Connections between human gut microbiome and gestational diabetes mellitus

Ya-Shu Kuang[1, †], Jin-Hua Lu[1,2, †], Sheng-Hui Li[1, †], Jun-Hua Li[3,4], Ming-Yang Yuan[1,2], Jian-Rong He[1,2], Nian-Nian Chen[1,2], Wan-Qing Xiao[1,2], Song-Ying Shen[1,2], Lan Qiu[1,2], Ying-Fang Wu[1,2], Cui-Yue Hu[1,2], Yan-Yan Wu[1,2], Wei-Dong Li[1,2], Qiao-Zhu Chen[5], Hong-Wen Deng[1,6], Christopher J Papasian[7], Hui-Min Xia[1,8*], Xiu Qiu[1,2 *]

[1]Division of Birth Cohort Study, Guangzhou Women and Children's Medical Center, Guangzhou Medical University, Guangzhou 510623, China
[2]Department of Women and Children's Health Care, Guangzhou Women and Children's Medical Center, Guangzhou Medical University, Guangzhou 510623, China
[3]BGI-Shenzhen, China National GeneBank-Shenzhen, Shenzhen 518083, China
[4]Shenzhen Key Laboratory of Human commensal microorganisms and Health Research, BGI-Shenzhen, Shenzhen 518083, China
[5]Department of Obstetrics and Gynecology, Guangzhou Women and Children's Medical Center, Guangzhou Medical University, Guangzhou 510623, China
[6]Center of Bioinformatics and Genomics, Department of Biostatistics and Bioinformatics, Tulane School of Public Health and Tropic Medicine, USA
[7]Department of Basic Medical Science, School of Medicine, University of Missouri – Kansas City, 2411 Holmes St., Kansas City, MO 64108
[8]Department of Neonatal Surgery, Guangzhou Women and Children's Medical Center, Guangzhou Medical University, Guangzhou 510623, China

E-Mails:
kuangyashu@126.com (YSK); lujinhua10@gmail.com (JHL*); lishenghui1005@gmail.com (SHL); lijunhua@genomics.cn (JHL); ymyeasy@126.com (MYY); hjr0703@163.com (JRH); cnnfezx@163.com (NNC); 15915711316@163.com (WQX); shsy_22@163.com (SYS); 38904442@qq.com (LQ); mywachy@163.com (YFW); bighcy@163.com (CYH); 153050569@qq.com (YYW); liweidong30303@163.com (WDL); bigcqz@163.com (QZC); hdeng2@tulane.edu (HWD); PapasianC@umkc.edu (CJP); huimin.xia876001@gmail.com (HMX); qxiu0161@163.com (XQ)

[†] The authors contributed equally to this work.

[*] Correspondence to:
Xiu Qiu, Division of Birth Cohort Study, Guangzhou Women and Children's Medical Center, Guangzhou Medical University, 9 Jinsui Road, Guangzhou 510623, China; Phone: 86 2038367162; Fax: 86 2038367162; E-mail: qxiu0161@163.com; xiu.qiu@bigcs.org.
Hui-Min Xia, Division of Birth Cohort Study, Guangzhou Women and Children's Medical Center, Guangzhou Medical University, 9 Jinsui Road, Guangzhou 510623, China; Phone: 86 2038076019; Fax: 86 2038076020; E-mail: huimin.xia876001@gmail.com;huimin.xia@bigcs.org.

Abstract

Background

The human gut microbiome can modulate metabolic health and affect insulin resistance, and may play an important role in the etiology of gestational diabetes mellitus (GDM). Here, we compared the gut microbial composition of 43 GDM patients and 81 healthy pregnant women via whole-metagenome shotgun sequencing of their fecal samples collected at 21-29 weeks, to explore associations between GDM and the composition of microbial taxonomic units and functional genes.

Results

Metagenome-wide association study (MGWAS) identified 154,837 genes, which clustered into 129 metagenome linkage groups (MLGs) for species description, with significant relative abundance differences between the two cohorts. *Parabacteroides distasonis*, *Klebsiella variicola*, etc., were enriched in GDM patients, whereas *Methanobrevibacter smithii*, *Alistipes* spp., *Bifidobacterium* spp. and *Eubacterium* spp. were enriched in controls. The ratios of the gross abundances of GDM-enriched MLGs to control-enriched MLGs were positively correlated with blood glucose levels. Random Forest model shows fecal MLGs have excellent discriminatory power to predict GDM status.

Conclusions

Our study discovered novel relationships between gut microbiome and GDM status, and suggested that changes in microbial composition may potentially be used to identify individuals at risk for GDM.

Background

The increasing prevalence of gestational diabetes mellitus (GDM), and its subsequent health outcomes, are a significant public health concern and a major challenge for obstetric practice [1]. GDM represents a heterogeneous group of metabolic disorders [2] which affects 3-14% of pregnancies, and 20-50% of these affected women are expected to develop type 2 diabetes（T2D） within 5 years [3, 4]. Emerging evidence has revealed a link between the gut microbiome and human metabolic health including T2D [5, 6], leading us to hypothesize that the gut microbiome may impact gestational metabolism and development of GDM.

Microbial dysbiosis in the human gut may be an important environmental risk factor for abnormal host metabolism, as recently exemplified in studies of obesity and T2D (reviewed by Karlsson, et. al) [7]. A study using an experimental animal model revealed that reduced numbers of *Bifidobacteria* led to enhanced endogenous lipopolysaccharide production, endotoxemia, and associated obesity and insulin resistance [8]. In humans, excessive weight gain and obesity in pregnancy resulted in deteriorated glucose tolerance and increased risk of GDM [9, 10]. *Prevotella copri* and *Bacteroides vulgatus* have been identified as the main species driving the association between biosynthesis of branched-chain amino acids, insulin resistance, and glucose intolerance [11]. *Bacteroides* spp. and *Staphylococcus aureus* are significantly more abundant in overweight women than in normal-weight women [12].

While the majority of previous studies have focused on associations between intestinal microbiota and obese states or T2D [6, 13-15], some recent studies have sought to characterize microbiota changes during pregnancy, with the goal of providing novel insights into the relationship between microbiota changes during pregnancy and potential metabolic consequences [16]. Studies based on sequencing of 16S ribosomal RNA have revealed novel relationships between gut microbiome composition and the metabolic hormonal environment in overweight and obese pregnant women in early gestation [17]. Koren et al. found that maternal gut microbiota changed from first to third trimesters, with a decline in butyrate-producing bacteria and increased *Bifidobacteria, Proteobacteria,* and lactic-acid producing bacteria [16]. Further, transplants of fecal material obtained during different trimesters were sufficient to confer different phenotypes in mouse models, with third-trimester fecal transplants leading to increased adiposity and

inflammation [16]. These studies suggest that pregnancy is associated with major shifts in the gut microbiome which may play an important role in observed increases in gestational inflammation, thereby potentially contributing to development of GDM. However, studies focusing on changes in the gut microbiome during pregnancy and development of GDM have not been reported so far.

Metagenomic shotgun sequencing, in which the full complement of genes present in the microbiome are sequenced, can furnish information about the relative abundance of genes in functional pathways and at all taxonomical levels [18]. In this study, we used whole-metagenome shotgun sequencing analyses of the gut microbiome during pregnancy to explore associations between GDM and the composition and abundance of microbial taxonomic units and functional genes. The objective was to obtain a comprehensive understanding of the connections between gut microbiome and the development of GDM.

Data description

Whole-metagenome shotgun sequencing was used to test gut microbial composition in fecal samples from 43 GDM patients and 81 healthy pregnant women based on the Illumina HiSeq2000 platform in BGI-Shenzhen, China. We constructed a paired-end library with insert size of 350 base pairs (bp) for every sample, and sequenced with 100bp read length from each end. Sequencing reads for fecal samples were independently processed for quality control and host sequences removal based on an in-house pipeline (see Methods), and a total of 795 Gbp high quality metagenomic data (average per sample, 6.4 Gbp) were generated for further analysis. We performed *de novo* assembly and gene calling for data of each sample and constructed a non-redundant gene catalogue of all pregnant women fecal samples containing 4,344,984 genes. This gene catalogue provided a suitable reference for metagenomic gene quantification, microbial diversity analysis, and metagenome-wide association study for the pregnant women fecal samples.

Results

Comparison of the gut microbiota between GDM patients and healthy pregnant women

First, we explored potential differences in the gut microbiome between 43 GDM patients and 81 healthy pregnant women. We obtained 795.3 Gb of high-quality data (6.4 ± 1.3 Gb per sample) via metagenomic shotgun sequencing of their fecal samples to perform this analysis. When we

quantified the microbial (alpha) diversity within each subject, the GDM patients showed significantly lower gene count and Shannon index compared with the healthy pregnant women ($P < 0.05$ for both indexes, Mann-Whitney U test).We then aligned the sequencing reads (43.8%) against available microbial genomes from the National Center for Biotechnology Information and generated taxonomic composition for all samples at the taxonomic levels of phylum, class, order, family, genus and species. Multivariate analysis based on Bray–Curtis distances between microbial genera revealed significant differences between GDM patients and healthy controls (Figure 1a). We then performed the Mann–Whitney U test to identify phylogenetic differences between GDM patients and healthy controls. Abundance at the phylum and class levels was similar between GDM patients and healthy controls; however, the order *Clostridiales* and the family *Coriobacteriaceae* were enriched in healthy controls. At the genus level, GDM patients had a significantly higher abundance of *Parabacteroides*, *Megamonas* and *Phascolarctobacterium*, while healthy controls were significantly enriched for *Ruminiclostridium*, *Roseburia*, *Eggerthella*, *Fusobacterium*, *Haemophilus*, *Mitsukella*, and *Aggregatibacter* (Figure 1b). We also found a number of bacterial species that differed significantly between GDM patients and healthy controls, consistent with the genus level observations (Table S2). These findings suggest dysbiosis of the gut microbiota among GDM patients.

Identification of GDM-associated markers from gut microbiome

To explore detailed signatures of the gut microbiome in GDM patients and heathy controls, we constructed a non-redundant gene catalogue consisting of 4.34 million genes, which allowed an average reads mapping rate of 79.5% for sequenced samples. We identified 154,837 genes that displayed significant abundance differences between the two groups (Mann-Whitney U test, $q<0.05$) (Figure S1 shows the P-value distribution between GDM patients and healthy pregnant women for all genes tested). About 68% of these genes were clustered into 129 metagenomic linkage groups (MLGs) (Table S3), which allowed species level description for the microbiome differences. The 71 MLGs enriched in GDM patients included *Parabacteroides distasonis*, *Klebsiella variicola*, *Catenibacterium mitsuokai*, *Coprococcus comes* and *Citrobacter spp.*, whereas the 58 MLGs enriched in healthy pregnant women included *Methanobrevibacter smithii*, *Alistipes spp.* (*A. shahii*, *A. senegalensis*), *Bifidobacterium spp.* (*B. animalis*, *B.*

pseudocatenulatum) and *Eubacterium spp.* (*E. siraeum*, *E. eligens*). The GDM-enriched and control enriched MLGs were highly positively interconnected within each group; however, only few negative connections were found between the two groups (Figure 2). Notably, GDM-enriched MLGs of *Enterobacteriaceae*, including *K. variicola*, *E. coli*, *Enterobacter cloacae* and *Citrobacter spp.*, were closely linked (correlation coefficients >0.40 between each other), representing a cooperative promoting function of *Enterobacteriaceae* to GDM development. Of particular interest, we also observed that the relative abundance of *Enterobacteriaceae* was positively associated with pre-pregnancy body mass index (PBMI, Figure S2).

Correlations between maternal blood glucose levels and gut microbiota

In order to explore the potential clinical paths by which changes in the microbiome might lead to GDM, we investigated whether the MLGs can affect blood glucose tolerance. The ratios of the gross abundances of GDM-enriched MLGs to those of control-enriched MLGs were obviously positively correlated with blood glucose levels during the second trimester of pregnancy (Figure 3), indicating that dysbiosis of the microbiome has a significant relationship with GDM status. Several GDM-enriched MLGs [e.g. GDM67, GDM64, *P. distasonis* (GDM1), *K. variicola* (GMD41) and *E. rectale* (GDM34)] were positively correlated with blood glucose levels, while most control-enriched MLGs were negatively correlated with blood glucose levels (Figure 4a). At the species level, *Eggerthella spp.*, *Megamonas spp.*, *Allofustis seminis* and several species in *Lachnospiraceae* and *Parabacteroides* were positively correlated with glucose tolerance, while several *Alistipes spp.* were negatively correlated with glucose tolerance (Figure 4b).

Functional characterization of gut microbiota in GDM

Next, we utilized KEGG pathway comparisons to explore potential differences in the functional composition of the microbiome of GDM patients vs. controls. Although the functional composition of GDM patients and controls were highly similar (Figure 5a), the microbiome of GDM patients showed a greater abundance in pathways of membrane transport and energy metabolism, while the microbiome of controls had higher abundance in amino acid metabolic pathways. We also found that the KEGG modules, including the phosphotransferase system (PTS) and lipopolysaccharide (LPS) biosynthesis and export systems, were associated with glucose

tolerance levels (Figure 5b).

Gut microbiota-based prediction of GDM

Finally, we utilized random forest models to assess the predictive ability of MLGs and species abundance profiles for GDM status. We found that certain 20 MLGs provided the best discriminatory power, as indicated by the area under the ROC curve (AUC) 0.91 (95% CI 0.87-0.96), which was higher than that achieved using species profiles with this model (the best AUC was 0.80; 95% CI 0.73-0.86) using 40 species (Figure 6a). The increased AUC for the MLG-based model may be due to the fact that MLGs furnish taxonomic and functional information for unknown or unanalyzable species. Bacterial species providing the highest discriminatory power were primarily members of the *Bacteroides* or *Parabacteroides* genera (Figure 6b-c), consistent with our observation that *Parabacteroides* is the predominant genus accounting for differences in the gut microbiome between GDM patients and controls (Figure 1b). Although PBMI is a predictor of GDM, it did not substantially improve the performance of MLGs. (Figure 6d and Figure S3).

Potential Implications

The gut microbiome can be considered both as an endocrine and metabolic organ, the dysfunction of which plays important roles in disease development. During gestation, profound hormonal, immunological and metabolic changes take place [19-21]. Our findings suggest that gut microbiota in pregnant women are sensitive to subtle changes in metabolism and increases in blood glucose levels. When taken together with results from previous studies on T2D [22], our findings suggest gut microbiota may be a potential predictor of T2D after pregnancy. Furthermore, data from our cohort indicate that women diagnosed with GDM also suffered from moderate gut bacterial dysbiosis and functional dysbiosis that was not restricted to certain microbial species. Although causality has not been demonstrated, it raises the possibility that susceptibility of postpartum metabolic (e.g. T2D) and immune dysfunction might be modified by reconditioning of gut microbiota. Given that the gut microflora can be modified by diet, altering the composition of gut microbiota in pregnant women may improve diabetes related outcomes. Future studies should explore how gut bacterial dysbiosis could be improved and evaluate the efficacy of potential

Commented [NN1]: Please add this section:
Authors should provide some additional comments about potential, more broad-ranging implications of their work that are not directly related to the current focus of their manuscript. This section is meant to promote discussion on possible ways the findings or data presented might be used in or have a relationship with other areas of research that may not be directly apparent in the work. It is not meant to provide 'proof of importance' of the work. Only to engender expansion of use to other areas.

Explicit personal opinions by the authors are permitted, but they should be made clear as such. References or related information to support the propositions should be included. These section should focus on work that can be done within the foreseeable future and specifically using the information within the manuscript, not provide speculation on how it will relate to far-reaching goals of the research area.

Reply： done as requested

interventions, such as probiotics and dietary manipulation among pregnant women.

Discussion

In the present metagenomics study, we observed associations between gut microbiome and GDM status. Specifically, *Parabacteroides distasonis*, *Klebsiella variicola*, etc. were enriched in GDM patients, whereas *Methanobrevibacter smithii*, *Alistipes* spp., *Bifidobacterium* spp. and *Eubacterium* spp. were enriched in controls. The distribution of MLGs in GDM patients differed from that in the control group. Functional analysis showed a greater abundance of membrane transport, energy metabolism pathways, lipopolysaccharide and phosphotransferase systems in the microbiome of GDM patients, while the microbiome of controls was enriched in the amino acid metabolic pathways (Figure 7). To our knowledge, this is the first metagenomics study exploring roles of microbiota in the development of GDM.

Previous studies have shown the GDM-enriched bacteria that observed in our study are involved in gut flora dysbiosis. For example, GDM-enriched *Bacteroides spp.* and *Parabacteroides distasonis* are considered to be opportunistic pathogens in infectious diseases, with potential for developing antimicrobial drug resistance [23]. The family *Enterobacteriaceae* also occurred with a higher relative abundance in GDM patients than in healthy controls, which indicates a status of gut flora dysbiosis that may lead to a series of chronic diseases, such as colitis [24], Crohn's disease and acute cholecystitis [25]. Previous studies have shown that *Enterobacteriaceae* instigate inflammation to induce colitis [24], and the endotoxin–producing bacterium *Enterobacter* contributed to the development of obesity in gnotobiotic mice [26].

The decreased microbes in GDM patients included *Bifidobacterium* spp. (including *B. pseudocatenulatum*, *B. animalis* and one unclassified MLG), *Eubacterium* spp. (*E. siraeum*, *E. eligens* and two unclassified *Eubacterium* MLGs) and *Roseburia* spp. (Tables S2 and S3). Similar findings were reported in previous studies on a variety of chronic diseases, including T2D [22], liver cirrhosis [27], Crohn's disease [28] and ulcerative colitis [29]. These bacteria can produce lactate or butyrate, which could regulate gut permeability and induce the gut inflammatory response that precedes the development of diabetes [30, 31].

Our data demonstrated the ratio of gross abundances of the GDM-enriched to control-enriched MLGs was positively correlated with blood glucose tolerance levels, suggesting that microbiome

dysbiosis might have a direct association with GDM pathophysiology. Functional analysis showed that the LPS biosynthesis and export systems were involved in regulation of glucose levels. Previous studies have shown that the higher systemic LPS levels were associated with low-grade chronic inflammation in obesity, metabolic syndrome and T2D [8, 32, 33]. Based on current knowledge, the possible pathways linking LPS levels to glucose metabolism may include the increases in intestinal permeability, the changes in the relative amounts of gram negative vs. gram positive bacteria and a low-grade chronic inflammatory state. LPS is a bacterial cell wall component in gram-negative bacteria and can stimulate an inflammatory response [34, 35]. Gut microbiome dysbiosis can facilitate LPS entry into the systemic circulation through increasing gut permeability, which leads to inflammation and metabolic dysfunction [36]. Our results were concordant with a previous report [22] which found that gut microbiota dysbiosis in T2D was characterized by a decrease in gram-positive butyrate producing *Clostridium* species that lack LPS and an increase in gram-negative opportunistic pathogens including some *Bacteroidetes* and *Proteobacteria* species that contain LPS. The functional analysis in the present study found that membrane transport, energy metabolic and PTS pathways were enriched in the GDM patients. PTS pathways are responsible for transporting glucose through outer and inner membranes and catalyzing the uptake of carbohydrates. The increased relative abundance of these pathways may indicate gut environment of a GDM status may stimulate bacterial accelerated usage of glucose as energy.

There were several limitations in our study. First, the sample size is relatively small. Second, we only analyzed one stool sample per participant, which was collected in the second trimester of pregnancy. It is well known that immune and metabolic changes occur throughout pregnancy, and that the gut microbiota shifts from first to third trimesters [16]. In the present study, we are unable to clarify the causal relationship between the microbiome and the development of GDM due to the cross-sectional design. Consequently, data at multiple time points are needed to provide further insights into their dynamic relationship. Third, we did not have information on several factors such as life style and diet may further affect both blood glucose levels and gut microbiota composition. In order to more confirm the associations observed in the current study, a large prospective cohort investigation, with analysis of other potentially significant variables, will be necessary. Besides, due to the lack of serum samples, we could not measure LPS levels and

describe the real endotoxemia level of the patients.

In summary, this is the first study to demonstrate an association between the gut microbiota dysbiosis, functional changes and GDM. Our findings contribute to the understanding of GDM pathophysiology and may have important implications for identifying patients at risk for development of GDM.

Methods

Study population and sampling

As part of the Born in Guangzhou Cohort Study (BIGCS) [37], fecal samples were obtained from 298 pregnant women during their second trimester in Guangzhou Women and Children's Medical Center (GWCMC) between 1st August, 2012 and 31st Aug, 2013. The inclusion criteria of current study were as follows: 1) without diseases which might affect glucose metabolism or microbiome composition such as pre-pregnancy diabetes, hypertension, thyroid disorders, asthma, lipid metabolic disorders, inflammatory bowel disease, irritable bowel syndrome and celiac disease; 2) had not received any antibiotic treatment 1 month before sample collection; 3) had not taken probiotics 2 weeks before sample collection. Of the 287 eligible women, 43 had a diagnosis of GDM and were included in the present study as the case group, and 81 women of non-GDM were randomly selected as the control group. Basic characteristics of the 124 pregnant women included in the study are summarized in Table S1. Compared to non GDM women, women with GDM were more likely to be older and multiparous and have higher pre-pregnant weight, pre-pregnancy body mass index (BMI) , gestational weight gain during pregnancy and premature delivery incidence. Fecal samples were frozen at -20°C freezers immediately (within 30 minutes) and transferred to -80 °C freezers within 24 hours after collected.

This study received approval from the Ethics Committee of GWCMC, and written informed consent was obtained from all participating pregnant women. Participants underwent a standard 2h 75g oral glucose tolerance test (OGTT) between 21–29 weeks' gestation by collection of 2ml blood samples fasting, 1h, and 2h after a 75g glucose load, using NaF/EDTA tubes. After centrifugation, plasma glucose was measured by a hexokinase method using Beckman Coulter AU5800 automatic analyzer (Beckman Coulter, California, US). The laboratory previously

achieved ISO15189 certification by China National Accreditation Service for Conformity Assessment. GDM was defined using the Chinese diagnostic criteria [38], which is in agreement with the one-step approach endorsed by the American Diabetes Association [39]. Pregnant women were diagnosed as having GDM if one or more of the following glucose levels were elevated: fasting ≥5.1 mmol/L, 1h ≥10.0 mmol/L, and 2h ≥8.5 mmol/L [38]. None of these women was treated with insulin or glyburide. Maternal age, pre-pregnancy weight and height were extracted from clinical records of the Hospital Information Systems (HIS) used in GWCMC. Pre-pregnancy body mass index (PBMI) was calculated from height and weight information.

DNA extraction and metagenomic sequencing

Total bacterial DNA was extracted from about 180-200 mg of feces using Qiagen QIAamp DNA Stool Mini Kit (Qiagen) following the manufacturer's instructions [40]. Extracted DNA of each sample was kept frozen at −20°C until used. Illumina HiSeq 2000 was used to sequence the samples. We constructed a paired-end library with insert size of 350 base pairs (bp) for every sample, and sequenced with 100bp read length from each end. Illumina sequencing reads for fecal samples from pregnant women were independently processed for quality control using FASTAX Toolkit[41]. The following criteria were used for quality control: (1) reads were removed if they contain more than 3 N bases or more than 50 bases with low quality (<Q20); (2) reads were trimmed in the end with low quality (<Q20) or assigned as N. The remaining reads were then mapped to the human genome using SOAPalinger2 (SOAPaligner/soap2 , RRID:SCR_005503) [42] to remove contaminating human DNA. After QC, an average of 1.9% of low-quality or human genome reads were removed for the 124 samples.

***De novo* assembly, gene calling and gene catalogue construction**

To determine the best assembling method for the obtained high-quality Illumina sequencing reads, we compared the performance of two assemblers, SOAPdenovo v2.04 (SOAPdenovo2 , RRID:SCR_014986) (as previously used in the MetaHIT and IGC projects) [43, 44] and IDBA-UD v1.1.1 (a *de novo* assembler for metagenomic sequences) [45]. For the SOAPdenovo, we tested the k-mer length ranging from 23bp to 123bp by 10bp step for each sample, and selected the assembled contig set with longest N50 length. For the IDBA-UD, parameters "--mink 21 --maxk 81 --step 20 --pre_correction" were used. For most samples, IDBA-UD obtained a better

assembled contig set than SOAPdenovo. This could be attributable to the relative efficiency of IDBA-UD in assembling bacterial genomes within regions of highly uneven depth in metagenomic samples. As a result, we obtained an average 197.9 ± 50.3 Mbp (mean ± SD) contig sets for each pregnant women sample, with N50 length 8.8 ± 3.9 kbp. Unassembled reads from these samples were pooled and re-assembled by using IDBA-UD for further analysis.

Genes were predicted by MetaGeneMark [46] based on parameter exploration by the MOCAT pipeline (MOCAT, RRID:SCR_011943) [47]. A non-redundant gene catalogue of pregnant women samples was constructed using CD-HIT (CD-HIT, RRID:SCR_007105) [48], through which, genes with >90% overlap and >95% nucleic acid similarity (no gap allowed) were removed as redundancies. A pregnant women gene catalogue containing 4,344,984 non-redundant genes was generated for fecal samples collected from these 124 pregnant women. This gene catalogue was further combined with the previous integrated gene catalogue (IGC) [44] by removing redundancies (2,621,398 genes) in the same manner as above. In the end, 39.6% (1,723,586) of the genes in the pregnant women gene catalogue were identified as novel.

Quantification of metagenomic genes

The abundance of genes in the combined non-redundant gene catalogue (combining the pregnant women gene catalogue and IGC) was quantified as relative abundance of reads. First, high-quality reads from each sample were aligned against the gene catalogue using SOAP2.21 [42], with thresholds that allowed a maximum of two mismatches in the initial 32bp seed sequence and 90% similarity over the whole reads. Only two types of alignments were accepted: (1) the entire paired-end read can be mapped onto a gene with the correct insert-size; (2) one end of the paired-end read can be mapped onto the end of a gene, only if the other end of read was mapped outside the genic region. The relative abundance of a gene in a sample was estimated by dividing the number of reads that uniquely mapped to that gene by the length of the gene region and by the total number of reads from the sample that uniquely mapped to any gene in the catalogue. The resulting set of gene relative abundances of a sample was its gene profile.

Richness

We used the gene count and Shannon index to represent the richness and evenness of the gut

microbiota for each sample. As defined previously [5], the gene counts of a metagenomic sample were calculated based on their reads mapping number on the non-redundant gene catalogue. To eliminate the influence of sequencing depth fluctuation, an equal number of 11 million reads for all samples were randomly extracted for mapping, and then, the mean number of genes over 30 random drawings was generated. The Shannon index (within sample diversity) was calculated as previously described [22].

Taxonomical and functional analyses

Taxonomical classification of genes. Reference microbial genomes were downloaded from the NCBI-genome database (version May-2015), which included 8,953 bacterial/archaea genomes (of which, 2,785 genomes were complete and 6,168 were draft genomes), and 4,400 viral genomes. Genes from the non-redundant gene catalogue were aligned to reference genomes using BLASTN (BLASTN, RRID:SCR_001598) with parameters "-word_size 16 -evalue 1e-10 -max_target_seqs 5000". At least 70% alignment coverage of each gene was needed. Based on the parameter exploration of sequence similarity across phylogenetic ranks [49], we used 85% identity as the threshold for genus assignment, and 65% for phylum assignment.

Functional annotation of genes. The Kyoto Encyclopedia of Genes and Genomes (KEGG orthologous, version Apr-2015) (KEGG, RRID:SCR_012773) and evolutionary genealogy of genes: Non-supervised Orthologous Groups (eggNOG, v4) databases (eggNOG, RRID:SCR_002456) were used for functional annotation of genes. Translated amino acid sequences of genes were searched against these databases using USEARCH v8.0.1616 [50] (evalue < 1e-5, query_cov > 0.70) with a minimum similarity of 30%. Each protein was assigned a KEGG orthologue (KO) or eggNOG orthologue group (OG) based on the best-hit gene in the database. Using this approach, 43.6% and 71.9% of the genes in the combined gene catalogue could be assigned a KO or OG, respectively. As a final step, the abundance profiles of KEGG and eggNOG were calculated by summing up the relative abundance of genes annotated to a feature.

Metagenome-wide association study (MGWAS)

We used the MGWAS methodology to identify gene markers that showed significant abundance differences between the GDM and control individuals. The MGWAS was performed using

methodology developed by Qin et al [22]. Briefly, gene relative abundance profiles were initially adjusted for population stratifications using the modified EIGENSTRAT method [51] that allows the use of covariance matrices estimated from abundance levels instead of genotypes. Then, a two-tailed Mann-Whitney U test was performed in the adjusted gene profiles, and the Benjamin-Hochberg procedure [52] was subsequently used to correct the p-values to generate the false discovery rate (FDR, known as "q-value") for each gene.

Metagenomic linkage group (MLG) analysis

Co-abundance genes were clustered into MLGs based on the previously described methodology [22]. Taxonomic assignment and abundance profiling of the MLGs were performed according to the taxonomy and the relative abundance of their constituent genes as previously described [22]. Briefly, assignment to species requires 90% of genes in an MLG to align with the species' genome with 95% identity and 70% overlap of query. Assigning an MLG to a genus requires 80% of its genes to align with a genome with 85% identity in both DNA and protein sequences. MLGs were further interconnected according to Spearman's correlation coefficient ($\rho>0.4$ or $\rho<-0.4$) between their abundances in all GMD and control samples, and the co-occurrence network of MLGs was visualized by Cytoscape 3.0.2 (Cytoscape, RRID:SCR_003032) [53]. The direction of enrichment was determined by the Mann-Whitney U test ($p<0.05$).

Statistical analysis

Statistical analysis was implemented using the R platform. Distance-based redundancy analysis (dbRDA) was performed using the "vegan" package [54] based on the Bray-Curtis distances on normalized taxa relative abundance matrices, then visualized using the "ggplot2" package. Permutational multivariate analysis of variance (PERMANOVA) was performed using the "vegan" package, and the permuted *p*-value was obtained by 10,000 permutations.

The Random Forest model has been shown [6] to be a suitable model for exploiting metagenomic data. Random Forest models were trained using the "randomForest" package (default parameters and 10,000 trees) to identify GDM status in a subset of GDM patients and control group by using the abundance profiles of species and MLGs. Performance of the predictive model was evaluated with cross-validation error. Variable importance by mean decrease

in accuracy was calculated for the Random Forest models using the full set of species or MLGs. Based on the rank of variables by importance, concise models were constructed that contained only the most important variables.

Receiver operator characteristic (ROC) analysis was performed using the "pROC" package, we then computed the 95% confidence interval (CI) of the area under the ROC curve (AUC) with 10,000 bootstrap replicates to assess the variability of the measure. Rarefaction analysis was performed to assess the gene richness of metagenomic samples, implemented by in-house Perl scripts.

Availability of supporting data and materials

All raw sequencing data has been deposited in the EBI Sequence Read Archive (SRA) under accession number ERP020710. Further supporting data is available in the GigaScience repository, GigaDB [REF#]

Abbreviations

MLGs, metagenome linkage groups; GDM, gestational diabetes mellitus;
MGWAS, Metagenome-wide association study;
PTS, phosphotransferase system; LPS, lipopolysaccharide; T2D, type 2 diabetes.

Competing interests

The authors declare that they have no competing interests.

Funding

This study is supported by the National Natural Science Foundation of China (81673181), Guangzhou Science and Technology Bureau, Guangzhou, China (201508030037), and Shenzhen Municipal Government of China (JSGG20140702161403250, CXB201108250098A and JSGG20160229172752028). The sponsors had no role in design and conduct of the study; collection, management, analysis, and interpretation of the data; or in preparation, review, or approval of the manuscript; or the decision to submit the manuscript for publication.

Authors' contributions

XQ and HX designed the birth cohort on which this study was based. XQ and HX designed the

Commented [NN2]: Please add this section –
Note that All the raw sequencing data needs to be deposited in the SRA and the accession numbers quoted in the manuscript.
highlighting that supporting data is available in the GigaSCience repository, GigaDB [Ref#]
Reply： done as requested

Commented [NN3]: Please list all abbreviations used more than once in alphabetical order.
Reply： done as requested

Commented [NN4]: Please declare all competing interests or state "The authors' declare that they have no competing interests".
Reply： done as requested

Commented [NN5]: ll sources of funding for the research reported should be declared. The role of the funding body in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript should be declared. Please use FundRef to report funding sources and include the award/grant number, and the name of the Principal Investigator of the grant.
Reply： done as requested

Commented [NN6]: The individual contributions of authors to the manuscript should be specified in this section. Guidance and criteria for authorship can be found in our editorial policies. We would recommend you follow some kind of standardised taxonomy like the CASRAI CRediT (Contributor Roles Taxonomy).
Reply： done as requested

study and directed its implementation. YK, MY, JH, JL*, NC, WX, SS, LQ, YW, CH, QC, WL and YW were involved in study design and sample collection. YK, JL* and SL analyzed the data and drafted the manuscript. XQ, HD, JL and CP revised the manuscript. All authors critically revised the manuscript, and approved the final version. JL* represents Jin-Hua Lu.

Acknowledgements

We thank all the pregnant women who participated in the Born in Guangzhou Cohort Study and all staff in the cohort team for their contribution to this study, particularly the research nurses and midwives and other recruiting staff for their excellent work.

Commented [NN7]: Please acknowledge anyone who contributed towards the article who does not meet the criteria for authorship including anyone who provided professional writing services or materials.
Authors should obtain permission to acknowledge from all those mentioned in the Acknowledgements section. If you do not have anyone to acknowledge, please write "Not applicable" in this section.
Reply： done as requested

Reference

1. Ferrara A: **Increasing prevalence of gestational diabetes mellitus: a public health perspective**. *Diabetes care* 2007, **30 Suppl 2**:S141-146.
2. Landon MB, Gabbe SG: **Gestational diabetes mellitus**. *Obstetrics and gynecology* 2011, **118**(6):1379-1393.
3. Kim C, Newton KM, Knopp RH: **Gestational diabetes and the incidence of type 2 diabetes: a systematic review**. *Diabetes care* 2002, **25**(10):1862-1868.
4. Allalou A, Nalla A, Prentice KJ, Liu Y, Zhang M, Dai FF, Ning X, Osborne LR, Cox BJ, Gunderson EP *et al*: **A Predictive Metabolic Signature for the Transition From Gestational Diabetes Mellitus to Type 2 Diabetes**. *Diabetes* 2016, **65**(9):2529-2539.
5. Le Chatelier E, Nielsen T, Qin J, Prifti E, Hildebrand F, Falony G, Almeida M, Arumugam M, Batto JM, Kennedy S *et al*: **Richness of human gut microbiome correlates with metabolic markers**. *Nature* 2013, **500**(7464):541-546.
6. Karlsson FH, Tremaroli V, Nookaew I, Bergstrom G, Behre CJ, Fagerberg B, Nielsen J, Backhed F: **Gut metagenome in European women with normal, impaired and diabetic glucose control**. *Nature* 2013, **498**(7452):99-103.
7. Karlsson F, Tremaroli V, Nielsen J, Backhed F: **Assessing the human gut microbiota in metabolic diseases**. *Diabetes* 2013, **62**(10):3341-3349.
8. Cani PD, Neyrinck AM, Fava F, Knauf C, Burcelin RG, Tuohy KM, Gibson GR, Delzenne NM: **Selective increases of bifidobacteria in gut microflora improve high-fat-diet-induced diabetes in mice through a mechanism associated with endotoxaemia**. *Diabetologia* 2007, **50**(11):2374-2383.
9. Chu SY, Callaghan WM, Kim SY, Schmid CH, Lau J, England LJ, Dietz PM: **Maternal obesity and risk of gestational diabetes mellitus**. *Diabetes care* 2007, **30**(8):2070-2076.
10. Hedderson MM, Williams MA, Holt VL, Weiss NS, Ferrara A: **Body mass index and weight gain prior to pregnancy and risk of gestational diabetes mellitus**. *American journal of obstetrics and gynecology* 2008, **198**(4):409 e401-407.

Commented [NN8]: Please move all URLs in the main paper to the reference list. Then renumber the references list in the paper and below.
Reply： done as requested

11. Pedersen HK, Gudmundsdottir V, Nielsen HB, Hyotylainen T, Nielsen T, Jensen BA, Forslund K, Hildebrand F, Prifti E, Falony G *et al*: **Human gut microbes impact host serum metabolome and insulin sensitivity**. *Nature* 2016, **535**(7612):376-381.
12. Collado MC, Isolauri E, Laitinen K, Salminen S: **Distinct composition of gut microbiota during pregnancy in overweight and normal-weight women**. *The American journal of clinical nutrition* 2008, **88**(4):894-899.
13. Vrieze A, Van Nood E, Holleman F, Salojarvi J, Kootte RS, Bartelsman JF, Dallinga-Thie GM, Ackermans MT, Serlie MJ, Oozeer R *et al*: **Transfer of intestinal microbiota from lean donors increases insulin sensitivity in individuals with metabolic syndrome**. *Gastroenterology* 2012, **143**(4):913-916 e917.
14. Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, Gordon JI: **An obesity-associated gut microbiome with increased capacity for energy harvest**. *Nature* 2006, **444**(7122):1027-1031.
15. England LJ, Dietz PM, Njoroge T, Callaghan WM, Bruce C, Buus RM, Williamson DF: **Preventing type 2 diabetes: public health implications for women with a history of gestational diabetes mellitus**. *American journal of obstetrics and gynecology* 2009, **200**(4):365 e361-368.
16. Koren O, Goodrich JK, Cullender TC, Spor A, Laitinen K, Backhed HK, Gonzalez A, Werner JJ, Angenent LT, Knight R *et al*: **Host remodeling of the gut microbiome and metabolic changes during pregnancy**. *Cell* 2012, **150**(3):470-480.
17. Gomez-Arango LF, Barrett HL, McIntyre HD, Callaway LK, Morrison M, Dekker Nitert M, Group ST: **Connections Between the Gut Microbiome and Metabolic Hormones in Early Pregnancy in Overweight and Obese Women**. *Diabetes* 2016, **65**(8):2214-2223.
18. Wang J, Jia H: **Metagenome-wide association studies: fine-mining the microbiome**. *Nature reviews Microbiology* 2016, **14**(8):508-522.
19. Lain KY, Catalano PM: **Metabolic changes in pregnancy**. *Clinical obstetrics and gynecology* 2007, **50**(4):938-948.
20. Mor G, Cardenas I: **The immune system in pregnancy: a unique complexity**. *American journal of reproductive immunology* 2010, **63**(6):425-433.
21. Newbern D, Freemark M: **Placental hormones and the control of maternal metabolism and fetal growth**. *Current opinion in endocrinology, diabetes, and obesity* 2011, **18**(6):409-416.
22. Qin J, Li Y, Cai Z, Li S, Zhu J, Zhang F, Liang S, Zhang W, Guan Y, Shen D *et al*: **A metagenome-wide association study of gut microbiota in type 2 diabetes**. *Nature* 2012, **490**(7418):55-60.
23. Boente RF, Ferreira LQ, Falcao LS, Miranda KR, Guimaraes PL, Santos-Filho J, Vieira JM, Barroso DE, Emond JP, Ferreira EO *et al*: **Detection of resistance genes and susceptibility patterns in Bacteroides and Parabacteroides strains**. *Anaerobe* 2010, **16**(3):190-194.
24. Garrett WS, Gallini CA, Yatsunenko T, Michaud M, DuBois A, Delaney ML, Punit S, Karlsson M, Bry L, Glickman JN *et al*: **Enterobacteriaceae act in concert with the gut microbiota to induce spontaneous and maternally transmitted colitis**. *Cell host & microbe* 2010, **8**(3):292-300.
25. Liu J, Yan Q, Luo F, Shang D, Wu D, Zhang H, Shang X, Kang X, Abdo M, Liu B *et al*: **Acute cholecystitis associated with infection of Enterobacteriaceae from gut microbiota**. *Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases* 2015, **21**(9):851 e851-859.

26. Fei N, Zhao L: **An opportunistic pathogen isolated from the gut of an obese human causes obesity in germfree mice**. *The ISME journal* 2013, **7**(4):880-884.
27. Qin N, Yang F, Li A, Prifti E, Chen Y, Shao L, Guo J, Le Chatelier E, Yao J, Wu L *et al*: **Alterations of the human gut microbiome in liver cirrhosis**. *Nature* 2014, **513**(7516):59-64.
28. Gevers D, Kugathasan S, Denson LA, Vazquez-Baeza Y, Van Treuren W, Ren B, Schwager E, Knights D, Song SJ, Yassour M *et al*: **The treatment-naive microbiome in new-onset Crohn's disease**. *Cell host & microbe* 2014, **15**(3):382-392.
29. Machiels K, Joossens M, Sabino J, De Preter V, Arijs I, Eeckhaut V, Ballet V, Claes K, Van Immerseel F, Verbeke K *et al*: **A decrease of the butyrate-producing species Roseburia hominis and Faecalibacterium prausnitzii defines dysbiosis in patients with ulcerative colitis**. *Gut* 2014, **63**(8):1275-1283.
30. Peng L, Li ZR, Green RS, Holzman IR, Lin J: **Butyrate enhances the intestinal barrier by facilitating tight junction assembly via activation of AMP-activated protein kinase in Caco-2 cell monolayers**. *The Journal of nutrition* 2009, **139**(9):1619-1625.
31. Vaarala O, Atkinson MA, Neu J: **The "perfect storm" for type 1 diabetes: the complex interplay between intestinal microbiota, gut permeability, and mucosal immunity**. *Diabetes* 2008, **57**(10):2555-2562.
32. Sun L, Yu Z, Ye X, Zou S, Li H, Yu D, Wu H, Chen Y, Dore J, Clement K *et al*: **A marker of endotoxemia is associated with obesity and related metabolic disorders in apparently healthy Chinese**. *Diabetes care* 2010, **33**(9):1925-1932.
33. Jayashree B, Bibin YS, Prabhu D, Shanthirani CS, Gokulakrishnan K, Lakshmi BS, Mohan V, Balasubramanyam M: **Increased circulatory levels of lipopolysaccharide (LPS) and zonulin signify novel biomarkers of proinflammation in patients with type 2 diabetes**. *Molecular and cellular biochemistry* 2014, **388**(1-2):203-210.
34. Manco M, Putignani L, Bottazzo GF: **Gut microbiota, lipopolysaccharides, and innate immunity in the pathogenesis of obesity and cardiovascular risk**. *Endocrine reviews* 2010, **31**(6):817-844.
35. Abreu MT: **Toll-like receptor signalling in the intestinal epithelium: how bacterial recognition shapes intestinal function**. *Nature reviews Immunology* 2010, **10**(2):131-144.
36. Brun P, Castagliuolo I, Di Leo V, Buda A, Pinzani M, Palu G, Martines D: **Increased intestinal permeability in obese mice: new evidence in the pathogenesis of nonalcoholic steatohepatitis**. *American journal of physiology Gastrointestinal and liver physiology* 2007, **292**(2):G518-525.
37. Qiu X, Lu J-H, He J-R, Lam K-bH, Shen S-Y, Guo Y, Kuang Y-S, Yuan M-Y, Qiu L, Chen N-N: **The Born in Guangzhou Cohort Study (BIGCS)**. 2017.
38. Zhu WW, Yang HX: **Diagnosis of gestational diabetes mellitus in China**. *Diabetes Care* 2013, **36**(6):e76.
39. Association. AD: **Standards of medical care in diabetes--2011**. *Diabetes Care* 2011, **34 Suppl 1**:S11-61.
40. Mirsepasi H, Persson S, Struve C, Andersen LO, Petersen AM, Krogfelt KA: **Microbial diversity in fecal samples depends on DNA extraction method: easyMag DNA extraction compared to QIAamp DNA stool mini kit extraction**. *BMC research notes* 2014, **7**:50.
41. **FASTAX Toolkit .**http://hannonlab.cshl.edu/fastx_toolkit/.
42. Li R, Yu C, Li Y, Lam TW, Yiu SM, Kristiansen K, Wang J: **SOAP2: an improved ultrafast tool for**

short read alignment. *Bioinformatics* 2009, **25**(15):1966-1967.

43. Li R, Zhu H, Ruan J, Qian W, Fang X, Shi Z, Li Y, Li S, Shan G, Kristiansen K *et al*: **De novo assembly of human genomes with massively parallel short read sequencing**. *Genome Res* 2010, **20**(2):265-272.

44. Li J, Jia H, Cai X, Zhong H, Feng Q, Sunagawa S, Arumugam M, Kultima JR, Prifti E, Nielsen T *et al*: **An integrated catalog of reference genes in the human gut microbiome**. *Nature biotechnology* 2014, **32**(8):834-841.

45. Peng Y, Leung HC, Yiu SM, Chin FY: **IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth**. *Bioinformatics* 2012, **28**(11):1420-1428.

46. Zhu W, Lomsadze A, Borodovsky M: **Ab initio gene identification in metagenomic sequences**. *Nucleic acids research* 2010, **38**(12):e132.

47. Kultima JR, Sunagawa S, Li J, Chen W, Chen H, Mende DR, Arumugam M, Pan Q, Liu B, Qin J *et al*: **MOCAT: a metagenomics assembly and gene prediction toolkit**. *PloS one* 2012, **7**(10):e47656.

48. Li W, Godzik A: **Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences**. *Bioinformatics* 2006, **22**(13):1658-1659.

49. Arumugam M, Raes J, Pelletier E, Le Paslier D, Yamada T, Mende DR, Fernandes GR, Tap J, Bruls T, Batto JM *et al*: **Enterotypes of the human gut microbiome**. *Nature* 2011, **473**(7346):174-180.

50. Edgar RC: **Search and clustering orders of magnitude faster than BLAST**. *Bioinformatics* 2010, **26**(19):2460-2461.

51. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D: **Principal components analysis corrects for stratification in genome-wide association studies**. *Nature genetics* 2006, **38**(8):904-909.

52. Storey JD: **A direct approach to false discovery rates Journal of the Royal Statistical Society: Series B (Statistical Methodology) Volume 64, Issue 3**. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 2002, **64**(3):479-498.

53. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T: **Cytoscape: a software environment for integrated models of biomolecular interaction networks**. *Genome research* 2003, **13**(11):2498-2504.

54. Dixon P: **VEGAN, a package of R functions for community ecology Journal of Vegetation Science Volume 14, Issue 6**. *Journal of Vegetation Science* 2003, **14**(6):927-930.

Figure legends

Figure 1 | Difference in microbial composition between GDM and healthy pregnant women. (**a**) Distance-based redundancy analysis (dbRDA) based on Bray–Curtis distances between microbial genera, revealing a GDM dysbiosis which overlaps only in part with taxonomic composition in GDM patients and healthy controls. The first two principal components (PCs) and the ratio of variance contributed by them is shown. Lines connect samples in the same group, and colored circles cover the samples near the center of gravity for each group. Genus (blue square), as the main contributors, are plotted by their loading in the PCs. (**b**) Boxplot shows genera that differ significantly between GDM patients and healthy controls. Genera with $q<0.05$ (Mann-Whitney U test corrected by the Benjamini-Hochberg method) are shown. Red and green boxes represent GDM patients and healthy controls, respectively. Only the genera with average relative abundances greater than 0.05% in all the samples are shown for clarity. The boxes represent the interquartile range (IQR) between first and third quartiles and the line inside represents the median. The whiskers denote the lowest and highest values within 1.5 times IQR from the first and third quartiles, respectively. The circles represent outliers beyond the whiskers.

Figure 2 | Interconnection of GDM-associated MLGs. A co-occurrence network deduced from GDM-enriched and control-enriched MLGs is shown. Nodes depict MLG's with their taxonomic assignment or ID shown. The size of each node indicates the number of genes within the MLG. Connecting lines represent Spearman correlation coefficient $\rho > 0.40$ (gray line) or < -0.40 (red line). Classified MLGs are colored (red: GDM-enriched; green: control-enriched) and grouped according to their taxonomic information. Only MLGs with >100 genes are shown for clarity of presentation and visualization, and the detailed information of all 129 MLGs are given in Table S2.

Figure 3 | Association of gross abundance of GDM-enriched and control-enriched MLGs with blood glucose levels 0, 60, and 120 minutes after an oral glucose tolerance test. Scatter plots of all samples (including GDM patients and healthy controls) are shown with lines indicating linear fit.

Figure 4 | Correlation of blood glucose levels 0, 60, and 120 minutes after an oral glucose tolerance test (only OGTT results are shown, I do not see the results for HbA1C in this figure) with MLGs (a) and species (b). Spearman's rank correlation coefficients and P-values for the correlations are shown. '+' denotes P<0.05; '++' denotes P<0.01. Only MLGs or species with average relative abundances greater than 0.001% and correlated (P<0.05) with at least one index are shown for clarity.

Figure 5 | Association of microbial genetic functional pathway composition in GDM patients and healthy pregnant women. (a) Distributions of relative abundances of KEGG pathway categories in GDM patients and healthy controls. '*' denotes q<0.05 (Mann-Whitney U test corrected by the Benjamini-Hochberg method) (b) Correlation of blood glucose levels 0, 60, and 120 minutes after an oral glucose tolerance test, with PTS system and LPS biosynthesis and transport system. Spearman's rank correlation coefficients and P-values for the correlations are shown. '+' denotes P<0.05; '++' denotes P<0.01.

Figure 6 | Classification of GDM status by the relative abundance of MLGs and species. (a) Classification performance of a random forest model using MLG or species abundance assessed by AUC. The performance was explored for different numbers of explanatory variables, ordered in importance. (b-c) The 30 most discriminant MLGs (b) and species (c) in the models classifying GDM and controls. The bar lengths in b and c indicate the importance of the variable, and colors represent enrichment in GDM (red shades) or controls (blue shades). (d) ROC analysis for classification of GDM status by MLGs and PBMI.

Figure 7 | **A schematic diagram showing the main bacteria and functions of the gut microbes that had a predicted GDM association.** Red and orange columns and text denotes enriched bacteria and their putative functions in GDM patients; green columns and text denotes depleted bacteria and their putative functions in GDM patients.

Figure 1

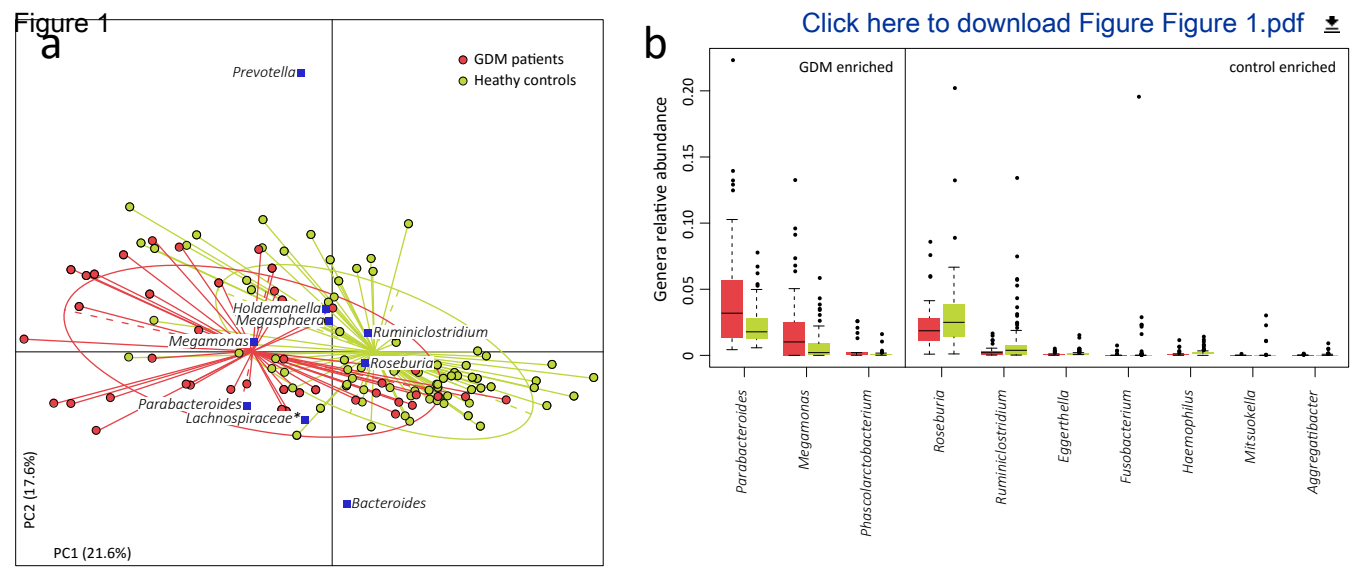

a
GDM patients
Heathy controls
Prevotella
Holdemanella
Megasphaera
Megamonas
Ruminiclostridium
Roseburia
Parabacteroides
Lachnospiraceae *
Bacteroides
PC2 (17.6%)
PC1 (21.6%)
b
Click here to download Figure Figure 1.pdf
GDM enriched
control enriched
Genera relative abundance
0
0.05
0.10
0.15
0.20
Parabacteroides
Megamonas
Phascolarctobacterium
Roseburia
Ruminiclostridium
Eggerthella
Fusobacterium
Haemophilus
Mitsuokella
Aggregatibacter

Figure 2

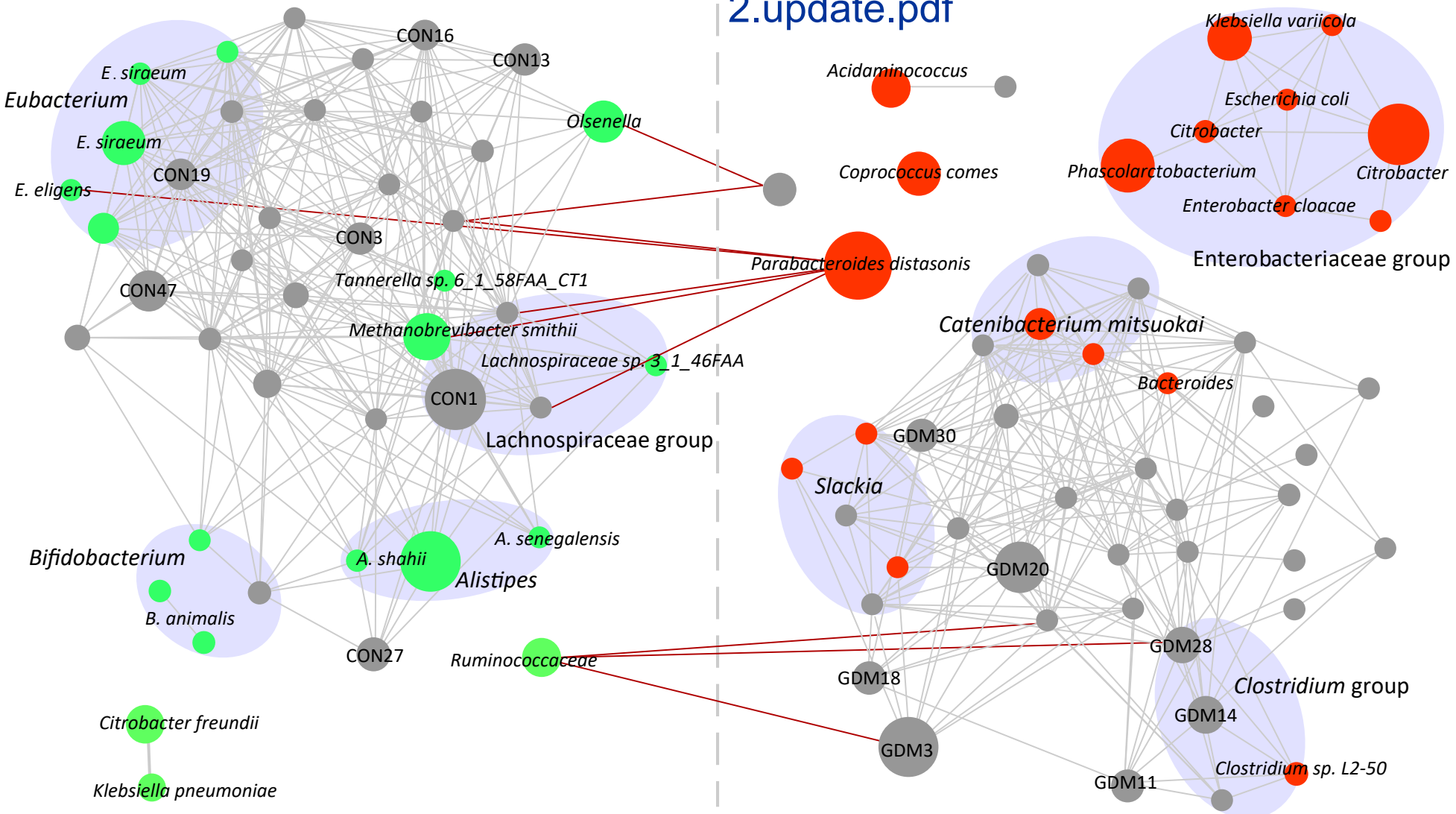

Control-enriched MLGs
CON16
CON13
E. siraeum
Eubacterium
Olsenella
E. siraeum
CON19
E. eligens
CON3
CON47
Tannerella sp. 6_1_58FAA_CT1
Methanobrevibacter smithii
Lachnospiraceae sp. 3_1_46FAA
CON1
Lachnospiraceae group
Bifidobacterium
A. shahii
A. senegalensis
Alistipes
B. animalis
CON27
Ruminococcaceae
Citrobacter freundii
Klebsiella pneumoniae
Click here to download Figure Figure 2.update.pdf
GDM-enriched MLGs
Acidaminococcus
Coprococcus comes
Parabacteroides distasonis
Klebsiella variicola
Escherichia coli
Citrobacter
Phascolarctobacterium
Citrobacter
Enterobacter cloacae
Enterobacteriaceae group
Catenibacterium mitsuokai
Bacteroides
GDM30
Slackia
GDM20
GDM28
GDM18
Clostridium group
GDM14
GDM3
Clostridium sp. L2-50
GDM11

Figure 3

Click here to download Figure figure3.pdf

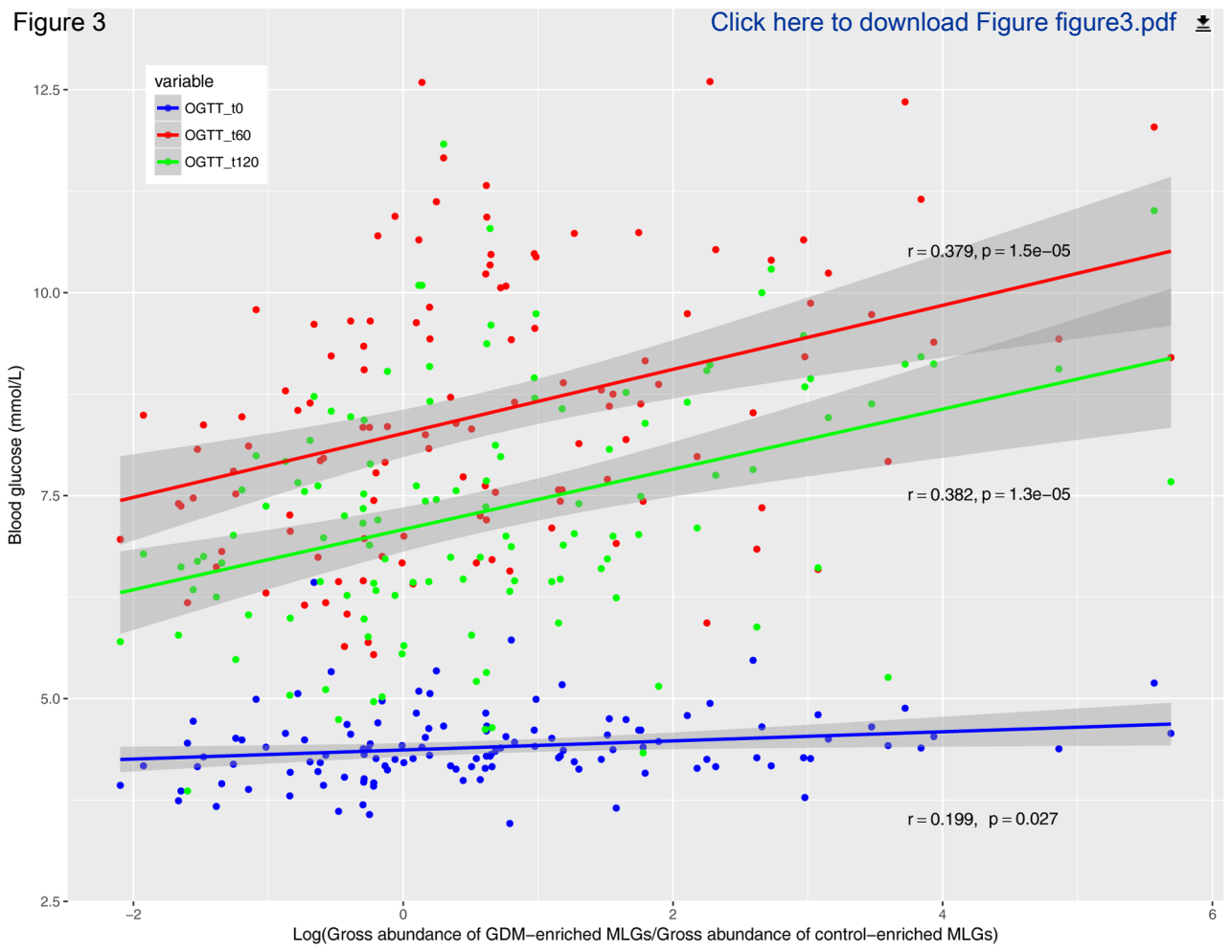


Figure 4

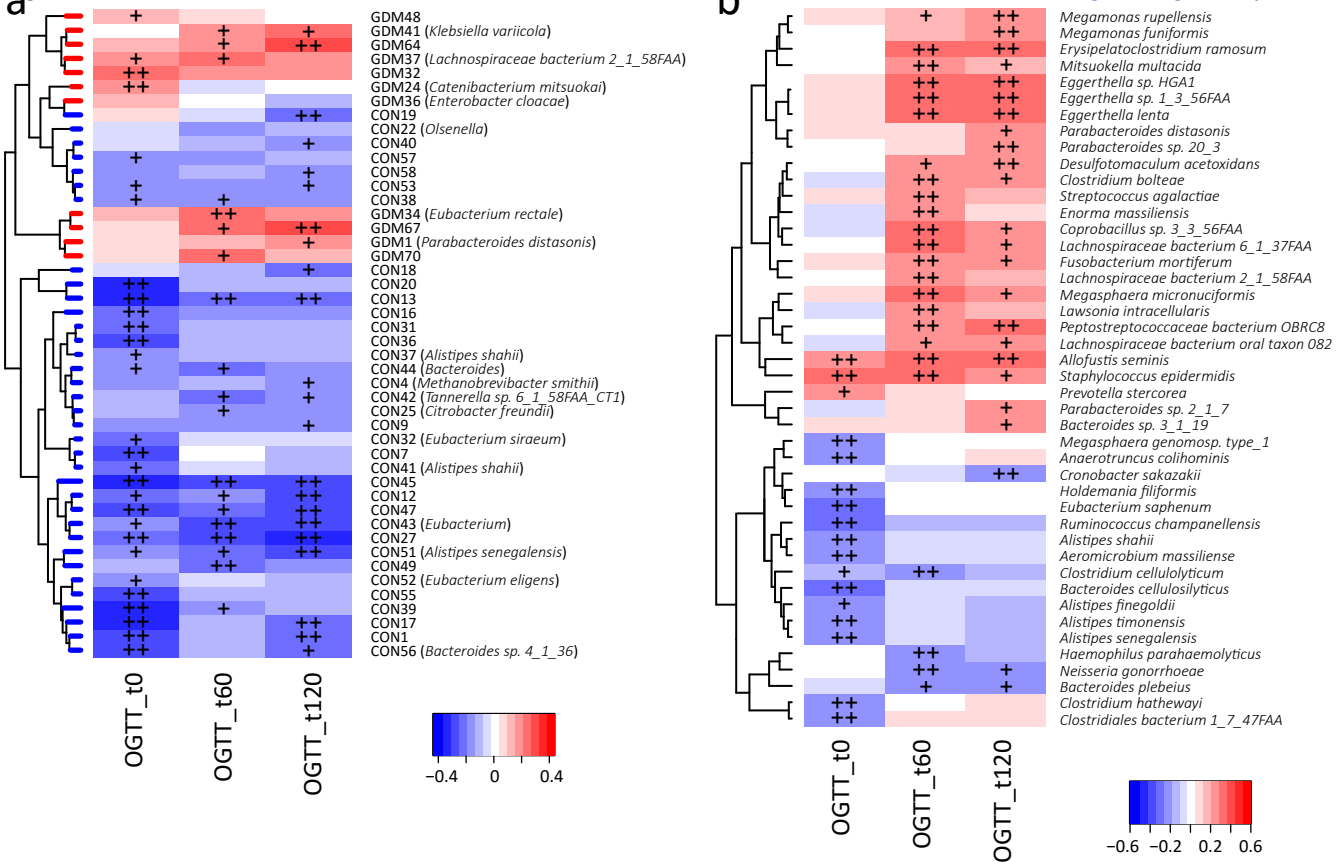
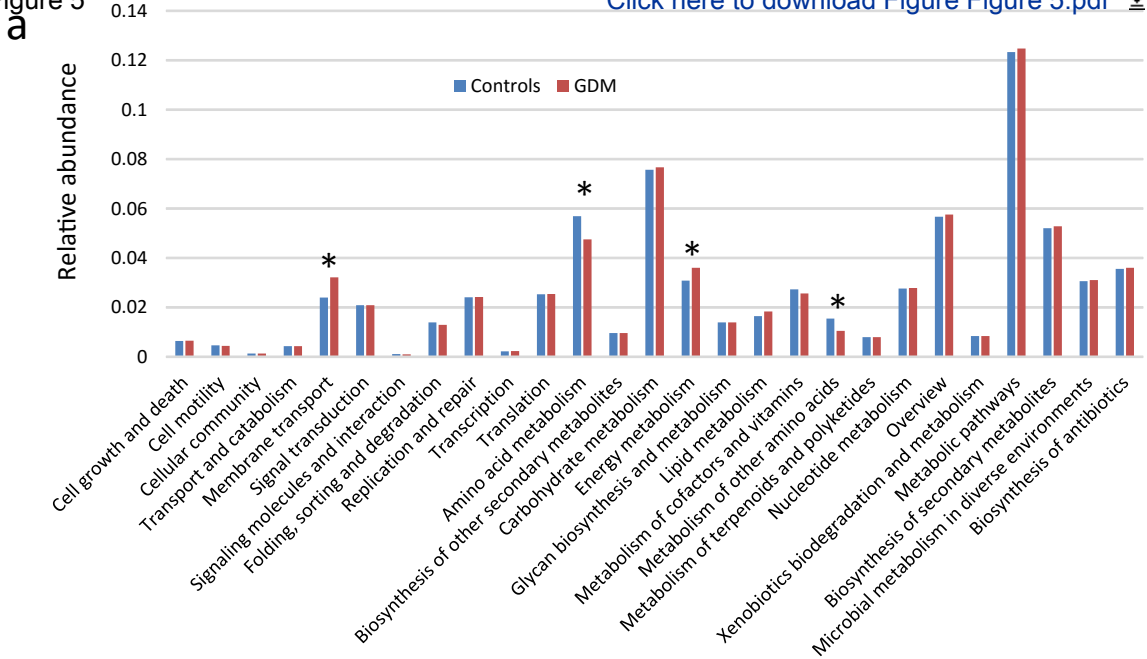

a
GDM48
GDM41 (Klebsiella variicola)
GDM64
GDM37 (Lachnospiraceae bacterium 2_1_58FAA)
GDM32
GDM24 (Catenibacterium mitsuokai)
GDM36 (Enterobacter cloacae)
CON19
CON22 (Olsenella)
CON40
CON57
CON58
CON53
CON38
GDM34 (Eubacterium rectale)
GDM67
GDM1 (Parabacteroides distasonis)
GDM70
CON18
CON20
CON13
CON16
CON31
CON36
CON37 (Alistipes shahii)
CON44 (Bacteroides)
CON4 (Methanobrevibacter smithii)
CON42 (Tannerella sp. 6_1_58FAA_CT1)
CON25 (Citrobacter freundii)
CON9
CON32 (Eubacterium siraeum)
CON7
CON41 (Alistipes shahii)
CON45
CON12
CON47
CON43 (Eubacterium)
CON27
CON51 (Alistipes senegalensis)
CON49
CON52 (Eubacterium eligens)
CON55
CON39
CON17
CON1
CON56 (Bacteroides sp. 4_1_36)
OGTT_t0
OGTT_t60
OGTT_t120
−0.4
0
0.4
b
Click here to download Figure Figure 4.pdf
Megamonas rupellensis
Megamonas funiformis
Erysipelatoclostridium ramosum
Mitsuokella multacida
Eggerthella sp. HGA1
Eggerthella sp. 1_3_56FAA
Eggerthella lenta
Parabacteroides distasonis
Parabacteroides sp. 20_3
Desulfotomaculum acetoxidans
Clostridium bolteae
Streptococcus agalactiae
Enorma massiliensis
Coprobacillus sp. 3_3_56FAA
Lachnospiraceae bacterium 6_1_37FAA
Fusobacterium mortiferum
Lachnospiraceae bacterium 2_1_58FAA
Megasphaera micronuciformis
Lawsonia intracellularis
Peptostreptococcaceae bacterium OBRC8
Lachnospiraceae bacterium oral taxon 082
Allofustis seminis
Staphylococcus epidermidis
Prevotella stercorea
Parabacteroides sp. 2_1_7
Bacteroides sp. 3_1_19
Megasphaera genomosp. type_1
Anaerotruncus colihominis
Cronobacter sakazakii
Holdemania filiformis
Eubacterium saphenum
Ruminococcus champanellensis
Alistipes shahii
Aeromicrobium massiliense
Clostridium cellulolyticum
Bacteroides cellulosilyticus
Alistipes finegoldii
Alistipes timonensis
Alistipes senegalensis
Haemophilus parahaemolyticus
Neisseria gonorrhoeae
Bacteroides plebeius
Clostridium hathewayi
Clostridiales bacterium 1_7_47FAA
OGTT_t0
OGTT_t60
OGTT_t120
−0.6
−0.2
0.2
0.6

Figure 5

Click here to download Figure Figure 5.pdf



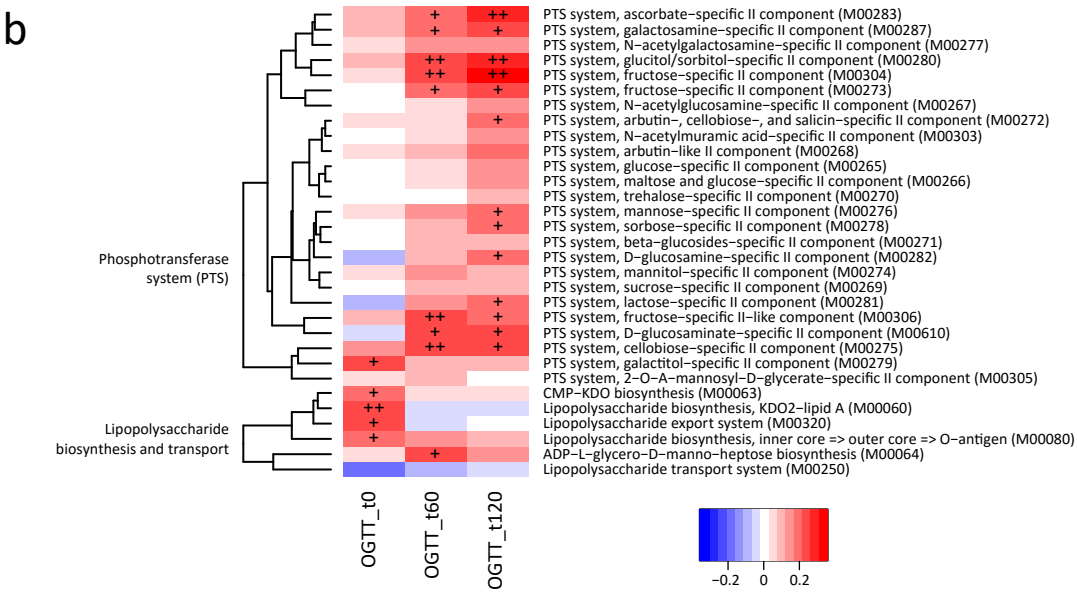
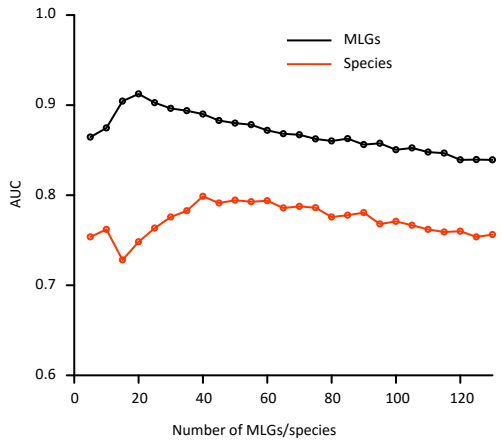
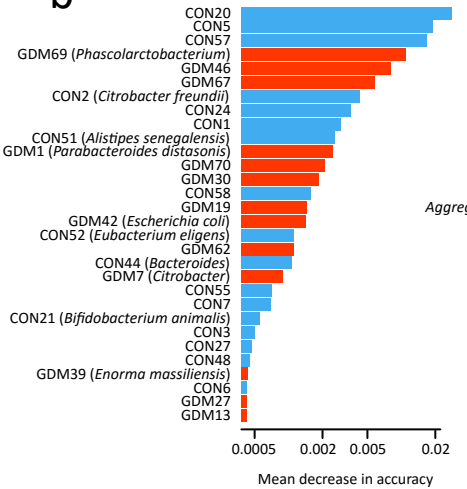
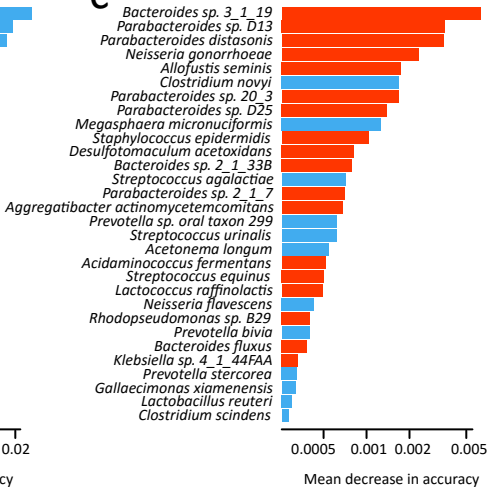


Figure 6

Click here to download Figure Figure 6.pdf







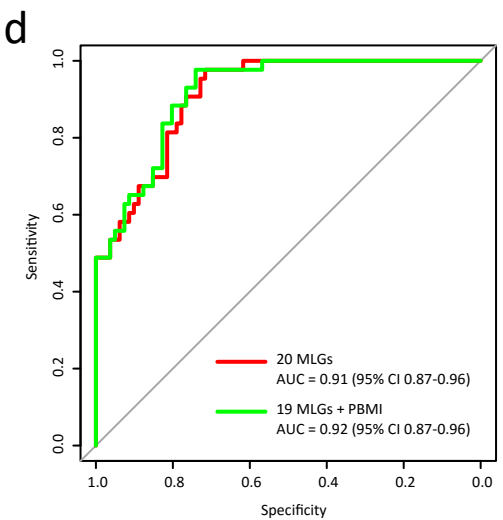
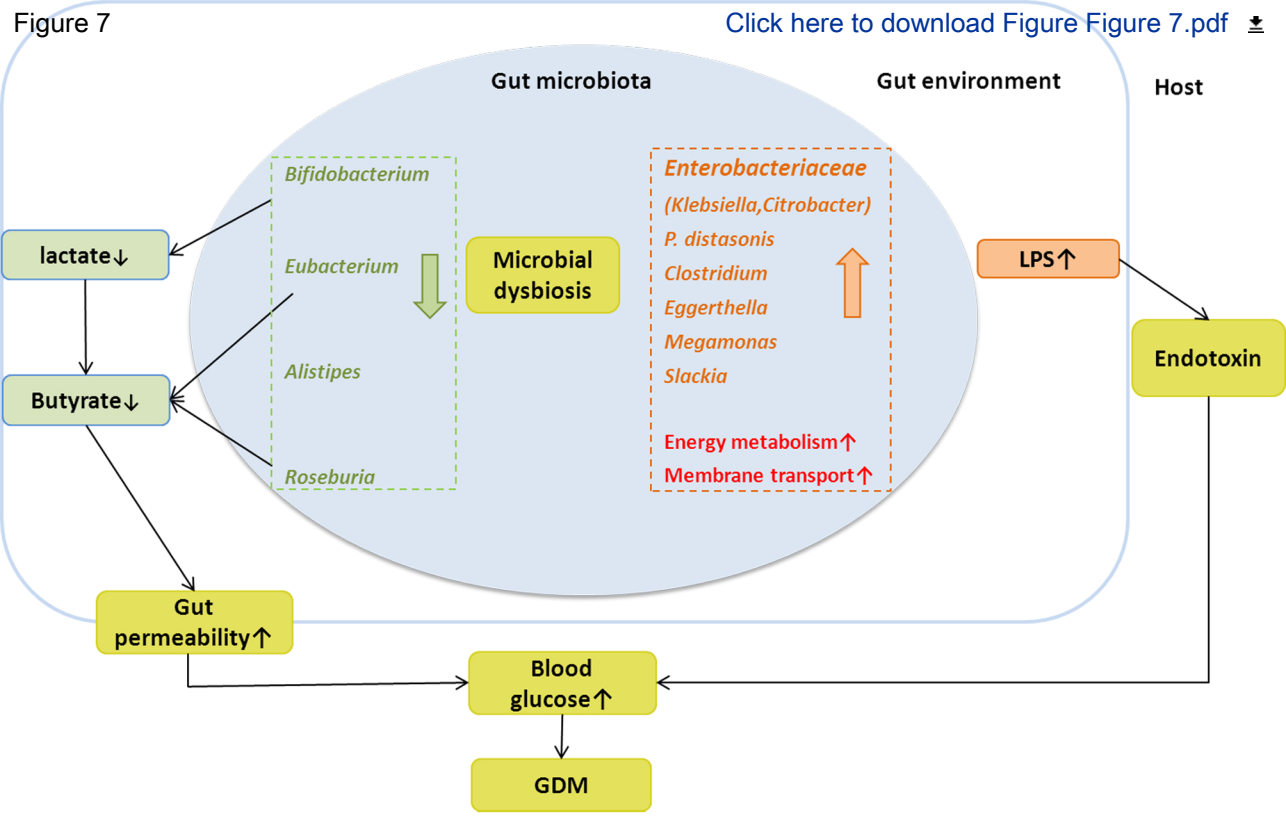


Figure 7



Additional file 1

Additional file2

letter of consent

Dear Dr. Nogoy,

Thank you for inviting us to submit a revised manuscript. We are grateful for the reviewers' comments.

We have addressed each of the points raised by the reviewer and the Gigascience Editorial team, and outlined how we have dealt with these in the attached revised manuscript.

We look forward to hearing from you regarding the suitability of this manuscript for publication in your journal.

Sincerely,

Xiu Qiu, M. D., Ph. D.
Division of Birth Cohort Study,
Guangzhou Women and Children's Medical Center, Guangzhou Medical University
9 Jinsui Road, Tianhe District, Guangzhou 510623, China
Tel: 86 2038367162
Fax: 86 2038367162
Email: qxiu0161@163.com; xiu.qiu@bigcs.org.