

# Causal relationships among the gut microbiome, short-chain fatty acids and metabolic diseases

Serena Sanna <sup>ID</sup><sup>1,10\*</sup>, Natalie R. van Zuydam<sup>2,3,10</sup>, Anubha Mahajan <sup>ID</sup><sup>2,3,10</sup>, Alexander Kurilshikov <sup>ID</sup><sup>1</sup>, Arnau Vich Vila <sup>ID</sup><sup>1,4</sup>, Urmo Võsa<sup>1</sup>, Zlatan Mujagic<sup>5</sup>, Ad A. M. Masclee<sup>5</sup>, Daisy M. A. E. Jonkers<sup>5</sup>, Marije Oosting<sup>6</sup>, Leo A. B. Joosten <sup>ID</sup><sup>6</sup>, Mihai G. Netea <sup>ID</sup><sup>6</sup>, Lude Franke <sup>ID</sup><sup>1</sup>, Alexandra Zhernakova<sup>1</sup>, Jingyuan Fu <sup>ID</sup><sup>1,7</sup>, Cisca Wijmenga <sup>ID</sup><sup>1,8,11\*</sup> and Mark I. McCarthy <sup>ID</sup><sup>2,3,9,11\*</sup>

**Microbiome-wide association studies on large population cohorts have highlighted associations between the gut microbiome and complex traits, including type 2 diabetes (T2D) and obesity<sup>1</sup>. However, the causal relationships remain largely unresolved. We leveraged information from 952 normoglycemic individuals for whom genome-wide genotyping, gut metagenomic sequence and fecal short-chain fatty acid (SCFA) levels were available<sup>2</sup>, then combined this information with genome-wide-association summary statistics for 17 metabolic and anthropometric traits. Using bidirectional Mendelian randomization (MR) analyses to assess causality<sup>3</sup>, we found that the host-genetic-driven increase in gut production of the SCFA butyrate was associated with improved insulin response after an oral glucose-tolerance test ( $P = 9.8 \times 10^{-5}$ ), whereas abnormalities in the production or absorption of another SCFA, propionate, were causally related to an increased risk of T2D ( $P = 0.004$ ). These data provide evidence of a causal effect of the gut microbiome on metabolic traits and support the use of MR as a means to elucidate causal relationships from microbiome-wide association findings.**

Increasing evidence indicates that the human gut microbiome plays a role in immune function and metabolic disease<sup>1,4,5</sup>. Manipulation of the gut microbiome offers an alternative to pharmacological interventions, provided that altering microbiota composition and/or function (for example, through personalized nutrition) can be demonstrated to have clinical benefit. To demonstrate such benefit, it is essential to discriminate microbiome features that are causal for disease from those that are a consequence of disease or its treatment, and from those that show a statistical correlation due to confounding or pleiotropy.

Animal studies support a causal role for the gut microbiome in the development of T2D, insulin resistance and obesity<sup>6,7</sup>, but translating these findings to humans and identifying the specific bacterial species responsible has proven challenging<sup>8</sup>. Cross-sectional studies have confirmed that the composition of the gut microbiota is altered in subjects with prediabetes or T2D compared with controls,

and fecal-transplantation studies have shown that insulin sensitivity increases in obese subjects with metabolic syndrome after the transfer of gut microbiota from lean donors<sup>4,5,9,10</sup>. Although the specific microbiome features identified as being responsible for these effects differ among studies, one consistent finding in T2D subjects is a shift in the microbiome composition away from species able to produce butyrate. Butyrate and other SCFAs, such as acetate and propionate, are produced by gut bacterial fermentation of undigested food components. After absorption by the colonocytes, these SCFAs either are used locally as fuel for colonic mucosal epithelial cells or enter the portal bloodstream<sup>11</sup>. Although the bulk of evidence suggests that increased SCFA production benefits the host by exerting antiobesity and antidiabetic effects<sup>4,10,12–14</sup>, some in vitro and in vivo studies have indicated that overproduction or accumulation of SCFAs in the bowel may also lead to obesity, owing to increased energy accumulation<sup>15,16</sup>. Resolution of these conflicting data requires a detailed understanding of the causal relationships among gut-microbiome composition, SCFA abundance and host energy metabolism.

Using an MR approach<sup>3</sup>, we set out to identify whether any bacterial species or pathways, i.e., sets of species grouped according to their specific functions in the gut, have a causal effect on metabolic traits. We and others have recently shown that it is possible to detect variants in the host genome that influence the composition of the gut microbiota<sup>2,17,18</sup>. These findings allowed us to deploy an MR approach to infer causal relationships by asking whether genetic predictors of microbiome content influence metabolic traits—or the reverse. This formulation holds even though the quantitative contribution of host genetics to variations in microbiome composition may be limited<sup>19</sup>.

We assembled genome-wide genetic data, gut metagenomic sequencing, measurements of fecal SCFAs and clinical phenotypes for 952 normoglycemic individuals from the LifeLines-DEEP (LL-DEEP) cohort. From consortium websites (GIANT, MAGIC and DIAGRAM; see URLs), we also gathered publicly available genome-wide-association summary statistics for 17 anthropometric

<sup>1</sup>Department of Genetics, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands. <sup>2</sup>Wellcome Centre for Human Genetics, University of Oxford, Oxford, UK. <sup>3</sup>Oxford Centre for Diabetes Endocrinology and Metabolism, Churchill Hospital, University of Oxford, Oxford, UK. <sup>4</sup>Department of Gastroenterology and Hepatology, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands.

<sup>5</sup>Maastricht University Medical Center, Division Gastroenterology-Hepatology, NUTRIM School for Nutrition, and Translational Research in Metabolism, Maastricht, the Netherlands. <sup>6</sup>Department of Internal Medicine, Radboud Institute of Molecular Life Sciences (RIMLS) and Radboud Center for Infectious Diseases (RCI), Radboud University Medical Center, Nijmegen, the Netherlands. <sup>7</sup>Department of Pediatrics, Groningen, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands. <sup>8</sup>K.G. Jebsen Coeliac Disease Research Centre, Department of Immunology, University of Oslo, Oslo, Norway. <sup>9</sup>Oxford NIHR Biomedical Research Centre, Oxford University Hospitals NHS Foundation Trust, John Radcliffe Hospital, Oxford, UK.

<sup>10</sup>These authors contributed equally: Serena Sanna, Natalie R. van Zuydam, Anubha Mahajan. <sup>11</sup>These authors jointly supervised this work: Cisca Wijmenga, Mark I. McCarthy \*e-mail: [s.sanna@umcg.nl](mailto:s.sanna@umcg.nl); [c.wijmenga@umcg.nl](mailto:c.wijmenga@umcg.nl); [mark.mccarthy@drl.ox.ac.uk](mailto:mark.mccarthy@drl.ox.ac.uk)

and glycemic traits<sup>20–27</sup> (Supplementary Table 1). We focused our analyses on 245 microbiome features (2 fecal SCFA levels, 57 unique taxa, and 186 pathways) that were, in LL-DEEP, correlated (false discovery rate (FDR)<0.1) with at least one of the measured anthropometric and metabolic traits (Methods and Supplementary Tables 2 and 3).

For each of these features, we sought genetic predictors—that is, independent genetic variants ( $r^2 \leq 0.1$ ) associated ( $P < 1 \times 10^{-5}$ ) with the respective features—by using genome-wide association study (GWAS) data from LL-DEEP, which were reprocessed from our previous study<sup>2</sup> (Methods). The threshold  $P < 1 \times 10^{-5}$  for variant inclusion was identified by maximizing the amount of genetic variance explained by the genetic predictors in 445 independent normoglycemic individuals (the 500FG cohort)<sup>28</sup> (Methods and Supplementary Fig. 1), and it was designed to capture sets of variants likely to be enriched in association. On average, in LL-DEEP, the identified genetic predictors explained 13% (range 2–30%) of the variance in their respective microbiome features. The average  $F$  statistic, another measure of the strength of these genetic predictors, was 21.7 (range 15.3–25.5); an  $F$  statistic>10 is considered sufficiently informative for MR analyses<sup>29</sup>.

We used the inverse-variance weighted (IVW) test to identify causal relationships among the 245 microbiome features and the 17 traits of interest in a two-sample bidirectional MR analysis using pairs of GWAS summary statistics (one from a microbiome feature and one from a metabolic/anthropometric trait)<sup>29</sup>. On the basis of principal component analysis and cluster analyses conducted on the microbiome and metabolic and anthropometric traits (Methods and Supplementary Fig. 2), we adopted a conservative multiple-testing-adjusted threshold of  $P < 1.3 \times 10^{-4}$  to declare a causal relationship significant. Because the presence of horizontal pleiotropy (in which a genetic predictor has independent effects on the diseases through multiple traits) could bias the MR estimates, we investigated the robustness of our significant findings to pleiotropy by using three additional MR tests: MR-PRESSO<sup>30</sup>, the weighted median test<sup>31</sup> and MR-Egger<sup>32</sup>. We formally examined the presence of horizontal pleiotropy by using the MR-PRESSO Global test<sup>30</sup> and modified Rücker's Q' test<sup>33,34</sup>. Finally, we sought to validate these causal relationships in an independent cohort (UK Biobank)<sup>35</sup> (Fig. 1).

We observed a significant causal influence for one specific microbiome feature, a microbial pathway involved in 4-aminobutyrate (GABA) degradation (MetaCyc designation PWY-5022: 4-aminobutyrate degradation V) on increased insulin secretion, specifically the ratio of the areas under the curve for insulin and glucose ( $AUC_{\text{insulin}}/AUC_{\text{glucose}}$ ) measured during an oral glucose-tolerance test (oGTT) (Fig. 2a). Using nine genetic predictors (variance explained = 16%;  $F$  statistic = 21; Supplementary Table 4), we estimated that each 1-s.d. increase in the abundance of PWY-5022 would generate a 0.16 mU/mmol increase in  $AUC_{\text{insulin}}/AUC_{\text{glucose}}$  ( $P = 9.8 \times 10^{-5}$ ; Supplementary Table 5 and Supplementary Fig. 3). This causal relationship was robust when additional MR tests were performed ( $P_{\text{MR-PRESSO}} = 0.02$ ,  $P_{\text{weighted-median}} = 0.02$  and  $P_{\text{MR-Egger}} = 0.02$ ), and there was no evidence of horizontal pleiotropy ( $P_{\text{MR-PRESSOGlobal}} = 0.18$  and  $P_{\text{RückerQ' (modified)}} = 0.77$ ) (Supplementary Fig. 4). The reverse MR analysis (testing the relationship between genetic predictors of  $AUC_{\text{insulin}}/AUC_{\text{glucose}}$  and PWY-5022 abundance) was not significant ( $P > 0.1$ ; Supplementary Table 6). There was no evidence of causality with seven metabolic and anthropometric traits (body-mass index (BMI), body-fat percentage, waist-hip ratio (WHR), visceral adipose tissue, abdominal subcutaneous adipose tissue, obesity and T2D) in an MR analyses that used UK Biobank summary statistics (Supplementary Table 7); insulin-secretion phenotypes after oGTT were not available. We also found evidence ( $P < 0.05$ ) supporting a causal effect of this pathway on other insulin-response parameters (Fig. 2b). Although other types

of causal relationships are possible, these data are consistent with a model in which host genetic variation influences gut-microbiome composition so as to modulate GABA degradation activity, which in turn increases the ability of the pancreatic islets to secrete insulin in response to a physiological glucose challenge.

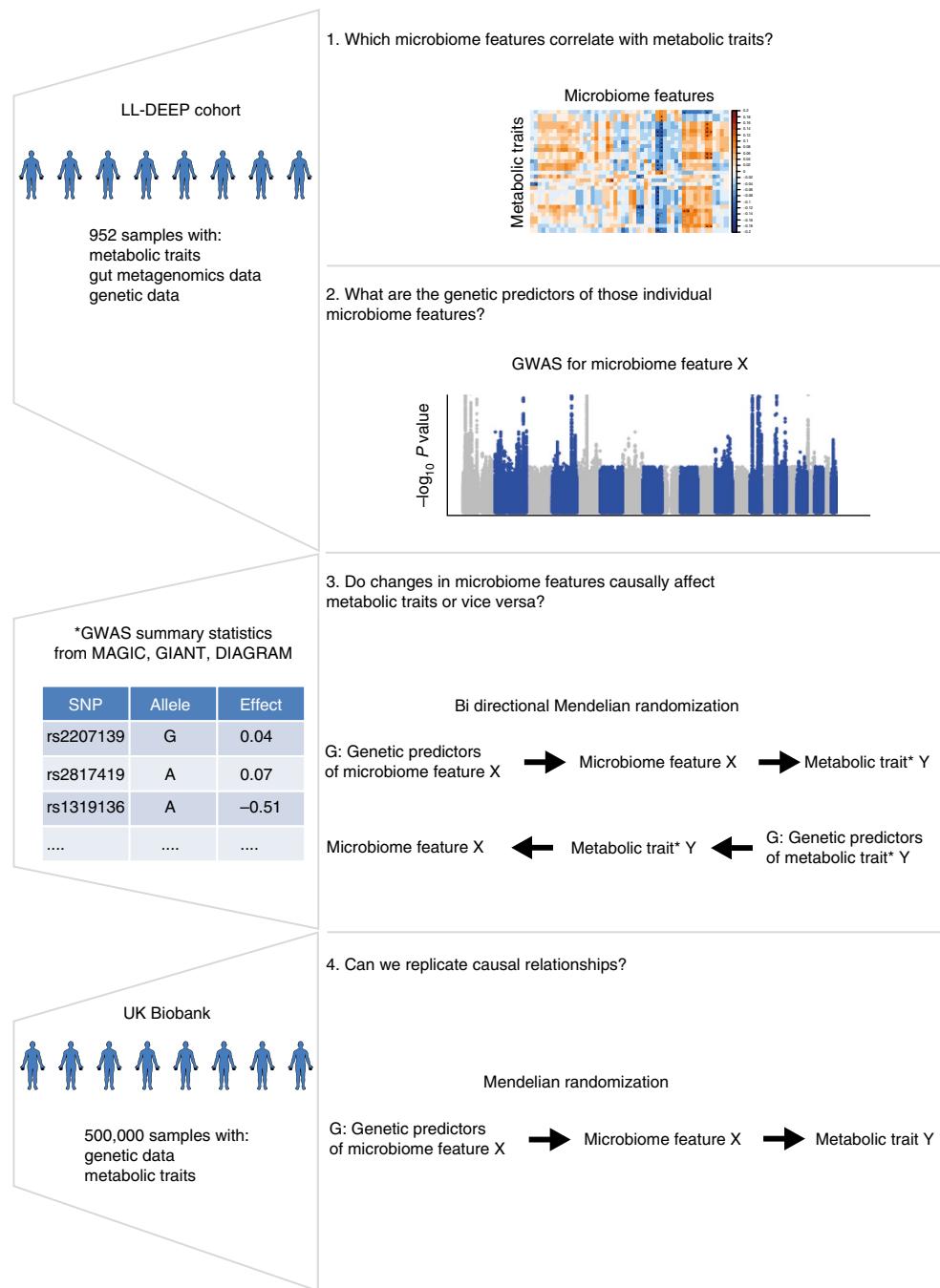
Butyrate and acetate are products of GABA degradation. In our taxonomic analyses, the bacterial species most correlated with the abundance of PWY-5022 were *Eubacterium rectale* and *Roseburia intestinalis* (Spearman  $\rho = 0.52$  and 0.30, respectively; Fig. 2c), both of which are well-known butyrate-producing bacteria<sup>36,37</sup>. We did not measure plasma butyrate levels in our study, because the current assays are challenging to perform and provide unreliable estimates<sup>38</sup>. Although we considered the abundance of the PWY-5022 pathway to act as a proxy for butyrate production in the gut, we were unable to directly link PWY-5022 abundance to the amount of butyrate absorbed by the host. The abundance of PWY-5022 was poorly correlated with fecal butyrate levels (Spearman  $\rho = 0.1$ ), and we did not detect any causal relationships between fecal butyrate and the 17 traits ( $P > 0.05$ ), thus indicating that fecal levels are a poor proxy for butyrate production and absorption.

These results suggest a causal role of gut-produced butyrate that is focused on the dynamic insulin response to food ingestion rather than on the homeostatic mechanisms involved in the maintenance of glucose metabolism in the fasted state. Independent clinical studies support this hypothesis. For example, an intervention study evaluating the role of *Bifidobacteria*-increasing prebiotics (fructooligosaccharides) in 35 healthy individuals has shown that prebiotics decrease the levels of butyrate-producing bacteria and have an adverse effect on glucose metabolism after an oGTT<sup>39</sup>.

The PWY-5022 finding led us to consider the roles of other SCFAs in metabolic and anthropometric traits. In our cross-sectional analysis within LL-DEEP, we detected associations between fecal propionate levels and BMI (FDR<0.1). Propionate is produced by different bacteria from those producing butyrate<sup>40</sup>, and its three genetic predictors (variance explained = 6.3%,  $F$  statistic = 21) were independent of those implicated in PWY-5022 abundance (Supplementary Tables 4 and 8). In MR analyses for the 17 traits of interest, we found that each standard-deviation increase in fecal propionate levels was causally associated with an 0.03-s.d. increase in BMI ( $P = 0.0068$ ) and an odds ratio of 1.15 for T2D ( $P = 0.004$ ) (Supplementary Table 9), although these did not pass the adjusted significance threshold described above. No associations were evident in the reverse MR analysis testing the effects of T2D and BMI on fecal propionate levels ( $P > 0.1$ ; Supplementary Table 10).

Of the two observed effects of fecal propionate on BMI and T2D, the latter was more robust. The causal relationship for increased T2D risk was robust when other MR tests were performed ( $P_{\text{MR-PRESSO}} = 0.03$ ,  $P_{\text{weighted-median}} = 0.03$ ), and there was no evidence of pleiotropy ( $P_{\text{MR-PRESSOGlobal}} = 0.75$ ,  $P_{\text{RückerQ' (modified)}} = 0.50$ ) (Supplementary Fig. 5). In contrast, the effect of propionate on increased BMI was not significant when we used other MR tests, and there was also evidence of pleiotropy ( $P_{\text{MR-PRESSOGlobal}} = 2.0 \times 10^{-3}$ ,  $P_{\text{RückerQ' (modified)}} = 9.2 \times 10^{-4}$ ; Supplementary Table 9 and Supplementary Fig. 6). The pleiotropy in the BMI effect could be accounted for by SNP rs7142308 (NC\_000014.8: g.79482379A>G) ( $P_{\text{MR-PRESSOOutlierTest}} = 0.01$ ), located within a BMI-associated locus<sup>20</sup> but independent of the lead variant (rs7141420 (NC\_000014.8: g.79899454C>T),  $r^2 = 0.01$  with rs7142308 in 1000 Genomes Europeans).

By applying MR analyses to UK Biobank summary statistics, we replicated the relationship between fecal propionate levels and increased T2D risk ( $P_{\text{IVW}} = 0.01$ ,  $P_{\text{MR-PRESSO}} = 0.007$ ,  $P_{\text{weighted-median}} = 0.04$ ;  $P_{\text{IVWcombined}} = 4 \times 10^{-5}$ ; Fig. 3), and there was no evidence of pleiotropy ( $P_{\text{MR-PRESSOGlobal}} = 0.97$ ,  $P_{\text{RückerQ' (modified)}} = 0.99$ ). The relationship between fecal propionate and BMI was again not robust to pleiotropy, thus highlighting the need for caution in interpreting this effect as causal (Supplementary Table 11).

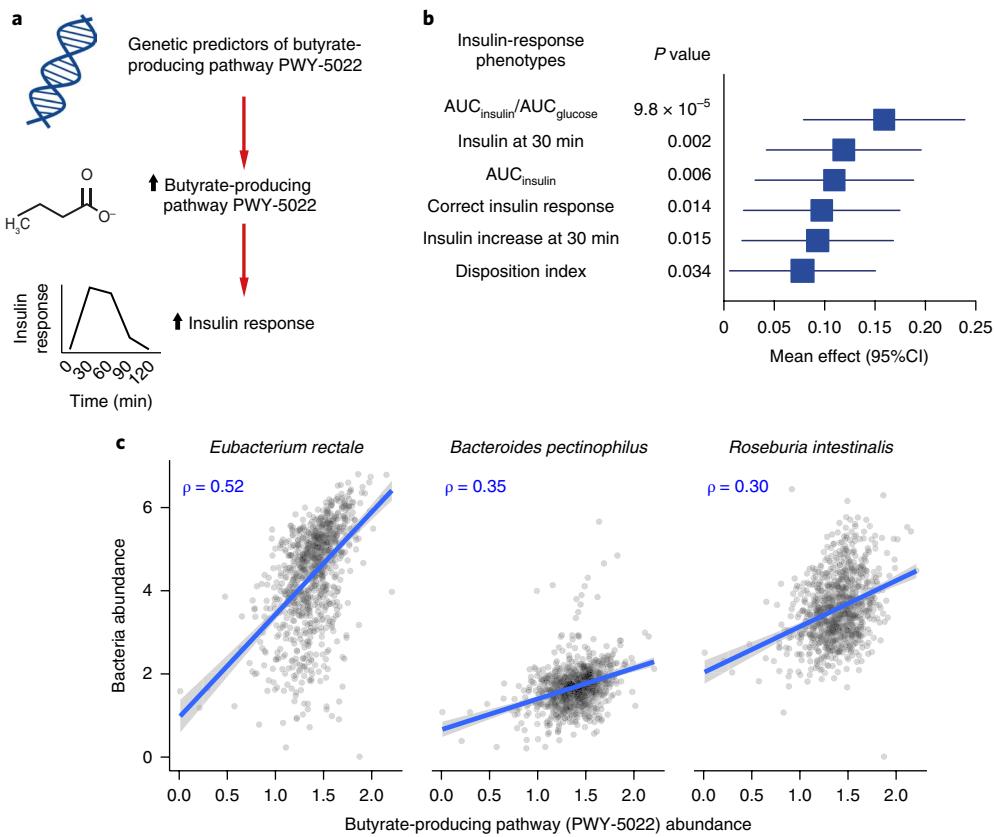


**Fig. 1 | Schematic representation of the study.** The schematic representation of our study highlights, for each step, the research question that we sought to answer, the analysis workflow and the data used. We first aimed to identify which microbiome feature (taxa, microbiome pathway or SCFA) correlated with metabolic traits in the LL-DEEP cohort (Step 1). We then performed genome-wide association analysis in LL-DEEP to identify genetic predictors of those microbiome features (Step 2) and used the genetic predictors to estimate causal relationships through bidirectional MR analysis and effect sizes for metabolic traits extracted from the summary statistics of large GWAS (Step 3). Finally, we validated our causality results by using the UK Biobank (Step 4).

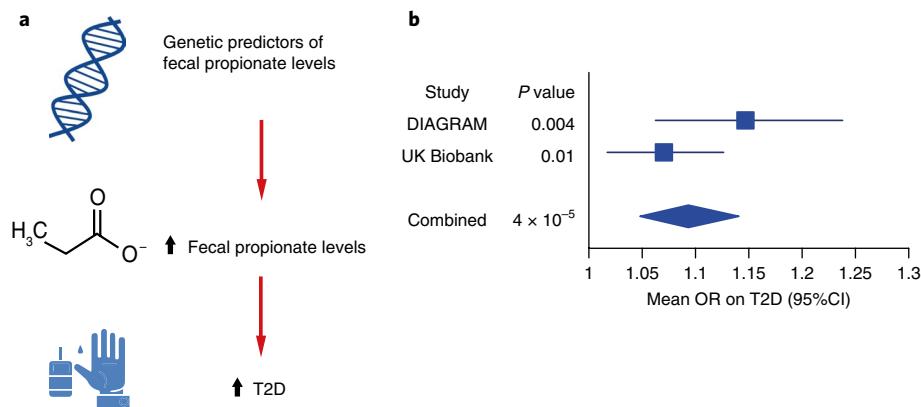
More than 95% of gut-produced SCFAs are absorbed by the host<sup>41</sup>, such that increases in fecal propionate levels may be a consequence of either increased production or decreased absorption. The latter (which would link increased fecal propionate to diminished circulating levels) would be more consistent with the preponderance of evidence indicating that SCFAs have a largely beneficial effect on energy balance and metabolic homeostasis<sup>4,10,12–14</sup>. As with plasma butyrate, plasma propionate levels were not measured in our cohorts. Further studies are warranted to

explore the mechanisms underlying this relationship between fecal propionate levels and T2D.

In summary, these data are consistent with a causal role of gut-produced SCFAs, specifically butyrate and propionate, with respect to energy balance and glucose homeostasis in humans. We showed that a genetically influenced shift in the gut microbiome toward increased production of butyrate has beneficial effects on beta-cell function, although we did not detect an effect on T2D risk. We also demonstrated that host genetic variation resulting in increased fecal



**Fig. 2 | Causal effect of butyrate-producing activity of the gut on the glucose-stimulated insulin response.** **a**, Schematic representation of the MR analysis results: genetic predisposition to higher abundance of the butyrate-producing microbiome pathway PWY-5022 (4-aminobutanoate degradation V pathway) is associated with insulin response after glucose challenge. The causal effect of PWY-5022 was also seen for other insulin-response parameters. **b**, Forest plot representing the magnitude of the effect on each parameter per 1-s.d. increase in pathway abundance, as estimated in the IVW MR analysis. MR analysis was carried out with up to nine genetic predictors and their effect sizes from LL-DEEP (952 samples) and MAGIC summary statistics (trait-specific sample sizes:  $AUC_{\text{insulin}}/AUC_{\text{glucose}} = 4,213$ ; insulin at 30 min = 4,409;  $AUC_{\text{insulin}} = 4,324$ ; correct insulin response = 4,789; insulin increase at 30 min = 4,447; disposition index = 5,130) (Methods and Supplementary Tables 4 and 5). Corresponding two-sided P values from the IVW MR test are shown. CI, confidence interval. **c**, Correlation plots with PWY-5022 abundance and the bacteria correlating the most with this abundance in 950 LL-DEEP samples (subset of the 952 normoglycemic samples for which presence of those bacteria was detected). The Spearman correlation coefficient  $\rho$  is given in blue in each panel.



**Fig. 3 | Causal effect of fecal propionate on T2D.** **a**, Schematic representation of the MR analysis results: genetic predisposition to higher fecal propionate levels is associated with increased risk of T2D. **b**, Forest plot depicting the magnitude of the causal effect on T2D for each 1-s.d. increase in fecal butyrate levels, as estimated by IVW MR analysis. The MR analysis was carried out by using the three genetic predictors derived in LL-DEEP and their effects in the discovery dataset (DIAGRAM; 26,676 T2D cases and 132,532 controls) and in the replication cohort (UK Biobank; 19,119 T2D cases and 423,698 controls). Corresponding two-sided P values from the IVW MR test are given. The effect derived by combining the two causal effects (from discovery and replication) with an inverse-variance-weighted meta-analysis approach, and the corresponding combined two-sided P values are shown at the bottom. OR, odds ratio.

propionate levels (reflecting some combination of increased production or impaired absorption) affects T2D risk.

Although the LL-DEEP cohort is the largest population study to date on the genetics of the microbiome<sup>2,17,18</sup>, it is still underpowered to capture the limited genetic component that has been estimated for microbiome features<sup>19</sup>. The results from this and other microbiome GWAS<sup>2,17,18</sup> show only limited direct overlap, thus highlighting the need for standardized protocols for data analyses and for larger sample sizes<sup>42</sup>. These will be crucial also in the context of MR analyses, because expanded GWAS carried out with standardized protocols will deliver more robust genetic predictors<sup>43</sup>. A better understanding of the complex interplay between the gut microbiome and host metabolism will require an expansion of current analyses and the ability to include measures of circulating SCFAs. Nevertheless, this study demonstrates that microbiome GWAS provide a route to causal inference that can guide and complement more direct experimental approaches, such as those based on fecal transplantation and animal models. We predict that with expanded microbiome-genetic studies (for example, the MiBioGen consortium<sup>44</sup>), MR will become a standard tool for systematically screening a large number of hypotheses generated in current and future microbiome-wide association studies.

**URLs.** MAGIC, <https://www.magicinvestigators.org/>; GIANT, [http://portals.broadinstitute.org/collaboration/giant/index.php/Main\\_Page](http://portals.broadinstitute.org/collaboration/giant/index.php/Main_Page); DIAGRAM, <http://www.diagram-consortium.org/>; UK Biobank, <http://www.ukbiobank.ac.uk/>; Human Functional Genomics Project, <http://www.humanfunctionalgenomics.org/>; Bracken, <https://github.com/jenniferlu717/Bracken/>; MetaCyc metabolic-pathway database, <http://www.metacyc.org/>; PLINK, [www.cog-genomics.org/plink2/](http://cog-genomics.org/plink2/); Michigan imputation server, <https://imputationserver.sph.umich.edu/>; R, <https://www.r-project.org/>; LDScore, <https://github.com/bulik/ldsc/>; MR-PRESSO, <https://github.com/rondolab/MR-PRESSO/>.

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at <https://doi.org/10.1038/s41588-019-0350-x>.

Received: 13 June 2018; Accepted: 10 January 2019;

Published online: 18 February 2019

## References

1. Zhernakova, A. et al. Population-based metagenomics analysis reveals markers for gut microbiome composition and diversity. *Science* **352**, 565–569 (2016).
2. Bonder, M. J. et al. The effect of host genetics on the gut microbiome. *Nat. Genet.* **48**, 1407–1412 (2016).
3. Evans, D. M. & Davey Smith, G. Mendelian randomization: new applications in the coming age of hypothesis-free causality. *Annu. Rev. Genomics Hum. Genet.* **16**, 327–350 (2015).
4. Larsen, N. et al. Gut microbiota in human adults with type 2 diabetes differs from non-diabetic adults. *PLoS One* **5**, e9085 (2010).
5. Karlsson, F. H. et al. Gut metagenome in European women with normal, impaired and diabetic glucose control. *Nature* **498**, 99–103 (2013).
6. Ley, R. E. et al. Obesity alters gut microbial ecology. *Proc. Natl Acad. Sci. USA* **102**, 11070–11075 (2005).
7. Kreznar, J. H. et al. Host genotype and gut microbiome modulate insulin secretion and diet-induced metabolic phenotypes. *Cell Rep.* **18**, 1739–1750 (2017).
8. Brunkwall, L. & Orho-Melander, M. The gut microbiome as a target for prevention and treatment of hyperglycaemia in type 2 diabetes: from current human evidence to future possibilities. *Diabetologia* **60**, 943–951 (2017).
9. Kootte, R. S. et al. Improvement of insulin sensitivity after lean donor feces in metabolic syndrome is driven by baseline intestinal microbiota composition. *Cell Metab.* **26**, 611–619.e6 (2017).
10. Zhang, X. et al. Human gut microbiota changes reveal the progression of glucose intolerance. *PLoS One* **8**, e71108 (2013).
11. Ríos-Covián, D. et al. Intestinal short chain fatty acids and their link with diet and human health. *Front. Microbiol.* **7**, 185 (2016).
12. Pingitore, A. et al. The diet-derived short chain fatty acid propionate improves beta-cell function in humans and stimulates insulin secretion from human islets in vitro. *Diabetes Obes. Metab.* **19**, 257–265 (2017).
13. Chambers, E. S. et al. Effects of targeted delivery of propionate to the human colon on appetite regulation, body weight maintenance and adiposity in overweight adults. *Gut* **64**, 1744–1754 (2015).
14. Zhao, L. et al. Gut bacteria selectively promoted by dietary fibers alleviate type 2 diabetes. *Science* **359**, 1151–1156 (2018).
15. Peng, L., He, Z., Chen, W., Holzman, I. R. & Lin, J. Effects of butyrate on intestinal barrier function in a Caco-2 cell monolayer model of intestinal barrier. *Pediatr. Res.* **61**, 37–41 (2007).
16. Schwierz, A. et al. Microbiota and SCFA in lean and overweight healthy subjects. *Obesity (Silver Spring)* **18**, 190–195 (2010).
17. Turpin, W. et al. Association of host genome with intestinal microbial composition in a large healthy cohort. *Nat. Genet.* **48**, 1413–1417 (2016).
18. Goodrich, J. K. et al. Genetic determinants of the gut microbiome in UK Twins. *Cell Host Microbe* **19**, 731–743 (2016).
19. Rothschild, D. et al. Environment dominates over host genetics in shaping human gut microbiota. *Nature* **555**, 210–215 (2018).
20. Locke, A. E. et al. Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197–206 (2015).
21. Shungin, D. et al. New genetic loci link adipose and insulin biology to body fat distribution. *Nature* **518**, 187–196 (2015).
22. Manning, A. K. et al. A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. *Nat. Genet.* **44**, 659–669 (2012).
23. Strawbridge, R. J. et al. Genome-wide association identifies nine common variants associated with fasting proinsulin levels and provides new insights into the pathophysiology of type 2 diabetes. *Diabetes* **60**, 2624–2634 (2011).
24. Soranzo, N. et al. Common variants at 10 genomic loci influence hemoglobin A<sub>1</sub>(C) levels via glycemic and nonglycemic pathways. *Diabetes* **59**, 3229–3239 (2010).
25. Prokopenko, I. et al. A central role for GRB10 in regulation of islet function in man. *PLoS Genet.* **10**, e1004235 (2014).
26. Saxena, R. et al. Genetic variation in GIPR influences the glucose and insulin responses to an oral glucose challenge. *Nat. Genet.* **42**, 142–148 (2010).
27. Scott, R. A. et al. An expanded genome-wide association study of type 2 diabetes in Europeans. *Diabetes* **66**, 2888–2902 (2017).
28. Li, Y. et al. A functional genomics approach to understand variation in cytokine production in humans. *Cell* **167**, 1099–1110.e14 (2016).
29. Burgess, S., Butterworth, A. & Thompson, S. G. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet. Epidemiol.* **37**, 658–665 (2013).
30. Verbanck, M., Chen, C. Y., Neale, B. & Do, R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nat. Genet.* **50**, 693–698 (2018).
31. Bowden, J., Davey Smith, G., Haycock, P. C. & Burgess, S. Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. *Genet. Epidemiol.* **40**, 304–314 (2016).
32. Bowden, J., Davey Smith, G. & Burgess, S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int. J. Epidemiol.* **44**, 512–525 (2015).
33. Bowden, J. et al. Improving the accuracy of two-sample summary data Mendelian randomization: moving beyond the NOME assumption. Preprint at <https://www.biorxiv.org/content/early/2018/10/11/159442> (2018).
34. Rücker, G., Schwarzer, G., Carpenter, J. R., Binder, H. & Schumacher, M. Treatment-effect estimates adjusted for small-study effects via a limit meta-analysis. *Biostatistics* **12**, 122–142 (2011).
35. Bycroft, C. et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
36. Duncan, S. H., Hold, G. L., Barcenilla, A., Stewart, C. S. & Flint, H. J. *Roseburia intestinalis* sp. nov., a novel saccharolytic, butyrate-producing bacterium from human faeces. *Int. J. Syst. Evol. Microbiol.* **52**, 1615–1620 (2002).
37. Pryde, S. E., Duncan, S. H., Hold, G. L., Stewart, C. S. & Flint, H. J. The microbiology of butyrate formation in the human colon. *FEMS Microbiol. Lett.* **217**, 133–139 (2002).
38. Jakobsdóttir, G., Bjørregaard, J. H., Skovbjerg, H. & Nyman, M. Fasting serum concentration of short-chain fatty acids in subjects with microscopic colitis and celiac disease: no difference compared with controls, but between genders. *Scand. J. Gastroenterol.* **48**, 696–701 (2013).
39. Liu, F. et al. Fructooligosaccharide (FOS) and galactooligosaccharide (GOS) increase *Bifidobacterium* but reduce butyrate producing bacteria with adverse glycemic metabolism in healthy young population. *Sci. Rep.* **7**, 11789 (2017).
40. Louis, P. & Flint, H. J. Formation of propionate and butyrate by the human colonic microbiota. *Environ. Microbiol.* **19**, 29–41 (2017).

41. den Besten, G. et al. The role of short-chain fatty acids in the interplay between diet, gut microbiota, and host energy metabolism. *J. Lipid Res.* **54**, 2325–2340 (2013).
42. Kurihikov, A., Wijmenga, C., Fu, J. & Zhernakova, A. Host genetics and gut microbiome: challenges and perspectives. *Trends Immunol.* **38**, 633–647 (2017).
43. Taylor, A. E. et al. Mendelian randomization in health research: using appropriate genetic variants and avoiding biased estimates. *Econ. Hum. Biol.* **13**, 99–106 (2014).
44. Wang, J. et al. Meta-analysis of human genome-microbiome association studies: the MiBioGen consortium initiative. *Microbiome* **6**, 101 (2018).

## Acknowledgements

We thank the participants and staff of the LL-DEEP cohort for their collaboration, the UMCG Genomics Coordination center, the UG Center for Information Technology and their sponsors BBMRI-NL and TarGet for storage and compute infrastructure. We are also grateful to M. J. Bonder for help in formatting summary statistics; to R. K. Weersma and Y. Li for discussions; and to K. Mc Intyre for editing the manuscript. Part of this work was conducted by using the UK Biobank resource under application no. 9161. This project was funded by IN-CONTROL CVON grant CVON2012-03 to M.G.N., A.Z., L.A.B.J. and J.F.; Top Institute Food and Nutrition (TiFN, Wageningen, the Netherlands) grant TiFN GH001 to C.W.; the Netherlands Organization for Scientific Research (NWO) grants NWO-VENI 016.176.006 to M.O., NWO-VIDI 864.13.013 to J.F. and NWO-VIDI 016.Vidi.178.056 to A.Z.; NWO Spinoza Prizes SPI 92-266 to C.W. and SPI 94-212 to M.G.N.; European Research Council (ERC) starting grant ERC no. 715772 to A.Z.; FP7/2007-2013/ERC Advanced Grant (agreement 2012-322698) to C.W.; ERC Consolidator Grant ERC no. 310372 to M.G.N.; Tripartite Immunometabolism consortium (TrIC)-Novo Nordisk Foundation grant NNF15CC0018486 to M.I.M.; and Wellcome grants 090532, 098381, 106130 and 203141 to M.I.M. A.Z. is also supported by a Rosalind Franklin Fellowship from the University of Groningen. M.I.M. is supported as a Wellcome Senior Investigator and a National Institute of Health Research Senior Investigator. The funders had no role in study design, data collection and analysis,

decision to publish, or preparation of the manuscript. The views expressed in this article are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health.

## Author contributions

S.S. performed statistical analyses on the LifeLines and 500FG cohorts; N.R.v.Z. and A.M. performed statistical analyses on UK Biobank and DIAGRAM studies; A.K. and A.V.V. processed raw microbiome data in Lifelines-DEEP and 500FG; U.V. and L.F. downloaded and harmonized the summary statistics from the GIANT, MAGIC and DIAGRAM consortia; L.F., and C.W. provided LifeLines-DEEP data; Z.M., A.A.M.M. and D.M.A.E.J. provided critical input in manuscript revisions; M.O., L.A.B.J. and M.G.N. provided 500FG data; S.S., N.R.v.Z. and M.I.M. wrote the manuscript, to which J.F., A.Z. and C.W. provided critical input; S.S., N.R.v.Z., A.M., C.W. and M.I.M. designed the study. All authors read, revised and approved the manuscript.

## Competing interests

M.I.M. serves on advisory panels for Pfizer, NovoNordisk and Zoe Global; has received honoraria from Pfizer, NovoNordisk and Eli Lilly; has stock options in Zoe Global; and has received research funding from Abbvie, Astra Zeneca, Boehringer Ingelheim, Eli Lilly, Janssen, Merck, NovoNordisk, Pfizer, Roche, Sanofi Aventis, Servier and Takeda. All other authors declare no competing financial interests.

## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41588-019-0350-x>.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Correspondence and requests for materials** should be addressed to S.S., C.W. or M.I.M.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2019

## Methods

**Study samples.** The discovery cohort of this study is LL-DEEP, a population-based cohort of 1,539 individuals from the northern Netherlands (age range 18–84 years) that is a subset of the largest Lifelines biobank ( $n=167,000$ ). For all LL-DEEP volunteers, an extensive dataset of measured and self-reported phenotypic information has been collected, as well as blood and stool specimens, as described previously<sup>45,46</sup>. Measurement of SCFAs in stool was carried out through gas chromatography–mass spectrometry according to ref. <sup>47</sup>.

To identify the appropriate threshold for the selection of genetic predictors of microbiome features, we used the 500 Functional Genomics (500FG) cohort<sup>28</sup>, an independent cohort of 534 healthy individuals from the Netherlands (age range 18–75 years). The protocols for stool collection and metagenomic sequencing were similar to those used in LL-DEEP, as previously described<sup>48</sup>.

All participants from both studies signed an informed consent form. The LL-DEEP study was approved by the institutional ethics review boards of the UMCG (ClinicalTrials.gov NCT00775060). The 500FG study was approved by the Ethical Committee of Radboud University Nijmegen (NL42561.091.12, 2012/550).

To replicate our findings, we used genotype and phenotype data from the UK Biobank, a study of 500,000 subjects from the United Kingdom who were 45–65 years of age<sup>35</sup>. Each participant provided a blood sample for DNA extraction and completed a detailed questionnaire providing baseline data. Individuals are also linked to electronic medical records on a number of traits including BMI and T2D.

**Data generation and preprocessing.** *Genotyping.* Genotype data were available for 1,268 LL-DEEP volunteers, as previously described<sup>2,45</sup>. In brief, genotyping was carried out with two Illumina arrays, HumanCytoSNP-12 BeadChip and ImmunoChip. After standard per-sample and per-SNP quality-control filters, data from the two arrays were merged, and additional markers were imputed with HRC reference panel v1.1 (ref. <sup>49</sup>) on the Michigan server (see URLs). In our analyses, we focused on 15,001,957 variants with imputation accuracy RSQR > 0.3. In the 500FG cohort, 516 samples were genotyped with the Illumina Human OmniExpress Exome-8 v1.0 SNP chip and, after standard quality-control checks<sup>28</sup>, were imputed with the same procedure and reference panel used with LL-DEEP. The UK Biobank samples were genotyped with the Affymetrix UK BiLEVE Axiom array on an initial 50,000 participants. The remaining 450,000 participants were genotyped with the Affymetrix UK Biobank Axiom array<sup>35</sup>. Quality control on samples and genotypes was performed centrally, and subsequent imputation was performed with the HRC reference panel at Wellcome Centre Human Genetics.

*Metagenomic sequencing.* Metagenomic sequencing of the gut microbiome was performed with the Illumina HiSeq platform on 1,179 LL-DEEP samples. After application of per-sample and per-read quality filters<sup>2</sup>, the profile of microbial composition was determined with the Bracken pipeline (see URLs). In total, 903 taxonomies were identified and normalized with log transformation; normalized nonzero values were then adjusted for age, sex and read depth with linear regression.

Functional profiling was performed with HUMAnN2 (v 0.4.0), which maps reads to a customized database of functionally annotated pangenomes<sup>50</sup>. This analysis identified 742 pathways from the MetaCyc metabolic-pathway database<sup>51</sup>. Similarly to the process for taxonomy data, pathway abundance values were normalized through log transformation, and the normalized nonzero values were corrected for age, sex and read depth. We considered only nonzero values for analyses and therefore restricted analyses to microbiome features (taxonomies and pathways) that had nonzero values in less than 50% of the samples and retained only one member of pairs of pathways or bacteria showing >0.99 Spearman correlation. This filtering resulted in a final set of 796 features (273 taxa and 523 pathways) that were used for analyses.

We further confined all statistical analyses to normoglycemic samples with good-quality genetic and microbiome data. Normoglycemic status was assigned to samples from individuals not reported to have diabetes or to be taking oral antidiabetes medications and who had fasting glucose levels <7 mmol/L. We also removed individuals who were taking antibiotics at the time of stool collection. This filtering resulted in a final set of 952 samples available for analyses. In the 500FG cohort, we used the same filters and selected 445 normoglycemic samples with both genetic and microbiome data for analyses.

*Genome-wide association scans of anthropometric and glycemic traits.* We downloaded full GWAS summary statistics from nine studies representing 17 GWAS for different anthropometric and glycemic traits. These traits were BMI and WHR, fasting glucose, insulin and proinsulin, 2-h glucose, HOMA-derived measurements of insulin resistance (HOMA-IR) and sensitivity (HOMA-B), glycated hemoglobin (HbA<sub>c</sub>), T2D and seven insulin-response parameters measured during an oGTT (Supplementary Table 1 and URLs). SNP names and genomic positions were aligned to the genomic build GRCh37/hg19.

**Statistical analysis.** *Correlation of SCFAs and microbiome features with anthropometric and glycemic traits.* We correlated five SCFAs (acetate, butyrate, propionate, caproate and valerate) and 796 other microbiome features (taxa or pathways) with measured anthropometric (BMI and WHR) and glycemic traits

(fasting glucose, insulin, HbA<sub>c</sub>, HOMA-IR and HOMA-B) in the LL-DEEP cohort. Anthropometric and glycemic traits were adjusted for age, sex and BMI (except for BMI phenotype). We used the nonparametric Spearman correlation test (*cor.test(method = "Spearman")* function in R (v3.3)) and considered results significant when the multiple-testing-adjusted two-sided *P* value was <0.1. The multiple-testing-adjusted *P* value, FDR *P*, was calculated with the Benjamini–Hochberg procedure in the *p.adjust()* function in R (v3.3) (see URLs).

*Genome-wide association analyses of SCFAs and microbiome features.* For each microbiome feature and SCFA, we performed a genome-wide-association scan in LL-DEEP samples by reprocessing data from our previous study in a different manner<sup>2</sup>. In particular: (i) we remapped metagenomic reads to a more recent database, (ii) we restricted analyses to only normoglycemic samples and those from subjects not taking antibiotics and (iii) we performed genetic analyses with a linear mixed model accounting for population structure instead of the Spearman correlation method. In particular, for genetic analyses we used EPACTS (v3.2.6)<sup>32</sup>, a program that performs a linear mixed model adjusted with a genomic-based kinship matrix calculated with all quality-checked genotyped autosomal SNPs with minor allele frequency >1%. The advantage of this model is that the kinship matrix encodes a wide range of sample structures, including both cryptic relatedness and population stratification, thus producing more robust results than standard linear regression. All traits were inverse-quantile normalized before genetic analysis. Specifically for SCFAs, age, sex, chromogranin A, stool type according to the Bristol scale and BMI were added as covariates.

The variance explained (adjusted  $r^2$ ) and the *F* statistic for each microbiome feature were extracted from a linear model that fitted all the selected genetic predictors on the normalized, covariate-adjusted microbiome feature.

*MR analyses with 17 GWAS traits.* The MR procedure consists of two steps: (i) identification of proper instrumental variables or genetic predictors, i.e., variants independently associated with the exposure factor and (ii) calculation of causal estimates. For each GWAS summary statistic, we first selected independent SNPs with the clumping procedure in PLINK v1.9 (see URLs), setting a linkage-disequilibrium threshold of  $r^2 < 0.1$  in a 500-kb window. Linkage disequilibrium was calculated with the LL-DEEP cohort when the clumping procedure was run on the GWAS of microbiome features and SCFAs, whereas for GWAS of anthropometric and glycemic traits, we used the linkage-disequilibrium estimates from the 1000 Genomes phase 3 European samples.

Furthermore, because most of the downloaded GWAS were based on the HapMap2 genetic map, for each independently associated variant, we identified the best HapMap2 proxy ( $r^2 > 0.8$ ) or discarded that variant if no proxy was available.

Finally, we selected only variants that showed association at  $P < 1 \times 10^{-5}$ . We identified this as the optimal *P*-value threshold to use for selection of genetic predictors associated with microbiome features, because this threshold led to a larger variance explained, on average, of the same microbiome features in the 500FG cohort (Supplementary Fig. 1). For consistency, we used the same threshold and procedure for selecting genetic predictors from the downloaded GWAS on anthropometric and glycemic traits.

To calculate causal estimates, we used the IVW method<sup>32</sup> as a two-sample MR analysis of summary association statistics of the exposure and the outcome. Specifically, we estimated the causal effect in a fixed-effect meta-analysis framework, i.e., as a sum of single-SNP causal effects (derived as a ratio of the SNP effect on the outcome by the SNP effect on the exposure), weighted by the inverse of their variance (derived as a squared ratio of the SNP standard deviation on the outcome to the SNP effect on the exposure). The *P* value was calculated as  $P = 2 \times (1 - \Phi(Z))$ , where  $\Phi(Z)$  is the standard normal cumulative distribution function, and  $Z$  is the ratio of the combined (with inverse-variance weights) causal effect and its standard error. Of note, the causal estimate is equivalent to that obtained as a weighted linear regression of the outcome SNP effects on the exposure SNP effects with a fixed intercept of 0 and with the inverse of the variance of the effect sizes on the outcome as weights. For analyses, we set the effect allele of the genetic predictors to be the allele with the positive direction. We also calculated causal estimates with additional MR methods: MR-PRESSO<sup>30</sup>, which removes pleiotropy by identifying and discarding influential outlier predictors from the IVW test and uses a *t* test to calculate *P* values; the weighted-median test<sup>31</sup>, which uses a statistical estimator robust to the presence of pleiotropy in a subset (<50%) of the predictors; and MR-Egger<sup>32</sup>, which adjusts for average horizontal pleiotropy and assumes that >50% of the predictors have pleiotropy. Furthermore, we specifically evaluated the presence of pleiotropy with the MR-PRESSO Global test<sup>30</sup> and modified Rücker's *Q'* test<sup>33</sup>.

*Calculation of significance threshold.* To define our significance threshold for the IVW-based MR analyses, we first ran a principal component analysis of the 245 microbiome features and observed that the total variability could be explained by the first 57 principal-component axes. To derive the number of independent anthropometric and metabolic traits out of the 17 of interest, we used pairwise genetic correlation calculated with LDScore regression (LDScore v1.0.0). Variants were restricted to those from HapMap3, and precomputed LD scores estimated in subjects of European descent were used as recommended by the authors<sup>53</sup>. Traits

were hierarchically clustered on the basis of genetic correlation values,  $\rho_g$ , with the dissimilarity metric  $(1 - \rho_g)/2$  (Supplementary Fig. 2). The number of resulting clusters was used to define the number of independent traits. Because genetic correlation could not be calculated with four insulin-secretion traits, we counted those as fully independent traits. We set our multiple-testing significance threshold at  $1.3 \times 10^{-4}$  ( $0.05/(57 \times 7)$ ).

**MR analyses in UK Biobank.** We first calculated the association of the 12 genetic predictors (nine for PWY-5022 and three for fecal propionate) with seven metabolic and anthropometric traits (BMI, body-fat percentage, WHR, visceral adipose tissue, subcutaneous adipose tissue, obesity and T2D) with a linear mixed model, as implemented in BOLT-LMM (v2.3.2)<sup>54</sup>. T2D status was defined according to the definition used in ref.<sup>55</sup>; BMI was defined according to that used by the GIANT consortium<sup>20</sup>, and obesity was defined by ICD code 278. Analyses were restricted to 442,817 individuals of European descent and were adjusted for age, sex, genotyping array and six genetic principal components; WHR was also adjusted for BMI. We then used the summary statistics at these 12 variants to estimate causal relationships and investigate presence of pleiotropy by applying the same statistical tests used with the GWAS summary statistics and described in the previous paragraph.

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

### Data availability

The LifeLines-DEEP metagenomic sequencing data are available at the European Genome-phenome Archive (EGA) under accession code EGAS00001001704. Genotype and phenotype data can be requested from the Lifelines Biobank at <https://www.lifelines.nl/researcher/biobank-lifelines/application-process/>.

Summary statistics for metabolic traits were downloaded from the MAGIC, GIANT and DIAGRAM websites (see URLs).

### References

45. Tigchelaar, E. F. et al. Cohort profile: LifeLines DEEP, a prospective, general population cohort study in the northern Netherlands: study design and baseline characteristics. *BMJ Open* **5**, e006772 (2015).
46. Li, N. et al. Pleiotropic effects of lipid genes on plasma glucose, HbA1c, and HOMA-IR levels. *Diabetes* **63**, 3149–3158 (2014).
47. García-Villalba, R. et al. Alternative method for gas chromatography-mass spectrometry analysis of short-chain fatty acids in faecal samples. *J. Sep. Sci.* **35**, 1906–1913 (2012).
48. Schirmer, M. et al. Linking the human gut microbiome to inflammatory cytokine production capacity. *Cell* **167**, 1125–1136.e8 (2016).
49. McCarthy, S. et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* **48**, 1279–1283 (2016).
50. Franzosa, E. A. et al. Species-level functional profiling of metagenomes and metatranscriptomes. *Nat. Methods* **15**, 962–968 (2018).
51. Vatanen, T. et al. Variation in microbiome LPS immunogenicity contributes to autoimmunity in humans. *Cell* **165**, 842–853 (2016).
52. Kang, H. M. et al. Variance component model to account for sample structure in genome-wide association studies. *Nat. Genet.* **42**, 348–354 (2010).
53. Bulik-Sullivan, B. K. et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
54. Loh, P. R. et al. Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat. Genet.* **47**, 284–290 (2015).
55. Eastwood, S. V. et al. Algorithms for the capture and adjudication of prevalent and incident diabetes in UK Biobank. *PLoS One* **11**, e0162388 (2016).

## 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](#).

### Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a	Confirmed
<input type="checkbox"/>	<input checked="" type="checkbox"/> The exact sample size ( <i>n</i> ) for each experimental group/condition, given as a discrete number and unit of measurement
<input type="checkbox"/>	<input checked="" type="checkbox"/> An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
<input type="checkbox"/>	<input checked="" type="checkbox"/> The statistical test(s) used AND whether they are one- or two-sided <i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>
<input type="checkbox"/>	<input checked="" type="checkbox"/> A description of all covariates tested
<input type="checkbox"/>	<input checked="" type="checkbox"/> A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
<input type="checkbox"/>	<input checked="" type="checkbox"/> A full description of the statistics 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)
<input type="checkbox"/>	<input checked="" type="checkbox"/> For null hypothesis testing, the test statistic (e.g. <i>F</i> , <i>t</i> , <i>r</i> ) with confidence intervals, effect sizes, degrees of freedom and <i>P</i> value noted <i>Give P values as exact values whenever suitable.</i>
<input checked="" type="checkbox"/>	<input type="checkbox"/> For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
<input checked="" type="checkbox"/>	<input type="checkbox"/> For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
<input type="checkbox"/>	<input checked="" type="checkbox"/> Estimates of effect sizes (e.g. Cohen's <i>d</i> , Pearson's <i>r</i> ), indicating how they were calculated
<input type="checkbox"/>	<input checked="" type="checkbox"/> Clearly defined error bars <i>State explicitly what error bars represent (e.g. SD, SE, CI)</i>

*Our web collection on [statistics for biologists](#) may be useful.*

### Software and code

Policy information about [availability of computer code](#)

Data collection	NA
Data analysis	<p>For data analyses we used the following open-source software: HUMAN2 (v 0.4.0) and Bracker for metagenomic analyses, EPACTS v(3.2.6) , Plink (v 1.9), BOLT-LMM (v2.3.2) and LDscore regression (v1.0.0) for genetic analyses. For statistical analyses (correlation, implementation of MR tests) we used functions from the open source R software (v3.3) For all software, the corresponding references, the link for download and version used are specified in the text</p>

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 upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

## Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

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

The LifeLines-DEEP metagenomics sequencing data are available at the European Genome-phenome Archive (EGA), with access code EGAS00001001704. Genotype and phenotype data can be requested from the Lifelines Biobank <https://www.lifelines.nl/researcher/biobank-lifelines/application-process>. Summary statistics for anthropometric and metabolic traits were downloaded from MAGIC, GIANT and DIAGRAM websites (see URLs).

## Field-specific reporting

Please select 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/authors/policies/ReportingSummary-flat.pdf](https://nature.com/authors/policies/ReportingSummary-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 this study. We used all available Lifelines-DEEP samples that were collected from previous studies (Bonder et al Nat Gen 2016), and reprocessed using a different statistical method and sample exclusion criteria. This is clearly described in the text and in the Methods section
Data exclusions	For analyses, we first started from all Lifelines-DEEP and 500FG samples from Bonder et al Nat Gen 2016 with quality controlled microbiome and genetic data, and then retained only normo-glycaemic individuals. Normo-glycaemic status was assigned to samples not reported to have diabetes or to be taking oral anti-diabetes medications and who had fasting glucose levels <7 mmol/L. We also removed individuals who were taking antibiotics at the time of the stool collection. Exclusions are clearly described in the Methods section
Replication	We searched for replication of our findings in the UK Biobank cohort. We replicated our observation in the discovery data set except for one trait (insulin response after an oral glucose tolerance test) which was not measured in this cohort.
Randomization	This is not an experimental study. Randomization is not applicable
Blinding	This is not an experimental study. Blinding is not applicable

## Reporting for specific materials, systems and methods

### Materials & experimental systems

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Unique biological materials
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input type="checkbox"/>	<input checked="" type="checkbox"/> Human research participants

### Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

## Human research participants

Policy information about [studies involving human research participants](#)

### Population characteristics

This study includes data from 3 population cohorts: Lifelines-DEEP, 500FG and UK Biobank. The LifeLines-DEEP (LL-DEEP), is a population-based cohort of 1,539 individuals from Northern Netherlands (age range 18–84 years) that is a subset of the largest Lifelines biobank (N=167,000). The 500 Functional Genomics (500FG) cohort is a population-based cohort of 534 healthy subjects from the Netherlands (age-

range 18–75 years).

The UK Biobank is a population-based cohort of 500,000 subjects from the United Kingdom aged 45–65 years of age.

## Recruitment

Individuals were recruited independently of this study, and was based on voluntary participation after an invitation letter.