicated individuals separately, with the shaded area corresponding to the 95% confidence intervals for the Bact2 regression. Adjustment for multiple testing (P_{adj}) was performed using the Benjamini–Hochberg method.



Extended Data Fig. 10 | Probability of carrying a Bact2 enterotype microbiota as a function of CRP levels and statin intake in the obese BMIS cohort. Association between systemic inflammation (measured by hsCRP levels) and having a faecal microbiota of the Bact2 enterotype, according to statin medication status. Binomial logistic regression (lines with 95% confidence intervals as shaded area) was performed for statin-medicated and non-statin-medicated individuals separately ($n = 462$ biologically independent samples).

nature research

Corresponding author(s): Jeroen Raes

Last updated by author(s): Mar 16, 2020

Reporting Summary

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

Statistics

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

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

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

Software and code

Policy information about availability of computer code

Data collection	[Metagenomic data] Sequencing reads, obtained from ion-proton technology (ThermoFisher Scientific), were cleaned using Alien Trimmer (v.0.4.0). Filtered reads were mapped to the 9.9 million-gene catalogue using Bowtie (v2.2.6). A gene abundance table was generated by means of a two-step procedure using METEOR, and posterior normalization was performed using the MetaOMineR R package using the FPKM strategy. Metagenomic species (MGS) or co-abundant gene groups were previously computed on 1267 human gut metagenomes used to construct the 9.9 million-gene catalogue, their abundance profiles for the MetaCardis cohort was estimated as the mean abundance of the 50 genes defining a robust centroid of the cluster, with taxonomic assignment by NCBI blast N on the NCBI database (November 2016 version). [Faecal microbial loads] Flow cytometry analysis was performed using a C6 Accuri flow cytometer (BD Biosciences, New Jersey, USA), using the BD Accuri CFlow software (v1.0.264.21) for gating and events counting.
Data analysis	[Metagenomic data] Gut metabolic modules (GMM) profiles were calculated using the software Omixer-RPM v1.0 and the newest version of the GMMs (v.2.0), which includes a specific set of six modules zooming in on bacterial TMA metabolism. [Quantitative microbiota profiles] QMP profiles were created using the phyloseq R package to rarefy the profiles to even sampling depth, sampling depth, defined as the ratio between sampling size (average mOTU marker genes coverage44) and microbial load (average total cell count per gram of frozen faecal material). [Statistical analyses] Statistical analyses were performed on Rstudio v1.1.456, using the following R packages: vegan v2.5-3, phyloseq v1.26.0, FSA v0.8.24, coin v1.2-2, DirichletMultinomial v1.24.0, Hmisc v4.1-1, car v3.0-2, sjstats v0.17.5, and nnet v7.3-12.

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

Data

Policy information about availability of data

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

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

Raw amplicon sequencing data used in this study have been deposited in the EMBL-EBI European Nucleotide Archive (ENA) under accession number PRJEB37249 [public access]. The metadata and processed microbiome data required for re-analysis of results presented in the manuscript are respectively provided as Extended Data 1 (tab separated file) and downloadable at http://raeslab.org/software/BMIS/.

Field-specific reporting

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

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

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

Life sciences study design

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

Sample size	No sample size calculation was performed prior to cohort recruitment. Based on the baseline prevalence of Bact2 enterotype (with baseline defined as lean/overweight individuals P(Bact2)=14%) in the amplicon-sequenced Flemish Gut Flora Project cohort, the present study cohort size allowed to identify a minimum difference of 7.4% in Bact2 prevalence between the two groups lean/overweight (N=414) vs obese (N=474) as significant (power=80%, alpha=0.05). Because the FGFP cohort was a population microbiome monitoring effort, while the BMIS cohort was actively recruited to have a balanced representation over a wide BMI range, therefore the sample size would be sufficient to detect a smaller prevalence difference.
Data exclusions	This study targeted the analysis of metabolic alterations associated with body mass index ranging from normal to severe obesity. The BMIS cohort (N=888) was selected from the MetaCardis consortium study cohort, by exclusion of cardiovascular patients (as defined in the MetaCardis consortium study protocol as patient groups 4, 5, 6 and 7) and any individual with type-2 diabetes (T2D). T2D diagnosis was defined using the American Diabetes Association (ADA) definition: fasting glycemia > 6.9 mmol/l and/or 2h values in the oral glucose tolerance test > 11 mmol/l and/or haemoglobin A1c (HbA1c, glycated haemoglobin) ≥ 6.5% and/or use of any anti-diabetic treatment. but excluding volunteers with a diagnosis of diabetes as to avoid the potentially confounding effect of the associated medication. This exclusion criteria was pre-established for this manuscript. The same criteria for inclusion and exclusion that were used for MetaCardis recruitment were applied to the FGFP validation cohort: inclusion of age ranging from 18 to 75 years old, exclusion of acute or chronic inflammatory or infectious diseases – notably diagnosis of inflammatory bowel disease and recent gastro-enteritis, and exclusion of diabetes patients – defined as above as diabetes diagnosis or elevated glycated haemoglobin (HbA1c ≥ 6.5%) or use of any anti-diabetic treatment.
Replication	All microbiome observations regarding Bact2 prevalence associated with obesity and statin therapy were successfully replicated in another subset of the MetaCardis cohort - the cardiovascular patients [CVD] (as defined in the MetaCardis consortium study protocol as patient groups 4, 5, 6 and 7), and in an independent cross-sectional cohort - the Flemish Gut Flora Project cohort [FGFP].
Randomization	Not applicable: this was a cross-sectional study, not a randomized study. No intervention was performed on subjects, and therefore no random allocation into groups. Potentially confounding covariates were identified as variables with significant association to the response or dependent variables and were added to a multivariate model to validate the findings.
Blinding	Not applicable: this was a cross-sectional study, not a randomized study. The investigators were not blinded during data collection or data analysis.

Reporting for specific materials, systems and methods

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

Materials & experimental systems

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

Methods

n/a	Involved in the study
[x]	[ ] ChIP-seq
[ ]	[x] Flow cytometry
[x]	[ ] MRI-based neuroimaging

Human research participants

Policy information about studies involving human research participants

Population characteristics	A complete description of the study participants can be found in Supplementary Table 1 (blood panel and current medication intake). The BMIS cohort consisted of N=888 participants (574 Females, 314 Males), with a median BMI of 31.49 [17.95-73.26] and a median age of 54 [18-76], recruited in 3 countries (294 DE, 247 DK, 347 FR). None of the included participants were diagnosed as diabetic.
Recruitment	The N=888 transnational Body Mass Index Spectrum (BMIS) cohort was assembled as part of the overall MetaCardis recruitment efforts. Participants were recruited between 2013 and 2015 in the clinical departments of the Pitié-Salpêtrière Hospital (Paris, France), the Integrated Research and Treatment Center for Adiposity Diseases (Leipzig, Germany), and in the Novo Nordisk Foundation Center for Basic Metabolic Research (Copenhagen, Denmark). Potential participants were evaluated for suitability according to standardized inclusion and exclusion criteria across the three recruitment centers. Exclusion criteria included history of abdominal cancer/radiation therapy on the abdomen, history of intestinal resection (except for appendectomy), acute or chronic inflammatory or infectious diseases (including VHC, VHB, and HIV), history of organ transplantation or receiving immunosuppressive therapy, severe kidney failure (MDRD glomerular filtration rate <50 ml (min 1.73m^2)-1), or drug or alcohol addiction. All study participants had to be free of any antibiotic use in the three months prior to inclusion. We do not expect any (self-)selection bias that would have an impact on the results.
Ethics oversight	Ethical approval was obtained from the Ethics Committee CPP Ile-de France, Ethics Committee at the Medical Faculty of the University of Leipzig, and the Ethical Committees of the Capital Region of Denmark. Study design complied with all relevant ethical regulations, aligning with the Helsinki Declaration and in accordance with European privacy legislation. All participants provided written informed consent.

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

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJEguidelines for publication of clinical research and a completedCONSORT checklist must be included with all submissions.

Clinical trial registration	The study protocol was registered at clinicaltrial.gov (NCT02059538).
Study protocol	The study protocol is available at https://clinicaltrials.gov/ct2/show/NCT02059538
Data collection	The N=888 transnational Body Mass Index Spectrum (BMIS) cohort was assembled as part of the overall MetaCardis recruitment efforts. Participants were recruited between 2013 and 2015 in the clinical departments of the Pitié-Salpêtrière Hospital (Paris, France), the Integrated Research and Treatment Center for Adiposity Diseases (Leipzig, Germany), and in the Novo Nordisk Foundation Center for Basic Metabolic Research (Copenhagen, Denmark).
Outcomes	The hypotheses tested in this manuscript were not listed as part of the planned NCT02059538 study outcomes. The primary predefined outcome of the MetaCardis project (description of differences in gut microbiota signatures between MetaCardis study groups using metagenomic sequencing) is not addressed in the present manuscript.

Flow Cytometry

Plots

Confirm that:

[x] The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

[x] The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

[x] All plots are contour plots with outliers or pseudocolor plots.

[x] A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	0.2 g frozen (-80°C) faecal aliquots were dissolved in physiological solution to a total volume of 100 mL (8.5 g/L NaCl). Subsequently, the slurry was diluted 1,000 times. Samples were filtered using a sterile syringe filter (pore size of 5 µm). 1 mL of the microbial cell suspension obtained was stained with 1 µL SYBR Green I (1:100 dilution in DMSO; shaded 15 min incubation at 37°C).
Instrument	C6 Accuri flow cytometer (BD Biosciences, New Jersey, USA).
Software	BD Accuri CFlow software v1.0.264.21 (BD Biosciences, New Jersey, USA).
Cell population abundance	Not applicable. No sorting of fractions was performed.
Gating strategy	Fluorescence events were monitored using the FL1 533/30 nm and FL3 >670 nm optical detectors. In addition, also forward and sideward-scattered light was collected. The BD Accuri CFlow software was used to gate and separate the microbial fluorescence events on the FL1/FL3 density plot from background. A threshold value of 2000 was applied on the FL1 channel. The gated fluorescence events were evaluated on the forward/sideward density plot, as to exclude remaining background events. Instrument and gating settings were kept identical for all samples.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.