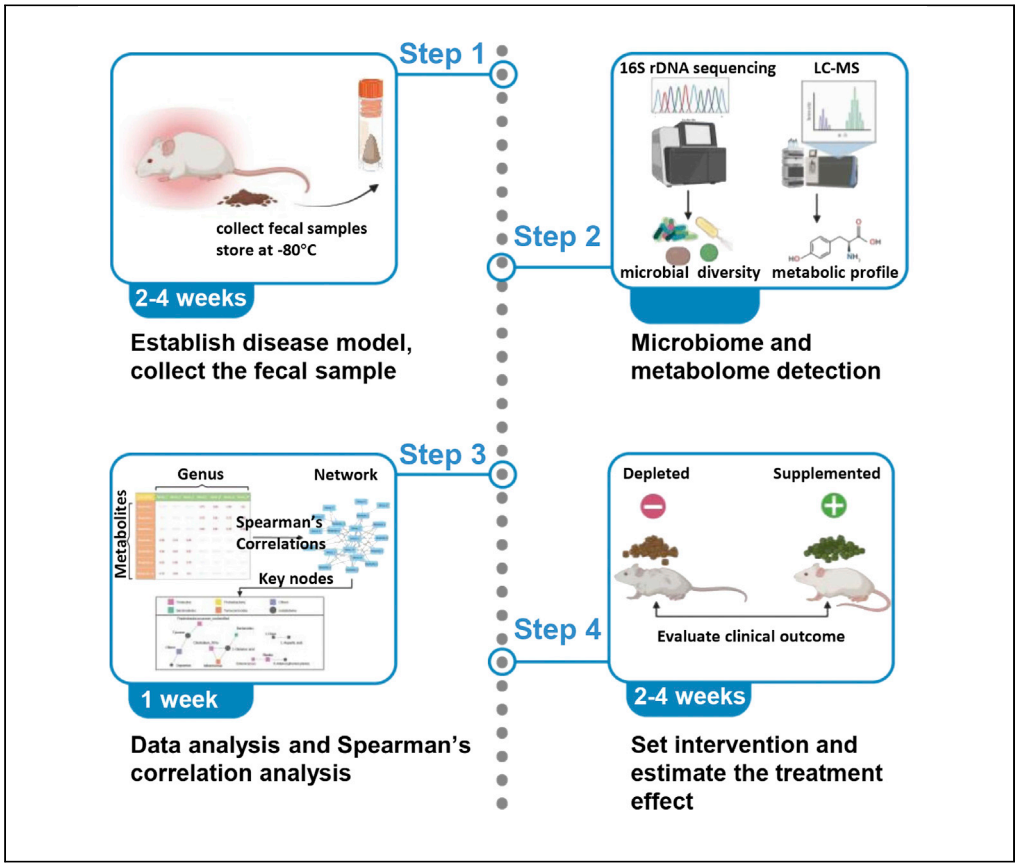
STAR Protocols

CellPress
OPEN ACCESS

Protocol

Protocol for correlation analysis of the murine gut microbiome and meta-metabolome using 16S rDNA sequencing and UPLC-MS



The gut microbiota and metabolites play pivotal roles in the pathobiology of various diseases. Here, we describe a protocol to profile the gut microbiome and meta-metabolome of a mouse disease model for acute graft-versus-host disease. We describe steps for fecal sample collection and processing for 16S sequencing and UPLC-MS. Finally, we detail the steps for data analysis and exhibit multi-omic associations to correlate with pathology.

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

Xiaoqing Li, Peng Wu, Xiangjun Zeng, Qiulei Lang, Yu Lin, He Huang, Pengxu Qian

linyu517@zju.edu.cn (Y.L.)
huanghe@zju.edu.cn (H.H.)
axu@zju.edu.cn (P.Q.)

Highlights

Fecal 16S rDNA gene sequencing and LC-MS decodes gut microenvironment during disease

Screening and correlation analysis between microbiome and metabolome

Joint analysis of multi-omics data and correlation with pathology

Li et al., STAR Protocols 3, 101494
September 16, 2022
https://doi.org/10.1016/j.xpro.2022.101494

STAR Protocols

CellPress
OPEN ACCESS

Protocol

Protocol for correlation analysis of the murine gut microbiome and meta-metabolome using 16S rDNA sequencing and UPLC-MS

Xiaoqing Li,[1,2,3,4,7] Peng Wu,[6,7] Xiangjun Zeng,[1,2,3,4] Qiulei Lang,[6] Yu Lin,[1,2,3,4,*] He Huang,[1,2,3,4,*] and Pengxu Qian[2,3,4,5,8,9,*]

[1]Bone Marrow Transplantation Center, the First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, Zhejiang, China

[2]Liangzhu Laboratory, Zhejiang University Medical Center, Hangzhou, China

[3]Institute of Hematology, Zhejiang University, Hangzhou, China

[4]Zhejiang Engineering Laboratory for Stem Cell and Immunotherapy, Hangzhou, Zhejiang, China

[5]Center of Stem Cell and Regenerative Medicine, and Bone Marrow Transplantation Center of the First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou 310058, China

[6]LC-BioTechnology Co., Ltd., Hangzhou, Zhejiang, China

[7]These authors contributed equally

[8]Technical contact

[9]Lead contact

*Correspondence: linyu517@zju.edu.cn (Y.L.), huanghe@zju.edu.cn (H.H.), axu@zju.edu.cn (P.Q.)
https://doi.org/10.1016/j.xpro.2022.101494

SUMMARY

The gut microbiota and metabolites play pivotal roles in the pathobiology of various diseases. Here, we describe a protocol to profile the gut microbiome and meta-metabolome of a mouse disease model for acute graft-versus-host disease. We describe steps for fecal sample collection and processing for 16S sequencing and UPLC-MS. Finally, we detail the steps for data analysis and exhibit multi-omic associations to correlate with pathology.

For complete details on the use and execution of this protocol, please refer to Li et al. (2020).

BEFORE YOU BEGIN

The protocol below describes the specific steps for correlation analysis of the gut microbiome and meta-metabolome using murine acute graft-versus-host disease (aGVHD). As our colleagues have successfully extended this protocol in an aging murine model and various murine tumor models, our approach could be generalized to other studies.

Multiple articles have described the procedure of characterizing the microbial community structure by using next-generation sequencing (NGS) in the STAR protocol (Reitmeier et al., 2020; Marti et al., 2021). Most studies have characterized microbiome diversity based on a single gene, the small subunit ribosomal RNA (rRNA) gene, which is encoded by highly conserved 16S ribosomal DNA (rDNA). Characterized by high expression, stability and conservatism, 16S rDNA is the most common method for estimation among different species of bacteria and archaea (Konstantinidis et al., 2006; Muhamad Rizal et al., 2020). Using Illumina high-throughput sequencing of 16S rDNA, we can fully describe bacterial diversity and community composition. As delays or failures in the identification of pathogens could occur via bacterial culture and biochemical testing in clinical diagnostic microbiology (Muhamad Rizal et al., 2020), NGS-based bacterial detection methods are emerging and becoming a mainstream technology in oncology. 16S rDNA sequencing provides an ideal

method for the identification of unculturable and fastidious bacteria achieving more rapid and predictable turn-around time with a streamlined identification protocol. In our study, the diversity and abundance of intestinal flora among different aGVHD murine groups were analyzed using 16S rDNA high-throughput sequencing. Based on our results obtained from 16S rDNA from six replicates per group, the data were large enough to provide sufficient statistical power to identify significant microbiota.

Interactions between the host and microbiota occur primarily through a multitude of metabolites. The development of metabolomic profiling offered by chromatographic separation (ex. liquid chromatography) coupled to mass spectrometry (MS) (LC–MS) introduced a new method for examining the dynamic multipara metric response to pathophysiological stimuli or genetic modification (Nicholson and Lindon, 2008; Gika et al., 2014). As the information complements the other two major omics genomics and proteomics, metabolomics identifies novel markers where changes face biological challenges. These variations function in unforeseen and/or unexpected ways with respect to their biochemical function. However, due to non-standardization of methods, instrumental drifts, MS detection instability, etc., untargeted LC–MS-based metabolic profiling still has difficulty achieving unbiased analysis, and it is difficult to compare datasets between different experiments or different laboratories (Gika et al., 2014; Seger and Salzmann, 2020; Kang et al., 2020). In such cases, we encourage researchers to focus on meticulous experimental design (ex. number of replicates, control selections, identification strategy and significance determination) and research (quality control and validation).

In our murine-based study, the control and test groups should ideally have the same genetic background and should be matched for age and sex, including being raised under the same feeding and environmental conditions. Due to the high sensitivity and specificity of LC–MS resulting in large errors in the identification or quantification of metabolites, we increased the number of replicates up to 6 samples/group and strictly controlled the sampling conditions and procedures. After comparing metabolomic profiling between the control and aGVHD groups, we obtained a large discrepancy between batches in the subsequent tyrosine validation experiments, which forced us to optimize the analytic strategies and only concentrate on the metabolites displaying significance in the former result.

Integrating analyses of the microbiome and meta-metabolome should facilitate understanding the role of the microbiota in human diseases. The complex microbial ecosystem inhabiting the intestinal tract produces a wide range of metabolites that may be absorbed and influence physiological processes (Raja et al., 2021). On the other hand, diet can change the intestinal microflora and consequently affect overall host health. In this study, we performed Spearman's correlation analysis to identify microbe–metabolite relationships in different omics layers. Our study first determined the most significantly changed factors, which was tyrosine, based on the microbiota and meta-metabolome data between the control and aGVHD disease groups. After discovering this correlation, specific interventions, including tyrosine-supplemented and deprived diets, were used as confirmatory tests to assess treatment efficiency. For better understanding, the representative data from the TCD-BM (control) and TCD-BM+T (aGVHD) groups on Day 14 are provided as an example (Li et al., 2020). This protocol aimed to provide the potential of combining microbiota and metabolite interventions as a means of achieving person specific, integrated and efficient therapy.

Institutional permissions (experimental animal ethics)

For animal experiments, the researchers are supposed to acquire approval form the Animal Experimentation Ethics Committee of the relevant institutions. In this study, all mice experiments were conducted under specific pathogen-free conditions in the Laboratory Animal Center of Zhejiang University.

Murine model establishment

Timing: 2 days

1. Prepare the model (aGVHD murine model).

Note: Specific pathogen-free (SPF) conditions are free from a selection of common pathogens to which the species are exposed in the wild. We achieve our experiment using SPF methods consisting of a controlled, health-monitored and more natural environment of indigenous gut flora. Although SPF gut microbiome heterogeneity may fail to translate into human trials, we still believe that SPF manipulation of the microbiome-host relationship has itself become a confounding variable in biomedical research (Dobson et al., 2019).

Note: To protect against infection, recipient mice (BALB/c) are often prescribed drinking water with antibiotics (penicillin and streptomycin antibiotic mix 100–200 μg/mL) for 1–2 weeks before BMT. However, administration of antibiotics affects bacterial communities and may lead to dysbiosis of the intestinal microbiota, and a standard process for clearance of gut bacteria has been proposed for single-bacteria research. In this study, we abandon antibiotic prophylaxis but increase the frequency of changing food and cages.

Note: Hematopoietic stem cells can reconstitute the entire hematopoietic system following lethal irradiation. Recipient mice (BALB/c) are treated with myeloablative conditioned with a dose of 7.5 cGy or split doses of 2×4.0 Gy for 4 h. The gut commensal microflora of SPF mice is not removed using total body irradiation (TBI).

⚠ CRITICAL: Variations could be observed between different irradiation instruments or even different runs on the same instruments, and therefore, the effect of lethal irradiation should be tested.

2. Bone marrow transplantation (BMT).

Note: Habituate the animals to the testing environment prior to experiments.

Note: We recommend the investigator who scores and groups the animals to be blinded to the experimental setting.

Note: All animal experiments must be performed in accordance with approved animal protocols and the guidelines and ethical approval. For the procedures described here, approval of the local Institutional Animal Care and Use Committee was received.

a. Flush bone marrow cells from the femurs and tibias of donor mice,
b. Remove $CD3^+$ T cells by magnetic bead sorting (CD3 Selection Kit, Miltenyi Biotec).
c. Purify splenic T cells by negative sorting using magnetic microbeads from the donor's spleen (Pan T-Cell Isolation Kit, Miltenyi Biotec).
d. Deliver 5×10^6 T-cell-depleted bone marrow (TCD-BM) cells with or without 5×10^5 splenic T cells to recipients on day 0 after lethal irradiation, while T cells are used as a source for aGVHD induction.
e. Assess the animals for weight loss, clinical condition and disease progression twice a week according to the criteria of the disease model.

Note: We suggest a group of investigators who are blind to the treatments for the assessment.

f. Use recording symptoms for assessing peak clinical disease severity so that the sampling time points can be determined.

Intervention setting

Timing: 2–4 weeks

3. Intervention determination.

Note: Mice transplanted with TCD-BM+ T cells exhibited the typical clinical manifestations of aGVHD, such as weight loss and high mortality rates, while mice receiving TCD-BM only (non-aGVHD) serve as the healthy control group.

a. Evaluate the changes in 16S rDNA sequencing and LC–MS-based meta-metabolomes between the two groups during early illness (day 14 after BMT) and during the critical stage (day 28 after BMT) to identify differences with the same variation trend at different stages of disease development.

Note: Such screening criteria aim to ensure the efficiency and sustainability of the intervention. For example, our results provide an increased abundance of bacteria previously associated with disease states, such as Enterococcus, in aGVHD mice (Stein-Thoeringer et al., 2019; Holler et al., 2014) but illustrate a dramatic and sustained decrease in many novel metabolites, especially tyrosine.

b. Select the pattern exhibiting the most significant change as the intervention, and determine dose limits or delivery way according to the previous studies or pharmacology.

Note: In our research, we select tyrosine as the intervention target and design tyrosine-supplemented and deprive diets as a confirmatory test (Figure 1).

4. Dietary intervention.

Note: Based on the omics results between the disease and control groups, tyrosine is the most significant item among the changed metabolites and is chosen for further experimental examination. According to a pilot study, a tyrosine content of 2%–2.5% is even considered the upper limit within the standard safety margin.

a. Set 2% tyrosine diet as the highest supplementary group and a diet deprived of tyrosine as a comparison group (normal standard 0.7% tyrosine), in order to amplify the differences (Messineo et al., 2018).

Note: The 0% or 2% tyrosine diets are administered (Hangzhou LiLeng Biotechnology Co., Ltd) one week before BMT.

b. Weigh feed every two to three days to record individual feed intake until the end of the experiment.

Note: The fecal samples collect at indicated time points (Days 14 and 28 after BMT) are subjected to microbiome and metabolomics analyses to assess the outcome of the intervention.

5. Microbiota supplementation.

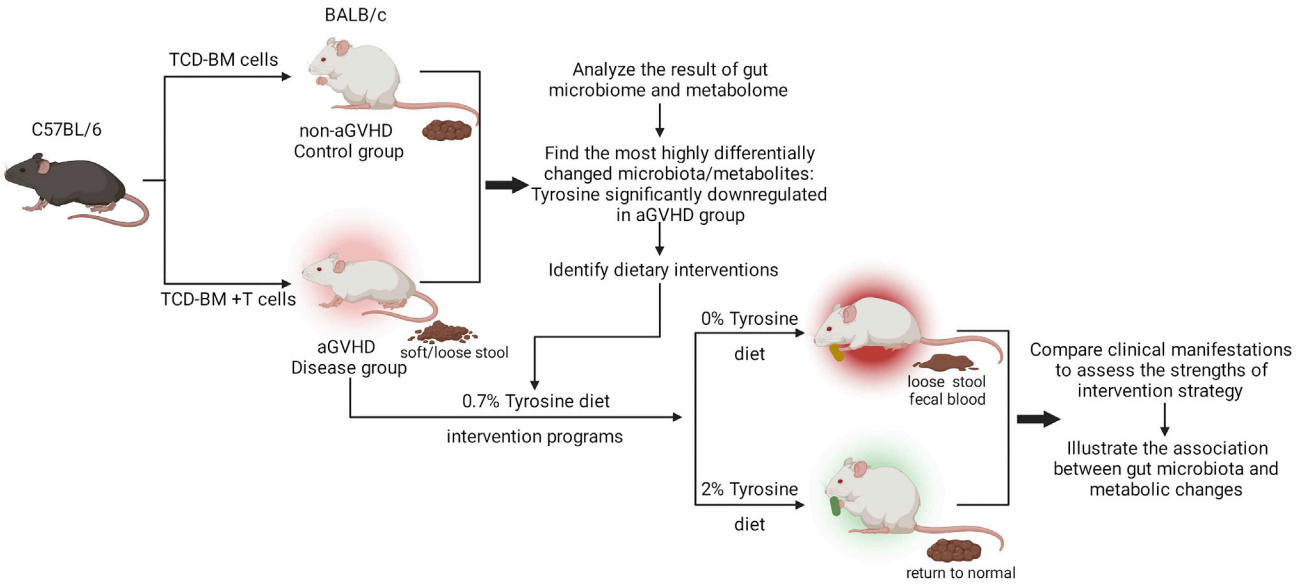


Figure 1. A schematic outline for identifying the intervention strategy

Analysis of the gut microbiota profiles and fecal metabolites from the preliminary experiment between the non aGVHD (TCD-BM cells only) and aGVHD (TCD-BM+T cells) groups, which aimed to highlight the differences and further select the intervention items. Either microbiota or metabolites could be potential modifiable factors in the intervention strategy that contribute to affecting mouse clinical outcomes (tyrosine is shown here as an example). Correlation analysis is performed on the microbiota and metabolites that changed significantly in response to treatment.

Note: After evaluation of the gut flora in a murine disease model, researchers directly affect the composition of the microflora by applying a suspension of bacterial strains.

a. Grow the selected strains in aerobic or anaerobic conditions in accordance with their specific requirements. All bacterial cultures are mixed with glycerol and added to a final concentration of 20%. Aliquots (1 mL) are individually frozen and stored at −80°C (Mathewson et al., 2016).
b. Administrate the selected strains via intragastric gavage every other day to naive mice beginning 14 days prior to allo-BMT with continued administration for 21 days post-BMT.

Note: For intervention dose, frequency and time interval from disease onset, we follow the protocol of Mathewson et al. (Mathewson et al., 2016). Microbial intervention can also be used in this protocol other than dietary intervention, while we use our dietary-intervened dataset as an example.

⚠ CRITICAL: We recommend all vehicle groups undergo high stringency settings, such as a controlled feed ratio and solvent control.

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Experimental models: Organisms/strains		
BALB/c (H-2Kd) Male, 8–12 weeks	Shanghai SLAC Laboratory Animal Co., Ltd	http://www.slaccas.com/
C57BL/6 (H-2Kb) Male, 6–8 weeks	Shanghai SLAC Laboratory Animal Co., Ltd	http://www.slaccas.com/

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Critical commercial assays		
Mouse CD3 cell isolation Kit	BioLegend	Cat.480031
Pan T-Cell Isolation Kit	Miltenyi Biotec	Cat.130-0950139
E.Z.N.A.® Stool DNA Kit (200 preps)	Omega	Cat#D4015-02
Phusion® Hot Start Flex 2× Master Mix	NEB	Cat#M0536L
AMPure XP beads	Beckman Coulter	Cat#A63881
PhiX Control Kit v3	Illumina	Cat#FC-110-3001
NovaSeq 6000 SP Reagent Kit v1.5 (500 cycles)	Illumina	Cat#20028402
Qubit dsDNA HS Assay Kits	Thermo Fisher Scientific	Cat#Q32851
Oligonucleotides		
Modified amplicon PCR 341F Primer: CCTACGGGNGGCWGCAG (The 5′ end of the primer is tagged with specific barcodes for sequencing)	(Klindworth et al., 2013)	https://academic.oup.com/nar/article/41/1/e1/1164457
Modified amplicon PCR 805R Primer: GACTACHVGGGTATCTAATCC (The 5′ end of the primer is tagged with specific barcodes for sequencing)	(Klindworth et al., 2013)	https://academic.oup.com/nar/article/41/1/e1/1164457
Software and algorithms		
FLASH	(Magoc and Salzberg, 2011)	http://ccb.jhu.edu/software/FLASH/
fqtrim (v 0.9.4)	(Pertea, 2018)	http://ccb.jhu.edu/software/fqtrim/
VSEARCH (v2.3.4)	(Rognes et al., 2016)	https://github.com/torognes/vsearch
QIIME (v 1.8.0)	(Caporaso et al., 2010)	https://qiime.org/
R (v 3.6.1)	R Core Team	https://www.r-project.org
Ribosomal Database Project (v 11.5)	(Cole et al., 2014)	http://rdp.cme.msu.edu/
MAFFT (v 7.310)	(Fischer et al., 2013)	https://mafft.cbrc.jp/
LEfSe analysis	(Segata et al., 2011)	http://huttenhower.sph.harvard.edu/lefse/
XCMS	(Forsberg et al., 2018)	https://www.nature.com/articles/nprot.2017.151
CAMERA	(Kuhl et al., 2012)	https://pubs.acs.org/doi/10.1021/ac202450g
MSConvert software (ProteoWizard)	(Chambers et al., 2012)	https://proteowizard.sourceforge.io/index.html
MetaX	(Wen et al., 2017)	https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-017-1579-y
MetaboAnalyst	(Li et al., 2020)	https://dev.metaboanalyst.ca/MetaboAnalyst/home.xhtml
Cytoscape	(Li et al., 2020)	https://cytoscape.org/
Chemicals, peptides, and recombinant proteins		
Normal control feed (0.7% Tyrosine)	LiLeng Biotechnology Co., Ltd	n/a
0% or 2% tyrosine feed	LiLeng Biotechnology Co., Ltd	n/a
Acetonitrile LC–MS grade	Merck	Cat#1000304000
Methanol LC–MS grade	Merck	Cat#1060354000
Formic acid	Thermo Fisher Scientific	Cat#A117-50
Deposited data		
Raw and analyzed data	This paper	SRA: PRJNA637751
Other		
Vanquish Flex UHPLC Systems	Thermo Fisher Scientific	Cat#IQLAAAGABHFAPUMBJC
ACQUITY UPLC HSS T3 column (2.1 mm ×100 mm, 1.8 μm)	Waters	Cat#186003539
TripleTOF 5600plus Mass Spectrometer	SCIEX	https://sciex.com/products/mass-spectrometers/qtof-systems/tripletof-systems/tripletof-5600-system

STEP-BY-STEP METHOD DETAILS

Part 1. Fecal sample collection and storage

Timing: 1–2 h/each time point

Table 1. Master mix for amplicon PCR

Reagent	Initial concentration	Volume (μL/sample)	Volume (μL/24 samples)
Modified Amplicon PCR 341F Primer	1 μM	2.5	70
Modified Amplicon PCR 805R Primer	1 μM	2.5	70
Phusion® Hot Start Flex 2× Master Mix	2×	12.5	350
Total	n/a	25 (including 7.5 μL of sample)	n/a

1. Collect fecal samples into fecal tubes or cryopreservation tubes with at least 150 mg/tube and then flash-freeze in liquid nitrogen for storage at −80°C.

   Note: If researchers plan to detect the fecal microbiota and metabolites at the same time point, we recommend that fecal samples should be collected and then split into two tubes. The previous literature by Okahashi N. et al. can be referred to for details (Okahashi et al., 2021).

   ⚠ CRITICAL: To analyze all changes in the intestinal environment, fresh feces are collected within 2 h or by stimulating defecation.

Part 2. PCR amplification of the 16S rRNA gene V3V4 region library preparation

PCR amplification

Timing: approximately 7–8 h for 24 samples

This section describes the PCR setup and cycling conditions for sequencing the bacterial 16S rRNA gene V3V4 region.

2. Extract genomic DNA from fecal samples using an Omega E.Z.N.A.® Stool DNA Kit following the manufacturer's instructions.
3. Prepare the Master Mix (Table 1) for amplicon PCR.

   Note: In order to enhance the amplification and sequencing efficiencies, we modified amplicon PCR primers, which are incorporated with the adapters and sample-specific barcode sequences.

   a. Transfer 17.5 μL of the prepared Master Mix into each well of a 96-well plate and add 7.5 μL of fecal DNA sample (on average, this contains 25 ng of DNA), mock DNA, or negative control per well. Pipette up and down several times to ensure proper mixing.
   b. Seal and spun the plate for 30 s to collect the liquid at the bottom of the wells.
   c. Place the 96-well plate in a Thermo cycler, and run the amplicon PCR program following the settings shown in Table 2.
4. Perform PCR product clean-up using AMPure XP Beads s= according to the manufacturer's instructions.
   a. Add 1 μL AMPure XP beads per 1 μL PCR product.

   Note: We found that a 1 × volume of AMPure XP beads to the PCR product is the correct ratio to remove primer dimers and unused primers while leaving the PCR product intact. Users of this protocol may need to perform pilot experiments to ensure that this ratio also works for their experiments.

5. Verify that the generated amplicons have the correct size by running the PCR products on a 2% agarose gel at 120 V.

Table 2. PCR cycling conditions

Steps	Temperature	Time	Cycles
Initial Denaturation	98°C	30 s	1
Denaturation	98°C	10 s	32
Annealing	54°C	30 s	
Extension	72°C	45 s	
Final extension	72°C	10 min	1
Hold	4°C	Forever	

Note: The expected amplicon size is ~630 bp.

16S rRNA gene amplicon sequencing

Timing: 3 days

This section describes the Illumina sequencing conditions. These methods are performed at LC-Bio Technology Co., Ltd. in Hangzhou, China (www.lc-bio.com).

6. Quantify the concentration of the PCR amplicons using Qubit dsDNA HS Assay Kits according to the manufacturer's instructions.

Note: Typical DNA concentrations range from 2 to 5 ng/μL.

7. Pool the amplicons equally such that approximately 25 ng of DNA per sample is loaded for sequencing.
8. Perform the sequencing with a NovaSeq 6000 System (NovaSeq 6000 SP Reagent Kit v1.5).

Note: The sequencing run should be carried out as 2×250 cycles utilizing 30% Phi-X Control V3.

Part 3. LC–MS-based untargeted metabolomics

Metabolite extraction

Timing: approximately 4–5 h for 24 samples

9. Remove the fecal sample and weigh 50 mg.
10. Transfer the sample to a 1.5 mL Eppendorf tube, add 500 μL of 50% (v/v) methanol/ultrapure water solution, and mix them thoroughly using an ultrasonic processor for 10 min at 4°C with 100 Hz.
11. Add 500 μL of acetonitrile to the sample solution and vortexed for 1 min to enhance the extraction efficiency.

Note: To reduce the complexity of the sample and remove proteins, store the sample at −20°C for 2 h and then centrifuge at 20,000 × g for 10 min at 4°C;

12. Collect the supernatant and freeze-dried at −50°C and 0.2 Pa. vacuum;
13. Dissolve the sample in 100 μL of acetonitrile, and pipet 10 μL of each sample solution to prepare a pooled quality control sample.

Run LC–MS

Timing: approximately 16 h for 24 samples

Table 3. LC gradient elution conditions

Time (min)	Flow rate (mL/min)	Solvent A (%)	Solvent B (%)
0–0.5 min	0.4	95	5
0.5–7 min;	0.4	95–0	5–100
7–8 min;	0.4	0	100
8–8.1 min	0.4	0–95	100–5
8.1–10 min	0.4	95	5

This section describes the process of chromatographic separation and LC–MS analysis. These methods are performed at LC-Bio Technology Co., Ltd. in Hangzhou, China (www.lc-bio.com).

Note: All chromatographic separations are performed using a Vanquish Flex UHPLC System with an ACQUITY UPLC HSS T3 column used for reversed-phase separation.

14. Inject 4 μL of each sample, set the flow rate is 0.4 mL/min, and the mobile phase consisted of solvent A (0.1% (v/v) formic acid/ultrapure water) and solvent B (0.1% (v/v) formic acid/acetonitrile).

Note: The gradient elution conditions are set as shown in Table 3.

15. Maintain the column at 35°C in the oven.
16. Detect metabolites eluted from the column by TripleTOF 5600 plus.
17. Operate Q-TOF in positive ion and negative ion modes separately.
    a. For the positive-ion mode, the ion spray floating voltage is set at 5 kV, and for the negative-ion mode, it is set at - 4.5 kV. The MS data are acquired in the IDA mode.
    b. The curtain gas pressure is set at 30 PSI, and the ion source gas1 and gas2 pressures are set at 60 PSI. The interface heater temperature is set at 650°C.
    c. The TOF mass range is 60–1200 Da. Survey scans are acquired every 150 ms, and as many as 12 signal ion scans are collected if the threshold of 100 counts/s is exceeded with a 1+ charge state. The entire collection cycle time is fixed at 0.56 s.
18. Sum four time bins for each scan at a pulse frequency of 11 kHz by monitoring the 40 GHz multi-channel TDC detector with four-anode/channel detection.

Note: Set the dynamic exclusion time of the scan to 4 s.

⚠ CRITICAL: During the collection process, the accuracy of the instrument is calibrated every 20 samples. Furthermore, a QC sample is analyzed every 10 samples to evaluate the stability of the LC–MS.

EXPECTED OUTCOMES

Sequence data processing and microbiota diversity

Paired-end reads are assigned to samples according to the barcodes, which are further truncated through the cutoff of barcodes and primer sequences and merged using FLASH. Quality filtering of the raw tags is performed under specific filtering conditions to obtain high-quality clean tags by fqtrim (v 0.9.4). Chimeric sequences are filtered using Vsearch software (v2.3.4), and operational taxonomic units (OTUs) are generated from sequences with at least 97% similarity. Taxonomic classification is conducted using the ribosomal database project (RDP) classifier (v11.5). Sequences are aligned using the MAFFT (v 7.310).

The normalization rarefying is achieved by QIIME (v 1.8.0), which makes each sample have the equal numbers of sequences for comparison. Recently, several methods have been developed to avoid making assumptions regarding the variation within a taxonomic group, which target the weakness of

Table 4. Statistics of the alpha diversity index between TCD-BM (control) and TCD-BM+T (aGVHD) groups

Gut microbiota diversity index	Control group		aGVHD group		
Day 14	Mean (SD)	Median (IQR)	Mean (SD)	Median (IQR)	p-value
Observed species	82.33 ± 5.95	82.00 (61.00–100.00)	109.67 ± 7.92	113.00 (76.00–135.00)	0.026
Chao1	130.83 ± 12.61	124.00 (90.25–172.30)	104.71 ± 24.28	93.00 (39.25–208.38)	0.3095
Shannon	1.83 ± 0.23	1.57 (1.35–2.83)	1.91 ± 0.26	2.11 (1.01–2.59)	0.999
Simpson	0.50 ± 0.18	0.44 (0.33–0.77)	0.57 ± 0.20	0.64 (0.32–0.77)	0.824
Good coverage	1.00 ± 0.00	1.00 (1.00–1.00)	1.00 ± 0.00	1.00 (1.00–1.00)	>0.9999

OTUs. Amplicon sequence variants (ASVs) attempt to model the error of the sequencer and to cluster reads so that their distribution within clusters is consistent with the error model (Caruso et al., 2019).

Alpha diversity

Alpha diversity represents the overall functional biodiversity of a location or within an ecosystem (Lozupone and Knight, 2008). Use Chao1 and Observed_species to estimate the total number of species in a community, while Goods_coverage, Shannon and Simpson consider both species richness and evenness. The above indices are analyzed using QIIME (v 1.8.0) and the sample data are made available in Table 4.

Beta diversity

Beta diversity indicates differences in species composition among different habitats and communities using QIIME software (v 1.8.0). Apply distance-based statistical tests (weighted and unweighted UniFrac distances clustered) to test the association of the cluster composition with environmental and biological factors. Unweighted UniFrac is a presence–absence distance and counts the fraction of branch lengths unique to either community, which exhibits efficient in detecting abundance change in rare lineages. Weighted UniFrac distance considers species abundance differences and weights that measure branch length with abundance information (Lozupone and Knight, 2005) (Figure 2 (Li et al., 2020)).

Metabolite annotation

The acquired LC–MS data pretreatment is performed using XCMS software. Raw data files are converted into mzXML format using MSConvert software (ProteoWizard) and then processed using the XCMS, CAMERA and metaX toolbox included in R software.

Utilize retention time (RT) combined with m/z to identify each ion. The intensity of every peak is recorded to generate a three-dimensional matrix, which contained randomly assigned peak indices (RT-m/z pairs), sample names (observations) and the intensity information of ions (variables). Identify metabolites at the MS2 level by matching their molecular formula through an in-house fragment

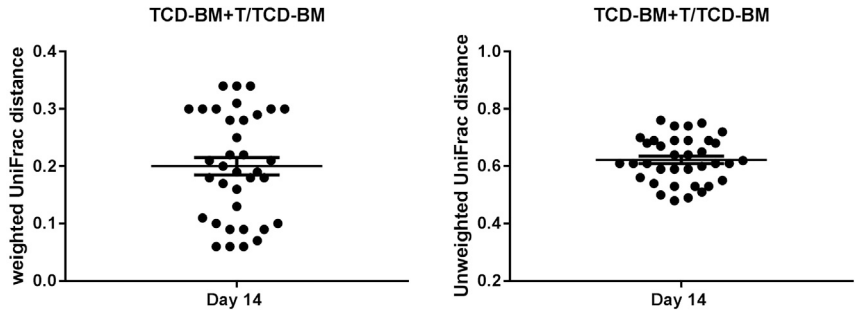


Figure 2. Weighted UniFrac and unweighted UniFrac analysis

Inter sample variability in community structure is determined by calculating the difference in beta diversity using unweighted and weighted UniFrac distance metrics.

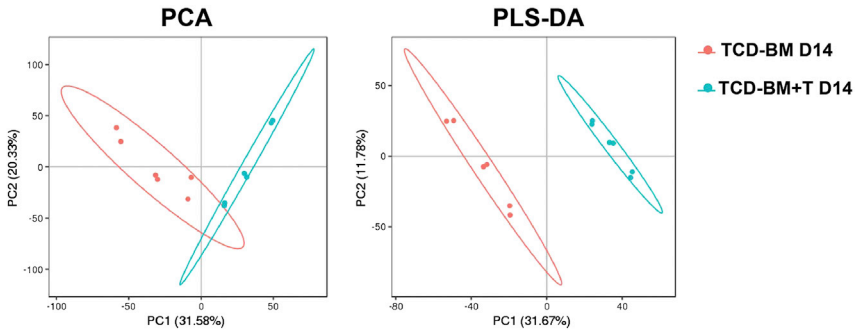


Figure 3. Score plots of the PCA and PLS-DA models

The PC1 and PC2 axes represent the first two principal coordinates. Clear discriminations between the non aGVHD (blue dots) and aGVHD (red dots falling in other quadrants) groups are observed in the PCA and PLS-DA score plots, with PC1 and PC2 explaining approximately 30% and 20% of the total variance, respectively.

spectrum library and public database. Adopt the Human Metabolome Database (HMDB) and online Kyoto Encyclopedia of Genes and Genomes (KEGG) to annotate metabolites through matching between the precise molecular mass data (m/z) of samples and those from database values below 10 ppm. Perform a QC-based signal-correction method, engaging nonlinear local polynomial regression (LOESS), to minimize the drift of signal intensity over time using the MetaX R package (Wen et al., 2017).

Note: The relative standard deviations (SDs) of metabolic features among all QC samples are calculated, among which those > 30% are excluded.

Use principal component analysis (PCA) to detect potential outliers and evaluate batch effects due to experimental covariates, while supervised partial least squares discriminant analysis (PLS-DA) is conducted on variables using a discriminant profiling statistical method to identify more specific differences between the groups (Figure 3).

QUANTIFICATION AND STATISTICAL ANALYSIS

As our study is comparing between two groups (healthy control vs. aGVHD group, tyrosine-deprived group vs. tyrosine-supplemented group), all analyses are based on a two-group comparison by an unpaired Student's t test (2-sided) is applied.

⚠ CRITICAL: Apply Mann–Whitney U test if the data did not meet the criteria of normality. For experiments include more than two groups, Wilcoxon-paired test corrected for multiple test comparisons using the Benjamini–Hochberg procedure.

All of the above analyses are conducted using the R statistical package (available as functions 't.test' and 'wilcox.test').

Differential species screening

1. Use linear discriminant analysis (LDA) effect size (LEfSe) analysis on the grouped samples using the nonparametric factorial Kruskal–Wallis (KW) sum-rank test to identify the significantly different genera between assigned taxa (Segata et al., 2011).

   Note: An LDA threshold of 3.5 is used for all biomarkers, and all tests of significance are two-sided, with p values less than 0.05 considered statistically significant.

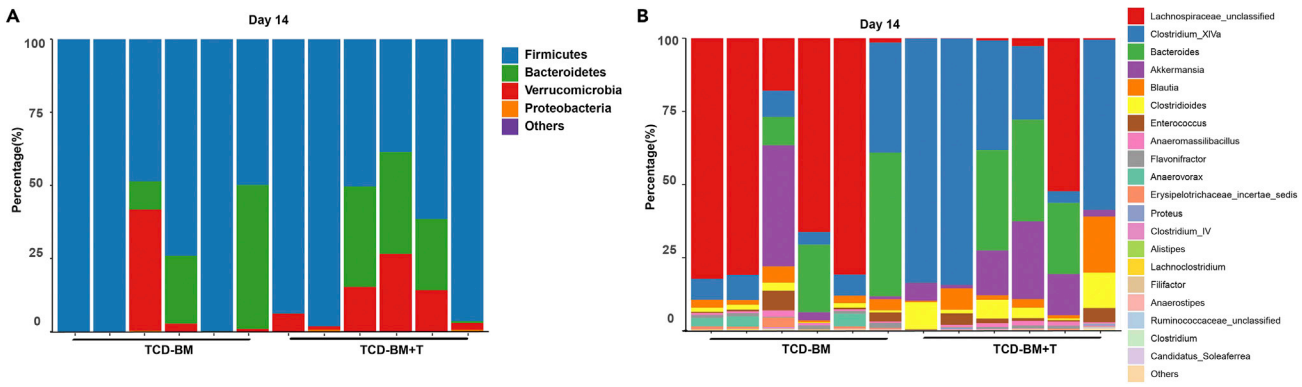


Figure 4. Relative abundance of gut microbiota at the phylum and genus levels

(A and B) 16S rDNA gene sequencing of fecal microbiota from TCD-BM (n=6) and TCD-BM+T-cell groups (n=6) and the relative abundance of gut microbiota at the phylum level (A) and genus level (B) on day 14 after transplantation. Reprinted with permission from (Li et al., 2020).

2. Differential species analysis can theoretically be performed at the level of domain, phylum, class, order, family, genus and species.
   a. In our study, downstream analysis is performed at the phylum and genus levels (Figures 4A and 4B (Li et al., 2020)). The abundance of top 20 different genus had took up over 99%, which are selected and organized the data into a structured matrix.
   b. Each row is the name of the species, and each column is the microbiota abundance and fold change value in each sample (Table 5).

Differential metabolite selection

3. Screen the metabolites meeting the requirements for the further study.
   a. Variants with ratio ≥ 2 or $\leq 1/2$,
   b. VIP (Multivariate statistical analysis with PLS-DA to obtain Variable Important for the Projection, VIP) ≥ 1,
   c. q value ≤ 0.05,
   d. Annotated at MS2 level.

   Note: The summary for different metabolites is collated and summarized in Table 6. Each row is the type of metabolite, and each column represents the metabolic relative components in a sample.

4. Develop a heatmap of the top 50 differentially expressed metabolites to organize and display the data clearly (Figure 5 (Li et al., 2020)).
5. Perform analysis of the relative enrichment of KEGG pathways using the web-based tool MetaboAnalyst 4.0 (Figures 6A and 6B), which contribute to understanding the biological functions of various metabolites.

Correlation analysis

6. Use Spearman's correlation coefficient to measure the strength of a linear relationship between two variables using the R package Hmisc, which ranges from -1 to 1. The sign reflects either a negative or a positive correlation, respectively.

   Note: Depending on the purpose of the experiment and the actual situation, the appropriate threshold can be determined to identify the significantly related bacterial-metabolite pair. In our study, only correlations with |Correlation coefficient | > 0.4 and p value < 0.01 are indicated and considered to be displayed. After calculating the correlation coefficient in each microbiota-metabolite, microbiota-microbiota and metabolite-metabolite pair (Table 7).

Table 5. Gut microbiota in healthy control (TCD-BM) and aGVHD (TCD-BM+T) groups

Genus	TCD-BM_D14 (Mean abundance)	TCD-BM+T_D14 (Mean abundance)	Fold change Day 14	p value of abundance
Lachnospiraceae_unclassified	54.957	9.4476	0.1719	0.0277
Bacteroides	13.6581	15.5655	1.1397	0.8624
Clostridium_XlVa	12.2968	48.7368	3.9634	0.0396
Akkermansia	7.5159	10.9121	1.4519	0.6767
Blautia	2.8294	5.4473	1.9252	0.4202
Enterococcus	1.9151	2.0063	1.0476	0.9462
Anaerovorax	1.771	0.0332	0.0188	0.086
Clostridioides	1.4102	5.5145	3.9105	0.0806
Flavonifractor	1.0631	0.3846	0.3618	0.0287
Erysipelotrichaceae_incertae_sedis	0.9833	0.1265	0.1287	0.1256
Anaeromassilibacillus	0.8568	0.9289	1.0841	0.8471
Clostridium_IV	0.2106	0.2111	1.0024	0.9943
Lachnoclostridium	0.1478	0.0708	0.479	0.058
Anaerostipes	0.1219	0.067	0.5498	0.1055
Ruminococcaceae_unclassified	0.0618	0.1008	1.632	0.3518
Delftia	0.0422	0	0	0.3436
Clostridium	0.0421	0.0217	0.5146	0.0945
Porphyromonadaceae_unclassified	0.0197	0.0248	1.2576	0.7447
Erysipelatoclostridium	0.0148	0	0	0.0723
.....				

Note: Gut microbiota accorded with significant change ($p<0.05$, highlighted in yellow) and abundance over 1%

7. Select an appropriate pattern of data visualization. Heatmap or network graphs can help us to visualize important relationships.

   Note: For the network graph generated by using Cytoscape, each node represents a metabolite (circular) or microbiota (square), and the node-to-node relationship (edges) represents the interaction between these biomolecules (Figures 7A and 7B (Li et al., 2020)).

8. Represent the parameters by different components in the network after the initial generation from the input data,

   Note: For example, the volume size refers to the strength of the correlation, and red and blue colors indicate positive and negative correlations, respectively (Figures 4C and 4D (Li et al., 2020)).

LIMITATIONS

One challenge is that fecal metabolomics cannot distinguish between endogenous and exogenous metabolites, leading us to rely on the reported literature or characterization of species, which is accomplished primarily using isolated, cultural, biochemical and molecular methods. However, there are many difficulties with experiments using pure cultures: most bacteria cannot be isolated in pure cultures, synthetic communities constructed from pure cultures might not represent any natural community, and there is no general standardization for knocking out specific taxa from natural communities (Rivett and Bell, 2018). Therefore, many studies of microbiota have been based on microflora with clarified properties.

TROUBLESHOOTING

Problem 1: Batch differences

We could generally separate our research into two parts: changes discovered and metabolite intervention. When we compared the latter result to the former data, we found that the outcomes are poor with respect to homogeneity and characterized differences. Large variations in the microbiome

Table 6. Different metabolic profile between TCD-BM (control) and TCD-BM+T (aGVHD) groups on day 14

		Control group	aGVHD group	Fold change	q Value	VIP
MS2superclass	Metabolites	Day 14 (mean value)	Day 14 (mean value)	Day 14	Day 14	Day 14
Phenylpropanoids and polyketides	Dehydroneotenone	100.61352	588.23156	5.8464462	0.0010913	2.6696723
	4-Hydroxycinnamic acid	9213.1937	781.732	0.0848492	0.0208994	3.3757817
	Caffeic acid	8433.7234	2545.4257	0.3018152	0.021606	2.1227442
Organoheterocyclic compounds	Riboflavin	56724.319	21366.578	0.376674	0.0014216	2.1782754
	Urocanic acid	16116.717	34136.885	2.1181041	0.0123058	1.6555307
	6-Hydroxynicotinic acid	3823.8213	654.00322	0.1710339	0.0141133	3.0237255
Organic oxygen compounds	Erythromycin	3751458.7	1854299.6	0.4942876	0.0039898	1.6508956
	Methacrolein	38248.643	2346.5117	0.0613489	0.0133858	3.4746616
Organic acids and derivatives	Tyrosine	3997.4834	1202.9267	0.300921	0.0182199	1.9353743
	L-Aspartic acid	10803.05	10391.73	0.9619256	0.8656363	1.0143129
	L-Glutamic acid	38549.17	30655.81	0.7952392	0.0970525	1.7928728
Lipids and lipid-like molecules	Floionolic acid	21504.28	7555.0427	0.3513274	7.20E-05	2.3758445
	Physalin P	9607.4026	2579.4691	0.2684877	0.0013269	2.2378511
	Anabsin	4722.1024	1496.2174	0.3168541	0.0013382	1.9491503
	Acylcarnitine 18:0	4008.0528	24523.682	6.1186024	0.0084062	3.3540335
	Glycocholic acid	1545.3678	3826.1	2.4758507	0.0084969	1.7927763
Benzenoids	Phenyl acetate	5304.0912	2548.4334	0.4804656	0.0054596	1.2035083
	2-Phenylethanaminium	1831.6042	12447.273	6.7958312	0.0176947	3.0204275
	Styrene	4818.0634	29328.697	6.0872377	0.0226978	2.8442328
	Tyramine	23230.42	24770.6	1.0663001	0.7209578	1.5157371
	Dopamine	5221.277	7967.126	1.5258961	0.0468903	1.2107321
Nucleosides, nucleotides, and analogs	S-Adenosylhomocysteine	2953.48	3625.85	1.2276534	0.1285364	0.7623467
.....						

identified between batches are apparent even at the family level, although we tried our best to control all of the variables, including mouse biological condition, feed ratio and sample collection.

Potential solution

Batch differences remain technically challenging in 16S rDNA sequencing and LC–MS. We recommend increasing the replicates in each group, when laboratory conditions and funding permit. If available, getting animals from the same parental colony would be better. Pre-screening of fecal microbiota community before any intervention could be helpful to minimize this variability and improve the consistency of results in subsequent experiments. For metabolite extraction, it requires a stable temperature, extraction time and space to ensure accuracy.

Problem 2: Dietary intervention falls short of expectation

Altering the content of a single metabolite in a diet might be not sufficient to reduce the risk of disease or to improve the manifestation, which leads to unsatisfactory microbiome and metabolomics results.

Potential solution

To be most effective, any supplementation plan should include a broad-spectrum metabolite combination as replenishment. For example, tyrosine is one of the predominant essential amino acids in the aromatic amino acid family, including phenylalanine and tryptophan. Therefore, a series of aromatic amino acids could be an appropriate choice for dietary intervention.

Problem 3: Selection of intestinal flora intervention

In some cases, an inappropriate increase in beneficial bacteria or the complete elimination of harmful bacteria might lead to counterproductive outcomes.

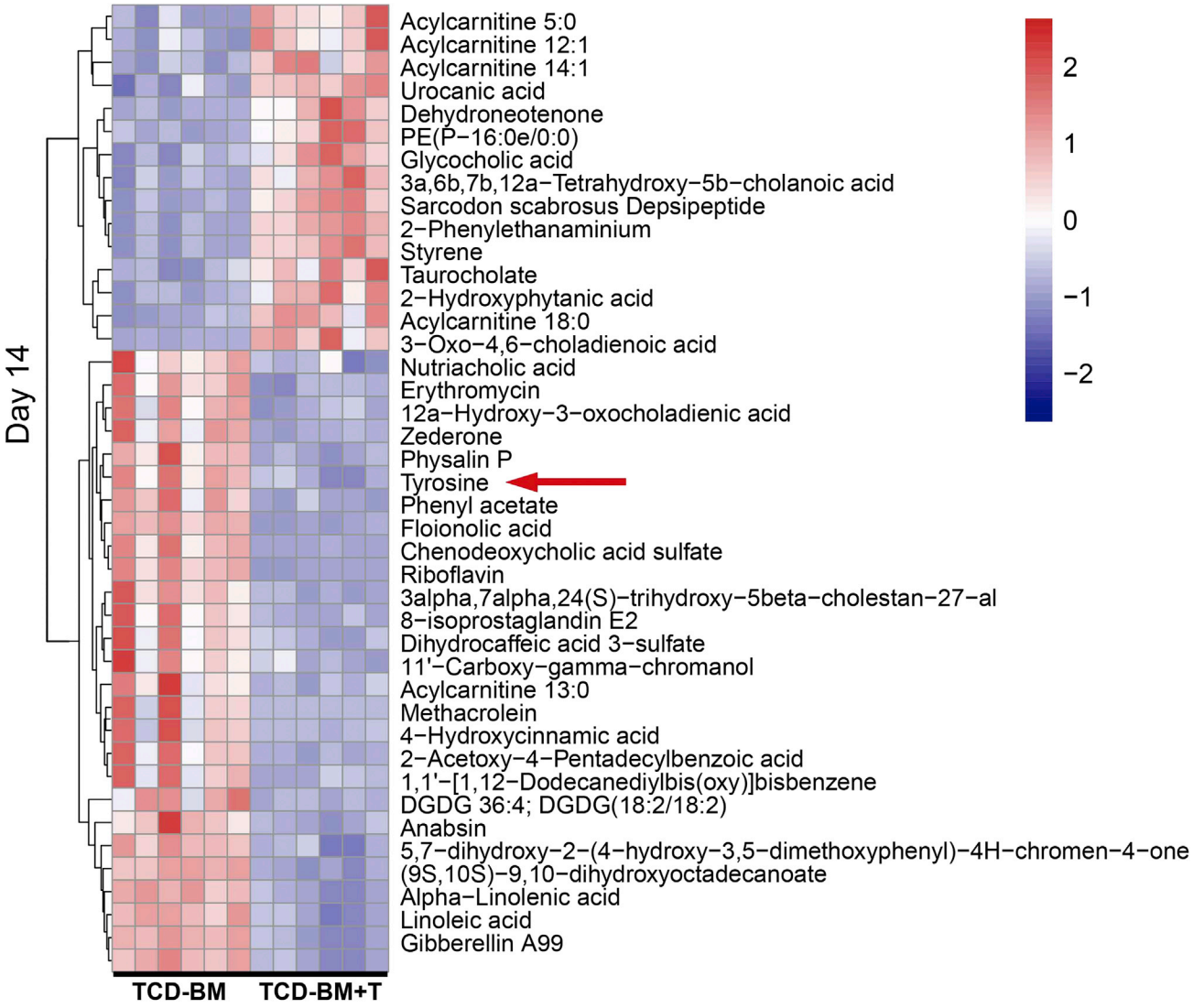


Figure 5. Changes in the metabolomics of the TCD-BM and TCD-BM+ T cell groups

The 50 most important metabolites after Student's t test and the hierarchical clustering of samples shown in the heatmap. Reprinted with permission from (Li et al., 2020).

Potential solution

The population of microorganisms in the intestine is a balanced phase. Hence, careful consideration should be given not only of the significance between groups but also their abundances. We recommend dose escalation experiments to determine the maximum tolerated dose and effective dose of the microbiota.

Problem 4: Identifying the metabolite under positive and negative ion scanning mode

Metabolites are acquired in both positive and negative ion modes, while some metabolites could only be identified in one mode. In addition, the molecular weight of the metabolite could be significantly different between positive and negative ion modes.

Potential solution

Basic metabolites are normally positively charged, and acidic metabolites are negatively charged. Identification of each metabolite is performed as the mass-to-charge ratio (m/z) of adducts. If a

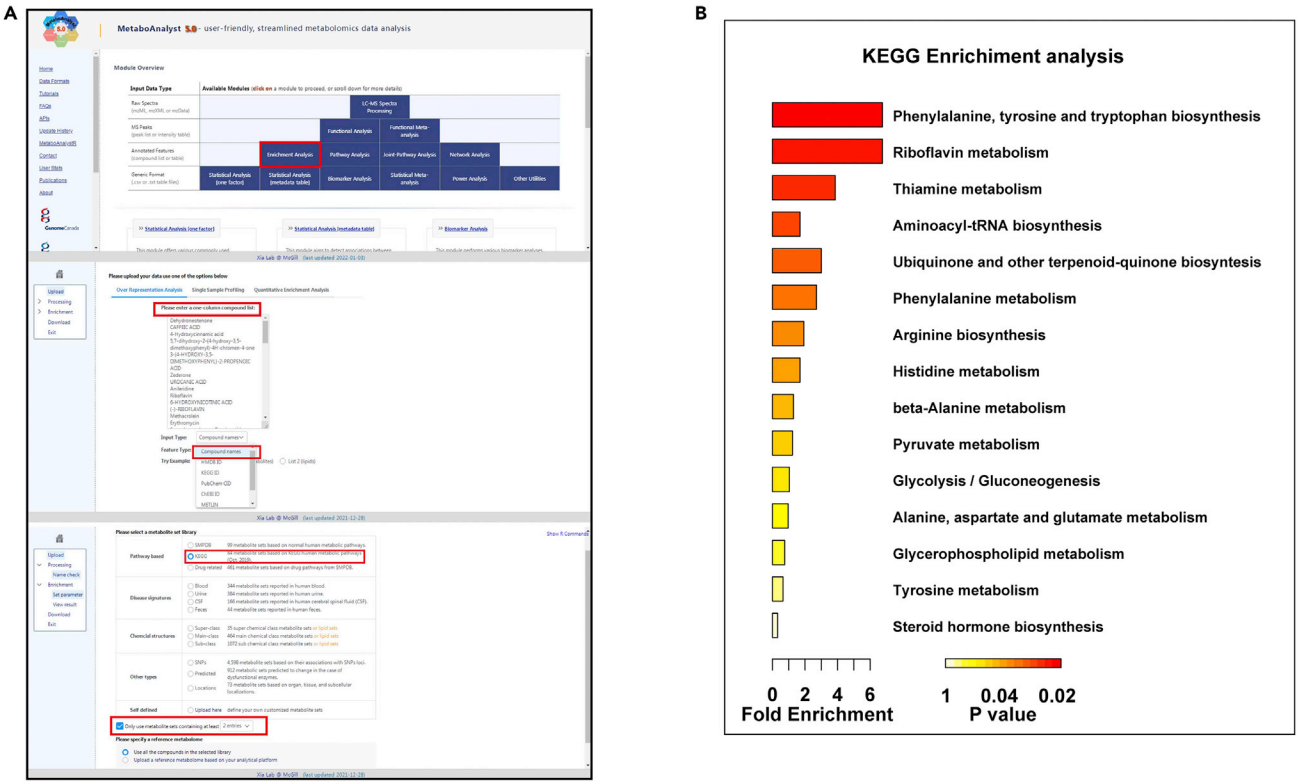


Figure 6. Enrichment analysis of selected metabolites

(A) Submitted selected metabolite data (MS2 level name) as compound name (input type) to the MetaboAnalyst Enrichment Analysis module. Pathway analysis based on the KEGG database as the library. We recommend displaying only metabolite sets containing at least two entries.

(B) The degree of KEGG enrichment ranked by enrichment score and assessed by the p value. Reprinted with permission from (Li et al., 2020).

metabolite exhibited greater variation in positive and negative ion modes, this phenomenon is attributable to different forms of adducts representing negatively and positively charged groups. Indeed, it is more accurate to identify metabolites by identifying ions from the same metabolite.

Problem 5: Limited data suggested from Spearman's correlation

Spearman's correlation analysis suggested limited information on the roles and mechanism between microbiota and metabolites, resulting in difficulty in validating predicted functions.

Potential solution

Spearman's correlations assume independence between interactions, simplifying the estimation procedure by reducing it to a combination of independent two-dimensional problems. However,

Table 7. Correlation network analysis of different comparisons on day 14

Node1	Relation	Node2	Rho	p-value
Clostridium_XlVa	neg	L-Glutamic acid	−0.8	0.0027
Bacteroides	pos	L-Glutamic acid	0.76	0.004
Akkermansia	pos	L-Glutamic acid	0.62	0.0373
Peptostreptococcaceae_unclassified	neg	L-Tyrosine	−0.75	0.0048
Blautia	neg	S-Adenosylhomocysteine	−0.65	0.0259
Others	pos	L-Tyrosine	0.69	0.013
Akkermansia	neg	Clostridium_XlVa	−0.86	0.013
Enterococcus	pos	Blautia	0.85	0.013
L-Aspartic acid	pos	L-Dopa	0.87	0.013
Others	neg	Dopamine	−0.72	0.0081

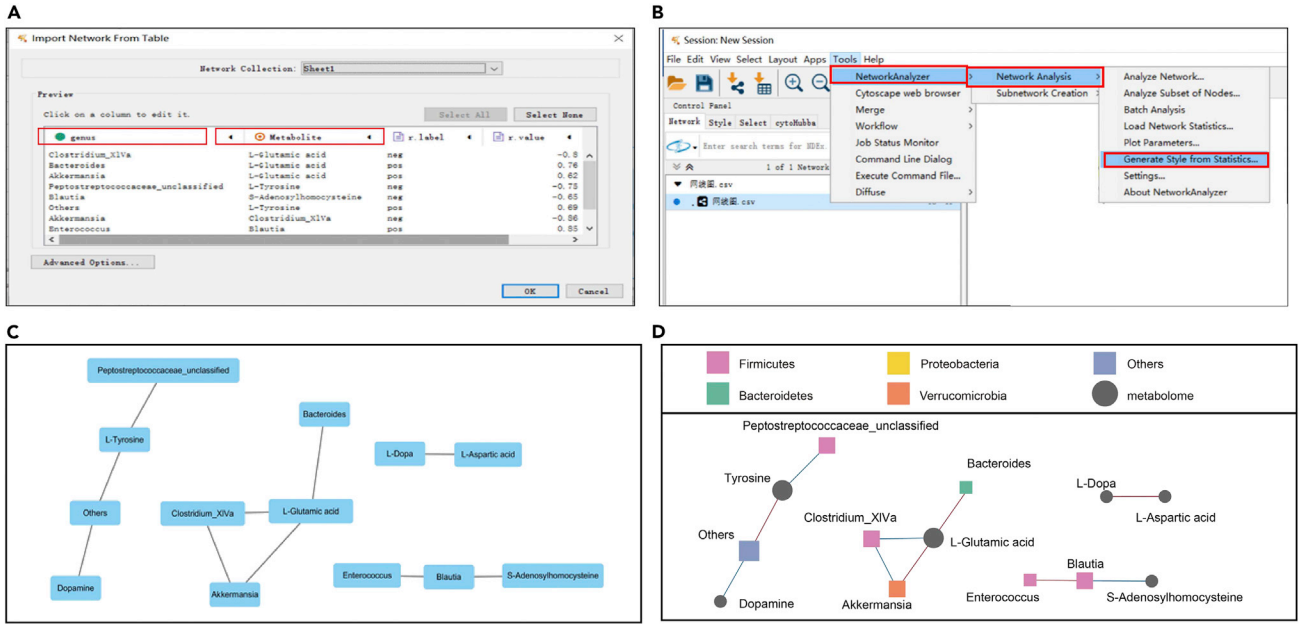


Figure 7. A network diagram is created using Cytoscape

(A) The correlation data are imported into Cytoscape. The genus is set as the source node, and the metabolite is set as the target node.

(B) Generate style form statistics by Network Analyzer menu under the Tools menu is used.

(C) Obtaining the initial network diagram.

(D) The size and color of the nodes and the color and thickness of the line can be set as some related parameters, such as the correlation coefficient and p value. Reprinted with permission from (Li et al., 2020).

these simplifications are not statistically valid for compositional data, especially for both microbiome and mass spectrometry datasets (Gloor et al., 2017). Recently, Morton et al. described a novel method for performing an integrated analysis of the microbiome and meta-metabolome, named microbe-metabolite vectors (MMVEC). Using neural networks (https://github.com/biocore/mmvec), MMVEC can estimate the conditional probability that each molecule is present given the presence of a specific microorganism (Morton et al., 2019).

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact Pengxu Qian (axu@zju.edu.cn).

Materials availability

This study did not generate new unique reagents.

Data and code availability

The published article (https://www.sciencedirect.com/science/article/pii/S2352396420304242) includes all datasets generated or analyzed during this study.

ACKNOWLEDGMENTS

We thank the members of the LC-Bio Technology Co. Ltd. for helpful discussions and advice. This work is supported by grants from the National Key R&D Program of China, Stem Cell and Translation Research (2018YFA0109300), the National Natural Science Foundation of China (81870080, 91949115, 82161138028, 82000187, 81900176), the Zhejiang Provincial Natural Science Foundation of China (LR19H080001), and the Leading Innovative and Entrepreneur Team Introduction

Program of Zhejiang (2020R01006). We are thankful for the technical support by the Core Facilities, Zhejiang University School of Medicine.

AUTHOR CONTRIBUTIONS

Conceptualization, methodology, and original draft writing, X.L., P.W., and X.Z.; writing, reviewing, editing, and overall supervision, Q.L., Y.L., H.H., and P.Q.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Peña, A.G., Goodrich, J.K., Gordon, J.I., et al. (2010). QIIME allows analysis of high-throughput community sequencing data. Nat. Methods *7*, 335–336. https://doi.org/10.1038/nmeth.f.303.

Caruso, V., Song, X., Asquith, M., and Karstens, L. (2019). Performance of microbiome sequence inference methods in environments with varying biomass. mSystems *4*. e00163-18.

Chambers, M.C., Maclean, B., Burke, R., Amodei, D., Ruderman, D.L., Neumann, S., Gatto, L., Fischer, B., Pratt, B., Egertson, J., et al. (2012). A cross-platform toolkit for mass spectrometry and proteomics. Nat. Biotechnol. *30*, 918–920. https://doi.org/10.1038/nbt.2377.

Cole, J.R., Wang, Q., Fish, J.A., Chai, B., Mcgarrell, D.M., Sun, Y., Brown, C.T., Porras-Alfaro, A., Kuske, C.R., and Tiedje, J.M. (2014). Ribosomal Database Project: data and tools for high throughput rRNA analysis. Nucleic Acids Res. *42*, D633–D642. https://doi.org/10.1093/nar/gkt1244.

Dobson, G.P., Letson, H.L., Biros, E., and Morris, J. (2019). Specific pathogen-free (SPF) animal status as a variable in biomedical research: have we come full circle? EBioMedicine *41*, 42–43. https://doi.org/10.1016/j.ebiom.2019.02.038.

Fischer, C., Koblmüller, S., Gülly, C., Schlötterer, C., Sturmbauer, C., and Thallinger, G.G. (2013). Complete mitochondrial DNA sequences of the threadfin cichlid (Petrochromis trewavasae) and the blunthead cichlid (Tropheus moorii) and patterns of mitochondrial genome evolution in cichlid fishes. PLoS One *8*, e67048. https://doi.org/10.1371/journal.pone.0067048.

Forsberg, E.M., Huan, T., Rinehart, D., Benton, H.P., Warth, B., Hilmers, B., and Siuzdak, G. (2018). Data processing, multi-omic pathway mapping, and metabolite activity analysis using XCMS Online. Nat. Protoc. *13*, 633–651. https://doi.org/10.1038/nprot.2017.151.

Gika, H.G., Wilson, I.D., and Theodoridis, G.A. (2014). LC-MS-based holistic metabolic profiling. Problems, limitations, advantages, and future perspectives. J. Chromatogr. B Anal. Technol. Biomed. Life Sci. *966*, 1–6. https://doi.org/10.1016/j.jchromb.2014.01.054.

Gloor, G.B., Macklaim, J.M., Pawlowsky-Glahn, V., and Egozcue, J.J. (2017). Microbiome datasets are compositional: and this is not optional. Front. Microbiol. *8*, 2224. https://doi.org/10.3389/fmicb.2017.02224.

Holler, E., Butzhammer, P., Schmid, K., Hundsrucker, C., Koestler, J., Peter, K., Zhu, W., Sporrer, D., Hehlgans, T., Kreutz, M., et al. (2014). Metagenomic analysis of the stool microbiome in patients receiving allogeneic stem cell transplantation: loss of diversity is associated with use of systemic antibiotics and more pronounced in gastrointestinal graft-versus-host disease. Biol. Blood Marrow Transplant. *20*, 640–645. https://doi.org/10.1016/j.bbmt.2014.01.030.

Kang, L., Weng, N., and Jian, W. (2020). LC-MS bioanalysis of intact proteins and peptides. Biomed. Chromatogr. *34*, e4633. https://doi.org/10.1002/bmc.4633.

Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M., and Glöckner, F.O. (2013). Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. Nucleic Acids Res. *41*, e1. https://doi.org/10.1093/nar/gks808.

Konstantinidis, K.T., Ramette, A., and Tiedje, J.M. (2006). The bacterial species definition in the genomic era. Philos. Trans. R. Soc. Lond. B Biol. Sci. *361*, 1929–1940. https://doi.org/10.1098/rstb.2006.1920.

Kuhl, C., Tautenhahn, R., Böttcher, C., Larson, T.R., and Neumann, S. (2012). CAMERA: an integrated strategy for compound spectra extraction and annotation of liquid chromatography/mass spectrometry data sets. Anal. Chem. *84*, 283–289. https://doi.org/10.1021/ac202450g.

Li, X., Lin, Y., Li, X., Xu, X., Zhao, Y., Xu, L., Gao, Y., Li, Y., Tan, Y., Qian, P., and Huang, H. (2020). Tyrosine supplement ameliorates murine aGVHD by modulation of gut microbiome and metabolome. EBioMedicine *61*, 103048. https://doi.org/10.1016/j.ebiom.2020.103048.

Lozupone, C., and Knight, R. (2005). UniFrac: a new phylogenetic method for comparing microbial communities. Appl. Environ. Microbiol. *71*, 8228–8235. https://doi.org/10.1128/aem.71.12.8228-8235.2005.

Lozupone, C.A., and Knight, R. (2008). Species divergence and the measurement of microbial diversity. FEMS Microbiol. Rev. *32*, 557–578. https://doi.org/10.1111/j.1574-6976.2008.00111.x.

Magoc, T., and Salzberg, S.L. (2011). FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics *27*, 2957–2963. https://doi.org/10.1093/bioinformatics/btr507.

Marti, M., Spreckels, J.E., Jenmalm, M.C., and Abrahamsson, T. (2021). A protocol for characterization of extremely preterm infant gut microbiota in double-blind clinical trials. STAR Protoc *2*, 100652. https://doi.org/10.1016/j.xpro.2021.100652.

Mathewson, N.D., Jenq, R., Mathew, A.V., Koenigsknecht, M., Hanash, A., Toubai, T., Oravecz-Wilson, K., Wu, S.R., Sun, Y., Rossi, C., et al. (2016). Gut microbiome-derived metabolites modulate intestinal epithelial cell damage and mitigate graft-versus-host disease. Nat. Immunol. *17*, 505–513. https://doi.org/10.1038/ni.3400.

Messineo, A.M., Gineste, C., Sztal, T.E., Mcnamara, E.L., Vilmen, C., Ogier, A.C., Hahne, D., Bendahan, D., Laing, N.G., Bryson-Richardson, R.J., et al. (2018). L-tyrosine supplementation does not ameliorate skeletal muscle dysfunction in zebrafish and mouse models of dominant skeletal muscle alpha-actin nemaline myopathy. Sci. Rep. *8*, 11490. https://doi.org/10.1038/s41598-018-29437-z.

Morton, J.T., Aksenov, A.A., Nothias, L.F., Foulds, J.R., Quinn, R.A., Badri, M.H., Swenson, T.L., Van Goethem, M.W., Northen, T.R., Vazquez-Baeza, Y., et al. (2019). Learning representations of microbe-metabolite interactions. Nat. Methods *16*, 1306–1314. https://doi.org/10.1038/s41592-019-0616-3.

Muhamad Rizal, N.S., Neoh, H.M., Ramli, R., A/LK Periyasamy, P.R., Hanafiah, A., Abdul Samat, M.N., Tan, T.L., Wong, K.K., Nathan, S., Chieng, S., et al. (2020). Advantages and limitations of 16S rRNA next-generation sequencing for pathogen identification in the diagnostic microbiology laboratory: perspectives from a middle-income country. Diagnostics *10*, 816. https://doi.org/10.3390/diagnostics10100816.

Nicholson, J.K., and Lindon, J.C. (2008). Systems biology: metabonomics. Nature *455*, 1054–1056. https://doi.org/10.1038/4551054a.

Okahashi, N., Ueda, M., Yasuda, S., Tsugawa, H., and Arita, M. (2021). Global profiling of gut microbiota-associated lipid metabolites in antibiotic-treated mice by LC-MS/MS-based analyses. STAR Protoc *2*, 100492. https://doi.org/10.1016/j.xpro.2021.100492.

Pertea, G. (2018). gpertea/fqtrim: Fqtrim Release v0. 9.7 (Zenodo).

Raja, G., Gupta, H., Gebru, Y.A., Youn, G.S., Choi, Y.R., Kim, H.S., Yoon, S.J., Kim, D.J., Kim, T.J., and Suk, K.T. (2021). Recent advances of microbiome-associated metabolomics profiling in liver disease: principles, mechanisms, and applications. Int. J. Mol. Sci. *22*, 1160. https://doi.org/10.3390/ijms22031160.

Reitmeier, S., Kiessling, S., Neuhaus, K., and Haller, D. (2020). Comparing circadian rhythmicity in the

human gut microbiome. STAR Protoc *1*, 100148. https://doi.org/10.1016/j.xpro.2020.100148.

Rivett, D.W., and Bell, T. (2018). Abundance determines the functional role of bacterial phylotypes in complex communities. Nat. Microbiol. *3*, 767–772. https://doi.org/10.1038/s41564-018-0180-0.

Rognes, T., Flouri, T., Nichols, B., Quince, C., and Mahé, F. (2016). VSEARCH: a versatile open source tool for metagenomics. PeerJ *4*, e2584. https://doi.org/10.7717/peerj.2584.

Segata, N., Izard, J., Waldron, L., Gevers, D., Miropolsky, L., Garrett, W.S., and Huttenhower, C. (2011). Metagenomic biomarker discovery and explanation. Genome Biol. *12*, R60. https://doi.org/10.1186/gb-2011-12-6-r60.

Seger, C., and Salzmann, L. (2020). After another decade: LC-MS/MS became routine in clinical diagnostics. Clin. Biochem. *82*, 2–11. https://doi.org/10.1016/j.clinbiochem.2020.03.004.

Stein-Thoeringer, C.K., Nichols, K.B., Lazrak, A., Docampo, M.D., Slingerland, A.E., Slingerland, J.B., Clurman, A.G., Armijo, G., Gomes, A.L.C., Shono, Y., et al. (2019). Lactose drives Enterococcus expansion to promote graft-versus-host disease. Science *366*, 1143–1149. https://doi.org/10.1126/science.aax3760.

Wen, B., Mei, Z., Zeng, C., and Liu, S. (2017). metaX: a flexible and comprehensive software for processing metabolomics data. BMC Bioinf. *18*, 183. https://doi.org/10.1186/s12859-017-1579-y.