



Interaction Between Functionally Activate Endometrial Microbiota and Host Gene Regulation in Endometrial Cancer

Peigen Chen¹, Yingchun Guo¹, Lei Jia¹, Jing Wan², TianTian He², Cong Fang^{1*} and Tingting Li^{1*}

OPEN ACCESS

¹ Reproductive Medicine Center, The Sixth Affiliated Hospital, Sun Yat-sen University, Guangzhou, China, ² The Third Affiliated Hospital, Sun Yat-sen University, Guangzhou, China

Edited by:

Maria Jose Saez Lara,
University of Granada, Spain

Reviewed by:

Xie Dewei,
Meijo University, Japan
Fang Lixin,
Nagoya University, Japan
Reza Nedaeinia,
Isfahan University of Medical
Sciences, Iran

*Correspondence:

Tingting Li
litt33@mail.sysu.edu.cn
Cong Fang
fangcong@mail.sysu.edu.cn

Specialty section:

This article was submitted to
Molecular and Cellular Pathology,
a section of the journal
Frontiers in Cell and Developmental
Biology

Received: 18 June 2021

Accepted: 01 September 2021

Published: 23 September 2021

Citation:

Chen P, Guo Y, Jia L, Wan J, He T,
Fang C and Li T (2021) Interaction
Between Functionally Activate
Endometrial Microbiota and Host
Gene Regulation in Endometrial
Cancer.
Front. Cell Dev. Biol. 9:727286.
doi: 10.3389/fcell.2021.727286

Objective: In this study, we mainly explored two questions: Which microorganisms were functionally active in the endometrium of patients with endometrial cancer (EC)? What kind of response did the human host respond to functionally active microorganisms?

Methods: Nine endometrial cancer patients and eight normal subjects were included in this study. HMP Unified Metabolic Analysis Network 3 (HUMAN3) was used to obtain functional information of microorganisms. In addition, metaCyc-based GSEA functional enrichment analysis was used to obtain information on the metabolic pathways of the human host. At the same time, the O2PLS model and Spearman correlation analysis were used to analyze the microorganisms–host interaction.

Results: With the novel metatranscriptome analysis pipeline, we described the composition of more than 5,000 functionally active microorganisms and analyzed the difference in microorganisms between the EC and the normal group. Our research found that these microorganisms were involved in part of the metabolic process of endometrial cancer, such as 6-sulfo-sialyl Lewis x epitope, *N*-acetyl-beta-glucosaminyl. In addition, the host–microbiota crosstalk of EC endometrium also included many biological processes, mainly functions related to tumor migration and the Apelin signaling pathway.

Conclusion: The functionally active microorganisms in the EC endometrium played an essential role in the occurrence and migration of tumors. This meant that functionally active microorganisms could not be ignored in the treatment of endometrial cancer. This study helped to better understand the possible role of endometrial functional, active microorganisms in the occurrence and development of EC in patients with endometrial cancer and provided new information for new attempts to treat EC.

Keywords: endometrial cancer, host–microbe interactions, gene regulation, microbiome, tumor migration

INTRODUCTION

Endometrial cancer (EC) is the fourth most common malignant tumor in women (Miller et al., 2020). The number of new endometrial cancers worldwide is about 382,100 per year, and nearly 89,900 patients die from endometrial cancer each year (Ferlay et al., 2019). In addition, there are nearly 64,000 new cases in China every year, resulting in 16,000 deaths (Chen W. et al., 2018). Moreover, this number was still rising, which was a considerable challenge.

For an extended period, the uterus was considered sterile. With the development of detection technology, the endometrium was proved to have its resident microbiota, the microbial community (Koedooder et al., 2019). The microbiota was closely related to the occurrence and development of cancer. For example, it could cause cancer by inhibiting cell apoptosis, stimulating proliferation, and interfering with genome stability (Baker et al., 2018). A study based on 16S sequencing technology by Walther-Antonio et al. (2016) showed that endometrial cancer patients and endometrial hyperplasia patients had changes in the microbiota structure of the endometrium, and there was a significant difference in the structure of the endometrial microbiota in mild cases. Another study by Lu et al. (2021) suggested that the endometrial microbiota were closely associated with the disorder of inflammatory cytokines in EC. These studies confirmed that the endometrial microbiota might play an essential role in EC. However, the host's response mechanism had not yet been elucidated.

In addition, the current standard methods of microbial research, such as 16s rRNA and metagenome analyses, were all based on DNA sequences. However, the existence of characteristic DNA sequences was not equal to the existence of living microorganisms. Furthermore, since DNA molecules might exist for decades (Glassing et al., 2016), the DNA sequence might originate from the decomposition of microorganisms, such as DNA from dead microorganisms (Aagaard et al., 2014). Therefore, research methods based on DNA sequence could not confirm the existence of microbiomes (Glassing et al., 2016; de Goffau et al., 2018; Sola-Leyva et al., 2021).

In this study, based on the functional, active endometrial microbiota mapping pipeline using metatranscriptome (meta-RNA sequencing analysis) (Macklaim and Gloor, 2018; Liu et al., 2020; Sola-Leyva et al., 2021), we explored the changes and activity of endometrial microbiota in EC. Metatranscriptomics could analyze microbial transcript profiles using RNA-seq data to identify living microorganisms and their functions (Macklaim and Gloor, 2018; Liu et al., 2020). At the same time, we combined with the host endometrial transcriptome to analyze the effects of the microbiota on the endometrium or the host's response to changes in the microbiota.

MATERIALS AND METHODS

Study Material

The data of nine endometrial cancer patients and eight control subjects were included in this study. The raw SRA files of RNA-seq were extracted from PRJNA612305 from the Gene Expression

Omnibus (GEO) repository. As mentioned earlier (DiGuardo et al., 2021), the Qubit RNA BR assay (Invitrogen, Carlsbad, CA, United States) was used to quantify total RNA, and the DV200 value (percentage of RNA fragments >200 bases in length) was used to assess the quality using the 2100 Bioanalyzer RNA Nano kit (Agilent Technologies, Santa Clara, CA, United States). The clinical data of the subject cohort included in this study were: EC cohort: $n = 9$, age: 68.2 ± 10.3 years (mean \pm standard deviation) and CON cohort: $n = 8$, age: 61.8 ± 13.5 years.

Functionally Active Endometrial Microbiota Mapping

First, the raw SRA files were converted into FASTQ format by using SRA toolkit (parameter `-split-files`) (Staff, 2011). Then, these FASTQ files were processed by using fastp (Chen S. et al., 2018) to remove reads containing adapters, more than 10% unknown nucleotides (N), and more than 50% of low-quality (Q -value ≤ 20) bases. Subsequently, the remaining sequences were aligned to the human reference genome GRCh38 from Gencode v26 using HISAT2 (Kim et al., 2015). The comparison results were stored in a separate .sam file. At the same time, the sequence (non-human sequence) that cannot be aligned with the human reference genome was stored as a separate FASTQ file (parameter `-un-conc-gz`). Next, the Kraken2 reference database was used to align these non-human sequences and output in mpa style (parameter `-use-mpa-style`) (Wood et al., 2019). Next, the reference database (including Bacteria, Archaea, and Viruses library) was downloaded using the Kraken-build utility. The scope of the database included the classification information of the National Center for Biotechnology Information (NCBI) and the complete genome sequence of RefSeq. Then metagenomeSeq R package (Paulson et al., 2013) was used to identify the different species in microorganismal communities of EC and normal (control) group with the directly assigned read counts. First, the results of the Kraken2 comparison were further carried out to remove rare species (at least in four samples, the count was greater than one) and normalized by using the Cumulative Sum Scale (CSS) method (Paulson et al., 2013). The process of the CSS algorithm was mainly to obtain a percentile by dividing the raw count by the cumulative sum of the counts. In this way, the relatively constant count distribution in the dataset was captured to process the raw count. The advantage of this treatment was that compared to ratio-based normalization or random sampling methods, CSS had a higher sensitivity (Paulson et al., 2013). Subsequently, the normalized data were used for differential species abundance analysis, using a zero-inflated log-normal model in metagenomeSeq.

Functional Enrichment Analysis of Endometrium Microbial Metabolic Pathways

The HMP Unified Metabolic Analysis Network 3 (HUMAN3) (Franzosa et al., 2018) was used to describe the metabolic potential of members of a microbial population based on the MetaCyc database. Then we used the HUMAN3_renorm_table script to normalize the data for count per million (CPM). Finally, all the results were merged into one file through

the HUMAnN_join_tables utility. Next, different metabolic pathways were screened by STAMP software (version 2.1.3) (Parks et al., 2014).

Host Transcriptome Analysis

Samtools (Li et al., 2009) was first used to transfer the sam file (as mentioned before, aligned to the human reference genome) to bam file. We then used featureCounts (Liao et al., 2014) to quantify the gene expression value. Based on the raw count calculated by featureCounts, DEseq2 R package (Love et al., 2014) was used to select the differential expression genes between EC and normal group with \log_2 -fold-change $\geq |1|$ and adj p -value < 0.05 . Since microbial metabolic pathways were analyzed based on the MetaCyc database, we used the Pathway Tools utility to obtain all human metabolic pathways (HUMAN2cyc, version: 24.5) in the MetaCyc database. The obtained pathway .col file contains a total of 390 human metabolic pathways. Then we converted the .col file into a gene set enrichment analysis (GSEA) input file in GMT format and used the GSEAPreranked method in GSEA to perform enrichment analysis of human metabolic pathways based on the results of the difference analysis (Subramanian et al., 2005).

Comprehensive Analysis of Interactions Between Endometrial Microbial Community Changes and Host Gene Dysregulations

The Two-way Orthogonal PLS (O2PLS) model (Trygg and Wold, 2003) showed excellent performance in the integrated analysis of multi-omics data, and its estimated value was close to the actual

parameters in both low-dimensional and high-dimensional data (Bouhaddani et al., 2016). This study used the O2PLS model to conduct a comprehensive analysis of endometrial microbial community changes and host gene dysregulations using OmicsPLS R package (Bouhaddani et al., 2018). First, we used the crossval_o2m_adjR2() function to determine the best parameters through sevenfold cross-validation. Then o2m() function was used for modeling. Next, ClusterProfiler R package (Yu et al., 2012) was performed for functional enrichment analysis of the top 50 genes loading.

In order to analyze the activities of each different species, Spearman correlation analysis was used to explore the relationship between species and species, species and host genes with correlation coefficient ≥ 0.6 , and p -value ≤ 0.05 . Compared with other methods (such as Pearson), the normalized count (gene expression) and component data (relative abundance of microbiota) of Spearman's correlation analysis performed better (Weiss et al., 2016). Furthermore, to study the host's response to the species, we used the metascape (Zhou et al., 2019) tool to perform functional analysis on the relevant genes of the species. Finally, the Cytoscape software (version 3.8.2) was used to build the interaction network (Shannon et al., 2003).

RESULTS

Mapping the Functional Endometrial Microbiota

Through the metatranscriptome analysis pipeline, we identified and analyzed the RNA of microorganisms present in the endometrium. Among the detected RNA sequences, the

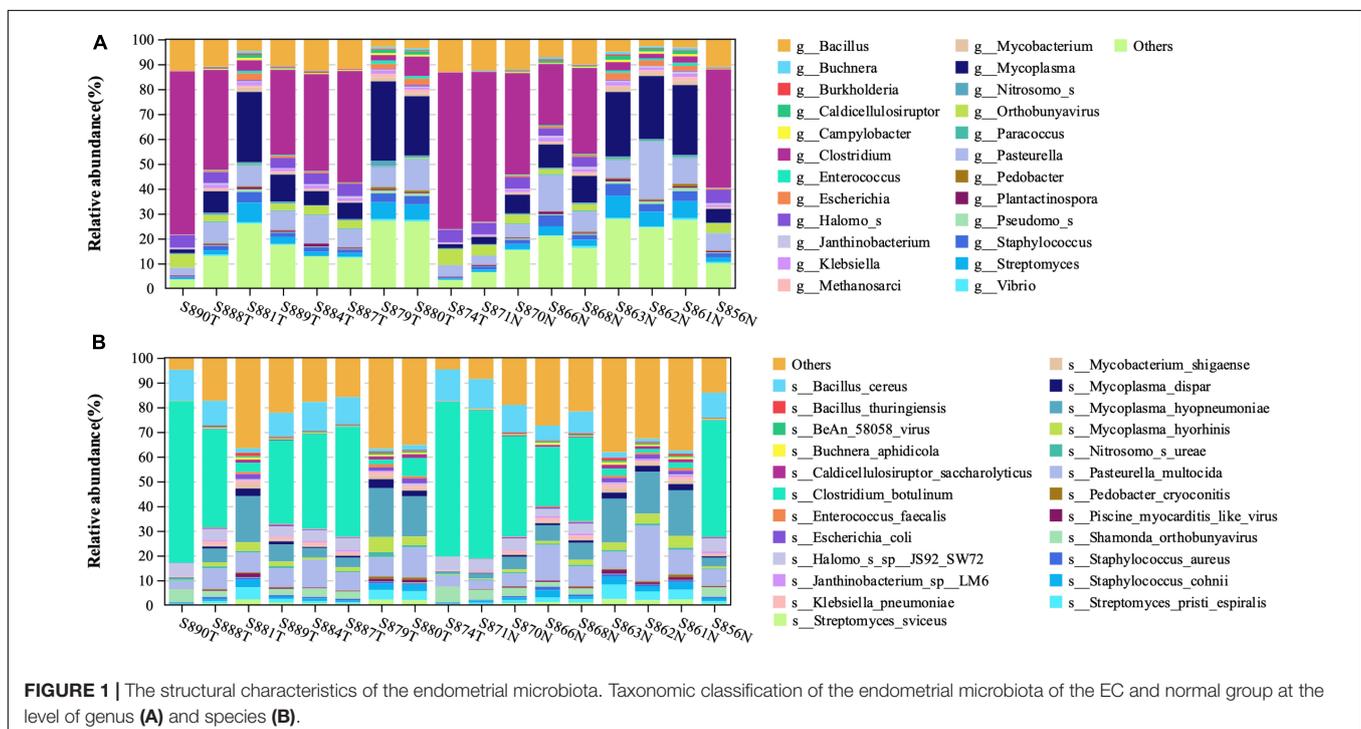


TABLE 1 | The species identified by metagenomeSeq.

Species	LogFC	p-values
<i>Borrelia coriaceae</i>	1.5971	0.0009
<i>Mycetocola sp. 449</i>	1.5590	0.0187
<i>Anaerostipes hadrus</i>	1.3859	0.0292
<i>Altererythrobacter namhicola</i>	1.2837	0.0100
<i>Leptospira biflexa</i>	1.2646	0.0251
<i>Nitrospirillum amazonense</i>	1.2608	0.0319
<i>Modestobacter marinus</i>	1.2418	0.0420
<i>Gramella forsetii</i>	1.2058	0.0194
<i>Pantoea ananatis</i>	1.1819	0.0079
<i>Hymenobacter nivis</i>	-1.2132	0.0022
<i>Potamipivirus A</i>	-1.2217	0.0426
<i>Acinetobacter equi</i>	-1.2241	0.0481
<i>Corynebacterium flavescens</i>	-1.2801	0.0451
<i>Streptococcus mitis</i>	-1.2959	0.0298
<i>Mycoplasma californicum</i>	-1.3346	0.0405
<i>Pannonibacter phragmitetus</i>	-1.7912	0.0287
<i>Corynebacterium ureiceleivorans</i>	-2.0840	0.0039

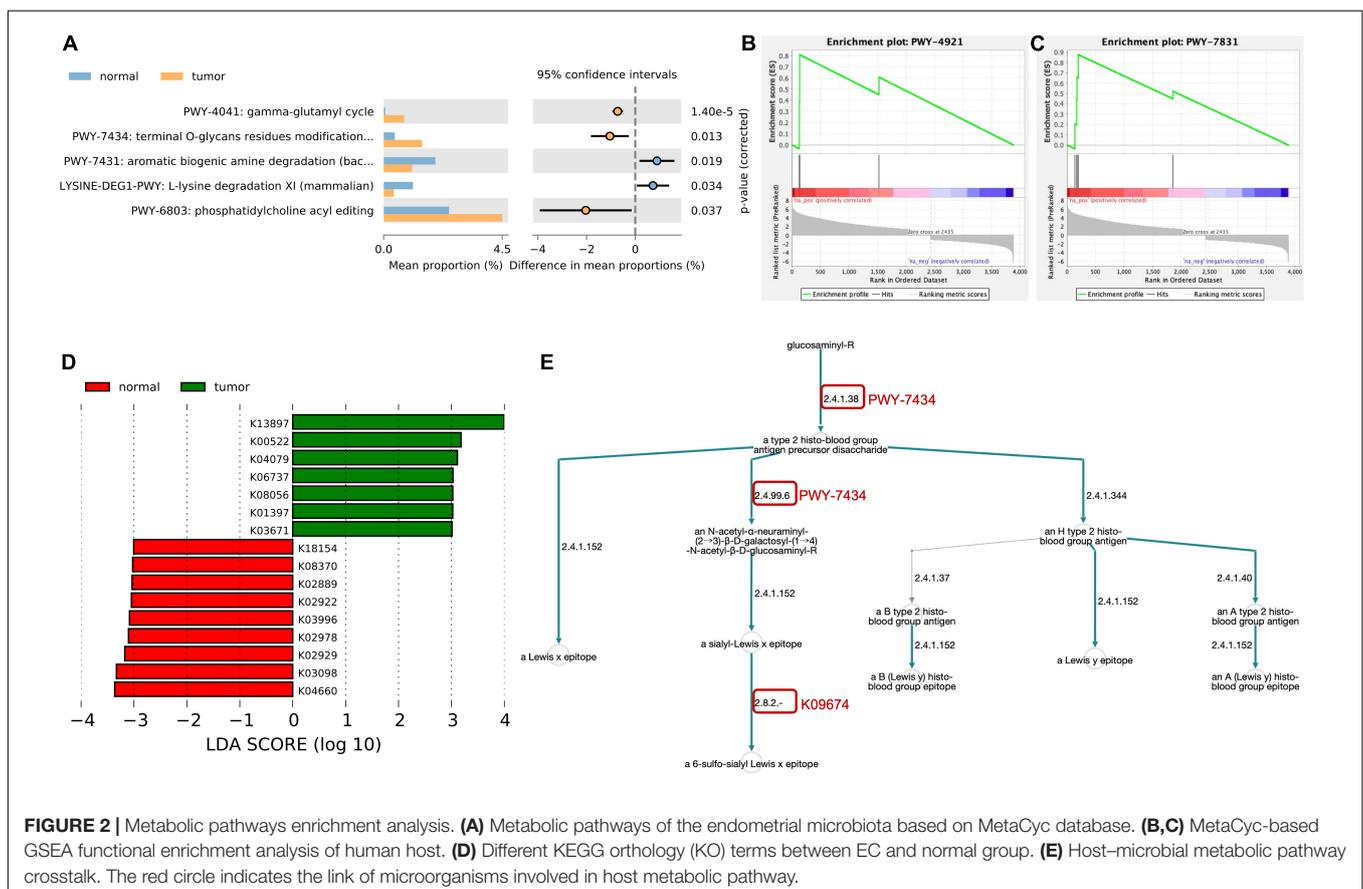
sequences that failed to be aligned to the human reference genome were about 10%. By using Kraken2 for comparison, a total of 5576 kinds of transcriptionally active *Bacteria* and 381 kinds of transcriptionally active *Archaea* were

identified (Supplementary Table 1). Figure 1 shows the top 25 abundance genus (Figure 1A) and species (Figure 1B) in each sample. The top four most abundant bacterial species in the endometrium of EC patients were *Clostridium botulinum*, *Mycoplasma hyopneumoniae*, *Bacillus cereus*, and *Pasteurella multocida*. At the same time, we noticed that the relative abundance of most species is less than 1%, which was consistent with the low abundance of endometrial biomass (Garcia-Grau et al., 2019).

Next, we discussed the species differences in the endometrium between EC and normal groups. Among the transcriptionally active microorganisms detected in endometrial samples, we found significant differences in the abundance of 17 kinds of microorganisms in EC endometrium and normal endometrium (Table 1).

Metabolic Pathway Enrichment Analysis Revealed Possible Host–Microbe Interactions

To study the possible role of transcriptionally active microorganisms in the endometrium, we analyzed the metabolic pathways of the host and microbiota based on the MetaCyc database. Among the detected microorganisms, three metabolic pathways were significantly enriched in the endometrium of EC patients (Figure 2A and Supplementary Table 2). In the



human host, based on the results of DESeq2 analysis, 3885 mRNAs were found to be differentially expressed between the EC group and the normal group (\log_2 -fold-change $\geq |1|$, $FDR \leq 0.05$, **Supplementary Table 3**). The metabolic pathway enrichment analysis of differential genes found that two pathways were significantly enriched, namely, PWY-4921 (protein citrullination, **Figure 2B**) and PWY-7831 (ABH and Lewis epitopes biosynthesis type 2 precursor disaccharide, **Figure 2C**). Furthermore, as shown in **Figure 2E**, we found that the endometrial microbiota of the EC group were involved in multiple links in the metabolic pathway of the host PWY-7831 (marked by the red circle in the figure).

HUMAN3 was also used for KEGG orthology (KO) enrichment analysis (**Supplementary Table 4**) and used linear discriminant analysis effect size (LEfSe) (Segata et al., 2011) to screen for differences in KO between the two groups. As a result, we obtained seven KOs that were significantly enriched in the endometrium of the EC group: cystatin-SN (K13897), ferritin heavy chain (K00522), molecular chaperone HtpG (K04079), tumor-associated calcium signal transducer

1 (K06737), protein disulfide isomerase family A, member 3 (K08056), matrix metalloproteinase-7 (matrilysin and uterine) (K01397), and thioredoxin 1 (K03671) ($LDA \geq 3$, **Figure 2D**).

Integrated Analysis of Interactions Between Host and Microbiota

To further explore the relationship between the activities of endometrial microorganisms and host endometrial gene disorders, we constructed an O2PLS model (**Figures 3A,B**). Through sevenfold cross-validation, it was determined that the modeling parameters were $n = 2$, $nx = 1$, $ny = 1$, and the $MSE = 127.60$ at this time. The model's loading diagram (**Figure 3A**) and the evaluation parameters of the model ($R^2X: 0.942$, $R^2Y: 0.519$) explain that the model construction is relatively satisfactory. The top 10 loading microorganisms and the top 20 loading genes are displayed in **Figure 3B**.

We conducted a KEGG enrichment analysis of the top 50 loading genes and found that they were mainly enriched

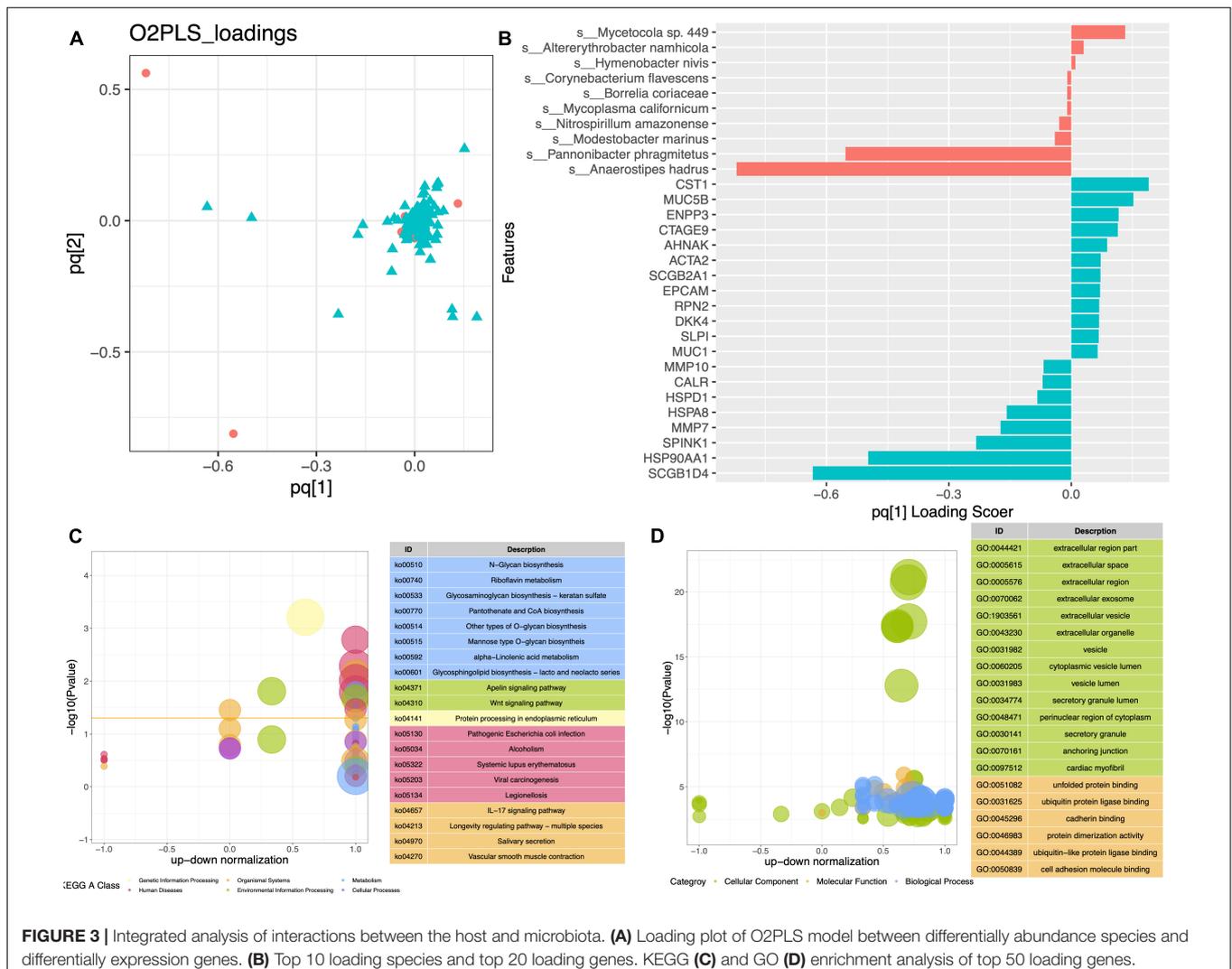
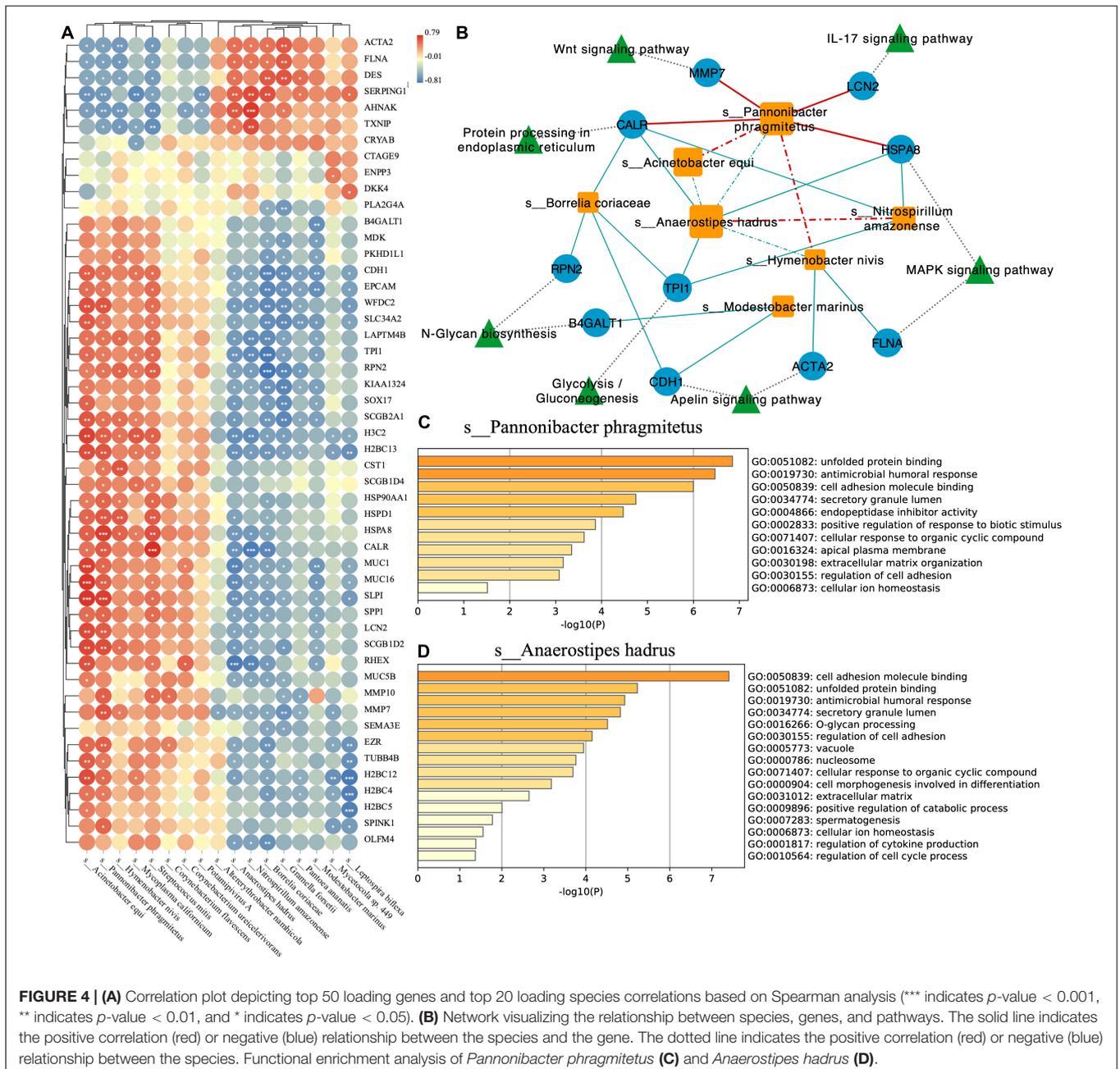


FIGURE 3 | Integrated analysis of interactions between the host and microbiota. **(A)** Loading plot of O2PLS model between differentially abundance species and differentially expression genes. **(B)** Top 10 loading species and top 20 loading genes. KEGG **(C)** and GO **(D)** enrichment analysis of top 50 loading genes.

in metabolic-related pathways. In addition, the Wnt signaling pathway, Apelin signaling pathway, and IL-17 signaling pathway were also significantly enriched (Figure 3C). From the results of GO enrichment analysis (Figure 3D), we could see that the main focus was on the functions related to cell adhesion, such as extracellular region part, extracellular space, and extracellular region.

To explore the interaction between species and the host and the relationship between species, we conducted Spearman correlation analysis with the correlation coefficient ≥ 0.6 , p -value ≤ 0.05 as threshold (Figure 4A). At the same time, we used Cytoscape software to visualize the network

(Figure 4B). We found that *Pannonibacter phragmitetus*, which had a high abundance in EC endometrium, was significantly positively correlated with Wnt signaling pathway, IL-17 signaling pathway, and MAPK signaling pathway (Figure 4B). For the two species with high abundance in the two groups (EC: *Anaerostipes hadrus*, Normal: *P. phragmitetus*), we used the metascape tool (Zhou et al., 2019) to perform functional enrichment analysis on their positively (*P. phragmitetus*, Figure 4C) or negatively (*A. hadrus*, Figure 4D) related genes. We found that the two have standard functions, cell adhesion molecule binding (GO: 0050839). In addition, we also found that the negatively related genes of *A. hadrus* were



significantly enriched in other functional items related to cell adhesion, such as extracellular matrix and regulation of cell adhesion (Figure 4C).

DISCUSSION

For a long time, microorganisms, including bacteria and viruses, have been thought to play an essential role in the occurrence and development of cancer. However, the elucidation of related mechanisms is challenging. Current animal experiments show that the carcinogenic effects of microorganisms may be more related to the overall changes in the microbiome rather than due to a single species (Schwabe and Jobin, 2013).

DNA is a very stable molecule, and unlike RNA that is rapidly degraded, it can exist for decades. Microorganisms are found in almost all body parts, including the endometrium (Glassing et al., 2016), so it is difficult for us to distinguish whether the detected species is active when using DNA-based detection technology for microbial detection.

In this study, our first concern was the composition and structure of functionally active microorganisms in the endometrium of patients with endometrial cancer and the difference between the structures of the microorganisms in the average population. We used the novel metatranscriptomics pipeline to study the functional and active microorganisms of the endometrium of patients with endometrial cancer. The analytical reliability of this novel analysis pipeline had been confirmed in the study of Sola-Leyva et al. (2021). Based on the analysis results, we found that the top five phyla that were significantly enriched in the endometrium of patients with endometrial cancer were *Firmicutes*, *Proteobacteria*, *Tenericutes*, *Actinobacteria*, and *Bacteroidetes* (Supplementary Figure 1). This was consistent with the research results of Walther-Antonio et al. (2016), which once again confirmed the reliability of the novel pipeline used in this study. At the same time, we also got the difference between the two groups of microorganisms. As far as we know, this was the first study to describe the difference in functional species composition at the species level.

Another issue we were concerned about was how these different species interact with the human host. Through enrichment analysis of metabolic pathways based on the MetaCyc database, we found that the microorganisms of EC endometrium participated in the metabolic reaction process of EC endometrium from *N-acetyl-beta-glucosaminyl* to *6-sulfo-sialyl Lewis x epitope* and promoted *6-sulfo -Biosynthesis of sialyl Lewis x epitope*. *6-Sulfo-sialyl Lewis x epitope* was highly expressed in a variety of tumors, and its role was mainly to promote tumor metastasis by adhering tumor cells to blood endothelial cells (Nakagoe et al., 2002). However, it had not been reported in endometrial cancer. *N-acetyl-beta-glucosaminyl* metabolism key enzyme β -*N-acetylglucosaminylglycopeptide* β -*1,4-galactosyltransferase* was closely related to the occurrence of epithelial-to-mesenchymal transition (EMT) in tumors (Zhang et al., 2011; Khan et al., 2018). In our last part of the results, we also found that the highly abundant species in EC had an inhibitory effect on the biosynthesis of N-Glycan. Therefore,

we speculated that endometrial microorganisms participate in tumor migration by affecting the metabolic activities of the host's endometrium.

In addition, by establishing a high-performance O2PLS model, we had further studied the functions of different species. The enrichment analysis results of the top 50 loading genes were roughly the same as the enrichment results of metabolic pathways, such as functions related to tumor migration (*N-glycan* biosynthesis and *Glycosaminoglycan* biosynthesis). At the same time, we found that these species with high abundance in the EC endometrium were closely related to activating the Apelin signaling pathway, which had been confirmed to be related to the increased risk of endometrial cancer (Yang et al., 2016). Another important finding was that species with high abundance in EC had an inhibitory effect on the processing of endoplasmic reticulum proteins. Thus, we speculated that endometrial microorganisms also played an essential role in the unfolded protein response (UPR).

The uterus was a site with very low microbial biomass. In this study, we used a novel analysis pipeline to describe the composition of functionally active microorganisms in the endometrium of endometrial cancer. Exploring the interaction mechanism between these active microorganisms and the host was the focus of our research. As far as we know, this was the first study on the mechanism of microbe–host interaction. We believe that this could provide new ideas for the treatment of endometrial cancer. However, this article still had limitations. The sample size included in this study was small, and the sample size needed to be expanded in the future for more in-depth research.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

AUTHOR CONTRIBUTIONS

CF and PC carried out the study. PC and YG analyzed and interpreted the data. TL and TH drafted the manuscript. CF, PC, and TL coordinated the study, participated in the design, and reviewed the manuscript. TL, LJ, JW, and TH took part in revising the manuscript. All authors read and approved the final manuscript.

FUNDING

This work was supported by the National Natural Science Foundation of China (Grant No. 81871214), the National Key R&D Program of China (Grant No. 2017YFC1001603), and the National Natural Science Foundation of China (Grant No. 81070495).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fcell.2021.727286/full#supplementary-material>

Supplementary Figure 1 | Pie chart of taxonomic classification of EC patients at the phylum level.

Supplementary Table 1 | Abundance of functionally active microorganisms in all samples.

Supplementary Table 2 | Metabolic pathways of functionally active microorganisms.

Supplementary Table 3 | Differentially expressed genes between the EC group and the normal group.

Supplementary Table 4 | KEGG orthology (KO) enrichment analysis.

REFERENCES

- Aagaard, K., Ma, J., Antony, K. M., Ganu, R., Petrosino, J., and Versalovic, J. (2014). The placenta harbors a unique microbiome. *Sci. Transl. Med.* 6:237ra65. doi: 10.1126/scitranslmed.3008599
- Baker, J. M., Chase, D. M., and Herbst-Kralovetz, M. M. (2018). Uterine microbiota: residents, tourists, or invaders? *Front. Immunol.* 9:208. doi: 10.3389/fimmu.2018.00208
- Bouhaddani, S. E., Houwing-Duistermaat, J., Salo, P., Perola, M., Jongbloed, G., and Uh, H. W. (2016). Evaluation of O2PLS in omics data integration. *BMC Bioinformatics* 17(Suppl. 2):11. doi: 10.1186/s12859-015-0854-z
- Bouhaddani, S. E., Uh, H. W., Jongbloed, G., Hayward, C., Klaric, L., Kielbasa, S. M., et al. (2018). Integrating omics datasets with the OmicsPLS package. *BMC Bioinformatics* 19:371. doi: 10.1186/s12859-018-2371-3
- Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34, i884–i890.
- Chen, W., Sun, K., Zheng, R., Zeng, H., Zhang, S., Xia, C., et al. (2018). Cancer incidence and mortality in China, 2014. *Chin. J. Cancer Res.* 30, 1–12.
- de Goffau, M. C., Lager, S., Salter, S. J., Wagner, J., Kronbichler, A., Charnock-Jones, D. S., et al. (2018). Recognizing the reagent microbiome. *Nat. Microbiol.* 3, 851–853. doi: 10.1038/s41564-018-0202-y
- DiGuardo, M. A., Davila, J. I., Jackson, R. A., Nair, A. A., Fadra, N., Minn, K. T., et al. (2021). RNA-Seq reveals differences in expressed tumor mutation burden in colorectal and endometrial cancers with and without defective DNA-Mismatch repair. *J. Mol. Diagn.* 23, 555–564. doi: 10.1016/j.jmoldx.2021.01.008
- Ferlay, J., Colombet, M., Soerjomataram, I., Mathers, C., Parkin, D. M., Pineros, M., et al. (2019). Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. *Int. J. Cancer* 144, 1941–1953. doi: 10.1002/ijc.31937
- Franzosa, E. A., McIver, L. J., Rahnvard, G., Thompson, L. R., Schirmer, M., Weingart, G., et al. (2018). Species-level functional profiling of metagenomes and metatranscriptomes. *Nat. Methods* 15, 962–968. doi: 10.1038/s41592-018-0176-y
- Garcia-Grau, I., Simon, C., and Moreno, I. (2019). Uterine microbiome-low biomass and high expectations dagger. *Biol. Reprod.* 101, 1102–1114. doi: 10.1093/biolre/iy257
- Glassing, A., Dowd, S. E., Galandiuk, S., Davis, B., and Chiodini, R. J. (2016). Inherent bacterial DNA contamination of extraction and sequencing reagents may affect interpretation of microbiota in low bacterial biomass samples. *Gut Pathog.* 8:24.
- Khan, G. J., Gao, Y., Gu, M., Wang, L., Khan, S., Naeem, F., et al. (2018). TGF-beta1 causes EMT by regulating N-Acetyl glucosaminyl transferases via downregulation of non muscle myosin II-A through JNK/P38/PI3K pathway in lung cancer. *Curr. Cancer Drug Targets* 18, 209–219. doi: 10.2174/1568009617666170807120304
- Kim, D., Langmead, B., and Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* 12, 357–360. doi: 10.1038/nmeth.3317
- Koedooder, R., Mackens, S., Budding, A., Fares, D., Blockeel, C., Laven, J., et al. (2019). Identification and evaluation of the microbiome in the female and male reproductive tracts. *Hum. Reprod. Update* 25, 298–325. doi: 10.1093/humupd/dmy048
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352
- Liao, Y., Smyth, G. K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30, 923–930. doi: 10.1093/bioinformatics/btt656
- Liu, Y. X., Qin, Y., Chen, T., Lu, M., Qian, X., Guo, X., et al. (2020). A practical guide to amplicon and metagenomic analysis of microbiome data. *Protein Cell* 12, 315–330. doi: 10.1007/s13238-020-00724-8
- Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15:550.
- Lu, W., He, F., Lin, Z., Liu, S., Tang, L., Huang, Y., et al. (2021). Dysbiosis of the endometrial microbiota and its association with inflammatory cytokines in endometrial cancer. *Int. J. Cancer* 148, 1708–1716. doi: 10.1002/ijc.33428
- Macklaim, J. M., and Gloor, G. B. (2018). From RNA-seq to biological inference: using compositional data analysis in meta-transcriptomics. *Methods Mol. Biol.* 1849, 193–213. doi: 10.1007/978-1-4939-8728-3_13
- Miller, K. D., Fidler-Benaoudia, M., Keegan, T. H., Hipp, H. S., and Jemal, A. (2020). Siegel RL. cancer statistics for adolescents and young adults, 2020. *CA Cancer J. Clin.* 70, 443–459. doi: 10.3322/caac.21637
- Nakagoe, T., Fukushima, K., Itoyana, N., Ikuta, Y., Oka, T., Nagayasu, T., et al. (2002). Expression of ABH/Lewis-related antigens as prognostic factors in patients with breast cancer. *J. Cancer Res. Clin. Oncol.* 128, 257–264. doi: 10.1007/s00432-002-0334-5
- Parks, D. H., Tyson, G. W., Hugenholtz, P., and Beiko, R. G. (2014). STAMP: statistical analysis of taxonomic and functional profiles. *Bioinformatics* 30, 3123–3124. doi: 10.1093/bioinformatics/btu494
- Paulson, J. N., Stine, O. C., Bravo, H. C., and Pop, M. (2013). Differential abundance analysis for microbial marker-gene surveys. *Nat. Methods* 10, 1200–1202. doi: 10.1038/nmeth.2658
- Schwabe, R. F., and Jobin, C. (2013). The microbiome and cancer. *Nat. Rev. Cancer* 13, 800–812.
- Segata, N., Izard, J., Waldron, L., Gevers, D., Miropolsky, L., Garrett, W. S., et al. (2011). Metagenomic biomarker discovery and explanation. *Genome Biol.* 12:R60.
- Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi: 10.1101/gr.1239303
- Sola-Leyva, A., Andres-Leon, E., Molina, N. M., Terron-Camero, L. C., Plaza-Diaz, J., Saez-Lara, M. J., et al. (2021). Mapping the entire functionally active endometrial microbiota. *Hum. Reprod.* 36, 1021–1031. doi: 10.1093/humrep/deaa372
- Staff, S. (2011). *Using the Sra Toolkit to Convert Sra Files into Other Formats*. Bethesda, MD: National Center for Biotechnology Information (US).
- Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U S A* 102, 15545–15550. doi: 10.1073/pnas.0506580102
- Trygg, J., and Wold, S. (2003). O2-PLS, a two-block (X–Y) latent variable regression (LVR) method with an integral OSC filter. *J. Chemometrics* 17, 53–64. doi: 10.1002/cem.775
- Walther-Antonio, M. R., Chen, J., Multini, F., Hokenstad, A., Distad, T. J., Cheek, E. H., et al. (2016). Potential contribution of the uterine microbiome in the development of endometrial cancer. *Genome Med.* 8:122.
- Weiss, S., Van Treuren, W., Lozupone, C., Faust, K., Friedman, J., Deng, Y., et al. (2016). Correlation detection strategies in microbial data sets vary widely

- in sensitivity and precision. *ISME J.* 10, 1669–1681. doi: 10.1038/ismej.2015.235
- Wood, D. E., Lu, J., and Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. *Genome Biol.* 20:257.
- Yang, Y., Lv, S. Y., Ye, W., and Zhang, L. (2016). Apelin/APJ system and cancer. *Clin. Chim. Acta* 457, 112–116. doi: 10.1016/j.cca.2016.04.001
- Yu, G., Wang, L. G., Han, Y., and He, Q. Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 16, 284–287. doi: 10.1089/omi.2011.0118
- Zhang, H., Meng, F., Wu, S., Kreike, B., Sethi, S., Chen, W., et al. (2011). Engagement of I-branching {beta}-1, 6-N-acetylglucosaminyltransferase 2 in breast cancer metastasis and TGF- β signaling. *Cancer Res.* 71, 4846–4856. doi: 10.1158/0008-5472.can-11-0414
- Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A. H., Tanaseichuk, O., et al. (2019). Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat. Commun.* 10:1523.
- Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- Publisher’s Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Chen, Guo, Jia, Wan, He, Fang and Li. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.