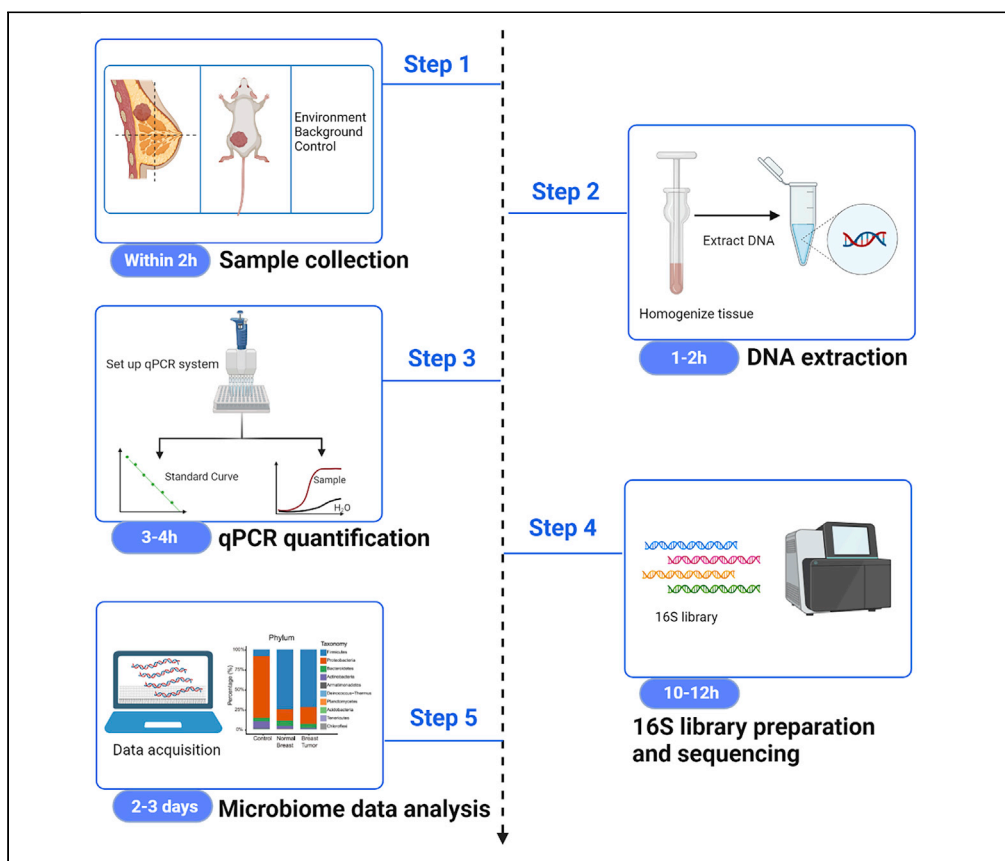# STAR Protocols

Protocol

Quantification and characterization of mouse and human tissue-resident microbiota by qPCR and 16S sequencing

Bingqing Yao, Tingting Dong, Aikun Fu, Hang Li, Chuhan Jiang, Nan Li, Shang Cai

yaobingqing@westlake.edu.cn (B.Y.)
caishang@westlake.edu.cn (S.C.)

Highlights

An optimized protocol for profiling tissue-resident bacteria by qPCR and 16S sequencing

Detailed steps for DNA extraction of low-biomass tissue-resident bacteria

Sensitive detection of 10^3–10^4 equivalent bacteria per gram of tissue with accuracy

The tissue-resident microbiota is an integral component of multiple tumor types, but it remains challenging to characterize its abundance and composition due to its low biomass. Here, we describe an optimized protocol for quantification and profiling of tissue-resident microbiota. The major optimized steps include DNA extraction, qPCR, 16S library construction, and bioinformatics analysis. This protocol enables robust and accurate characterization of the dynamics of normal and tumor tissue-resident microbiota at its physiological abundance from both mouse and human origins.

Publisher's note: Undertaking any experimental protocol requires adherence to local institutional guidelines for laboratory safety and ethics.

Protocol

Quantification and characterization of mouse and human tissue-resident microbiota by qPCR and 16S sequencing

Bingqing Yao,[1,2,3,5,*] Tingting Dong,[1,2,3] Aikun Fu,[2,3] Hang Li,[4] Chuhan Jiang,[2,3] Nan Li,[4] and Shang Cai[1,2,6,*]

[1]Fudan University, Shanghai 200000, China

[2]Westlake Laboratory of Life Sciences and Biomedicine, Hangzhou, Zhejiang 310024, China

[3]School of Life Sciences, Westlake University, Hangzhou, Zhejiang 310024, China

[4]Westlake University High-Performance Computing Center, Westlake University, Hangzhou, Zhejiang 310024, China

[5]Technical contact

[6]Lead contact

*Correspondence: yaobingqing@westlake.edu.cn (B.Y.), caishang@westlake.edu.cn (S.C.)
https://doi.org/10.1016/j.xpro.2022.101765

SUMMARY

The tissue-resident microbiota is an integral component of multiple tumor types, but it remains challenging to characterize its abundance and composition due to its low biomass. Here, we describe an optimized protocol for quantification and profiling of tissue-resident microbiota. The major optimized steps include DNA extraction, qPCR, 16S library construction, and bioinformatics analysis. This protocol enables robust and accurate characterization of the dynamics of normal and tumor tissue-resident microbiota at its physiological abundance from both mouse and human origins.

For complete details on the use and execution of this protocol, please refer to Fu et al. (2022).

BEFORE YOU BEGIN

The tissue-resident microbiota has a much lower biomass than the gut microbiome (10^{10} bacteria/gram) by several orders of magnitude. Therefore, there is a seriously higher host genome contamination when analyzing tissue-resident microbiota. Additionally, as bacterial load diminishes, environmental contamination becomes a major issue that can mask the real microbiota signal (Eisenhofer et al., 2019). Therefore, it is particularly challenging to quantify the absolute abundance and to profile the microbial community within tissues with accuracy and high sensitivity (Davis et al., 2018; de Goffau et al., 2018; Jervis-Bardy et al., 2015; Kim et al., 2017; Laurence et al., 2014; Salter et al., 2014).

The protocol below has been specifically optimized to characterize the abundance and composition of the tissue-resident microbiota using breast tissue. We also tested it in other tissues such as lung and skin, as well as on dissociated primary cells with success. The protocol describes the specific steps of how to efficiently extract DNA from tissue samples, how to quantify bacterial abundance sensitively and precisely by qPCR, and how to reliably construct 16S rRNA gene libraries and perform downstream bioinformatics analysis. This optimized protocol uses the QIAamp PowerFecal (pro) DNA kit (QIAGEN-#51804) for bacterial DNA isolation and the Illumina 16S rRNA gene amplicon sequencing protocol (Part #15042322) for library preparation.

We recommend reading the original protocols and literature before following the optimized protocol (Caporaso et al., 2011; Suzuki et al., 2000) (Amplicon, 2013; QIAGEN).

Institutional permissions (if applicable)

All animal experiments described in this protocol were carried out in compliance with Chinese laws and regulations. The local institutional animal ethics board of Westlake University (Institutional Animal Care and Use Committee) approved all mouse experiments (permission numbers: 19-001-CS). The committee approving the experiments and confirming that all experiments conform to the relevant regulatory standards are included under "Institutional permissions". All mice used were female.

All human samples were collected and analyzed after informed consent was obtained from the patients and according to IRB-approved protocols: IRB-2019-99 and IRB-2020-634. The sex, gender, and information about age is provided for all study participants in informed consent. Readers who wish to conduct their animal and human work as described in this protocol will need to obtain the permission from their IACUC and IRB committees following their institutional guidelines and regulations.

Overall guidelines for reagents and operations

1. All the tubes, pipette tips, razor blades, glass homogenizers, tweezers and scissors must be sterilized by autoclave in advance.

   ⚠ CRITICAL: Autoclave can only kill germs, but does not necessarily remove genetic material. Therefore, all the tools should be cleaned by bleach, 70% ethanol and water, before autoclaving. Tubes and tips should be sterile, DNase and RNase free.

2. As there are potential microbial contaminants in the environment within aerosols, dust, all the operations are required to be performed in a clean room and in a biosafety cabinet following all the sterilization rules required for cell culture and to avoid sample exposure to the environment as much as possible. At the same time, a tissue surrogate (usually same amount of PBS) must be included as an environment background control (EBC) and go through all the steps together with samples.
3. Besides the environmental contaminants, we noticed another important source of contamination is from reagents, especially from the DNA polymerase used for qPCR. Therefore, pre-screening and selection of a clean and efficient qPCR kit is key to the final sensitivity and accuracy of the experiment.

Prepare standards using *E. coli*

⏱ Timing: 2 days

To estimate the abundance of tissue-resident bacteria, we used *Escherichia coli* (TIANGEN # CB101-02) as positive control to establish the correlation between bacterial DNA concentration in the sample and the number of bacteria as colony forming units (CFU). This assay was adapted from Ibekwe et al. with optimization at specific steps (Ibekwe et al., 2002) (Figure 1).

4. Inoculate a single colony of *E. coli* into 4 mL LB medium and culture 12–16 h (overnight) at 37°C with agitation at 200 rpm.
5. On day 2, inoculate 20 μL *E. coli* overnight culture into 1 mL fresh LB medium at a ratio of 1:50. Grow *E. coli* at 37°C with agitation at 200 rpm until OD=0.8.

   Note: OD=0.8 means that bacteria growth reaches log phase. It usually takes around 2.5–3.5 h.
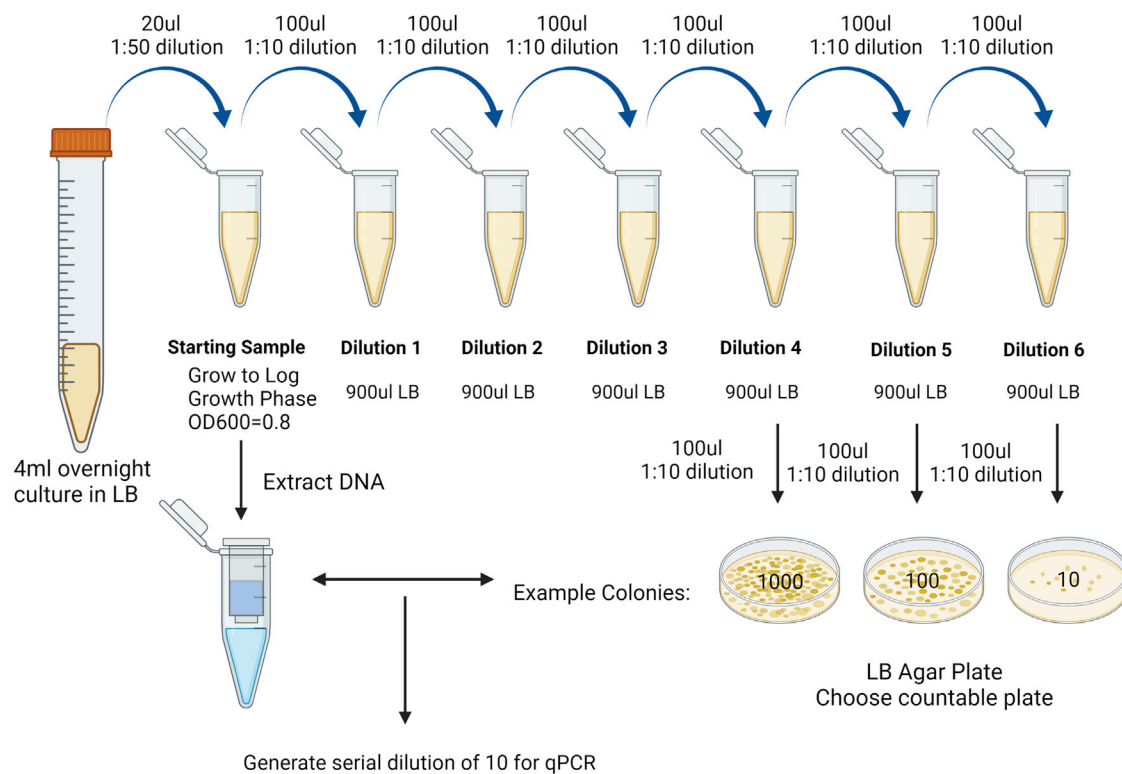
Figure 1. Workflow of setting up bacterial genome standards for qPCR

6. Set up six 1.5 mL Eppendorf tubes (EP tubes) with 900 μL of LB media for 10-fold serial dilutions 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}, 10^{-6} and label the tubes.

7. Add 100 μL starting culture into 900 μL LB medium labeled as Dilution 1 and vortex tube to mix gently but thoroughly.

8. Repeat step 4 to set up Dilution (N) by transferring 100 μL of Dilution (N-1) into 900 μL LB medium as shown in Figure 1.

9. Spread 100 μL bacteria from Dilutions 4, 5, and 6 onto LB plates. Each dilution needs triplicate plates.

10. Incubate the plates at 37°C for 12–16 h, then count colonies for each dilution of *E. coli* samples. Calculate the CFU of the original *E. coli* sample of OD=0.8.

   Note: The number of CFU of bacteria was determined by sample dilutions that give rise to countable colonies (usually between 10-1000 colonies per plate) on the plates. Samples with OD=0.8 usually contain 10^8 *E. coli*.

11. At the same time, centrifuge 1 mL bacteria culture of starting sample at 12,000 rpm for 5 min for DNA extraction by using QIAamp PowerFecal (pro) DNA kit (QIAGEN-#51804).

   Note: The DNA yield from 1 mL bacteria (OD=0.8) is around 24 ng/ μL (100 μL) in our experiment. Aliquot genomic DNA into several tubes and store at −20°C to avoid freeze and thaw cycles.

12. Dilute the bacterial genomic DNA with a 10-fold serial dilution. Use the serial dilution of bacterial genomic as standards to evaluate commercial qPCR kit.

   Note: To evaluate the quality of the qPCR kit, a 10-fold serial dilution of 10^8 bacteria from 24 ng/μL to 0.024 pg/μL was utilized.
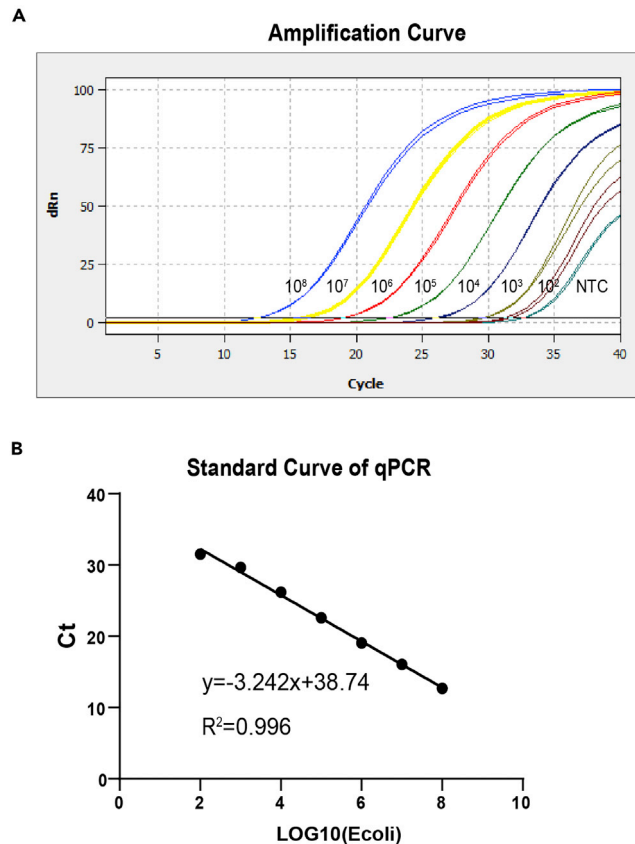
Figure 2. Standard curve of qPCR
(A) Amplification curves by qPCRsoft384 1.1 automatically.
(B) Standard curve of qPCR using a dilution series of *E. coli* reference genomic DNA.

Screen the contamination-low qPCR kit

🕐 Timing: 3–4 h

As qPCR reagents also contain a significant amount of bacterial DNA contamination, we set out to screen the best commercial qPCR kit with the best sensitivity (the lowest amount of bacteria DNA it can faithfully detect), specificity (e.g., Taqman probe qPCR has a higher specificity than SYBR green qPCR), and stability (smallest variation between various experiments). We strongly recommend testing each batch of commercial qPCR kits by standard curve before quantification of the real tissue sample.

To test the quality of qPCR kit, briefly, 20 μL reaction mix containing Premix Ex Taq (probe qPCR), 750 nM of forward primer, 500 nM reverse primer and 250 nM probe, and 2 μL sample DNA was loaded on the qPCR machine (Jena# Qtower3G or qTOWEP384/G). The reaction was programmed as follows: denaturation at 95°C for 2 min, 40 cycles of 95°C for 15 s, 60°C for 60 s. The standard of bacterial serial dilution genome can be used to evaluate commercialized qPCR kit (Figure 2). Refer to the qPCR quantification section for experimental details.

> **Note:** Various commercialized qPCR kits were tested for their sensitivity, specificity and stability for qPCR quantification of *E. coli* std dilutions. Then the Ct values were plotted over log scale of *E. coli* amount, and further subjected to linear regression. The slope of the standard curve was calculated by linear regression analysis, which can represent PCR amplification

efficiencies (E = ($10^{-1/slope}$) − 1). A reaction with 100% efficiency will generate a slope of −3.32. PCR efficiency of 90%–110% is acceptable (3.1<slope<3.6). R^2 of linear regression, which represents the variation of the data, should be more than 0.99. The sensitivity was determined by the lowest *E. coli* quantity that can be faithfully detected within the linear range. The Ct value of lowest *E. coli* std should be less than the Ct value of NTC, which represents bacteria DNA contamination in the qPCR reagents (Figure 2A). A good Ct value of NTC is usually greater than 32, therefore, any sample with Ct value greater than that of NTC is beyond the detection limit.

Breast tumor tissue collection and storage

⏱ Timing: within 2 h (tissue dissection time is not included)

13. For human breast tissue sample collection:
    a. Collect human samples in the sterile gross room and immediately transfer fresh tissues to germ-free conical tubes with 10 mL sterile DMEM culture medium.
    b. Transport the samples in a portable cooler with ice packs and process them within hours in the clean and sterile biosafety cabinet with autoclaved dissection tools.

Note: All human samples need to be de-identified. New DMEM should be aliquoted in the biosafety cabinet to keep it sterile.

14. For mouse breast tissue sample collection:

Collect normal or tumor breast tissue into a sterile EP tube after mice are sacrificed by cervical dislocation in the clean and sterile biosafety cabinet with autoclaved dissection tools. Keep the samples on ice.

Note: Fresh tissue for microbiota analysis is preferred. However, tissues can be stored in −80°C for future batch extraction of DNA. If no live cells are expected from the sample, flash frozen of the sample is recommended for sample preservation.

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Human breast tumor tissue sample, adult, female	The First Affiliated Hospital of Zhejiang University	IRB-2019-99
Human breast tumor tissue sample, adult, female	Zhejiang Cancer Hospital	IRB-2020-634
Bacterial and virus strains		
Escherichia coli	TIANGEN	CB101-02
Critical commercial assays		
QIAamp DNA Microbiome kit	QIAGEN	Cat# 51704
QIAamp PowerFecal (pro) DNA kit	QIAGEN	Cat#51804
Takara2 Premix Ex Taq	Takara	Cat#RR390A
TruePrep® Index Kit V3 for Illumina®	Vazyme	Cat#TD203
Dynabeads™ MyOne™ Streptavidin C1 Beads	Thermo Fisher Scientific	Cat #5002
Qubit DNA assay	Thermo Fisher Scientific	
AMPure XP beads	Beckman Coulter	Cat #A63881
KAPA Hyper Prep Kits	KAPA	Cat #KK8505

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
16S rRNA gene amplicon sequencing	BioProject: PRJNA681060	https://dataview.ncbi.nlm.nih.gov/object/PRJNA681060?reviewer=7v2h87ups0iauqrpdmae8ue64j
Experimental models: Organisms/strains		
FVB/N-Tg(MMTV-PyVT)634Mul/J PyMT+, female, 14w	The Jackson Laboratory	Cat#002374
Oligonucleotides		
16S rRNA qPCR forward primer	Suzuki et al. (2000)	CGGTGAATACGTTCYCGG
16S rRNA qPCR reverse primer	Suzuki et al. (2000)	GGWTACCTTGTTACGACTT
16S rRNA qPCR probe	Suzuki et al. (2000)	CTTGTACACACCGCCCGTC
biotinylated 515 Forward primer for 16S rRNA library preparation	This paper	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGTGYCAGCMGCCGCGGTAA
biotinylated 806 Reverse primer for 16S rRNA library preparation	This paper	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGGACTACNVGGGTWTCTAA
Software and algorithms		
FastQC	Babraham Institute	https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
Seurat v4	Satija Lab	https://github.com/satijalab/seurat
vsearch v2.14.2	Rognes et al. (2016)	N/A
usearch v10	Edgar (2010)	N/A
Other		
TissueLyser II	QIAGEN	Cat#85300
qTOWEP384/G-Analytik	Jena	Cat# SIA-PCR006
Fragment Analyzer-12/96	AATI	Cat#GENE-QC006
5 mL glass homogenizer	AZKKA	Cat#HZ3153
Axygen® 384-well Polypropylene PCR Microplate	Axygen	Cat# PCR-384M2-C
Eppendorf BioPhotometer D30	Eppendorf	Cat#BioPhotometer D30
Thermo Scientific Freso 21	Thermo Fisher Scientific	Cat#75002476

STEP-BY-STEP METHOD DETAILS

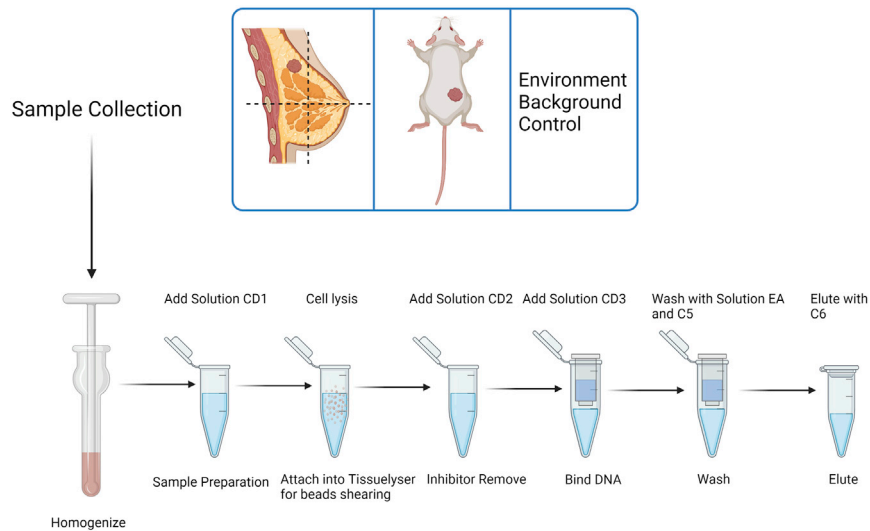### Tissue homogenization

⏱ Timing: 5–10 min per sample

Before DNA extraction, we need to prepare the environmental background control and tissue homogenates, which is critical for DNA extraction (Figure 3).

1. Set up environmental background control (EBC) by adding 200 μL PBS to glass homogenizer. After douncing, transfer PBS to EP tube.

   Note: As contaminants in the reagents are not easy to eliminate, it is essential to set up proper negative controls. We set up an environmental background control (EBC), using PBS as a tissue surrogate to undergo the same procedures as tissue samples. This control sample will help to determine the contamination landscape.

2. Preparation of tissue homogenates.
   a. Cut a small piece of tissue and weigh it in the 1.5 mL EP tube.
      i. Use a dissection scissor to mince the tissue into smaller pieces.
      ii. Add 1 mL sterile ice-cold PBS into the tube.

   Note: The optimal tissue size is about 200 mg. Overweight tissue (>200 mg) would significantly decrease the efficiency of genome extraction due to the limited binding capacity of the spin column.

Figure 3. Workflow of DNA extraction

b. Transfer the tissue with PBS using wide-orifice tips to the glass homogenizer.

Note: The glass homogenizer needs to be pre-chilled on ice, and the homogenization should be performed on ice. The wide-orifice tips can be prepared by cutting the tip with scissors.

c. Homogenize the tissue until the tissue are fully disassembled.

Note: Usually more than 10 dounces and around 1–3 min. The douncing times depend on tissue type, grinding force and grinding frequency. For example, human tissue is grinded harder than mouse tissue. Normal breast tissue is grinded harder than tumor tissue. Homogenize the tissue until the tissue are fully disassembled and become homogenate, and there is only some white connective tissue.

d. Transfer the tissue slurry back to the 1.5 mL EP tube.
e. Centrifuge the tissue homogenate with the table-top centrifuge at 4°C for 10 min.

Note: The centrifuge should be turned on in advance and pre-chilled to 4°C. The minimum speed is 16,000 g.

f. Discard the supernatant and save the pellet for further DNA extraction.

Note: For the PBS control samples, leave no more than 0.1 mL in the tube. At this step, tissue pellets can be stored at −80°C for future DNA extraction.

⚠ CRITICAL: Tumor dissection and processing should be strictly carried out in the clean and sterile biosafety cabinet with autoclaved dissection tools.

⚠ CRITICAL: All samples, reagents and homogenizer should be on ice all the time.

DNA extraction from tissue

⊙ Timing: 1–2 h

See Troubleshooting 1.

In this step, tissue-resident bacterial DNA is extracted following the QIAamp® PowerFecal® Pro DNA Kit Handbook (QIAamp PowerFecal Pro DNA Kit qiagen.com). The QIAamp PowerFecal (pro) procedure comprises four steps including lysis, binding, washing and elution. We optimized the procedure as indicated italicizing in tissue homogenization and below (Figure 3).

3. Sample lysis:
    a. Transfer the tissue and lysis buffer to PowerBead Pro Tube.
        i. Spin the PowerBead Pro Tube to ensure that the beads have settled at the bottom.
        ii. Transfer 200 mg of tissue pellet and 800 μL of Solution CD1 to PowerBead Pro Tube and vortex to mix.

    Note: If samples are not freshly made, thaw the tissue pellets from step 2c on ice from −80°C.

    b. Perform the beads shearing for the samples in the PowerBead Pro Tubes with a TissueLyser II for 10 min at 30 Hz.

    Note: Adapters of TissueLyser II should be pre-chilled to −20°C.

    Note: Shaking is critical for complete shearing and cell lysis. Microbes are lysed by a combination of chemical agents in Solution CD1 and mechanical forces from beads. Vigorous shaking causes the collision between microbial cells and the beads, facilitating the microbes to be broken apart, which is essential for Gram+ bacteria DNA extraction.

    c. Centrifuge the PowerBead Pro Tube at 15,000 × *g* for 1 min.
    d. Transfer the supernatant to a clean 2 mL Microcentrifuge Tube (kit provided).
4. Preparation of binding suspension.
    a. Add 200 μL of Solution CD2 to the supernatant and vortex for 5 s.

    Note: Solution CD2 contains Inhibitor Removal Technology (IRT), which can precipitate non-DNA organic and inorganic material including polysaccharides, cell debris and proteins. It is important to remove contaminating organic and inorganic substances that may reduce DNA purity and influence downstream DNA applications.

    b. Centrifuge at 15,000 × *g* for 1 min.
    c. Transfer up to 700 μL of supernatant to a clean 2 mL Microcentrifuge Tube (provided) and discard the pellet.

    Note: Expect 500–600 μL of the supernatant.

    Note: The pellet at this point contains non-DNA organic and inorganic material including polysaccharides, cell debris and proteins. For the best DNA yields and quality, avoid transferring any of the pellet.

    d. Add 600 μL of Solution CD3 to the supernatant and vortex for 5 s.
    e. Load 650 μL of the supernatant onto an MB Spin Column.
    f. Centrifuge at 15,000 × *g* for 1 min and discard the flow-through.
    g. Load the rest of the supernatant onto an MB Spin Column and centrifuge at 15,000 × *g* for 1 min. Discard the flow-through.
5. Preparation of washing.
    a. Carefully place the MB Spin Column into a clean 2 mL Collection Tube (provided).

    Note: Avoid splashing any flow-through onto the MB Spin Column.

  b. Add 500 μL of Solution EA to the MB Spin Column. Centrifuge at 15,000 × *g* for 1 min.

  c. Discard the flow-through and place the MB Spin Column back into the same 2 mL Collection Tube.

  d. Add 500 μL of Solution C5 to the MB Spin Column. Centrifuge at 15,000 × *g* for 1 min.

  e. Discard the flow-through and place the MB Spin Column into a new 2 mL Collection Tube (provided).

6. Preparation of elution.

  a. Centrifuge at 16,000 × *g* for 2 min. Carefully place the MB Spin Column into a new 1.5 mL Elution Tube (provided).

  b. Add 100 μL of Solution C6 to the center of the white filter membrane, incubate at around 25°C for 5 min.

  c. Centrifuge at 15,000 × *g* for 1 min. Discard the MB Spin Column. The DNA is now ready for downstream applications.

> *Note:* We recommend storing the DNA frozen (–20°C or –80°C) as Solution C6 does not contain EDTA.

> *Note:* The final DNA concentration from 200 mg tissue is usually around 400–600 ng/μL.

qPCR quantification

⊙ Timing: 3–4 h

To quantify the total bacteria, the primer 1369F and 1492R, which cross V9 region is regarded as best and universal primer set for bacterial SSU rDNA quantification and is chosen. 1369 Forward 5′-CGGTGAATACGTTCYCGG-3′, 1492 Reverse 5′-GGWTACCTTGTTACGACTT-3′, and Probe 5′-CTTGTACACACCGCCCGTC-3′ (5′ FAM and 3′ TAMRA) (Suzuki et al., 2000). Use the validated qPCR kit in 'before you begin' session. Change tips for each sample well and the experiment should be done in the RNA operation hood.

7. Preparation for qPCR assay. Troubleshooting 2.

  a. Irradiate the RNA operation hood including all tools and consumables to be used with UV light for 30 min.

  b. Thaw Premix Ex Taq on ice. Mix, and centrifuge each component before use.

> *Note:* Set up metal bath at 4°C in advance and put them on 4°C metal bath.

  c. Prepare the qPCR primer pool as following and mix gently:

qPCR primer pool

Primer pool	20× store concentration	Final concentration
16S-V9-F(1369F)	15 μL (100 μM)	750 nM
16S-V9-R(1492R)	10 μL (100 μM)	500 nM
16S-V9-Probe(1389F)	5 μL (100 μM)	250 nM
ddH$_2$O	70 μL	
Total reaction volume	100 μL	

> *Note:* The probe has fluorophores, so it's better to wrap the primer pool in aluminum foil to protect it from light.

8. Set up the following reaction of Premix Ex Taq (probe qPCR) mix and primer pool.

a. Calculate the number of samples including the standard curve, NTC, EBC and tissue samples.

Note: Each sample need at least 2 replicates. Given the pipetting error, the total number of samples should be more than required.

b. Combine the following reagents according to the sample size.
   i. Mix gently, but thoroughly and centrifuge the reaction mix.
   ii. Aliquot evenly into 8-strip PCR tube (Nest# 404001).

qPCR reaction

Reagent	Amount (µL) (1×)	Amount (µL) (110×)
Premix Ex Taq	10	
Primer pool	1	
ddH$_2$O	7	
Total reaction volume	18	

c. Use multichannel pipette to dispense 18 µL reaction mix accurately into the wells. For instance, the 96-well plate can be filled as follows:

	1	2	3	4	5	6	7	8	9	10	11	12
A	Sample1	Sample1	Sample1	Sample2	Sample2	Sample2	Sample3	Sample3	Sample3	Sample4	Sample4	Sample4
B	Sample5	Sample5	Sample6	Sample7	Sample7	Sample7	Sample8	Sample8	Sample8	Sample9	Sample9	Sample9
C	Sample	Sample	Sample	Sample	Sample	Sample	Sample	Sample	Sample	Sample	Sample	Sample
D	Sample	Sample	Sample	Sample	Sample	Sample	Sample	Sample	Sample	Sample	Sample	Sample
E	Sample	Sample	Sample	Sample	Sample	Sample	Sample	Sample	Sample	Sample	Sample	Sample
F	Sample	Sample	Sample	Sample	Sample	Sample	Sample	Sample	Sample	Sample	Sample	Sample
G	Ecoli8	Ecoli8	Ecoli8	Ecoli7	Ecoli7	Ecoli7	Ecoli6	Ecoli6	Ecoli6	Ecoli5	Ecoli5	Ecoli5
H	Ecoli4	Ecoli4	Ecoli4	Ecoli3	Ecoli3	Ecoli3	EBC	EBC	EBC	NTC	NTC	NTC

Note: The 96-well plate contains the sample in triplicate. You can choose 8-strip qPCR tube (NEST#403012), 96 well-plate (Bio-Rad# HSP9655), or 384 plate (Axygen#PCR-384M2-C) for various numbers of samples.

⚠ CRITICAL: It is important to use fresh ddH$_2$O aliquot (Invitrogen# 10977015) every time.

d. Add 2 µL of sample DNA, standards of *E. coli* and ddH$_2$O as NTC, EBC to each corresponding qPCR plate well.

Reagent	Amount (µL)
Template DNA	2

Note: It is important to pipet accurately for qPCR accuracy. Be carefully to pipet and check the volume during all PCR set up.

⚠ CRITICAL: Due to the low biomass of tissue-resident bacteria, it is critical to set up controls including NTC and EBC and use filter tips for qPCR to minimize contamination.

e. Seal the plate (Axygen# PCR-384M2-C) with plate sealing film (Bio-Rad #MSB1001), mix by gentle vortex, and spin down.

Note: It's better to choose transparent or semi-transparent plate. After spin down, double check the PCR volume, mark the suspicious wells.

⚠ CRITICAL: Seal carefully to avoid cross-contamination and smudging the surface of the film.

9. Quantify by qPCR.

Load the plate to the qPCR machine (Jena# Qtower3G or qTOWEP384/G) and run the following program:

qPCR program			
Steps	Temperature	Time	Cycles
Initial Denaturation	95°C	2 min	1
Denaturation	95°C	15 s	40 cycles
Annealing	60°C	60 s	
Hold	4°C	forever	

10. Analyze the qPCR data.
    a. Check the NTC, EBC wells for any amplification. Check the samples for good amplification and remove any low-quality replicates.
    b. Generate the standard curve from *E. coli* genomic DNA by plotting the Ct values over the log (*E. coli* CFU) (Figure 2).
    c. Analyze the PCR efficiency and R^2 (Figure 2B). The acceptable efficiency of PCR is 90%–110% (equal to -3.58 to -3.1 of the slope) and the R^2 >0.9.

Note: The slope of the standard curve was calculated by linear regression analysis, which can represent PCR amplification efficiencies ($E = (10^{-1/slope}) - 1$). A reaction with 100% efficiency will generate a slope of -3.32. PCR efficiency of 95%–105% is recommended. R^2 of linear regression, which represents the variation of the data, should be more than 0.99. No Template Control (NTC) represents bacterial DNA contamination of qPCR reagents. The Ct value of NTC should be higher than standards (Figure 2A).

    d. Calculate the equivalent bacteria of tissue samples according to a bacterial standard curve produced with *E. coli* DNA.

Note: The abundance of bacteria is calculated by the following equation:

$A_{sample} = Q_{sample}/V_{sample} * V_{Total}/W_{sample}$.
A_{sample}: equivalent bacteria quantity per gram tissue (CFU/gram).
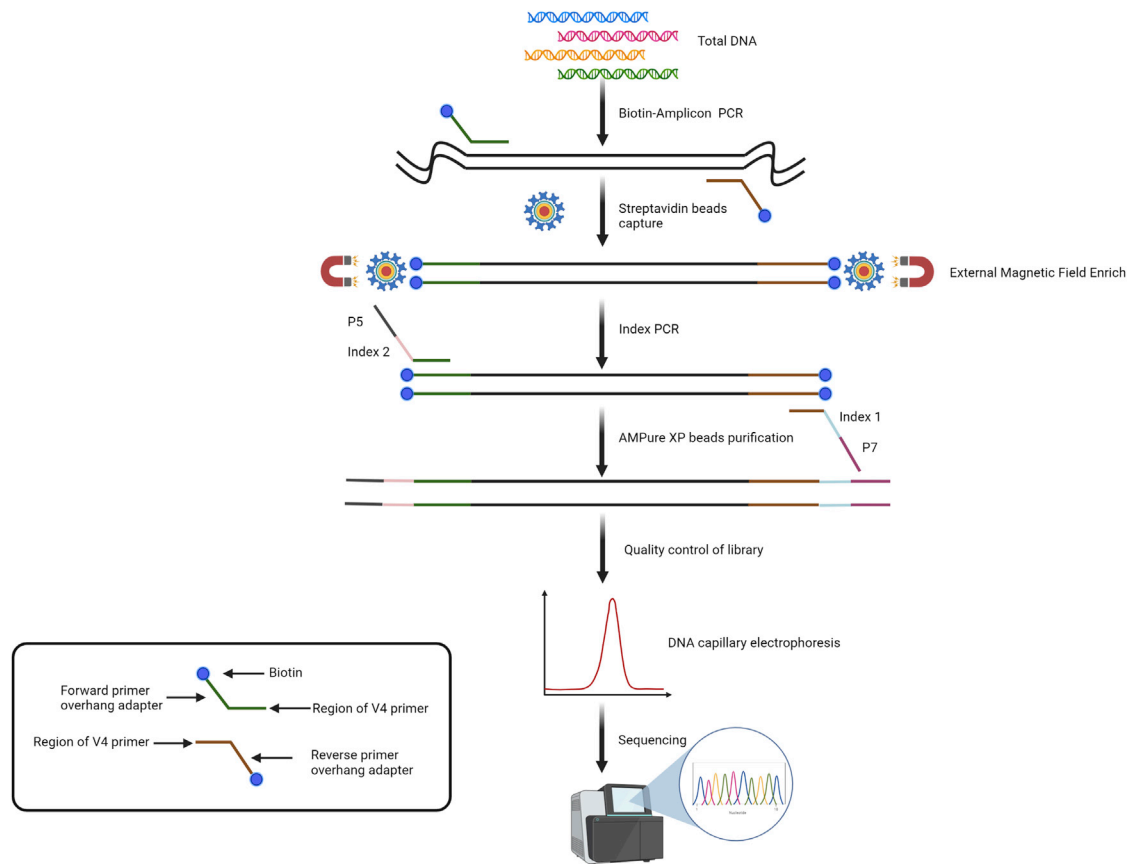Q_{sample}: equivalent bacteria quantity according to standard curve (CFU).
V_{sample}: Volume of genome sample used for qPCR (μL).
V_{total}: Volume of total sample elution (μL).
W_{sample}: Weight of sample for DNA extraction (g).

16S library preparation and purification

⏱ Timing: 10–12 h

Figure 4. Workflow of 16S library preparation

See Troubleshooting 3.

To profile the bacteria community, in this step, we refer to the paper of Caporaso et al. and choose the hypervariable region 4 (V4 region) to prepare the library. The PCR primers F515/R806 developed against the V4 region of the 16S rRNA gene, are determined to yield most diverse bacteria (Caporaso et al., 2011). The sequencing libraries of the extracted bacterial DNA from tissue are constructed according to the Illumina 16S metagenomics protocol (Part #15042322). The procedure is optimized by adding a biotin enrichment step due to the low biomass of tissue resident bacteria (Figure 4). To avoid contamination, prepare the sequencing libraries in an RNA hood and change tips between samples.

11. First-round amplicon PCR (2 h).
    a. Irradiate the clean batch including all tools and consumables with UV light to be used for 30 min.
    b. Thaw 2× KAPA HiFi HotStart ReadyMix on ice. Mix, and centrifuge each component before use. Put them on 4°C metal bath.

    Note: Set up metal bath at 4°C in advance.

    c. Calculate the amount of reagents according to the sample size (including the NTC and EBC controls) and set up the following reaction of 2× KAPA HiFi HotStart ReadyMix according to the table below, aliquot 14.5 μL of master mix into each well in a 96 well plate:

Amplicon PCR system

Reagent	Amount (μL) (1×)	Amount (μL) (110×)
2× KAPA HiFi HotStart ReadyMix	12.5	
Biotin-Forward primer (10 μM)	1	
Biotin-Reverse primer (10 μM)	1	
Total reaction volume	14.5	

Note: 2× KAPA HiFi HotStart ReadyMix should be set up on ice or at 4°C since the high proofreading activity of the enzyme will result in rapid primer degradation at around 25°C (room temperature).

Note: Use fresh PCR-grade water every time.

d. Add 10.5 μL of sample DNA into each well.

Reagent	Amount (μL) (1×)
DNA template	10.5

⚠ CRITICAL: It is vital to avoid DNA cross-contamination. We recommend using filter pipette tips for library construction to minimize contamination. Do remember to change tips for each well.

e. Seal the 96 well-plate (Thermo Fisher Scientific# AB0800) with plate sealing film (Bio-Rad #MSB1001), vortex gently, and spin down.
f. Load the plate in a thermocycler and run the program as outlined below:

Note: It's better to choose transparent or semi-transparent plate. After spinning down, check the PCR volume and mark the suspicious wells.

Amplicon PCR program

Steps	Temperature	Time	Cycles
Initial Denaturation	95°C	3 min	1
Denaturation	95°C	30 s	30 cycles
Annealing	55°C	30 s	
Extension	72°C	30 s	
Final extension	72°C	5 min	1
Hold	4°C	forever	

12. Biotin amplicon enrichment (1.5 h).
    a. Centrifuge the Amplicon PCR plate at 1,000 × g at around 25°C (room temperature) for 1 min, carefully remove the seal.
    b. Add ddH$_2$O to 50 μL for each well.
    c. Vortex the C1 beads (Invitrogen, #65001) for 30 s to ensure a homogenous slurry. Use 10 μL beads for each sample.

Note: Calculate the total amount of beads required for all the samples. Prepare more beads than needed considering the pipetting error.

d. Wash the C1 beads with 2×B&W buffer twice and resuspend beads with 50 μL 2 × B&W (per sample).

Note: Prepare B&W solution before library preparation and aliquots. Use fresh aliquot in each experiment.

2×B&W buffer		
Component	Volume	Final Conc
Tris-HCL (pH=7.5) 1 M	500 µL	10 mM
EDTA 0.5 M	200 µL	1 mM
NaCl (5 M)	20 mL	2 M
ddH$_2$O	29.3 mL	
Total volume	50 mL	

e. Aliquot 50 µL resuspended beads in B&W buffer to each well of amplicon PCR plate. (50 µL resuspended beads + 50 µL DNA).

f. Incubate on the rotator (MX-RL-Pro#8031402101) at around 25°C (room temperature) for 1 h.

g. After incubation, put each tube on a magnetic stand for 30 s, and discard the supernatant.

h. Keep each tube in the magnetic stand. Wash with 180 µL 1× B&W buffer once and ddH$_2$O twice, discard the supernatant.

i. Add 15 µL ddH$_2$O to each sample well using the multichannel pipette.

j. Take the plate off from the magnetic stand, gently resuspend beads by multichannel pipette.

Note: These beads samples can be used for second-round amplification index PCR.

13. Second-round amplification by index PCR (1 h).

This step attaches dual indices and Illumina sequencing adapters to the amplicon using the TruePrep® Index Kit V3 for Illumina.

a. Set up the following reaction master mix of Index 1 and 2 primers, 2× KAPA HiFi HotStartReadyMix according to the sample size.

   i. Aliquot 35 µL of master mix into each sample well in the 96-well plate containing 15 µL of sample beads.

   ii. Gently pipette up and down 10 times to mix.

⚠ CRITICAL: It is important to take Notes of the index information used for bioinformatic analysis.

Index PCR	
Reagent	Amount (µL)
2× KAPA HiFi HotStart ReadyMix	25
N6XX	5
N8XX	5
Total reaction volume	35

b. Seal the plate (Axygen# PCR-384M2-C) with plate sealing film (Bio-Rad #MSB1001), vortex gently, and spin down. Load the plate into a thermocycler and run the program as outlined below:

PCR cycling conditions

Steps	Temperature	Time	Cycles
Initial Denaturation	95°C	3 min	1
Denaturation	95°C	30 s	8 cycles
Annealing	55°C	30 s	
Extension	72°C	30 s	
Final extension	72°C	5 min	1
Hold	4°C	forever	

14. Amplicon Purification (1.5 h).
    a. Spin down and place the plate on a magnetic stand for 2 min until the supernatant is clear.
    b. Carefully transfer the supernatant into a new 96 well plate.
    c. Vortex the AMPure XP beads for 30 s to form a homogenous slurry.
    d. Aliquot 50 μL of resuspended AMPure XP beads (1×) to each well of the Index PCR plate.
    e. Gently pipette up and down 10 times to mix.
    f. Incubate at around 25°C (room temperature) without shaking for 5 min.
    g. Place the plate on a magnetic stand for 2 min until the supernatant is clear.
    h. With the Amplicon PCR plate on the magnetic stand, discard the supernatant using multi-channel pipet.
    i. With the Amplicon PCR plate on the magnetic stand, wash the beads twice with 200 μL of freshly prepared 80% ethanol each time in each sample well. Incubate the plate on the magnetic stand for 30 s. Carefully discard the supernatant.
    j. Spin down and use a multichannel pipette with fine pipette tips (20 μL) to remove excess ethanol.
    k. With the Amplicon PCR plate still on the magnetic stand, allow the beads to air-dry for 10 min.
    l. Remove the Amplicon PCR plate from the magnetic stand. Using a multichannel pipette, add 27.5 μL of ddH$_2$O to each well of the Amplicon PCR plate to elute the amplicon.
    m. Gently pipette up and down 10 times. Make sure that beads are fully resuspended.
    n. Incubate at around 25°C (room temperature) for 2 min.
    o. Place the plate on the magnetic stand for 2 min until the supernatant is clear.
    p. Carefully transfer 25 μL of the supernatant from the Amplicon PCR plate to a new 96-well PCR plate.

    ⅠⅠ **Pause point:** Library at this step can be sealed and stored at −20°C for up to a week.

15. Library quality control and sequencing (2 h).
Refer to the user guide of Fragment Analyzer 474 and the High Sensitivity NGS Fragment Analysis Kit (1 bp – 6,000 bp).
    a. Prepare Gel, Inlet buffer, Capillary Condition Solution.
    b. Prime the instrument according to the manufactory's instructions.
    c. Run 1 μL of the library on a Fragment Analyzer 474 to verify the size of amplicon.

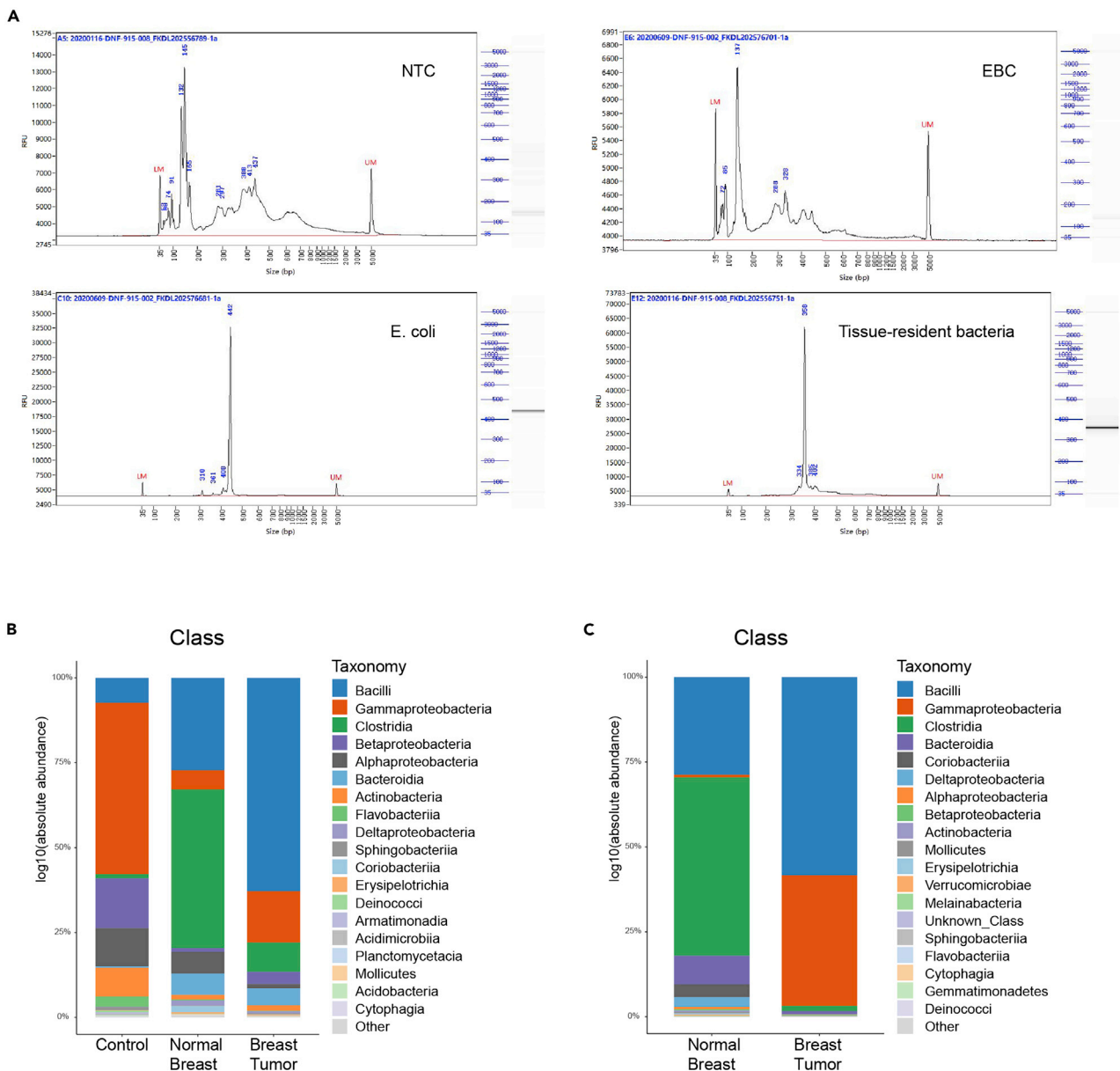   *Note:* The expected size of V4 amplicon of the final sample on the fragment analyzer is ∼300–450 bp (Figure 5A).

    d. Outsource the library to sequencing company for normalization, pooling and paired-end PE250 sequencing.

Microbiome data analysis

⏱ Timing: 2–3 days

See Troubleshooting 4.

A

B

C

Figure 5. Expected results of 16S library preparation and analysis

(A) The capillary electrophoresis results of the 16S library from NTC, EBC, *E. coli* and tissue samples for quality control. The X axis represents the size of library. The size of our bacterial 16S library is 300–450 bp as shown in the samples of *E. coli* and tissue-resident bacteria. While in NTC, EBC, the major peak is under 200 bp, which suggests most is primer dimer and the bacterial signal is relatively low. For y axis, the number represents the relative fluorescence units, which depends on the concentration of each peak.

(B) Stacked plot of relative abundance of bacteria at the class level in environment control, normal breast and PyMT breast tumor.

(C) Stacked plot of absolute abundance of bacteria at the class level in normal breast and PyMT breast tumor by correction of control qPCR data.

The BioProject accession number of the data used for this optimized protocol in NCBI is PRJNA681060. In addition, the website is https://dataview.ncbi.nlm.nih.gov/object/PRJNA681060?reviewer=7v2h87ups0iauqrpdmae8ue64j. The sequencing data were split using de-multiplexing tools bcl2fastq2 to generate fastq format files for every sample using barcode sequences. All reads were analyzed using a standardized metagenome bioinformatic pipeline using vsearch v2.14.2 (Rognes et al., 2016) and usearch v10 (Edgar, 2010) (Figure 6).
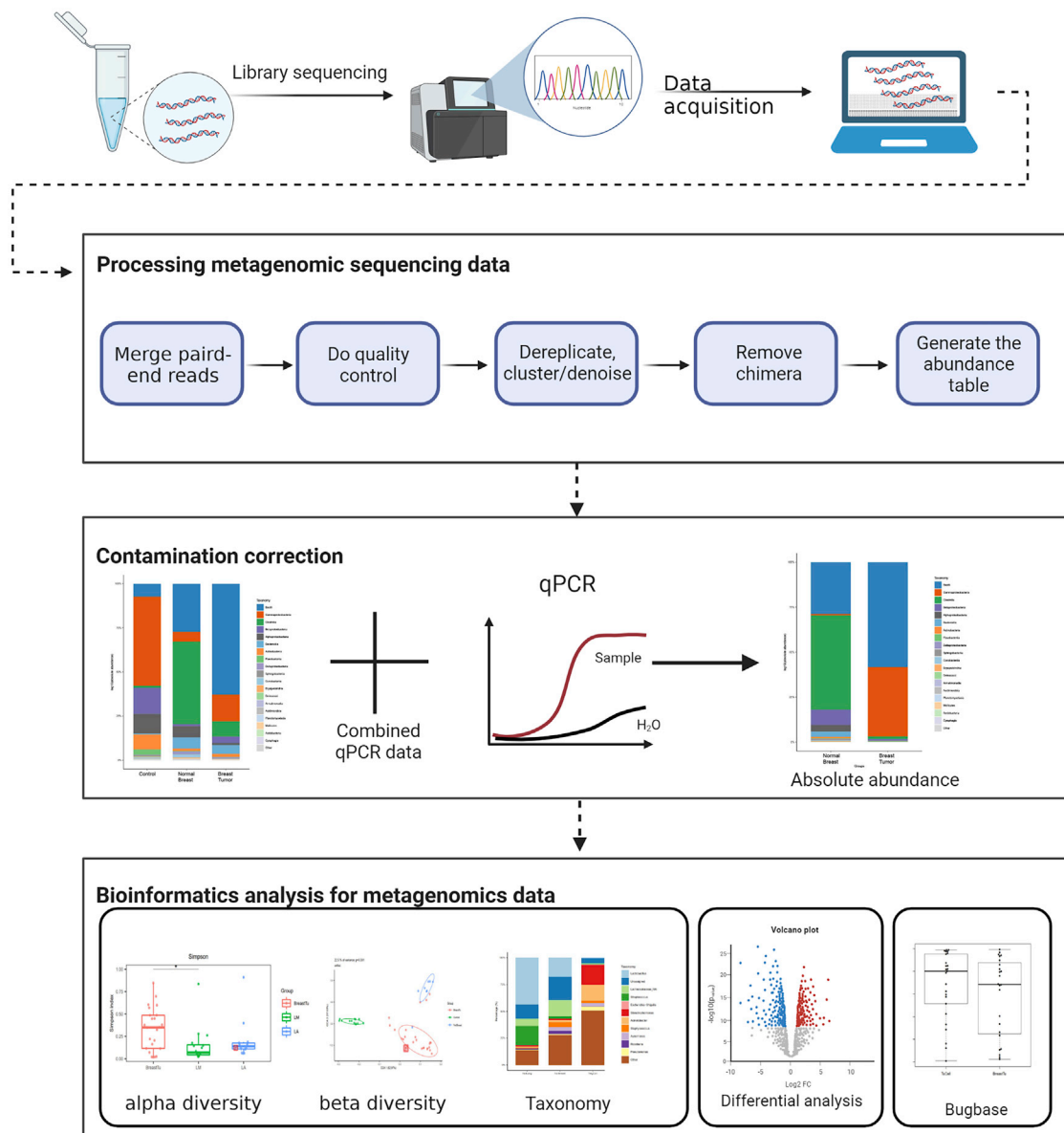
Figure 6. Workflow of 16S Bioinformatics analysis

Processing 16S amplicon sequencing data (also see Methods video S1).

16. Cleaning and clustering of sequencing reads.
    a. Merge paired-end reads and rename them by vsearch. Do the same for each sample.

```
> vsearch –fastq_mergepairs $sample_1.fq \

>       –reverse $sample_2.fq \

>       –fastqout $sample.merged.fq \

>       –relabel $sample
```

b. Do quality control: trim and remove adaptor and low-quality reads by "vsearch –fastx_filter" with maximum expected error rate 0.01.

```
> mkdir -p temp/ outdir/raw

> cat *merged.fq >temp/all.fq

> vsearch –fastx_filter temp/all.fq \

>       -fastq_stripleft 19 \

>       -fastq_stripright 20 \

>       -fastq_maxee_rate 0.01 \

>       -fastaout temp/filtered.fa
```

c. Dereplicate reads using vsearch and discard singleton reads.

```
> vsearch –derep_fulllength temp/filtered.fa \

>       -output temp/uniques.fa \

>       -relabel Uni \

>       -minuniquesize 2 \

>       -sizeout
```

d. Denoise all remaining unique sequences by unoise3 to obtain candidate sequence features.

Note: The ASV approach of unoise3 denoises the original data without setting a threshold, which is equivalent to 100% clustering.

```
> usearch –unoise3 temp/uniques.fa \

>       -zotus temp/zotus.fa

> sed 's/Zotu/ASV_/g' temp/zotus.fa > temp/otus.fa
```

e. Remove chimeric features through referring the SILVA Release 123 by using vsearch (Quast et al., 2013).

```
> vsearch –uchime_ref temp/otus.fa \

>       -db silva_16s_v123.fa \

>       -nonchimeras outdir/otus.fa
```

17. The abundance table generation.
    a. Align all reads with good quality control to the template sequence features by using "vsearch –usearch_global" with an identity threshold of 0.97.

```
> vsearch –usearch_global temp/filtered.fa \

>       -db outdir/otus.fa \

>       -otutabout outdir/raw/otutab.txt \

>       -id 0.97
```

   b. Classify sequence features by ''vsearch –sintax'' (Edgar, 2016) with cutoff set to 0.6.

   c. Filter non-specific sequence features in eukaryote, chloroplast and mitochondria.

Note: Rule out the samples with prokaryote specific reads less than 3,000 and 1,000 for mouse and human samples respectively.

```
> vsearch –sintax outdir/raw/otus.fa \

>       –db silva_16s_v123.fa \

>       –otutabout outdir/otus.sintax \

>        –sintax_cutoff 0.6

>sed -i -r -e 's/\t{4}/\t{3}/' outdir/otus.sintax

>Rscript script/filter_nonBac.r

>bash script/post_tax.sh
```

   d. Normalize all left samples into the same number of reads for a fair comparison by using the ''rrarefy'' function of R package ''vegan'' (Oksanen et al., 2019).

```
> Rscript script/otutab_rare.r \

>       –input outdir/otutab.txt \

>     –depth 3000

> Rscript script/out_count.r

> for i in p c o f g s;do

> usearch10 -sintax_summary outdir/otus.sintax \

>       -otutabin outdir/otutab_rare.txt \

>       -rank $i\

>       -output tax_rare/sum_${i}.txt

>sed -i 's/(//g;s/)//g;s/\"//g;s/\#//g;s/\/Chloroplast//g' tax_rare/sum_*.txt

> usearch10 –alpha_div outdir/otutab_rare.txt \

>       -output diversity/alpha/alpha_index.txt

> usearch10 –alpha_div_rare outdir/otutab_rare.txt \

>       -output alpha/alpha_rare.txt \

>       -method without_replacement

> usearch10 –cluster_agg outdir/otus.fa \

>       -treeout outdir/otus.tree

> usearch10 –beta_div outdir/otutab_rare.txt \

>       -tree outdir/otus.tree \

>       -filename_prefix diversity/beta/sampling_
```

Contamination correction

18. For mouse data, correct them with qPCR data (Figures 5B and 5C):

a. Obtain the relative abundance of each bacterial species from the metagenomics sequencing data.

b. Calculate the absolute amount of bacteria in each species of each sample according to the qPCR quantification and relative abundance.

c. Measure the contamination bacteria from background environment as the median absolute amount of a species among all negative control samples, formulated as below: $C_{ij}=c_{ij}*Q_i$
$C_j=Median(C_{ij})$.

Note: c_{ij} is the relative abundance of species j from sample i, i.e., the percentage of species j data among the data of all species in sample i.

Note: Q_i is the qPCR quantification result of sample i in unit of CFU/g.

Note: C_j is the overall contamination effect of species j, as the median relative abundance of species j among all environment samples.

d. The correction simply took off the overall contamination effect of species j from the measured species j in each sample: $A_{ij}= a_{ij}*Q_i-C_j$.

Note: a_{ij} is the relative abundance of species j from sample i; Q_i is quantification of total bacteria of sample i; A_{ij} is the corrected amount for species j of sample i.

```
> Rscript script/qPCR_adjust_script.r

> usearch10 -alpha_div qpcr_value/otu.txt \

>     -output alpha/alpha_index.txt

> usearch10 -beta_div qpcr_value/otu.txt \

>     -tree outdir/otus.tree \

>     -filename_prefix qpcr_value/beta/qPCR_
```

19. For human data, execute a binomial test:

Employ binomial test on each taxa's prevalence. x=the number of non-negative control samples with the taxa, n=the number of non-negative control samples, p=taxa's prevalence in negative control samples. Only keep taxa that pass p-value less than 0.05.

```
> Rscript script/binom_adjust.r

> usearch10 -alpha_div binom/otu.txt \

>     -output binom/alpha/alpha_index.txt

> usearch10 -beta_div binom/otu.txt \

>     -tree outdir/otus.tree \

>     -filename_prefix binom/beta/binom_
```

Bioinformatics analysis for 16S sequencing data

20. For differential analysis among samples, analyze contamination-corrected abundance data by using edgeR (Robinson et al., 2010). Obtain volcano plots from edgeR output.

Note: Taxonomies with FDR<0.25 are considered enriched or depleted.

21. Generate heatmaps of the contamination-corrected abundance data by R package ''pheatmap'' (Kolde, 2019) with manhattan distance.

    Note: Present the mean and standard error of each cluster's abundance as a bar plot.

22. Calculate Alpha diversity matrixes and correspondent rarefaction analysis by using usearch ''-alpha_div'' and ''-alpha_div_rare''.

    Note: Use the Wilcoxon test to do a statistical analysis of alpha diversity between two groups.

23. Calculate Beta diversity matrixes by using usearch ''-beta_div''.
    a. Perform Principal Coordinate Analysis (PCoA) and Constrained Principal Coordinate Analysis (CPCoA) by using the function ''beta_pcoa'' and ''beta_cpcoa_dis'' in the R package ''amplicon'' (Liu et al., 2021) with default parameter.
    b. Perform multiple Response Permutation Procedure (MRPP) test by R package ''vegan'' to compare the differences in the community composition between groups,
24. Regenerate the abundance table by "vsearch –usearch_global" (identity cutoff 0.97) and reference the database of Greengene release 13_8 (DeSantis et al., 2006) for microbiome phenotypes prediction.
25. Use Bugbase (Ward et al., 2017) for phenotype prediction.
26. Use the Wilcoxon test to do a statistical analysis of groups on given phenotypes.

    Note: Apply the abundance table to the contamination-corrected approach above mentioned.

27. To evaluate the taxonomy difference among sample groups, use a straightforward group mean approach on abundance tables to generate one abundance table for each sample group.
    a. Apply Beta diversity analysis to the group abundance table for the distance matrix between groups.
    b. Transform the distance matrix into the similarity matrix using the following reciprocal transformation: Sij=1/(1+Dij).

    Note: Sij is the similarity between the two groups and Dij is the beta diversity distance. The similarity matrix is used as input to the R package ''pheatmap'' for heatmap and clustering analysis.

28. For more microbial analysis, our previous study can be referred to (Fu et al., 2022).

EXPECTED OUTCOMES

qPCR was used to quantify the abundance of tissue-resident bacteria, as described in step 7 (Figure 2), we normally get 10^4 equivalent bacteria/gram for normal mouse and human breast tissue, and 10^5 equivalent bacteria/gram for mouse PyMT tumor and human breast tumors. The library construction of tissue samples yields 20–30 nmol/L (according to the qPCR quantification from the company) of DNA with a size distribution of 300–450 bp (Figure 5A) For more microbial analysis, our previous study can be referred to (Fu et al., 2022).

LIMITATIONS

With this optimized protocol, we can quantify and construct a 16S amplicon library for as low as 10^3 -10^4 equivalent bacteria per gram of tissue, which is essential for characterizing tissue-resident bacteria. However, there is still residual contamination that masks rare tissue bacteria. Further

improvements in quantification and library preparation protocols are required. For example, the improvement on decreasing the systemic bacterial contamination and the enrichment of tissue-resident bacteria genomic DNA are in need. In addition, we may design to probe more conserved regions to better target bacterial 16S genes (Nejman et al., 2020). Also, 16S amplicon sequencing of the v4 region often only allows characterization of the bacteria at the genus level, not down to the species level. This approach also does not measure gene expression which is necessary to determine how the bacteria are functioning within different tissue environments. These types of information may be revealed by other sequencing techniques like shotgun metagenomics together with meta transcriptomics.

TROUBLESHOOTING

Problem 1 (for DNA extraction)
The final DNA yields from tissues are low.

Potential solution

- The maximum tissue amount used for DNA extraction should not exceed 0.25 g. Too much tissue will significantly lower the lysis and extraction efficiency due to the limitation of DNA extraction column capacity.
- When dealing with fibrotic tissue that is very hard to grind, shred tissue to even smaller pieces in a 1.5 mL EP tube with scissors before grinding.
- During sample processing, after centrifugation of the tissue homogenates, remove excess liquid as much as possible. Keep samples on ice in homogenization and at 4°C during centrifugation. Pre-heat the C6 elution buffer at 42°C before adding into column, and extend the incubation time to 5 min.

Problem 2 (for qPCR)
Some of the following problems might be encountered during qPCR:

- The Ct value of tissue samples are higher than that in NTC and EBC control.
- PCR efficiency is too low (<90%).
- The standard curve of E. *coli* for quantification is not precise as before.

Potential solution
Some of the following solutions might be useful (corresponding to each bullet point of problem for qPCR):

- This might be caused by the contamination of bacterial genome in PCR system. Be sure to use fresh ddH$_2$O every time, do experiment in clean room and hood, and use filter tips. When encountering with this problem, double-check the reaction reagents for contaminations.
- DNA samples may contain PCR inhibitors that need to be eliminated or diluted; freeze and thaw cycles may deteriorate the reagents; the primer pool may have gone bad after repetitive uses. When encountering with this problem, purify the samples, or set up fresh reactions.
- Freeze and thaw cycles may deteriorate the standards. When encountering with this problem, set up a new standard curve with newly extracted E. *coli* DNA. Aliquot the standard DNA for future use.

Problem 3 (for 16S library construction)
Some of the following problems might be encountered during 16S library construction:

- Environmental background contamination is too high.
- The library concentration is too low.
- There are multiple peaks of the library on the fragment analyzer.

Potential solution

Some of the following solutions might be useful (corresponding to each bullet point of problem for 16S library construction):

- Make sure the room and hood are clean; make sure to use filter pipettes; the glass homogenizer must be cleaned by bleach and extensively washed by ddH$_2$O to remove the residual trace amount of DNA.
- Make sure the sample DNA concentration is within a reasonable range (usually 400–600 ng/μL). Make sure that positive control works. If the same problem still occurs, the phenomenon suggests the bacteria in the sample is below the detection limit.
- Purify the sample by size selection of AMPure XP beads.

Problem 4 (for data analysis)

For microbial analysis, the final prokaryote of each sample has limited reads. In certain cases, background contamination during library construction and sequencing becomes a considerable factor.

Potential solution

- Down-sampled each data from 20%-100%, and calculate alpha diversities to evaluate their saturation.
- Compare the constitution of control and treatment samples and conform that they are statistically different.
- Evaluate the background microbes by calculating the absolute quantities of microbiota in control sample using qPCR data. Subtract background microbial species from tissue samples.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Shang Cai (caishang@westlake.edu.cn).

Materials availability

This study did not generate new unique reagents.

Data and code availability

16S amplicon sequencing data have been deposited at SRA and are publicly available. The accession number for 16S amplicon reported in this paper is SRA: PRJNA681060.

This protocol does not report original code.

Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.xpro.2022.101765.

ACKNOWLEDGMENTS

AUTHOR CONTRIBUTIONS

Conceptualization, S.C. and B.Y.; methodology, S.C., B.Y., T.D., and A.F.; investigation, B.Y., T.D., A.F., and C.J.; formal analysis, H.L. and N.L.; writing – original draft, S.C., B.Y., T.D., and H.L.; supervision, S.C.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

Amplicon, P. 16S Metagenomic Sequencing Library Preparation. 2013 (Illumina).

Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Lozupone, C.A., Turnbaugh, P.J., Fierer, N., and Knight, R. (2011). Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proc. Natl. Acad. Sci. USA 108 (Suppl 1), 4516–4522.

Davis, N.M., Proctor, D.M., Holmes, S.P., Relman, D.A., and Callahan, B.J. (2018). Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. Microbiome 6, 226.

de Goffau, M.C., Lager, S., Salter, S.J., Wagner, J., Kronbichler, A., Charnock-Jones, D.S., Peacock, S.J., Smith, G.C.S., and Parkhill, J. (2018). Recognizing the reagent microbiome. Nat. Microbiol. 3, 851–853.

DeSantis, T.Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E.L., Keller, K., Huber, T., Dalevi, D., Hu, P., and Andersen, G.L. (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl. Environ. Microbiol. 72, 5069–5072.

Edgar, R.C. (2010). Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26, 2460–2461.

Edgar, R.C. (2016). SINTAX: a simple non-Bayesian taxonomy classifier for16S and ITS sequences. bioRxiv. https://doi.org/10.1101/074161.

Eisenhofer, R., Minich, J.J., Marotz, C., Cooper, A., Knight, R., and Weyrich, L.S. (2019). Contamination in low microbial biomass microbiome studies: issues and recommendations. Trends Microbiol. 27, 105–117.

Fu, A., Yao, B., Dong, T., Chen, Y., Yao, J., Liu, Y., Li, H., Bai, H., Liu, X., Zhang, Y., et al. (2022). Tumor-resident intracellular microbiota promotes metastatic colonization in breast cancer. Cell 185, 1356–1372.e26.

Ibekwe, A.M., Watt, P.M., Grieve, C.M., Sharma, V.K., and Lyons, S.R. (2002). Multiplex fluorogenic real-time PCR for detection and quantification of Escherichia coli O157:H7 in dairy wastewater wetlands. Appl. Environ. Microbiol. 68, 4853–4862.

Jervis-Bardy, J., Leong, L.E.X., Marri, S., Smith, R.J., Choo, J.M., Smith-Vaughan, H.C., Nosworthy, E., Morris, P.S., O'Leary, S., Rogers, G.B., and Marsh, R.L. (2015). Deriving accurate microbiota profiles from human samples with low bacterial content through post-sequencing processing of Illumina MiSeq data. Microbiome 3, 19.

Kim, D., Hofstaedter, C.E., Zhao, C., Mattei, L., Tanes, C., Clarke, E., Lauder, A., Sherrill-Mix, S., Chehoud, C., Kelsen, J., et al. (2017). Optimizing methods and dodging pitfalls in microbiome research. Microbiome 5, 52.

Kolde, R. (2019). pheatmap: Pretty Heatmaps. R package version 1.0.12. https://CRAN.R project.org/package=pheatmap.

Laurence, M., Hatzis, C., and Brash, D.E. (2014). Common contaminants in next-generation sequencing that hinder discovery of low-abundance microbes. PLoS One 9, e97876.

Liu, Y.X., Qin, Y., Chen, T., Lu, M., Qian, X., Guo, X., and Bai, Y. (2021). A practical guide to amplicon and metagenomic analysis of microbiome data. Protein Cell 12, 315–330.

Nejman, D., Livyatan, I., Fuks, G., Gavert, N., Zwang, Y., Geller, L.T., Rotter-Maskowitz, A., Weiser, R., Mallel, G., Gigi, E., et al. (2020). The human tumor microbiome is composed of tumor type-specific intracellular bacteria. Science 368, 973–980.

n.d. QIAGEN. (n.d.) QIAamp PowerFecal Pro DNA Kits (QIAGEN).

Oksanen, J., Guillaume Blanchet, F., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., et al. (2019). vegan: Community Ecology Package. R package version 2.5-6. https://CRAN.R project.org/package=vegan.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., and Glöckner, F.O. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 41, D590–D596.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26, 139–140.

Rognes, T., Flouri, T., Nichols, B., Quince, C., and Mahé, F. (2016). VSEARCH: a versatile open source tool for metagenomics. PeerJ 4, e2584.

Salter, S.J., Cox, M.J., Turek, E.M., Calus, S.T., Cookson, W.O., Moffatt, M.F., Turner, P., Parkhill, J., Loman, N.J., and Walker, A.W. (2014). Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. BMC Biol. 12, 87.

Suzuki, M.T., Taylor, L.T., and DeLong, E.F. (2000). Quantitative analysis of small-subunit rRNA genes in mixed microbial populations via 5'-nuclease assays. Appl. Environ. Microbiol. 66, 4605–4614.

Ward, T., Larson, J., Meulemans, J., Hillmann, B., Lynch, J., Sidiropoulos, D., Spear, J., Caporaso, G., Blekhman, R., Knight, R., et al. (2017). BugBase predicts organism level microbiome phenotypes. bioRxiv, 1–19.