ARTICLE

doi:10.1038/nature18646

Human gut microbes impact host serum metabolome and insulin sensitivity

Helle Krogh Pedersen[1]*, Valborg Gudmundsdottir[1]*, Henrik Bjørn Nielsen[1]*, Tuulia Hyotylainen[2,3,4]*, Trine Nielsen[5]*, Benjamin A. H. Jensen[6], Kristoffer Forslund[7], Falk Hildebrand[7,8,9], Edi Prifti[10,11], Gwen Falony[9,12], Emmanuelle Le Chatelier[10], Florence Levenez[10], Joel Doré[10,13], Ismo Mattila[4,14], Damian R. Plichta[1], Päivi Pöhö[4,15], Lars I. Hellgren[1], Manimozhiyan Arumugam[5], Shinichi Sunagawa[7,16], Sara Vieira-Silva[9,12], Torben Jørgensen[17,18], Jacob Bak Holm[6], Kajetan Trošt[14], MetaHIT Consortium†, Karsten Kristiansen[6,19], Susanne Brix[1], Jeroen Raes[8,9,12], Jun Wang[6,19,20,21,22], Torben Hansen[5,23], Peer Bork[7,24,25,26], Søren Brunak[1,27], Matej Oresic[3,4,14], S. Dusko Ehrlich[10,28] & Oluf Pedersen[5,17]

Insulin resistance is a forerunner state of ischaemic cardiovascular disease and type 2 diabetes. Here we show how the human gut microbiome impacts the serum metabolome and associates with insulin resistance in 277 non-diabetic Danish individuals. The serum metabolome of insulin-resistant individuals is characterized by increased levels of branched-chain amino acids (BCAAs), which correlate with a gut microbiome that has an enriched biosynthetic potential for BCAAs and is deprived of genes encoding bacterial inward transporters for these amino acids. *Prevotella copri* and *Bacteroides vulgatus* are identified as the main species driving the association between biosynthesis of BCAAs and insulin resistance, and in mice we demonstrate that *P. copri* can induce insulin resistance, aggravate glucose intolerance and augment circulating levels of BCAAs. Our findings suggest that microbial targets may have the potential to diminish insulin resistance and reduce the incidence of common metabolic and cardiovascular disorders.

Insulin resistance (IR) and metabolic syndrome are risk factors for both type 2 diabetes and ischaemic cardiovascular diseases, pathologies that are in epidemic growth worldwide. Mounting evidence suggests a link between the gut microbiome and human metabolic health[1–4], with transferability of insulin resistance phenotypes through faecal microbiome transplants[5,6]. These effects may partly be mediated through the metabolome[7]. Serum levels of amino acids, most consistently the BCAAs[8,9], triacylglycerols with low carbon number and double bonds[10,11], as well as specific membrane phospholipids[12], have previously been associated with IR and future risk of metabolic and cardiovascular morbidities. However, the origin of the abnormal IR-associated serum metabolome is largely unknown[13].

To explore the relationships between the fasting serum metabolome and the gut microbiome in the states of IR and metabolic syndrome, we examined 291 non-diabetic Danish adults, of whom 277 had available gut microbial data from the MetaHIT[1] study population. IR was estimated by homeostatic model assessment (HOMA-IR), which in epidemiological studies is a widely applied measure of IR, primarily as an estimate of hepatic IR[14]. Since IR is largely influenced by body mass index (BMI), we also estimated the IR index adjusted for BMI ($HOMA\text{-}IR_{BMIadj}$). Metabolic syndrome was defined according to recommendations by the International Diabetes Federation[15] (see Methods). For comparative analyses we further included 75 Danish type 2 diabetes patients[4] with quantitative gut metagenomics data generated by the same experimental protocol. Characteristics of the study samples are given in Supplementary Table 1 and Extended Data Fig. 1.

Results

In the present study, we assessed the role of the gut microbiome as a source for key features of the serum metabolome profile, predicting metabolic and cardiovascular disorders in non-diabetic lean and obese people. Untargeted metabolome profiles were generated on fasting serum samples, applying two mass spectrometry-based analytical platforms and providing information about 325 polar metabolites (94 known, 231 unknown) and 876 molecular lipids (289 known, 587 unknown), respectively (here collectively termed serum metabolites). These were binned into 74 co-abundance clusters across all individuals (Fig. 1a, Supplementary Tables 2 and 3). We found 19 of the 74 metabolite clusters (26%), comprising 26 polar metabolites and 367 lipids, to be significantly associated with both IR and metabolic syndrome, with consistent directionality across the 291 non-diabetic

[1]Center for Biological Sequence Analysis, Dept. of Systems Biology, Technical University of Denmark, DK-2800 Kongens Lyngby, Denmark. [2]University of Örebro, SE-702 81 Örebro, Sweden. [3]Turku Centre for Biotechnology, University of Turku and Åbo Akademi University, FI-20520 Turku, Finland. [4]VTT Technical Research Centre of Finland, FI-02044 Espoo, Finland. [5]The Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, DK-2200 Copenhagen, Denmark. [6]Laboratory of Genomics and Molecular Biomedicine, Department of Biology, University of Copenhagen, DK-2100 Copenhagen, Denmark. [7]European Molecular Biology Laboratory, 69117 Heidelberg, Germany. [8]Department of Bioscience Engineering, Vrije Universiteit Brussel, 1050 Brussels, Belgium. [9]Center for the Biology of Disease, VIB, 3000 Leuven, Belgium. [10]MGP MetaGénoPolis, INRA, Université Paris-Saclay, 78350 Jouy en Josas, France. [11]Institute of Cardiometabolism and Nutrition (ICAN), 75013 Paris, France. [12]Department of Microbiology and Immunology, Rega Institute, KU Leuven, 3000 Leuven, Belgium. [13]Micalis Institute, INRA, AgroParisTech, Université Paris-Saclay, 78350 Jouy-en-Josas, France. [14]Steno Diabetes Center, DK-2820 Gentofte, Denmark. [15]Faculty of Pharmacy, University of Helsinki, FI-00014 Helsinki, Finland. [16]Institute of Microbiology, ETH Zurich, CH-8092 Zurich, Switzerland. [17]Faculty of Health and Medical Sciences, University of Copenhagen, DK-2200 Copenhagen, Denmark. [18]Research Centre for Prevention and Health, Centre for Health, Capital region, Glostrup Hospital, DK-2600 Glostrup, Denmark. [19]BGI-Shenzhen, 518083 Shenzhen, China. [20]Princess Al Jawhara Albrahim Center of Excellence in the Research of Hereditary Disorders, King Abdulaziz University, Jeddah, Saudi Arabia. [21]Macau University of Science and Technology, Avenida Wai long, Taipa, Macau. [22]Department of Medicine and State Key Laboratory of Pharmaceutical Biotechnology, University of Hong Kong, Hong Kong. [23]Faculty of Health Sciences, University of Southern Denmark, DK-5000 Odense, Denmark. [24]Molecular Medicine Partnership Unit, University of Heidelberg and European Molecular Biology Laboratory, 69120 Heidelberg, Germany. [25]Max Delbrück Centre for Molecular Medicine, D-13125 Berlin, Germany. [26]Department of Bioinformatics, University of Wuerzburg, D-97074 Würzburg, Germany. [27]Novo Nordisk Foundation Center for Protein Research, Disease Systems Biology, Faculty of Health and Medical Sciences, University of Copenhagen, DK-2200 Copenhagen, Denmark. [28]King's College London, Centre for Host–Microbiome Interactions, Dental Institute Central Office, Guy's Hospital, SE1 9RT London, UK.

*These authors contributed equally to this work.

†Lists of participants and their affiliations appear in the Supplementary Information.

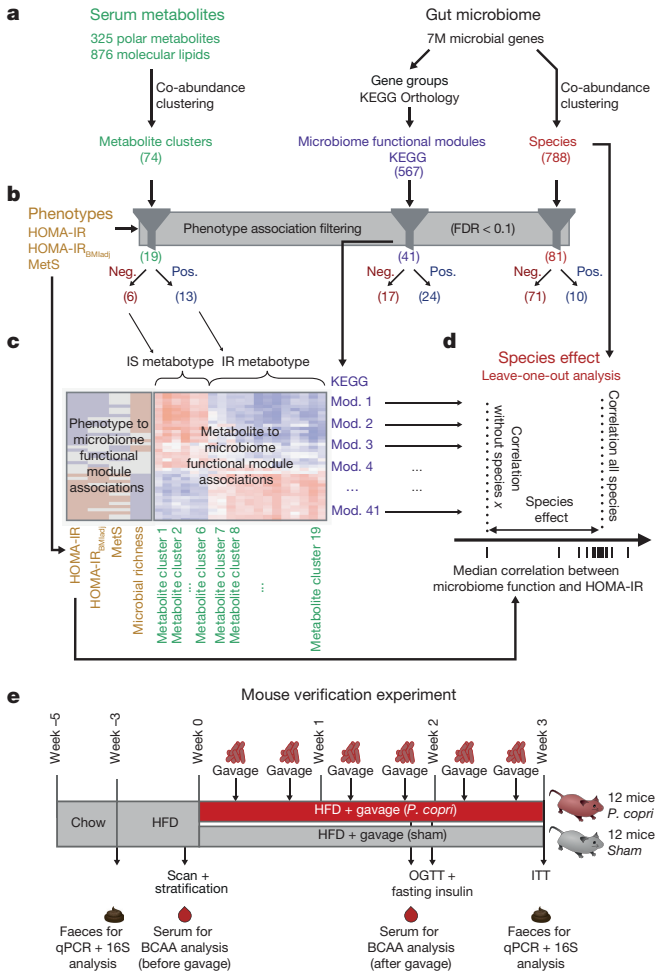


Figure 1 | Overview of the workflow integrating human phenotypes, fasting serum metabolome, gut microbiome data and mouse feeding experiments. a, Metabolites were summarized as co-abundance clusters, and functional module and species abundance profiles extracted from gut microbiome data. **b**, Features were filtered for significant positive or negative associations with HOMA-IR, HOMA-IR$_{BMIadj}$ or metabolic syndrome. **c**, Metabolite clusters were divided into IR- and IS-metabotypes and associated with microbiome functional modules. **d**, Microbial driver species for the functional module associations with HOMA-IR were identified using a leave-one-out analysis. **e**, Experimental design for *P. copri* feeding mouse experiment. HFD, high-fat diet; OGTT, oral glucose tolerance test; ITT, insulin tolerance test; MetS, metabolic syndrome.

individuals (Spearman rank correlation test, false discovery rate (FDR) < 0.1, Extended Data Fig. 2 and Supplementary Tables 4 and 5). These 19 metabolite clusters separated into two groups that were either positively or negatively correlated with both IR and metabolic syndrome, referred hereafter to as the IR-metabotype and insulin sensitivity (IS)-metabotype, respectively (Fig. 1b, Supplementary Table 6, Supplementary Results and Extended Data Fig. 3). Most of these associations were confirmed in type 2 diabetes patients (see Supplementary Results for details).

In accordance with previous reports, the identifiable serum metabolites constituting the IR-metabotype mainly comprised amino acids (including BCAAs; see Supplementary Discussion), tricarboxylic acid cycle metabolites, and a large number of triacylglycerols (Supplementary Table 6, previous findings referenced therein). Interestingly, we found that serum BCAAs clustered with hydrocinnamic and indole-3-lactic acids (metabolite cluster M10, Supplementary Table 2), both of which belong to a class of phenolic compounds that are associated with gut microbial metabolism[16]. Serum levels of hydrocinnamic acid are affected by gut microbiota in mice[17] and hydrocinnamic acid is a known inhibitor of branched-chain α-keto acid dehydrogenase kinase, which regulates the breakdown of BCAAs[18]. By contrast, the IS-metabotype consisted solely of serum lipids, primarily phospholipids and triacylglycerols with odd carbon number and high double bond content (Supplementary Table 6). Hence, our results confirm previous reports of odd-chain fatty acids and polyunsaturated fatty acids as associated with reduced risk of type 2 diabetes[10,19]. Plasma odd-chain fatty acids are often considered to be a product of microbial metabolism or a marker of dietary intake, especially of dairy products, although recent findings show that C15:0 and C17:0 are also endogenously produced in adipocytes through α-oxidation of palmitic and stearic acid[20,21]. Based on our data, however, it is not possible to distinguish if these odd-chain fatty acids originate from dietary intake or bacterial metabolism. Surprisingly, the IS-metabotype also included C16:0-ceramide and sphingomyelin. Elevated circulating concentrations of ceramides[22], including the C16:0 ceramide[23,24], are generally implicated in the pathogenesis of IR. However, our data suggest that C16:0-ceramide may play a more complex role in the regulation of host insulin sensitivity than recognized until now.

Some of the IR- and metabolic-syndrome-associated fasting serum metabolites, such as the essential BCAAs, cannot be synthesized by humans and must therefore originate from ingested food or gut microbial synthesis. Therefore, we investigated the metabolic potential of the gut microbiome in relation to these metabolites using KEGG functional modules (Fig. 1a), that is, a set of manually curated functional and pathway units from the KEGG database, each consisting of a number of KEGG orthologous gene groups[25].

Forty-one of the 567 microbiome functional modules (7%) were significantly associated (Mann–Whitney *U*-test, FDR < 0.1) with one or more of the IR and metabolic syndrome phenotypes (Extended Data Fig. 2, Supplementary Table 7). All 41 functional modules were furthermore associated with the IR- and IS-metabotypes (Figs 1c and 2), with a majority also differing in abundance in the expected direction in type 2 diabetes patients (Supplementary Results). A majority (23/33) of the HOMA-IR associated modules remained significant after adjusting for BMI, while six additional modules were identified upon the BMI adjustment (Extended Data Fig. 2, Supplementary Table 7). The results suggest that, in this sample of lean and obese individuals, the functional microbial shift in IR is largely independent of BMI, which should be further validated in population-based cohorts. Furthermore, we found this functional shift to also partly overlap with functional characteristics of a low richness gut microbiome, as the IR associations of many functional modules (24/33) were attenuated after adjusting for microbial gene richness (Supplementary Results). We observed only a minor separation in the microbiome or metabolomic data corresponding to patient subgroups based on the five clinical components of the metabolic syndrome (Supplementary Results).

The cross-domain associations between the IR and metabolic syndrome phenotypes, the serum metabolome and the gut microbiome described above may suggest functional relationships. Notably, the functional modules of the microbiome that were positively associated with both HOMA-IR and the IR-metabotype contained enzymes for biosynthesis of BCAAs, cofactors, vitamins, lipopolysaccharides and various transport systems, some of which have elevated expression in gut microbiomes transplanted into mice from obese donors[6]. In contrast, the microbiome negatively associated with HOMA-IR and the IR-metabotype contained functional modules important for methanogenesis, pyruvate oxidation and transport systems, including inward transport of BCAAs. Noticeably, the microbiome functional modules that were associated with the two metabotypes showed a large overlap with the metabolic pathways essential for microbiome–host interactions mediated through metabolites, including the BCAAs, as highlighted in a recent *Caenorhabditis elegans* study[26].

Coherently, the majority (60/81) of the microbial species associated with IR or metabolic syndrome (Supplementary Table 8) correlated significantly with the IR- and IS-metabotypes (Supplementary Results,

Extended Data Figs 4 and 5). To identify the species that contributed the most to the associations between the microbiome functional modules and HOMA-IR, we iterated the correlation analysis, leaving out genes from a single species in each iteration (see Fig. 1d and Methods for details). For several functional modules, the correlations with HOMA-IR were driven by only a few species (Supplementary Table 9). Notably, the positive associations between HOMA-IR and the biosynthesis modules of BCAA, tryptophan and lipopolysaccharides were largely driven by *P. copri* followed by *B. vulgatus*, as they were almost abolished when genes from these species were omitted. Although these two species were the main drivers for many of the same functional modules, they were at the same time significantly anti-correlated (Spearman correlation coefficient $= -0.38$, $P = 9.0 \times 10^{-11}$, Extended Data Fig. 6), consistent with the previously reported trade-off between these genera[27]. Still, the abundance of each of these species (when present) correlated positively with HOMA-IR (Extended Data Fig. 7, Supplementary Table 10), as did their combined abundance across all individuals (Spearman correlation coefficient $= 0.20$, $P = 0.001$, Extended Data Fig. 6). The possible importance of *P. copri* is further highlighted by its positive correlation with the BCAA-containing metabolite cluster M10 in the 94 individuals with detectable levels of the species (Spearman correlation coefficient $= 0.22$, $P = 0.034$, Extended Data Fig. 7, Supplementary Table 10). However, the correlations of *P. copri* to the BCAA-containing metabolite cluster M10 and HOMA-IR were abolished in the full cohort of 277 non-diabetic individuals ($P = 0.44$ and $P = 0.67$, respectively), suggesting that in the absence of *P. copri* other species may influence the levels of the M10 metabolite cluster and HOMA-IR. *P. copri*, a hallmark species of the *Prevotella* enterotype[28] (see Supplementary Results for further discussion), has been associated with rheumatoid arthritis[29] but not with IR in humans. *B. vulgatus* is reported to correlate positively with fat mass and other markers of IR in humans[30], whereas a mouse study suggested protection against the development of metabolic disorders, although only in the context of a low-fat diet[6]. Thus, the role of *B. vulgatus* in IR and metabolic disorders seems complex and is probably context-specific. Among other driver species for the microbial functional modules positively correlated with HOMA-IR were *Escherichia coli*, *Sutterella wadsworthensis* and unclassified *Sutterella* and *Prevotella* species (Extended Data Fig. 8, Supplementary Table 9).

In contrast to the functional modules enriched in the gut microbiome of insulin-resistant individuals, which were driven by relatively few microbial species each with high impact, all microbial functional modules associated with increased insulin sensitivity were driven by multiple species, all with minor effect (Supplementary Results).

The microbial functional module analysis presented in Fig. 2 highlighted an increased BCAA biosynthesis potential, but depletion for genes encoding the transport system for bacterial BCAA uptake (referred to as inward BCAA transport) in the gut microbiome of insulin-resistant individuals (Fig. 3a). These findings synergistically contribute to an increased BCAA pool and are of considerable interest

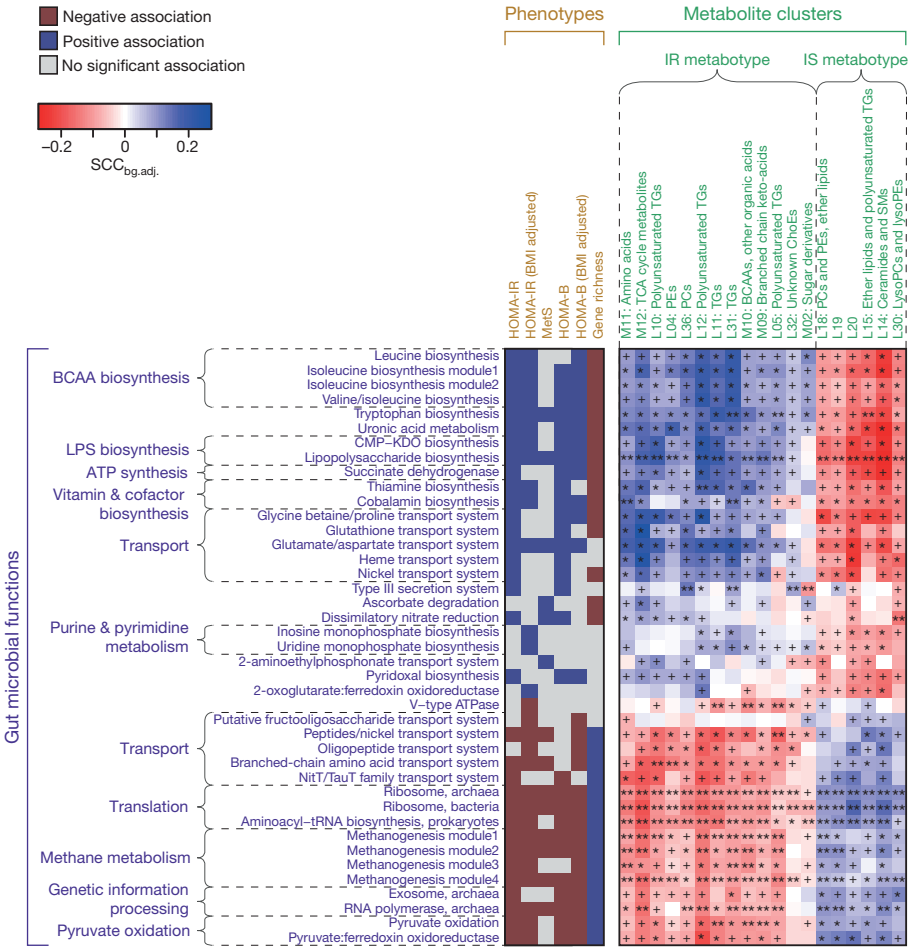


Figure 2 | Association map of the three-tiered analyses integrating the phenome, the gut microbiome and the fasting serum metabolome in 277 non-diabetic individuals with available metagenomic data. The left panel shows significant associations (Mann–Whitney *U*-test FDR < 0.1) between microbial functional modules and the indicated phenotypes; colouring indicates direction of association. The right panel shows associations between the same modules and serum metabolite clusters. Colouring represents the median Spearman correlation coefficient between metabolite clusters and the indicated functional modules, corrected for background distribution ($SCC_{bg.adj.}$, see Methods), where MWU FDRs are denoted: +, FDR < 0.1; *, FDR < 0.01; **, FDR < 0.001.

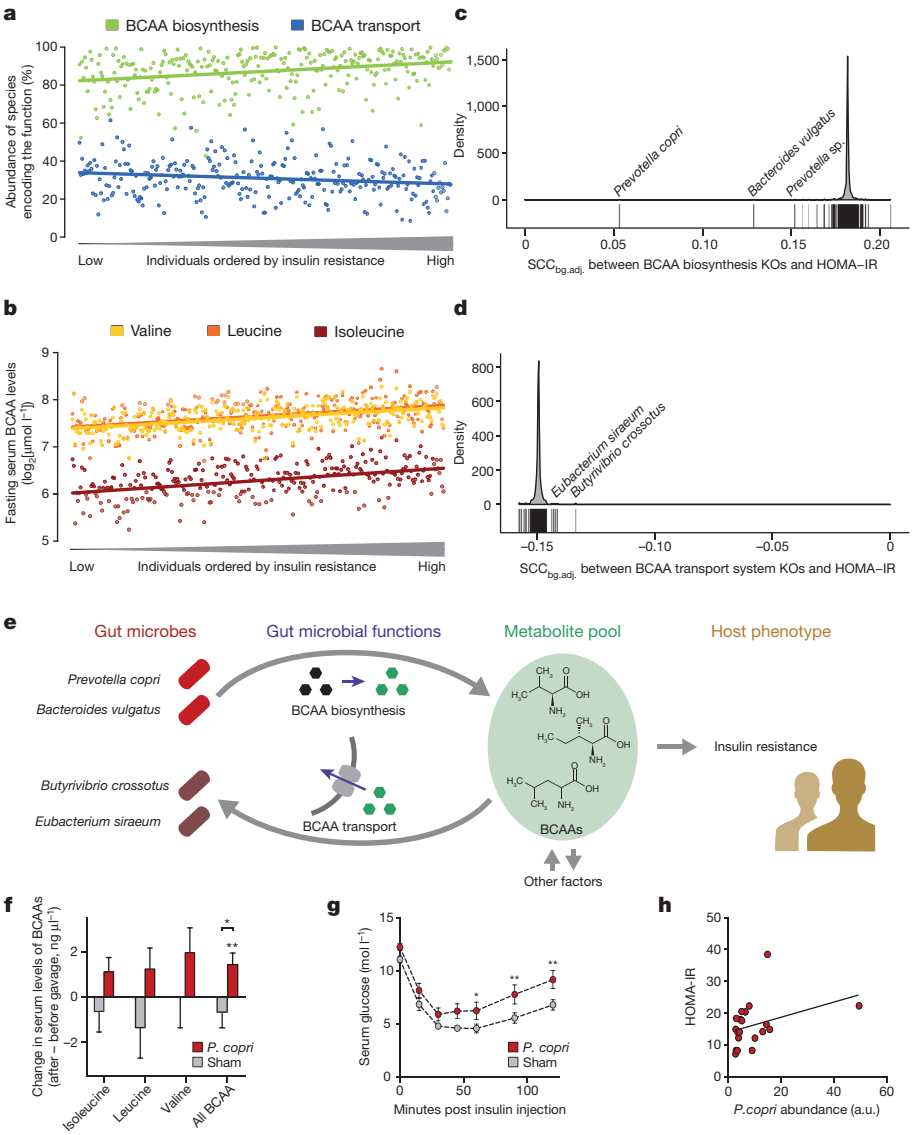


Figure 3 | The ratio between the gut microbiome potential for BCAA biosynthesis and inward transport is linked to fasting serum BCAA levels and IR in humans and can be attributed to a few driver species, including *P. copri* that induces an aggravation of glucose intolerance in mice concomitantly with elevated serum BCAA levels. a, The total abundance of all microbial species containing genes coding for BCAA biosynthesis (Spearman correlation coefficient (SCC) = 0.30, $P = 5.3 \times 10^{-7}$) and inward transport (SCC = −0.17, $P = 4.5 \times 10^{-3}$) potential is shown in green and blue, respectively, for 277 non-diabetic individuals ordered by their HOMA-IR levels. The slopes are significantly different ($P = 4.0 \times 10^{-8}$, see Methods). **b**, Fasting serum BCAA levels in 291 non-diabetic individuals ordered by HOMA-IR levels (SCC: leucine = 0.40, isoleucine = 0.44, valine = 0.49, $P < 4 \times 10^{-12}$). **c**, **d**, The effect of specific microbial species on associations between BCAA biosynthesis and transport, respectively, and HOMA-IR in 277 non-diabetic individuals; illustrated by the change in background-adjusted median SCC ($SCC_{bg.adj}$) between HOMA-IR and the BCAA biosynthesis/transport functional modules when a given species has been excluded from the analysis (see Methods). **e**, Suggested model of the microbiome contribution to serum BCAA levels and IR. Fasting serum BCAA levels are influenced by microbial BCAA biosynthesis and uptake; these levels in turn may influence host insulin sensitivity. **f**, Changes in mouse serum BCAA levels (mean ± s.e.m.) after *P. copri* or sham gavaging for two weeks before challenge. Asterisks indicate significance between *P. copri*-gavaged ($n = 12$) and sham-gavaged control mice ($n = 12$) and significance relative to before challenge (likelihood ratio test, $^*P < 0.05$; $^{**}P < 0.01$). **g**, Insulin tolerance test. The *P. copri*-gavaged mice ($n = 12$) had significantly higher serum glucose levels compared to sham-gavaged controls ($n = 12$, $P = 0.032$, repeated measures two-way ANOVA) after three weeks challenge. Mean ± s.e.m. is depicted. Asterisks indicate significant differences at individual time points (repeated measures two-way ANOVA): $^*P < 0.05$; $^{**}P < 0.01$. **h**, Faecal *P. copri* abundance (arbitrary units, quantitative PCR normalized 16S rDNA) as a function of HOMA-IR two weeks post bacterial challenge (SCC = 0.46, $P = 0.040$).

in the light of the report demonstrating that serum BCAA levels are increased in mice following transplantation of stools from insulin-resistant individuals[6]. The described gut microbiome features are consistent with the elevated serum levels of BCAAs in insulin-resistant individuals (Fig. 3b), suggesting that an increased ratio between gut microbial biosynthesis and uptake of BCAAs contributes to elevated serum BCAA concentrations, which are known to associate with IR[31]. Moreover, the difference between the potential for BCAA biosynthesis and inward BCAA transport of the gut microbiome was even more significantly associated with the serum BCAA concentrations than either of the two functional modules alone (Supplementary Table 11). Importantly, while the IR-associated increase in the microbial potential for BCAA biosynthesis was largely driven by *P. copri* or *B. vulgatus* (see above), the reduced capacity for bacterial BCAA uptake was driven by reduced abundance of multiple species, including *B. crossotus* and *E. siraeum* (Fig. 3c, d).

Consistent correlations with HOMA-IR were observed for almost all of the enzymes and transport system subunits within the BCAA

biosynthesis and transport modules respectively, lending further credence to the importance of these modules (Extended Data Fig. 9, Supplementary Table 12). Notably, of the BCAA biosynthesis enzymes, D-citramalate synthase showed the strongest individual positive correlation with HOMA-IR (Spearman correlation coefficient = 0.24, $P = 4.89 \times 10^{-5}$). D-citramalate synthase constitutes the first step in a less common threonine-independent isoleucine biosynthesis pathway[32]. While most microorganisms synthesize isoleucine from threonine[33], 87 gut microbial species identified in our cohort (including both *P. copri* and *B. vulgatus*, Supplementary Table 13), contained the D-citramalate synthase gene, highlighting the importance of this less common pathway in the human gut microbiome.

Our findings are summarized in the model shown in Fig. 3e, which offers a mechanistic explanation for the increased circulating BCAA concentrations that are known to correlate positively with IR[31] and type 2 diabetes[8]. The difference in functional potential between microbial BCAA biosynthesis and BCAA uptake explained 2.4–3.1% of the variance in serum BCAA levels (Supplementary Table 11). In comparison, less than 1% of variance is explained by currently identified variants in the human genome, although the heritability of fasting serum BCAA levels[34] is estimated to be 0.41–0.49.

P. copri was the strongest driver species for the positive association between microbial BCAA biosynthesis in the gut and IR, suggesting a possible causal relation. This assumption is further supported by several reports associating Prevotellaceae[35–37], and more specifically *P. copri*[29,38], with intestinal inflammation, an important contributor to host IR. To experimentally address the relation between *P. copri* and abnormal glucose metabolism, we compared *P. copri* ($n = 12$) and sham-gavaged ($n = 12$) C57BL/6J male mice on a high-fat diet (Fig. 1e). Notably, two weeks of *P. copri* challenge aggravated glucose intolerance (Extended Data Fig. 10a) and increased serum total BCAA levels (compared with baseline: likelihood ratio test, $P = 0.004$, Fig. 3f). Insulin sensitivity was reduced after three weeks of *P. copri* challenge (Fig. 3g, $P = 0.032$), suggesting a potentially causal role of *P. copri* in mediating the increase in serum BCAA and IR. The relatively modest effect of the *P. copri* inoculation on serum BCAA levels may reflect that *P. copri* correlated more strongly to HOMA-IR than to BCAAs in the human cohort (Supplementary Table 10). Importantly, upon glucose challenge, the plasma insulin excursion rates were indistinguishable between groups (Extended Data Fig. 10b; $P = 0.80$), hence negating beta cell malfunction as a confounding variable. It should be noted that Prevotellaceae (and thus *P. copri*) were a minor component of the mouse microbiome in both groups upon feeding with a high-fat diet, as evidenced by faecal 16S rDNA amplicon analysis (Extended Data Fig. 10c). Despite minor compositional changes in the gut microbiota (adonis, $P = 0.04$), we found no alterations of the relative abundances of any single family component thereof to be significant upon gavage (Extended Data Fig. 10c, FDR > 0.05). In contrast, *P. copri* was significantly increased (Extended Data Fig. 10d, $P = 0.0058$), suggesting that it contributed to the observed pathophysiological changes. This conclusion is further supported by the direct correlation between *P. copri* abundances and HOMA-IR (Fig. 3h, Spearman correlation coefficient = 0.46, $P = 0.040$). Interestingly, an inverse relationship between glucose intolerance and *P. copri* challenge was recently reported[39]. However, in the latter study a completely different dietary regimen, high in fibre and low in fat, was used, suggesting that the effect of *P. copri* is diet-dependent. Furthermore, it should be noted that the feed used in the present study contained soy protein, which is low in BCAAs, as the protein source, contrasting the use of casein, high in BCAAs, in the study by Kovatcheva-Datchary *et al.*[39].

Discussion

By integrating data on host insulin sensitivity and metabolic syndrome, gut microbiome, and fasting serum metabolome, we were able to demonstrate clear metabolome signatures of IR phenotypes among non-diabetic individuals and validate them in patients with type 2 diabetes. The IR-associated metabolome associates with functional components of the IR-linked gut microbiome: notably the increased potential for lipopolysaccharide and BCAA biosynthesis, and reduced potential for BCAA transport into bacterial cells, methanogenesis and pyruvate oxidation.

Studies demonstrating the association of fasting serum BCAA levels with incident type 2 diabetes[8], the normalization of BCAA levels in obese individuals after bariatric surgery[40] and the development of IR in rats after BCAA diet supplementation[31] suggest a potentially causative role of the BCAAs, or their breakdown products[41,42], in metabolic disorders, although the reason as to why they are elevated in the first place is not well established. Potential explanations include reduced BCAA breakdown in adipose tissue[43,44] and liver[45]. Consistent relationships in our data between serum concentrations of BCAAs, bacterial BCAA biosynthesis and inward transport potentials, and the severity of IR phenotypes are reinforced by the intriguing findings that BCAAs co-vary tightly with fasting serum metabolites known to be of microbial origin. This suggests that the gut microbiota is another independent contributing source of elevated serum BCAA levels in common human states of IR. Furthermore, positive correlations between microbial functions—including BCAA biosynthesis—and IR are largely driven by a few species only, notably *P. copri* and *B. vulgatus*, suggesting that they may directly impact host metabolism. We tested this hypothesis in mice fed a high-fat diet, and found that a challenge with *P. copri* led to increased circulating serum levels of BCAAs, insulin resistance and an aggravation of glucose intolerance. We conclude that dysbiosis of the human gut microbiota impacts the serum metabolome and contributes to insulin resistance.

Importantly, while our findings relate to non-diabetic individuals and type 2 diabetes patients with preserved insulin secretion, they cannot yet be generalized to patients with impaired insulin secretion (Supplementary Discussion). Future physiological studies are needed to elucidate how the intestinal BCAAs and other amino acids enter the bloodstream and from which intestinal location they are absorbed. Furthermore, investigations of how dietary changes alone or in combination with microbial or pharmacological interventions may impact the microbiome and, in particular, influence *P. copri* modulation of serum BCAA levels will open novel avenues to counter the pathogenesis of IR and its linked epidemics of common metabolic and cardiovascular disorders.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 18 January; accepted 14 June 2016.
Published online 13 July 2016.

1. Le Chatelier, E. *et al.* Richness of human gut microbiome correlates with metabolic markers. *Nature* **500,** 541–546 (2013).
2. Qin, J. *et al.* A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature* **490,** 55–60 (2012).
3. Karlsson, F. H. *et al.* Gut metagenome in European women with normal, impaired and diabetic glucose control. *Nature* **498,** 99–103 (2013).
4. Forslund, K. *et al.* Disentangling type 2 diabetes and metformin treatment signatures in the human gut microbiota. *Nature* **528,** 262–266 (2015).
5. Vrieze, A. *et al.* Transfer of intestinal microbiota from lean donors increases insulin sensitivity in individuals with metabolic syndrome. *Gastroenterology* **143,** 913–916 (2012).
6. Ridaura, V. K. *et al.* Gut microbiota from twins discordant for obesity modulate metabolism in mice. *Science* **341,** 1241214 (2013).
7. Holmes, E., Li, J. V., Marchesi, J. R. & Nicholson, J. K. Gut microbiota composition and activity in relation to host metabolic phenotype and disease risk. *Cell Metab.* **16,** 559–564 (2012).
8. Wang, T. J. *et al.* Metabolite profiles and the risk of developing diabetes. *Nat. Med.* **17,** 448–453 (2011).
9. Lee, C. C. *et al.* Branched-chain amino acids and insulin metabolism: The Insulin Resistance Atherosclerosis Study (IRAS). *Diabetes Care* **39,** 582–588 (2016).
10. Rhee, E. P. *et al.* Lipid profiling identifies a triacylglycerol signature of insulin resistance and improves diabetes prediction in humans. *J. Clin. Invest.* **121,** 1402–1411 (2011).

11. Kotronen, A. *et al.* Serum saturated fatty acids containing triacylglycerols are better markers of insulin resistance than total serum triacylglycerol concentrations. *Diabetologia* **52,** 684–690 (2009).
12. Floegel, A. *et al.* Identification of serum metabolites associated with risk of type 2 diabetes using a targeted metabolomic approach. *Diabetes* **62,** 639–648 (2013).
13. Roberts, L. D., Koulman, A. & Griffin, J. L. Towards metabolic biomarkers of insulin resistance and type 2 diabetes: progress from the metabolome. *Lancet Diabetes Endocrinol*. **2,** 65–75 (2014).
14. Matthews, D. R. *et al.* Homeostasis model assessment: insulin resistance and beta-cell function from fasting plasma glucose and insulin concentrations in man. *Diabetologia* **28,** 412–419 (1985).
15. 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).
16. Aura, A. M. Microbial metabolism of dietary phenolic compounds in the colon. *Phytochem. Rev.* **7,** 407–429 (2008).
17. Velagapudi, V. R. *et al.* The gut microbiota modulates host energy and lipid metabolism in mice. *J. Lipid Res.* **51,** 1101–1112 (2010).
18. Tso, S. C. *et al.* Structure-based design and mechanisms of allosteric inhibitors for mitochondrial branched-chain α-ketoacid dehydrogenase kinase. *Proc. Natl Acad. Sci. USA* **110,** 9728–9733 (2013).
19. Forouhi, N. G. *et al.* Differences in the prospective association between individual plasma phospholipid saturated fatty acids and incident type 2 diabetes: the EPIC-InterAct case-cohort study. *Lancet Diabetes Endocrinol*. **2,** 810–818 (2014).
20. Roberts, L. D., Virtue, S., Vidal-Puig, A., Nicholls, A. W. & Griffin, J. L. Metabolic phenotyping of a model of adipocyte differentiation. *Physiol. Genomics* **39,** 109–119 (2009).
21. Collins, J. M. *et al.* De novo lipogenesis in the differentiating human adipocyte can provide all fatty acids necessary for maturation. *J. Lipid Res.* **52,** 1683–1692 (2011).
22. Summers, S. A. Ceramides in insulin resistance and lipotoxicity. *Prog. Lipid Res.* **45,** 42–72 (2006).
23. Raichur, S. *et al.* CerS2 haploinsufficiency inhibits β-oxidation and confers susceptibility to diet-induced steatohepatitis and insulin resistance. *Cell Metab.* **20,** 687–695 (2014).
24. Turpin, S. M. *et al.* Obesity-induced CerS6-dependent C16:0 ceramide production promotes weight gain and glucose intolerance. *Cell Metab.* **20,** 678–686 (2014).
25. Kanehisa, M. *et al.* KEGG for linking genomes to life and the environment. *Nucleic Acids Res.* **36,** D480–D484 (2008).
26. Watson, E. *et al.* Interspecies systems biology uncovers metabolites affecting *C. elegans* gene expression and life history traits. *Cell* **156,** 759–770 (2014).
27. Yatsunenko, T. *et al.* Human gut microbiome viewed across age and geography. *Nature* **486,** 222–227 (2012).
28. Arumugam, M. *et al.* Enterotypes of the human gut microbiome. *Nature* **473,** 174–180 (2011).
29. Scher, J. U. *et al.* Expansion of intestinal *Prevotella copri* correlates with enhanced susceptibility to arthritis. *eLife* **2,** e01202 (2013).
30. Dewulf, E. M. *et al.* Insight into the prebiotic concept: lessons from an exploratory, double blind intervention study with inulin-type fructans in obese women. *Gut* **62,** 1112–1121 (2013).
31. Newgard, C. B. *et al.* A branched-chain amino acid-related metabolic signature that differentiates obese and lean humans and contributes to insulin resistance. *Cell Metab.* **9,** 311–326 (2009).
32. Charon, N. W., Johnson, R. C. & Peterson, D. Amino acid biosynthesis in the spirochete Leptospira: evidence for a novel pathway of isoleucine biosynthesis. *J. Bacteriol.* **117,** 203–211 (1974).
33. Umbarger, H. E. Amino acid biosynthesis and its regulation. *Annu. Rev. Biochem.* **47,** 532–606 (1978).
34. Shin, S.-Y. *et al.* An atlas of genetic influences on human blood metabolites. *Nat. Genet.* **46,** 543–550 (2014).
35. Elinav, E. *et al.* NLRP6 inflammasome regulates colonic microbial ecology and risk for colitis. *Cell* **145,** 745–757 (2011).
36. Henao-Mejia, J. *et al.* Inflammasome-mediated dysbiosis regulates progression of NAFLD and obesity. *Nature* **482,** 179–185 (2012).
37. Heimesaat, M. M. *et al.* Gram-negative bacteria aggravate murine small intestinal Th1-type immunopathology following oral infection with Toxoplasma gondii. *J. Immunol.* **177,** 8785–8795 (2006).
38. Dillon, S. M. *et al.* Gut dendritic cell activation links an altered colonic microbiome to mucosal and systemic T-cell activation in untreated HIV-1 infection. *Mucosal Immunol.* **9,** 24–37 (2016).
39. Kovatcheva-Datchary, P. *et al.* Dietary fiber-induced improvement in glucose metabolism is associated with increased abundance of *Prevotella*. *Cell Metab.* **22,** 971–982 (2015).
40. Laferrère, B. *et al.* Differential metabolic impact of gastric bypass surgery versus dietary intervention in obese diabetic subjects despite identical weight loss. *Sci. Transl. Med.* **3,** 80re2 (2011).
41. Lynch, C. J. & Adams, S. H. Branched-chain amino acids in metabolic signalling and insulin resistance. *Nat. Rev. Endocrinol.* **10,** 723–736 (2014).
42. Jang, C. *et al.* A branched-chain amino acid metabolite drives vascular fatty acid transport and causes insulin resistance. *Nat. Med.* **22,** 421–426 (2016).
43. Herman, M. A., She, P., Peroni, O. D., Lynch, C. J. & Kahn, B. B. Adipose tissue branched chain amino acid (BCAA) metabolism modulates circulating BCAA levels. *J. Biol. Chem.* **285,** 11348–11356 (2010).
44. Pietiläinen, K. H. *et al.* Global transcript profiles of fat in monozygotic twins discordant for BMI: pathways behind acquired obesity. *PLoS Med.* **5,** e51 (2008).
45. Shin, A. C. *et al.* Brain insulin lowers circulating BCAA levels by inducing hepatic BCAA catabolism. *Cell Metab.* **20,** 898–909 (2014).

Supplementary Information is available in the online version of the paper.

Acknowledgements The authors wish to thank S. Castillo, M. Sysi-Aho, A. Ruskeepää, U. Lahtinen, A. Forman, T. Lorentzen, B. Andreasen, G. J. Klavsen, M. J. Nielsen, B. Pedersen, M. T. F. Damgaard and L. B. Rosholm for technical assistance, D. R. Mende and J. R. Kultima for their help in data processing and tool provision, C. Ekstøm and S. Ditlevsen for statistical and mathematical assistance, respectively, and T. F. Toldsted and G. Lademann for management assistance. C. B. Newgard and A. Vaag are thanked for critical comments on our manuscript. The present study is initiated and funded by the European Community's Seventh Framework Program (FP7/2007-2013): MetaHIT, grant agreement HEALTH-F4-2007-201052. Additional funding came from The Lundbeck Foundation Centre for Applied Medical Genomics in Personalized Disease Prediction, Prevention and Care (LuCamp, http://www.lucamp.org), Metagenopolis grant ANR-11-DPBS-0001 and FP7 METACARDIS HEALTH-F4-2012-305312. J.R., S.V.-S. and G.F. are funded by the Rega institute for Medical Research, KU Leuven, the Agency for Innovation by Science and Technology (IWT), Marie Curie Actions FP7 People COFUND - Proposal 267139 and the Fund for Scientific Research Flanders (FWO). M.O. is also supported by Academy of Finland (Centre of Excellence in Molecular Systems Immunology and Physiology Research, Decision No. 250114) and EU FP7 Project TORNADO (project 222720). F.H. has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 600375. The Center for Biological Sequence Analysis and the Novo Nordisk Foundation Center for Basic Metabolic Research have in addition received support from the Innovative Medicines Initiative Joint Undertaking under grant agreement no. 115317 (DIRECT), resources of which are composed of financial contribution from the European Union's Seventh Framework Programme (FP7/2007-2013) and EFPIA companies' in kind contribution. The Novo Nordisk Foundation Center for Protein Research received funding from the Novo Nordisk Foundation (grant agreement NNF14CC0001). The Novo Nordisk Foundation Center for Basic Metabolic Research is an independent Research Center at the University of Copenhagen partially funded by an unrestricted donation from the Novo Nordisk Foundation (http://www.metabol.ku.dk).

Author Contributions O.P., S.D.E. and P.B. devised the study. O.P., S.D.E., S.Bru. and H.B.N. designed the study protocol and supervised all parts of the project. H.B.N. and S.Bru. led the data integration, the bioinformatics analyses and did the primary interpretation of analytical outcomes in close collaboration with H.K.P. and V.G. H.K.P., V.G., B.A.H.J., T.Hy., E.P., D.P., S.S., F.H., K.F., J.B.H. and S.V.-S. performed data analyses. T.N., T.Ha. and O.P. composed the clinical protocol, carried out phenotyping of study participants including collection of biological samples and physiological data generation and interpretation. F.L. performed DNA extraction and J.D. supervised DNA extraction. J.W. supervised DNA sequencing and gene profiling. M.O., T.Hy., I.M., K.T. and P.P. performed profiling of serum metabolomics and serum lipidomics. B.A.H.J., K.K., J.B.H. and S.Bri. performed mouse experiments. H.B.N., H.K.P. and V.G. drafted the first versions of the paper with critical and substantial contributions from O.P., S.Bru., T.N., J.R., K.F., F.H., M.O., I.I.H., D.P., G.F., P.B. and S.D.E. All authors approved the final version. MetaHIT consortium members provided support and constructive criticism throughout MetaHIT research operations.

Author Information Raw nucleotide data can be found for all samples used in the study in the European Nucelotide Archive (accession numbers: ERP003612, ERP004605, MetaHIT samples; ERP014713, 16S rDNA from mouse experiment). The metabolomics data has been deposited in the MetaboLights database (http://www.ebi.ac.uk/metabolights/) under accession number: MTBLS351. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.Bru. (soren.brunak@cpr.ku.dk), M.O. (matej.oresic@gmail.com), S.D.E. (dusko.ehrlich@jouy.inra.fr) or O.P. (oluf@sund.ku.dk).

Reviewer Information *Nature* thanks J. Garrett, L. Groop, C. Lozupone, G. Siuzdak and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Human study population. The study population comprised 291 normoglycaemic and middle-aged Danish MetaHIT individuals as previously reported[1], as well as 75 Danish type 2 diabetes patients[4]. All analyses involving gut microbiome data were restricted to 75 type 2 diabetes patients and 277 out of the 291 of the non-diabetic individuals whose data passed all quality control criteria (see details below), whereas all non-microbiome analyses involved the full set of 291 non-diabetic individuals. No statistical methods were used to predetermine sample size.

Of the 75 type 2 diabetes patients, ten patients (13%) received no hyperglycaemic medications, 58 patients (77%) received the biguanide compound, metformin; of these 28 patients (37%) received metformin as the only anti-hyperglycaemic medication, 10 patients (13%) received sulfonylurea alone or in combination with metformin, 14 patients (19%) received a combination of oral anti-diabetic drugs and insulin treatment and 4 patients (5%) were on insulin treatment only. Eleven patients (15%) were treated with dipeptidyl peptidase-4 (DPP4) inhibitors or glucagon-like peptide-1 (GLP-1), all of them in combination with metformin. Patients were reported as receiving anti-hypertensive treatment if at least one of the following drugs were reported: spironolactone, thiazides, loop diuretics, beta-blockers, calcium channel blockers, moxonidin or drugs affecting the renin-angiotensin system (in total $n = 55$; 73%). Patients receiving statins were reported as receiving lipid-lowering medication (in total $n = 56$; 75%). We have previously shown that metformin has impact on the configuration and functional potential of human gut microbiota whereas no significant differences were observed for other anti-diabetic medications, diuretics, or lipid lowering- and blood pressure lowering drugs[4].

In short, all study participants (Caucasian Danes) were recruited from the population-based Inter99 study[46] or from the outpatient clinic at Steno Diabetes Center. Study volunteers were invited for two visits with approximately 14 days apart. At the first visit the participants were examined in the morning after an overnight fast of at least 10 h and without prior morning physical activity. At the second visit a DEXA scan was performed to evaluate body fat percentage. All samples (from both non-diabetic and diabetic individuals) were generated within the Danish part of the MetaHIT consortium (http://www.metahit.eu), and the infrastructure of this project was constructed specifically to make all samples comparable by ensuring similar handling and by minimizing the risk of batch effects. All of these samples were collected individually as participant recruitment proceeded, by the same staff and using the same protocol for both stool and blood samples. Serum for metabolomics and stool for metagenomics were stored at −80 °C until the planned analyses took place.

The study was approved by the Ethical Committees of the Capital Region of Denmark (HC-2008-017 and H-15000306) and was in accordance with the principles of the Declaration of Helsinki. All individuals gave written informed consent before participating in the study.

Phenotyping. Clinical phenotyping and fasting biochemistry were performed and analysed as reported[1]. HOMA-IR was calculated as: (fasting plasma glucose (mmol/l) × fasting serum insulin (mU/l))/22.5[14] and HOMA-B (a measure of pancreatic beta-cell function) calculated as (20 × fasting plasma insulin)/(fasting plasma glucose − 3.5)[14]. Due to the confounding effects of anti-diabetic treatment, HOMA-IR and HOMA-B were only assessed in non-diabetic individuals. BMI was calculated as weight (kg) divided by height (m^2) and obesity defined as BMI >30 kg/m^2. Metabolic syndrome was defined according to the International Diabetes Federation[15]. Non-diabetic individuals classified as having the metabolic syndrome in accordance with the IDF ($n = 163$) were characterized by central obesity defined by gender-specific increased waist circumference as well of two of four additional risk factors for cardiovascular disease: raised levels of fasting plasma glucose or triglycerides, or reduced fasting plasma level of HDL-cholesterol, or raised blood pressure. In order to account for differences in age, sex and BMI, we matched the 75 Danish type 2 diabetes patients with 75 age-, sex- and BMI-matched individuals chosen from the non-diabetic study samples of 277 Danish individuals (Supplementary Table 1).

Generation of serum metabolome data set. The metabolomics analyses were performed on all human samples together, in randomized order, both related to the sample preparation and instrumental analysis. Similarly, all 48 mouse samples were analysed for BCAAs together in randomized order. Control serum samples consisting of pooled human serum were included in equal intervals as part of the sample run for both human and mouse samples ($n = 30$ for humans and $n = 4$ for mice), as well as a set of pure standard samples (pure standards in a solvent for evaluation of instrument performance and robustness), extracted standard samples (standards going through the same sample preparation to monitor the performance on the whole analytical procedure), blank samples (only solvent to observe possible background peaks) and calibration samples.

Metabolomics analyses were also conducted on previously reported Danish type 1 diabetes patients[4] ($n = 30$ for polar metabolites, $n = 31$ for molecular lipids). These data were only used for the co-abundance clustering of metabolites (see below) to increase sample size for finding inter-variable conditions. Type 1 diabetes patients are distinctive from type 2 diabetes patients as they represent both a very different pathology and lifelong medical/lifestyle treatment and were consequently not used for any other analyses.

Human and mouse samples were analysed at VTT Technical Research Centre of Finland (Espoo, Finland) and Steno Diabetes Center (Gentofte, Denmark), respectively, using the same type of instrument with identical analytical conditions.

Analysis of fasting serum polar metabolites by GC × GC-TOFMS. Serum polar metabolites were analysed in 396 human and 48 mouse serum samples (24 pre- and 24 post-gavage samples) using comprehensive two-dimensional gas chromatography combined with time-of-flight mass spectrometry (GC × GC-TOFMS, a LECO Pegasus 4D equipped with a cryogenic modulator from LECO Corp.) with a method described previously[47]. Specifically, 400 μl methanol and 10 μl internal standard mixture (C17:0 (186.5 mg/l), deuterated valine (37 mg/l) and succinic acid-d_4 (63 mg/l)) were added to 30 μl of serum samples. The samples were vortex mixed and centrifuged for 5 min at 10,000 r.p.m. and half of the supernatant was evaporated to dryness. This was followed by two-step derivatization using methoximation and silylation by first adding 25 μl methoxamine (45 °C, 60 min) and then 25 μl *N*-trimethylsilyl-*N*-methyl trifluoroacetamide (45 °C, 60 min). Finally, a retention index standard mixture (*n*-alkanes, 25 μl, concentration = 8 mg/l) and an injection standard (4,4′-dibromooctafluorobiphenyl, 50 μl, concentration = 10 mg/l), both in hexane, were added to the mixture. The calibration consisted of six points for each quantified metabolite.

The columns were as follows: a methyl-deactivated retention gap (1.5 m × 0.53 mm i.d.) was connected to 10 m × 0.18 mm Rxi-5MS (phase thickness 0.18 μm) and to 1.5 m × 0.1 mm BPX-50 (phase thickness 0.1 μm). Helium was used as the carrier gas at a constant pressure mode (40 psig). A 4-s separation time was used in the second dimension. The temperature program was as follows for the first-dimension column: 50 °C (2 min), at 7 °C/min to 240 °C and at 25 °C/min to 300 °C (3 min). The second-dimension column temperature was 20 °C higher than the corresponding first-dimension column throughout the program.

The analytical method used allows for combined targeted and untargeted analysis, where a selected subset of metabolites can be quantified[47]. In the present study, quantitation of 27 target metabolites (stearic acid, oleic acid, linoleic acid, palmitic acid, citric acid, glutamic acid, 3,4-dihydroxybutanoic acid, 3-hydroxybutyric acid, threonine, 2,4-dihydroxybutanoic acid, phenylalanine, tyrosine, 2-hydroxybutyric acid, serine, glyceric acid, methionine, glycine, leucine, isoleucine, valine, ornithine, proline, cholesterol, arachidonic acid, alanine and aspartic acid) was done by external calibration curves for each individual metabolite.

ChromaTOF vendor software (LECO) was used for within-sample data processing, and the Guineu software[47] was used for alignment, normalization and peak matching across samples. The normalization was performed by correction for internal standards and specific target metabolites were additionally quantified using external calibration curves. Other mass spectra from the GC × GC-TOFMS analysis were searched against the NIST Mass Spectral Library and Golm Metabolome Database[48], also using retention index data in the identification. Artefact peaks due to chemical background and compounds outside the linear range of the method were removed from the data set. Control serum samples ($n = 30$ for human, $n = 4$ for mouse analyses) were analysed together with the samples. The relative standard deviation (RSD) for internal standards, spiked into the samples, was on average of 12.8% for the human serum sample analyses. The RSD% of the quantified metabolites in the control serum samples ($n = 30$) was on average 18.5% for the analysis of human serum samples. For quantitative analysis of BCAAs in mice serum samples, the RSD% was 13.3% for isoleucine, 11.2% for leucine and 24.2% for valine in control serum samples ($n = 4$), and for internal standards in mice serum the RSD was on average 16.6%. Neither sample preparation nor analysis order showed any significant effect on the results.

All serum metabolite peaks that were present (non-zero value) in more than 50% of samples were included in the data analyses, including the unidentified ones. We reasoned that inclusion of complete data as obtained from the platform best represent the global metabolome as covered by the platform. The unidentified peaks were annotated with their structural class from the Golm Metabolome Database[48] using functional group prediction based on the fragmentation patterns[47]. In total 325 (94 known and 231 unknown) serum polar metabolites were measured in the human study.

Analysis of serum molecular lipids using UHPLC-QTOFMS. Serum molecular lipids were analysed in 397 human serum samples using ultra-high-performance liquid chromatography coupled with time-of-flight mass spectrometry (UHPLC-QTOFMS Q-Tof Premier mass spectrometer, Waters, Inc.) using electrospray ionization in positive ionization mode with a methodology described earlier[49]. Specifically, a standard mixture 1 (20 μl) containing LPC(17:0), PC(17:0/17:0), PE(17:0/17:0) and Cer(d18:1/17:0) (Avanti Polar Lipids, Inc.)

and TG(17:0/17:0/17:0) (Larodan Fine Chemicals AB), all at concentration of 10 mg/l, was added to 10 μl of serum samples and the samples were extracted with chloroform and methanol (2:1; 100 μl). The lower phase (60 μl) was collected after centrifugation and 20 μl of internal standard mixture 2 (LPC(16:1D_3), concentration: 10 mg/l, PC(16:1/16:1-D_6), concentration: 15 mg/l, and TG(16:0/16:0/16:0-^{13}C3), concentration: 15 mg/l, was added. The extracts were analysed on an Acquity UPLCTM BEH C18 2.1 × 100 mm column packed with 1.7 μm particles. The solvent system included 1% 1 M NH_4Ac, 0.1% HCOOH and acetonitrile/isopropanol (1:1, 1% 1M NH_4Ac, 0.1% HCOOH) in gradient elution mode with a flow rate of 0.4 ml/min. The serum lipid profiling was carried out at a mass range of m/z 300–1,200 with scan duration of 0.2 s.

The data processing using MZmine 2 (ref. 50) included alignment of peaks, peak integration, isotopic grouping, normalization, and peak identification. Serum lipids were identified using an internal spectral library or with tandem MS and data were normalized using the internal standards representatives of each class of serum lipid present in the samples as previously described[49]. Sphingomyelins were normalized with the PC standard. All lipid peaks were included in the data analyses, including the unidentified ones. Control serum samples ($n = 30$), extracted standard samples ($n = 22$) as well as blanks and pure standard samples were analysed together with the study samples. The standard deviation (RSD) for internal standards, spiked into the samples, was 8–13% in serum samples ($n = 397$) and 9.1% in the extracted standards ($n = 22$). The RSD% of the identified lipids in the control serum samples was on average 12.8%. In total 876 molecular serum lipids, 289 known and 587 unknown, were measured. Neither sample preparation nor analysis order showed any significant effect on the results.

Clustering of co-abundant serum metabolites. Clusters of co-abundant serum metabolites were identified using the R package WGCNA[51] (Fig. 1a). Serum polar metabolites and serum molecular lipids (see above) were analysed separately but are collectively referred to as serum metabolites. Signed, weighted metabolite co-abundance correlation (biweight midcorrelations (a median-based correlation measure that is more robust to outliers than Pearson correlation[52,53]) after $\log_2$ transformation) networks were calculated across all examined individuals. A scale-free topology criterion was used to choose the soft threshold $\beta = 13$ for the polar metabolites and $\beta = 14$ for molecular lipid correlations. Clusters were identified with the dynamic hybrid tree-cutting algorithm[54], using deepSplit of 4 and a minimum cluster size of 3 and 5 for polar metabolite and molecular lipid clusters, respectively. Serum metabolites that did not fit the clustering criteria were combined in a group named 'remaining'. The profile of each serum metabolite cluster was summarized by the cluster eigenvector (that is, the first principal component of the metabolite abundances across individuals). Similar clusters were subsequently merged if the biweight mid-correlation between the cluster's eigenvectors exceeded 0.8 and 0.75 for serum polar metabolite and serum molecular lipid clusters, respectively. The serum polar metabolite and serum molecular lipid clusters (labelled M00–M35 and L00–L39, respectively) are collectively termed metabolite clusters.

Construction of a non-redundant metagenomics reference gene catalogue. The reference gene catalogue used in the present study was constructed from Illumina shotgun sequencing data from 620 publically available faecal samples[4,55]. The data was processed using the MOCAT software package (version 1.1)[56]. Reads were trimmed (option read_trim_filter) using a quality and length cut-off of 20 and 30 bp, respectively. Trimmed reads were subsequently screened against a custom database of Illumina adapters (option screen_fastafile) and the human genome version 19 using a 90% identity cut-off (option screen). The resulting high-quality reads were assembled (option assembly) and assemblies revised (option assembly_revision). Genes were predicted on scaftigs with a minimum length of 500 bp (option gene_prediction).

Predicted protein-coding genes with a minimum length of 100 bp were clustered at 95% sequence identity using CD-HIT (version 4.6.1)[57] with parameters set to: -c 0.95, -G 0 -aS 0.9, -g 1, -r 1). The representative genes of the resulting clusters were 'padded' (that is, extended up to 100 bp at each end of the sequence using the sequence information available from the assembled scaftigs), resulting in the final reference gene catalogue used in this study. The reference gene catalogue was functionally annotated using SmashCommunity (version 1.6)[58] after aligning the amino acid sequence of each gene to the KEGG database (version 62)[25].

Profiling of metagenomics samples. We obtained metagenomics data for the 291 non-diabetic individuals and 75 type 2 diabetes patients as previously described[1,4]. The DNA extraction protocol was standardized for all samples and all sequencing was done using Illumina platforms, using paired-end reads (Supplementary Table 14). Of the 291 non-diabetic samples, 14 that were sequenced with the Illumina Solexa technology had average read length below 42 nt (leading to less certain mappings) and were thus excluded from further analyses to avoid unnecessary biases, resulting in a final set of 277 non-diabetic individuals with gut microbiome data. All 75 type 2 diabetes samples passed all quality control criteria. Sequencing statistics for each sample have previously been described[4] and are repeated in Supplementary Table 14. Raw insert (sequenced fragments of DNA represented by single or paired-end reads) count profiles were generated using MOCAT[56] by mapping high-quality reads from each metagenome to the reference gene catalogue (option screen) using an alignment length and identity cut-off of 45% and 95%, respectively. The abundance profile for each catalogue gene was calculated as the sum of uniquely mapped sequence reads (97% identity) using 7M sequence reads per sample (downsized). KEGG Orthology (KO) gene group abundance profiles were calculated by summing the abundance of genes annotated to the respective KO gene group. KOs were mapped to microbiome functional KEGG modules based on annotations downloaded on 14 January 2014 from the KEGG BRITE database. Gene richness was calculated as the number of genes with abundance higher than zero in the given sample (downsized). For negative binominal statistics all uniquely mapped sequence reads were used to estimate abundances.

Metagenomics species construction. The gene catalogue was clustered by co-abundance as described by Nielsen *et al.*[59], which defined 10,754 co-abundance gene groups (CAGs) with high correlations (Pearson correlation coefficient >0.9). The 925 largest of these, with more than 700 genes, were considered as metagenomics species (MGS) and referred to as species throughout the article. The abundance profiles of the CAGs and MGS were determined as the medium gene abundance (7M reads per sample) throughout the samples. Furthermore, the CAGs and MGSs were taxonomically annotated, by summing up the taxonomical annotation of their genes as described by Nielsen *et al.*[59]. Each gene was annotated by sequence similarity to known reference genomes (blastN, E-value < 0.001 against imomi4 genome databank, August 2012 release).

Mouse intervention study with *Prevotella copri*. *P. copri* (CB7, DSMZ) was cultivated under anaerobic conditions in PYG medium (SSI Diagnostica), harvested in log phase, aliquoted and stored in 10% glycerol at −80 °C until use. Mice (24 male WT C57BL/6J, Taconic, 10 weeks of age) were housed in cages of four and kept at 22 °C in a specific-pathogen-free facility under a 12 h light-dark cycle, fed *ad libitum* and had free access to water. Number of mice (sample size) was selected based on previous experience in measurement of dysglycaemic phenotype in a mouse model, but was not arrived at through formal analysis. After 2 weeks of acclimatization on a standard chow diet (Altromin 1310, Altromin), mice were transferred to a soy protein-based 60% high-fat diet low in BCAAs (12492D, Ssniff) in order to prime for later development of glucose intolerance and insulin resistance and sham-gavaged (1% glycerol in PBS) twice a week. After three weeks on high-fat diet, mice were divided into two equal groups of 12 mice each (three cages per group, minimizing the influence of potential cage effects). The groups were not randomized but stratified by magnetic resonance (MR)-based fat mass and weight. For the following three weeks, mice were gavaged twice weekly with 100 μl of either *P. copri* (5×10^8 CFU/mouse, 1% glycerol in PBS) or sham while remaining on the high-fat diet. Blood for BCAA measurements was drawn 3 days before diet change and 11 days post diet change in EDTA-coated tubes kept on ice and centrifuged at 4 °C for 10 min at 1,000*g* before storage at −80 °C until metabolomics analysis.

An oral glucose tolerance test (OGTT) was performed after two weeks of *P. copri* challenge. The mice were fasted for 5 h with free access to water and gavaged with 3 g glucose/kg lean mass. Blood glucose was measured in tail vein blood before and 15, 30, 45, 60, 90, and 120 min post the glucose bolus, using standard Contour Next Test Strips that met 2013 accuracy criteria requirements from International Organization for standardization of diagnostic tests (Bayer Contour).

Insulin secretion capacity was evaluated by measuring plasma insulin before (T0) and 15 min post glucose bolus (T15) using an electrochemiluminescence assay (Meso Scale Diagnostics) following the manufacturer's instructions.

An insulin tolerance test was performed after 3 weeks of bacterial challenge. The mice were acclimatized in clean cages without bedding for 2 h and injected intra peritoneally with 0.75 U insulin/kg lean mass in alternating order. Insulin was diluted in succinulated gelatin (Gelofusine B. Braun Melsungen AG), allowing increased efficacy of insulin administration, thus minimizing intra experimental standard variations. Blood glucose was measured immediately before and 15, 30, 45, 60, 90 and 120 min post insulin bolus, using standard Contour Next Test Strips that met accuracy criteria requirements from International Organization for standardization of diagnostic tests (Bayer Contour).

In all mouse experiments performed, mice were handled in alternating order. During ITT and OGTT the operator was semi-blinded as each mouse was assigned a random number corresponding to the time (minute) for insulin/glucose bolus administration.

For quantitative PCR and 16 rDNA amplicon sequencing, bacterial DNA from faecal samples were extracted using a NucleoSpin soil kit (Macherey-Nagel) according to manufacturer's instructions. DNA yield and integrity

were assessed using a Nanodrop and agarose gel electrophoresis, respectively. Quantitative PCR analyses were performed using the SYBR Green qPCR Master mix (Thermo Scientific) and the Stratagene Mx3000P qPCR System. The relative abundance of *P. copri* was determined by normalization to the reference 16S rDNA gene using the comparative C_T method[75] for relative gene expression. Primer sequences were previously published[29] and are as follows: Universal 16S forward: 5′-ACTCCTACGGGAGGCAGCAGT-3′, Universal 16S reverse: 5′-ATTACCGCGGCTGCTGGC-3′, *P. copri* forward: 5′-CCGGACTCCTGCCCCTGCAA-3′, *P. copri* reverse: 5′-GTTGCGCCA GGCACTGCGAT-3′.

The PCR-based library formation for the 16S rDNA amplicon sequencing was performed using 10 ng DNA, 0.2 μM of each barcoded forward and reverse primer, 0.2 mM dNTPs and 0.5 units Phusion high fidelity DNA polymerase (Thermo Scientific) in a total volume of 25 μl. To target the 16S rRNA gene's variable region 4 (V4) a forward primer 515F (5′-AATGATACGGCGACCACCGAGATCTACAC NNNNNNNN TATGGTAATTGTGTGCCAGCMGCCGCGGTAA-3′; 'N' indicates the nucleotides of the barcode sequence) and a reverse primer 806R (5′-CAAGCAGAAGACGGCATACGAGAT NNNNNNNNNNNNAGTCAGT CAG CC GGACTACHVGGGTWTCTAAT-3′) were used, both with Illumina adaptor sequences in the 5′ end[60,61]. Cycling condition was as follows: 98 °C for 30 s followed by 35 cycles of 98 °C for 5 s, 56 °C for 20 s and 70 °C for 20 s. PCR products were purified using Agencourt AMPure XP (Beckman Coulter) beads. Subsequently, samples were pooled in equal concentrations and sequenced using an Illumina MiSeq with V2 PE500 cartridge (500 cycles).

For 16S rDNA data analysis, generated sequences were analysed using QIIME v.1.9.1 with default settings. Chimaera checking was performed using UCHIME[62] and *de novo* OTU-picking was performed using UCLUST[63] with 97% sequence similarity. Representative sequences were assigned taxonomy against the Greengenes database v.13_8[64] using the RDP-classifier[65]. Subsequent analyses were performed in R v.3.2.3 using the metagenomeSeq[66], PhyloSeq[67], vegan[68] and ggplot2[69] packages. Data was filtered for low-abundance OTUs by removal of OTUs present in fewer than 3 of the 48 samples and with a relative abundance across all samples 0.005%. Analyses in R were performed with an average of 51,912 ± 15,173 (s.d.) sequences per sample after filtering. For bacteria, differential abundance analysis we normalized read counts with metagenomeSeq[66] using cumulative-sum scaling. Statistical analysis was performed on data filtered based on effective sample sizes where taxa were not included if they had fewer than X effective number of positive samples, where X is the median of estimated effective samples per feature calculated using metagenomeSeq.

All animal experiments were conducted in accordance with national Danish guidelines (Amendment #1306 of 23 November 2007) as approved by the Danish Animal Inspectorate, Ministry of Justice, permission #2014-15-2934-01027. Mice were kept under specific-pathogen-free conditions and experimental protocols were validated by in-house standard operation procedures.

Statistical analysis. Apart from the mouse experiment, in which Graphpad Prism version 6, Treestar, was used, all statistical analyses were performed in the statistical computing language R[70]. As the non-diabetic study sample comprised both non-obese and obese individuals, a number of the physiological variables were not normally distributed (Extended Data Fig. 1). Consequently, non-parametric tests were applied in the statistical analysis. For all analyses involving microbiome data, we only included microbial species and KOs present in at least three non-diabetic individuals to avoid artefactual results. All comparisons with a Benjamini–Hochberg false discovery rate (FDR) <0.1 were considered significantly different throughout the analysis. To focus the analysis on phenotype-relevant variables, we employed an initial filtering of significant phenotype association of the serum metabolite clusters, microbial species and microbiome-derived KEGG modules (Fig. 1b). Serum metabolite clusters and species associations with HOMA-IR were conducted with a Spearman rank correlation test. Partial Spearman rank correlation tests[71] were used to adjust the correlations for BMI ($\text{HOMA-IR}_{\text{BMIadj}}$), or gene richness ($\text{HOMA-IR}_{\text{GeneRichness.adj}}$) where applicable, using the ppcor R package[72]. Serum metabolite clusters and species abundance differences between individuals with and without metabolic syndrome were tested with a Mann–Whitney U-test. We focused the downstream analysis on the serum metabolite clusters associated with all three phenotypes (HOMA-IR, $\text{HOMA-IR}_{\text{BMIadj}}$ and metabolic syndrome), but other metabolite clusters were significantly associated with one or more of the three phenotypes that were not retained for further study. Owing to the low number of microbiome functional modules and species associated with all three phenotypes, we included all functional modules and species associated with HOMA-IR, $\text{HOMA-IR}_{\text{BMIadj}}$ or metabolic syndrome in further analysis.

All associations for microbial functional modules were identified using a Mann–Whitney U-test, where the ranks of KOs within a given KEGG functional module were compared with the ranks of all other KOs. For associations of microbial functional modules with HOMA-IR, gene richness and serum metabolite clusters the ranks were based on Spearman correlation coefficients, whereas partial Spearman correlation coefficients were used to rank KOs for association with $\text{HOMA-IR}_{\text{BMIadj}}$, and $\text{HOMA-IR}_{\text{GeneRichness.adj}}$, as it allows adjusting for BMI and gene richness, respectively. To quantify the shift in (partial) Spearman correlation for a given KEGG module compared to the background distribution, we calculated the background adjusted median Spearman correlation ($\text{SCC}_{\text{bg.adj}}$) for a given KEGG module m as:

$$\text{SCC}_{\text{bg.adj}} = \text{median}(\text{SCC}_{\text{KOs}\in\text{KEGG module } m}) - \text{median}(\text{SCC}_{\text{KOs}\notin\text{KEGG module } m})$$

where SCC_{KO} is the (partial) Spearman correlation coefficient between the KO and HOMA-IR, $\text{HOMA-IR}_{\text{BMIadj}}$, $\text{HOMA-IR}_{\text{GeneRichness.adj}}$, gene richness or serum metabolite clusters. For associations of microbial functional modules with metabolic syndrome and type 2 diabetes, the ranks were based on Wald statistics for testing differentially abundant KOs with a negative binomial test, using the DESeq2 R package[73]. In contrast to all other tests, analyses using the negative binomial test were based on the non-rarefied gene counts.

To identify the main microbial species driving the association between functional KEGG modules and HOMA-IR described above, the calculation of the KO gene group abundance profiles was iterated excluding the genes from a different species, in each iteration (leave-one-out analysis). The effect of a given species on a specified association was defined as the change in median Spearman correlation coefficient between KOs and HOMA-IR when genes from the respective species were left out (DeltaSCC), that is, for species s and KEGG module m:

$$\text{DeltaSCC}_m^s = \text{median}\left(\text{SCC}_{\text{KOs}\in\text{KEGG module } m}^{\text{all species}}\right) - \text{median}\left(\text{SCC}_{\text{KOs}\in\text{KEGG module } m}^{\text{excluding species } s}\right)$$

For visualizing the species' effect on the association between HOMA-IR and a given KEGG module m, $\text{SCC}_{\text{bg.adj},m}$ (the background-adjusted median SCC between HOMA-IR and KOs in module m when including all species) was used as reference, that is:

$$\begin{aligned}\text{Plottedvalue}_m^s &= \text{SCC}_{\text{bg.adj},m} - \text{DeltaSCC}_m^s \\ &= \text{median}\left(\text{SCC}_{\text{KOs}\in\text{KEGG module } m}^{\text{excluding species } s}\right) - \text{median}\left(\text{SCC}_{\text{KOs}\notin\text{KEGG module } m}^{\text{all species}}\right)\end{aligned}$$

In addition, the magnitude of each species' relative contribution to the correlation was summarized as the percentage change compared to the original median Spearman correlation coefficient (% SCC), that is:

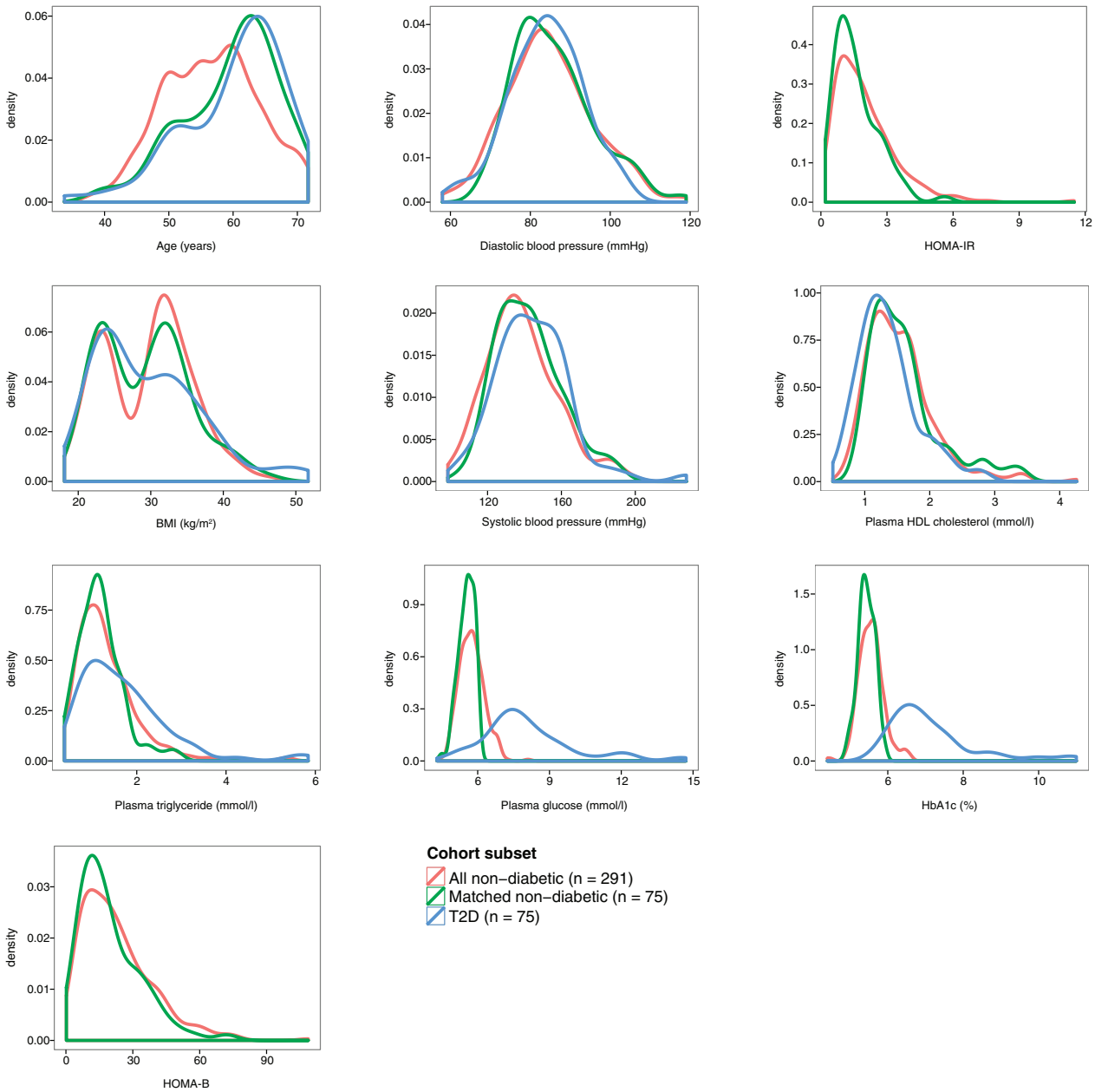
$$\%\text{SCC}_m^s = \frac{\text{DeltaSCC}_m^s}{\text{SCC}_{\text{bg.adj},m}} \times 100$$

When testing whether regression slopes were significantly different we used the z score as suggested by Paternoster *et al.*[74] and converted it to a P value using a 2-sided t distribution.

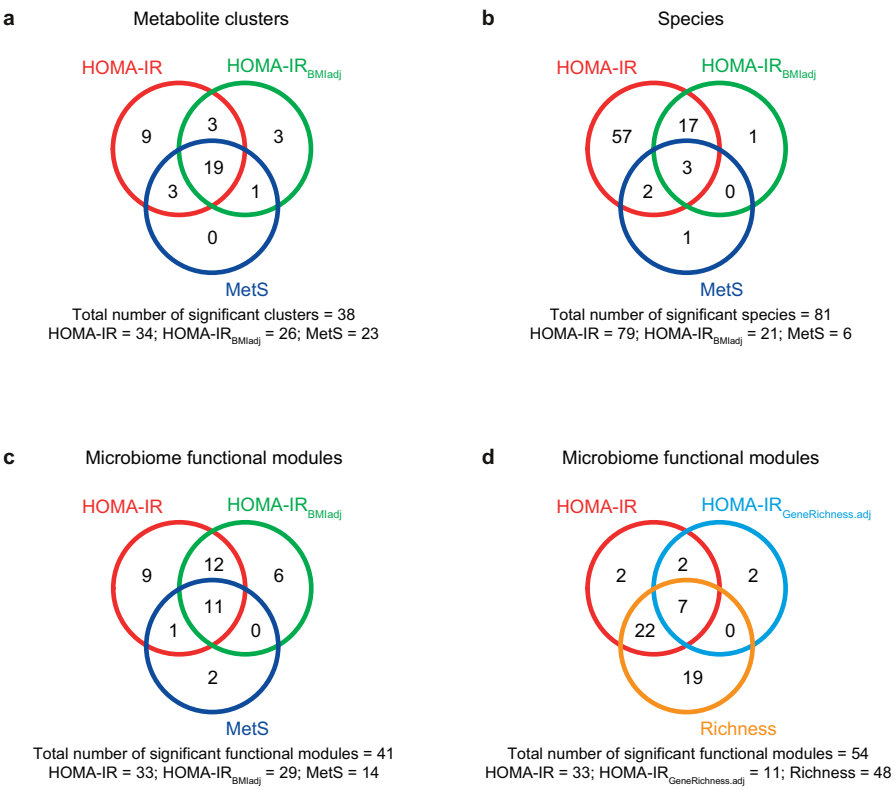
For analysis of the data from the mouse intervention, the repeated blood glucose measurements from the OGTT were compared between the two groups of mice using repeated measurements two-way ANOVA, Tukey's post hoc test (Sigmaplot). The area under the curve (AUC) (GraphPad) was compared using a Mann–Whitney U-test (GraphPad). The effect of the *P. copri* gavage challenge on the combined BCAA measures (leucine, isoleucine and valine) was tested on log-transformed metabolite values using a likelihood ratio test (linear model using generalized least squares), which does not assume independence between the amino acid measurements from the same mouse. For this, the variances of the three amino acids were modelled independently. To test for differences in the gut microbiota composition between sham and *P. copri*-gavaged mice we performed adonis test on the 16S data using weighted UniFrac distance. To test for bacteria families differentially abundant between sham and *P. copri*-gavaged mice we used the fitZig function in the metagenomeSeq R package[66].

46. Jørgensen, T. *et al.* A randomized non-pharmacological intervention study for prevention of ischaemic heart disease: baseline results Inter99. *Eur. J. Cardiovasc. Prev. Rehabil.* **10,** 377–386 (2003).
47. Castillo, S., Mattila, I., Miettinen, J., Orešič, M. & Hyötyläinen, T. Data analysis tool for comprehensive two-dimensional gas chromatography/time-of-flight mass spectrometry. *Anal. Chem.* **83,** 3058–3067 (2011).
48. Kopka, J. *et al.* GMD@CSB.DB: the Golm Metabolome Database. *Bioinformatics* **21,** 1635–1638 (2005).
49. Nygren, H., Seppänen-Laakso, T., Castillo, S., Hyötyläinen, T. & Orešič, M. Liquid chromatography-mass spectrometry (LC-MS)-based lipidomics for studies of body fluids and tissues. *Methods Mol. Biol.* **708,** 247–257 (2011).
50. Pluskal, T., Castillo, S., Villar-Briones, A. & Oresic, M. MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics* **11,** 395 (2010).
51. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9,** 559 (2008).

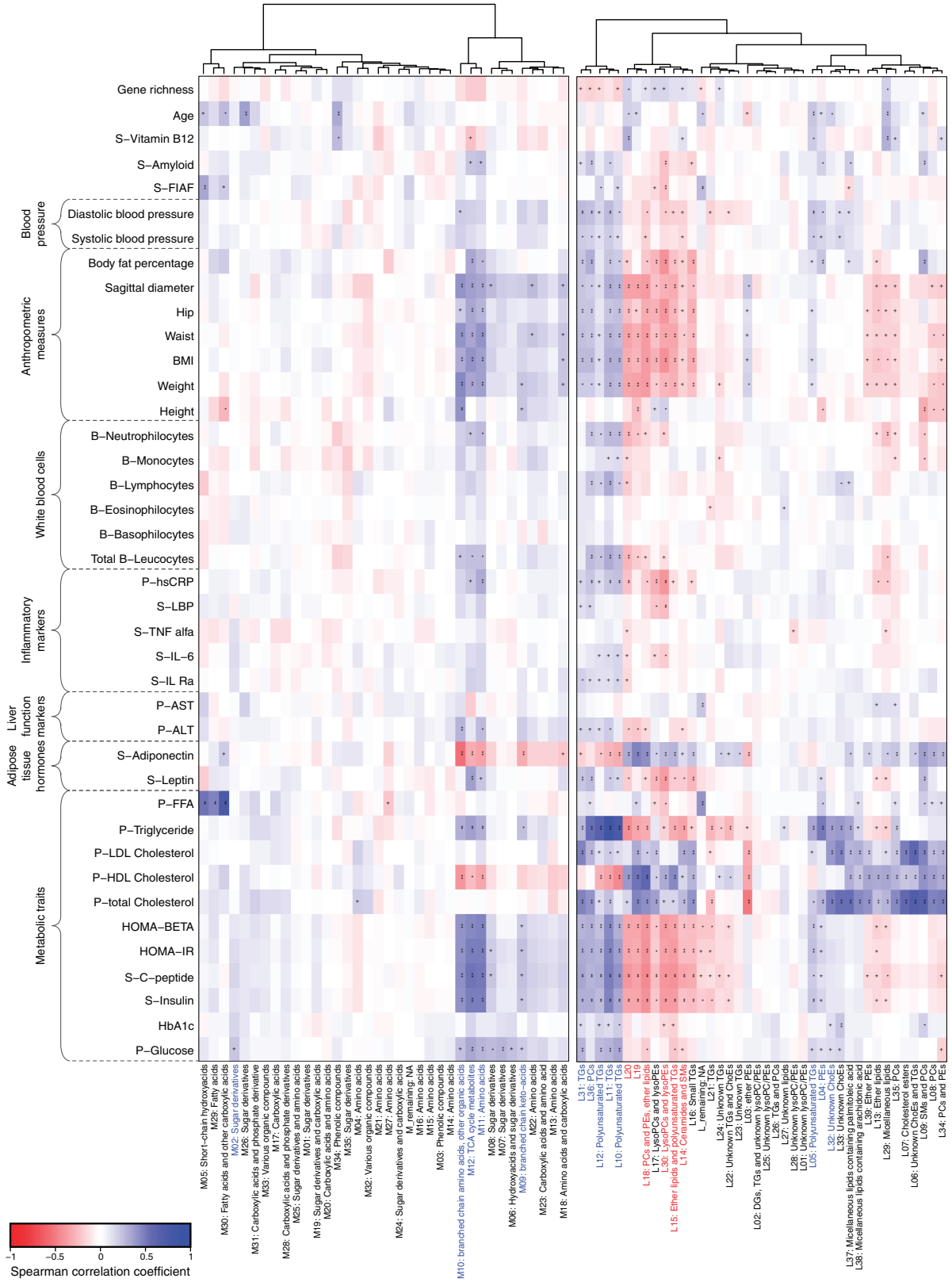
52. Langfelder, P. & Horvath, S. Fast R functions for robust correlations and hierarchical clustering. *J. Stat. Softw.* **46,** i11 (2012).
53. Mosteller, F. & Tukey, J. W. *Data Analysis and Regression. A Second Course in Statistics*, 203–209 (Addison–Wesley, 1977).
54. Langfelder, P., Zhang, B. & Horvath, S. Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. *Bioinformatics* **24,** 719–720 (2008).
55. Li, J. *et al.* An integrated catalog of reference genes in the human gut microbiome. *Nat. Biotechnol.* **32,** 834–841 (2014).
56. Kultima, J. R. *et al.* MOCAT: a metagenomics assembly and gene prediction toolkit. *PLoS One* **7,** e47656 (2012).
57. Li, W. & Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22,** 1658–1659 (2006).
58. Arumugam, M., Harrington, E. D., Foerstner, K. U., Raes, J. & Bork, P. SmashCommunity: a metagenomic annotation and analysis tool. *Bioinformatics* **26,** 2977–2978 (2010).
59. 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).
60. Caporaso, J. G. *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **7,** 335–336 (2010).
61. Kozich, J. J., Westcott, S. L., Baxter, N. T., Highlander, S. K. & Schloss, P. D. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl. Environ. Microbiol.* **79,** 5112–5120 (2013).
62. Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C. & Knight, R. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **27,** 2194–2200 (2011).
63. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26,** 2460–2461 (2010).
64. DeSantis, T. Z. *et al.* Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl. Environ. Microbiol.* **72,** 5069–5072 (2006).
65. Wang, Q., Garrity, G. M., Tiedje, J. M. & Cole, J. R. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* **73,** 5261–5267 (2007).
66. Paulson, J. N., Stine, O. C., Bravo, H. C. & Pop, M. Differential abundance analysis for microbial marker-gene surveys. *Nat. Methods* **10,** 1200–1202 (2013).
67. McMurdie, P. J. & Holmes, S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* **8,** e61217 (2013).
68. Oksanen, J. *et al.* vegan: Community Ecology Package. R package version 2.3-3. https://cran.r-project.org/web/packages/vegan/index.html (2016).
69. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. (Springer, 2009).
70. R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. (2012).
71. Whittaker, J. *Graphical models in applied multivariate statistics.* (John Wiley & Sons, 1990).
72. Seongho, K. ppcor: Partial and semipartial (Part) correlation. R package version 1.0. https://cran.r-project.org/web/packages/ppcor/index.html (2012).
73. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15,** 550 (2014).
74. Paternoster, R., Brame, R., Mazerolle, P. & Piquero, A. Using the correct statistical test for the equality of regression coefficients. *Criminology* **36,** 859–866 (1998).
75. Lee, S., Chang, J. & Blackstone, C. FAM21 directs SNX27–retromer cargoes to the plasma membrane by preventing transport to the Golgi apparatus. *Nat. Commun.* **7,** http://dx.doi.org/10.1038/ncomms10939 (2016).



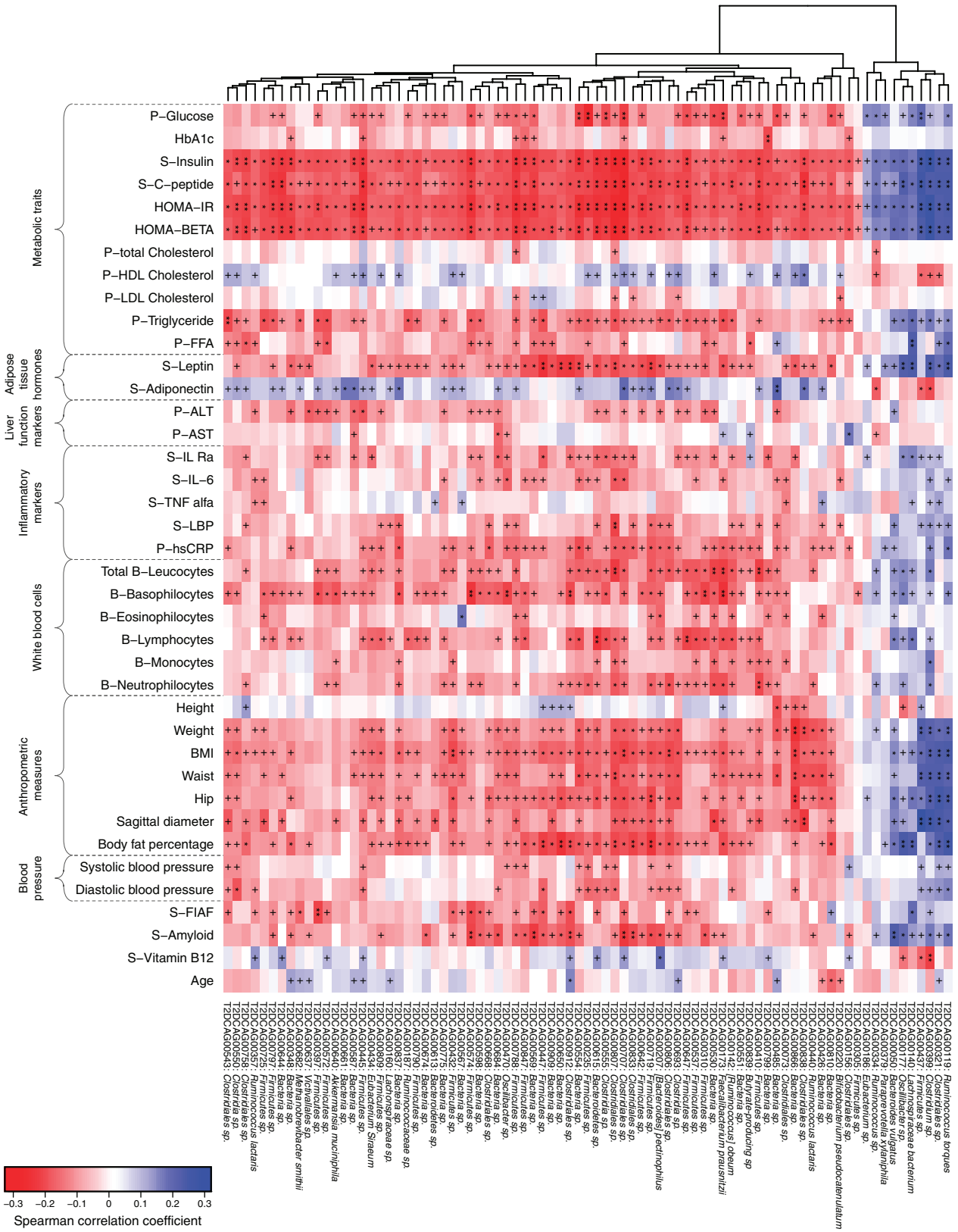
Extended Data Figure 1 | Distributions of continuous physiological traits for the 291 non-diabetic individuals, 75 type 2 diabetes patients and 75 matched non-diabetic controls. An overview of the same traits is shown in Supplementary Table 1. The 75 non-diabetic controls are a subset of the 291 non-diabetic individuals matched to the type 2 diabetes patients by age, sex and BMI and used for comparative analyses.



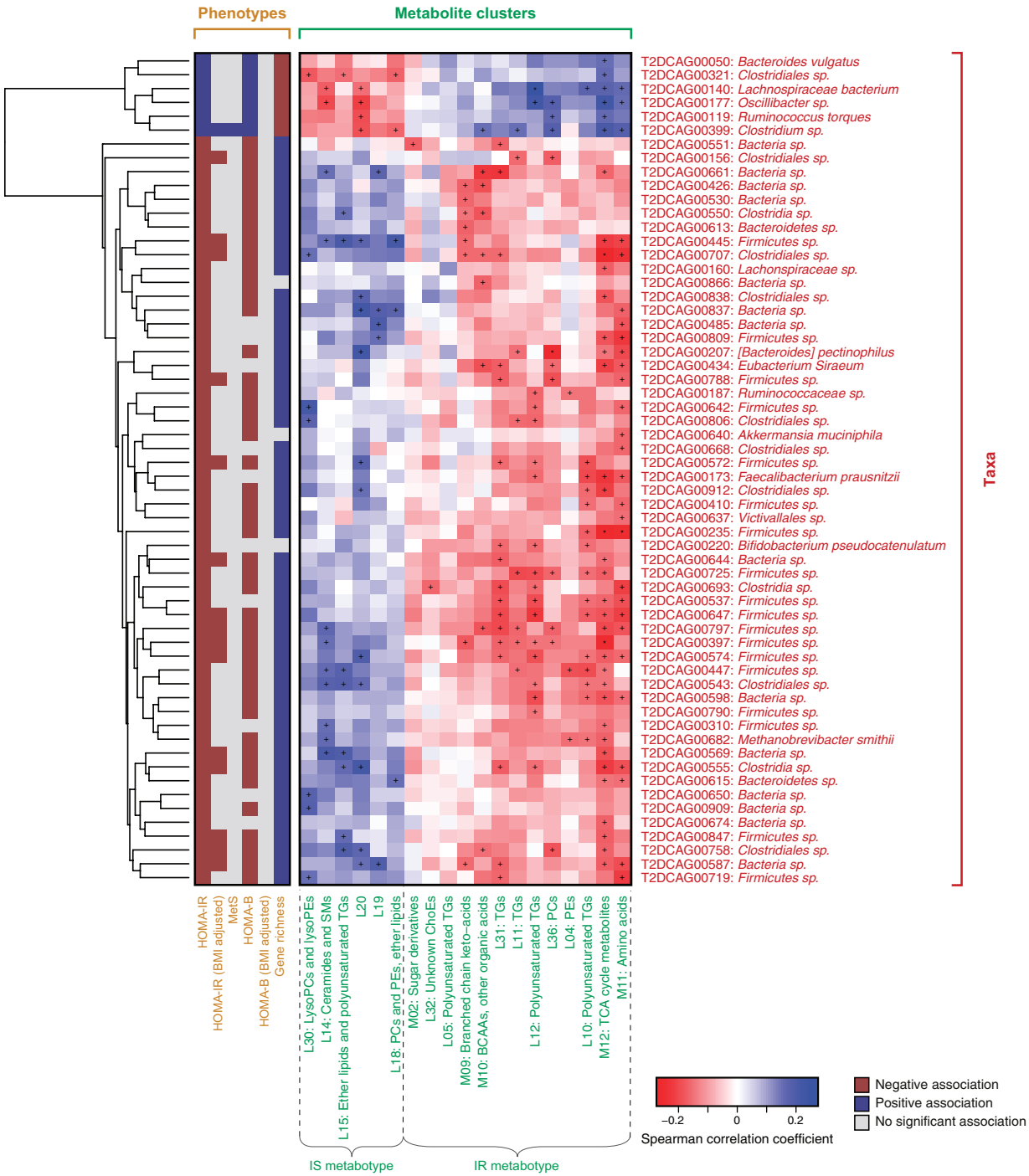
Extended Data Figure 2 | The number of metabolite clusters, species and microbiome functional modules significantly associated with HOMA-IR, HOMA-IR$_{BMIadj}$, and metabolic syndrome. a–c, Venn diagrams resuming the number of serum metabolite clusters (**a**), species (**b**) and microbiome functional modules (**c**) that are associated with the three HOMA-IR, HOMA-IR$_{BMIadj}$ and metabolic syndrome phenotypes at FDR < 0.1. **d**, The number of microbiome functional modules associated with HOMA-IR, gene richness and HOMA-IR$_{GeneRichness.adj}$. The metabolite cluster associations are based on all 291 non-diabetic individuals whereas the species and KEGG module associations were estimated on the 277 non-diabetic individuals with microbiome data. MetS, metabolic syndrome.



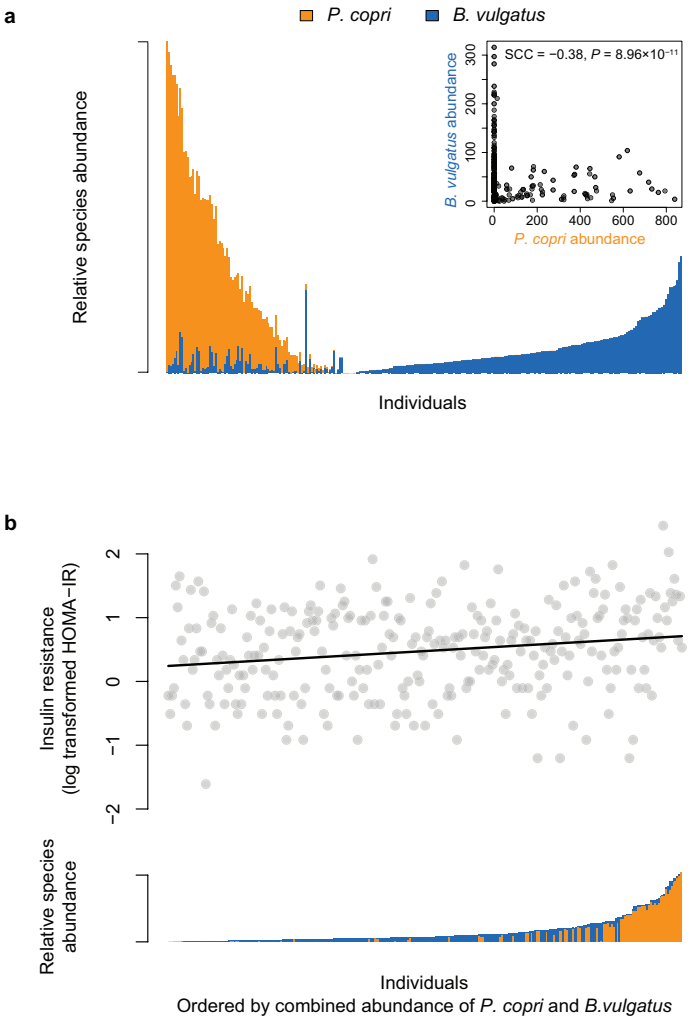
Extended Data Figure 3 | Fine-grained correlation profile of fasting serum metabolite clusters and physiological traits in 291 non-diabetic individuals. Spearman correlations between all fasting serum metabolite clusters (top panel, molecular lipids; bottom panel, polar metabolites) and clinical phenotypes. The metabolites in each panel are clustered by their correlation profile (see dendrogram). The colour represents positive (blue) or negative (red) correlations and FDRs are denoted: +, FDR < 0.1; *, FDR < 0.01; **, FDR < 0.001. The names of the 19 metabolite clusters making up the IR- and IS-metabotypes are highlighted with blue (IR-metabotype) and red (IS-metabotype), respectively.



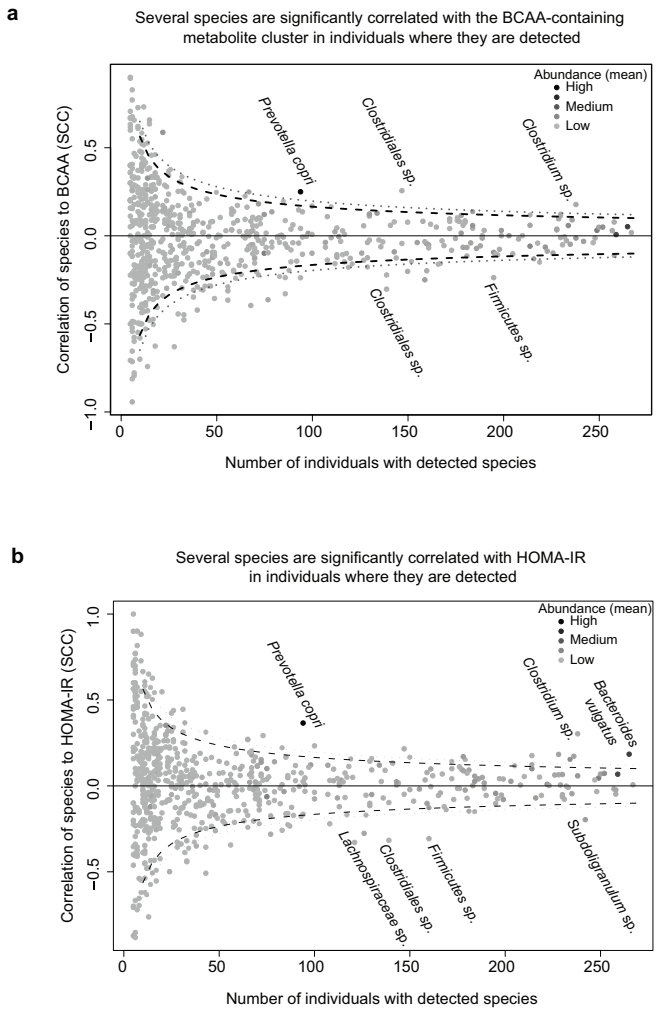
Extended Data Figure 4 | Fine-grained correlation profile of IR- and metabolic-syndrome-associated microbial species and physiological traits in 277 non-diabetic individuals. Spearman correlations between continuous physiological traits and the 81 species significantly associated (FDR < 0.1) with HOMA-IR, HOMA-IR$_{BMIadj}$ or metabolic syndrome phenotypes (Extended Data Fig. 1). The species are clustered by their correlation profile. The colour represents positive (blue) or negative (red) correlations and FDRs are denoted: +, FDR < 0.1; *, FDR < 0.01; **, FDR < 0.001.



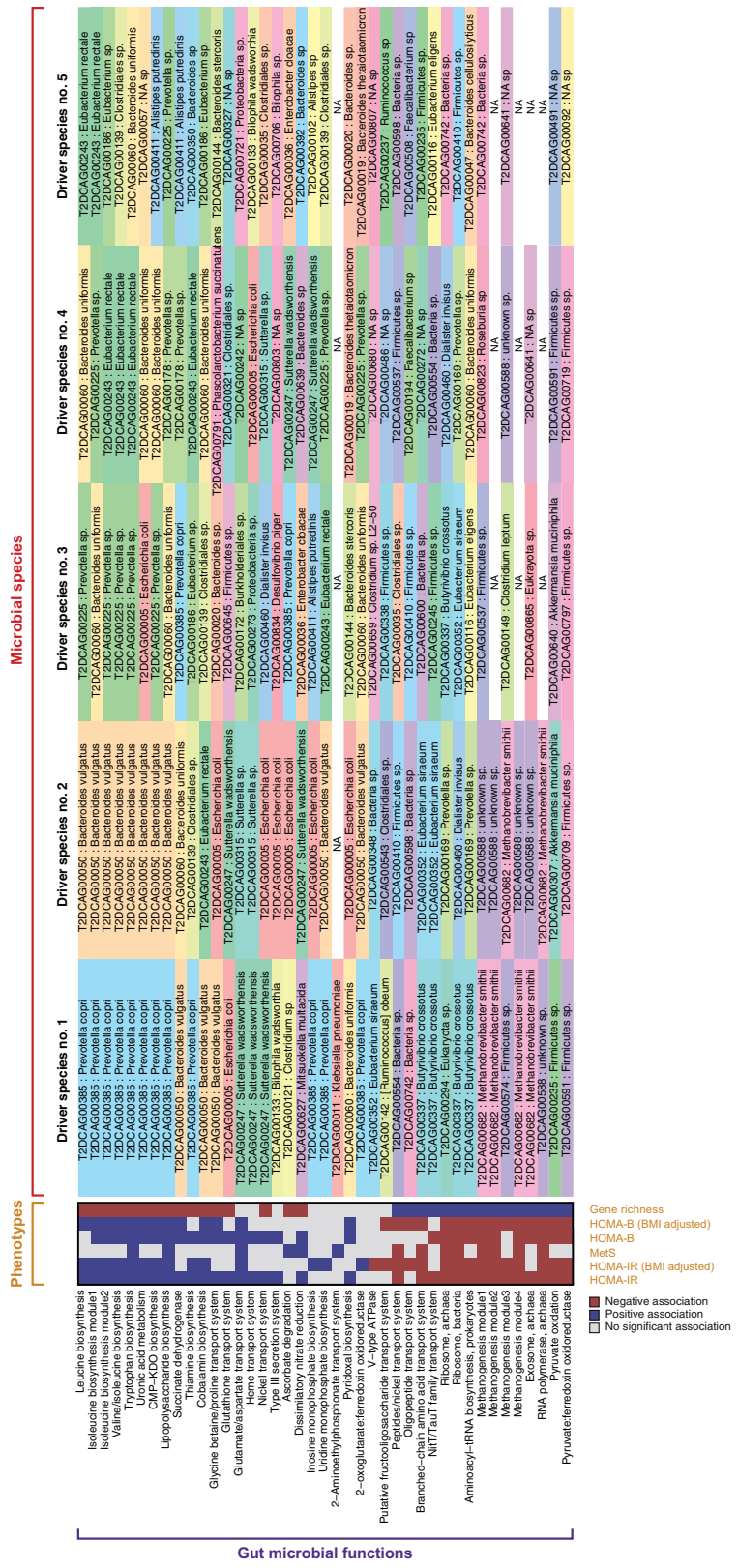
Extended Data Figure 5 | Correlations between IR- and metabolic syndrome-associated microbial species and fasting serum metabolite clusters in 277 non-diabetic individuals. Spearman correlations between species that were significantly associated (FDR < 0.1) with HOMA-IR, HOMA-IR$_{BMIadj}$ or metabolic syndrome phenotypes and the 19 metabolite clusters making up the IR- and IS-metabotypes. The colour represents positive (blue) or negative (red) correlations and FDRs are denoted: +, FDR < 0.1; *, FDR < 0.01; **, FDR < 0.001. The left sidebar represents positive (blue) or negative (red) correlations between the species and the indicated phenotypes (FDR < 0.1). MetS, metabolic syndrome.



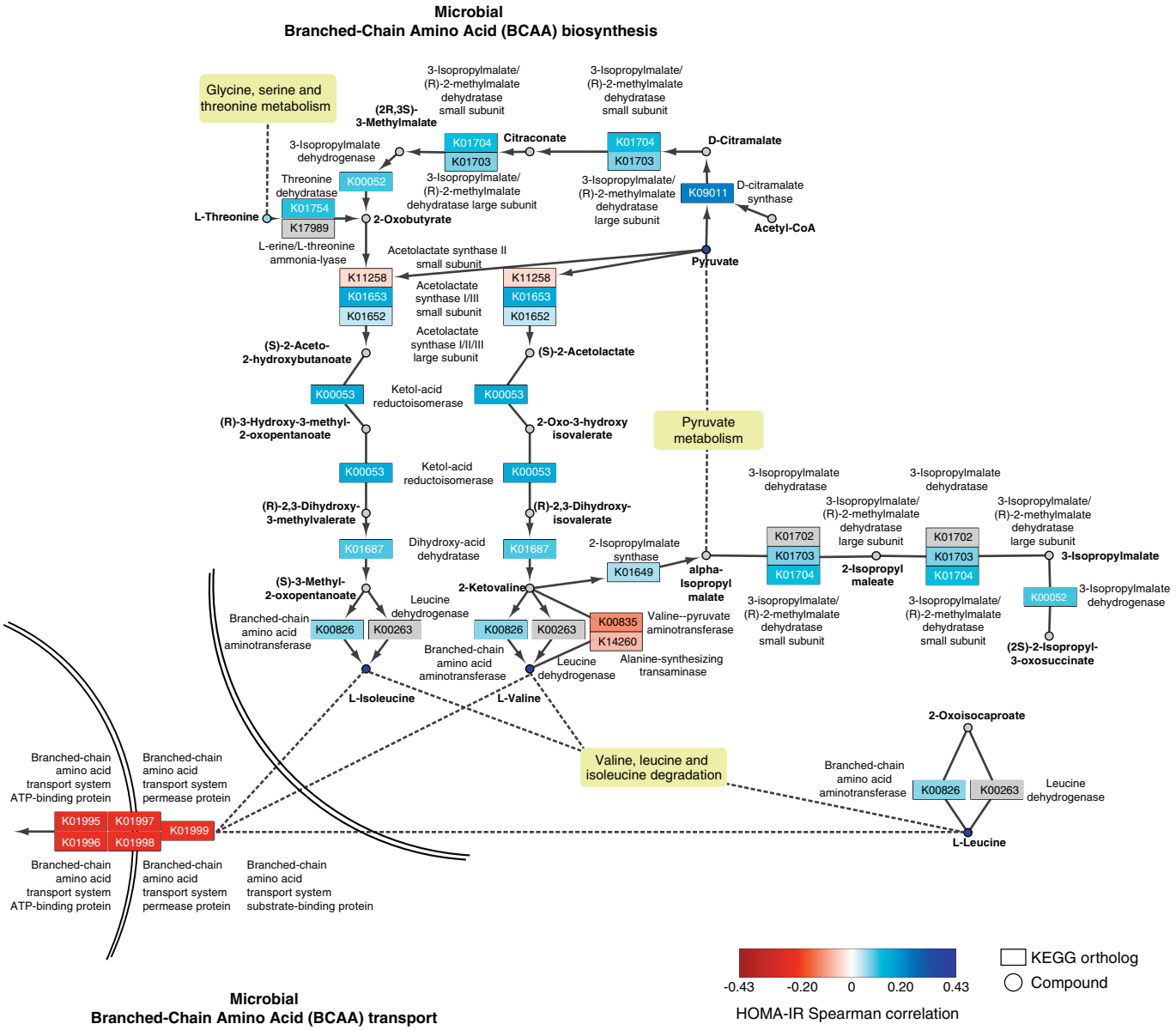
Extended Data Figure 6 | Abundances of *P. copri* and *B. vulgatus* anti-correlate and their combined abundance correlates with HOMA-IR in 277 non-diabetic individuals. a, **b**, The abundances of T2DCAG00385: *P. copri* (orange) and T2DCAG00050: *B. vulgatus* (blue), shown for all non-diabetic individuals arranged by decreasing *P. copri* abundance and increasing *B. vulgatus* abundance (**a**), and arranged by total abundance of both species with HOMA-IR levels shown above (**b**).



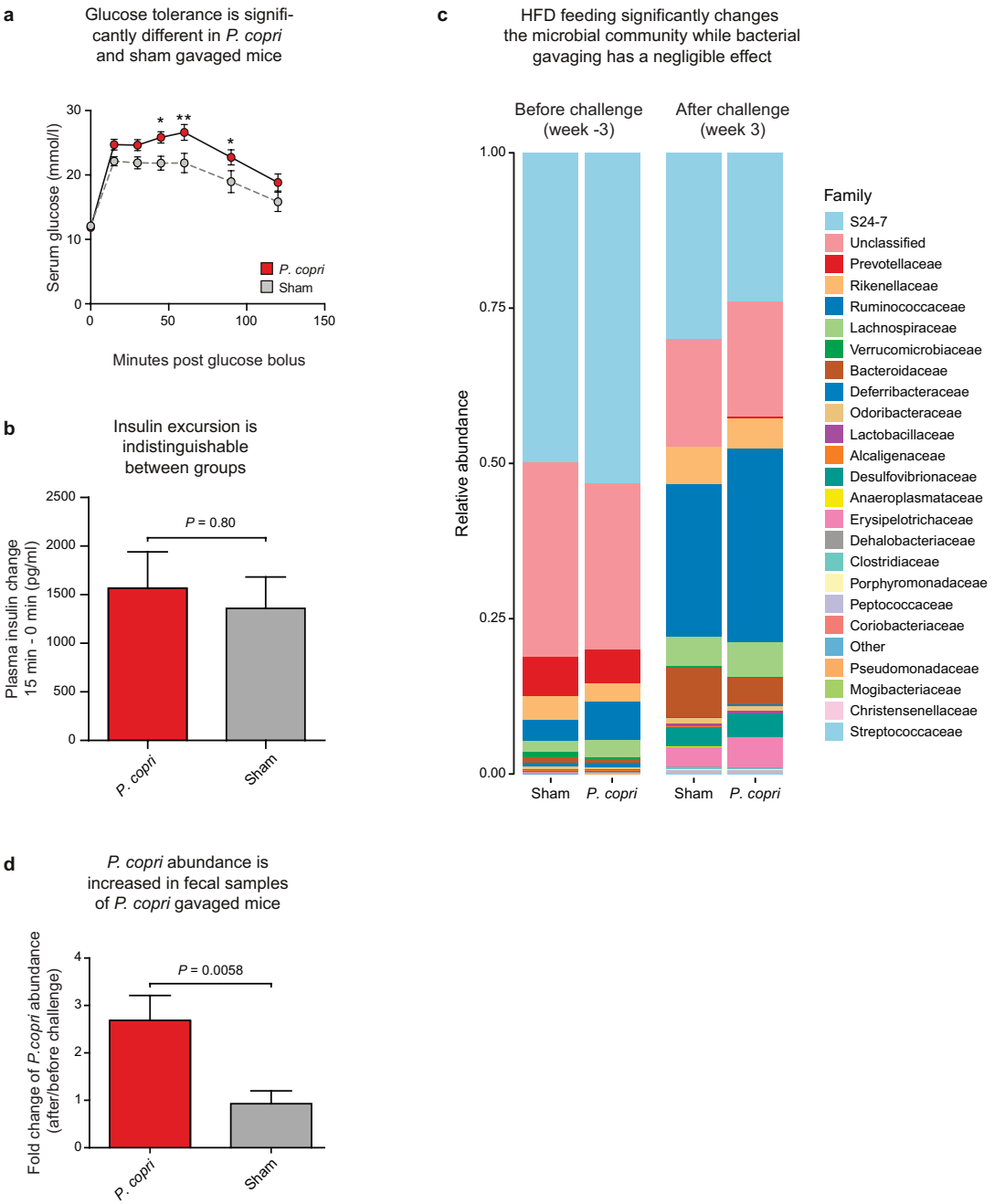
Extended Data Figure 7 | Correlations between microbial species and both HOMA-IR and the BCAA-containing metabolite cluster (M10) in 277 non-diabetic individuals. a, **b**, Spearman correlations between species and both the BCAA-containing metabolite cluster (**a**, M10) and insulin resistance (**b**, HOMA-IR) in individuals with detectable abundances of the respective species. FDRs of 0.1 and 0.05 are denoted with dotted and dashed lines, respectively. Colour intensity represents mean species abundance in individuals where the species was observed.



Extended Data Figure 8 | Microbial driver species for associations between microbiome functional modules and insulin resistance in 277 non-diabetic individuals. The five most important microbial species driving the association between the indicated microbiome functional modules and insulin resistance (HOMA-IR) are shown (see Supplementary Table 9 for effect sizes). Each species is highlighted with a different colour. The left sidebar represents positive (blue) or negative (red) associations between the functional modules and the indicated phenotypes (FDR <0.1). MetS, metabolic syndrome.



Extended Data Figure 9 | An in-depth view of the microbial BCAA biosynthesis pathway and BCAA inward transport system, illustrating the correlations between microbial KEGG orthologous gene groups and serum metabolites with human insulin resistance. KEGG orthologous gene groups (squares) and metabolites (circles) are coloured by their Spearman correlation with HOMA-IR in the non-diabetic individuals ($n = 277$ for KEGG orthologous gene groups, $n = 291$ for metabolites), or coloured grey if no information was available. The network is adapted from KEGG pathway maps (pathways hsa00290 and hsa02010).



Extended Data Figure 10 | Oral glucose tolerance test after two weeks of *P. copri* or sham gavaging and 16S rDNA amplicon sequencing of faecal samples from mice after three weeks of treatment with *P.copri* or *sham*. **a**, Oral glucose tolerance test. The *P. copri*-gavaged mice ($n = 12$) had significantly higher serum glucose levels compared to sham-gavaged controls ($n = 12$, $P = 0.02$, Mann–Whitney U-test for AUC) after two weeks of the gavage challenge. Mean ± s.e.m. is depicted. Stars indicate significant differences at individual time points (repeated measurements two-way ANOVA): $^{*}P < 0.05$; $^{**}P < 0.01$. **b**, Plasma insulin was measured before and 15 min post glucose bolus, $P = 0.80$, Mann–Whitney U-test, bars represents mean ± s.e.m., $n = 12$. **c**, Microbiota taxa summary plots on family level after the two given time points, that is, pre high-fat diet (HFD) and post HFD plus gavage. HFD feeding significantly changed the microbial community (adonis $P = 0.001$) while bacterial gavaging had negligible effect. Data represent mean values. $n = 12$ per group (one sample from the sham group at time point −3 weeks did not go successfully through the 16S rDNA amplicon sequencing and is therefore represented by 11 samples). 'Unclassified' refers to reads that could not be classified to any taxonomy. 'Other' refers to reads that could not be classified at family level. **d**, *P. copri* changes in mouse faecal samples after *P. copri* gavaging as determined by quantitative PCR. Bars show the change in *P. copri* levels relative to before *P. copri* or sham challenge (bars represents mean ± s.e.m., $n = 12$, $P = 0.0058$, Mann–Whitney U-test).