doi:10.1038/nature15766

# Supplementary Discussion

## 1. Validation of the computational approach

### 1.1. Comparison of computational pipelines in present and previous studies

The previous T2D metagenome studies of Qin et al.[3] and Karlsson et al.[4] each use subtly different bioinformatics analysis approaches. Karlsson et al. first determined taxonomic composition by aligning reads directly to 2382 reference genomes. Next, reads were additionally aligned to a novel gene catalog made from assembly of their novel samples merged with the previous MetaHIT gene catalog. Wherever abundant genes were correlated in abundance across subjects, the resulting groups were treated as metagenomic clusters (MGCs) identified from the data. These were then linked to known taxa where possible through BLAST searches. The resulting two representations of taxonomy were used in the study as features for prediction of T2D status. Functional profiles were determined by assigning gene catalog members to the KEGG database through USEARCH searches and then deriving KEGG pathway abundances based on the gene abundances. Qin et al. also created a novel reference gene catalog based on their novel samples together with the previous MetaHIT gene catalog. About 21% of the genes in this catalog were assigned to a bacterial genus by BLAST sequence search in a reference gene catalog of 2890 bacterial genomes, 47% to KEGG orthologs and 61% to eggNOG orthologs. Analogously to the MGCs in Karlsson et al. correlated genes were grouped together as metagenomic linkage groups (MLGs) and the abundances of these groups were used to carry out the analysis. Thus, all three studies use separate but roughly similar methods for inferring groups of correlated genes from metagenomes directly – MGS, MGC, MLG, respectively.

Of these the MGSs/CAGs[34] are the most recently published and we consider them most reliable. In order to make use of the MGSs, therefore, the analysis in the present study was constrained to use the most comprehensive gene catalog (from 620 MetaHIT samples) for which they were available. While this means that rare gene variants from Sweden or China may remain undetected in the current analysis, no presently publicly available gene catalog was derived from all three datasets. To avoid false conclusions based on this limitation, we therefore focused the present analysis on similarity rather than difference; searching for whatever signal of T2D pathology or medication instead might find robust support across these datasets. While we likely miss further markers and correlates of T2D or metformin treatment in this manner, it should not reduce the reliability of the markers we do detect and report. We anticipate future studies using further expanded gene catalog resources which might validate and extend our findings. However, we note that the average fraction of inserts mapped to genes in the presently chosen gene catalog (Supplementary Table S1) is very similar (89-91%) for all three datasets, suggesting that only minor contributors are missed.

On a level of functional profiles, the present analysis differs from the previous two in that it uses a database (GMMs) which extends and curates the otherwise often eukaryote-centric KEGG resource to better represent functional pathways in human gut microbes. Inferring pathway abundances from KEGG KO members, the approach in the two previous studies risks concluding presence of a pathway from the presence of a single abundant homolog of any of the pathway members or the presence of a pathway present only when genomes of several species are pooled, but not in any one genome individually. To avoid this, we instead adopted the present approach which draws on the MGS resource as a scaffold for possible pathway presence as described here and in the Methods. In short, functional profiles were determined by investigating each MGS/CAG for completeness of every pathway, then using the abundance of these units of correlated genes to determine the final

functional profiles. This approach may be slightly less sensitive and miss presence of rare functional pathways, but should more seldom falsely conclude presence or high abundance of a pathway, which meets our objective of finding a robust "core" signal of functional dysbiosis in T2D.

On a level of taxonomic composition, the recently introduced mOTU technology was chosen over the approaches of the two source studies (mapping reads directly to genomes, mapping called genes directly to reference genomes); it utilizes a set of marker genes curated and benchmarked to reliably occur in single copy across bacterial genomes as well as not to transfer horizontally, thereby making the method more accurate in quantifying taxonomic composition also for poorly characterized taxa with few or no sequenced genomes which we expected to be relevant in the context of the gut; mOTUs are further assigned to known taxa where possible using a database larger than used in either source study. While the MLGs, MGCs or MGSs each to some extent may capture uncharacterized taxa, they will involve a more diverse, as well as non-overlapping, collection of genes for each. For this reason, we chose to use the mOTUs rather than MGSs for taxonomic composition estimate, but the MGSs, rather than the mOTUs, for functional profile quantification. As seen below, some of the taxonomic associations previously reported involve microbial strains which are not represented as distinct mOTUs, meaning they cannot in principle be replicated or can only be replicated at a broader level. In any such case assignment to a reference genome would occur at a level of sequence similarity higher than that used for mapping sequencing reads for the purpose of assessing taxonomic composition. Such assignments therefore imply a taxonomic specificity which is not borne out when using the reference database for mapping reads, suggesting greater specificity than the metagenomic data actually allows. For this reason, we consider the lower granularity of some mOTUs to be a benefit rather than a drawback, and to ensure we are on

the side of precision rather than recall, providing another motivation for the current approach rather than that of either source study.

### 1.2. Comparison of reanalysis to published results on SWE and CHN datasets

We applied our analysis pipeline (filtering and quality control, mapping of samples to mOTU marker genes, quantification of taxonomic composition, dataset-blocked test for significant differences in composition) to the data from the two earlier source studies we incorporate, to see to what extent previously reported associations are possible to replicate using the same samples. Comparing all T2D and ND CTRL samples, the results of this conceptual replication are shown in Supplementary Table S2. Out of 41 associations reported in the two studies, 22 achieved significance in the same direction also in the joint dataset, either with the taxon represented as a mOTU or otherwise significance being seen at the corresponding higher level or for multiple mOTUs of the same genus if it is not. For 7 previously seen associations the microbial feature is exactly represented but does not differ significantly between T2D and ND CTRL under the blocked test, in some cases reflecting signals that are unique to each dataset. Remaining discrepancies largely follow from these previous studies attempting to separately quantify taxa that do not correspond to distinct mOTUs (e.g. multiple strains of the same species with very high marker gene sequence identity to each other, or else taxa not present within the collection of gut microbiomes used to derive the mOTU marker genes). Reanalysis of the samples from each source study taken separately and compared with the originally reported findings from each show on average the same degree of replication, but with more (19/26) associations replicated from Qin et al. (2012)[3] (Supplementary Table S2) than from Karlsson et al. (2013)[4] (3/15). The latter is based on fewer samples, and many of the reported associations reference specific unique strains of *Clostridium* and

*Lactobacillus* which are too similar in sequence to be represented as different mOTUs, therefore making direct replication harder. All in all, therefore, while a majority of previously reported associations achieved significance here, not all did, and this follows from how the three studies delineate taxa differently. See above for discussion on why mOTUs and the set of curated universal single-copy marker genes were used here instead of e.g. mapping to reference genomes directly.

### 1.3. Benchmarking of the metagenome meta-analysis pipeline using positive and negative simulated controls

Since the approach of integrating multiple datasets in the way used here is novel for metagenomic analysis, we evaluated the performance of the overall analysis pipeline under positive and negative control conditions as follows.

A negative control for a meta-analysis is done by applying the method to data where it is known no signal exists. To generate such data we randomly reshuffled sample labels within each source dataset, thereby eliminating any actual signal, and reran the T2D metformin+ vs T2D metformin- comparison analysis. No (spurious) significant differences resulted, demonstrating the meta-analysis holds up with regards to negative controls.

A positive control for a meta-analysis is done by applying the method to data where it is known a signal does exist. We therefore added an artificial signal to the negative (shuffled) benchmark set by transforming *Akkermansia* abundances on a scale similar to that of the *Escherichia* metformin treatment effect as seen from the MHD dataset (metformin-treated samples roughly 150% as likely to have nonzero abundance, with a roughly threefold higher abundance where present). These results are shown in Extended Data Figure 1a. Across 100 randomizations, this artificial signal is

reliably detected at FDR scores comparable to those of the *Escherichia* signal from the main dataset.

Given these controls, we are confident that our approach is capable of detecting signals of a strength such as those reported, and that this capacity does not come at the price of any strongly increased risk of false positives resulting from the meta-analysis pipeline itself.

## 2. Initial analysis of the joint dataset

### 2.1. Metagenomic, ecological and demographic comparisons of the three cohorts

Integrating the two previously published sets of T2D metagenomes[3,4] (with the Danish MetaHIT diabetes samples) proved challenging. Phenotypically we observed significant differences between the three study samples. The SWE samples comprise only postmenopausal women and the CHN T2D patients were younger than MHD and SWE patients. BMI was higher in MHD samples than in SWE or CHN samples (Extended Data Figure 2a). For clinically relevant markers, we found glycated haemoglobin A1c (HbA1c), reflecting long-term blood glucose levels, to be lowest in MHD T2D and highest in SWE ND CTRL samples ($p<2x10^{-16}$). Fasting plasma glucose levels were highest in CHN T2D patients and lowest in CHN ND CTRL, compared to corresponding SWE & MHD samples ($p=5.6x10^{-4}$). Based on these clinical metadata, the cohorts are clearly disjoint (Supplementary Table S2).

On a microbiome level, the composition of samples from these cohorts was also very different: various ecological indices (Shannon diversity, richness, chao1 and evenness as measured on gene

level) were lowest in the Chinese samples and highest in the SWE samples (all indices KW p=0 for equivalent distributions). Using a multivariate PERMANOVA test ($10^5$ randomizations), all cohorts were significantly ($p<1x10^{-5}$) different from each other, as shown in a PCoA of Bray-Curtis divergence between samples (Extended Data Figure 2b). Further, beta diversity was significantly different between cohorts ($p=1.6x10^{-17}$), CHN samples being more dispersed than MHD samples ($p=3x10^{-8}$) and MHD samples more dispersed than SWE samples ($p=4.7x10^{-7}$). The compositional differences were largely driven by an increase in *Bacteroides* ($p=2.2x10^{--73}$) and *Proteobacteria* ($p=3.9x10^{-46}$) in CHN samples, while *Firmicutes* ($p=1x10^{-79}$), *Actinobacteria* ($1.3x10^{-37}$), *Verrucomicrobia* ($p=3.5x10^{-45}$), and *Euryarcheaota* ($p=7.7x10^{-28}$) were increased in SWE samples with MHD samples being intermediate (Supplementary Table S2).

### 2.2. Technical variability between datasets

While all cohorts were sequenced using the Illumina platform, technological artifacts such as the specifics of microbial DNA extraction, stool sample collection and handling likely contribute to these differences, but it cannot be ruled out that geographical separation and disjoint participant demographics (see above) underlie some of these between cohort differences.

We recognize that extraction protocols may influence the recovered gut microbiome composition, as there is variation between different kits/methods in how they target different bacterial groups. However, we note that this variation has been reported to consistently be below the inter-individual variation[47] including studies of direct reextraction of the same 11 samples using both the Chinese and the MetaHIT protocol[26]. Furthermore, the greatest difference seems to come down to the lysis method of choice. As repeatedly reported[54-56], beat beating is crucial for recovering high amounts of

DNA and a correct gram-positive quantification. In this context, we note that alongside the MetaHIT protocol, both the Swedish and the Chinese protocol use beat beating, at comparable duration and using exactly the same 0.1 mm zirconium-silica beads. Both protocols then use an isopropanol precipitation for recovering the DNA. Here, the Swedish protocol differs by subsequently using a QIAamp DNA stool mini kit column for purifying DNA. We cannot exclude that this may introduce a bias, but have no reason to expect such an effect. We therefore estimate that extraction protocol differences are unlikely to explain the bulk of dataset differences. Even where they do play a part, our statistical approach which controls for source dataset in each test is specifically intended to prevent any such factors from affecting the conclusions drawn when comparing cases/controls or treated/untreated samples.

Beyond protocol and differences in individual host demographics (age, sex, BMI, fraction of controls versus cases...) we also have to consider the possibility of genuine geographic differences in the human microbiome. Diet is an obvious cause, but we cannot easily test its impact in the absence of diet data in the source studies. Another factor which may play a part is exposure to antibiotic compounds, which may shift microbiota towards taxa more prone to carry antibiotic resistance genes. Previous work (reviewed in Forslund et al. 2013[57]) show an effect at the community level wherein exposure of patients or farm animals to antibiotics increases the carriage of such genes more generally in human communities. Furthermore, a previous survey[57] including the SWE and CHN cohorts, as well as many of the MHD samples, showed higher carriage of antibiotic resistance genes in the Chinese metagenomes than the Swedish or Danish, matching correspondingly higher use of antibiotics in medicine and animal husbandry in China than in northern Europe. Reexamining those results (intermediate results from Forslund et al. 2013[57] available on request) also show how the CHN samples are significantly shifted towards bacterial

taxa where antibiotic resistance gene carrying strains are known compared to the SWE and MHD samples, indicating that the taxonomic difference between the cohorts to a significant extent can be understood as a shift towards those bacteria which are more prone to antibiotic resistance. Notably, these include *Escherichia*, which may explain why no strong shift in this taxon is seen under metformin treatment in the CHN cohort – the healthy samples already have *Escherichia* levels comparable to those of metformin-treated SWE or MHD samples, something which then may be a consequence of the higher degree of antibiotic exposure at the population level.

Thus, despite all three cohorts stemming from shotgun sequencing-based quantitative metagenomics studies that focused on exploring the effect of T2D on the gut microbiome, from a medical perspective as well as from a microbiological perspective there were substantial differences between datasets that needed to be taken into account for any further statistical analysis. To address these issues, for the present study a methodology was chosen that explicitly models study source as a confounder in all applicable tests[42].

### 2.3. Initial analysis of T2D microbiome signatures unstratified for treatment

Comparing gut microbial taxonomic and functional profiles between T2D (n = 199) and control (n = 554) metagenomic samples, we recover a majority of associations previously reported as significant[3,4], also taking into account the novel Danish MetaHIT T2D samples (Supplementary Table S3). Metagenomes from all T2D patients regardless of medication status, when compared with non-diabetic samples (ND CTRL), show enrichment of the bacterial genus *Lactobacillus* and depletion of the newly characterized genus *Intestinibacter*[15]. However, the inferred associations overlap only partly between the three cohorts, with 7 of 26 genus-level associations not supported

from multiple datasets (Supplementary Table S3). Significant demographic and microbial (functional, taxonomic and diversity-associated) differences exist between the study sets (see above Supplementary Discussion 2.1, Supplementary Table S2) which could contribute to this divergence. It is currently difficult to determine to what extent these reflect genuine geographic, ethnic or demographic microbiome differences, and to which extent they follow from the different sampling and DNA extraction protocols used in the studies (see above Supplementary Discussion 2.2). To account for this uncertainty, all statistical tests were explicitly blocked for study source as confounder[42].

## 3. Influence of diet and medication on the gut microbiome

### 3.1. Effects of medications taking statistical power into account

Testing the influence of various glucose-lowering drugs prescribed to T2D patients, we found that at genus level, therapy-attributable variability in our joint dataset could be largely explained from only metformin treatment status (no significant differences were found testing for compositional changes as a result of any of the other drugs taken by diabetic patients; see Supplementary Table S4, though smaller effects could have been missed due to lower statistical power for the other medications tracked). We therefore analyzed the detection power of the study using two different strategies. First, we replicated the analysis both for effects of metformin and for those of other tracked medications, but using a subset of samples generated by randomly selecting an equal, set number of cases and controls. This was repeated five times for each such sample count, and in each case, the number of bacterial genera found significantly different under different FDR thresholds

was counted. As the size of these subsampled datasets increase, two bacterial genera (*Escherichia* and *Intestinibacter*) were robustly found significantly different in abundance under metformin treatment, but with no such taxa for other treatments (Extended Data Figure 1b). Effects of the other tested drugs are therefore likely smaller than those of metformin. To explicitly test this conclusion, we first calculated standardised (Cohen's) effect sizes for the two bacterial genera impacted by metformin. We then used the G*Power application[58] to compute the statistical power a WRS test would achieve for the other three treatments assuming the same effect sizes and a type I error rate of 0.05, and with case/control numbers corresponding to each other treatment. The results are shown in Supplementary Table S4. An effect as strong as that of metformin on *Intestinibacter* would likely have been detected, given the expected power for the other three treatments ranged between 0.66 and 0.9. In contrast, a relatively weaker effect such as that of metformin on *Escherichia*, would likely have been missed, as the expected power under those conditions was only 0.18-0.28.

### 3.2. Analysing relative contributions of treatment, disease and cohort to overall taxonomic composition

To be able to quantify the relative effects if different factors affecting the gut microbiomes in the study, we measured Canberra distances between all T2D metformin+, T2D metformin- and ND CTRL samples. For each such pairwise distance we also determined whether the samples had the same treatment status (any drug vs no drugs), disease status (T2D vs control) or came from the same source dataset or not (and thereby same country or not). The distances were modeled using the R aov function, specifying a model encoding treatment, disease and country similarity separately and as pairwise and threeway interactions, as the predictor variables. The results are shown in Extended Data Table 1a. Broadly speaking, country of origin explains roughly 9% of intrasample

distances, treatment roughly 3%, and diabetes itself roughly 1%. The bulk of intrasample distances cannot be explained by any of these factors, but rather reflect individual variability or other uncontrolled-for factors.

### 3.3. Analysis of dietary data for a subset of the MetaHIT cohort

Diet is known to impact both diabetes development and the microbiome itself. We therefore investigated whether any clear trends can be seen between sample categories within the MHD cohort with regards to diet, such that they might mediate the microbiome signals we observe. Food frequency questionnaire data was possible to obtain for a subset of the MHD samples as described in the Methods, and was transformed into quantitative estimates of intake for a set of macronutrient and food categories. These estimates were compared between sample categories using general linear models, with results shown in Supplementary Table S16. No significant differences were found between T2D metformin+ and T2D metformin- samples. Comparing all T2D samples with ND CTRL samples reveal the T2D patients consume significantly less saccharose and milk products, but significantly more meat, where the first two likely correspond to compliance with dietary interventions in diabetes. It is conceivable that the increased meat consumption is associated with the observed decrease in butyrate-producing gut bacteria but no definite conclusions can be drawn from this very limited dataset. Likewise no conclusions can as yet be drawn on how much diet plays a role in the differences between the three source studies, in the absence of dietary data for non-MHD samples.

**4. Species-level breakdown of taxonomic correlates of T2D and metformin treatment**

The main manuscript discusses results achieved when grouping taxa together at the genus level. The underlying mOTU resource further allows for species-level resolution in many cases, though novel or poorly characterized mOTUs may have only broad taxonomic identity assigned and therefore may be difficult to interpret, which was why they were not reported in the main manuscript. We repeated the univariate tests for contrasting groups of samples (FDR-adjusted KWT tests blocked for study source, FDR-adjusted WRS tests post-hoc for comparing sample subsets) on a mOTU level as well. Results of these tests are shown in Supplementary Table S9.

For the comparison of T2D metformin- with ND CTRL samples, e.g. analysis of the T2D-associated microbiota in absence of metformin treatment, the genus-level signal is largely reflected at the mOTU level with representatives for each differentially abundant genus also seen significantly different in mOTU abundance. The *Lactobacillus* genus-level difference appears to be driven by *L. amylovorus* and *L. salivarius*, the *Haemophilus* difference by *H. parainfluenzae*, whereas mOTU-level resolution of the various butyrate producers include mOTUs for *Roseburia* and several poorly characterized groups. Beyond the signals also seen at the genus level, certain other taxa also emerge significantly different in abundance at the mOTU level, including strains of *Faecalibacterium prausnitzii*. For *I. bartlettii* as well as for eight uncharacterized mOTUs, there is a significant difference both when comparing T2D metformin- with T2D metformin+ samples, and comparing T2D metformin- with ND CTRL samples, in eight cases consistent with a reduction or inversion of the T2D pattern upon metformin treatment.

Comparing T2D metformin+ and T2D metformin- samples, the *Intestinibacter* and *Escherichia* signals stand out clearly, and are driven respectively by *I. bartlettii* and some closely related taxon,

and by *Escherichia coli*. Of remaining mOTUs affected by metformin, as stated many are uncharacterized at the genus level, and several of them show a significant and reverse trend when comparing T2D metformin+ with T2D metformin- samples. Our finding that microbial SCFA production in the small intestine may lead to improved glucose control may thus also involve some of these uncharacterized *Firmicutes*, most of whom are *Clostridiales*.

## 5. T1D control analysis

### 5.1. Characterization of the adult T1D microbiome

Notably a recent study of children at risk for T1D showed a reduction in microbial diversity just prior to onset[59], the present findings then suggest that it does not necessarily remain in adults under long-term blood glucose lowering through insulin and lifestyle interventions. The T1D samples were distinguishable with high ROC-AUC values from both 277 MHD healthy and 75 MHD T2D samples using an SVM classifier (Supplementary Table S6). With or without adjusting for community richness differences (using both a regression model and selective subsampling), T1D samples were enriched (Wilcoxon rank-sum test FDR < 0.1) for several taxa including *Methanobrevibacter* and *Ruminococcus* (Supplementary Table S11).

### 5.2. Microbiome–based classification of type 1 diabetes and type 2 diabetes from controls, the latter without stratification on metformin treatment

Type 1 diabetes (T1D) classification from controls performed poorly using bacterial family abundances as features and we achieved a ROC-AUC (Area Under (Receiver Operating) Curve metric) of around 60% using CAGs[34] as input data. We also evaluated bacterial genus and family level taxonomic abundance matrices as predictive features. Using family level community composition as input, predictors performed no better than random; using genus-level taxonomic data as input the classifier reached ~79% ROC-AUC including 13 genera as features (Supplementary Table S8). Listed in order of descending importance for the classifier, these were: *Methanobrevibacter*, *Odoribacter*, *Faecalibacterium*, *Oscillibacter*, *Ruminococcaceae*, *Dorea*, *Collinsella*, *Bifidobacterium*, *Erysipelotrichaceae*, *Coprococcus*, *Oscillospiraceae*, *Dialister*, *Sutterellaceae*. This set is largely overlapping with bacterial genera found significantly different between T1D and ND CTRL samples (Supplementary Table S10). On mOTU level, the prediction was similar: 77% ROC-AUC was achieved using 22 mOTUs as features and 75% using 8 mOTUs. Here, *Methanobrevibacter* OTU's were not important in the prediction (rank 97), but more so an unclassified *Faecalibacterium* (SpecI Cluster1577), a *Bacteroides* (SpecI Cluster1104) and a *Lachnospiraceae* (SpecI Cluster1608).

T2D classifications performed more poorly. A good separation was achieved between T2D metformin+ and T2D metformin- samples (ROC-AUC at 75-90% on the three datasets using *Intestinibacter* only), but not between ND CTRL and T2D metformin-, where the best classification with ROC-AUC at 64-84% across datasets was achieved by including 63 MGS (Supplementary Table S8).

## 6. Independent replication of metformin-associated gut microbial composition changes

To validate our finding of an association between metformin and particular changes in gut microbial composition, specifically an increase in *Escherichia* and a decrease in *Intestinibacter* carriage, in a fourth cohort (see Methods), we selected a set of 30 T2D patients from an ongoing study, 21 of which were treated with metformin, extracted DNA from frozen faecal samples as for the MHD samples, then carried out 16S amplicon sequencing taxonomic profiling (see Methods) and assigning a taxonomy against the SILVA 119 rRNA database[39]. Two OTUs that were classified as *Escherichia/Shigella* and unknown *Intestinibacter*, respectively, were then searched against the NCBI rRNA database[40] which is more complete with respect to *Intestinibacter* rRNA; we found >= 99% identity match over the whole sequence to *I. bartlettii* and *Escherichia/Shigella* 16S rRNA. Testing specifically for the two signals concluded that under metformin treatment, both a depletion of *Intestinibacter* and an enrichment of *Escherichia* are significant (MWU test, Extended Data Table 1b, Extended Data Figure 1c), though *Intestinibacter* was at the detection threshold, which could suggest a poor primer amplification.

## 7. A potential role of *Escherichia* enrichment in metformin adverse effects

Several functional modules, both under SEED[60] or the GMM system, were significantly enriched for in T2D metformin+ compared to T2D metformin- samples, and the increased abundance of these gene modules could largely be traced to higher *Escherichia* abundance. This was indirectly demonstrated by removing all contributions to the functional profiles from putative *Enterobacteriaceae* genes, which eliminated significant differences between groups (Supplementary Table S15). Thus, the observed symptoms of gastrointestinal distress may derive from the

*Escherichia*-associated increase in lipopolysaccharide synthesis potential[61], or the corresponding increase for 14 SEED virulence subterms (including virulence regulators, adhesion and biofilm formation genes, antibiotic resistance genes and fimbriae), as well as from enhanced sulfate metabolism potential (Supplementary Table S14). Increased genetic potential for hydrogen production will not only contribute to further enhancement of hydrogen sulfide production, but also to bloating, a clinically well-known side effect of metformin treatment[62]. Further, intestinal bile acid absorption is reduced in metformin-treated patients; this might cause an increased availability of sulfate from conjugated bile steroids, leading to increased sulfate reduction[63]. Sulphur metabolism in itself is associated with intestinal bloating and discomfort[64].

## Supplementary table legends

**Supplementary Table S1 –** Samples used in study. “Sample subsets” tab lists diabetes and metformin treatment status for all 784 samples where this is available. “Phenotypes” tab lists body mass index, fasting plasma glucose, fasting serum insulin, and Hba1c for all samples in the MHD, SWE and CHN datasets. To convert mU/l to pmol/l for insulin levels a factor of 6.0 was used. To convert HbA1c levels from % (DCTT) to mmol/mol (IFCC) the following formula was applied ((HbA1c % - 2.15)*10.929). “Sequencing statistics” tabs show metagenome sample sizes and measurements, including QC measures, for all samples in the three study sets.

**Supplementary Table S2 –** Significant differences in taxonomic composition (genus and phylum level, KWT) and gene richness indices (KWT) for T2D patients between the three geographically separated datasets used in the study. Question mark notation in a taxon description indicates a novel taxon at that hierarchical level, e.g. “Bacteria; Firmicutes;?;?;?;?” indicates a currently uncharacterized *Firmicutes*. A post-hoc test was used to determine between-subset significances, where “=” refers to n.s. differences and “>”, “>>” and “>>>” refer to p-values < 0.05, 0.01 and 0.001, respectively. the “Metadata” tab summarizes some metadata differences between the three country datasets (data are shown as median and interquartile range).

**Supplementary Table S3** – Significant (WRS P-values, and BH FDR-corrected Q-values blocked for country, taking significance at FDR < 0.1, these cells are shaded in yellow) differences in taxonomic (genus, family and mOTU level) and functional (GMMs and KEGG modules) composition between T2D (n=199) and ND CTRL (n=554) samples disregarding medication status. These results are compared in subsequent sub-tables with those reported in the source studies (with

red/green shading noting enrichment/depletion in T2D), either using only each set of source study samples separately or the full set of samples, with tests blocked for study source.

**Supplementary Table S4** – Investigating the influence of medications including statin, insulin and blood pressure lowering drugs on gut microbial genus abundance, tested for T2D samples. No significant differences (WRS P-values, and BH FDR-corrected Q-values blocked by source country, taking significance at FDR < 0.1) were found between treated and untreated T2D patients. Blocking also for metformin treatment yielded no further significant differences. The power calculations sub-table shows post-hoc calculations of effective statistical power of a WRS test in detecting taxonomic changes with effect sizes comparable to those found under metformin treatment, at a type I error rate of 0.05 and with sample sizes instead equivalent to those of the other medications tracked.

**Supplementary Table S5 –** Multivariate significant differences in taxonomic composition between sample subsets. FDR values shown were determined using a betadisper test between T2D metformin- (n=106), T2D metformin+ (n=93) and ND CTRL (n=554) samples with $10^4$ repetitions, blocked for country effect on different taxonomic levels, using Canberra distances.

**Supplementary Table S6 –** Performance of robust recursive feature elimination support vector machine (rRFE-SVM) predictors built to separate sample subsets based on diabetes and metformin treatment status. The table shows cross-validated ROC-AUC classifier performance using a selected feature subset for their separation of A) T2D metformin- (n=106), T2D metformin+ (n=93) or ND CTRL (n=554) states and B) MHD T1D (n=31) from MHD ND CTRL (n=277) samples.

**Supplementary Table S7 –** Performance of robust recursive feature elimination support vector machine (rRFE-SVM) predictors built to separate sample subsets based on diabetes status ignoring

metformin treatment status. The table shows ROC-AUC classifier performance (%) when training the classifier on each country data subset, then testing the resulting classifier on the same or on each of the two other countries. Number of features used by each classifier is also shown, as is the set of features used.

**Supplementary Table S8 –** Bacterial genera tested for significantly (study source adjusted WRS P-values, and BH FDR-corrected Q-values, taking significance at FDR < 0.1, these cells are shaded in yellow) different abundance between T2D metformin- (n=106) and ND CTRL (n=554) samples. All tests were corrected for study source. Question mark notation in a taxon description indicates a novel taxon at that hierarchical level, e.g. “Bacteria; Firmicutes;?;?;?;?” indicates a currently uncharacterized *Firmicutes*.

**Supplementary Table S9 –** Bacterial species (mOTU taxonomic entities) tested for significantly (study source adjusted WRS P-values, and BH FDR-corrected Q-values, taking significance at FDR < 0.1, these cells are shaded in yellow) different abundance between T2D metformin- (n=106) and ND CTRL (n=554) samples, and between T2D metformin+ (n=93) and T2D metformin- (n=106) samples. All tests were corrected for study source.

**Supplementary Table S10** – Bacterial functional modules (GMM and SEED modules, respectively) tested for significantly (study source adjusted WRS P-values, and BH FDR-corrected Q-values, taking significance at FDR < 0.1, these cells are shaded in yellow) different abundance between T2D metformin- (n=106) and ND CTRL (n=554) samples. All tests were corrected for study source.

**Supplementary Table S11** – Bacterial genera that were significantly (WRS P-values, and BH FDR-corrected Q-values, taking significance at FDR < 0.1) different in abundance between T1D

(n=31) and MHD ND CTRL (n=277) samples. Question mark notation in a taxon description indicates a novel taxon at that hierarchical level, e.g. "*Bacteria; Firmicutes;?;?;?;?*" indicates a currently uncharacterized *Firmicutes*.

**Supplementary Table S12 –** Clinical markers and taxonomic diversity indices tested for significant differences between sample sets based on metformin treatment or diabetes status. Shown are median values and standard deviations for subsets under comparison, WRS P-values, and BH FDR-corrected Q-values, taking significance at FDR < 0.1.

**Supplementary Table S13** – Two bacterial genera which were significantly different between T2D metformin+ (n=106) and T2D metformin- (n=93) samples (*Escherichia* and *Intestinibacter*) were retested with a WRS-test (taking significance at FDR < 0.1) correcting for the following potential confounders: BMI, gender, fasting plasma glucose and fasting serum insulin. All tests were corrected for study source.

**Supplementary Table S14** – Bacterial functional modules (GMM and SEED modules, respectively) tested for significantly (study source adjusted WRS P-values, and BH FDR-corrected Q-values, taking significance at FDR < 0.1, these cells are shaded in yellow) different abundance between T2D metformin+ (n=93) and T2D metformin- (n=106) samples. All tests were corrected for study source.

**Supplementary Table S15** – Bacterial functional modules (GMM and SEED modules, respectively) tested for significantly (study source adjusted WRS P-values, and BH FDR-corrected Q-values, taking significance at FDR < 0.1) different abundance between T2D metformin+ (n=93) and T2D metformin- (n=106) samples while excluding contributions from *Enterobacteriaceae* (see Methods). All tests were adjusted for study source.

**Supplementary Table S16** – Analysis of dietary differences in Danish MHD individuals between T2D metformin- (n=16), T2D metformin+ (n=50) samples, and between T2D (either metformin status, n=66) and ND CTRL (n=194) samples; Beta-values are estimated using a general linear model also incorporating age, sex and energy intake. Fat, carbohydrate and protein energy intake tests were adjusted for age and sex only. The following dietary components were log-transformed in order to approximate a normal distribution: Energy intake, total fat intake, carbohydrate intake, saccharose, milk, cheese, vegetables, potatoes, poultry, egg and protein-rich food.

**Supplementary Discussion references**


54 Salonen A, et al. Comparative analysis of fecal DNA extraction methods with phylogenetic microarray: effective recovery of bacterial and archaeal DNA using mechanical cell lysis. *J Microbiol Methods.* **81**:127-34. (2010)

55 Li F et al. Optimization of terminal restriction fragment polymorphism (TRFLP) analysis of human gut microbiota. *J Microbiol Methods*. **68**:303-11. (2007)

56 Zoetendal EG, et al. Temperature gradient gel electrophoresis analysis of 16S rRNA from human fecal samples reveals stable and host-specific communities of active bacteria. *Appl Environ Microbiol*. **64**:3854-9. (1998)

57 Forslund K., et al. Metagenomic insights into the human gut resistome and the forces that shape it. *Bioessays*. **36:**316-29. (2014)

58 Faul F, Erdfelder E, Lang AG, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav Res Methods*. **39**:175-91. (2007)

59 Kostic AD, et al. The dynamics of the human infant gut microbiome in development and in progression toward type 1 diabetes. *Cell Host Microbe*.**17**:260-73. (2015)

60 Overbeek, R. *et al.* The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic acids research* **42**, D206-214, doi:10.1093/nar/gkt1226 (2014).

61 Rhee, S. H. Lipopolysaccharide: basic biochemistry, intracellular signaling, and physiological impacts in the gut. *Intestinal research* **12**, 90-95, doi:10.5217/ir.2014.12.2.90 (2014).

62 Florez, H. *et al.* Impact of lifestyle intervention and metformin on health-related quality of life: the diabetes prevention program randomized trial. *Journal of general internal medicine* **27**, 1594-1601, doi:10.1007/s11606-012-2122-5 (2012).

63 Van Eldere, J., Robben, J., De Pauw, G., Merckx, R. & Eyssen, H. Isolation and identification of intestinal steroid-desulfating bacteria from rats and humans. *Applied and environmental microbiology* **54**, 2112-2117 (1988).

64 Manichanh, C. *et al.* Anal gas evacuation and colonic microbiota in patients with flatulence: effect of diet. *Gut* **63**, 401-408, doi:10.1136/gutjnl-2012-303013 (2014).

## Additional MetaHIT consortium members:

Mathieu Almeida[1,1a], Maria Antolin[2], François Artiguenave[3,4,5], Jean-Michel Batto[1,1a], Marcelo Bertalan[6], Hervé Blottière[1,1a], Natalia Boruel[2], Christian Brechot[7], Thomas Bruls[3,4,5], Kristoffer Burgdorf[3], Francesc Casellas[2], Antonella Cultrone[1a], Willem M. de Vos[9,10], Christine Delorme[1a], Gérard Denariaz[11], Muriel Derrien[11], Rozenn Dervyn[1a], Qiang Feng[12], Niels Grarup[8], Francisco Guarner[2], Eric Guedon[1a], Florence Haimet[1,1a], Alexandre Jamet[1a], Agnieska Juncker[6], Catherine Juste[1a], Sean Kennedy[1,1a], Ghalia Khaci[1a], Michiel Kleerebezem[9], Jan Knoll[13], Séverine Layec[1a], Marion Leclerc[1a], Pierre Leonard[1,1a], Denis LePaslier[3,4,5], Christine M'Rini[7], Emmanuelle Maguin[1a], Chaysavanh Manichanh[2], Daniel Mende[14,15], Alexandre Mérieux[7], Raish Oozeer[13], Julian Parkhill[16], Eric Pelletier[3,4,5], Nicolas Pons[1,1a], Junjie Qin[12], Simon Rasmussen[6], Pierre Renault[1a], Maria Rescigno[17], Nicolas Sanchez[1a], Thomas Sicheritz-Ponten[6], Julien Tap[11], Sebastian Tims[9], Antonio Torrejon[2], Keith Turner[16], Maarten van de Guchte[1a], Johan E.T. van Hylckama Vlieg[11], Gaetana Vandemeulebrouck[1a], Encarna Varela[2], Patrick Viega[11], Jean Weissenbach[3,4,5], Yohanan Winogradski[1a], Takuji Yamada[14], Erwin G. Zoetendal[9]

1 Institut National de la Recherche Agronomique, US1367 Metagenopolis

1a Institut National de la Recherche Agronomique, UMR14121 MICALIS

2 Digestive System Research Unit, University Hospital Vall d'Hebron, Ciberehd, 08035 Barcelona, Spain

3 Commissariat à l'Energie Atomique, Genoscope, 91000 Evry, France

4 Centre National de la Recherche Scientifique, UMR8030, 91000

5 Evry, France, Université d'Evry Val d'Essone 91000 Evry, France

6 Center for Biological Sequence Analysis & Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, DK-2800 Kongens Lyngby, Denmark

7 Institut Mérieux,17 rue Burgelat,69002 Lyon, France.

8 The Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, DK-2200 Copenhagen, Denmark

9 Laboratory of Microbiology, Wageningen University, The Netherlands

10 Department of Bacteriology and Immunology, University of Helsinki, FIN-00014 Finland

11 Danone Research, 91120 Palaiseau, France

12 BGI-Shenzhen, Shenzhen517083, China

13 Gut Biology & Microbiology, Danone Research, Centre for specialized nutrition, Wageningen, The Netherlands

14 European Molecular Biology Laboratory, Meyerhofstrasse1,69117 Heidelberg, Germany

15 Daniel K Inouye Center for Microbial Oceanography Research and Education (C-MORE), University of Hawai'i at Manoa, Honolulu, HI, USA

16 The Wellcome Trust Sanger Institute, Hinxton, Cambridge CB101SA, UK

17 IEO - Istituto Europeo di Oncologia, Via Giuseppe Ripamonti, 435, 20141 Milano, Italy