Supplementary File S1. Methodology for TRACE-seq-based metatranscriptomic analysis of host response and microbiome in vagina of women.

1. Metatranscriptomic library preparation and sequencing of vaginal samples

The whole procedure was summarized in Supplementary File S2. Generally, vaginal swabs from patients and healthy controls were inactivated before further handling of the specimens as per China biosafety requirements. Total nucleic acids were extracted using QIAamp Viral RNA Mini Kit (Qiagen) following the manufacturer's instructions. Metatranscriptomic libraries were prepared following the workflow of TRACE-seq, using TruePrep® RNA Library Prep Kit for Illumina (Vazyme, TR502-01) according to the manufacturer's instructions but with several modifications: 1) total RNA was extracted from each vaginal swab and directly used for each library without removing rRNA; 2) gDNA removal was performed at 42℃ for 10 min, instead of 2 min; 3) reverse transcription was conducted in the presence of random hexamers and oligo(dT)20VN primers; and 4) the final concentration of each PCR primer (N5 and N7 primers) in the PCR reaction was reduced to 0.2 μM. After 18 PCR cycles, the TRACE-seq libraries were purified by 0.8X Agencourt AMPure XP beads (Beckman Coulter) and eluted in 20 μL nuclease-free water. The concentration of resulting libraries was determined by Qubit 2.0 fluorometer with the Qubit dsDNA HS Assay kit (Invitrogen) and the size distribution of libraries was assessed by Agilent 4200 TapeStation instrument (Agilent Technologies). Finally, libraries were sequenced on Hiseq X TEN platform (Illumina) which generated 2 x 150 bp of paired-end raw reads.

2. Data pre-processing

Raw sequencing reads were firstly subjected to Trim Galore (v0.6.4_dev) (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) to trim adaptor

sequence and bases with sequencing quality lower than 20. In addition, Poly A tails were trimmed by Trim Galore --poly A. After all the trimming processes, reads longer than 20nt were utilized for downstream analyses.

3. Host transcriptional profiling analysis

Reads mapped to human rRNA were firstly filtered out by Bowtie2 (v2.2.9) [1]. Remaining reads were mapped to human genome (hg19) and transcriptome using STAR (v2.7.1a) [2]. The expression level of annotated genes was determined by cuffnorm (v2.2.1) [3], and genes with FPKM > 1 were considered to be expressed. Multi-dimensional scaling (MDS) and differential gene expression analysis were performed using EdgeR (v3.28.1) [4], based on gene count data generated by HTSeq (v0.11.2) [5]. The identified significantly up- and down-regulated genes were respectively subjected to Gene Ontology Enrichment Analysis using DAVID (v6.8) [6]. GO terms were further simplified using REViGO [7]. The top 10 enriched representative GO terms of biological process were displayed. Protein interaction analysis was performed using STRING [8].

4. Discrimination of sars-cov-2 from meta-transcriptomic data

After removing human reads, the remaining reads were aligned to the reference genome of Wuhan-Hu-1 (GenBank accession number: NC_045512) using Bowtie2 (v2.2.9) [1] for SARS-CoV-2 identification.

5. Microbiome analysis

After human reads removal, microbial taxonomic classification was performed based on the remaining reads using Kraken2 (v2.0.8-beta) [9] with a custom database.

To build the custom database, standard RefSeq complete bacterial genomes were downloaded through “kraken2-build --download-library bacteria” and complete genomes of human viruses and genome assemblies of fungi were downloaded from NCBI’s RefSeq and added to the custom database’s genomic library using the “--add-to-library” switch. The rarefaction curves were plotted by rarecure function from vegan package version 2.5-6 (https://CRAN.R-project.org/package=vegan). Principal coordinate analysis (PCoA) was done using cmdscale command in R based on relative abundance of microbial taxa at the genus level. Distances between samples were determined using clark dissimilarity index by vegdist command from vegan package version 2.5-6 (https://CRAN.R-project.org/package=vegan). The expression profiles of antimicrobial resistance genes were demonstrated by aligning non-host reads to the Comprehensive Antibiotic Resistance Database (CARD) [10]. Antimicrobial resistance genes with more than 10 completely matching reads were considered to be expressed. All corresponding graphs were plotted using custom R scripts by RStudio (v1.2.5033) (https://rstudio.com/).

6. Data availability

The raw sequence data was deposited in China National Microbiology Data Center (www.nmdc.cn) with accession numbers NMDC40003555 to NMDC40003575.

References

[1] Langmead B, Salzberg SL. Fast gapped-read alignment with bowtie 2. Nat Methods, 2012, 9: 357-359

[2] Dobin A, Davis CA, Schlesinger F*, et al.* Star: Ultrafast universal rna-seq aligner. Bioinformatics, 2013, 29: 15-21

[3] Trapnell C, Williams BA, Pertea G*, et al.* Transcript assembly and quantification by rna-seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol, 2010, 28: 511-515

[4] Robinson MD, McCarthy DJ, Smyth GK. Edger: A bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics, 2010, 26: 139-140

[5] Anders S, Pyl PT, Huber W. Htseq--a python framework to work with high-throughput sequencing data. Bioinformatics, 2015, 31: 166-169

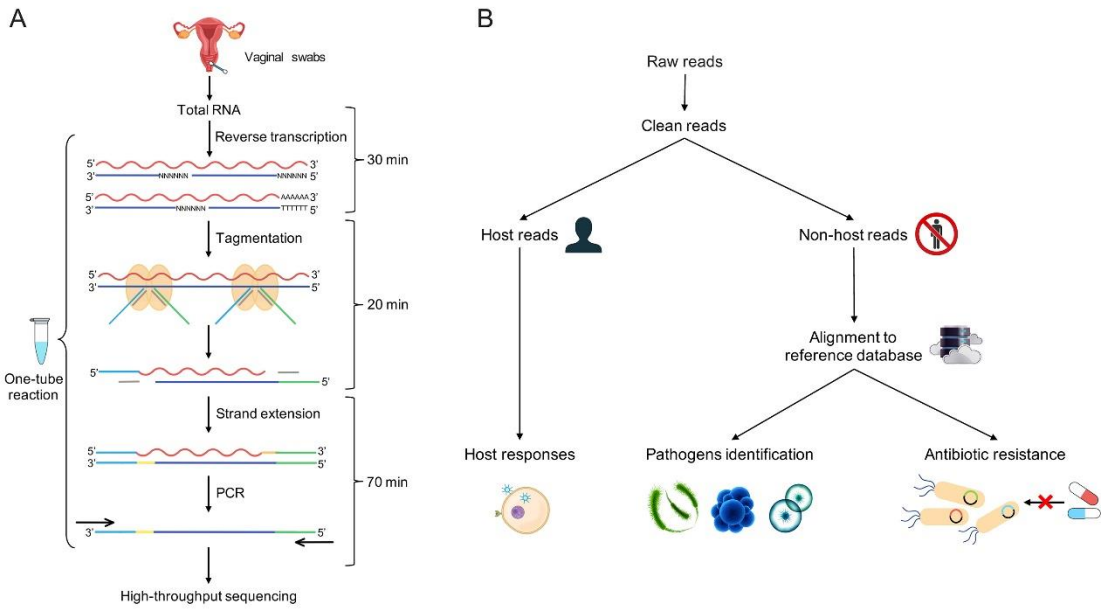
[6] Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using david bioinformatics resources. Nat Protoc, 2009, 4: 44-57

[7] Supek F, Bosnjak M, Skunca N*, et al.* Revigo summarizes and visualizes long lists of gene ontology terms. PLoS One, 2011, 6: e21800

[8] Szklarczyk D, Gable AL, Lyon D*, et al.* String v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res, 2019, 47: D607-D613

[9] Wood DE, Lu J, Langmead B. Improved metagenomic analysis with kraken 2. Genome Biol, 2019, 20: 257

[10] McArthur AG, Waglechner N, Nizam F*, et al.* The comprehensive antibiotic resistance database. Antimicrob Agents Chemother, 2013, 57: 3348-3357

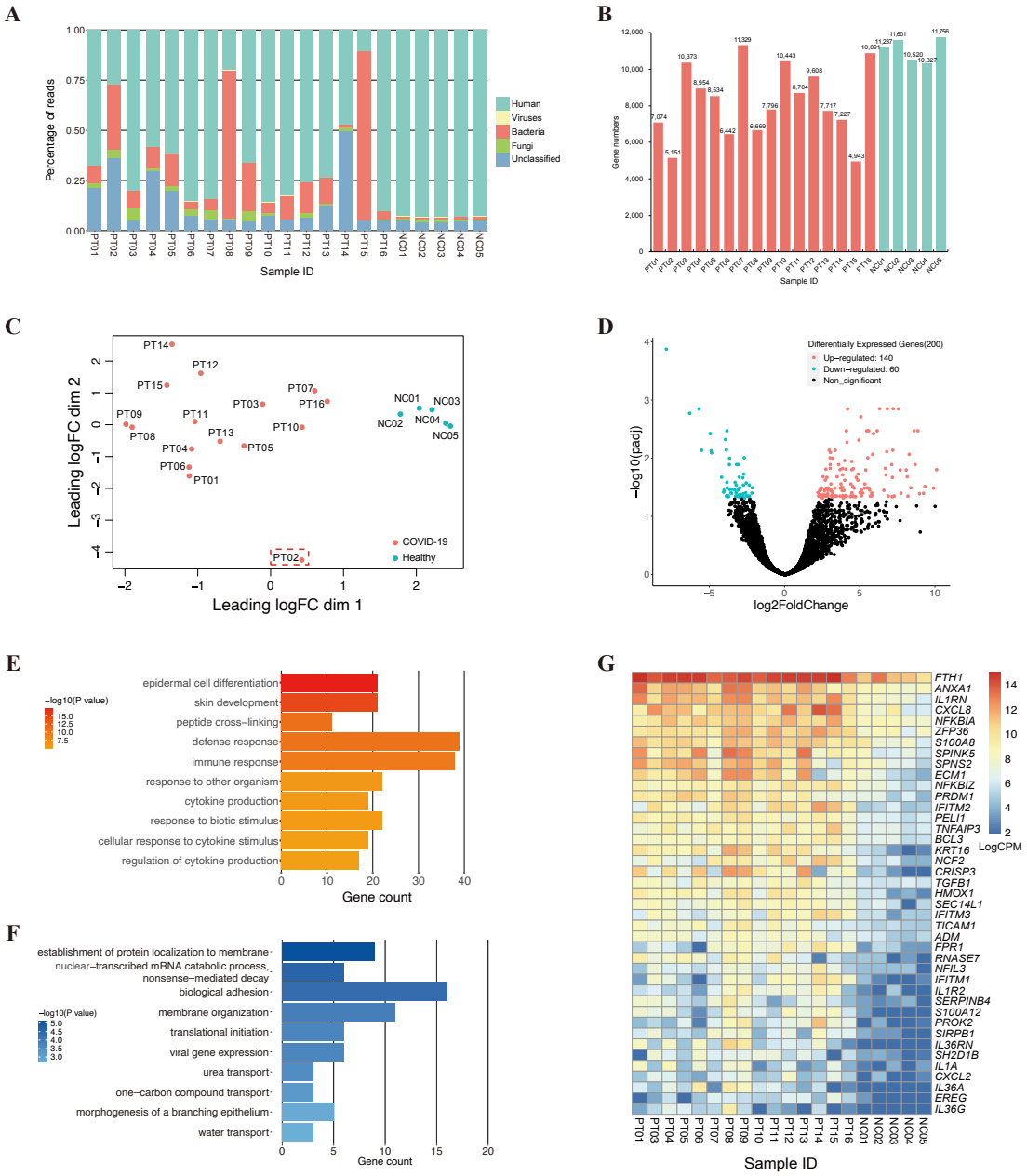


Supplementary File S2. Workflow of TRACE-seq for metatranscriptomic analysis.

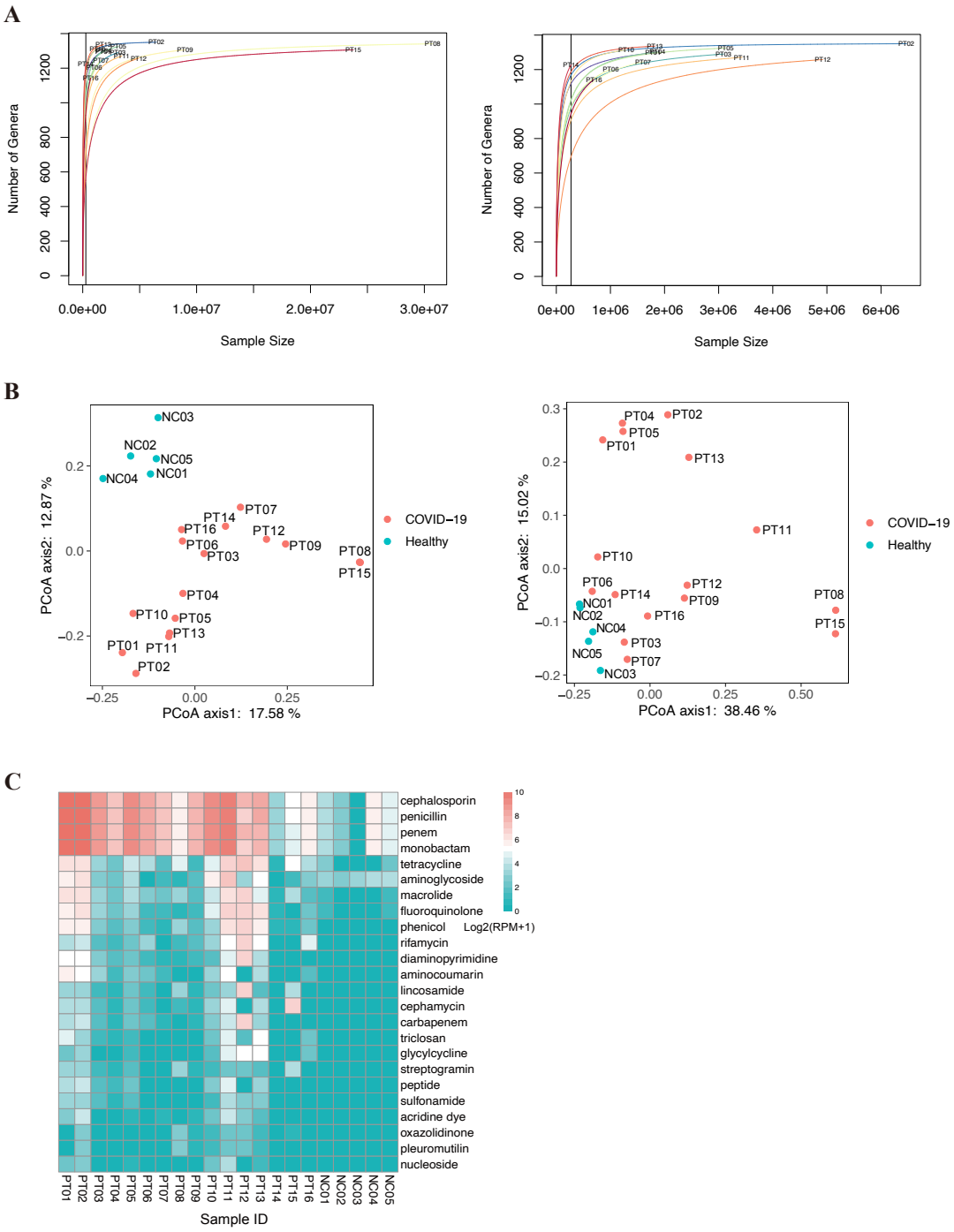
A. The wet-lab protocol of TRACE-seq starting with total RNA extracted from vaginal swab samples of COVID-19 patients. B. The dry-lab pipeline consisting of pathogens identification, antibiotic resistance profiling and host response characterization.

Supplementary File S3. Clinical features of women patients and healthy individuals involved in this study.

Individual ID	Age	Menopause	Primary disease	Illness onset	Initial symptoms	Chest CT-Bilateral ground-glass opacities	Diagnosed by	Severity of COVID-19	Days from onset to ICU	Mechanical ventilation	Lowest oxygen saturation %	Antiviral therapy	Use of broad range antimicrobials	Outcome	Days from onset to Sample collection
COVID-19 patients (n=16)															
PT01	63	Yes	—	Jan-28-2020	Cough, fever, diarrhea	Yes	Serology	Critical	17	Yes	92	Yes	Yes	Dead	27
PT02	55	Yes	Obesity	Jan-23-2020	Cough, fever	Yes	RT-PCR, serology	Critical	14	Yes	80	Yes	Yes	Hospitalized, Brain dead	32
PT03	66	Yes	—	Jan-28-2020	Cough, fever	Yes	RT-PCR, serology	Critical	19	Yes	93	Yes	Yes	Dead	29
PT04	71	Yes	Hypothyroidism	Jan-25-2020	Cough, fever	Yes	RT-PCR, serology	Critical	22	Yes	94	Yes	Yes	Dead	33
PT05	73	Yes	Diabetes, hypothyroidism	Feb-07-2020	Cough, fever	Yes	RT-PCR, serology	Critical	9	Yes	91	Yes	Yes	Recovered	20
PT06	57	Yes	—	Feb-01-2020	Cough, fever, diarrhea	Yes	Serology	Critical	7	Yes	65	Yes	Yes	Dead	27
PT07	65	Yes	Hypertension, diabetes, coronary disease	Jan-30-2020	Cough, fever	Yes	Serology	Critical	22	Yes	70	Yes	Yes	Recovered	29
PT08	52	Yes	Hypertension, coronary disease	Jan-25-2020	Cough, fever	Yes	RT-PCR, serology	Critical	19	Yes	94	Yes	Yes	Recovered	34
PT09	80	Yes	—	Jan-23-2020	Cough, fever	Yes	RT-PCR, serology	Critical	25	Yes	84	Yes	Yes	Dead	43
PT10	75	Yes	Hypertension, diabetes, coronary disease	Feb-06-2020	Cough, fever	Yes	RT-PCR, serology	Critical	31	No	99	Yes	Yes	Dead	34
PT11	72	Yes	Hypertension	Feb-03-2020	Cough, fever	Yes	RT-PCR, serology	Critical	38	Yes	70	Yes	Yes	Hospitalized, Unconscious	45
PT12	74	Yes	Hypertension	Jan-27-2020	Cough, fever	Yes	RT-PCR, serology	Critical	51	Yes	80	Yes	Yes	Dead	66
PT13	67	Yes	Hypertension	Jan-25-2020	Cough, fever	Yes	RT-PCR, serology	Critical	25	Yes	98	Yes	Yes	Dead	68
PT14	58	Yes	Hypertension, diabetes	Jan-28-2020	Cough	Yes	RT-PCR, serology	Critical	38	Yes	96	Yes	Yes	Dead	65
PT15	86	Yes	COPD, Senile dementia	Jan-22-2020	Cough, fever	Yes	RT-PCR, serology	Critical	59	No	99	Yes	Yes	Dead	71
PT16	68	Yes	Hypertension, diabetes	Feb-10-2020	Cough	Yes	RT-PCR, serology	Critical	27	Yes	96	Yes	Yes	Dead	52
Healthy individuals (n=5)															
NC01	57	Yes	Benign thyroid nodule	—	—	—	—	—	—	—	—	—	No	—	—
NC02	64	Yes	—	—	—	—	—	—	—	—	—	—	No	—	—
NC03	77	Yes	Parkinsonism, coronary insufficiency	—	—	—	—	—	—	—	—	—	No	—	—
NC04	68	Yes	Sjogren's syndrome	—	—	—	—	—	—	—	—	—	No	—	—
NC05	77	Yes	—	—	—	—	—	—	—	—	—	—	No	—	—



Supplementary File S4. Host transcriptome profiles in vaginal samples collect in the present study. A. The proportion of reads mapped to human, viruses, bacteria and fungi for the individual samples. B. Numbers of detected host genes in each vaginal sample. C. MDS plot based on host transcriptome profiles with raw gene expression count. The outlier is marked by the red dotted box. D. Volcano plot displaying differentially expressed host genes between SARS-CoV-2 positive and negative samples. E. Bar chart showing the top 10 up-regulated GO terms for biological process. F. Bar chart showing the top 10 down-regulated GO terms for biological process. G. Heatmap depicting the expression levels of genes involved in the immune response.



Supplementary File S5. Microbiome transcriptome profiles in vaginal samples collect in the present study. A. Genus-level rarefaction curves of all patient samples (left) and all patient samples except PT08, PT09 and PT15 (right). B. Principal Coordinates Analysis(PCoA) based on total vaginal microbial (left) and mycobiome (right) compositions. C. Antimicrobial resistance gene expression profiling (RPM: reads per million non-host reads).