Supplementary Methods

Contents

- Primary cohort sample collection
- GDM diagnosis method
- Additional cohorts (for FMT experiments)
- Serology: primary cohort
- Microbiome sequencing and pre-processing
- Short-chain fatty acids profiling
- Untargeted metabolomics
- FMT into germ-free mice
    - FMT for additional cohorts
- Serology: FMT with primary cohort samples
- Prediction
- Statistical analysis
- Ethics statement
- Patient and public involvement

Primary cohort sample collection

Blood samples (15ml) were collected in EDTA tubes from the primary cohort at recruitment. The tubes were centrifuged at 4°C for 10 min at 3600 rpm; plasma was then stored in coded 1.7 ml tubes at -80°C until processing. Fecal samples from this cohort were collected as previously published [1], from participants at 11-14 gestational weeks, close to recruitment, and frozen immediately at -80°C until processing. These samples were used to profile the gut microbiome composition, short-chain fatty acids (SCFAs) and the metabolome.

GDM diagnosis method

In Israel, GDM diagnosis is done with the 2-step procedure [2]: glucose challenge test (GCT) and an oral glucose tolerance test (OGTT). GCT screening is universal at 24-28 weeks and followed by a diagnostic OGTT if GCT>140mg/dl. Sometimes for women with risk factors (BMI, FPG, Past GDM) the OGTT, without GCT, is done late in the first trimester. In our study no participants underwent this testing prior to providing samples. Exact timing of GDM diagnosis does not affect the

findings of our study, and our microbiota, metabolome, and inflammation markers provide an earlier window into disease risk than current diagnostic practices.

Additional cohorts

In addition to the primary and secondary cohorts, for fecal microbiome transplant (FMT) experiments in germ-free (GF) models, described below, fecal samples from pregnant women in T1 who would and would not go on to develop GDM from two additional independent cohorts were included: samples from a Finnish cohort [3–5] and the Stanford Outcomes Research in Kids (STORK) study [6]. The subjects in the Finnish cohort were recruited during the first trimester of pregnancy to a randomized, controlled trial (ClinicalTrials NCT00167700) assessing the impact of dietary counseling and a probiotic intervention on various maternal and infant outcomes [3–5]. Pregnant women with chronic disease including metabolic abnormalities were excluded from the study. Fecal samples were collected at recruitment during T1 prior to intervention, and again during the third trimester (T3; the latter were not used in this study). Consequently, T1 samples were collected from metabolically healthy women. GDM was diagnosed by a two-hour OGTT, which was performed at 24-28 weeks of gestation in subjects at increased risk including women with excessive weight gain, age >40 years, glucosuria, increased fetal growth or a history of GDM or macrosomic newborn(s) in previous pregnancies. GDM was diagnosed if plasma glucose concentration was ≥4.8 mmol/L at baseline, ≥10.0 mmol/L at 1 h or ≥8.7 mmol/L at 2 h during the OGTT. The STORK study is a multiethnic birth cohort from California [6]; for this study, a total of 6 mothers diagnosed with GDM in T1 were matched on gestational age, maternal age and race/ethnicity to 6 controls with fecal samples collected between 10.5 to 23 (mean 16) weeks of pregnancy.

Serology: primary cohort

Cytokine (TNF-α, IFN-γ, GM-CSF, IL-2, IL-4, IL-6, IL-8, IL-10) and hormone (insulin and leptin) levels were measured in plasma using the Bio-Plex Pro Human Cytokine 8-Plex Panel (Bio-Rad Laboratories Inc., Irvine, CA, USA) according to the manufacturer's instructions. Serum was not available for all women in the GDM group such that 35 of the 44 pre-GDM women (and 78 controls) were included in this analysis.

Microbiome sequencing and pre-processing

DNA was extracted using the PowerSoil DNA extraction kit (MoBio, Carlsbad, CA, USA) according to the manufacturer's instructions and following a 2-minute bead beating step (BioSpec, Bartlesville, OK, USA). Purified DNA was used for PCR amplification of the variable V4 region using the 515F and 806R barcoded primers following the Earth Microbiome Project protocol [7]. For each PCR reaction, the following materials were added: 4µl (~40ng/ µl) DNA (sample), 2 µl 515F (forward, 10µM) primer, 2 µl 806R (reverse, 10µM) primer, and 25 µl PrimeSTAR Max PCR Readymix (Takara, Mountain View, CA, USA). PCR reactions were carried out by 30 cycles of denaturation at 98°C for 10 seconds, annealing at 55°C for 5 seconds, extension at 72°C for 20 seconds and then a final elongation at 72°C for 1 minute. Amplicons were purified using AMPure magnetic beads (Beckman Coulter, Brea, CA, USA) and quantified using the Picogreen dsDNA quantitation kit (Invitrogen, Carlsbad, CA, USA). Then, equimolar amounts of DNA from individual samples were pooled and sequenced using the Illumina MiSeq platform and MiSeq Reagent Kit V2 (500 cycles) at the Genomic Center at the Bar-Ilan University Azrieli Faculty of Medicine, Israel.

Microbial diversity and composition were assessed using QIIME2 version 2019.4 [8] First, single-end sequences were imported (qiime import) and demultiplexed (qiime demux) with golay error correction. Next, sequences were denoised using DADA2 [9] (qiime dada2 denoise-single), trimming the first 5 bases and truncating each sequence at position 215. Feature tables and representative sequences from the different sequencing runs were then merged. A phylogenetic tree was constructed using the fragment-insertion method (qiime fragment-insertion sepp [10]). Taxonomic classification was done using a naïve-based classifier trained on the 99% Greengenes 13_8 V4 reference set [11] (qiime feature-classifier classify-sklearn [12]). In order to remove low-confidence features, only features with a frequency higher than 50 in at least 5 samples were kept. In addition, features that contained mitochondria or chloroplast sequences or that were not assigned to a phylum were filtered out. Data were then rarefied to 8,000 (human) or 16,000 (mouse) sequences per sample (qiime diversity core-metrics-phylogenetic).

Short-chain fatty acids profiling

Short chain fatty acid (SCFA) extraction and analysis was performed at the MIGAL Galilee Research Institute, Israel. An aliquot of 0.25 gr of wet feces from 20

age matched pre-GDM and control pairs was thawed and suspended in 1 ml of an orthophosphoric acid solution (8% v/v) and kept at room temperature for 10 min with occasional shaking. The mixture was homogenized for 2 min, and the suspension was centrifuged at 4°C for 15 min at 14,000 rpm. The supernatant was filtered by additional centrifugation at 4°C for 15 min at 14,000 rpm. Next, 225 µl of the supernatant were transferred into a polypropylene tube, and 25 µl of 2-methyl-butyric-acid (Sigma-Aldrich (Merck), St. Louis, MO, USA) were added as an internal standard (IS) to a final concentration of 0.001M and transferred to a chromatographic vial for gas chromatography analyses. The IS was used to correct for injection variability between samples and for minor changes in the instrument response. Vials were stored at -20°C before GC analysis. A standard mix (WSFA-4, Sigma-Aldrich, St. Louis, MO, USA) was used to determine the concentrations of propionic acid. Standard curves for acetic acid and butyric acid (Sigma-Aldrich, St. Louis, MO, USA) were prepared using stock solutions of both acids, separately.

Gas chromatography analysis was then performed. Chromatographic analyses were carried out using the Agilent Technologies 6890, a GC system with a mass selective detector. A fused-silica capillary column with a free fatty acid phase (DB-FFAP 122-3232, 30 m×0.25 mm×0.25 um) was used. The carrier gas was helium at a flow rate of 13.6 mL/min. The initial oven temperature was 70°C, raised to 100°C at a rate of 20°C/min, then raised to 180°C at 8°C/min and held for 3 min, before then being raised to 230°C at 20°C/min. The injection volume was 1 µL and the run time of a single analysis was 17 min.

Untargeted metabolomics

Untargeted metabolomics was performed at the MIGAL Galilee Research Institute, Israel and Tel Hai College, Israel on fecal samples from 15 pairs of BMI- and age-matched women who would and would not go on to develop GDM. Fecal samples were extracted using methanol (0.333 mg/ml of MeOH), vortexed, and centrifuged. The supernatant was collected and filtered before injection to the LC-MS/MS instrument. A pooled matrix prepared by mixing a small volume (20 µl) of each experimental sample was used as a quality control (QC) for batch normalization and compound identification.

The samples were injected (5 µL) into UHPLC connected to a photodiode array detector (Dionex Ultimate 3000, Thermo Fisher Scientific, Sunnyvale, CA, USA), with

a reverse-phase column (ZORBAX Eclipse Plus C18; Agilent, Santa Clara, CA, USA; 100*3.0 mm; 1.8 µm). The mobile phase consisted of (A) DDW with 0.1% formic acid and (B) acetonitrile containing 0.1% formic acid. The gradient was initiated with 2% B which was increased to 30% B over 4 min, and then increased to 40% B over 1 min before being kept isocratic at 40% B for another 3 min. Then, the gradient increased to 50% over 6 min, and to 55% over another 4 min and to 95% over 5 min and kept isocratic for 7 min. Finally phase B was returned to 2% over 3 min and the column was allowed to equilibrate at 2% B for 3 min before the next injection. The flow rate was 0.4 mL/min. Blank (methanol) and QC samples were injected at the start of the sequence, after every 10 samples, and at the end of the sequence.

LC–MS/MS analysis was performed with a Heated Electrospray ionization (HESI-II) source connected to a Q Exactive™ Plus Hybrid Quadrupole-Orbitrap™ Mass Spectrometer, Thermo Scientific™, Germany. ESI capillary voltage was set to 3500 V, capillary temperature to 300°C, gas temperature to 350°C and gas flow to 10 mL/min. The mass spectra (m/z 100–1500) were acquired using both positive and negative ion modes. Data dependent MS2 analysis was generated for the QC samples and used for compound identification. Downstream analysis and data processing were performed with the Thermo Scientific™ Compound Discoverer™ program, version 3.1.0.305 (mass tolerance ≤ 5ppm; intensity tolerance≤ 30%; S/N threshold= 3; minimum peak intensity=1,000,000; RT tolerance≤0.2min). Databases used for identification were Chemspider [13], MzCloud [14] and KEGG [15].

FMT into germ-free mice

Transplantation experiments were performed using fecal samples from the primary prospective cohort and the two additional cohorts (described in the supplementary methods). First, for samples from the primary cohort, we used the model of fecal gut microbiome transplants to germ-free mice as conducted previously [1,16,17]. Briefly, germ-free (GF) female Swiss Webster mice (8 weeks old for the Israeli cohort; other FMT experiments detailed below) were maintained in isolators under a strict 12h light:12h dark cycle with estrous cycles synchronized to minimize mouse hormonal variation. Mice were fed an autoclaved chow diet (Harlan-Teklad, Madison, WI) ad libitum. Stool samples from T1 pregnant women who were and were not later diagnosed with GDM were selected based on age- and BMI-matching without a priori knowledge of bacterial diversity. Fecal matter (0.1 g) was

suspended in 1.5 ml of reduced sterile PBS, vortexed for 5 min and settled for 5 min to allow larger particles to settle to the bottom of the tube. Handling of human fecal samples was performed under anaerobic conditions. Mice were divided into two groups with equal weights and then immediately gavaged with 200 µl of fecal slurries from the 2 study groups. Each fecal slurry was gavaged into a single mouse and the mice were then placed in ventilated cages, 3-4 mice per cage (divided by treatment group) and followed for 4 weeks.

Body weight and chow consumption were monitored weekly. Fecal pellets were collected on days 7, 14 and 21, snap-frozen in liquid N_2 and stored at −80°C for analysis of microbial communities. On day 21, intraperitoneal glucose tolerance test (ipGTT) was performed by an injection of 2 g/kg body weight glucose after an 8 h fast. Tail blood samples were collected at 0, 15, 30, 60, 90, and 120 minutes and blood glucose levels determined. On day 29, mice were sacrificed, and blood samples and ceca were collected.

FMT for additional cohorts

FMT experiments using samples from the additional cohorts were performed in the Cornell University animal facility. Deviations from the above protocol, used with primary cohort samples, are outlined here: For the Finnish cohort, two sets of experiments were performed. PGD1 study: stool samples were obtained in T1 from 6 women diagnosed with GDM matched to 6 healthy controls from Finland [18]. Twelve 6-8 week old female GF Swiss Webster mice were gavaged with stool sample slurries prepared under anoxic conditions as previously described. Here, an OGTT was administered 12 days post inoculation. Glucose dose was 2g/kg; readings were at 0, 30, 60, 120 minutes, via ACCU CHEK Compact Plus (Roche, Mannheim, Germany). PGD2: repeat of PGD1 with 12 mice aged 6-8 weeks and 12 mice aged 11-13 week old mice (total of 48). OGTT was performed on day 19.

For the STORK cohort, PGD3, 8 week old mice and 12 week old mice were gavaged with stool collected on gestational age of 16 weeks. As above, each donor sample was inoculated into one 8 week old and one 12 week old female mouse. An OGTT was administered 12 days post inoculation. Glucose dose was 2g/kg; readings were at 0, 30, 60, 120 minutes, as above.

Serology: FMT with primary cohort samples

Mouse cytokine (IL-1b, IL-6, IL-10, IL-17A, IFN-γ, TNF-α) and hormone (insulin and leptin) levels were measured in plasma (see experimental design below) of mice transplanted with samples from the primary cohort using the BioPlex Mouse Cytokine 8-plex Immunoassay (Bio-Rad Laboratories Inc., Irvine, CA, USA) according to the manufacturer's instructions. The fluorescent signals were measured on a Bio-Plex MAGPIX Multiplex Reader (Bio-Rad Laboratories Inc., Irvine, CA, USA). Analyte concentrations were calculated using standard curves in the Bio-Plex Manager Software. Values out of range (below/above) were imputed with the minimal/maximal in range values respectively.

Prediction

To predict GDM, we developed a prediction model using our prospective cohort based on all T1 information. We checked each combination of the following components: 1) cytokines, 2) microbiome, 3) general clinical information and 4) food questionnaires (15 total combinations). The accuracy of the prediction was assessed using the Area Under Curve (AUC) of the test set, in a 20%/80% test/training set division and a five fold cross validation.

The microbiome was merged into a genus level representation, log transformed and merged using the standard parameters of the MIPMLP pipeline [19]. For the other components, all non-numerical values were replaced by a one-hot representation. All missing values were replaced by the median value as of the same category. All values were z-scored to an average of 0 and a standard deviation of 1.

We used a binary XGBoost [[20] with a learning rate of 0.001, 200 estimators, gblinear classifiers, a logistic loss function, a lambda regularization of 0.01, and gamma regularization of 0.1 with the XGBclassifier function. All other parameters were the default of the function. The binary outcome was whether the woman later developed GDM. When combining different types of inputs for the classification, the inputs were concatenated.

We limited the external feature analysis to features informative on the training set in the first cross validation (Pearson correlation in the training test with the outcome of p value <0.1). The resulting feature used were:

Smoking (Yes/No/Past); BMI; FGT T1 (fasting glucose test, 1st trimester); Sleeping hours T1; GOT_T1 (serum glutamic-oxaloacetic transaminase); GPT T1 (serum

glutamate-pyruvate transaminase); PAPP-A[mU/L](Pregnancy-associated plasma protein A); Aspirin T1 (binary); medications_T1 (chronic medications).

When performing a two-step approach, we first predicted GDM using only the external features above. We then performed a microbiome only classification on the entire test set of the first stage classification. This resulted in two scores. One for the clinical features and one for the microbiome, further denoted: S(Clinical) and S(Microbiome). Samples with a low S(Clinical) or S(Microbiome) value were defined as negative. Samples with high scores for both S(Clinical) and S(Microbiome) values were defined as positive.

To examine the generalizability of our model, we applied the XGBoost classifier to an independently published 16S rRNA dataset from a cohort in China [21]. Briefly, this cohort included 98 pairs of pregnant women with and without GDM (matched) that provided a fecal sample in week 10-15 of pregnancy. We applied the same hyperparameters that were used on the primary cohort. The processing of the microbial data was similar to that mentioned above. Further, we built a model based on the intersection of bacteria between the two cohorts (67 shared microbes at the genus level). We trained the model on the main cohort and tested the model performance on the Chinese cohort. To improve the performance of the transfer learning, we used the iMic model [22]. The iMic model uses the taxonomy structure of the microbiome to translate the microbiome into images. Then convolutional neural networks (CNNs) are applied to the images. For the cross validity with iMic, we used the data at the species level.

Statistical analysis

Demographic and lifestyle characteristics of women with and without GDM were compared using t-tests or Fisher's Exact tests as appropriate; serum levels of cytokines and hormones, concentrations of short chain fatty acids were compared using the non-parametric Mann-Whitney U test followed by FDR correction as implemented in the scipy stats library of python [23].

Microbiome data was preprocessed (see above) and then diversity analysis was performed. Differences in alpha-diversity (Shannon's diversity index) were tested using a Kruskal-Wallis test (implemented in qiime diversity alpha-group-significance). Un/weighted UniFrac [24], Bray Curtis and Jaccard were used as metrics of paired distance between samples (beta-diversity), and the permutation-based PERMANOVA

test was performed (qiime diversity beta-group-significance) to test whether distances between samples within a group (GDM/control), were more similar to each other than they were to samples from the other group. To incorporate the major risk factors (BMI and age) into the model, we fit a distance matrix and used diagnosis, BMI and age as explanatory variables using adonis2 [25,26]. Since the model inputs variables sequentially, we fit 6 different models to examine all of the different orders of the variables.

Mantel's correlation between features were performed and unweighted UniFrac distances was used as the metric for microbial dissimilarity. For all other features, data was log transformed and min-max normalized, and Euclidean distance was used as the distance metric. 9999 permutations of label mixing were done; the p-value was calculated as the proportion of these permutations that lead to a higher explained variance than the measured explained variance.

To associate microbial features with GDM, features were collapsed to the different taxonomic levels from phylum to species. Spearman rank correlations were used to identify associations between the disease state for each microbial feature at each taxonomic level. Disease state labels were mixed 1000 times to receive a background distribution, and only correlations with $p<0.01$ were preserved. To control for the main risk factors of GDM, age and BMI, we adjusted associations by building a linear model and performed Spearman rank correlations on the linear regression residuals. Specifically, we regressed the disease state label over the age and BMI and computed the residual. We then computed the correlation between the different bacteria and the residual and compared that with the results obtained when scrambling the residuals among patients.

When considering the fecal microbiota profiles from the FMT study, microbial features were associated with GDM donors using MaAsLin2 [27] to perform per feature linear mixed effects (LME) modeling (see supplementary methods). Features were first log transformed and were subjected to cumulative sum scaled (CSS) normalization. Disease state (GDM/control) and days-post-FMT were used as fixed effects while cage and donor were included as random effects following Eq.1.

Eq 1: feature ~ disease + days_after_fmt + (1 | cage) + (1 | donor).

MetaCyc pathway abundances per sample were predicted using PICRUSt2 [28] with default parameters and using DADA2 amplicon sequence variants as inputs. Pathways with more than 25% zeros across samples were

removed. We next applied LME models (Eq. 2) to identify differentially abundant pathways between mice with FMT from women with vs. without a later diagnosis of GDM. Finally, we used Wald chi-square tests to determine the significance of the 'disease' fixed effect in each model and applied FDR corrections to all generated p-values.

Eq 2: pathway ~ disease + days_after_fmt + (1 | mice_ID)

For untargeted metabolomics, differential abundances of the metabolites between the groups were identified by log transformation of the peak areas followed by student's t-tests and FDR correction. Short peptides were manually curated using the metabolite name and using a list of dipeptides downloaded from the PubChem database [29]. Enrichment of amino acids was calculated using Fisher's exact test with the following contingency table groups: amino acid of interest, all other amino acids, peptides enriched in GDM, peptides not enriched in GDM.

Ethics statement

Informed consent was obtained from all participants of the main and secondary cohorts in accordance with Clalit's institutional review board approval No.0135-15-COM for the main cohort and with Rabin Medical Center institutional review board approval No.0263-15-RMC for the secondary cohort.

All experiments involving mice were performed using protocols approved by the local animal ethics committee at Bar-Ilan University (number 33-04-2018) and the IRB at Cornell University (Number NCT00167700). The STORK subjects were approved under Stanford IRB protocol number 17756.

Patient and public involvement

Participants were not involved in developing the research question nor the design of this study. Main and secondary cohort members were first involved when they were enrolled; their informed consent included time required to participate and no expectation of involvement in the dissemination of results.

1 Uzan-Yulzari A, Morr M, Tareef-Nabwani H, *et al.* The intestinal microbiome, weight, and metabolic changes in women treated by adjuvant chemotherapy for breast and gynecological malignancies. *BMC Med* 2020;**18**:281.

2 Kalter-Leibovici O, Freedman LS, Olmer L, *et al.* Screening and diagnosis of gestational diabetes mellitus: critical appraisal of the new International Association of Diabetes in Pregnancy Study Group recommendations on a national level. *Diabetes Care* 2012;**35**:1894–6.

3 Laitinen K, Poussa T, Isolauri E, *et al.* Probiotics and dietary counselling contribute to glucose regulation during and after pregnancy: a randomised controlled trial. *Br J Nutr* 2009;**101**:1679–87.

4 Ilmonen J, Isolauri E, Poussa T, *et al.* Impact of dietary counselling and probiotic intervention on maternal anthropometric measurements during and after pregnancy: A randomized placebo-controlled trial. Clinical Nutrition. 2011;**30**:156–64. doi:10.1016/j.clnu.2010.09.009

5 Luoto R, Laitinen K, Nermes M, *et al.* Impact of maternal probiotic-supplemented dietary counselling on pregnancy outcome and prenatal and postnatal growth: a double-blind, placebo-controlled study. *Br J Nutr* 2010;**103**:1792–9.

6 Ley C, Sanchez M de la L, Mathur A, *et al.* Stanford's Outcomes Research in Kids (STORK): a prospective study of healthy pregnant women and their babies in Northern California. *BMJ Open* 2016;**6**:e010810.

7 Consortium THMP, The Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. Nature. 2012;**486**:207–14. doi:10.1038/nature11234

8 Bolyen E, Rideout JR, Dillon MR, *et al.* Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol* 2019;**37**:852–7.

9 Callahan BJ, McMurdie PJ, Rosen MJ, *et al.* DADA2: High-resolution sample inference from Illumina amplicon data. Nature Methods. 2016;**13**:581–3. doi:10.1038/nmeth.3869

10 Janssen S, McDonald D, Gonzalez A, *et al.* Phylogenetic Placement of Exact Amplicon Sequences Improves Associations with Clinical Information. *mSystems* 2018;**3**. doi:10.1128/mSystems.00021-18

11 DeSantis TZ, Hugenholtz P, Larsen N, *et al.* Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* 2006;**72**:5069–72.

12 Bokulich NA, Kaehler BD, Rideout JR, *et al.* Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. *Microbiome* 2018;**6**:90.

13 Pence HE, Williams A. ChemSpider: An Online Chemical Information Resource. Journal of Chemical Education. 2010;**87**:1123–4. doi:10.1021/ed100697w

14 mzCloud – Advanced Mass Spectral Database. https://www.mzcloud.org/ (accessed 7 Jul 2021).

15 Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 2000;**28**:27–30.

16 Uzan-Yulzari A, Turta O, Belogolovski A, *et al.* Neonatal antibiotic exposure impairs child growth during the first six years of life by perturbing intestinal microbial colonization. *Nat Commun* 2021;**12**:443.

17 Binyamin D, Werbner N, Nuriel-Ohayon M, *et al.* The aging mouse microbiome has obesogenic characteristics. *Genome Med* 2020;**12**:87.

18 Koren O, Goodrich JK, Cullender TC, *et al.* Host remodeling of the gut microbiome and metabolic changes during pregnancy. *Cell* 2012;**150**:470–80.

19 Jasner Y, Belogolovski A, Ben-Itzhak M, *et al.* Microbiome Preprocessing Machine Learning Pipeline. *Front Immunol* 2021;**12**:677870.

20 Tianqi Chen University of Washington, Seattle, WA, USA, Carlos Guestrin University of Washington, Seattle, WA, USA. XGBoost. ACM Conferences. https://dl.acm.org/doi/10.1145/2939672.2939785 (accessed 9 Nov 2022).

21 Ma S, You Y, Huang L, *et al.* Alterations in Gut Microbiota of Gestational Diabetes Patients During the First Trimester of Pregnancy. *Front Cell Infect Microbiol* 2020;**10**:58.

22 Oshrit S, Haim I, Sondra T, *et al.* Image and graph convolution networks improve microbiome-based machine learning accuracy. Published Online First: 13 May 2022. doi:10.48550/arXiv.2205.06525

23 Virtanen P, Gommers R, Oliphant TE, *et al.* SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods* 2020;**17**:261–72.

24 Lozupone CA, Knight R. The UniFrac significance test is sensitive to tree topology. *BMC Bioinformatics* 2015;**16**:211.

25 Warton DI, Wright ST, Wang Y. Distance-based multivariate analyses confound location and dispersion effects. Methods in Ecology and Evolution. 2012;**3**:89–101. doi:10.1111/j.2041-210x.2011.00127.x

26 McArdle BH, Anderson MJ. FITTING MULTIVARIATE MODELS TO COMMUNITY DATA: A COMMENT ON DISTANCE-BASED REDUNDANCY ANALYSIS. Ecology. 2001;**82**:290–7. doi:10.1890/0012-9658(2001)082[0290:fmmtcd]2.0.co;2

27 Mallick H, Rahnavard A, McIver LJ, *et al.* Multivariable association discovery in population-scale meta-omics studies. *PLoS Comput Biol* 2021;**17**:e1009442.

28 Douglas GM, Maffei VJ, Zaneveld JR, *et al.* PICRUSt2 for prediction of metagenome functions. *Nat Biotechnol* 2020;**38**:685–8.

29 Kim S, Chen J, Cheng T, *et al.* PubChem in 2021: new data content and improved web interfaces. *Nucleic Acids Res* 2021;**49**:D1388–95.