# The Gut Microbiome in Polycystic Ovary Syndrome (PCOS) and its Association with Metabolic Traits

Kreete Lüll[1,2], Riikka K. Arffman[3], Alberto Sola-Leyva[4,5], Nerea M. Molina[4,5], Oliver Aasmets[1,2], Karl-Heinz Herzig[10], Stephen Franks[6], Laure Morin-Papunen[3], Juha S. Tapanainen[7,3], Andres Salumets[8,9], Signe Altmäe[4,5,9], Terhi T. Piltonen[3,10] and Elin Org[1*]

[1]Institute of Genomics, Estonian Genome Centre, University of Tartu, Tartu, Estonia;

[2]Department of Biotechnology, Institute of Molecular and Cell Biology, University of Tartu, Tartu, Estonia;

[3]Department of Obstetrics and Gynecology, PEDEGO Research Unit, Medical Research Centre, Oulu University Hospital, University of Oulu, Oulu, Finland;

[4]Department of Biochemistry and Molecular Biology, Faculty of Sciences, University of Granada, Granada, Spain;

[5]Instituto de Investigación Biosanitaria ibs. GRANADA, 18014 Granada, Spain;

[6]Faculty of Medicine, Department of Metabolism, Digestion and Reproduction, Imperial College London, London, United Kingdom;

[7]Department of Obstetrics and Gynecology, University of Helsinki and Helsinki University Hospital, Helsinki, Finland;

[8]Department of Obstetrics and Gynaecology, Institute of Clinical Medicine, University of Tartu, Tartu, Estonia;

[9]Competence Centre on Health Technologies, 50410 Tartu, Estonia;

[10]Department of Obstetrics and Gynecology, Reproductive Endocrinology and IVF Unit, Oulu, University Hospital, University of Oulu, Oulu, Finland.

11 Research Unit of Biomedicine, Medical Research Center, University of Oulu; University of Oulu, Oulu, Oulu University Hosptial, Oulu, Finland; Department of Paediatric Gastroenterology and Metabolic Diseases, Poznań University of Medical Sciences, Poznań, Poland.

*Corresponding author:

Elin Org, PhD

Institute of Genomics, Estonian Genome Centre
University of Tartu
Tartu 51010, Estonia
phone: (372) 737 4034; fax: (372) 737 4060; Email: elin.org@ut.ee

## ABSTRACT

**Purpose:** Despite gut microbiome being widely studied in metabolic diseases, its role in polycystic ovary syndrome (PCOS) has been scarcely investigated. The aim of our study was to test for possible associations between gut microbiome and PCOS in late fertile age women and investigate whether changes in the gut microbiome correlate with PCOS-related metabolic parameters.

**Methods:** We compared the 16S rRNA sequenced gut microbiome of 102 PCOS women with 201 age- and body mass index (BMI) matched non-PCOS women. Clinical and biochemical characteristics of the participants were assessed at ages 31 and 46 and analyzed in the context of gut microbiome data at the age of 46.

**Results:** Bacterial diversity indices did not differ significantly between PCOS and controls. We identified four genera whose balance helps to differentiate between PCOS and non-PCOS. In the whole cohort, the abundance of two genera from the order Clostridiales was correlated with several PCOS-related markers. When investigating the gut microbiome composition in PCOS women with different BMI and glucose tolerance groups, prediabetic PCOS women had significantly lower alpha diversity and markedly increased abundance of genus *Dorea* compared to women with normal glucose tolerance.

**Conclusions:** Our data indicate that PCOS and non-PCOS women at late fertile age with similar BMI do not signficantly differ in gut microbiota. However, there are significant microbial changes in PCOS individuals depending on their metabolic health. Further studies are needed in order to further understand these changes in more detail.

**Key Words:** PCOS, gut microbiome, metabolic traits, glucose tolerance

## INTRODUCTION

Polycystic ovary syndrome (PCOS) is one of the most prevalent endocrine and metabolic disorders in women, affecting 8–18% of women in reproductive age, depending on the studied population and applied diagnostic criteria (1–3). PCOS is characterized by persistent menstrual

irregularities, clinical or biochemical hyperandrogenism and polycystic ovarian morphology. It is a complex disorder associated with a variety of metabolic derangements, including obesity, insulin resistance and type 2 diabetes. However, regardless of obesity, insulin resistance is present in about 50% of women with PCOS (4,5). The etiology of PCOS remains unknown but is believed to be multifactorial where genetics, intrauterine environment, lifestyle factors and possibly alterations in the gut microbiome could have a role (6).

The human gut microbiota refers to all of the microorganisms inhabiting the gastrointestinal tract (7), where the majority belong to four bacterial phyla - Bacteroidetes, Firmicutes, Proteobacteria and Actinobacteria (8). Improved metabolic health has been related to increased microbial diversity and diverse microbiome gene content (9), while changes in the gut microbiome composition have been associated with a vast number of disease states beyond the gastrointestinal health, including type 2 diabetes (10–13), obesity (14–16), insulin resistance (17,18) and depression (19–21), among others.

Recently, the possible link between PCOS and gut microbiome has drawn increased attention. Previous studies investigating the relationship between gut microbiome and PCOS in women of different ethnicities have reported reduced bacterial diversity and overall composition as well as changes in the relative abundance of specific bacteria, mostly belonging to Bacteroidetes and Firmicutes phyla (22–25). Nevertheless, despite the reported links between the gut microbiota and clinical parameters in PCOS, the results are inconsistent and the community structure and function of the gut microbiome in women with PCOS remains unclear. Moreover, lack of body mass index (BMI) adjustment or BMI-matched controls, small sample size, ethnic differences, and variability in the methods used when conducting the microbiome analysis can lead to difficulties in reproducing the results. Hence, population-based cohorts, where individuals have been

evaluated during the follow-up visits, provide an important resource to define homogeneous phenotypes and detect changes in multiple pathophysiological pathways. Previous studies have primarily included PCOS women of reproductive age, whereas the effects of the microbiome on the health of women during the late reproductive and pre- or early menopausal years have remained unstudied.

The aim of our study was to test whether gut microbiome is associated with PCOS in late fertile age women including age- and BMI-matched controls. We also aimed to assess different metabolic and hormonal markers related to PCOS and their associations with the gut microbiome.

## MATERIAL AND METHODS

### *Study population*

The study population consisted of a subset of females in the longitudinal Northern Finland Birth Cohort 1966 (NFBC1966) which includes all expected births in 1966 in the two northernmost provinces of Finland (26). The NFBC1966 is a unique population-based cohort, which was established as a longitudinal research program to promote health and well-being of the population. Data collection procedures and the identification of PCOS cases in the NFBC1966 have been described earlier (27–29). In brief, at age 31 a questionnaire was sent to all women with known addresses in Finland and a majority of them also participated in a clinical examination with anthropometric measurements and blood samples. At age 46 the women were invited for a follow-up and to fill in a questionnaire. During the follow-up, clinical examinations with blood and fecal sample collection, oral glucose tolerance test (OGTT) and anthropometric measurements were performed. Women with PCOS in the cohort were identified by utilizing the questionnaires at ages 31 and 46. Briefly, at age of 31, women were asked whether they had oligo-amenorrhea (OA) or

hirsutism. If they answered “yes” to both, they were considered as PCOS cases. At 46, instead of OA and hirsutism questions, the women were asked if they had been diagnosed with polycystic ovaries and/or PCOS and those answering “yes” were considered PCOS cases. The total PCOS population consisted of women who reported hirsutism and oligo-amenorrhea at age 31 and/or PCO/PCOS at age 46. The validity of the PCOS diagnosis using the method described above has been verified in previous publications (28–30). The non-PCOS control population consisted of women with no PCOS symptoms at 31 nor PCO/PCOS by 46 years of age. Women who were on hormonal contraceptives or were pregnant at 31 years were excluded. The Finnish register for drug reimbursements was used to identify women who had been prescribed antibiotics, antimycotics, letrozole or tamoxifen within the 3 months preceding sample collection, who were then excluded. Also, women not permitting the use of their data were excluded. Two BMI-matched controls were chosen for each woman with PCOS. Women with PCOS from whom fecal samples were available were identified from the dataset. The total study population for the current study was 304 women with 102 women with PCOS and 202 non-PCOS control women. A flow-chart of the sample selection process is represented in Supplementary Figure 1. The study has been approved by the ethical committee of Northern Ostrobothnia hospital district. All participants of the NFBC1966 have given informed consent for the data and samples to be used for scientific purposes.

### *Sampling and laboratory measurements*

#### Hormonal measurements

Testosterone (T) was assayed from serum using Agilent triple quadrupole 6410 liquid chromatography–mass spectrometry equipment (Agilent Technologies, Wilmington, DE, USA). Elevated T level at ages 31 and 46 was defined according to the normal upper limit for T at these

respective ages based on the 97.5% percentile calculated in this population (2.3 nmol/l at age 31 and 1.7 nmol/l at age 46). Consequently, at age 31, women with T > 2.3 nmol/l were considered as women with elevated T levels and women with T ≤ 2.3 nmol/l as women with normal T levels. An elevated T level at age 46 was defined as serum level of T > 1.7 nmol/l and normal T level as serum T ≤ 1.7 nmol/l. Sex Hormone Binging Globulin (SHBG) was assayed by chemiluminometric immunoassay (Immulite 2000, Siemens Healthcare, Llanberis, UK). The free androgen index (FAI) was calculated using the following equation: T (nmol/L)/SHBG (nmol/L) × 100. For more detailed information please see Ollila *et al*. 2016 (29).

**Glucose metabolism assessment**

Plasma glucose was analyzed by an enzymatic dehydrogenase method (Advia 1800, Siemens Healthcare Diagnostic Inc., Tarrytown, NY, USA). Serum insulin was analyzed by a chemiluminometric immunoassay (Advia Centaur XP, Siemens Healthcare Diagnostics, Tarrytown, NY, USA). A 75 g glucose OGTT was performed with blood glucose and insulin measurements before glucose intake and 30 min, 60 min and 2 h later. Area under the curve (AUC) for glucose and AUC for insulin were calculated using the equation [(fasting X + (2 × (30min X) + (3 × 60min X) + (2 × 120min X)] × 15, where X states glucose (mmol/L) or insulin (μU/mL). The Secretion Index was calculated using the equation (fasting insulin (μU/mL) + 30min insulin (μU/mL)) × 6.945 / (fasting glucose (mmol/L) + 30 min glucose (mmol/L)). Matsuda index was calculated using the equation 10000 / sqrt ((fasting insulin (μU/mL) × fasting glucose (mmol/L) × 18) × (AUC glucose × 18 / 120 × AUC insulin / 120)). The Disposition Index was calculated by multiplying the Secretion Index by Matsuda Index.

Individuals were grouped according to WHO classification of diabetes based on their OGTT results. A person was categorized as preT2D if they had impaired fasting glucose (IFG: fasting plasma glucose from 6.1 to 6.9 mmol/L), impaired glucose tolerance (IGT: 2h glucose from 7.8 to 11.0 mmol/L), or both. A person was considered diabetic if they had fasting plasma glucose over 7.0 mmol/L or 2h plasma glucose over 11.2 mmol/L(31). Based on BMI scores, individuals were assigned into BMI groups: underweight (BMI below 18.4), normal weight (BMI 18.5-24.9), overweight (BMI 25-29.9), and obese (BMI over 30) (Supplementary Table 1).

***Gut microbiome analysis***

The fecal samples were collected at home by the study participants at age 46. It was recommended that the fecal sample should be delivered in a cooler on the day of collection. If that was not possible, the sample was stored for one or two days in a freezer at -20ºC until delivery. After delivery, the fecal samples were initially stored at -20ºC and then moved to -70ºC for long-term storage.

For bacterial DNA isolation, first the samples were homogenized in a Stomacher-400 blender. QIAamp Stool Mini Kit (Quiagen, Venlo, The Netherlands) was used for DNA extraction. The standard protocol was followed, with the exception that the samples were mixed with the lysis buffer and incubated at 95 ºC instead of 70 ºC in order to ensure the lysis of both Gram-negative and Gram-positive bacteria. The extracted DNA was quantified with NanoDrop ND-1000 spectrophotometer (Thermo Fisher Scientific, DE, USA). DNA yield was determined by measuring absorbance ratios spectrophotometrically, and included A260/280 nm for protein and A260/230 nm for salt and phenol contamination.

Microbiome was profiled by sequencing the V3-V4 regions of the 16S rRNA gene on an Illumina MiSeq instrument with forward 5'-CCTACGGGNGGCWGCA-3' and reverse 5'-GACTACHVGGGTATCTAATCC-3' primers. All PCR reactions were performed in 25µl volume containing 12.5µl 2X KAPA HiFi Hotstart ready mix (KAPA Biosystems, Woburn, MA, USA), 5µL of each primer (1µM), and 2.5µL of extracted DNA (10 ng) under the following cycling conditions: initial denaturation at 95ºC for 3 min, followed by 35 cycles of denaturation at 95ºC for 30 s, annealing at 55ºC for 30 s, and elongation at 72ºC for 30 s, with a final extension at 72ºC for 5 min. PCR clean-up was done with AMPure XP beads (Beckman Coulter, Indianapolis, IN, USA). Next, a PCR to index the amplicons was performed using the Nextera XT Index Kit (Illumina, San Diego, CA, USA) with conditions: 95ºC for 3 min; 8 cycles of 95ºC for 30 s, 55ºC for 30 s, 72ºC for 30 s, with a final extension step of 5 min at 72ºC, and hold at 4ºC. Next, the pooled PCR products were purified using AMPure XP beads (Beckman Coulter) before quantification. The final library was paired-end sequenced (2 x 300 bp) using a MiSeq Reagent Kit v.3 on the Illumina MiSeq sequencing system (Illumina).

**16S rRNA sequencing data analysis**

Raw sequences were demultiplexed with Illumina bcl2fastq2 Conversion Software v2.20 and raw data were imported into open-source software QIIME 2 2019.7 (32) using the q2-tools-import script with PairedEndFastqManifestPhred33 input format. In total, 16,063,617 reads were generated with the V3-V4 16S rRNA sequencing, with 5,501,498 reads for women with PCOS (average 53,936), and 10,562,119 (average 52,287) reads for healthy controls. Denoising was done with DADA2 (33), using a quality-aware model of Illumina amplicon errors to attain an abundance distribution of sequence variances, which have a difference of a single nucleotide. After retrieving quality scores, q2-dada2-denoise script was used to truncate the forward reads at position 288 and

trim at position 16. Reverse reads were truncated at position 241, trimming was not applied on the reverse read. Chimera removal was done using the "consensus" filter via q2-dada2-denoise in which chimeras are detected in samples individually and sequences found chimeric in a sufficient fraction of samples are removed. During this step forward and reverse reads are also merged. All amplicon sequence variants (ASVs) were aligned with MAFFT (34) via q2-alignment and phylogeny was constructed with FASTTREE2 (35) (via q2-phylogeny). Taxonomy was assigned to ASVs using the q2-feature-classifier (36) classify-sklearn naïve Bayes taxonomy classifier against the SILVA 16S V3-V4 v132_99 (37) along with a similarity threshold of 99%. The final 16S rRNA dataset comprised 16,063,617 reads with an average of 52,841 reads per sample. Altogether, 72,738 ASVs were identified (average of 680 ASVs for women with PCOS, and average of 670 ASVs for controls). A total number of 399 genera, 127 families, 56 orders, 29 classes and 19 phyla were detected. Data filtering steps included pruning samples to exclude samples with less than 10,000 reads after which 303 samples were left (102 women with PCOS and 201 control women). For beta diversity and association analysis we filtered out taxa that were detected in less than 30% of the remaining samples. The data were aggregated to genus level for further analysis. Alpha (Shannon's Index, Observed, and Inverse Simpson's Index) and beta [Principal Coordinate Analysis (PCoA), based on the UniFrac distance metrics (unweighted and weighted)] diversities were calculated using the phyloseq v.1.28.0 package (38). Weighted UniFrac metric accounts for taxa abundances in calculating the distances, whereas unweighted UniFrac only takes into account whether the taxa are present/absent and does not put emphasis on microbial abundances. For diversity comparisons, alpha diversity estimators were calculated and evaluated using ANOVA, one-way ANOVA tests with Tukey multiple pairwise-comparisons as

appropriate. ADONIS-2 function from the vegan package using 10,000 permutations for p-value calculations was used when testing differences in beta diversity.

### Statistical analysis

Statistical analyses and data visualisation were performed using the statistical software R v.3.6.1 (under RStudio v.1.2.1335). All of the visualizations were made using the ggplot2 v.3.3.0 (39) and corrplot v.084 (40) packages. A p-value below 0.05 was considered to be statistically significant and multiple testing was taken into account using the Benjamini and Hochberg False Discovery Rate (FDR) method (37,41). In the clinical biomarker's data comparison, the values were expressed as mean ± standard deviation. Wilcoxon sign rank test was used for testing the differences in the continuous variables between the study groups. Statistical analysis between groups of categorical variables were analyzed by Fisher's Exact test and the data were indicated as relative and absolute frequencies. The core microbiome was identified using the microbiome v.1.6.0 package (42) with a detection threshold of 0.01% and prevalence threshold of 95%. Associations between alpha diversity and clinical biomarkers were tested using Spearman correlation and adjusting for FDR using the *cor* function from the WGCNA v.1.69.package (43). ANOVA-Like Differential Expression tool (ALDEx2 v.1.16.0) (44) was used to identify differentially abundant taxa. *Selbal* v.0.1.0(45) was used to identify taxa whose balance, a log-contrast between two groups of taxa was predictive of PCOS status. Correlation analyses between taxa and biomarkers were performed using the function bicorAndPValue from the WGCNA v.1.69 package.

## RESULTS

### *Clinical characteristics of study participants*

A total of 303 women were included for analysis (102 PCOS and 201 age- and BMI-matched healthy controls). Table 1 summarizes the clinical characteristics of the study participants. Clinical examinations and blood sampling were done at 31 and 46 years of age, however, 68 study subjects had not attended the clinical examination at 31 years. Testosterone at 31 years ($p = 0.01$) and FAI at 46 years ($p = 0.01$) were higher in women with PCOS. As expected, matching for BMI resulted in a lack of significance in SHBG between the study groups, although the SHBG still showed some tendency for being lower in PCOS. Compared with controls, women with PCOS reported having fewer menses per year at age 31 and they experienced more infertility problems (lifetime) by age 31 and 46 years.

### *Landscape of microbiome composition and diversity*

The fecal microbial composition was profiled at the age of 46, at an age when the women were nearing the end of fertile age, by sequencing 16S rRNA gene V3-V4 regions. We first characterized the phylogenetic variation across all the samples ($n = 303$) at different taxonomic levels. We detected in total 72,738 ASVs, which resulted in a total richness of 399 genus-level taxonomic groups. After filtering out taxa present in less than 30% of the samples, the data consisted of 128 genera, 37 families, 18 orders, 14 classes and 8 phyla. We observed a typical Western diversity profile for gut microbiota, where Firmicutes (54.0%) and Bacteroidetes (31.9%) were the dominant phyla, followed by Proteobacteria (6.7%), Actinobacteria (3.4%), and Verrucomicrobia (2.4%) (Fig 1 A). The core microbiome (the number of taxa present in over 95% of individuals across the whole cohort) represent eight different genera belonging to two most abundant phyla, Bacteroidetes and Firmicutes. The most abundant genus was *Bacteroides* (19.9%),

followed by *Alistipes* (7.5%), *Faecalibacterium* (4.9%), *Roseburia* (2.5%), *Blautia* (2.5%), *Lachnoclostridium* (1.5%), *Ruminococcaceae* uncultured (1.2%), and *Oscillibacter* (1.1%) (Fig 1 B).

Next, we assessed differences in the gut microbiome diversity (alpha and beta diversities) between controls and PCOS samples. In alpha diversity analyses, that assesses the richness and evenness of the bacterial community, PCOS showed no differences from the controls in any of the assessed metrics ($P_{Shannon} = 0.979$, $P_{Observed} = 0.870$, $P_{InverseSimpson} = 0.248$) (Fig. 2A, Supplementary Fig 2). Beta diversity, which represents how much the community changes between controls and PCOS samples, the PCoA test based on both unweighted UniFrac (ADONIS $p = 0.175$, $r^2 = 0.004$) and weighted UniFrac measures (ADONIS $p = 0.44$, $r^2 = 0.003$), did not show any clustering with PCOS nor did the ADONIS show any difference between the groups (Fig 2B, C).

We then investigated whether bacterial richness (Shannon index, as alpha diversity) and beta diversity were associated with clinical, hormonal and metabolic parameters for all women. The Shannon index was negatively correlated with BMI ($p = 0.002$, FDR = 0.006), fasting insulin ($p = 0.01$, FDR = 0.01) and FAI ($p = 0.01$, FDR = 0.02) and positively correlated with SHBG ($p= 0.0002$, FDR = 0.0007), Matsuda Index ($p = 0.0001$, FDR = 0.0007) and Disposition Index ($p = 0.003$, FDR = 0.007) (Supplementary Table 2). Both unweighted and weighted UniFrac had statistically significant associations with BMI ($p_{weighted} = 2\times10^{-4}$, $p_{unweighted} = 1\times10^{-4}$), fasting insulin ($p_{weighted} = 0.006$, $p_{unweighted} = 0.001$), Matsuda Index ($p_{weighted} = 1\times10^{-4}$, $p_{unweighted} = 2\times10^{-4}$) and Disposition Index ($p_{weighted} = 1\times10^{-4}$, $p_{unweighted} = 0.007$). SHBG had a statistically significant association with beta diversity ($p_{weighted} = 0.003$, $p_{unweighted} = 0.034$) (Supplementary Table 2).

***Differences in fecal microbiome between the healthy women and women with PCOS***

The relative abundance of the 128 genera (present at least in 30% of the population) were compared between healthy control women and PCOS women (Supplementary Table 3). First, we used the ALDEx2 package (v1.16.0), which performs differential abundance analysis for the comparison of two or more conditions, PCOS vs. non-PCOS in this case. We did not observe any genera that showed statistically significant differences between the PCOS women and healthy control women after correcting for multiple testing (Supplementary Table 4).

With the *Selbal* analysis, we aimed to identify microbial signatures - groups of microbial taxa that are able to differentiate between PCOS and non-PCOS women. The identification of microbial signatures involves modeling the response variable and identifying the smallest number of taxa with the highest prediction accuracy. Therefore, these microbial signatures could be used to identify PCOS cases based on an individual's specific microbiome. *Selbal* analysis identified two groups of taxa consisting of *Paraprevotella*-*Streptococcus* and *Eubacterium ventriosum* group-*Bifidobacterium*, whose relative abundance or balance could differentiate between PCOS women and healthy control women (Fig 3). The discrimination value of the balance (AUC = 0.643) demonstrates a modest discrimination between the PCOS and healthy women. Interestingly, the *Eubacterium ventriosum* group, *Paraprevotella*, and *Streptococcus* were also among the top 10 genera that we identified in the ALDEx2 analysis. However, *Selbal* cannot be compared with ALDEx2 in terms of power and FDR since the aim of *Selbal* is to obtain the best model to predict the response (response being PCOS or control) in contrast to identifying all taxa that are associated with the response, which is the aim of ALDEx2.

### *Associations of taxa with hormonal and metabolic markers*

We correlated the top ten associated taxa to the markers that could possibly have a role in the development of PCOS. We did not detect any correlations that reached statistical significance, as all of the correlations found had a borderline significance (FDR ≤ 0.2). The abundance of *Ruminococcaceae UCG-002* was positively correlated with Matsuda Index (p = 0.010; FDR = 0.140), Disposition Index (p = 0.001; FDR = 0.065) and SHBG (p = 0.001; FDR = 0.065). The abundance of *Clostridiales Family XIII AD3011* group was positively correlated with Matsuda Index (p = 0.010; FDR = 0.140) and SHBG (p = 0.006; FDR = 0.140) and negatively correlated with HbA1c (p = 0.003; FDR = 0.099), 2 h glucose level (p = 0.006; FDR = 0.140) and BMI (p = 0.010; FDR = 0.140) (Fig 4, Supplementary Table 5).

### *Relationships between PCOS and metabolic traits*

Previous studies have shown that obesity has a critical role in PCOS and individuals with PCOS have an elevated risk in developing type 2 diabetes (T2D) (46). Therefore, we next aimed to evaluate the relationships between the gut microbiome and their potential effect on BMI and T2D in PCOS individuals. In order to study the relationships of PCOS and T2D together with gut microbiome, we focused on the women with PCOS and categorized them as having a normal glucose tolerance (NGT) or preT2D based on the values of fasting and 2-hour plasma glucose following the OGTT. The preT2D state includes individuals with IFG, IGT, or both (see "Methods" for details). In the PCOS group, there were 76 NGT individuals and 14 preT2D individuals. The four individuals who already had developed T2D based on their OGTT and eight individuals with no OGTT data were excluded from this analysis. We compared microbiota composition between PCOS women with NGT (n = 76) and with preT2D (n = 14). In diversity

analyses, both alpha (Shannon index; p = 0.018) and beta diversities (unweighted UniFrac; p = 0.003 and weighted UniFrac; p = 0.012) had statistically significant differences between the NGT and preT2D in the PCOS group (Fig 5 A, B; Supplementary Fig 3). These results are in line with previous studies which also report decreased microbial diversity in preT2D and T2D compared with NGT (47). As weighted UniFrac, as opposed to unweighted UniFrac, takes abundance into account, these results indicate that the association between preT2D and human gut microbiome could be induced by the presence/absence of more abundant taxa. In contrast, when analyzing control women, we did not observe differences in alpha or beta diversities between NGT and preT2D women (data not shown).

We also detected differentially abundant taxa between the NGT and preT2D, where the genus *Dorea* (Fig 5C) had elevated abundance in preT2D with the difference being statistically significant after FDR correction (FDR = 0.03). Genera *Bacteroides* (FDR = 0.07), *Ruminococcus torques* group (FDR = 0.07) and *Lachnospiraceae UCG-004* (FDR = 0.06) reported results of borderline significance (Supplementary Table 6, Supplementary Fig 4), all three had elevated abundance in preT2D group.

To study the effect of BMI on gut microbiome, we divided the PCOS individuals into three subgroups based on their BMI values: normal weight (n = 37), overweight (n = 38), and obese (n = 27). There were no statistically significant differences in diversity analysis (both alpha and beta diversity) between any of the sub-groups (Supplementary Fig 5 A-C). ALDEx2 analysis revealed no statistically significant differences in taxa between the normal weight and overweight groups. Analysis between normal weight and obese group as well as overweight and obese group revealed taxa with statistical significance, however these associations became insignificant after correcting for multiple testing (Supplementary Table 7).

## DISCUSSION

There is growing evidence that the gut microbiome contributes to women's health in addition to genetic and lifestyle factors (25,48,49). Metabolically active complex microbial community in the gut could either directly or indirectly affect the development of complex health disorders, such as PCOS (23,25). In this study, we investigated the relationship of gut microbiome with PCOS status in women nearing the end of fertile age, by assessing a total of 303 women from the population-based Northern Finland Birth Cohort 1966 (NFBC1966), all born in the same year and followed longitudinally from birth up to 46 years of age.

We first characterized the gut microbiome diversity and composition of the whole study cohort. The gut microbiome profile for Finnish females in their late fertile years is representative of the average gut microbiome profile in a population consuming a so-called Western-diet with the most prevalent phyla being *Firmicutes* and *Bacteroidetes*, followed by *Proteobacteria*, *Actinobacteria*, and *Verrucomicrobia* (50). The core microbiome (i.e., the genera shared by 95% of samples) in our cohort consists of 8 genera, of which 6 overlapped with the recent international MiBioGen consortium study findings where a total of 18,473 samples across 25 populations were analysed (51). Only two additional taxa, *Oscillibacter* and *Ruminococcaceae* uncultured, were identified as core members in our cohort. Both gender effect and geographic differences might explain the observed differences in core.

Next, we assessed the diversity and taxonomic differences between the women with PCOS and non-PCOS controls. The fecal samples for gut microbiome analysis were only collected at the age of 46, when the women were close to the end of their fertile age. Diversity analysis showed no significant difference in bacterial alpha and beta diversity between PCOS women and healthy

control women. In a previous work Qi *et al.* also reported no differences in alpha diversity when using whole-genome shotgun sequencing between the PCOS and controls, however, they did report significantly lower beta diversity in PCOS compared to the controls (52). Diversity differences between the PCOS and non-PCOS individuals are inconsistent throughout previous studies; while some were able to detect changes in alpha/beta diversities (22,23,25,49,53), others found no significant differences between groups (24,52,54,55). It is possible that PCOS itself does not alter the gut microbiome to such extent that it reflects changes in the whole microbiome community. Rather the changes in diversity may become significant when analyzing PCOS-related metabolic traits. We did, indeed, observe significant associations between diversity measures and PCOS-related hormonal and metabolic parameters such as BMI, SHBG levels, and insulin sensitivity as well as insulin resistance as measured by Matsuda and Disposition Indices.

Despite seeing no significantly different taxa between the PCOS and controls after multiple corrections, some of the top genera have previously been linked with PCOS-related traits. *Eubacterium ventriosum* group, that we detected in both single taxa and group-based modelling analysis, has previously shown to be enriched in obese individuals (56). The genus *Paraprevotella* has been formerly detected in prenatal androgenized rat models where *Paraprevotella* had significantly enriched levels in androgenized rodents (57,58). The fact that most PCOS cases also suffer from hyperandrogenism may explain the link between genus *Paraprevotella* and PCOS. However, in our results *Paraprevotella* was decreased in PCOS women. It is possible that this discrepancy is due to the fact that the women were already 46 years old and that their androgen levels did not differ from the age- and BMI-matched controls at this time point. The genus *Turicibacter* itself has not been reported to be associated with PCOS previously, but in the present analysis it was one of the top genera seen in PCOS. *Turicibacter* belongs to the family

*Erysipelotrichaceae* that has been reported to have elevated levels in PCOS both in mouse and human (22,59). In our analyses we observed three genera from the family *Ruminococcaceae* that belonged to the top taxa found between PCOS and controls. None of the genera from *Ruminococcaceae* which we observed have previously been linked with PCOS, although the family itself and some its genera have been associated with PCOS in earlier studies (24,25). It is possible that the different genera from *Ruminococcaceae* have similar metabolic functions that might affect the development of PCOS and its traits. Interestingly, in our study one particular genera, the *Ruminococcaceae UCG-002,* was positively associated with higher SHBG levels and insulin sensitivity and negatively associated with BMI, HbA1c and 2h glucose levels. Finally, using a modelling-based analysis, we were also able to identify two groups of microbial taxa whose relative abundance differentiate between PCOS and non-PCOS women. Although our data show modest discrimination between the PCOS and healthy women among the identified microbial groups, this analysis provides an important indication for future prediction studies.

PCOS is associated with metabolic derangements, where more than half of the women are overweight or obese and PCOS triples a woman's risk of developing type 2 diabetes and metabolic syndrome REF?. In our cohort of 102 PCOS women, 38 were overweight and 27 were obese, and normal weight, overweight and obese PCOS individuals showed no statistically significant changes in bacterial diversities or in taxonomical composition between the three groups. Many earlier studies have shown that BMI has a crucial role in the diversity of gut microbiome and that higher BMI is strongly associated with gut dysbiosis (60–62). The reason for the observed inconsistency could have been the small sample size in our BMI based analyses in the PCOS group, as significant associations between diversity and BMI were observed when the entire cohort was used. However, we were able to identify significant changes in both alpha and beta diversity when

comparing women with PCOS classified based on OGTT as either NGT or preT2D. This is consistent with previous work, showing decreased richness and diversity in individuals with preT2D and T2D(47,63). In addition, PCOS women with preT2D also showed higher abundance of genus *Dorea* compared to women in the NGT group. In a previous work, the genus *Dorea* has been linked to BMI and plasma metabolites such as glutamate and branched chain amino acids (BCAAs) which can predict BMI values (64). Naderpoor *et al.* showed, that the abundance of *Dorea* was positively correlated with fasting blood glucose (65). Interestingly, the abundance of *Dorea* is significantly associated with various metabolic traits, such as BMI, glucose and insulin levels also in our cohort (Supplementary Fig 6 A-F), providing additional support for the role of this genus in metabolic disorders. Insulin resistance (IR) is common in women with PCOS although it is not included in diagnostic criteria for PCOS (4,5). Cani *et al.* reported in their work that some Gram-negative bacteria have the ability to produce lipopolysaccharides which induce inflammation, IR, and obesity, all widely known traits of PCOS (66). Moreover, certain Gram-negative bacteria belonging to the genera *Bacteroides* and *Escherihia*/*Shigella* are significantly increased in PCOS (24). It is plausible that the gut microbiome is involved in the pathogenesis of PCOS by mediating IR and systemic inflammation. Our data support this hypothesis, as we detected significant associations between microbiome diversity as well as specific taxa and insulin resistance. Moreover, the effect of microbiome could also be mediated through metabolites produced by gut microbiome. For example, in a study in mice it was shown that the animals with diet-induced obesity showed improved insulin resistance after short chain fatty acid (SCFA) supplementation (67). SCFAs are produced through fermentation of non-digestible dietary fibers by the gut microbiome and they are an important energy source to the gut as well as performing as signaling molecules to affect the host metabolism (68). Therefore, it is possible that the

dysbiosis of the gut that affects SCFA production is in turn, at least to some extent, responsible for the insulin resistance seen in PCOS individuals.

The strength of our current study is that the cohort was homogenous, all women belonging to the unique Northern Finland Birth Cohort and were born in 1966 with minimal ethnic and geographic variation. PCOS cases and controls were all BMI-matched, thus the BMI effects on the comparisons between the cases and the controls are minimal. Furthermore, to our knowledge, the present study includes the largest number of PCOS women to date to investigate the associations between PCOS and the gut microbiome (69). Additionally, this is the first study assessing microbiota of women with PCOS in late fertile age.

Multiple reasons might explain why we could not confirm previously reported associations. The gut microbiome variation is affected by several factors such as diet, geography, medications, age, which all could possibly influence the results. In addition, the use of methods for sample collection, DNA extraction, 16S rRNA gene sequencing, and data analysis varies widely from study to study. Also, the use of small sample size in some previous studies may limit the ability to detect changes in specific bacteria. All of the above factors probably contributed to the discrepancies observed, but the identical age and large sample set are probably the most important factors that differed from previous studies. Earlier studies have been mainly focused on women of reproductive age  but the effects of the microbiome on the health of older women who are close to starting menopause might be different. At age 46, women with PCOS are less hyperandrogenic and metabolically more similar to non-PCOS women, who also have gained weight while ageing. It may be that at early and mid-reproductive years, PCOS women may more substantially differ in the gut microbiome. Limitations of our study are self-reporting of PCOS symptoms and lack of ultrasonography. However, the validity of self-reported PCOS phenotype has been established

(28–30) and also supported by recently published data (70) and genetic analysis (71). Furthermore, in the case of PCOS, it would be interesting to investigate the microbiome not only of the gut but also from the reproductive tract. It is possible that PCOS-related infertility problems might also be associated with microbiome status as suggested by some early data (72).

## CONCLUSIONS

This study, utilizing the unique NFBC66 cohort, revealed that women with PCOS in their late reproductive years have no large-scale difference in gut microbiome signature compared to age- and BMI-matched women. However, we did identify a number of associations between microbial diversity, as well as specific taxa, and PCOS-related hormones and metabolic traits. In addition, we show clear differences in microbiome profile in PCOS women in a pre-diabetic state compared to PCOS women with normal glucose tolerance. Further studies (metagenomic, metabolomics, and functional) would be required to clarify the link between gut microbiome, metabolites and development of PCOS.

## SUPPLEMENTARY DATA

Supplementary tables: https://docs.google.com/spreadsheets/d/1DuUUNnRhrQzaoRY9mu286xeiat05NIYPCRrwTTiU1YU/edit?usp=sharing

Supplementary figures: https://docs.google.com/document/d/1A-A50W5ZQMdQeiL75PK-aKw6QRL_p5hfQFwW_2vTcvc/edit#

## ACKNOWLEDGEMENTS

We thank all NFBC66 study participants.

**DECLARATIONS**

**Ethics approval and consent to participate.** All subjects have given written informed consent and the study was approved by the Ethics Committee of the Northern Ostrobothnia hospital district and was in accordance with the Helsinki Declaration.

**Availability of data and materials.** Individual-level 16S RNA sequencing data are submitted in the Sequence Read Archive (SRA) under accession number XXX

**Competing interest.** The authors declare that they have no competing interests.

**Funding.** This work was funded by Estonian Research Council grants PUT 1371 (to E.O.), EMBO Installation grant 3573 (to E.O.) …

E.O. was supported by European Regional Development Fund Project No. 15-0012 GENTRANSMED and Estonian Center of Genomics/Roadmap II project No 16-0125.

S.A. was supported by the Spanish Ministry of Economy, Industry and Competitiveness (MINECO) and European Regional Development Fund (FEDER): grants RYC-2016-21199 and ENDORE (SAF2017-87526-R); and by FEDER/Junta de Andalucía-Consejería de Economía y Conocimiento: MENDO (B-CTS-500-UGR18).

**Authors' contributions.** T.P., L.M.-P., J.S.T., S.F., A.S., S.A and E.O. supervised the study. R.K.A., T.P. and K.H-H. oversaw collection of samples. A.S.-L., N.M.M. and S.A. prepared microbiome sequencing analysis. K.L performed the data analysis. K.L., S.A., T.P. and E.O. prepared the manuscript with comments from other authors.

## REFERENCES


1. **March WA, Moore VM, Willson KJ, Phillips DIW, Norman RJ, Davies MJ.** The prevalence of polycystic ovary syndrome in a community sample assessed under contrasting diagnostic criteria. *Hum. Reprod.* 2010;25(2):544–551.
2. **Teede HJ, Misso ML, Costello MF, Dokras A, Laven J, Moran L, Piltonen T, Norman RJ, Network IP.** Recommendations from the international evidence-based guideline for the assessment and management of polycystic ovary syndrome. *Hum. Reprod.* 2018;33(9):1602–1618.
3. **Zeng X, Xie Y-J, Liu Y-T, Long S-L, Mo Z-C.** Polycystic ovarian syndrome: Correlation between hyperandrogenism, insulin resistance and obesity. *Clin. Chim. Acta* 2020;502:214–221.
4. **Chen L, Xu WM, Zhang D.** Association of abdominal obesity, insulin resistance, and oxidative stress in adipose tissue in women with polycystic ovary syndrome. *Fertil. Steril.* 2014;102(4):1167-1174.e4.
5. **Luque-Ramírez M, Escobar-Morreale HF.** Polycystic ovary syndrome as a paradigm for prehypertension, prediabetes, and preobesity. *Curr. Hypertens. Rep.* 2014;16(12):500.
6. **Tremellen K, Pearce K.** Dysbiosis of Gut Microbiota (DOGMA) – A novel theory for the development of Polycystic Ovarian Syndrome. *Med. Hypotheses* 2012;79(1):104–112.
7. **Coyte KZ, Schluter J, Foster KR.** The ecology of the microbiome: Networks, competition, and stability. *Science (80-. ).* 2015;350(6261):663–666.
8. **Arumugam M, Raes J, Pelletier E, Le Paslier D, Yamada T, Mende DR, Fernandes GR, Tap J, Bruls T, Batto J-M, Bertalan M, Borruel N, Casellas F, Fernandez L, Gautier L, Hansen T, Hattori M, Hayashi T, Kleerebezem M, Kurokawa K, Leclerc M, Levenez F, Manichanh C, Nielsen HB, Nielsen T, Pons N, Poulain J, Qin J, Sicheritz-Ponten T, Tims S, Torrents D, Ugarte E, Zoetendal EG, Wang J, Guarner F, Pedersen O, de Vos WM, Brunak S, Doré J, Consortium M, Antolín M, Artiguenave F, Blottiere HM, Almeida M, Brechot C, Cara C, Chervaux C, Cultrone A, Delorme C, Denariaz G, Dervyn R, Foerstner KU, Friss C, van de Guchte M, Guedon E, Haimet F, Huber W, van Hylckama-Vlieg J, Jamet A, Juste C, Kaci G, Knol J, Lakhdari O, Layec S, Le Roux K, Maguin E, Mérieux A, Melo Minardi R, M'rini C, Muller J, Oozeer R, Parkhill J, Renault P, Rescigno M, Sanchez N, Sunagawa S, Torrejon A, Turner K, Vandemeulebrouck G, Varela E, Winogradsky Y, Zeller G, Weissenbach J, Ehrlich SD, Bork P.** Enterotypes of the human gut microbiome. *Nature*

2011;473(7346):174–180.

9. **Sonnenburg JL, Bäckhed F.** Diet-microbiota interactions as moderators of human metabolism. *Nature* 2016;535(7610):56–64.
10. **Allin KH, Tremaroli V, Caesar R, Jensen BAH, Damgaard MTF, Bahl MI, Licht TR, Hansen TH, Nielsen T, Dantoft TM, Linneberg A, Jørgensen T, Vestergaard H, Kristiansen K, Franks PW, Hansen T, Bäckhed F, Pedersen O.** Aberrant intestinal microbiota in individuals with prediabetes. *Diabetologia* 2018;61(4):810–820.
11. **Forslund K, Hildebrand F, Nielsen T, Falony G, Le Chatelier E, Sunagawa S, Prifti E, Vieira-Silva S, Gudmundsdottir V, Krogh Pedersen H, Arumugam M, Kristiansen K, Voigt AY, Vestergaard H, Hercog R, Igor Costea P, Kultima JR, Li J, Jorgensen T, Levenez F, Dore J, Nielsen HB, Brunak S, Raes J, Hansen T, Wang J, Ehrlich SD, Bork P, Pedersen O.** Disentangling type 2 diabetes and metformin treatment signatures in the human gut microbiota. *Nature* 2015;528(7581):262–266.
12. **Karlsson FH, Tremaroli V, Nookaew I, Bergström G, Behre CJ, Fagerberg B, Nielsen J, Bäckhed F.** Gut metagenome in European women with normal, impaired and diabetic glucose control. *Nature* 2013;498(7452):99–103.
13. **Larsen N, Vogensen FK, Van Den Berg FWJ, Nielsen DS, Andreasen AS, Pedersen BK, Al-Soud WA, Sørensen SJ, Hansen LH, Jakobsen M.** Gut microbiota in human adults with type 2 diabetes differs from non-diabetic adults. *PLoS One* 2010;5(2). doi:10.1371/journal.pone.0009085.
14. **Bäckhed F, Ding H, Wang T, Hooper L V, Koh GY, Nagy A, Semenkovich CF, Gordon JI.** The gut microbiota as an environmental factor that regulates fat storage. *Proc. Natl. Acad. Sci. U. S. A.* 2004;101(44):15718–15723.
15. **Haro C, García-Carpintero S, Rangel-Zúñiga OA, Alcalá-Díaz JF, Landa BB, Clemente JC, Pérez-Martínez P, López-Miranda J, Pérez-Jiménez F, Camargo A.** Consumption of Two Healthy Dietary Patterns Restored Microbiota Dysbiosis in Obese Patients with Metabolic Dysfunction. *Mol. Nutr. Food Res.* 2017;61(12). doi:10.1002/mnfr.201700300.
16. **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.
17. **Pedersen HK, Gudmundsdottir V, Nielsen HB, Hyotylainen T, Nielsen T, Jensen BAH, Forslund K, Hildebrand F, Prifti E, Falony G, Le Chatelier E, Levenez F, Doré J, Mattila I, Plichta DR, Pöhö P, Hellgren LI, Arumugam M, Sunagawa S, Vieira-Silva**

**S, Jørgensen T, Holm JB, Trošt K, Consortium M, Kristiansen K, Brix S, Raes J, Wang J, Hansen T, Bork P, Brunak S, Oresic M, Ehrlich SD, Pedersen O.** Human gut microbes impact host serum metabolome and insulin sensitivity. *Nature* 2016;535(7612):376–381.

18. **Aron-Wisnewsky J, Prifti E, Belda E, Ichou F, Kayser BD, Dao MC, Verger EO, Hedjazi L, Bouillot J-L, Chevallier J-M, Pons N, Le Chatelier E, Levenez F, Ehrlich SD, Dore J, Zucker J-D, Clément K.** Major microbiota dysbiosis in severe obesity: fate after bariatric surgery. *Gut* 2019;68(1):70–82.
19. **Luna RA, Foster JA.** Gut brain axis: diet microbiota interactions and implications for modulation of anxiety and depression. *Curr. Opin. Biotechnol.* 2015;32:35–41.
20. **Schachter J, Martel J, Lin C-S, Chang C-J, Wu T-R, Lu C-C, Ko Y-F, Lai H-C, Ojcius DM, Young JD.** Effects of obesity on depression: A role for inflammation and the gut microbiota. *Brain Behav. Immun.* 2018;69:1–8.
21. **Valles-Colomer M, Falony G, Darzi Y, Tigchelaar EF, Wang J, Tito RY, Schiweck C, Kurilshikov A, Joossens M, Wijmenga C, Claes S, Van Oudenhove L, Zhernakova A, Vieira-Silva S, Raes J.** The neuroactive potential of the human gut microbiota in quality of life and depression. *Nat Microbiol* 2019;4(4):623–632.
22. **Insenser M, Murri M, Del Campo R, Martínez-García MÁ, Fernández-Durán E, Escobar-Morreale HF.** Gut Microbiota and the Polycystic Ovary Syndrome: Influence of Sex, Sex Hormones, and Obesity. *J. Clin. Endocrinol. Metab.* 2018;103(7):2552–2562.
23. **Lindheim L, Bashir M, Münzker J, Trummer C, Zachhuber V, Leber B, Horvath A, Pieber TR, Gorkiewicz G, Stadlbauer V, Obermayer-Pietsch B.** Alterations in Gut Microbiome Composition and Barrier Function Are Associated with Reproductive and Metabolic Defects in Women with Polycystic Ovary Syndrome (PCOS): A Pilot Study. *PLoS One* 2017;12(1):e0168390.
24. **Liu R, Zhang C, Shi Y, Zhang F, Li L, Wang X, Ling Y, Fu H, Dong W, Shen J, Reeves A, Greenberg AS, Zhao L, Peng Y, Ding X.** Dysbiosis of Gut Microbiota Associated with Clinical Parameters in Polycystic Ovary Syndrome. *Front. Microbiol.* 2017;8:324.
25. **Torres PJ, Siakowska M, Banaszewska B, Pawelczyk L, Duleba AJ, Kelley ST, Thackray VG.** Gut Microbial Diversity in Women With Polycystic Ovary Syndrome Correlates With Hyperandrogenism. *J. Clin. Endocrinol. Metab.* 2018;103(4):1502–1511.
26. **Rantakallio P.** The longitudinal study of the northern Finland birth cohort of 1966. *Paediatr. Perinat. Epidemiol.* 1988;2(1):59–88.

27. **Karjula S, Morin-Papunen L, Auvinen J, Ruokonen A, Puukka K, Franks S, Järvelin M-R, Tapanainen JS, Jokelainen J, Miettunen J, Piltonen TT.** Psychological Distress Is More Prevalent in Fertile Age and Premenopausal Women With PCOS Symptoms: 15-Year Follow-Up. *J. Clin. Endocrinol. Metab.* 2017;102(6):1861–1869.

28. **Koivuaho E, Laru J, Ojaniemi M, Puukka K, Kettunen J, Tapanainen JS, Franks S, Järvelin M-R, Morin-Papunen L, Sebert S, Piltonen TT.** Age at adiposity rebound in childhood is associated with PCOS diagnosis and obesity in adulthood-longitudinal analysis of BMI data from birth to age 46 in cases of PCOS. *Int. J. Obes.* 2019;43(7):1370–1379.

29. **Ollila M-ME, Piltonen T, Puukka K, Ruokonen A, Järvelin M-R, Tapanainen JS, Franks S, Morin-Papunen L.** Weight Gain and Dyslipidemia in Early Adulthood Associate With Polycystic Ovary Syndrome: Prospective Cohort Study. *J. Clin. Endocrinol. Metab.* 2016;101(2):739–747.

30. **Taponen S, Ahonkallio S, Martikainen H, Koivunen R, Ruokonen A, Sovio U, Hartikainen A-L, Pouta A, Laitinen J, King V, Franks S, McCarthy MI, Järvelin M-R.** Prevalence of polycystic ovaries in women with self-reported symptoms of oligomenorrhoea and/or hirsutism: Northern Finland Birth Cohort 1966 Study. *Hum. Reprod.* 2004;19(5):1083–1088.

31. **WHO.** Definition and diagnosis of diabetes mellitus and intermediate hyperglycaemia. 2006;ISBN: 978.

32. **Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, Alexander H, Alm EJ, Arumugam M, Asnicar F, Bai Y, Bisanz JE, Bittinger K, Brejnrod A, Brislawn CJ, Brown CT, Callahan BJ, Caraballo-Rodríguez AM, Chase J, Cope EK, Da Silva R, Diener C, Dorrestein PC, Douglas GM, Durall DM, Duvallet C, Edwardson CF, Ernst M, Estaki M, Fouquier J, Gauglitz JM, Gibbons SM, Gibson DL, Gonzalez A, Gorlick K, Guo J, Hillmann B, Holmes S, Holste H, Huttenhower C, Huttley GA, Janssen S, Jarmusch AK, Jiang L, Kaehler BD, Kang K Bin, Keefe CR, Keim P, Kelley ST, Knights D, Koester I, Kosciolek T, Kreps J, Langille MGI, Lee J, Ley R, Liu Y-X, Loftfield E, Lozupone C, Maher M, Marotz C, Martin BD, McDonald D, McIver LJ, Melnik A V, Metcalf JL, Morgan SC, Morton JT, Naimey AT, Navas-Molina JA, Nothias LF, Orchanian SB, Pearson T, Peoples SL, Petras D, Preuss ML, Pruesse E, Rasmussen LB, Rivers A, Robeson 2nd MS, Rosenthal P, Segata N, Shaffer M, Shiffer A, Sinha R, Song SJ, Spear JR, Swafford AD, Thompson LR, Torres PJ, Trinh P, Tripathi A, Turnbaugh PJ, Ul-Hasan S, van der Hooft JJJ,**

**Vargas F, Vázquez-Baeza Y, Vogtmann E, von Hippel M, Walters W, Wan Y, Wang M, Warren J, Weber KC, Williamson CHD, Willis AD, Xu ZZ, Zaneveld JR, Zhang Y, Zhu Q, Knight R, Caporaso JG.** Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat. Biotechnol.* 2019;37(8):852–857.

33. **Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP.** DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods* 2016;13(7):581–583.
34. **Katoh K, Standley DM.** MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 2013;30(4):772–780.
35. **Price MN, Dehal PS, Arkin AP.** FastTree 2--approximately maximum-likelihood trees for large alignments. *PLoS One* 2010;5(3):e9490.
36. **Bokulich NA, Kaehler BD, Rideout JR, Dillon M, Bolyen E, Knight R, Huttley GA, Gregory Caporaso J.** Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. *Microbiome* 2018;6(1). doi:10.1186/s40168-018-0470-z.
37. **Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO.** The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* 2013;41(Database issue):D590-6.
38. **McMurdie PJ, Holmes S.** phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS One* 2013;8(4):e61217.
39. **Gómez-Rubio V.** ggplot2 - Elegant Graphics for Data Analysis (2nd Edition). *J. Stat. Softw.* 2017;77(Book Review 2). doi:10.18637/jss.v077.b02.
40. **Wei T, Simko V.** R package "corrplot": Visualization of a Correlation Matrix (Version 0.84). 2017. Available at: https://github.com/taiyun/corrplot.
41. **Benjamini Y, Hochberg Y.** Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B* 1995;57(1):289–300.
42. **Lahti L, Shetty S.** Microbiome R package. Available at: url: http://microbiome.github.io.
43. **Langfelder P, Horvath S.** WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 2008;9(1). doi:10.1186/1471-2105-9-559.
44. **Fernandes AD, Macklaim JM, Linn TG, Reid G, Gloor GB.** ANOVA-Like Differential Expression (ALDEx) Analysis for Mixed Population RNA-Seq. *PLoS One* 2013;8(7):e67019.
45. **Rivera-Pinto J, Egozcue JJ, Pawlowsky-Glahn V, Paredes R, Noguera-Julian M, Calle ML.** Balances: a New Perspective for Microbiome Analysis. *mSystems* 2018;3(4).

doi:10.1128/mSystems.00053-18.

46. **Rubin KH, Glintborg D, Nybo M, Abrahamsen B, Andersen M.** Development and Risk Factors of Type 2 Diabetes in a Nationwide Population of Women With Polycystic Ovary Syndrome. *J. Clin. Endocrinol. Metab.* 2017;102(10):3848–3857.
47. **Gurung M, Li Z, You H, Rodrigues R, Jump DB, Morgun A, Shulzhenko N.** Role of gut microbiota in type 2 diabetes pathophysiology. *EBioMedicine* 2020;51:102590.
48. **Younes JA, Lievens E, Hummelen R, van der Westen R, Reid G, Petrova MI.** Women and Their Microbes: The Unexpected Friendship. *Trends Microbiol.* 2018;26(1):16–32.
49. **Zhou L, Ni Z, Cheng W, Yu J, Sun S, Zhai D, Yu C, Cai Z.** Characteristic gut microbiota and predicted metabolic functions in women with PCOS. *Endocr Connect* 2020;9(1):63–73.
50. **Senghora B, Sokhnab C, Ruimycde R, Lagiera J-C.** Gut microbiota diversity according to dietary habits and geographical provenance. *Hum. Microbiome J.* 2018;7–8(1):1–9.
51. **Kurilshikov A, Medina-Gomez C, Bacigalupe R, Radjabzadeh D, Wang J, Demirkan A, Le Roy CI, Raygoza Garay JA, Finnicum CT, Liu X, Zhernakova D V, Bonder MJ, Hansen TH, Frost F, Rühlemann MC, Turpin W, Moon J-Y, Kim H-N, Lüll K, Barkan E, Shah SA, Fornage M, Szopinska-Tokov J, Wallen ZD, Borisevich D, Agreus L, Andreasson A, Bang C, Bedrani L, Bell JT, Bisgaard H, Boehnke M, Boomsma DI, Burk RD, Claringbould A, Croitoru K, Davies GE, van Duijn CM, Duijts L, Falony G, Fu J, van der Graaf A, Hansen T, Homuth G, Hughes DA, Ijzerman RG, Jackson MA, Jaddoe VW V, Joossens M, Jørgensen T, Keszthelyi D, Knight R, Laakso M, Laudes M, Launer LJ, Lieb W, Lusis AJ, Masclee AAM, Moll HA, Mujagic Z, Qibin Q, Rothschild D, Shin H, Sørensen SJ, Steves CJ, Thorsen J, Timpson NJ, Tito RY, Vieira-Silva S, Völker U, Völzke H, Võsa U, Wade KH, Walter S, Watanabe K, Weiss S, Weiss FU, Weissbrod O, Westra H-J, Willemsen G, Payami H, Jonkers DMAE, Vasquez AA, de Geus EJC, Meyer KA, Stokholm J, Segal E, Org E, Wijmenga C, Kim H-L, Kaplan RC, Spector TD, Uitterlinden AG, Rivadeneira F, Franke A, Lerch MM, Franke L, Sanna S, D'Amato M, Pedersen O, Paterson AD, Kraaij R, Raes J, Zhernakova A.** Genetics of human gut microbiome composition. *bioRxiv* 2020:2020.06.26.173724.
52. **Qi X, Yun C, Sun L, Xia J, Wu Q, Wang Y, Wang L, Zhang Y, Liang X, Wang L, Gonzalez FJ, Patterson AD, Liu H, Mu L, Zhou Z, Zhao Y, Li R, Liu P, Zhong C, Pang Y, Jiang C, Qiao J.** Gut microbiota–bile acid–interleukin-22 axis orchestrates polycystic ovary syndrome. *Nat. Med.* 2019;25(8):1225–1233.

53. **Haudum C, Lindheim L, Ascani A, Trummer C, Horvath A, Münzker J, Obermayer-Pietsch B.** Impact of Short-Term Isoflavone Intervention in Polycystic Ovary Syndrome (PCOS) Patients on Microbiota Composition and Metagenomics. *Nutrients* 2020;12(6). doi:10.3390/nu12061622.
54. **Chu W, Han Q, Xu J, Wang J, Sun Y, Li W, Chen Z-J, Du Y.** Metagenomic analysis identified microbiome alterations and pathological association between intestinal microbiota and polycystic ovary syndrome. *Fertil. Steril.* 2020;113(6):1286-1298.e4.
55. **Eyupoglu ND, Ergunay K, Acikgoz A, Akyon Y, Yilmaz E, Yildiz BO.** Gut microbiota and oral contraceptive use in overweight and obese patients with polycystic ovary syndrome. *J. Clin. Endocrinol. Metab.* 2020. doi:10.1210/clinem/dgaa600.
56. **Kasai C, Sugimoto K, Moritani I, Tanaka J, Oya Y, Inoue H, Tameda M, Shiraki K, Ito M, Takei Y, Takase K.** Comparison of the gut microbiota composition between obese and non-obese individuals in a Japanese population, as analyzed by terminal restriction fragment length polymorphism and next-generation sequencing. *BMC Gastroenterol.* 2015;15(1). doi:10.1186/s12876-015-0330-2.
57. **Moreno-Indias I, Sánchez-Alcoholado L, Sánchez-Garrido MÁ, Martín-Núñez GM, Pérez-Jiménez F, Tena-Sempere M, Tinahones FJ, Queipo-Ortuño MI.** Neonatal Androgen Exposure Causes Persistent Gut Microbiota Dysbiosis Related to Metabolic Disease in Adult Female Rats. *Endocrinology* 2016;157(12):4888–4898.
58. **Sherman SB, Sarsour N, Salehi M, Schroering A, Mell B, Joe B, Hill JW.** Prenatal androgen exposure causes hypertension and gut microbiota dysbiosis. *Gut Microbes* 2018;9(5):400–421.
59. **Kelley ST, Skarra D V, Rivera AJ, Thackray VG.** The Gut Microbiome Is Altered in a Letrozole-Induced Mouse Model of Polycystic Ovary Syndrome. *PLoS One* 2016;11(1):e0146509.
60. **Barlow GM, Yu A, Mathur R.** Role of the Gut Microbiome in Obesity and Diabetes Mellitus. *Nutr. Clin. Pr.* 2015;30(6):787–797.
61. **Guirro M, Costa A, Gual-Grau A, Herrero P, Torrell H, Canela N, Arola L.** Effects from diet-induced gut microbiota dysbiosis and obesity can be ameliorated by fecal microbiota transplantation: A multiomics approach. *PLoS One* 2019;14(9):e0218143.
62. **Saad MJA, Santos A, Prada PO.** Linking Gut Microbiota and Inflammation to Obesity and Insulin Resistance. *Physiol.* 2016;31(4):283–293.
63. **Li Q, Chang Y, Zhang K, Chen H, Tao S, Zhang Z.** Implication of the gut microbiome composition of type 2 diabetic patients from northern China. *Sci. Rep.* 2020;10(1):5450.

64. **Ottosson F, Brunkwall L, Ericson U, Nilsson PM, Almgren P, Fernandez C, Melander O, Orho-Melander M.** Connection Between BMI-Related Plasma Metabolite Profile and Gut Microbiota. *J. Clin. Endocrinol. Metab.* 2018;103(4):1491–1501.

65. **Naderpoor N, Mousa A, Gomez-Arango L, Barrett H, Nitert MD, de Courten B.** Faecal Microbiota Are Related to Insulin Sensitivity and Secretion in Overweight or Obese Adults. *J. Clin. Med.* 2019;8(4):452.

66. **Cani PD, Amar J, Iglesias MA, Poggi M, Knauf C, Bastelica D, Neyrinck AM, Fava F, Tuohy KM, Chabo C, Waget A, Delmée E, Cousin B, Sulpice T, Chamontin B, Ferrières J, Tanti J-F, Gibson GR, Casteilla L, Delzenne NM, Alessi MC, Burcelin R.** Metabolic endotoxemia initiates obesity and insulin resistance. *Diabetes* 2007;56(7):1761–1772.

67. **Perry RJ, Peng L, Barry NA, Cline GW, Zhang D, Cardone RL, Petersen KF, Kibbey RG, Goodman AL, Shulman GI.** Acetate mediates a microbiome–brain–β-cell axis to promote metabolic syndrome. *Nature* 2016;534(7606):213–217.

68. **Flint HJ, Scott KP, Duncan SH, Louis P, Forano E.** Microbial degradation of complex carbohydrates in the gut. *Gut Microbes* 2012;3(4):289–306.

69. **Thackray VG.** Sex, Microbes, and Polycystic Ovary Syndrome. *Trends Endocrinol. Metab.* 2019;30(1):54–65.

70. **Chan JL, Pall M, Ezeh U, Mathur R, Pisarska MD, Azziz R.** Screening for Androgen Excess in Women: Accuracy of Self-Reported Excess Body Hair Growth and Menstrual Dysfunction. *J. Clin. Endocrinol. Metab.* 2020;105(10). doi:10.1210/clinem/dgz264.

71. **Day F, Karaderi T, Jones MR, Meun C, He C, Drong A, Kraft P, Lin N, Huang H, Broer L, Magi R, Saxena R, Laisk T, Urbanek M, Hayes MG, Thorleifsson G, Fernandez-Tajes J, Mahajan A, Mullin BH, Stuckey BGA, Spector TD, Wilson SG, Goodarzi MO, Davis L, Obermayer-Pietsch B, Uitterlinden AG, Anttila V, Neale BM, Jarvelin M-R, Fauser B, Kowalska I, Visser JA, Andersen M, Ong K, Stener-Victorin E, Ehrmann D, Legro RS, Salumets A, McCarthy MI, Morin-Papunen L, Thorsteinsdottir U, Stefansson K, Team 23andMe Research, Styrkarsdottir U, Perry JRB, Dunaif A, Laven J, Franks S, Lindgren CM, Welt CK.** Large-scale genome-wide meta-analysis of polycystic ovary syndrome suggests shared genetic architecture for different diagnosis criteria. *PLoS Genet.* 2018;14(12):e1007813–e1007813.

72. **Moreno I, Simon C.** Relevance of assessing the uterine microbiota in infertility. *Fertil. Steril.* 2018;110(3):337–343.

## FIGURES

**Figure 1**: Landscape of microbiome composition in the cohort. (A) Pie chart indicates the average relative abundances of the top major phyla in the cohort. (B) Mean relative abundance of the core microbiome (taxa present in over 95% of individuals) of the cohort at genus level.

**Figure 2**: Comparison of the gut microbiome diversity between control and PCOS samples. (A) Box-plots of the Shannon diversity Index (alpha diversity), median values and interquartile ranges have been indicated on the plots. Beta diversity is represented by Principal Coordinate Analysis (PCoA) based on unweighted UniFrac (B) and weighted UniFrac (C) distances. Each plot point represents a single individual, the shapes indicate study groups (rhombus: controls, star: PCOS).

**Figure 3**: *Selbal* analysis results. The balance is made out of two groups of taxa: *Paraprevotella-Streptococcus* and *Eubacterium ventriosum* group-*Bifidobacterium*. The boxplots characterize the distribution of the balance scores for PCOS women and healthy controls. The right part of the figure holds the ROC curve its AUC value (0.643) and density curve. Abbreviations: TPR: true positive rate; FPR: false positive rate; ROC: receiver operating characteristic curve; AUC: area under the curve; PCOS: polycystic ovary syndrome.

**Figure 4**: Correlation heatmap of hormonal and metabolic biomarkers at the age 46 and top 10 bacteria associated with differences between PCOS and control women. Key: * FDR ≤ 0.2; ** FDR ≤ 0.1. The color key indicates the correlation direction (blue: negative correlation, red: positive correlation). Abbreviations: FDR: false discovery rate; BMI: body mass index; Hb1Ac: glycated haemoglobin; SHGB: Sex Hormone Binding Globulin.

**Figure 5**: Comparison of the gut microbiome diversity between NGT (n = 76) and preT2D samples (n = 14) among women with PCOS. (A) Box-plots of the Shannon diversity Index (alpha diversity),

median values and interquartile ranges have been indicated in the plots ($p = 0.018$). (B) Beta diversity is represented by Principal Coordinate Analysis (PCoA) based on unweighted UniFrac distance ($p = 0.003$). (C) Genus *Dorea* with statistically significant differences ($p = 0.03$). NGT: normal glucose tolerance; preT2D: pre-type 2 diabetes.

**Table 1**. Clinical, metabolic, and hormonal characteristics of study population.

| | 31 years | | | 46 years | | |
|---|---|---|---|---|---|---|
| | Control n = 142-202 | PCOS n = 75-102 | p-value[a] | Control n = 200-202 | PCOS n = 99-102 | p-value[a] |
| BMI, kg/m$^2$ | 25.09±4.81 | 25.35±5.41 | 0.92 | 27.44±4.98 | 26.96±5.08 | 0.82 |
| SHBG, nmol/L | 50.60±15.09 | 40.71±13.90 | 0.21 | 58.76±30.99 | 55.76±27.07 | 0.64 |
| Testosterone, nmol/L | 1.14±0.61 | 1.33±0.62 | 0.01 | 0.86±0.33 | 0.93±0.32 | 0.08 |
| FAI | 3.51±3.82 | 5.70±3.07 | 0.29 | 1.77±1.11 | 2.04±1.10 | 0.01 |
| Menses per year | 12.15±1.25 | 10.43±2.73 | $1.10\times10^{-12}$ | ND | ND | ND |
| Infertility problems ever in life, *n* | 28 (14%) | 38 (38%) | $2.98\times10^{-06}$ | 24 (13%) | 37 (39%) | $1.21\times10^{-06}$ |
| How many times given birth | 1.51±1.26 | 1.5±1.19 | 0.97 | 2.38±1.41 | 2.25±1.56 | 0.13 |
| How many miscarriages | 0.21±0.51 | 0.24±0.63 | 0.90 | 0.46±1.05 | 0.47±0.97 | 0.71 |
| Fasting glucose, mmol/L | 4.96±0.49 | 5.15±1.24 | 0.19 | 5.05±0.92 | 5.38±0.58 | 0.55 |
| 2h glucose, mmol/L | ND | ND | ND | 5.88±1.63 | 5.69±1.45 | 0.24 |
| Fasting insulin, mU/L | 8.52±3.79 | 9.14±4.99 | 0.70 | 9.82±5.65 | 10.2±7.63 | 0.79 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 2h insulin, mU/L | ND | ND | ND | 63.03±50.61 | 57.34±41.81 | 0.62 |
| Matsuda Index | ND | ND | ND | 4.84±2.55 | 5.24±3.35 | 0.77 |
| Disposition Index | ND | ND | ND | 186.15±87.32 | 189.47±89.59 | 0.74 |

Data are presented as mean ± standard deviation for continuous traits and as absolute proportions and prevalence (%).

[a] Wilcoxon sign rank test (continuous variable) or Fisher's Exact test (categorical variable). The number of women in separate analyses varies due to non-response to some items. **Abbreviations**: BMI: body mass index; FAI: Free Androgen Index; *n*: number of individuals; SHBG: Sex hormone binding globulin; y: year; ND: no data.

Figure-1

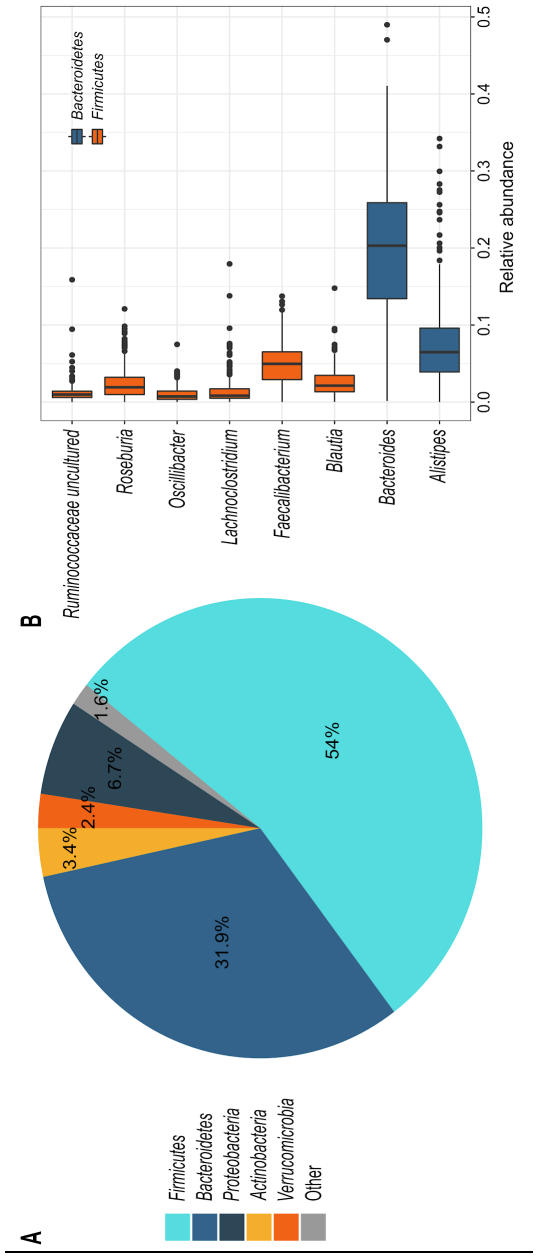

A
Firmicutes
Bacteroidetes
Proteobacteria
Actinobacteria
Verrucomicrobia
Other
54%
31.9%
3.4%
2.4%
6.7%
1.6%
B
Bacteroidetes
Firmicutes
Relative abundance
0.0
0.1
0.2
0.3
0.4
0.5
Ruminococcaceae uncultured
Roseburia
Oscillibacter
Lachnoclostridium
Faecalibacterium
Blautia
Bacteroides
Alistipes

Figure-2

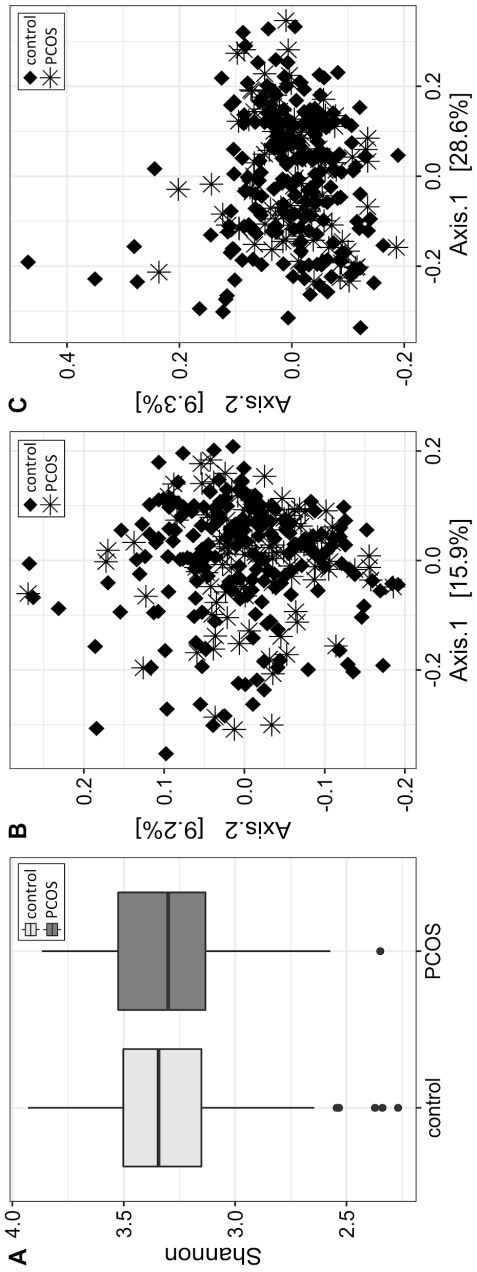

A
Shannon
4.0
3.5
3.0
2.5
control
PCOS
control
PCOS
B
Axis.2 [9.2%]
0.2
0.1
0.0
-0.1
-0.2
-0.2
0.0
0.2
Axis.1 [15.9%]
control
PCOS
C
Axis.2 [9.3%]
0.4
0.2
0.0
-0.2
-0.2
0.0
0.2
Axis.1 [28.6%]
control
PCOS

Figure-3

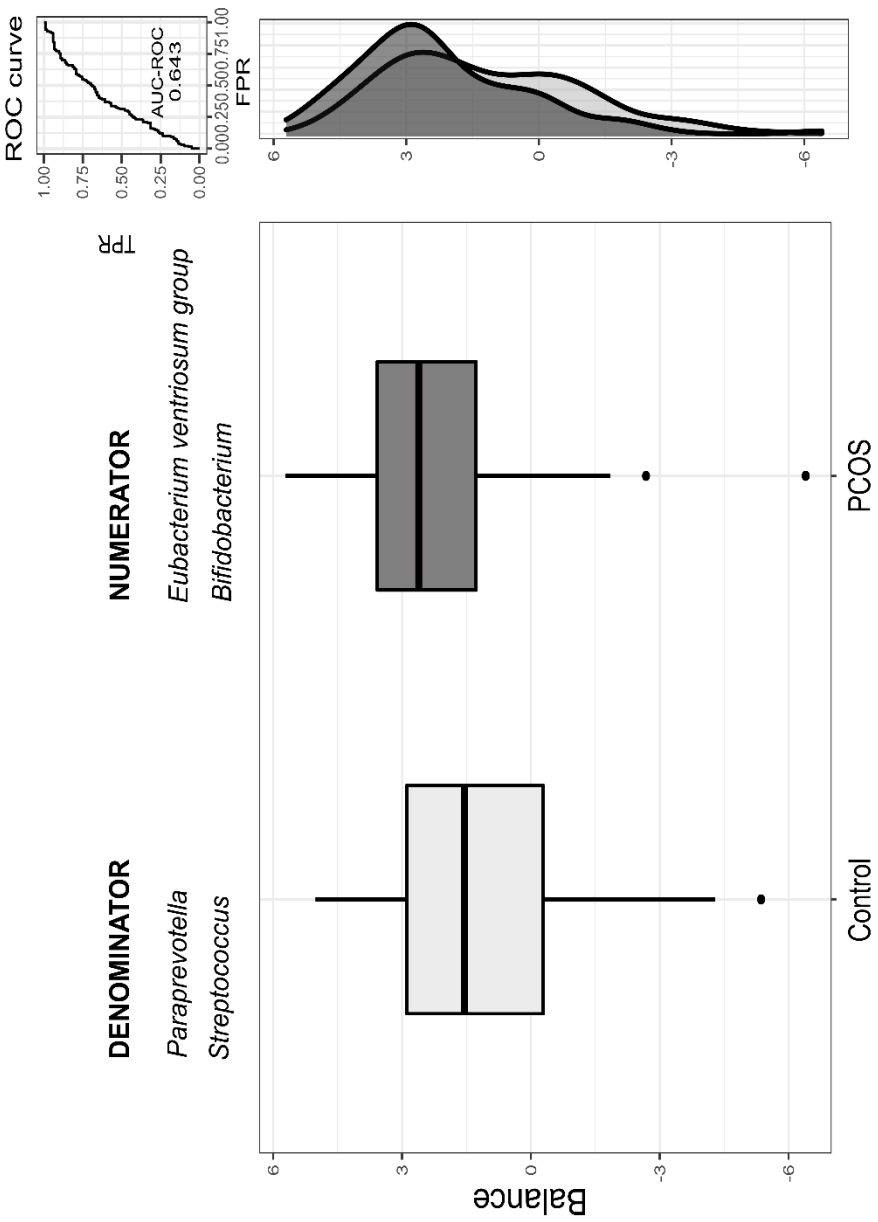

ROC curve
AUC-ROC
0.643
TPR
FPR
NUMERATOR
Eubacterium ventriosum group
Bifidobacterium
DENOMINATOR
Paraprevotella
Streptococcus
PCOS
Control
Balance

Figure-4

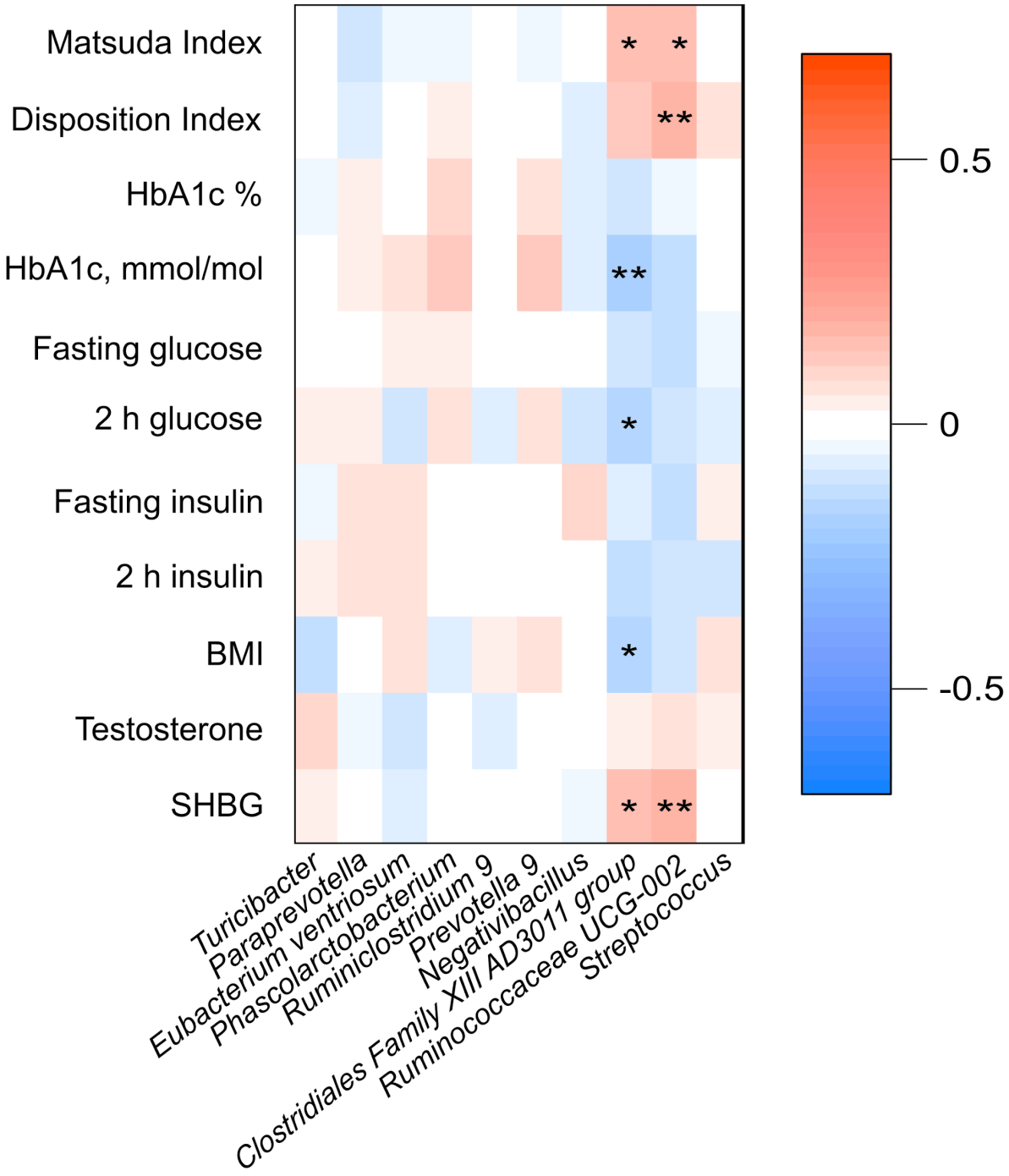
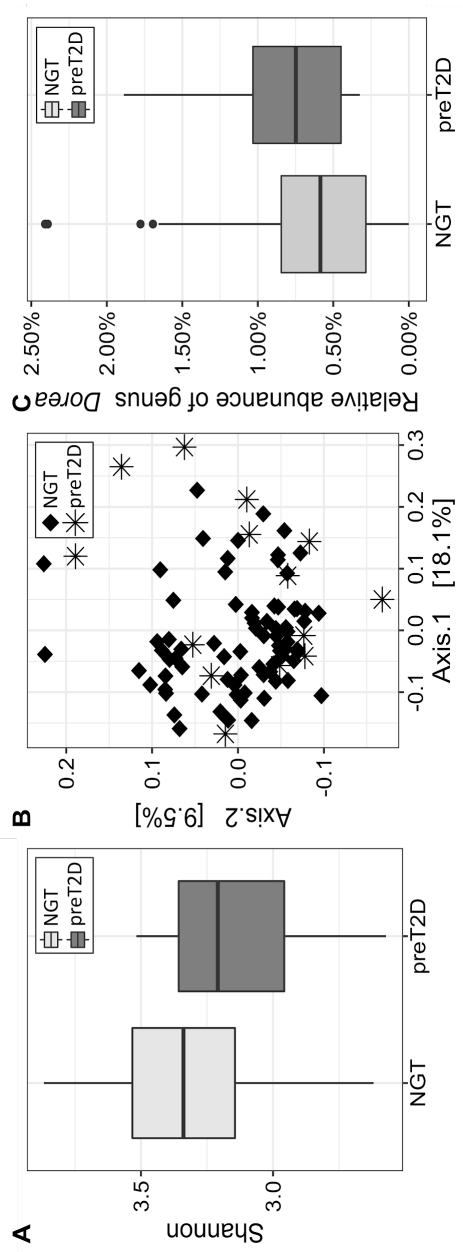

Matsuda Index
Disposition Index
HbA1c %
HbA1c, mmol/mol
Fasting glucose
2 h glucose
Fasting insulin
2 h insulin
BMI
Testosterone
SHBG
Turicibacter
Paraprevotella
Eubacterium ventriosum
Phascolarctobacterium
Ruminiclostridium 9
Prevotella 9
Negativibacillus
Clostridiales Family XIII AD3011 group
Ruminococcaceae UCG-002
Streptococcus
0.5
0
-0.5

Figure-5


A
Shannon
3.5
3.0
NGT
preT2D
NGT
preT2D
B
Axis.2 [9.5%]
0.2
0.1
0.0
-0.1
-0.1
0.0
0.1
0.2
0.3
Axis.1 [18.1%]
NGT
preT2D
C
Relative abunance of genus Dorea
2.50%
2.00%
1.50%
1.00%
0.50%
0.00%
NGT
preT2D
NGT
preT2D