

Manuscript Number:	GIGA-D-16-00167
Full Title:	Connections between human gut microbiome and gestational diabetes mellitus
Article Type:	Research
Abstract:	<p>Background 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 collecting at 21-29 weeks, to explore associations between GDM and the composition of microbial taxonomic units and functional genes.</p> <p>Results Metagenome-wide association study (MGWAS) identified 154,837 genes, which enabled to cluster into 129 metagenome linkage groups (MLGs) for species description, with significant abundance differences between 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 GDM-associated species showed correlations with maternal blood glucose, indicating a potential relationship between gut microbes and blood glucose tolerance. We further evaluated the performance of gut microbiota as biomarker to identify GDM status with a Random Forest model and demonstrated that fecal MLGs may offer new indicators for prognosis of GDM.</p> <p>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.</p>
Additional Information:	
Question	Response
Are you submitting this manuscript to a special series or article collection?	No
Experimental design and statistics	Yes
<p>Full details of the experimental design and statistical methods used should be given in the Methods section, as detailed in our Minimum Standards Reporting Checklist. Information essential to interpreting the data presented should be made available in the figure legends.</p> <p>Have you included all the information requested in your manuscript?</p>	
Resources	Yes
A description of all resources used, including antibodies, cell lines, animals and software tools, with enough information to allow them to be uniquely identified, should be included in the Methods section. Authors are strongly	

<p>encouraged to cite Research Resource Identifiers (RRIDs) for antibodies, model organisms and tools, where possible.</p> <p>Have you included the information requested as detailed in our Minimum Standards Reporting Checklist?</p>	
<p>Availability of data and materials</p> <p>All datasets and code on which the conclusions of the paper rely must be either included in your submission or deposited in publicly available repositories (where available and ethically appropriate), referencing such data using a unique identifier in the references and in the “Availability of Data and Materials” section of your manuscript.</p> <p>Have you have met the above requirement as detailed in our Minimum Standards Reporting Checklist?</p>	No
<p>If not, please give reasons for any omissions below.</p> <p>as follow-up to "Availability of data and materials</p> <p>All datasets and code on which the conclusions of the paper rely must be either included in your submission or deposited in publicly available repositories (where available and ethically appropriate), referencing such data using a unique identifier in the references and in the “Availability of Data and Materials” section of your manuscript.</p> <p>Have you have met the above requirement as detailed in our Minimum Standards Reporting Checklist?</p> <p>"</p>	We are planning to submit our datasets to GigaDB database, but now is not yet complete.

Connections between human gut microbiome and gestational diabetes mellitus

Yong Guo^{1,2,†}, Ya-Shu Kuang^{1,†}, Sheng-Hui Li^{1,†}, Jun-Hua Li^{3,4}, Ming-Yang Yuan^{1,2}, Jian-Rong He^{1,2}, Jin-Hua Lu^{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 Chen⁵, Hong-Wen Deng^{1,6}, Christopher J Papasian⁷, 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, Guangzhou Women and Children's Medical Center, Guangzhou Medical University, Guangzhou 510623, China

³ BGI-Shenzhen, Shenzhen 518083, China

⁴ BGI Hong Kong Research Institute, Hong Kong, 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:

geyong084@164.com (YG); kuangyashu@126.com (YSK); lishenghui1005@gmail.com (SHL); lijinhua@genomics.cn (JHL); ymyeasy@126.com (MYE); hjr0703@163.com (JRH); lujinhua10@gmail.com (JHL); cnfezx@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.

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 2038076019; E-mail: huimin.xia876001@gmail.com.

Abstract

Background

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 collecting 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 enabled to cluster into 129 metagenome linkage groups (MLGs) for species description, with significant abundance differences between 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 GDM-associated species showed correlations with maternal blood glucose, indicating a potential relationship between gut microbes and blood glucose tolerance. We further evaluated the performance of gut microbiota as biomarker to identify GDM status with a Random Forest model and demonstrated that fecal MLGs may offer new indicators for prognosis of GDM.

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

1 human metabolic health [5, 6], leading us to hypothesize that the gut microbiome may impact
2 gestational metabolism and development of GDM.
3

4 Microbial dysbiosis in the human gut may be an important environmental risk factor for
5 abnormal host metabolism, as recently exemplified in the studies of obesity and T2D (reviewed by
6 Karlsson, et. al)[7]. A study using experimental animal model revealed reduced numbers of
7 Bifidobacteria led to enhanced endogenous lipopolysaccharide production, endotoxemia, and
8 associated obesity and insulin resistance [8]. In humans, excessive weight gain and obesity in
9 pregnancy resulted in deteriorated glucose tolerance and increased risk of GDM [9, 10]. *Prevotella*
10 *copri* and *Bacteroides vulgatus* have been identified as the main species driving the association
11 between biosynthesis of branched-chain amino acids, insulin resistance, and glucose intolerance
12 [11], and *Bacteroides* spp. and *Staphylococcus aureus* are significantly more abundant in
13 overweight women than in normal-weight women [12].
14
15
16
17
18
19
20
21
22
23
24

25 While the majority of previous studies have focused on associations between intestinal
26 microbiota and obese states or T2D [6, 13-15], some recent studies have sought to characterize
27 microbiota changes during pregnancy, with the goal of providing novel insights into the
28 relationship between microbiota changes during pregnancy and potential metabolic consequences
29 [16]. Studies based on sequencing of 16S ribosomal RNA have revealed novel relationships
30 between gut microbiome composition and the metabolic hormonal environment in overweight and
31 obese pregnant women in early gestation [17]. Koren et al. found that maternal gut microbiota
32 changed from first to third trimesters, with a decline in butyrate-producing bacteria and increased
33 Bifidobacteria, *Proteobacteria*, and lactic-acid producing bacteria [16]. Further, transplants of
34 fecal material obtained during different trimesters were sufficient to confer different phenotypes in
35 mouse models, with third-trimester fecal transplants leading to increased adiposity and
36 inflammation [16]. These studies suggest that pregnancy is associated with major shifts in the gut
37 microbiome which may play an important role in observed increases in gestational inflammation,
38 thereby potentially contributing to development of GDM. However, studies focusing on changes
39 in the gut microbiome during pregnancy and development of GDM have not been reported so far.
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55

56 Metagenomic shotgun sequencing, in which the full complement of genes present in the
57 microbiome are sequenced, can furnish information about the relative abundance of genes in
58 functional pathways and at all taxonomical levels [18]. In this study, we used whole-metagenome
59
60
61
62
63
64
65

1 shotgun sequencing analyses of the gut microbiome during pregnancy to explore associations
2 between GDM and the composition and abundance of microbial taxonomic units and functional
3 genes. The objective was to obtain a comprehensive understanding of the gut microbiome's role in
4 the etiopathogenesis of GDM.
5
6
7
8
9

10 **Data description**

11 We obtained the fecal samples from 124 pregnant women, including 43 GDM patients and 81
12 healthy control individuals, during their second trimester in Guangzhou Women and Children's
13 Medical Center (GWCMC). Whole-metagenome shotgun sequencing of the samples were
14 preformed based on the Illumina HiSeq2000 platform in BGI-Shenzhen, China. We constructed a
15 paired-end library with insert size of 350 base pairs (bp) for every sample, and sequenced with
16 100bp read length from each end. Sequencing reads for fecal samples were independently
17 processed for quality controlling and host sequences removing based on an in-house pipeline (see
18 Methods), and totaling 795 Gbp high quality metagenomic data (average per sample, 6.4 Gbp)
19 were finally generated for further analysis. We performed *de novo* assembly and gene calling for
20 data of each sample, and constructed a non-redundant gene catalogue of all pregnant women
21 samples containing 4,344,984 genes. This gene catalogue provided a suitable reference for
22 metagenomic gene quantification, microbial diversity analysis, and metagenome-wide association
23 study for the pregnant women samples.
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40

41 **Results**

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

43 First, we explored potential differences in the gut microbiome between 43 GDM patients and 81
44 healthy pregnant women. In order to perform this analysis, we obtained 795.3 Gb of high-quality
45 data (6.4 ± 1.3 Gb per sample, Table S1) via metagenomic shotgun sequencing of their fecal
46 samples. We aligned the sequencing reads (43.8%) against available microbial genomes from the
47 National Center for Biotechnology Information and generated taxonomic composition for all
48 samples at the taxonomic levels of phylum, class, order, family, genus and species. Multivariate
49 analysis based on Bray–Curtis distances between microbial genera revealed significant differences
50 between GDM patients and healthy controls (Figure 1a). We then preformed the Mann–Whitney U
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1 test to identify phylogenetic differences between GDM patients and healthy controls. No
2 significant differences were found at the phylum and class levels. However, the order *Clostridiales*
3 was enriched in healthy controls. At the genus level, GDM patients had a significantly higher
4 abundance of *Parabacteroides*, *Megamonas* and *Phascolarctobacterium*, while healthy controls
5 were significantly enriched for *Ruminiclostridium*, *Roseburia*, *Eggerthella*, *Fusobacterium*,
6 *Haemophilus*, *Mitsukella*, and *Aggregatibacter* (Figure 1b). We also found a number of bacterial
7 species that differed significantly between GDM patients and healthy controls, consistent with the
8 genus level observations (Table S2). These findings suggest dysbiosis of the gut microbiota based
9 on GDM status.

21 Identification of GDM-associated markers from gut microbiome

22 To explore detailed signatures of the gut microbiome in GDM patients and healthy controls, we
23 constructed a non-redundant gene catalogue consisting of 4.34 million genes, which allowed an
24 average reads mapping rate of 79.5% for sequenced samples. We identified 154,837 genes that
25 displayed significant abundance differences between the two groups (Mann-Whitney U test,
26 $q < 0.05$). (Figure S1 shows the P-value distribution between GDM patients and healthy pregnant
27 women for all genes tested). ~68% of these genes were clustered into 129 MLGs (Table S3),
28 which allowed species level description for the microbiome differences. The 71 MLGs enriched in
29 GDM patients included *Parabacteroides distasonis*, *Klebsiella variicola*, *Catenibacterium*
30 *mitsukaii*, *Coprococcus comes* and *Citrobacter spp.*, whereas the 58 MLGs enriched in healthy
31 pregnant women included *Methanobrevibacter smithii*, *Alistipes spp.* (*A. shahii*, *A. senegalensis*),
32 *Bifidobacterium spp.* (*B. animalis*, *B. pseudocatenulatum*) and *Eubacterium spp.* (*E. siraeum*, *E.*
33 *elicens*). The GDM-enriched and control enriched MLGs were highly positively interconnected
34 within each group; however, only a few negative connections were found between the two groups
35 (Figure 2). Notably, GDM-enriched MLGs of *Enterobacteriaceae*, including *K. variicola*, *E. coli*,
36 *Enterobacter cloacae* and *Citrobacter spp.*, were closely linked (correlation coefficients > 0.40
37 between each other), representing a cooperative promoting function of *Enterobacteriaceae* to
38 GDM development. Of particular interest, we also observed that the relative abundance of
39 *Enterobacteriaceae* was positively associated with pre-gestational body mass index (PBMI,
40 Figure S2).

Correlations between maternal blood glucose 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 gross abundance of GDM-enriched and control-enriched MLGs were obviously associated with the blood glucose level during the second trimester of pregnancy (as measured by OGTT, Figure 3). 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 a large proportion of control-enriched MLGs were negatively correlated the blood glucose levels (Figure 4a). At the species level, *Eggerthella spp.*, *Megamonas spp.*, *Allofustis seminis* and several *Lachnospiraceae* and *Parabacteroides* species, were positively correlated with glucose tolerance, while several *Alistipes spp.*, among others, 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 microbiomes of GDM patients vs. controls. Although the functional composition of GDM patients and control subjects were highly similar (Figure 5a), the microbiome of GDM patients showed a greater abundance of membrane transport and energy metabolism pathways, while the microbiome of control subjects was enriched in the function of amino acid metabolic pathways. We also found that KEGG modules involving the phosphotransferase system (PTS, a major component of membrane transport) and lipopolysaccharide (LPS, a major component of the outer membrane of Gram-negative bacteria that induces a profound inflammatory response) biosynthesis and export systems, were associated with glucose tolerance levels (Figure 5b). These findings are consistent with those of a previous study [19] that reported an increase in microbiome functions for membrane transport and LPS metabolism in the microbiome of patients with type 2 diabetes.

Gut microbiota-based classification of GDM

Finally, we utilized random forest models to assess the ability of MLGs and species abundance

1 profiles to predict GDM status. We found that 20 MLGs provided the best discriminatory power,
2 as indicated by the area under the ROC curve (AUC) 0.91 (95% CI 0.87-0.96); this was higher
3 than that achieved using species profiles with this model (the best AUC was 0.80; 95% CI
4 0.73-0.86) using 40 species (Figure 6a). The increased AUC for the MLG-based model may be
5 due to the fact that MLGs furnish taxonomic and functional information for unknown or
6 unanalyzable species. Bacterial species providing the highest discriminatory power were primarily
7 members of the *Bacteroides* or *Parabacteroides* genera (Figure 6b-c), consistent with our earlier
8 observation that *Parabacteroides* is the predominant genus accounting for differences in the gut
9 microbiome between GDM patients and control subjects (Figure 1b). When age and PBMI of
10 pregnant women were included along with the 129 MLGs, PBMI was selected as a marker
11 together with the highest discriminatory MLGs, but the performance of the model did not
12 obviously improve (Figure S3 and Figure 6d). Therefore, fecal MLGs, by themselves, may
13 provide a mechanism and biomarkers for early detection of GDM or, potentially, for identifying
14 risk of developing GDM.

31 Discussion

32 To identify and understand alterations in the gut microbiome associated with GDM, we
33 characterized the genic, taxonomic, and functional repertoire of microbiomes of 43 GDM patients
34 and 81 healthy pregnant women. To our knowledge, this is the first metagenomics study on stools
35 of GDM patients, revealing significant dysbiosis, taxonomic shifts and functional changes in their
36 microbiome as compared with healthy pregnant women.

37 Our study furnished a powerful set of microbial markers for GDM prediction, which achieved
38 an AUC of 0.91, for identifying GDM status based on 20 species-level MLGs. The discriminatory
39 power of this set of markers was higher than prediction models based on genomic markers
40 identified by genome-wide association studies (GWAS) (AUC 0.5-0.7) [20]. Thus, as
41 demonstrated in the current study, analysis of fecal microbiota could be used for the early
42 diagnosis of GDM or, potentially, for identifying individuals at risk for developing GDM. Future
43 systematic assessment of the key species and gene markers identified here will be required to
44 further develop this tool.

45 *Enterobacteriaceae*, a bacterial family that contains many commensals as well as human

1 pathogens that causes pneumonia, diarrhea and urinary tract infections, occurred with higher
2 relative abundance in GDM patients than in healthy controls. Overgrowth of *Enterobacteriaceae*
3 usually indicates a status of gut flora dysbiosis, which may lead to a series of chronic diseases,
4 such as colitis [21], Crohn's disease and acute cholecystitis [22]. In this study, the GDM-enriched
5 *Enterobacteriaceae* species include *Klebsiella variicola*, *Escherichia coli*, *Enterobacter cloacea*
6 and three *Citrobacter* MLGs (Table S3). Based on these findings, the association between
7 *Enterobacteriaceae* and GDM is intriguing, but of uncertain pathogenic significance. This
8 intriguing observation warrants further studies.

9
10
11
12
13
14
15
16
17 In the current study, microbes found to be enriched in the gut of control subjects included
18 *Methanobrevibacter smithii*, *Bifidobacterium* spp. (including *B. pseudocatenulatum*, *B. animalis*
19 and one unclassified MLG), *Eubacterium* spp. (*E. siraeum*, *E. eligens* and two unclassified
20 *Eubacterium* MLGs) (Tables S2 and S3). Previous studies of a variety of chronic diseases,
21 including type 2 diabetes [19], liver cirrhosis [23], Crohn's disease [24] and ulcerative colitis [25],
22 have also found these bacteria to be enriched in the gut microbiome of healthy control subjects.
23 This suggests that alterations in gut microbiota resulting in decreased relative abundance of these
24 species may be a considerable risk factor for multiple metabolic syndromes, including GDM.
25 Thus, it is conceivable that alterations in the microbiome associated with GDM, contribute to the
26 pathogenesis of GDM.

27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

1 risk or protective factor for GDM development.

2 The main limitations of our study are the relatively limited sample size, and the fact that we
3 only analyzed one stool sample per subject, collected in the second trimester of pregnancy. It is
4 well known that immune and metabolic changes occur throughout pregnancy, and that the gut
5 microbiota shifts from first to third trimesters [16]. Consequently, associations between the
6 microbiome and GDM status need to be examined at other time points during pregnancy to
7 provide further insights into when the changes we observed at 21-29 weeks develop, and whether
8 they are sustained for the remainder of the pregnancy. In addition, metadata information
9 available on the effect of maternal GDM status and changes in microbiota composition in
10 pregnancy were limited. Confounding factors such as life style, diet, and antibiotic
11 treatment may further affect both blood glucose levels and gut microbiota composition. In order to more
12 definitively establish the associations observed in the current study, a large cohort investigation,
13 with analysis of other potentially significant variables, will be necessary. Further, the
14 observational, cross-sectional design of the current study precluded examination of potential
15 causality. Whether the microbiome impacts blood glucose levels, glucose levels impact the
16 microbiome, or the relationship is bidirectional, cannot be proven without further experimentation,
17 most likely involving animal models. Furthermore, to identify fecal metagenomic markers with
18 sufficient predictive power to identify GDM, future work will be necessary to refine the diagnostic
19 approach developed in our study, to identify additional markers with improved predictive value
20 and, eventually, to validate them in other larger cohorts.

21 In summary, our findings extend findings of earlier studies showing a correlation between gut
22 microbiota and various metabolic derangements. Specifically, we demonstrated an important
23 associations between the gut microbiota and GDM. Our results suggest that changes in
24 composition of the gut microbiome may be used to identify individuals with GDM, could
25 potentially be used to identify individuals at risk for GDM, and may contribute to the pathogenesis
26 of GDM.

27 **Methods**

28 **Study population and sampling**

29 As part of the Born in Guangzhou Cohort Study (BIGCS), fecal samples were obtained from 124

1 pregnant women during their second trimester in Guangzhou Women and Children's Medical
2 Center (GWCMC). Pregnant women who had severe obstetric complications (GDM excluded)
3 such as pregnancy-induced hypertension, preeclampsia, or eclampsia, were excluded. Study
4 subjects had not received any antibiotic treatment within 1 month of sample collection and had not
5 ingested yogurt within 2 weeks of sample collection. Samples were frozen immediately (within 30
6 minutes) and transferred to -80 °C freezers until further analysis.

7
8
9
10
11
12
13 This study received approval from the Ethics Committee of GWCMC, and written informed
14 consent was obtained from all participating pregnant women. Eligible participants underwent a
15 standard 2h 75g oral glucose tolerance test (OGTT) between 21–29 weeks' gestation by collection
16 of 2ml blood samples fasting, 1h, and 2h after a 75g glucose load, using NaF/EDTA tubes. After
17 centrifugation, plasma glucose was measured by a hexokinase method using Beckman Coulter
18 AU5800 automatic analyzer (Beckman Coulter, California, US). The laboratory previously
19 achieved ISO15189 certification by China National Accreditation Service for Conformity
20 Assessment. GDM was defined using the Chinese diagnostic criteria [28], which is in agreement
21 with the one-step approach endorsed by the American Diabetes Association [29]. Following the
22 OGTT, pregnant women were diagnosed as having GDM if one or more of the following glucose
23 levels were elevated: fasting ≥ 5.1 mmol/L, 1h ≥ 10.0 mmol/L, and 2h ≥ 8.5 mmol/L [28]. None of
24 these women was treated with insulin or glyburide. Maternal age, pre-pregnancy weight and
25 height were extracted from obstetric records of the Hospital Information Systems (HIS) used in
26 GWCMC. Pre-pregnancy body mass index (PBMI) was calculated from height and weight
27 information. Pregnant women were included in this study included 43 patients with GDM
28 (average age 30.5 years; average PBMI 21.9 kg/m²) and 81 healthy controls (average age 29.1
29 years; average PBMI 20.2 kg/m²).

50 DNA extraction and metagenomic sequencing

51 Total bacterial DNA was extracted from about 100 mg of feces using Qiagen QIAamp DNA Stool
52 Mini Kit (Qiagen) following the manufacturer's instructions. Extracted DNA of each sample was
53 kept frozen at -20°C until used. Illumina HiSeq 2000 was used to sequence the samples. We
54 constructed a paired-end library with insert size of 350 base pairs (bp) for every sample, and
55 sequenced with 100bp read length from each end. Illumina sequencing reads for fecal samples

1 from pregnant women were independently processed for quality control using FASTAX Toolkit
2 (http://hannonlab.cshl.edu/fastx_toolkit/). Quality control used the following criteria: (1) reads
3 were removed if they contain more than 3 N bases or more than 50 bases with low quality (<Q20);
4 (2) reads were trimmed in the end with low quality (<Q20) or assigned as N. The remaining reads
5 were then mapped to the human genome using SOAPalinger2 [30] to remove contaminating
6 human DNA. After t QC, an average of 1.9% of low-quality or human genome reads were
7 removed for the 124 samples.

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


16 To determine the best assembling method for the obtained high-quality Illumina sequencing reads,
17 we compared the performance of two assemblers, SOAPdenovo v2.04 (as previously used in the
18 MetaHIT and IGC projects) [31, 32] and IDBA-UD v1.1.1 (a *de novo* assembler for metagenomic
19 sequences) [33]. For the SOAPdenovo, we tested the k-mer length ranging from 23bp to 123bp by
20 10bp step for each sample, and selected the assembled contig set with longest N50 length. For the
21 IDBA-UD, parameters “--mink 21 --maxk 81 --step 20 --pre_correction” were used. For most
22 samples, IDBA-UD obtained a better assembled contig set than SOAPdenovo. This could be
23 attributable to the relative efficiency of IDBA-UD in assembling bacterial genomes within regions
24 of highly uneven depth in metagenomic samples. As a result, we obtained an average 197.9 ± 50.3
25 Mbp (mean \pm SD) contig sets for each pregnant women sample, with N50 length 8.8 ± 3.9 kbp.
26 Unassembled reads from these samples were pooled and re-assembled by using IDBA-UD for
27 further analysis.

28 Genes were predicted by MetaGeneMark [34] based on parameter exploration by the MOCAT
29 pipeline [35]. A non-redundant gene catalogue of pregnant women samples was constructed using
30 CD-HIT [36], through which, genes with >90% overlap and >95% nucleic acid similarity (no gap
31 allowed) were removed as redundancies. A pregnant women gene catalogue containing 4,344,984
32 non-redundant genes was generated for fecal samples collected from these 124 pregnant women.
33 This gene catalogue was further combined with the previous integrated gene catalogue (IGC) [32]
34 by removing redundancies in the same manner as above; 2,621,398 genes were removed as
35 redundant. In the end, 39.6% (1,723,586) of the genes in the pregnant women gene catalogue were
36 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 [30], 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 [19].

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 with parameters “-word_size 16 -evaluate 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 [37], 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) and evolutionary genealogy of genes: Non-supervised Orthologous Groups (eggNOG, v4) databases were used for functional annotation of genes. Translated amino acid sequences of genes were searched against these databases using USEARCH v8.0.1616 [38] (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 [19]. Briefly, gene relative abundance profiles were initially adjusted for population stratifications using the modified EIGENSTRAT method [39] 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 [40] 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 [19]. 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 [19]. 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

1 their abundances in all GDM and control samples, and the co-occurrence network of MLGs was
2 visualized by Cytoscape 3.0.2 [41]. The direction of enrichment was determined by the
3
4 Mann-Whitney U test ($p < 0.05$).
5
6

7 **Statistical analysis**

8
9
10 Statistical analysis was implemented using the R platform. Distance-based redundancy analysis
11 (dbRDA) was performed using the “vegan” package [42] based on the Bray-Curtis distances on
12 normalized taxa abundance matrices, then visualized using the “ggplot2” package. Permutational
13 multivariate analysis of variance (PERMANOVA) was performed using the “vegan” package, and
14 the permuted p -value was obtained by 10,000 permutations.
15
16
17
18
19

20 The Random Forest model has been shown [6] to be a suitable model for exploiting
21 metagenomic data. Random Forest models were trained using the “randomForest” package
22 (default parameters and 10,000 trees) to identify GDM status in a subset of GDM patients and
23 control subjects by using the abundance profiles of species and MLGs. Performance of the
24 predictive model was evaluated with cross-validation error. Variable importance by mean decrease
25 in accuracy was calculated for the Random Forest models using the full set of species or MLGs.
26 By ranking the variables by importance, smaller models were constructed that contained only the
27 most important variables.
28
29
30
31
32
33
34
35
36

37 Receiver operator characteristic (ROC) analysis was performed using the “pROC” package, we
38 then computed the 95% confidence interval (CI) of the area under the ROC curve (AUC) with
39 10,000 bootstrap replicates to assess the variability of the measure. Rarefaction analysis was
40 performed to assess the gene richness of metagenomic samples, implemented by in-house Perl
41 scripts.
42
43
44
45
46
47
48

49 **Reference**

- 50
51 1. Ferrara A: **Increasing prevalence of gestational diabetes mellitus: a public health**
52 **perspective.** *Diabetes care* 2007, **30 Suppl 2**:S141-146.
- 53 2. Landon MB, Gabbe SG: **Gestational diabetes mellitus.** *Obstetrics and gynecology* 2011,
54 **118**(6):1379-1393.
- 55 3. Kim C, Newton KM, Knopp RH: **Gestational diabetes and the incidence of type 2 diabetes:**
56 **a systematic review.** *Diabetes care* 2002, **25**(10):1862-1868.
- 57 4. Allalou A, Nalla A, Prentice KJ, Liu Y, Zhang M, Dai FF, Ning X, Osborne LR, Cox BJ,
58
59
60
61

- 1 Gundersen EP *et al*: **A Predictive Metabolic Signature for the Transition From**
2 **Gestational Diabetes Mellitus to Type 2 Diabetes.** *Diabetes* 2016, **65**(9):2529-2539.
- 3 5. Le Chatelier E, Nielsen T, Qin J, Prifti E, Hildebrand F, Falony G, Almeida M, Arumugam M,
4 Batto JM, Kennedy S *et al*: **Richness of human gut microbiome correlates with metabolic**
5 **markers.** *Nature* 2013, **500**(7464):541-546.
- 6 6. Karlsson FH, Tremaroli V, Nookaew I, Bergstrom G, Behre CJ, Fagerberg B, Nielsen J,
7 Backhed F: **Gut metagenome in European women with normal, impaired and diabetic**
8 **glucose control.** *Nature* 2013, **498**(7452):99-103.
- 9 7. Karlsson F, Tremaroli V, Nielsen J, Backhed F: **Assessing the human gut microbiota in**
10 **metabolic diseases.** *Diabetes* 2013, **62**(10):3341-3349.
- 11 8. Cani PD, Neyrinck AM, Fava F, Knauf C, Burcelin RG, Tuohy KM, Gibson GR, Delzenne
12 NM: **Selective increases of bifidobacteria in gut microflora improve high-fat-diet-induced**
13 **diabetes in mice through a mechanism associated with endotoxaemia.** *Diabetologia* 2007,
14 **50**(11):2374-2383.
- 15 9. Chu SY, Callaghan WM, Kim SY, Schmid CH, Lau J, England LJ, Dietz PM: **Maternal**
16 **obesity and risk of gestational diabetes mellitus.** *Diabetes care* 2007, **30**(8):2070-2076.
- 17 10. Hedderson MM, Williams MA, Holt VL, Weiss NS, Ferrara A: **Body mass index and weight**
18 **gain prior to pregnancy and risk of gestational diabetes mellitus.** *American journal of*
19 *obstetrics and gynecology* 2008, **198**(4):409 e401-407.
- 20 11. Pedersen HK, Gudmundsdottir V, Nielsen HB, Hyotylainen T, Nielsen T, Jensen BA, Forslund
21 K, Hildebrand F, Prifti E, Falony G *et al*: **Human gut microbes impact host serum**
22 **metabolome and insulin sensitivity.** *Nature* 2016, **535**(7612):376-381.
- 23 12. Collado MC, Isolauri E, Laitinen K, Salminen S: **Distinct composition of gut microbiota**
24 **during pregnancy in overweight and normal-weight women.** *The American journal of*
25 *clinical nutrition* 2008, **88**(4):894-899.
- 26 13. Vrieze A, Van Nood E, Holleman F, Salojarvi J, Kootte RS, Bartelsman JF, Dallinga-Thie GM,
27 Ackermans MT, Serlie MJ, Oozeer R *et al*: **Transfer of intestinal microbiota from lean**
28 **donors increases insulin sensitivity in individuals with metabolic syndrome.**
29 *Gastroenterology* 2012, **143**(4):913-916 e917.
- 30 14. Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, Gordon JI: **An**
31 **obesity-associated gut microbiome with increased capacity for energy harvest.** *Nature*
32 2006, **444**(7122):1027-1031.
- 33 15. England LJ, Dietz PM, Njoroge T, Callaghan WM, Bruce C, Buus RM, Williamson DF:
34 **Preventing type 2 diabetes: public health implications for women with a history of**
35 **gestational diabetes mellitus.** *American journal of obstetrics and gynecology* 2009,
36 **200**(4):365 e361-368.
- 37 16. Koren O, Goodrich JK, Cullender TC, Spor A, Laitinen K, Backhed HK, Gonzalez A, Werner
38 JJ, Angenent LT, Knight R *et al*: **Host remodeling of the gut microbiome and metabolic**
39 **changes during pregnancy.** *Cell* 2012, **150**(3):470-480.
- 40 17. Gomez-Arango LF, Barrett HL, McIntyre HD, Callaway LK, Morrison M, Dekker Nitert M,
41 Group ST: **Connections Between the Gut Microbiome and Metabolic Hormones in Early**
42 **Pregnancy in Overweight and Obese Women.** *Diabetes* 2016, **65**(8):2214-2223.
- 43 18. Wang J, Jia H: **Metagenome-wide association studies: fine-mining the microbiome.** *Nature*
44 *reviews Microbiology* 2016, **14**(8):508-522.

19. 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.
20. Lowe WL, Jr., Scholtens DM, Sandler V, Hayes MG: **Genetics of Gestational Diabetes Mellitus and Maternal Metabolism.** *Current diabetes reports* 2016, **16**(2):15.
21. Garrett WS, Gallini CA, Yatsunencko 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.
22. 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.
23. 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.
24. 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.
25. Machiels K, Joossens M, Sabino J, De Preter V, Arijis 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.
26. 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.
27. Shetty SA, Marathe NP, Lanjekar V, Ranade D, Shouche YS: **Comparative genome analysis of Megasphaera sp. reveals niche specialization and its potential role in the human gut.** *PloS one* 2013, **8**(11):e79353.
28. Zhu WW, Yang HX: **Diagnosis of gestational diabetes mellitus in China.** *Diabetes Care* 2013, **36**(6):e76.
29. Association. AD: **Standards of medical care in diabetes--2011.** *Diabetes Care* 2011, **34** Suppl 1:S11-61.
30. 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.
31. 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.
32. 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.
33. 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.
34. Zhu W, Lomsadze A, Borodovsky M: **Ab initio gene identification in metagenomic**

sequences. *Nucleic acids research* 2010, **38**(12):e132.

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
35. 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.
36. 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.
37. 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.
38. Edgar RC: **Search and clustering orders of magnitude faster than BLAST.** *Bioinformatics* 2010, **26**(19):2460-2461.
39. 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.
40. 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.
41. 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.
42. 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.

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.

This study is supported by the National Natural Science Foundation of China (81673181) and Guangzhou Science and Technology Bureau, Guangzhou, China (201508030037). 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.

Author contributions

XQ and HX designed the birth cohort on which this study was based. XQ and HX designed the study and directed its implementation. YG, YK, MY, JH, JL, NC, WX, SS, LQ, YW, CH, QC, WL and YW were involved in study design and sample collection. YG, YK 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.

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 coloured circles cover the samples near the center of gravity for each group. (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% of total abundance in all 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

1 IQR from the first and third quartiles, respectively. The circles represent outliers beyond the
2 whiskers.
3

4
5
6 **Figure 2 | Interconnection of GDM-associated MLGs.** A co-occurrence network deduced from
7 GDM-enriched (red circle) and control-enriched (green circle) MLGs is shown. Nodes depict
8 MLGs with their ID displayed in the center. The size of each node indicates the number of genes
9 within the MLG. Connecting lines represent Spearman correlation coefficient $\rho > 0.40$ (represented
10 by gray line) or < -0.40 (represented by red line). Only MLGs with > 200 genes or $> 0.1\%$ relative
11 abundance are shown for clarity of presentation and visualization. GDM-enriched MLGs for the
12 Enterobacteriaceae group is labeled.
13
14
15
16
17
18
19
20
21
22

23 **Figure 3 | Association of gross abundance of GDM-enriched and control-enriched MLGs**
24 **with blood glucose levels 0, 60, and 120 minutes after an oral glucose tolerance test.** Scatter
25 plots of samples are shown with lines indicating linear fit.
26
27
28
29
30

31 **Figure 4 | Correlation of blood glucose levels 0, 60, and 120 minutes after an oral glucose**
32 **tolerance test, and HbA1c levels with MLGs (a) and species (b).** Spearman's rank correlation
33 coefficients and P-values for the correlations are shown. '+' denotes $P < 0.05$; '++' denotes $P < 0.01$.
34 Only MLGs or species with average relative abundances greater than 0.001% and correlated
35 ($P < 0.05$) with at least one index are shown for clarity.
36
37
38
39
40
41
42
43

44 **Figure 5 | Association of microbial genetic functional pathway composition in GDM patients**
45 **and healthy pregnant women.** (a) Distributions of relative abundances of KEGG pathway
46 categories in GDM patients and healthy controls. '*' denotes $q < 0.05$ (Mann-Whitney U test
47 corrected by the Benjamini-Hochberg method) (b) Correlation of blood glucose levels 0, 60, and
48 120 minutes after an oral glucose tolerance test, and HbA1c levels, with PTS system and LPS
49 biosynthesis and transport system. Spearman's rank correlation coefficients and P-values for the
50 correlations are shown. '+' denotes $P < 0.05$; '++' denotes $P < 0.01$.
51
52
53
54
55
56
57
58
59

60 **Figure 6 | Classification of GDM status by the relative abundance of MLGs and species.** (a)
61

1 Classification performance of a random forest model using MLG or species abundance assessed
2 by AUC. The performance was explored for different numbers of explanatory variables, ordered in
3 importance. (b-c) The 30 most discriminant MLGs (b) and species (c) in the models classifying
4 GDM and controls. The bar lengths in b and c indicate the importance of the variable, and colors
5 represent enrichment in GDM (red shades) or controls (blue shades). (d) ROC analysis for
6 classification of GDM status by MLGs and PBMI.
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1 **Additional material**

2 **Additional file1:**

3
4 File format: .xlsx

5
6 Table S1: **Phenotype information of 43 GDM and 81 healthy pregnant women**

7
8 Table S2: **Bacterial species that differed significantly between two cohorts.**

9
10 Table S3: **Detailed information of 129 GDM-associated MLGs**

11
12
13
14
15
16 **Additional file 2:**

17
18 File format: .docx

19
20
21
22
23 **Figure S1: Density histogram showing the P-value distribution between GDM patients and**
24 **healthy pregnant women for all genes tested.** The horizontal line represents the expected
25 distribution of P-values, and the π_0 value indicates the proportion of genes under the null
26 hypothesis.

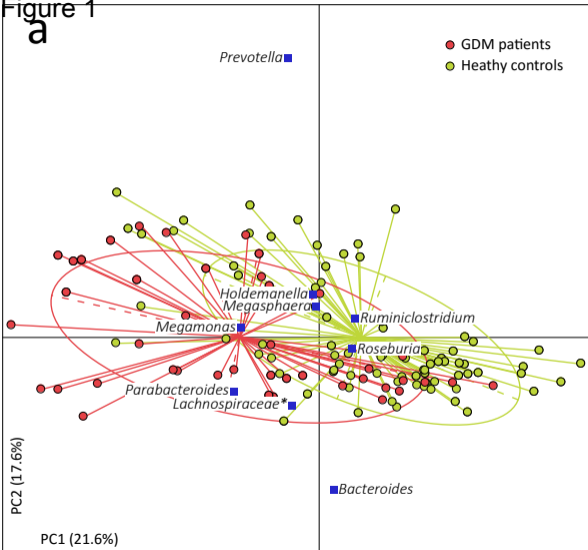
27
28
29
30 **Figure S2: Figure S2 | Correlation between Enterobacteriaceae relative abundance and**
31 **PBMI.** Scatter plots of samples are shown with lines indicating linear fit.

32
33
34
35 **Figure S3: Figure S3 | Classification of GDM status by abundance of MLGs and PBMI.**

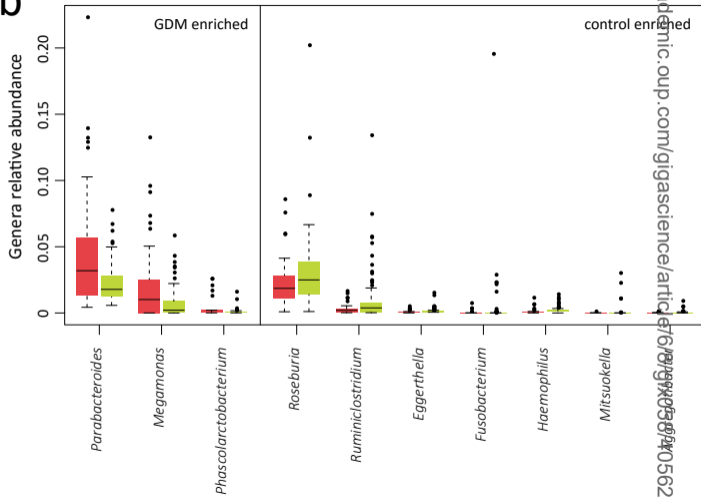
36
37 The 30 most discriminant MLGs or PBMI in the models for classifying GDM and controls. The
38 bar lengths indicate the importance of the variable, and colors represent enrichment in GDM (red
39 shades) or controls (blue shades)

Figure 1

a



b



[Click here to download Figure Figure 1.pdf](#)

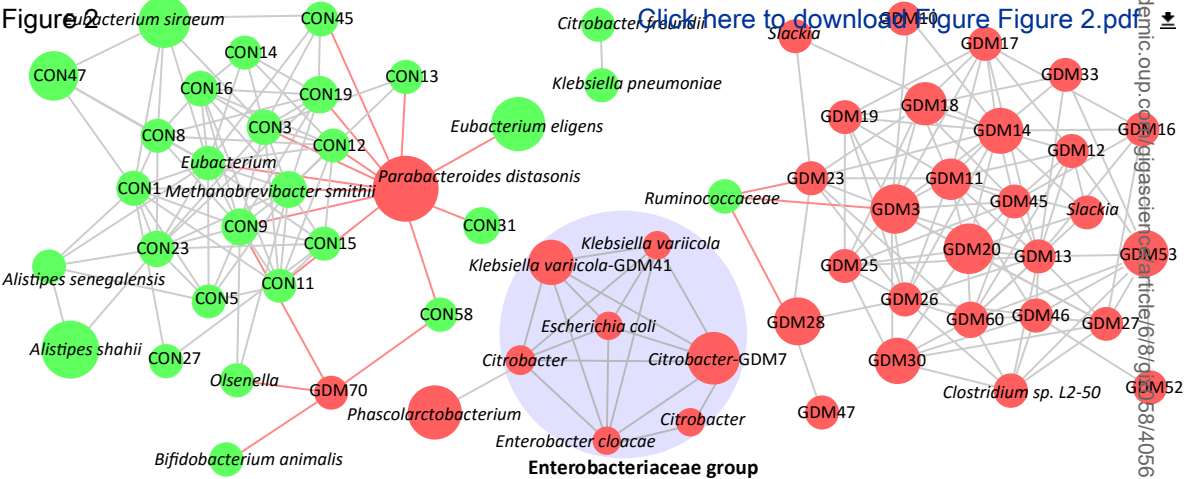
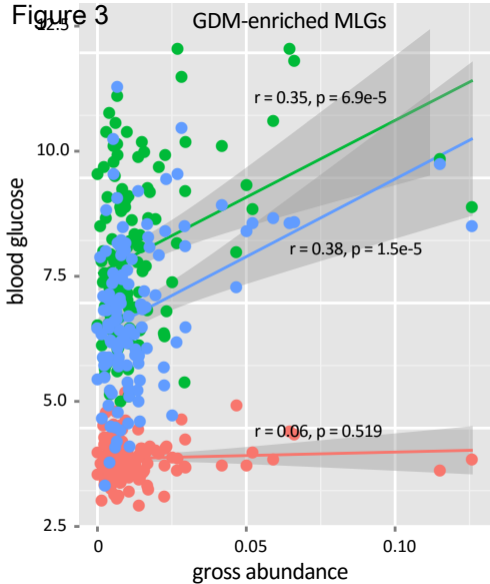


Figure 2

Figure 3



[Click here to download Figure 3.pdf](#)

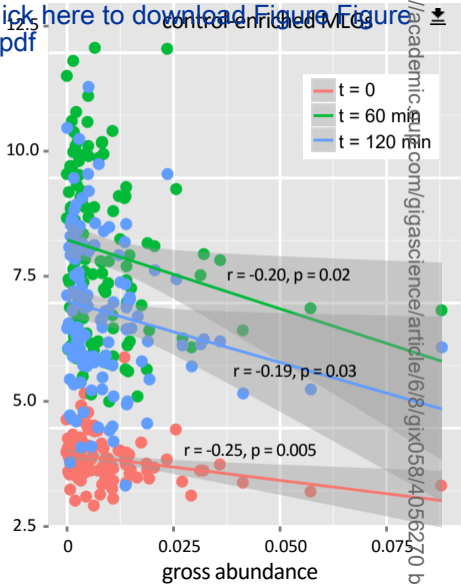
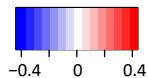
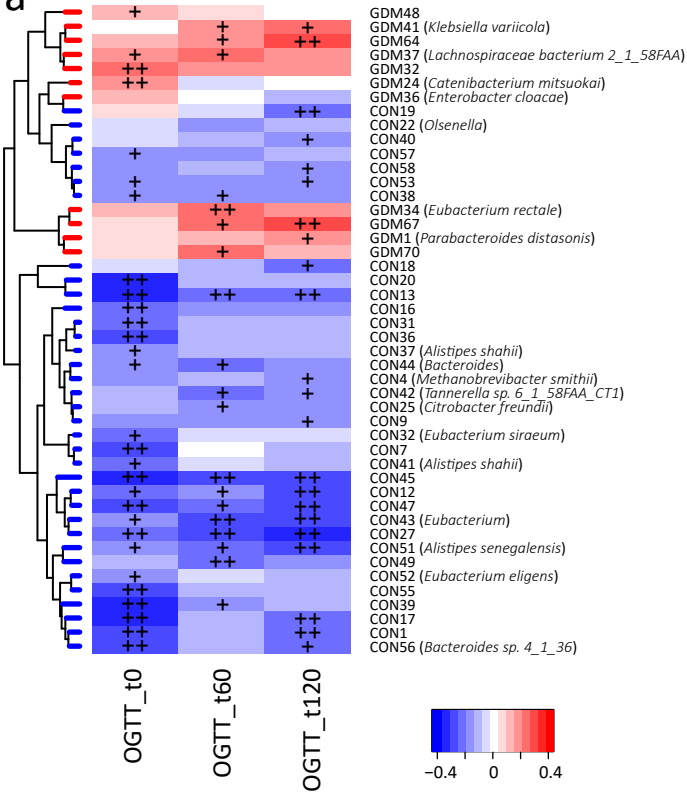
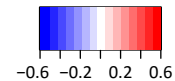
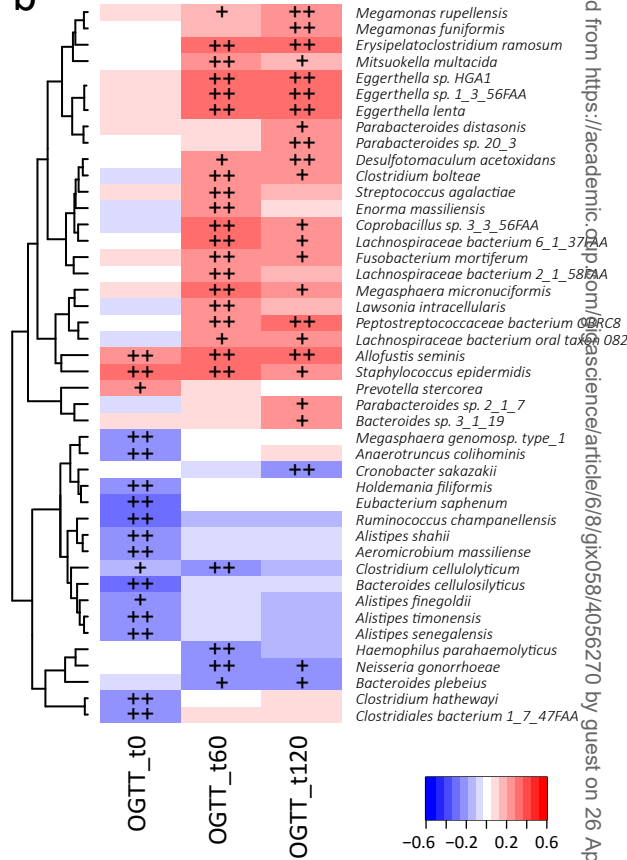


Figure 4

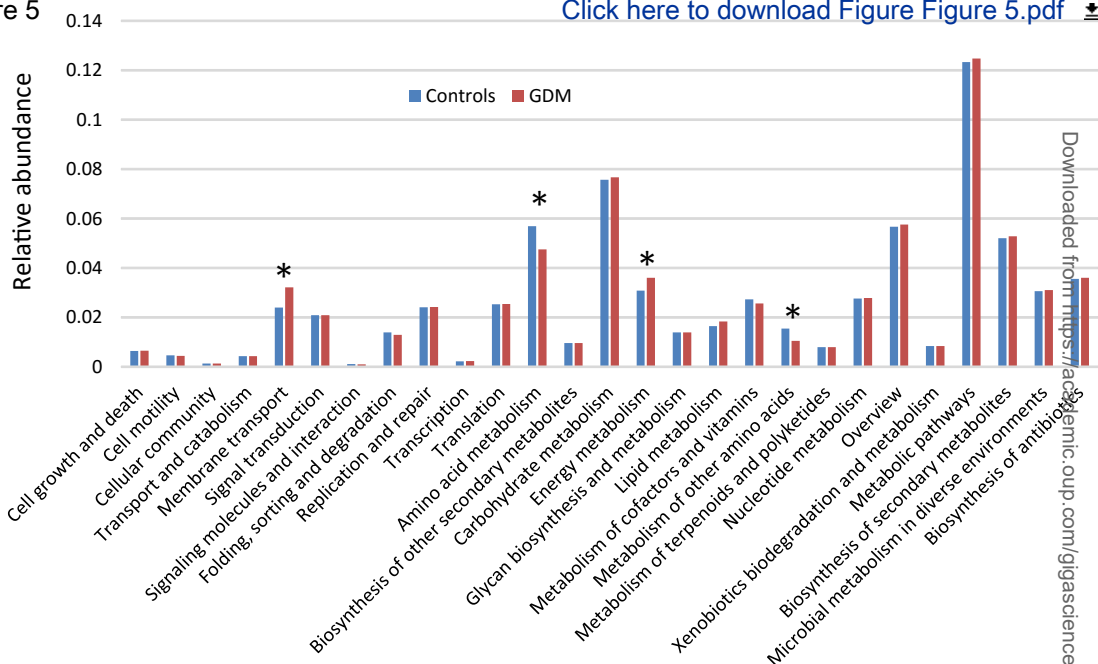


b

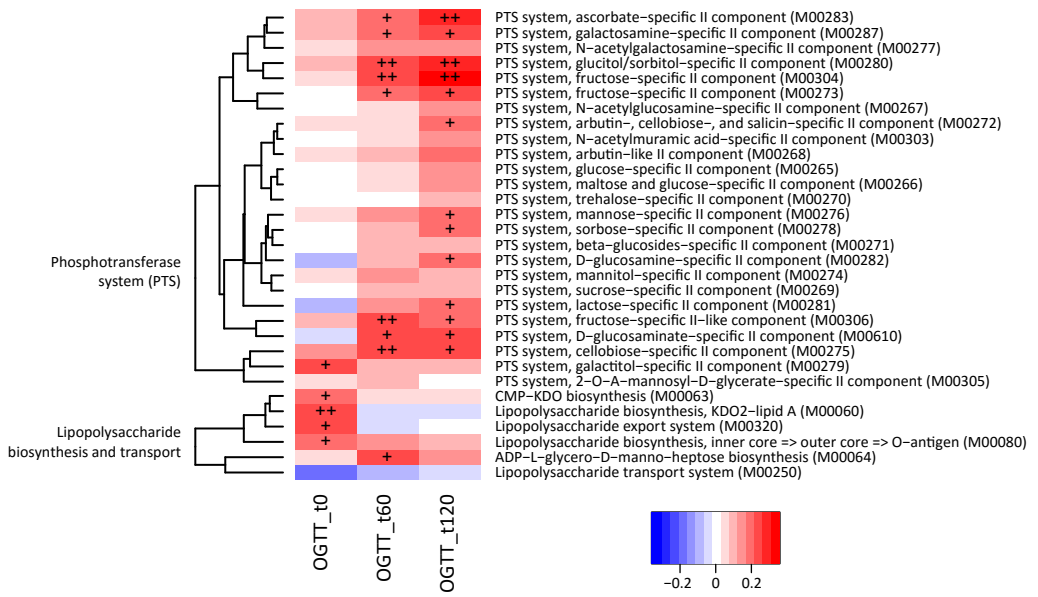
Click here to download Figure Figure 4.pdf



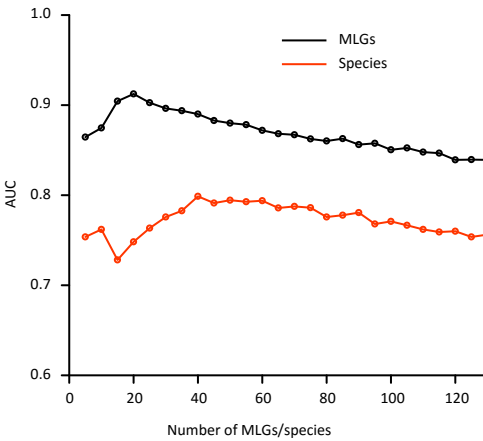
a



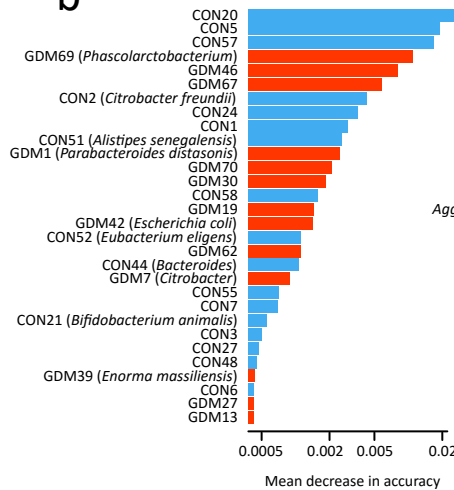
b



a Figure 6

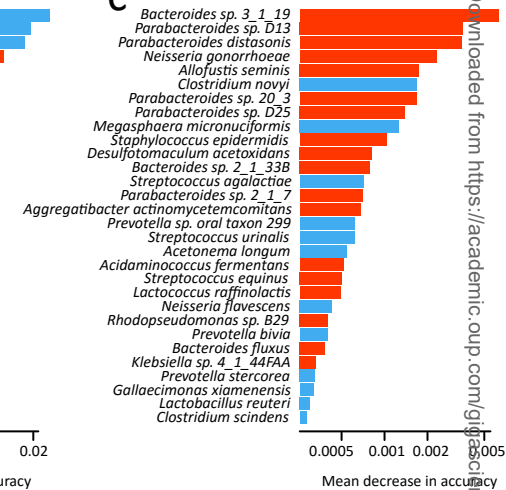


b

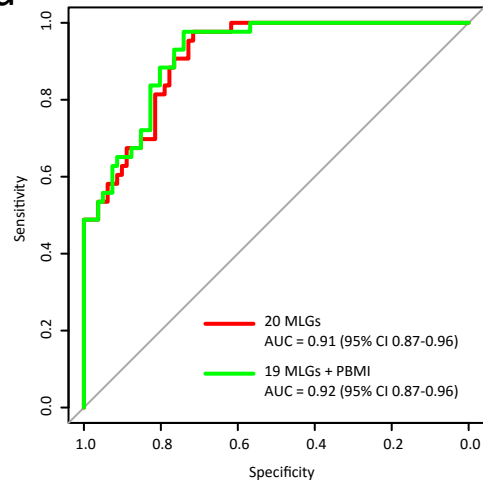


Click here to download Figure Figure 6.pdf


c




d



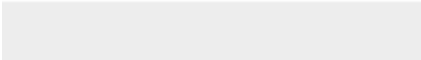

Downloaded from https://academic.oup.com/gib/advance-article/doi/10.1093/gib/giab001/6311001 by guest on 26 April 2025



Click here to access/download
Supplementary Material
Additional file 1.xlsx



Click here to access/download
Supplementary Material
Additional file2.docx



Dear Dr. Laurie Goodman,

Enclosed please find our manuscript entitled “**Connections between human gut microbiome and gestational diabetes mellitus**” for your consideration of publication as an Article in *GigaScience*.

Emerging evidence has revealed a link between the gut microbiome and human metabolic health. The gut microbiome can modulate metabolic health and affect insulin resistance, and may play an important role in the etiology of gestational diabetes mellitus (GDM). However, studies focusing on changes in the gut microbiome during pregnancy and development of GDM have not been reported so far.

In the present study, we performed 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 (43 GDM patients versus 81 healthy control subjects). The objective was to obtain a comprehensive understanding of the gut microbiome’s role in the etiopathogenesis of GDM.

Our data showed that GDM patients had significantly higher abundance of *Parabacteroides*, *Megamonas* and *Phascolarctobacterium*. Metagenome-wide association study (MGWAS) identified 154,837 genes with significant abundance differences between GDM patients and healthy pregnant women. Furthermore, the GDM microbiome showed greater abundance of membrane transport and energy metabolism pathways, involving the phosphotransferase system, and the lipopolysaccharide biosynthesis and export system, that may contribute to the potential relationship between gut microbes and blood glucose tolerance. Our data also imply that fecal metagenome linkage groups may provide a mechanism and biomarkers for early detection of GDM or, potentially, for identifying risk of developing GDM.

We believe that findings from our study represent an novel relationship between gut microbiome composition and GDM status, by which the gut microbiome may potentially be used to identify individuals at risk for GDM, and may contribute to the pathogenesis of GDM. And as such should

be of great interest to the general medical readership of your journal.

All authors have reviewed and approved the final version. Neither this manuscript, nor one with similar content, has been published or is being considered for publication elsewhere. None of the authors have conflict of interest with regard to the data presented in this manuscript.

Thank you for your time and consideration. We look forward to further discussion regarding this manuscript.

Respectfully Yours,

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