The Pennsylvania State University

The Graduate School

MICROBIAL PROFILING VIA MULTIMODAL SINGLE CELL BIOSENSORS

A Dissertation in

Biomedical Engineering

by

Jyong-Huei Lee

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

May 2024

The dissertation of Jyong-Huei Lee was reviewed and approved by the following:

Pak Kin Wong
Professor of Biomedical Engineering
Dissertation Advisor
Chair of Committee

Darrell William Cockburn
Assistant Professor of Food Science

Scott Medina
Associate Professor of Biomedical Engineering

Justin Pritchard
Associate Professor of Biomedical Engineering

Nanyin Zhang
Professor of Biomedical Engineering

Yuguo Lei
Associate Professor of Biomedical Engineering
Director of Graduate Studies

ABSTRACT

The human microbiota, consisting of trillions of microorganisms inhabiting various regions of our bodies, is integral to numerous physiological processes such as digestion, immune response, and hormone regulation. Microbial imbalances, or dysbiosis, have been implicated in a plethora of health conditions. Exploiting the microbiota's potential can revolutionize the development of diagnostic and therapeutic strategies aimed at enhancing treatment outcomes, reducing complications, and preventing disease recurrence. A significant barrier to clinical application lies in the absence of rapid microbiota analysis methods that can produce clinically relevant data. This thesis introduces innovative multimodal biosensors designed for precise single-cell microbiota profiling. These biosensors enable the detection and functional assessment of microbial populations at the single-cell level, thus facilitating a holistic understanding of the microbiome's role in health and disease. The proposed biosensors provide a comprehensive analysis toolkit, capable of determining absolute microbial abundance, viability, spatial distribution, and gene expression profiles. Furthermore, we apply these multimodal biosensors to conduct microbial community profiling on clinical samples, with a particular focus on exploring the gut microbiome variations in mouse models of familial Alzheimer's disease. This research bridges a critical gap in microbial diagnostics and paves the way for integrating microbiome analysis into personalized medicine and clinical care.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

As I reflect on my PhD journey, my heart swells with gratitude for the myriad of individuals who have been pivotal to my achievements over the last five years. Achieving this milestone would have been unattainable without their unwavering support. I extend my heartfelt thanks to my committee members, Dr. Cockburn, Dr. Medina, Dr. Pritchard, and Dr. Zhang, whose expertise and mentorship were invaluable to my development. Their insightful feedback and encouragement greatly enriched my academic experience. Special thanks are due to my supervisor, Dr. Pak Kin Wong, whose guidance has been instrumental in my research. Dr. Wong not only initiated research topics but also oversaw experimental designs and fostered engaging discussions. His trust in his students created a nurturing and inspiring environment, making my time in the lab both enjoyable and profoundly enriching. I am also deeply thankful for the collaboration and insights provided by Dr. Liao, Dr. Mach, and Dr. Yang from Stanford University, as well as Dr. Francisco Diaz, Dr. Linden, Dr. Zhang, and Dennis from Penn State University. Their contributions were crucial in bringing our research to fruition. I would like to express my appreciation to my lab mates - Chriss, Mona, Lilly, Alex, Samara, Eric, Peter, Sam, and Hui. Together, we cultivated a thriving academic atmosphere that fostered not only innovative ideas and stimulating discussions but also enjoyable leisure activities. Their camaraderie has been a cornerstone of my PhD experience. My journey was also supported by my friends, whose encouragement and discussions were vital in propelling me forward. I am especially grateful for the support from my partner, cat, parents, and brother. Their understanding and backing were indispensable to my perseverance and success. Lastly, I wish to acknowledge the Pennsylvania State University and the Department of Biomedical Engineering for

granting me the opportunity to pursue my studies. This journey has been one of the most fulfilling chapters of my life, and I am profoundly thankful for everyone who played a part in it.

This research was supported, in part, by the National Institutes of Health (NIH), grant number R01AI153133. The findings and conclusions do not necessarily reflect the view of the funding agency

Chapter 1 Introduction

Portions of this chapter are reproduced from: Jyong-Huei Lee, Siew Mei Chin, Kathleen E. Mach, April Bobenchik, Joseph C. Liao, Samuel Yang, Pak Kin Wong (2024). Translating Microbiota Analysis for Clinical Applications. Nature Reviews Bioengineering

The human body is estimated to host approximately 30 trillion microorganisms.[1] While conventional cultivation-based techniques failed to fully characterize the microbiota, i.e., the collection of microorganisms, newer molecular approaches have enabled the analysis of the uncultivable species with unprecedented resolution. The Human Microbiome Project revealed that many species associated with the body are obscure commensal organisms serving beneficial functions.[2,3] These microbes can prime the immune system, support metabolism, break down dietary components, and regulate hormonal activities.[4-6] For example, the gut microbiota modulates signaling pathways involved in intestinal mucosa homeostasis through fermentation and metabolite production, and it is associated with various physiological functions and diseases.[5] In addition to the gut, the significance of microbiota in other organs has also been revealed. For instance, while it was initially believed that the respiratory tract and urinary tract were sterile, recent explorations of the microbiome have demonstrated this is not the case.[7,8] These microbial communities not only help prevent the overgrowth of harmful pathogens but also are associated with various pathophysiological processes.

Analyzing microbial communities presents new opportunities in clinical management. Microbial patterns are observed when comparing the microbiota of healthy individuals and patients.[9,10] These patterns may help physicians in evaluating patient's health conditions,

susceptibility to diseases, and treatment responsiveness. To give a few examples, the composition and diversity of intestinal microbiota are associated with *Clostridioides difficile* infection, inflammatory bowel diseases, and colorectal cancer.[11-13] Similarly, urinary and pulmonary microbiota are linked to infections, kidney stone diseases, and cancer of the urinary and respiratory systems.[14,15] Importantly, the microbiota can contribute to diseases far beyond the organ it resides in.[16] In addition to diagnostics and prognostics, techniques for manipulating microbial communities, such as prebiotics, probiotics, and fecal transplants, hold promise for treating recurrent diseases and enhancing treatment efficacy.[17] For instance, fecal transplants are recommended for treating persistent *C. difficile* infections and are shown to improve the performance of immune checkpoint inhibitor.[18-20] Another remarkable example of microbial therapy is Bacille Calmette-Guérin (BCG), an attenuated strain of *Mycobacterium bovis* that has been approved as an intravesical therapy for intermediate- and high-risk non muscle invasive bladder cancer since 1990.[21] Notably, many medical conditions and treatment options, such as infections, stress, stone ablation, antibiotic usage, and dietary interventions, can also influence the microbiota.[22-24] These examples illustrate the significance of the microbiota in clinical management. Nevertheless, few clinical studies have monitored the dynamic changes of the microbiota resulting from medical intervention, colonization, microbe-microbe interactions, and host-microbe communication.[25] The microbiota is seldom integrated into the clinical decision-making process, despite its potential to optimize clinical care. The ability to rapidly analyze the microbiota in patients will facilitate the translation of microbiota analysis into medical management.

High throughput sequencing methods are powerful techniques for characterizing human microbiomes in exploration and discovery studies. However, despite their ability to improve our understanding of the microbiota, these techniques are often expensive, slow, cumbersome, and labor-intensive. While the scalability of sequencing methods can reduce costs and processing time,

conducting experiments in batches with a small number of samples remains costly and time-consuming. There are challenges in translating standard sequencing techniques for clinical applications. Rapid microbiota analysis systems for medical applications must be cost-effective, have a quick turnaround time, and be easy to use. In particular, the abundance of different types of microorganisms varies significantly between individuals, and a straightforward definition of a 'healthy' microbiota may not exist. Efficiently tracking the evolution of the microbiota in the same patient is required to understand the progression of a disease. The ability to analyze microbiota effectively is also crucial for optimizing the efficacy of medical treatments that may directly or indirectly modulate the microbiota. In addition to cost and time considerations, it is critical that the microbiota analysis system focuses on clinically relevant information. Complex microbiota data can be challenging to interpret in clinical settings and will have limited clinical utility. Furthermore, conventional sequencing techniques have limitations in absolute quantification, live/dead cell detection, multiomic analysis, and resolving the spatial distribution of microbiota. These issues hinder the translation of microbiota discoveries into clinical practice.

In this chapter, we analyze the potential of rapid microbiota analysis techniques for clinical translation, with a primary focus on bacterial communities that are extensively studied in recent years. We use the term 'microbiota' to describe the collection of microorganisms and 'microbiome' to encompasses not only the microorganisms themselves but also their surrounding environment, which includes microbial metabolites and components of the host that interact with these microorganisms. We first review genomic, proteomic, and cell-based technological platforms potentially suitable for microbiota analysis in at least some clinical applications (Figure 1-1). Then, we highlight valuable capabilities currently unavailable or challenging to implement in standard sequencing or other microbiota analysis techniques (

Table 1). Finally, we conclude by discussing the challenges and opportunities associated with integrating rapid microbiota analysis technologies into clinical decision workflows.

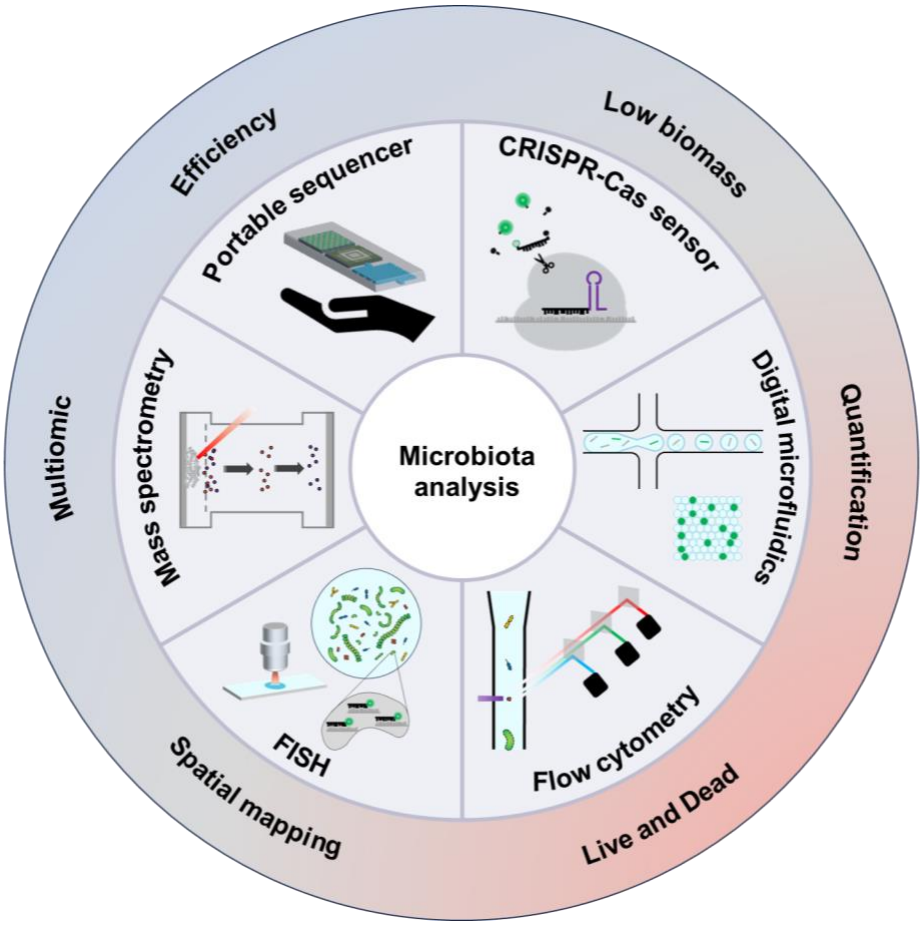


Figure 1-1 Rapid Microbiota Analysis Systems. Microbiota analysis systems can be classified into genomic, proteomic, and cell-based platforms. These technologies have inherent advantages and disadvantages. Ongoing efforts are dedicated to improving their cost and time efficiencies, as well as their abilities to quantify absolute abundance, handle low biomass, provide spatial mapping, distinguish live and dead microbes, and perform multiomic analysis.

	Genomics		Proteomics		Cell-based assay	
	Sequencing	**Biosensors (e.g., CRISPR, toehold)**	**Mass spec**	**Metaproteomics**	**Flow cytometry**	**FISH**
Assay time	Hours to weeks	< hours	< hours after colony formation	< hours	Hours	Hours
Assay cost	High	Low	Medium	Medium	Low	Low
Classification	Genus to strain	Genus to species	Species strain	Species strain	Genus to species	Genus to species
Instrumentation cost	Low to high	Low to medium	High	High	High	Medium
Quantification	Relative	Semi-quantitative	No	Relative	Single cell	Single cell
Low biomass	No	Yes	No	No	Yes	Yes
Spatial Mapping	No	No	Imaging MS	No	No	Yes
Viability	No	Maybe	No	No	Yes	Yes
Omics	RNA	RNA	Protein Metabolites	Protein Metabolites	RNA Immunostaining	RNA Immunostaining

Table 1 Comparison of microbiota analysis approaches

Rapid microbiota analysis techniques for clinical applications

In this section, we discuss microbiota analysis platforms that show promise as candidates for clinical translation. High throughput sequencing techniques undoubtedly stand out as a major enabling technology for microbiota analysis. Recently, other molecular methods, such as multiplex PCR, CRISPR-Cas sensors, DNA microarray, mass spectrometry, flow cytometry, and fluorescence in situ hybridization (FISH), have also gained traction for rapid microbiota analysis. While these techniques may not offer the same resolution as high throughput sequencing, a comprehensive analysis of the entire microbial community is not always necessary. These culture-free approaches may be suitable for situations that demand a rapid and cost-effective analysis of specific components of the microbiota (e.g., an increase in a probiotic or a change in one or more taxonomic groups). We classify microbiota analysis techniques into genomic, proteomic, and cell-based platforms, as the processing and analysis procedures in each approach are often associated with specific advantages and limitations (Figure 1-2). Additionally, we highlight selected

biosensing technologies that hold promise for microbiota analysis, even though they may not have been extensively applied in microbiota analysis. Comprehensive reviews of microbiological sensing platforms for other medical applications are available elsewhere.[26-28]

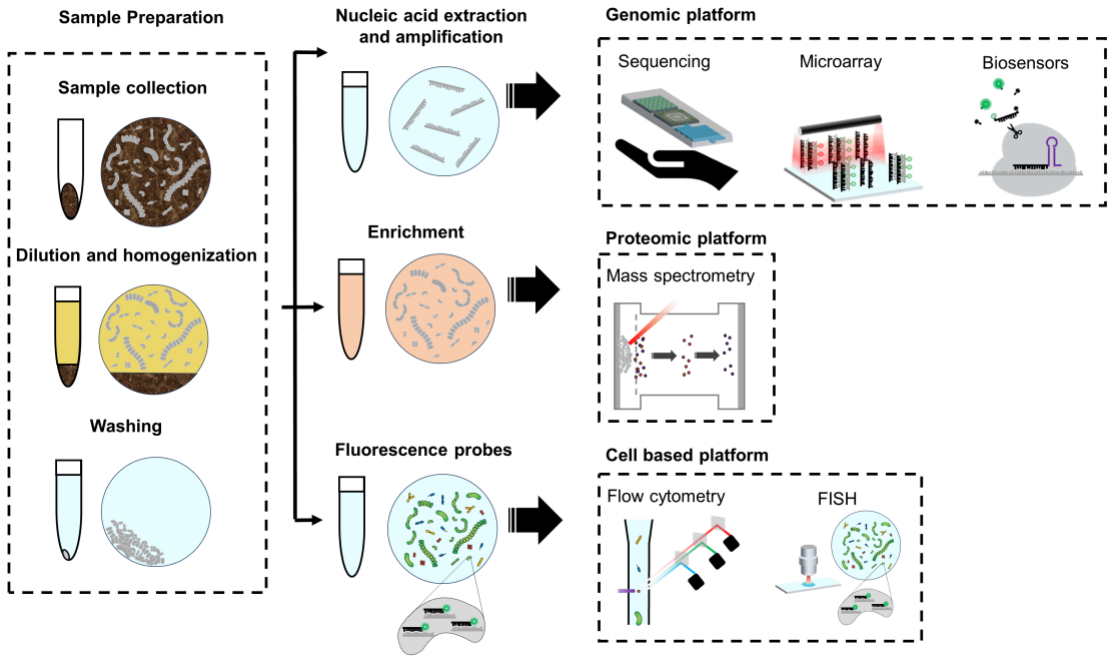


Figure 1-2 Sample preparation and processing for microbiota analysis. Sample preparation steps, such as homogenization and washing, are performed depending on the sample type. For genomic assays, cells are lysed. DNA or RNA targets are then extracted and processed to be characterized by sequencing, microarrays, or molecular biosensors. For proteomic assays, microbial cells are cultured to form colonies for mass spectrometry analysis. For cell-based assays, the cell samples are fixed and detected with nucleic acid probes or other reagents.

Genomic platforms

High-throughput sequencing techniques, coupled with advanced computational tools, have facilitated the culture-free analysis of genomic nucleic acids, enabling the characterization of compositions within microbial communities. Two primary methods for characterizing microbes are marker genes and metagenomic approaches.[29] Marker gene sequencing, utilizing genes like 16S rRNA, 18S rRNA, and internal transcribed spacer (ITS), targets hypervariable regions that can be compared to databases of known microbial sequences, allowing researchers to characterize

microbial communities. Well-established 16S rRNA databases are readily available for this purpose.[30] However, 16S rRNA sequencing is often limited to taxonomic categorization at the genus level, restricting the analysis of commensal, opportunistic, and pathogenic bacteria at the species and strain levels and omitting viral and fungal organisms.[31,32] Furthermore, 16S rRNA sequencing can introduce errors due to substitutions and gaps. On the other hand, metagenomic sequencing involves the random sequencing of short fragments of DNA from a mixed microbial sample, which are then computationally assembled and analyzed to identify and characterize the diverse microorganisms present.[33] While metagenomic sequencing incurs additional steps and higher costs compared to 16S rRNA sequencing, it provides a comprehensive readout of all genomic nucleic acids present. This untargeted approach has lower bias and enables the identification of both bacterial and non-bacterial organisms. Nevertheless, metagenomics can be more sensitive to host DNA contamination and requires greater sequencing depth for accurate calibration.[34] Remarkably, clinical metagenomics has been applied to analyze microbial cell-free DNA in plasma for infection diagnostics.[35] Metagenomics can also be combined with microfluidic droplet barcoding for single cell genomic sequencing of microbial communities.[36]

Other nucleic acid biosensing platforms have been applied for microbiological analysis. For example, multiplex PCR assays are available for the rapid identification of bacterial pathogens associated with respiratory and gastrointestinal diseases in clinical settings.[37,38] Microarrays provide a method for the multiplex detection of amplicons for scalable microbiota characterization.[39] They identify multiple species in parallel through an extensive array of probes arranged on a surface or microspheres. Phylogenetic DNA microarrays have been demonstrated for characterizing dysbiosis of fecal microbiota in Crohn's disease and airway microbiota in asthma.[40,41] However, microarrays can have limitations in quantification and are susceptible to high background levels due to cross-hybridization. Recently, CRISPR-Cas biosensors, such as CRISPR/Cas9-

triggered isothermal exponential amplification reaction (CAS-EXPAR), specific high sensitivity enzymatic reporter unlocking (SHERLOCK), and DNA endonuclease-targeted CRISPR trans reporter (DETECTR), have been demonstrated for the rapid analysis of microorganisms.[42,43] The sequence specificity and multiplexity make CRISPR biosensors promising candidates for microbiota analysis. These biosensors can be integrated with lateral flow devices, enabling microbial detection with the naked eye or fluorescent reporters. In addition to CRISPR, RNA toehold switch biosensors, which recognize and respond to target RNA molecules by changing their conformation, can also be applied for microbiota analysis.[44] An RNA toehold switch biosensor combined with nucleic acid sequence-based amplification (NASBA) has been developed to provide a low-cost platform for analyzing gut microbiota and host markers.[45]

Proteomic platforms

Despite the outstanding performance of genomic platforms in microbiota analysis, proteomic platforms also hold promise for clinical translation. Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) is a powerful technique for accurately identifying bacterial species. Indeed, MALDI-TOF MS has long been applied in the clinical workflow of microbiological analysis.[46] In MALDI-TOF MS-based bacterial identification, a bacterial colony is prepared from a culture of clinical specimens, such as blood, urine, or tracheal aspirate. A matrix is applied to the bacterial sample to desorb and ionize the bacterial proteins when exposed to laser energy. The ions produced travel through a flight tube and are separated based on their time of flight. Lighter ions reach the detector faster than heavier ions, and this information is used to determine the mass-to-charge ratios of the ions. The generated mass spectrum, which represents a fingerprint of the bacterial sample, is then compared with a database of reference spectra from known bacterial species to identify the bacterium.

MALDI-TOF MS bacterial identification is accurate and rapid, often providing results within minutes from an isolated bacterial colony. This technology can identify bacteria at the species level, significantly improving clinical microbiology by enhancing the speed and precision of bacterial identification compared to traditional methods like overnight biochemical reactions.[46] Mass spectroscopy is particularly valuable for diagnosing bacterial infections, guiding antibiotic treatment decisions, and monitoring the spread of bacterial pathogens in healthcare settings.[47] However, mass spectrometry has a limitation: it typically requires isolated bacterial colonies to be grown from polymicrobic mixtures. The process is time-consuming and is limited to culturable bacteria. Efforts have been devoted to computationally deconvoluting mixtures of bacteria.[48] Magnetic enrichment has also been demonstrated for detecting bacterial mixtures directly from urine for urinary tract infection diagnostics.[49] Recently, single cell mass spectroscopy is available for analysis molecular contents of individual cells, including bacteria.[50,51] Mass spectroscopy is also capable of metabolomic analysis and can be integrated with sequencing platforms to provide comprehensive microbiome analysis.[52,53] Advances in microfluidics, automated sample preparation, and computational analysis may further expand the application of mass spectrometry in microbiota analysis in clinical settings.

Metaproteomics, which involves large-scale protein identification and quantification, represents a new initiative for the characterization of the microbial communities.[54] Metaproteomics uses techniques such as MS and liquid chromatography along with bioinformatic database search workflows to separate and identify peptides and proteins to identify taxonomic composition and functions of a microbial community.[55] The metaproteomic approach not only allows characterization of the microbiota composition but also evaluates the functional proteins that are associated with immune, metabolic, and hormonal activities. Recently studies have demonstrated the use of metaproteomics for investigating gut, urinary, and respiratory microbiota functions in

patients.[56-58] To give an example, the impact of vaginal microbiota on the effectiveness of tenofovir gel microbicide in HIV prevention among women has been investigated using metaproteomics.[59] The study found that tenofovir was significantly more effective (61% reduction in HIV incidence) in women with *Lactobacillus*-dominant vaginal microbiota compared to those with non-*Lactobacillus* bacteria (18% reduction). The presence of non-*Lactobacillus* bacteria was associated with lower levels of detectable tenofovir due to rapid metabolism, linking vaginal bacteria to the microbicide's efficacy. Challenges of metaproteomics include the requirement of extensive reference datasets and advanced computational workflows.

Cell based platforms

Another category of rapid microbiota characterization techniques includes cell-based assays, such as FISH and flow cytometry, which maintain the integrity of microbial cells.[60,61] These assays have been applied in clinical microbiology for a long time. Detecting molecular markers in intact cells allows for microbial identification and quantification without the need for cultivation, addressing the challenge of dealing with unculturable species in microbiota analysis. Similar to other genomic-based methods, cell-based platforms can identify microbial species using 16S rRNA or other marker genes. Microbial cells are usually isolated and fixed to preserve their structure. The microbes are then hybridized with nucleic acid probes and detected using a fluorescence microscope or a flow cytometer. In flow cytometry, light scattering and fluorescence signals can also indicate sizes, cell structures, and molecular information of each cell. High-resolution microbiota flow cytometry based on bacterial shape and DNA content has been demonstrated as a rapid and low-cost method for characterizing fecal microbiota and predicting Crohn’s disease states.[62,63] Cell-based assays are particularly valuable for identifying bacteria in situations where quantification is important and when knowing the presence and location of specific bacteria is

crucial for diagnosis and treatment (e.g., biofilm formation in implant- and catheter-associated infections).[64,65]

Due to the small volume of a cell, one notable characteristic of intracellular sensing is its high effective concentration, eliminating the need for amplification in other genomic analysis techniques. Another advantage of preserving the integrity of bacterial cells is their ability to provide precise quantitative counts down to the level of single cells. Furthermore, it is possible to detect other molecular markers and distinguish between live and dead cells. It is also possible to reserve the cell viability for downstream phenotypic assays.[66] For instance, the combination of a monoclonal antibody and a viability dye has been shown to detect live *Salmonella enterica* Serovar Typhimurium in the presence of a large number of non-target and dead bacteria.[67] However, a challenge lies in developing a permeabilization protocol suitable for a wide spectrum of bacteria. Specifically, Gram-positive and Gram-negative bacteria have distinct membrane structures characterized by a thick layer of peptidoglycan in Gram-positive bacteria and a double membrane structure in Gram-negative bacteria. Another challenge is associated with the presence of strong autofluorescence, which can affect the interpretation of results.

Fluorescence detection is often limited by the number of colors available due to optical bandwidth. Remarkably, the multiplexity of FISH probes can be significantly enhanced through barcode schemes and sequential staining.[68,69] For example, the high-phylogenetic-resolution microbiome mapping technique known as fluorescence in situ hybridization (HiPR-FISH) employs a 10-color barcode scheme to distinguish 1,023 isolates of *Escherichia coli* and to conduct spatial mapping of microbial communities.[68] This technique has been applied to study the antibiotic disruption of the gut microbiome and the stability of human plaque biofilms.[68] By incorporating sequential FISH, the parallel sequential fluorescence in situ hybridization (par-seqFISH) technique has been developed for analyzing spatial transcriptomics at the single-cell resolution.[69] The

technique detects a set of 105 genes in *Pseudomonas aeruginosa* in planktonic and biofilm cultures. Furthermore, this platform enables the measurement of other characteristics, such as cell size and chromosome copy number, within the same cell.

Characteristics of microbiota analysis systems for medical applications

In this section, we discuss characteristics of microbiota analysis technologies that are important for clinical applications. We pay specific attention to aspects that are challenging or not achievable with conventional sequencing and microbiota analysis methods. Specifically, we highlight important considerations including cost and time efficiencies, absolute quantification of bacteria species, low biomass detection, spatial analysis, live and dead bacteria distinction, and multimodal analysis of the microbial community.

Cost, time, and implementation efficiencies

Diagnostic technologies must be highly effective to seamlessly integrate into clinical workflows. The most efficient approach, however, depends on the situation. In some cases, a comprehensive analysis of all microbial species may not be necessary. For example, the use of bioactive ingredients, such as probiotics, prebiotics, synbiotics (combinations of probiotics and prebiotics), and genetically modified organisms, holds promise for addressing healthcare challenges.[70] Monitoring commensal bacteria can aid in tracking adherence, growth, and colonization of beneficial species, optimizing the duration and route of treatment administration. Characterizing specific pathogens and opportunistic bacteria may optimize the treatment and prevention of recurrent diseases, such as *C. difficile* infection and urinary tract infections[71,72] Characterizing respiratory microbiota in ventilated patients in the intensive care unit has been

suggested as a way to manage ventilator-associated pneumonia.[73] Monitoring the retention of BCG may also benefit the optimization of intravesical therapy for bladder cancer. In these situations, a rapid assay targeting a relatively small panel of bacteria, while allowing effective measurement at multiple time points, may be most effective. On the contrary, some situations may necessitate a more comprehensive analysis of microbial composition and functions. Examples include evaluating the microbiome in patients undergoing fecal microbiota transplants. In a study of fecal transplant for ulcerative colitis patients, the investigators used 16S, metagenomic, and immunoglobulin A (IgA) sequencing to analyze samples from patients who received fecal microbiota transplants or a placebo.[74] The research revealed dynamic competitive interactions between donor and patient strains, demonstrating that transferred microbes are not static. Some patients experienced a loss of donor bacteria, coinciding with pathogenic bacterial growth and worsening symptoms. These studies highlight that the effectiveness of a microbiota analysis system is context dependent.

Recent efforts have been dedicated to the development of portable and rapid sequencers.[75] For instance, single molecule sequencing technology, e.g., nanopore sequencing, has enabled sequencing analyses in a highly portable format.[76] Since Oxford Nanopore Technologies introduced the first nanopore sequencer, MinION, a pocket-sized sequencer for DNA and RNA, sequencing technology and its applications in both fundamental and practical research have advanced significantly. While portable sequencing systems may have limitations in terms of biases and error rates, their portability and rapid workflow make them promising candidates for clinical applications. For example, the MinION system has demonstrated its capability to analyze fungal pathogens from sample preparation to interpretation in as little as 2.5 hours.[77] However, the implementation of high throughput sequencing technology generally can be limited by library preparation and data analysis. As automated library preparation techniques and cloud computing for data analysis become

available,[78] the cost and turnaround time will continuously decrease, potentially enabling high-throughput sequencing technology for a wide range of medical applications.

Adapting technologies into clinical workflows is challenging. Considering infectious disease diagnostics, multiple diagnostic platforms based on multiplex PCR are already available for the rapid analysis of syndromic-based panels of microorganisms.[79,80] These systems provide high-resolution molecular analysis in rapid and automated formats. Sample preparation, amplification, detection, and analysis for respiratory, bloodstream infections, meningitis, bone and joint infections, and gastrointestinal pathogens can be completed in one to two hours. These systems enable rapid diagnostics for *C. difficile* and ventilator-associated infections, reducing the empirical use of broad-spectrum antibiotics.[79,80] Nevertheless, the adoption of these technologies in routine medical practice remains slow. The cost, throughput, and accuracy are among the most important hurdles preventing the wide adoption of these platforms.[81,82] For microbiota analysis, additional hurdles may also include sample heterogeneity and challenges in data interpretation. Advancements in technologies, such as microfluidics, single-cell analysis, and artificial intelligence, may be required to improve the cost, assay time, and implementation efficiency of microbiota analysis.[83]

Absolute quantification

Quantification is an area of growing interest in microbiota research.[84] As the abundance of microbiota is associated with various diseases and medical treatments, quantification can help identify imbalances in microbial communities and optimize the efficiency of microbiota modulation strategies.[24,85-87] Quantifying the total bacterial load is crucial for data interpretation, sample quality control, and comparisons between different methods and studies.[88] Variations in

microbial load among healthy individuals can have up to a tenfold difference in magnitude.[89] When significant variations in bacterial load exist among samples, using relative measurements can be challenging for establishing associations between bacterial composition and metabolite production or health outcomes. For instance, the ratio between *Firmicutes* and *Bacteroidetes* is often considered an indicator of intestinal homeostasis and is associated with various medical conditions, such as inflammatory bowel disease.[90] However, an increase in the ratio could result from an increase in *Firmicutes*, a decrease in *Bacteroidetes*, or both. Interestingly, the bacterial load can be linked to enterotype distinctions.[89] A reduction in microbial abundance is a feature of the microbiome alterations for patients with Crohn's disease.

Sequencing typically provides relative abundance data based on the ratio of operational taxonomic units (OTUs) to the total number of sequences. The values can be influenced by various sample collection, storage, processing, library preparation, and data analysis steps.[91] In particular, variations in cell lysis and DNA extraction efficiency between different bacterial strains due to different cell wall and membrane structures can introduce bias in both relative and absolute measurements of genome abundance.[92] The copy number of 16S rRNA genes across taxa is another source of bias. Furthermore, the PCR amplification process may introduce significant biases, which can be attributed to a range of factors, including the properties of the PCR template (e.g., concentration and GC content), primer selection (encompassing coverage and mismatches), polymerase choice, and the particulars of the PCR protocol (including annealing temperature and cycle number).[93] Data processing and statistical analysis methods can also create additional uncertainties in the interpretation of the results.[94]

Quantifying the bacterial load in microbiota analysis can be achieved by digital PCR, flow cytometry, and FISH.[84] Digital PCR is a highly sensitive molecular biology technique used for

quantifying and analyzing nucleic acids. In microfluidic digital PCR, the sample is partitioned into thousands or even millions of droplets, each containing a minute quantity of the target nucleic acid. This partitioning allows for the quantification of the nucleic acid based on the number of positive and negative reactions, rather than relying on a standard curve in qPCR. Digital PCR has been demonstrated for strain specific quantification of a set of six probiotics.[95] By combining digital PCR and melt curve analysis, mixtures of 16 bacteria can be distinguished and quantified.[96] However, the loading of bacterial sample, which requires high singlets and low empty droplets and multiplets (i.e., droplets with more than one cell), present a tradeoff between dynamic range and throughput.

Cell-based assays, e.g., flow cytometry and FISH, are valuable tools for quantifying microbial groups or species within complex samples.[60,61] They provide useful information on the abundance of microbes by direct counting the microbes in the samples. Yet, accurate quantification and data analysis also require specialized equipment and expertise. Distinguishing bacteria species with a high resolution also requires advanced probe designs and imaging techniques. In quantitative microbiome profiling, high throughput sequencing and quantitative assays (e.g., flow cytometry, digital microfluidics, and qPCR) are combined to address the tradeoff between analysis resolution and quantification accuracy in microbiota analysis. For example, integrating amplicon sequencing and flow cytometry reveals large variation in bacterial loads of feces.[89] The combination of microfluidic droplet barcoding and genomic sequencing demonstrates single-cell genomic sequencing of microbial communities.[36]

Low biomass samples

A consideration related to quantification in microbiota analysis is associated with the large range of bacterial concentrations in physiological samples. The human microbiota is estimated to involve ~10^{13} microorganisms. The colon is one of the densest regions, with up to 10^{11} bacteria/g of luminal content. The concentration of an individual species in a stool sample can range from 10^4 to 10^8 bacteria/ml. In the respiratory and urinary tracts, which were originally considered to be sterile, the microbiota have significantly lower densities of 10^4 bacteria/g of lung tissue and 10^2-10^4 bacteria/ml of urine.[97,98] Nevertheless, most sequencing technologies are optimized for high microbial mass samples and recommend a large amount of DNA on the order of ~10 ng. Despite their diagnostic value, urinary and respiratory samples with low bacterial density are particularly prone to issues related to the efficiencies of DNA or RNA extraction, amplification, and detection. Low biomass samples can be easily overwhelmed by background DNA from the environment or the host. Multiple sources, such as collection procedures, sampling handling, and laboratory environment, may contribute to contamination and biases in the analysis of low biomass samples.[25] The low signal-to-noise ratio can lead to inaccurate or incomplete taxonomic profiles of the microbiome.[99] It has been suggested that low-density communities (below 10^6 bacteria/ml) can introduce significant error in microbiota analysis, and over 10^6 bacteria/ml is recommended for microbial analysis by sequencing.[100]

Various strategies have been applied for handling low biomass samples.[101] For instance, a method called 2bRAD sequencing for Microbiome (2bRAD-M) has been reported to address challenges in sequencing microbiome samples with low microbial biomass or degraded DNA.[102] This approach digests microbial genomes with type IIB restriction endonucleases to produce iso-length DNA fragments to reduce bias in PCR amplification. The technique sequences only about 1% of the microbiota but can provide detailed taxonomic profiles of bacteria, archaea, and fungi at

the species level. It is capable of accurately profiling difficult samples with minimal DNA (picogram level), high host DNA contamination, or severely fragmented DNA. Testing on various sample types, including stool, skin, environmental, and clinical formalin-fixed, paraffin-embedded samples, has shown successful reconstruction of comprehensive microbial profiles including low abundance species with high resolution and accuracy.

Digital microfluidics can provide an effective way to increase the signal to noise ratio for low biomass samples. Compartmentalized microbial DNA in droplets provide a way to separate the target from background DNA contamination, which increases the effective signal to noise ratio. It has been demonstrated that digital PCR can faithfully amplify a low amount of template (50-100 pg) and minimize the influence of DNA contamination in the analysis.[103] Similarly, analyzing 16S rRNA in intact cells with FISH and flow cytometry naturally creates an enclosed compartment, which eliminates the need of DNA amplifications and the influence of DNA contamination. These techniques provide promising approaches to studying bacteria in low biomass samples by directly characterizing bacteria at the single cell level.

Mapping spatial distribution of microbiota

The density and composition of the microbiota depend heavily on the location in the body. Spatial distribution is, therefore, a crucial parameter when applying microbiota analysis systems for medical applications. For instance, microbiota in the duodenum, jejunum, ileum, and colon differ by orders of magnitude, contributing to diverse functions.[104] Even in a similar location, the microbiota in stool, luminal contents, and biopsies can exhibit significant differences.[105,106] Similarly, microbiota in the kidney, bladder, ureter, and urethra play distinct roles in health and diseases. *Proteus mirabilis*, which produces urease, is associated with urea stone formation in the

kidney, while the same bacterium in the lower urinary tract represents a significant cause of catheter-associated infections.[107] Analyzing *P. mirabilis* and associated microbiota in different regions of the urinary tract may have important implications for evaluating the risk of development of different urological diseases.

The sample collection method is an important consideration for analyzing the microbiota in different locations of the body. When specific spatial locations are crucial, medical procedures like colonoscopy, bronchoscopy, and cystoscopy may be required to obtain specific samples for analysis. On the other hand, stool and urine, which can be collected non-invasively, may facilitate longitudinal microbiota analysis in patients. For the diagnosis of urinary tract infections, midstream urine samples are typically preferred to reflect bladder bacterial counts. This strategy minimizes contamination by bacteria colonizing the distal urethra and genital mucosa. Analyzing different portions of the urine sample may provide a strategy for analyzing the spatial distribution of the microbiota in a non-invasive manner.

In addition to spatial mapping at the organ scale, spatial mapping at the tissue-cell level is also crucial for clinical applications. For example, analyzing the biofilm on urinary catheters can provide insights into the initial site of bacterial biofilm formation on indwelling catheters and guide strategies for preventing catheter-associated complications.[108] In these situations, FISH can be performed to measure spatial distribution, bacterial composition, and gene expression profiles on the sample. However, autofluorescence of bacteria and biofilm can significantly influence the results. Advanced imaging protocols and noise reduction algorithms are needed to improve the accuracy of the analysis.[109] Additionally, imaging mass spectrometry is available for visualizing metabolites and microbe interactions in microbiological analysis down to the single cell level.[110,111]

An interesting example of spatial analysis is kidney stone microbiota. Kidney stone is associated with various bacteria, and stone obstruction of the urinary tract can lead to pyelonephritis, i.e., kidney infections.[107,112] However, urine culture is a poor predictor for the presence of bacteria and biofilms on kidney stones.[113-115] Patients with no growth urine culture can develop severe postoperative complications including urinary tract infections and urosepsis, which is associated with the stone microbiota. An enhanced culture technique (EQUC) and 16S rRNA gene sequencing have been demonstrated for characterizing stone microbiota of stone patients.[116] The two methods complement each other by compensating for each other's strengths and weaknesses. Deep 16S rRNA sequencing allows in-depth microbiota analysis without the need for bacterial cultivation. However, it struggles with precise identification below the genus level. In contrast, EQUC can capture bacteria for full-scale characterization, contingent on their cultivability. The isolated bacteria can be identified to the species level by MALDI-TOF MS or whole genome sequencing.

Live and Dead Bacteria

A limitation of genomic analysis is its inability to determine the viability of cells. In fresh feces, dead cells can constitute over 30% of the total microorganisms.[117] This issue could impede the analysis of how microbiota evolves under antibiotic treatment or other medical procedures. Live bacteria are crucial as they actively engage in and contribute to biomass production. Interestingly, while dead bacteria do not participate in biomass production, they represent unique characteristics of individuals and can reflect interactions with the host and medical intervention. This has been demonstrated in studies involving live, injured (dormant or inactive), and dead bacteria of the fecal microbiota.[118] Analyzing both live and dead bacteria is important to interpret the influence of the microbiota on patients' conditions and host conditions.

Several methods can be applied to discriminate between viable and non-viable microorganisms.[119] Culture-dependent approaches are often considered the gold standard, but they prove ineffective when dealing with unculturable bacteria, including microbes that only grow under specific conditions and viable but non-culturable (VBNC) bacteria.[119] Culture-independent approaches analyze either the intactness of the cell membrane or metabolic activity. Viability dyes, such as Propidium iodide (PI) and SYTO, selectively stain nucleic acids in dead cells depending on the intactness of the cell membrane. However, they may have limitations in effectively staining some species.[120] Metabolic assays, such as Calcein-AM for esterases and resazurin for mitochondrial enzymes, focus on measuring metabolic activity but might overlook metabolically dormant spores.

Both viability stains and metabolic assays are compatible with microbiota analysis. They can be implemented alongside cell-based assays, such as FISH and flow cytometry, and occupy one of the fluorescence channels. For example, a combination of a cell viability kit and flow cytometry has been applied to rapidly analyze microbial populations and sort extremely oxygen-sensitive species like *Faecalibacterium prausnitzii* from fecal materials.[121] Additionally, propidium monoazide (PMA) can be combined with genomic assays to analyze only viable cells.[122] In this method, PMA, a photoactivatable DNA-intercalating dye, is used to bind extracellular and non-viable DNA that is not protected by an intact cell membrane. The PMA treatment can be combined with PCR, 16S rRNA sequencing, and metagenomics to analyze the DNA from viable cells. However, the process may render subsequent analysis less quantitative.[123]

Multiomic and functional analysis

Clinical decision-making can benefit not only from the composition of microbiota but also from their functions and interactions with the host. Metagenomics describes what can be accomplished by microbes, while other omics analysis techniques, particularly metatranscriptomics, metaproteomics, and metabolomics, reveal their actual activities at different levels.[124] Even within the same strains, significant phenotypic heterogeneity exists and evolves over time.[125] Metatranscriptomics analyzes the RNA content of a microbial community, offering insights into the gene expression activities within the microbial community. Metaproteomics further analyzes protein expression in the microbial community. Metaproteomics has advantages over metatranscriptomics due to the greater stability of proteins compared to RNA, particularly those from prokaryotes, making it a valuable method for deciphering biological functions.[126] Furthermore, metabolomics measures a wide range of small molecules, including sugars, amino acids, lipids, and other organic molecules, directly examining the activity of the microbiota and their influence on the host. Metabolomics typically employs nuclear magnetic resonance (NMR) and mass spectrometry. Effective sample collection, storage, and preparation pose some of the challenges in metabolomics, as some metabolites are susceptible to degradation.[127] Distinguishing between host-origin and microbiota-origin metabolites presents another significant challenge.

In addition to genomic and proteomic assays, cell-based assays can be applied to evaluate the functions of microbiota with the resolution of single cells. This ability can reveal the phenotypic heterogeneity and resolve the group of microorganisms responsible for the phenotypic change, which may benefit the optimization of therapy or microbiota manipulation procedures. In the case of FISH, molecular probes quantify gene expression with high resolution, without the need for cell isolation and lysis.[128] Sequential FISH and barcode strategies can be applied to evaluate the numbers of genes.[68,69] However, counting the number of spots may underestimate the expression

level due to overlapping mRNA in the field. Integrating fluorescence intensity provides a more accurate quantification of the expression level. In addition, imaging mass spectroscopy can resolve locations of biomolecules within cells.[111] However, current studies are mostly performed using a small number of cells. Further development will be required to resolve functional activities in complex microbiota samples.

Outline

In the preceding sections, we explored the potential of emerging technologies for characterizing microbiota and their role in facilitating microbiota-associated clinical management strategies. Our focus was on technologies exhibiting promise for swift microbiota analysis in medical settings. After delving into essential features of microbiota analysis systems, such as implementation efficiency, absolute abundance measurement, low biomass detection, spatial distribution mapping, discrimination between live and dead cells, and functional analysis, we identified the crucial need for developing effective single-cell measurement methods in practical applications.

In **Chapter 2**, we propose a multimodal biosensor design tool, specifically tailored for single-cell microbiota profiling. The simplicity of construction and cost-effectiveness associated with these tools are anticipated to make microbial community profiling widely accessible and efficient for diverse scientific research purposes. Leveraging multimodal biosensors, we can now detect microbial populations and assess functionality at the single-cell level, opening up new possibilities for comprehensive analysis in scientific research. In **Chapter 3**, we focus on microbial community profiling via multimodal biosensor on clinical samples and in **Chapter 4** , we will make a conclusion and discuss future work.

Chapter 2 Multimodal Biosensor Design

Multimodal Biosensor Design Workflow

In this chapter, we will talk about the multimodal biosensor design workflow, as shown in Figure 2-1. This report will comprehensively cover the database, bioinformatic algorithms, biosensors binding kinetics, and detailed applications of the multimodal biosensor. The design workflow showcased for this biosensor holds versatile potential across various applications, including but not limited to 16S rRNA identifications and mRNA quantifications at the single-cell level. The discussions will delve into essential aspects pertaining to these applications.

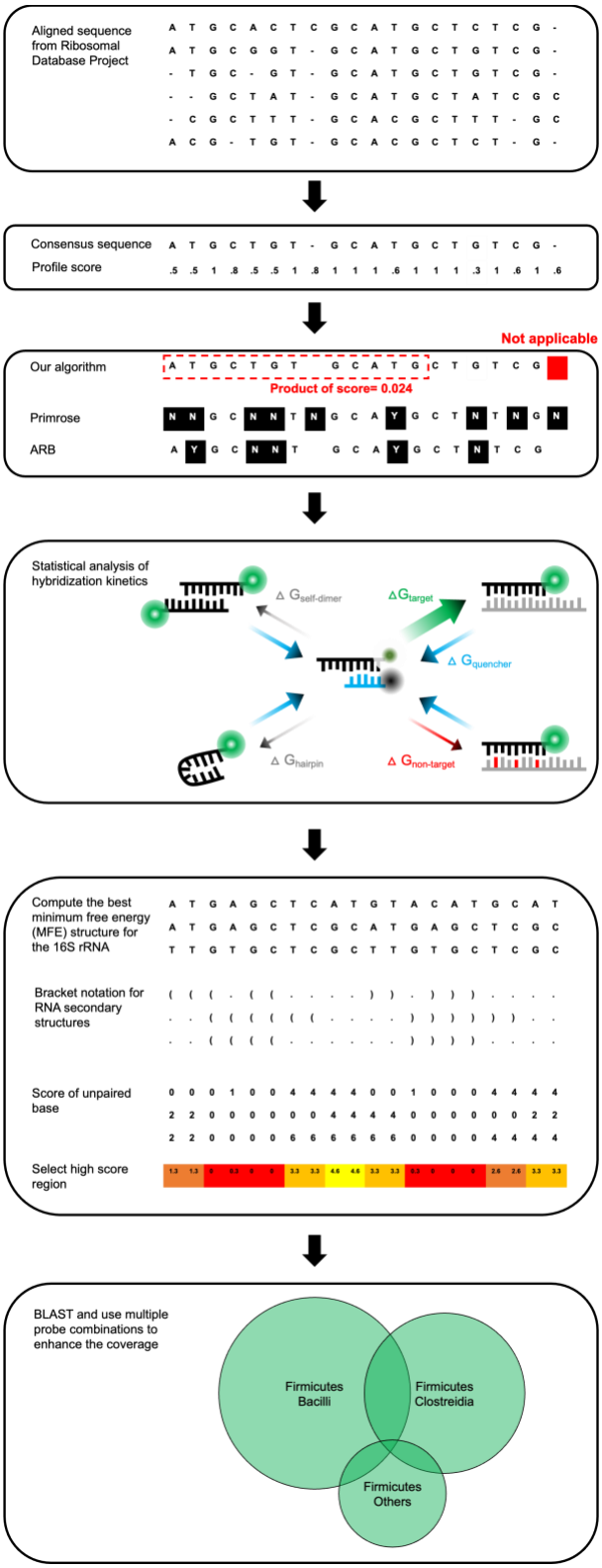


Figure 2-1 Multimodal biosensor design workflow

Consensus sequence of target taxon or gene

In our study, we focused on determining the consensus sequence of the target taxon or gene using specific algorithms as shown in Figure 2-2. Initially, we retrieved the 16S rRNA genes from the RDP database[129] and mRNA genes from Kyoto Encyclopedia of Genes and Genomes (KEGG)[130]. To create probes for detecting the murine microbiota in fecal sample, our approach involves referencing the Murine Microbiome Database[131] alongside our sequence data. This allows us to identify and select appropriate species or genera from the database for probe design. The 16S rRNA genes obtained from the RDP website were aligned upon download, while for mRNA genes sourced from KEGG, we utilized Clustal Omega at EMBL-EBI[132] via web services to align the sequences. Subsequently, we derived a consensus string by identifying the most prevalent nucleotides in each column of the aligned sequence matrix. Each column was assigned a profile score based on similarity, and the consensus string's probability was determined by multiplying the probabilities of individual nucleotides. Additionally, we computed the probabilities of potential donor candidates by multiplying the profile scores of specific lengths, enabling the identification of high-sensitivity candidates by setting probability thresholds.

It's crucial to highlight the difference in approach from methods like Primrose[133] and ARB[134], where the use of degenerate bases to generate consensus sequences often compromises the specificity of probes and might miss out on high-sensitivity candidates. In contrast to ARB, which typically removes gaps when they are most prevalent in a column, our gap threshold assigns positions as 'not applicable,' effectively bypassing these regions. To tackle this issue, we've introduced a threshold aimed at excluding gaps within the consensus sequence, restricting the usage to only the four bases for more precision.

This methodology ensures a nuanced approach to consensus sequence determination, emphasizing both sensitivity and specificity while strategically handling gaps in the sequence.

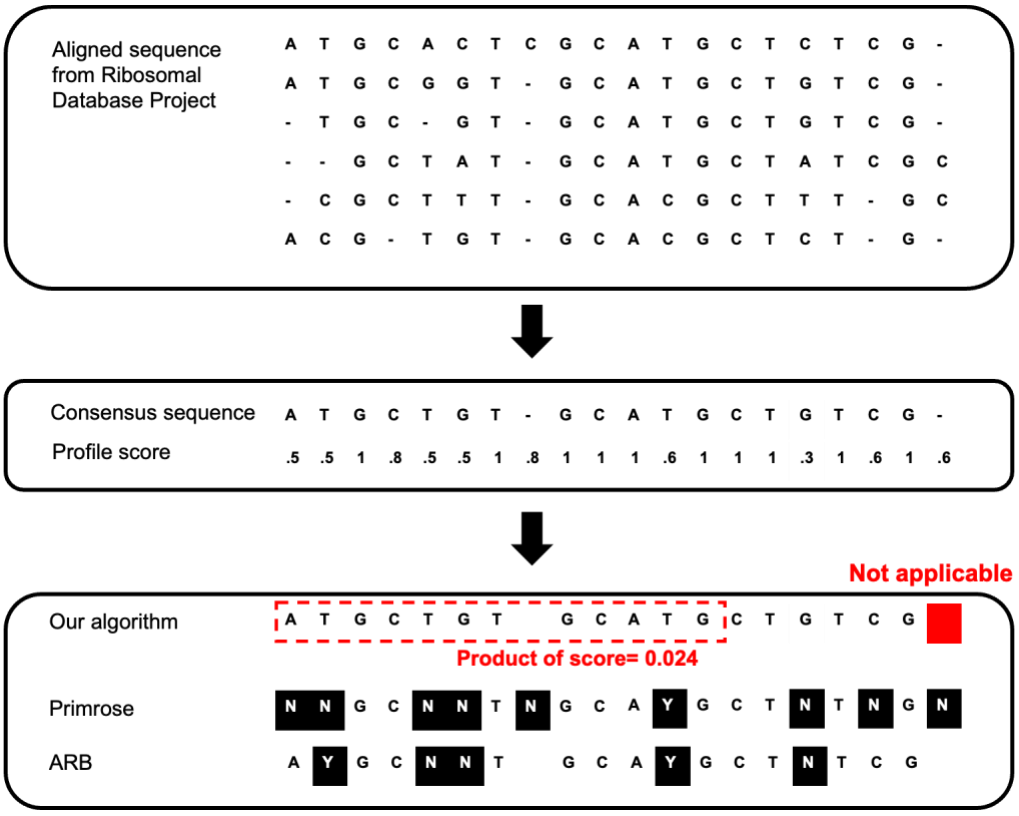


Figure 2-2 Consensus sequence of target taxon or gene. Starting with an aligned sequence from the Ribosomal Database Project, the process continues with the calculation of profile scores and the establishment of thresholds for selecting probe candidates.

Statistical analysis of binding energy

In the subsequent phase, we conduct a statistical analysis of the free energy change (ΔG) within both the target and non-target groups. To ensure probes exhibit exceptionally high sensitivities and specificities, adherence to the thermodynamic condition of $\Delta G_{target} < \Delta G_{quencher} < \Delta G_{non-target}$ is important (as illustrated in Figure 2-3). The robust hybridization of the target and 16S rRNA outperforms that of the quencher, leading the donor to bind to the 16S rRNA and emit light. Conversely, weaker hybridization of the non-target 16S rRNA compared to the quencher prompts the donor to bind to the quencher, diminishing the fluorescence signal. Utilizing functions within

the ViennaRNA package[135], we compute the minimum free energy among the donor, quencher, and target.

To statistically evaluate the sensitivity and specificity of each potential donor, we generate plots displaying the percentages of thermodynamically qualified species within both the target and non-target groups. Specifically, the percentage of species meeting the condition $\Delta G_{target} < \Delta G_{quencher}$ in the target group represents sensitivity, while the percentage adhering to $\Delta G_{quencher} < \Delta G_{non-target}$ in the non-target group signifies specificity. Hence, adjusting the quencher lengths allows for fine-tuning the probes' sensitivity and specificity. Notably, as the quencher length increases, sensitivity may decrease while specificity may augment.

By integrating the ViennaRNA package into the proposed R package, we aim to select donors exhibiting high sensitivities and specificities from a plethora of potential candidates. This selection process facilitates the determination of an appropriate quencher length for optimal probe performance. Furthermore, a comprehensive evaluation of the hairpin and self-dimer structures within the donor sequence is performed. The stability of these structures has the potential to significantly influence the binding efficiency of the biosensors.

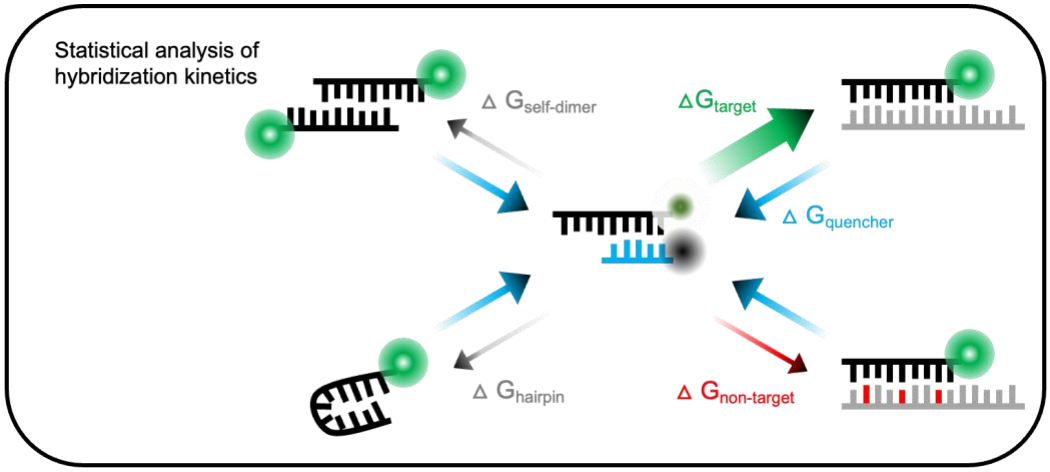


Figure 2-3 Statistical analysis of binding energy. Conducting a statistical analysis of the free energy change (ΔG) within both the target and non-target groups as well as self-dimer and hairpin structure.

Secondary structure prediction for biosensors selection

The secondary structure prediction for biosensors selection is shown in Figure 2-4. Following the statistical analysis of binding energy, numerous probe candidates are still viable, as depicted in Figure 2-5 (A). The Y-axis indicates the percentage of the target genus, *Lactobacillus*, showing lower binding energy, and all non-target bacteria (excluding those in the *Lactobacillus* genus but included in the mice stool sample database) displaying higher binding energy, relative to the quencher, in mice stool database. To pinpoint the most suitable probes, we delve into the secondary structures of our target. In the case of the 16S rRNA, the folding intricacies significantly impact its binding capability. However, assessing all targets individually, especially considering the potential multitude of species within a single phylum, is a formidable task. Therefore, we employ the ViennaRNA package[135] to derive bracket notations representing the minimum free energy (MFE) structure.

In quantifying the degree of freedom within specific target regions, we convert the bracket notations into scores based on continuity. For example, in instances illustrated in the figure, an open area containing four continuous bases would accrue a score of 4 for each of these bases. Subsequently, we calculate the average scores for all target regions and visualize them as a heat map, showcasing regions with the highest probability of continuous unpaired bases, as illustrated in Figure 2-5 (B). Moreover, we compute the entropy of each column, obtaining their averages, and present this data as a heat map Figure 2-5 (C). This visualization assists in enhancing the selection process by providing a clearer understanding of the entropy distribution for better-informed decision-making. This approach aids in the selection of the most optimal probes from the extensive pool of hundreds of candidates.

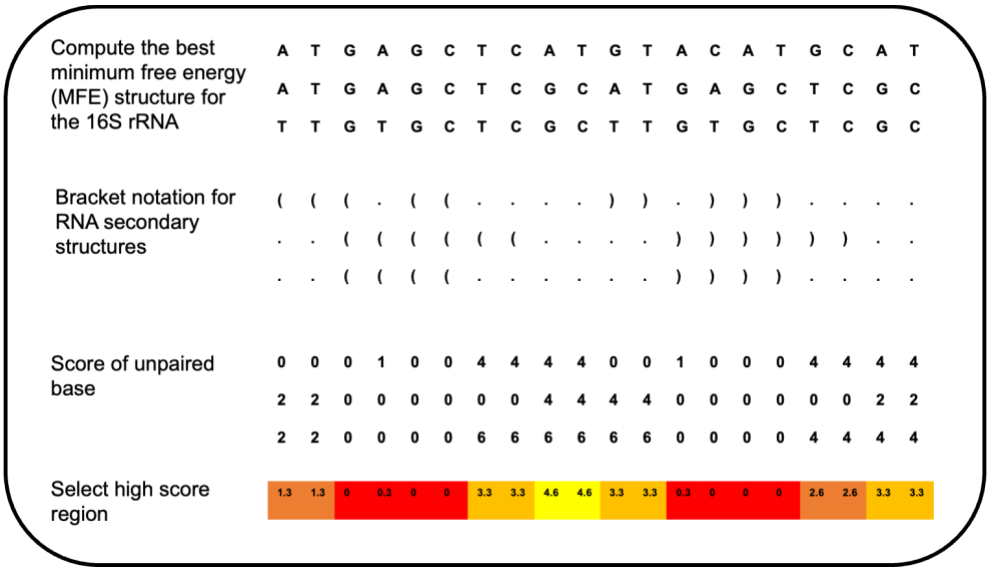


Figure 2-4 Secondary structure predictions. Employment of the ViennaRNA package135 to derive bracket notations representing the minimum free energy (MFE) structure.

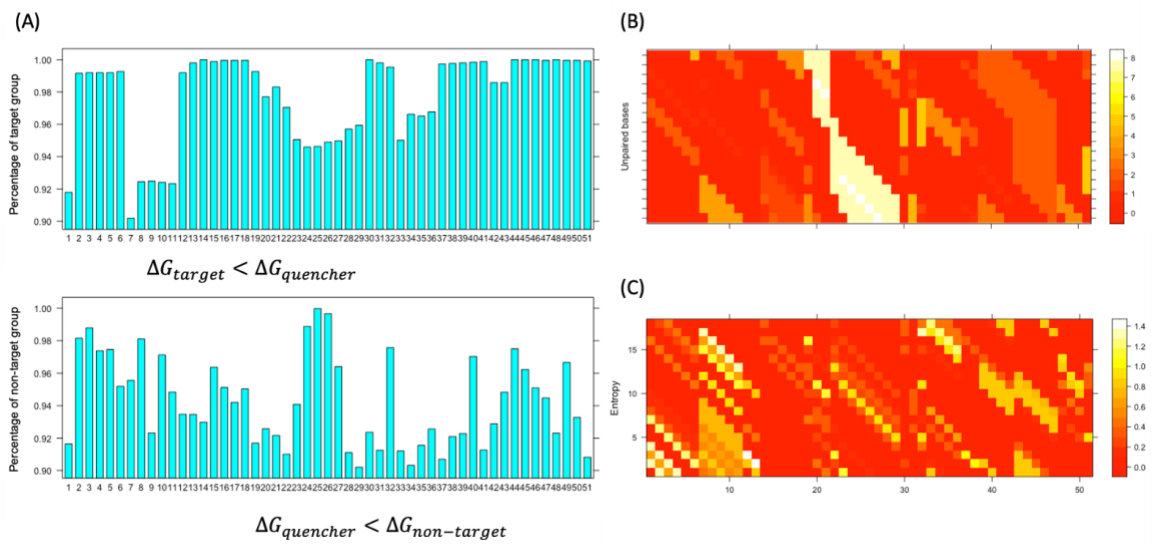


Figure 2-5 Example of Lactobacillus probe selection. The Y-axis indicates the percentage of the target genus, Lactobacillus, showing lower binding energy, and all non-target bacteria (excluding those in the Lactobacillus genus but included in the mice stool sample database) displaying higher binding energy, relative to the quencher, in mice stool database. The X-axis is the serial number of probe candidates.

Enhance the coverage by multiple probe combinations

Following the selection of probes exhibiting high sensitivity, specificity, and unpaired regions, we assess their efficacy using rBLAST. Generally, genus-level probes tend to achieve superior sensitivity and specificity. However, designing probes at the phylum level presents challenges as it involves encompassing a larger number of species while simultaneously excluding others. To address this complexity, we implement multiple probe combinations as a solution (Figure 2-6).

For instance, in the case of *Firmicutes* probes, we initially create two probes, Firm_1 targeting the *Bacilli* class and Firm_2 targeting the *Clostridia* class, which are the dominant classes in murine fecal samples. During the probe design phase, the absence of mutual exclusivity between the two probes is advantageous for the design process, as shown in Table 2. Subsequently, we design Firm_3 based on species not covered by Firm_1 and Firm_2. By employing these combinations, we achieve a coverage of 98.5% with high specificity. We apply a similar strategy to other predominant phyla in murine fecal samples, such as *Bacteroidota, Actinomycetota*, and *Proteobacteria*.

The probe combinations' BLAST results, as outlined in Table 3, demonstrate that these combinations significantly improve coverage to over 97% while maintaining less than 3% of nonspecificity. This approach effectively addresses the challenges associated with designing probes at the phylum level, enhancing both coverage and specificity in detecting the Murine Microbiota. The phylum probes are focused on the phylum level; there may be greater potential for nonspecificity if our focus were on more refined taxonomic levels. With probes targeting the genus level, we observe much lower nonspecificity (less than 1%).

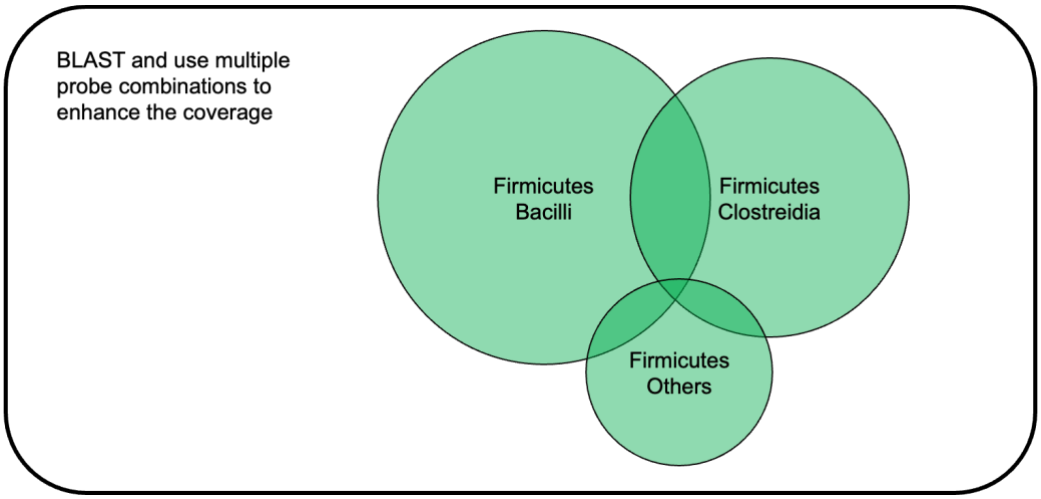


Figure 2-6 Enhance the coverage by multiple probe combinations

Taxa \ Probe	Bac_1	Bac_2	Firm_1	Firm_2	Firm_3	Actino_1	Actino_2	Proteo_1	Proteo_2
Bacteroidota	98.3%	98.9%	0.2%	1.0%	2.3%	0.2%	0.0%	0.6%	0.0%
Firmicutes Bacilli	0.1%	0.1%	97.3%	33.8%	3.2%	0.2%	0.0%	0.0%	0.0%
Firmicutes Clostridia	0.3%	0.2%	20.6%	82.3%	63.5%	0.3%	0.0%	0.0%	0.6%
Actinomycetota	0.2%	0.2%	0.2%	1.0%	0.2%	92.0%	22.9%	0.0%	0.0%
Proteobacteria Alphaproteobacteria	0.0%	0.0%	0.7%	0.0%	0.9%	0.4%	0.0%	98.8%	6.2%
Proteobacteria Gammaproteobacteria	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	18.3%	99.5%

Table 2 Coverage of individual probes

Taxa \ Probe	Bac_1 + Bac_2	Firm_1 + Firm_2 + Firm_3	Actino_1 + Actino_2	Proteo_1 + Proteo_2
Bacteroidota	99.2%	2.9%	0.2%	0.6%
Firmicutes	0.2%	98.5%	0.2%	0.3%
Actinomycetota	0.2%	1.2%	97.8%	0.0%
Proteobacteria	0.0%	0.9%	0.2%	99.4%

Table 3 Coverage of probe combinations

Incorporate the workflow into a single R package

To optimize biosensor design efficiency, we've consolidated the workflow into a unified R package. This package allows for easy modification of parameters such as the threshold, donor length, and quencher to enhance results. Integration with ViennaRNA and rBLAST packages ensures seamless functionality within the R Studio platform.

The most resource-intensive and time-consuming aspect involves utilizing the brute-force algorithm to search for the best-matched regions among probe candidates, targets, and non-targets, involving four layers of iterations. To expedite this process, we've implemented parallel computing on the R Studio server, a significant enhancement in computational speed. What originally took about a week to run on a personal computer now completes within a few hours on the servers. This adoption of parallel computing significantly reduces the computational load and expedites the identification of optimal probe regions, greatly streamlining the biosensor design process.

Multimodal Biosensor Design for Mouse Fecal Microbiota

Phylogenetic biosensors design for mouse fecal microbiota

Designing phylogenetic biosensors for mouse fecal samples involves targeting specific bacterial groups commonly found in stool. The selected phyla for the biosensors are *Bacteroidota, Firmicutes, Actinomycetota,* and *Proteobacteria*, as they are frequently observed in these samples. Additionally, we'll focus on *Lactobacillus* and *Bifidobacterium,* core genera often present in probiotic products, for probe development.

To assess the detection limits of the system, the chosen target species for evaluation is *Enterococcus faecalis*. This species is relatively uncommon in our mouse fecal samples, making it an ideal candidate to evaluate limit of detection of the biosensor system. *Enterococcus faecalis* holds significance in various contexts due to its versatile nature. As a commensal bacterium, it naturally resides in the gastrointestinal tract of humans and animals, contributing to the complex microbial community within the gut[136]. Despite being a part of the normal gut microbiota, *Enterococcus faecalis* can pose health concerns as it possesses traits of opportunistic pathogenicity, leading to infections, particularly in hospital settings[137,138].

Genus-level and species-level probes commonly display heightened discriminatory power, offering superior sensitivity and specificity in target identification. Conversely, the design of probes at the phylum level presents inherent challenges due to the vast taxonomic diversity it encompasses, necessitating the inclusion of numerous species while excluding others. To surmount these challenges, our methodology employs a strategy involving the creation of multiple probe combinations described in **Chapter 2**. This approach allows for a comprehensive coverage of the diverse range of species within the phylum while concurrently ensuring precise and selective targeting, thus optimizing the specificity and sensitivity of the detection system.

Gene biosensors design for single cell gene expression

The inhibitory neurotransmitter role of gamma-aminobutyric acid (GABA) in the context of the nervous system is fundamental. In the central nervous system of animals, GABA acts as the primary inhibitory neurotransmitter, regulating neuronal excitability by binding to GABA receptors on neurons. When GABA binds to these receptors, it triggers an inhibitory response, essentially calming the activity of neurons. This modulation of neuronal activity helps maintain a balance between excitation and inhibition, crucial for normal brain function.

In the realm of bacteria, while GABA functions primarily in metabolic and stress-related pathways, its role as an inhibitory neurotransmitter akin to its function in the nervous system of animals is not precisely replicated. However, the mechanisms involving GABA in bacterial cells contribute to regulating cellular responses to stressors and maintaining internal homeostasis. Within bacteria, GABA can act as a signaling molecule to regulate various cellular processes. It can facilitate responses to external stresses such as acid stress or oxidative stress, aiding in bacterial survival under adverse conditions. By modulating GABA levels, bacteria can adapt to changing environments, adjust their metabolic activity, and potentially resist external challenges.

Understanding the role of GABA and the pathways involved in its synthesis, including the gadB gene, provides insights into how bacteria manage stress responses and adapt to different environmental conditions. Manipulating GABA metabolism in bacterial systems might have implications beyond bacterial physiology, potentially influencing stress tolerance, probiotic development, or even strategies for controlling microbial populations.

Exploring the parallels and divergences between GABA's functions as an inhibitory neurotransmitter in animals and its roles in bacterial physiology remains an area of active research. This exploration not only deepens our understanding of microbial biology but also unveils potential

applications in biotechnology, where harnessing these mechanisms could lead to innovative solutions in various fields. Initially, we retrieved the gadB genes specific to the *Lactobacillus* genus from the KEGG. Subsequently, we employed Clustal Omega, accessible through EMBL-EBI's web services, to align these sequences. Following the alignment, we utilized our R package to develop probes following a standardized protocol. In this process, the 16S rRNA sequences served as the non-target group for the probe design, ensuring specificity to the gadB genes of *Lactobacillus*. The phylogenetic tree analysis of gadB genes of *Lactobacillus* are shown in Figure 2-7.

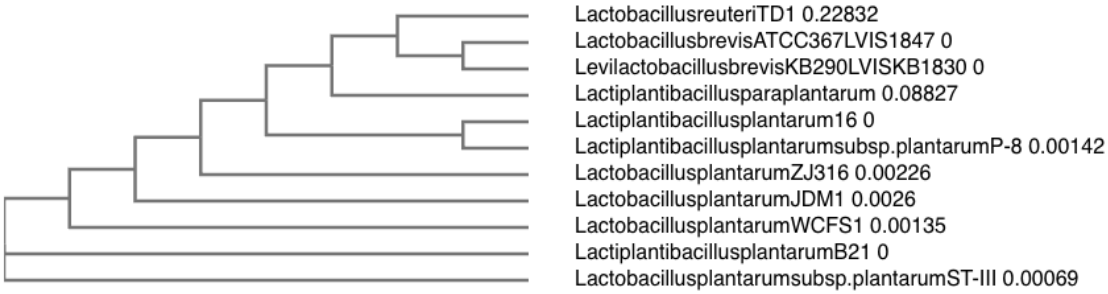


Figure 2-7 Phylogenetic tree analysis of gadB genes of Lactobacillus.

Chapter 3 Microbial Community Profiling via Multimodal Biosensor

Within this chapter, our aim is to introduce and elaborate on the application of phylogenic biosensors in studying the fecal microbiome, alongside an exploration of biosensor hybridization protocols. We'll thoroughly discuss the materials used, the intricacies of the sample preparation process, and the study design (Figure 3-1).

Furthermore, we will integrate the use of Live/Dead kits to discern and assess the viability of distinct taxa. A critical aspect of our exploration will be studying the limits of detection inherent within these methods, along with detailing the flow of image processing integral to this analytical process. We will also introduce the mRNA biosensor, which enables the evaluation of gene expression levels at the single-cell level. This innovation promises to significantly enhance our ability to understand and analyze the dynamic gene expression patterns within the microbiome.

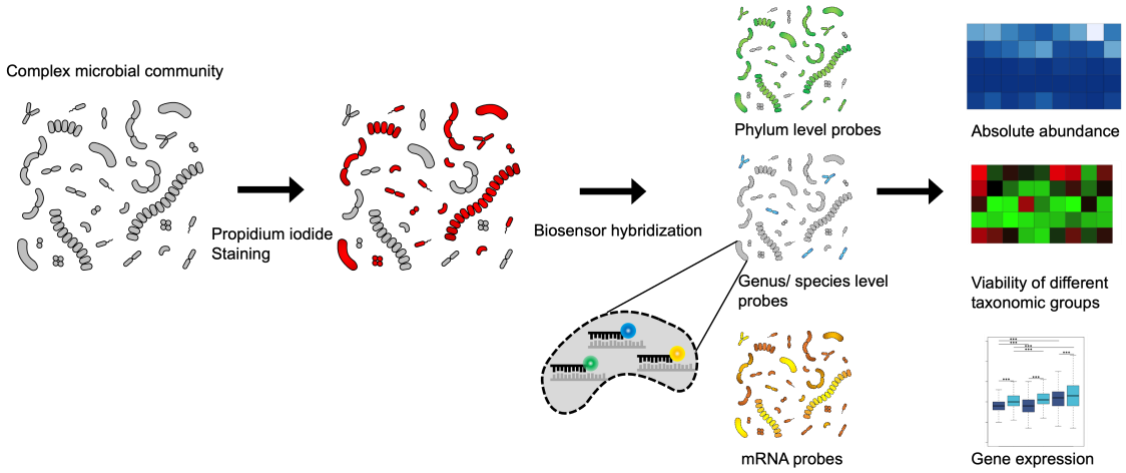


Figure 3-1 Microbial Profiling via Multimodal Biosensor

Materials and Methods

Fecal sample collection and bacteria isolation

We gather 9 mice fecal samples in a dual collection on the same day – one set for sequencing and the other for our biosensor profiling. Following immediate collection, we implement the fecal sample processing protocol.

1. Dilute the fecal sample with PBS at a ratio of 1:10.
2. Agitate the diluted sample using a vortex mixer then follow a sonicating water bath for 10-15 minutes to ensure thorough mixing.
3. Centrifuge the mixture at a low speed of 200 x g for 2 minutes to separate larger particles and debris.
4. Carefully collect the supernatant obtained after centrifugation and subject it to a higher centrifugal force of 14,000 rpm for 5-10 minutes to pellet smaller particles.
5. Resuspend debris in PBS by agitation using a vortex mixer then follow a sonicating water bath for 10-15 minutes to ensure thorough mixing.
6. Centrifuge the debris-PBS mixture at 200 x g for 2 minutes to separate bacteria
7. Collect the supernatant obtained and subject it to a centrifugal force of 14,000 rpm for 5-10 minutes to pellet any remaining solid materials.
8. Wash the resulting pellet twice using PBS to ensure cleanliness and remove impurities.

This protocol aims to isolate supernatant containing target components while removing debris and impurities through centrifugation and washing steps. Adjustments to timing, centrifuge speeds, and wash repetitions may be made based on specific research requirements or sample characteristics.

Viability staining using Live/Dead kit

To obtain information on single-cell viability, we utilized the LIVE/DEAD BacLight Bacterial Viability Kit (Cat. No. L7012). This specific kit serves the purpose of evaluating the viability of bacterial populations based on their cellular membrane integrity. Cells with compromised membranes, indicative of being dead or dying, exhibit a red stain, while cells with intact membranes exhibit a green stain. Our choice to utilize the red dye (propidium iodide) solely for staining dead bacteria is aiming to conserve spectral resources for phylogenetic biosensors. Considering the permeable step of our hybridization protocol, we perform the viability staining beforehand. Upon bacterial isolation, we conduct two washes with PBS and subsequently resuspend the bacteria in 1 ml of PBS. Following this, we add 1.5 µL of the dye mixture to each milliliter of the bacterial suspension. The suspension is then incubated at room temperature in darkness for 15 minutes. Subsequently, the bacteria undergo a modified fluorescence in situ hybridization (FISH) protocol to simultaneously gather information regarding both viability and phylogenetics.

Modified fluorescence in situ hybridization (FISH) protocol

Fluorescence in situ hybridization (FISH) is a powerful molecular biology technique used to detect and localize specific DNA or RNA sequences within cells or tissues. It utilizes fluorescently labeled probes that hybridize to complementary target sequences, allowing for the visualization and identification of these sequences under a fluorescence microscope.

In the context of bacteria, FISH can be particularly useful for studying microbial communities, identifying specific bacterial species, and examining their spatial distribution within environmental samples or host tissues. However, bacteria possess unique cell structures and

membranes that can make it challenging to access their genetic material for FISH analysis. Bacterial cell membranes contain rigid cell walls and other protective layers that can impede the penetration of FISH probes. To overcome this barrier and facilitate probe entry, various methods are employed to effectively lyse or permeabilize the bacterial cell wall. Enzymatic and/or heat lysis methods are commonly used to disrupt the tough bacterial membrane and allow the FISH probes to access the intracellular DNA or RNA.

Enzymatic lysis involves the use of enzymes, such as lysozyme, to degrade components of the bacterial cell wall. Lysozyme targets the peptidoglycan layer of the cell wall, while proteinase K breaks down proteins that contribute to the structural integrity of the membrane. These enzymes weaken the cell wall, making it more permeable to the FISH probes. Heat lysis, on the other hand, involves subjecting the bacterial cells to elevated temperatures, typically through heat shock or boiling, to disrupt the cell membrane. This process helps to soften the cell wall and increase its permeability, allowing the FISH probes to enter and bind to the target genetic material. After enzymatic or heat lysis, the fluorescently labeled probes are introduced to the sample and allowed to hybridize specifically with the complementary sequences within the bacterial cells. Unbound probes are washed away, and the samples are visualized using fluorescence microscopy. The fluorescence signals emitted by the bound probes indicate the presence and location of the targeted DNA or RNA sequences within the bacterial cells.

Overall, the combination of FISH with enzymatic or heat lysis methods enables the visualization and study of specific genetic material within bacterial cells, contributing to a better understanding of microbial communities, pathogen identification, and various other applications in microbiology and environmental sciences. Here is the optimized protocol for conducting FISH within bacterial cells.

1. Preparation of Hybridization Buffer in Water:

- NaCl: 0.9 M
- Tris-HCl (pH 7.6): 0.02 M
- Formamide: 20%
- Dextran Sulphate: 10%
- SDS: 0.01%
- CaCl2: 0.45 mM

2. Preparation of Probe:
    - Combine 1 µl of each probe with 99 µl of hybridization buffer to make 100 µl of the probe solution.
3. Bacterial Cell Preparation:
    - Harvest 50 µl of bacteria culture at 10^9 CFU/ml for each FISH test.
    - Wash the cells twice with 1X PBS.
4. Fixation and Cell Preparation:
    - Fix the bacteria cells by adding 150 µl of 4% PFA and incubate for 40 minutes at 4 °C.
    - Remove excess fixative (3% PFA) and wash the cells twice with 1X PBS.
5. Cell Treatment for FISH:
    - Heat the bacteria cells at 80°C for 10 minutes.
    - Immediately place the cells on ice for 1 minute.
    - Permeabilize the cells by treating them with 50 µl of a mixture containing 20 mg/ml lysozyme + 1 mg/ml lysostaphin for 15 minutes at 37°C. (For gram-positive target)
    - Wash the cells thoroughly with ice-cold 1X PBS.
6. Hybridization and Washing Steps:

- Resuspend the cells and incubate them with 50 µl of hybridization buffer containing 1 µm probe at 46 °C for 20 minutes.
- Wash the cells twice with 50 µl of washing buffer 1 (0.4x SSC) to eliminate excess probes, incubating at 46 °C for 3.5 minutes each time to reduce nonspecific binding.
- Perform additional washes with 50 µl of washing buffer 2 (2x SSC/0.05% Tween) twice.

7. Imaging:
    - Acquire images using a fluorescent microscope with a fluorescent exposure time of 1 second and a bright field image.

This optimized protocol for FISH in bacterial cells involves multiple steps for cell preparation, fixation, permeabilization, probe hybridization, and thorough washing to ensure specific and accurate visualization of targeted RNA sequences within the bacterial cells. Adjustments should be made based on specific experimental requirements and bacterial species characteristics (enzyme lysis for gram-positive species).

Image analysis protocol

Following viability staining and modified fluorescence in situ hybridization, dispense 1 µl of the prepared solution onto a glass slide and capture images using a fluorescent microscope set with a 1-second exposure time for fluorescent images and a bright-field image. All procedural steps are automated using the ImageJ Macro. Initially, the image is converted to 8-bits and the background is subtracted. Subsequently, a threshold is established and the image is converted to a mask to create a binary representation. Employing the watershed method, each object's center is determined via a morphological erode operation. A distance map is calculated from these object

centers to the object edges, simulating a "topological map" filled with simulated water. At junctures where two "Watersheds" meet, a dam is constructed to separate them. The watershed function effectively separates touching cells. Following this, particle analysis is executed using automatic particle counting. To eliminate small background artifacts generated by the threshold, a minimum size filter is applied. Regions of interest (ROIs) are then obtained and used to measure the mean fluorescent intensity within them, yielding results. Subsequently, an intensity threshold is set, and the data is plotted using R for further analysis and visualization. The process is shown in Figure 3-2.

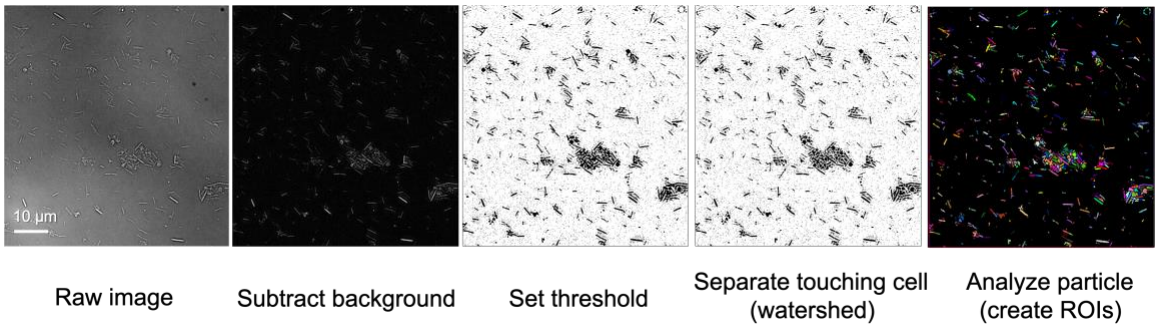


Figure 3-2 Image processing of bright field image.

Sampling from large population of bacteria in stool

Determining the required sample size from a large population of bacteria in stool to attain statistical significance is a pertinent academic query. In accordance with Dillman's formula[139], accommodations are applied to account for finite population size within the confidence interval formula. In this formula, n_p represents the total count of the larger population, n_s stands for the sample size, p denotes the proportion of the targeted group for detection, ε is the margin of sampling error, and $Z_{\alpha/2}$ refers to the z-score corresponding to the desired confidence level.

$$n_s = \frac{n_p p(1-p)}{(n_p - 1)\left(\frac{\varepsilon}{Z\alpha_{/2}}\right)^2 + p(1-p)}$$

In order to estimate the necessary sample size, several assumptions must be established. For a 95% confidence level, $Z\alpha_{/2}$ equals 1.96. The estimate number of the bacteria in healthy mouse feces is near 10^9 CFU per gram of stool. Considering an experiment with a total stool weight of 0.1 grams, n_p would equate to 10^8. We sampled 50,000 bacteria through the capture of 10 images, each comprising 5,000 bacteria. The table below illustrates the margin of sampling error achieved by setting different target proportions (Table 4). We can get different but acceptable sampling error whin dynamic range, subject to specific constraints and statistical parameters.

Target proportion	50%	10%	5%	1%	0.5%	0.1%	0.05%	0.01%
Sampling error	0.44%	0.26%	0.19%	0.087%	0.061%	0.028%	0.019%	0.0087%

Table 4 Target proportion and sampling error.

16S rRNA sequencing

To assess our sample's microbial composition, we engaged CD Genomics for 16S rRNA sequencing, a widely recognized method for profiling microbial communities. This approach involves amplifying prokaryotic 16S rDNA hypervariable regions via PCR and subsequently sequencing the amplicons on a high-throughput platform. Analysis of these distinct regions allows for determining the relative abundance of various taxa within a community and comparing

taxonomic profiles among different groups, aiding in understanding changes in microbial profiles over time or between treatments.

Traditionally, absolute quantification of specific microorganisms in environmental samples involved qPCR, yet this method often yielded unstable results and necessitated the design of specific primers. Presently, standard 16S rRNA sequencing provides relative abundance data by calculating the ratio of sequence variants to the total number of sequences. However, addressing the absolute quantification issue has garnered considerable attention.

For this purpose, we submitted a stool sample over 0.1g to CD Genomics, shipping it in a fresh state with dry ice. They isolated the DNA and performed Accu16STM (Accurate 16S Absolute Quantification Sequencing). The process began with DNA extraction using the FastDNA® SPIN Kit for Soil (MP Biomedicals, Santa Ana, CA) as per the manufacturer's instructions. Genomic DNA integrity, concentration, and purity were assessed via agarose gel electrophoresis, Nanodrop 2000, and Qubit3.0 Spectrophotometer.

To enable accurate quantification, synthetic spike-ins were created with conserved regions mimicking natural 16S rRNA genes but featuring randomly sequenced variable regions (~40% GC content). These spike-ins, with known gradient copy numbers, were mixed in appropriate proportions with the sample DNA. Subsequently, the V3-V4 hypervariable regions of the 16S rRNA gene and spike-ins were amplified using primers 341F (5'-CCTACGGGNGGCWGCAG-3') and 805R (5'-GACTACHVGGGTATCTAATCC-3') and sequenced using an Illumina NovaSeq 6000 sequencer.

In data processing, QIIME2 was employed. The cutadapt plugin removed adaptor and primer sequences, followed by DADA2 for quality control and amplicon sequence variant (ASV) identification. Taxonomic assignments were made with a pre-trained Naive Bayes classifier based on Greengenes (version 13.8) at a confidence threshold of 0.8. Subsequently, spike-in sequences

were identified, read counts recorded, and standard curves generated for each sample, correlating read counts to spike-in copy numbers. This facilitated calculating the absolute copy number of each ASV in the samples.

Given that spike-in sequences were not part of the sample's natural flora, they were removed in subsequent analyses. ASVs, differing from OTUs by removing repetitive sequences without performing similarity clustering, were generated post quality control using the DADA2 plugin. Taxonomic annotations of ASVs were achieved by comparing representative sequences with a database via QIIME2 software, using different annotation methods at a confidence threshold of 0.8 (16S confidence level). The resulting taxonomic annotations were obtained through various annotation methods: classify-sklearn, classify-consensus-blast, classify-consensus-vsearch, and default classify-sklearn.

Results and Discussions

Proof of concepts with culturable strain

To validate the concept with culturable strains, we tested *Bacteroides fragilis*, *Lactobacillus plantarum*, *Bifidobacterium longum*, and *Escherichia coli*, all stored at -80°C in a solution comprising 25% glycerol and 75% culture medium. The anaerobic strains (*Bacteroides fragilis*, *Lactobacillus plantarum*, and *Bifidobacterium longum*) were cultivated using BD GasPak™ EZ Gas Generating Systems and EZ Anaerobe Container System Sachets, taking several days to reach sufficient growth concentrations. *Bacteroides fragilis* was cultured in *Bacteroides* Phage Recovery Medium (BPRM), while *Lactobacillus plantarum* and *Bifidobacterium longum* were cultured in De Man–Rogosa–Sharpe (MRS) media supplemented with 0.1% cysteine to

tolerate oxygen exposure. These bacteria were incubated at 37°C without agitation. *Escherichia coli* was cultured in Mueller Hinton Broth 2 at 37°C with shaking. Subsequently, phylogenetic biosensors were applied to the mouse fecal microbiota using modified fluorescence in situ hybridization. The resulting BLAST analysis is presented in Table 5, while the phylogenetic biosensors' transformation efficiency is illustrated in the Figure 3-3. The under fluorescent microscope are shown in Figure 3-4.

	Bacteroidota	*Firmicutes*	*Lactobacillus*	*Actinomycetota*	*Bifidobacterium*	*Proteobacteria*
donor	CGACACCTCA CGGCACGAGC	CGTAGTTAGC CGTGACTTTC	GCAGGTTCGC TTCTCGTTGT	CGGCCATTGT AGCATGCGTG	ACCGTTAAGC GATGGACTTT	GCCCAGTAAT TCCGATTAAC
quencher	GCTCGTGCCG TGAGG	GAAAGTCACG GCTAA	ACAACGAGA AGCGAA	CACGCATGCT ACAAT	AAAGTCCATC GCTTA	GTTAATCGGA ATTAC
Bacteroides fragilis	GCTCGTGCCG TGAGGTGTCG	N	NA	NA	NA	NA
Lactobacillus plantarum	NA	GAAAGCCACG GCTAACTACG	ACAACGAGTT GCGAACTCGC	NA	NA	NA
Bifidobacterium longum	NA	NA	NA	CACGCATGCT ACAATGGCCG	AAAGTCCATC GCTTAACGGT	NA
Escherichia coli	NA	NA	NA	NA	NA	GTTAATCGGA ATTACTGGGC

Table 5 BLAST result of culturable strains

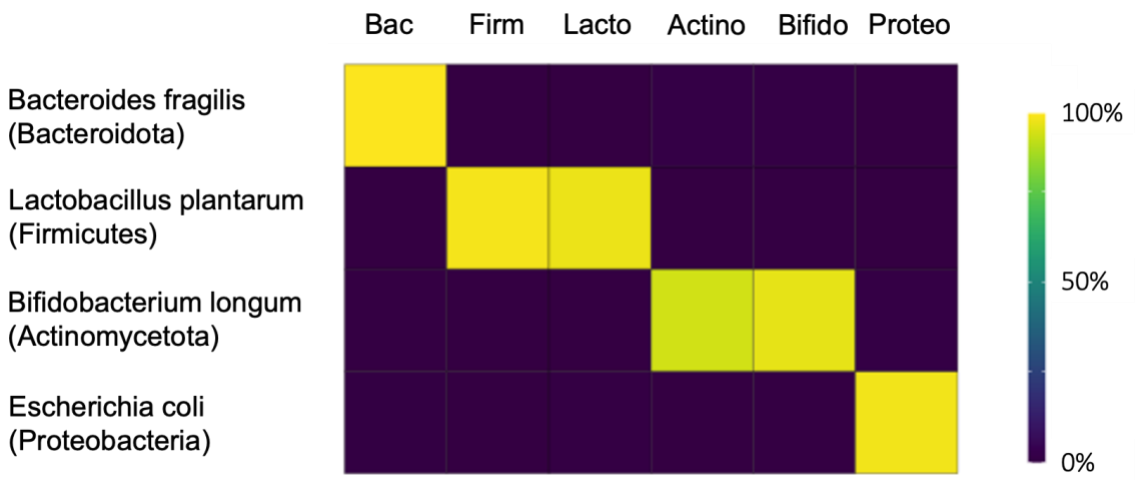


Figure 3-3 Phylogenetic biosensors' transformation efficiency

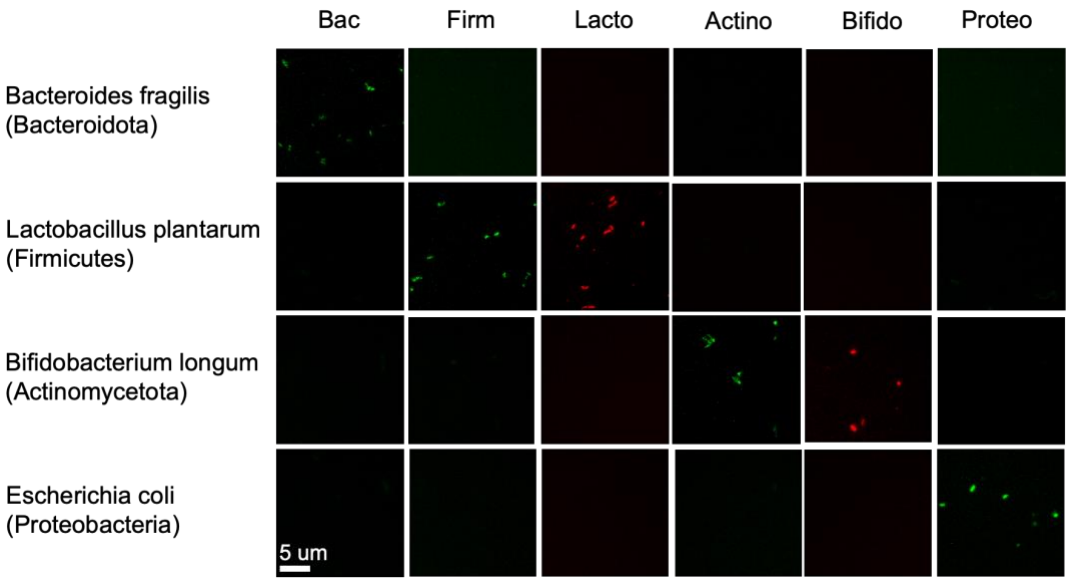


Figure 3-4 The bacteria after phylogenetic biosensor hybridization under fluorescent microscope

Limit of detection

We introduced *Enterococcus faecalis* into mouse stool samples and utilized *Enterococcus faecalis* probes to assess the system's limit of detection. *Enterococcus faecalis* was chosen as the

target due to its absence in certain samples, and the probes exhibited high specificity according to the BLAST results. Cultured *Enterococcus faecalis* was added, and ImageJ was employed to quantify cell concentrations. Through serial dilutions, we obtained various target proportions: 50%, 10%, 5%, 1%, 0.5%, 0.1%, 0.05%, and 0.01%. Subsequently, we conducted species probe hybridization and employed specific workflows to analyze the resultant images. Each data point was derived from 50,000 bacteria samples, captured across 10 images, each containing 5,000 bacteria. The experiments were replicated three times (n=3). The ratios of spiked samples and the measured target proportions using multimodal biosensors are detailed in the corresponding results (Table 6 and Figure 3-5). Following the limit of detection formula[140], the calculated limit of detection (LoD) is calculated as 0.0075%.

$$LoD = LoB + 1.645\ SD_{low\ concentration}$$

In this context, LoB denotes the limit of blank (which is zero in our case), and SD represents the selected standard deviation of 0.01%. The R-squared value is 0.9997, indicating a high degree of correlation. Additionally, the coefficient is 0.9369, closely aligning with the transformation efficiency. The observed errors in the results might originate from various factors, such as sampling errors discussed in the preceding paragraph, transformation efficiency, and/or probe coverages. When the calculated LOD is lower than the lowest amount actually measured in the tests, this might suggest that the assay has the theoretical capability to detect smaller quantities than those we have experimentally confirmed. Consequently, we evaluated the lower concentration level of 0.0075%, obtaining an average measurement of 0.0126% with a standard deviation of 0.0026%. This outcome is clearly distinguishable from the blank sample, demonstrating the substance's detectability at this concentration level.

Spiked sample ratio	100%	50%	10%	5%	1%	0.5%	0.1%	0.05%	0.01%	0.0075%	0%
Measurement by biosensor	94.34%	45.3%	9.23%	4.88%	1%	0.548%	0.07%	0.041%	0.013%	0.0126%	0%
Standard deviation	3.62%	3.51%	0.655%	0.299%	0.167%	0.112%	0.0196%	0.0102%	0.00461%	0.0026%	0%

Table 6 Explore limit of detection by spike cultured bacteria into mice stool sample

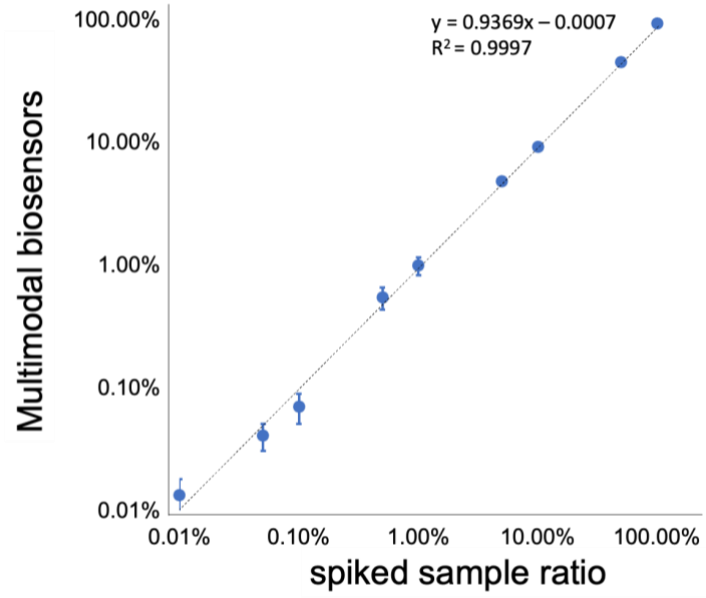


Figure 3-5 The relationship between target proportions and detection capability

Relative abundance and absolute abundance

In our taxonomic analysis, we concentrated on *Proteobacteria*, *Bacteroidota*, *Actinomycetota*, and *Firmicutes*, significant phyla prevalent in murine feces with crucial physiological roles. Furthermore, we specifically targeted *Lactobacillus* and *Bifidobacterium*, core genera commonly found in probiotic products, alongside *Enterococcus faecalis* due to its absence

or low abundance in select samples, enabling sensitivity testing of the biosensors. To validate the microbial composition of our samples, we collaborated with CD Genomics for 16S rRNA sequencing analysis. Using a biosensor platform, we characterized the microbial community profiles within mice fecal samples, displaying the abundance of bacteria at the phylum, genus, species level via a heatmap (Figure 3-6).

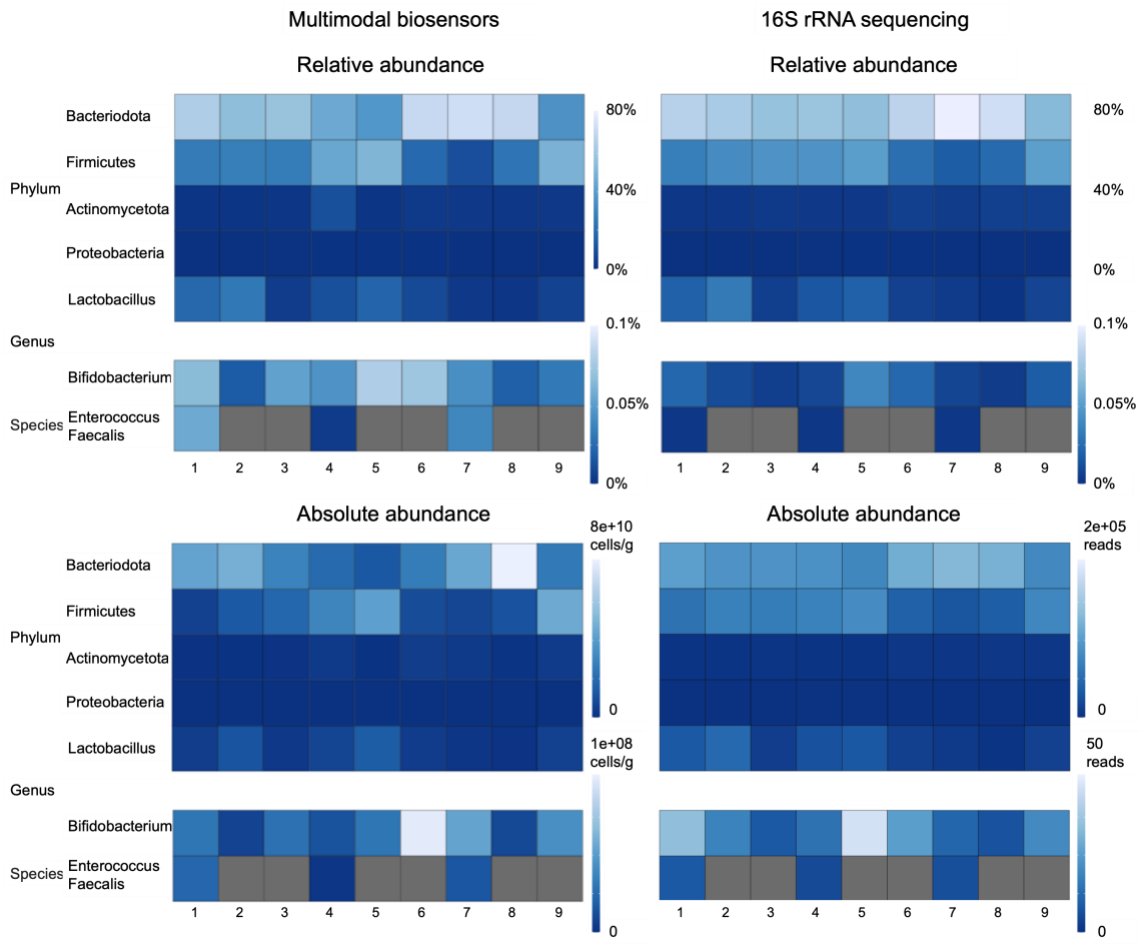


Figure 3-6 Heatmaps of the microbial community composition.

The findings exhibit strong concordance with the data obtained from 16S rRNA sequencing. Notably, the biosensor platform effectively detected a significant increase in *Bacteroidota* and a

substantial decrease in *Firmicutes* at the phylum level in samples 6, 7, and 8. At the genus level, distinct patterns emerged for *Lactobacillus* and *Bifidobacterium*, with both exhibiting notably high percentages (>40%) and very low percentages (<0.1%). Additionally, the presence or absence of *Enterococcus Faecalis* could be determined, even with its minimal abundance at the species level.

In terms of absolute abundance, as our sampling is based on actual cell numbers, it offers a more meaningful comparison of the actual bacterial quantities between samples. The absolute abundance particularly valuable when comparing different individuals. For instance, while samples 6, 7, and 8 display similar high relative abundances of *Bacteroidota*, an examination of absolute abundance reveals that sample 8 contains double and triple the amount of *Bacteroidota* per gram of stool compared to samples 6 and 7. This information was not captured by relative abundance analysis and absolute read counts in 16S rRNA sequencing, and may potentially provide useful insights into the microbial composition across samples.

Viability of different taxonomic group

One of the key advantages of our biosensors lies in their ability to assess the viability of various taxonomic groups, a capability not achievable through sequencing alone. To gain a comprehensive understanding, it is imperative to obtain simultaneous information regarding the identity and activity of microbial cells. Live bacteria actively contribute to biomass production, while inactive bacteria, although not participating in production, may still serve functions within the gastrointestinal tract, as evidenced by studies involving inactive probiotic bacteria.

Defining life and death in microbiology can be complex, with two main aspects of microbial viability being the intactness of the cell membrane and cellular metabolism. While

culture-dependent approaches are considered the gold standard, they prove ineffective when dealing with bacteria that cannot be cultured. Culture-independent methods encompass viability stains and metabolism-based approaches.

Viability stains, such as Propidium iodide (PI), typically rely on the integrity of the cell membrane. In our workflow, we have incorporated PI by staining the bacteria before permeabilization. Subsequently, we hybridize the biosensors with distinct fluorescent colors, enabling us to obtain both viability and taxonomic information simultaneously, as illustrated in

Figure 3-7 (C-E).

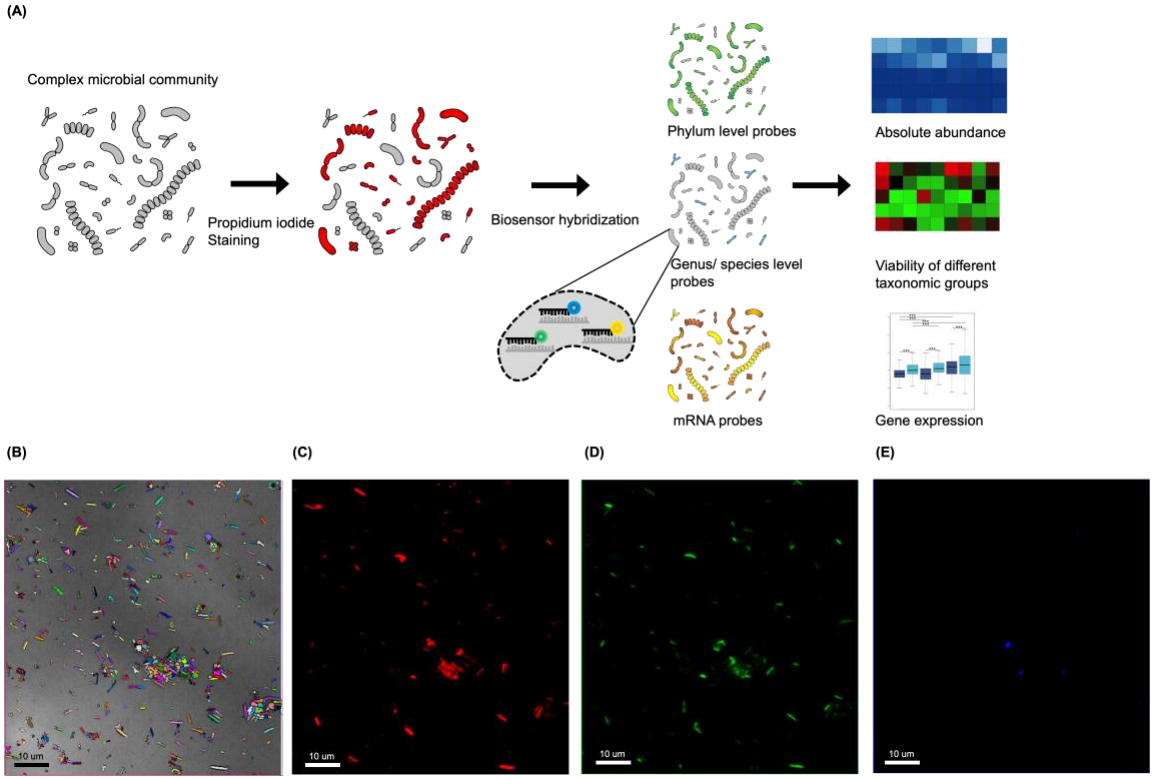
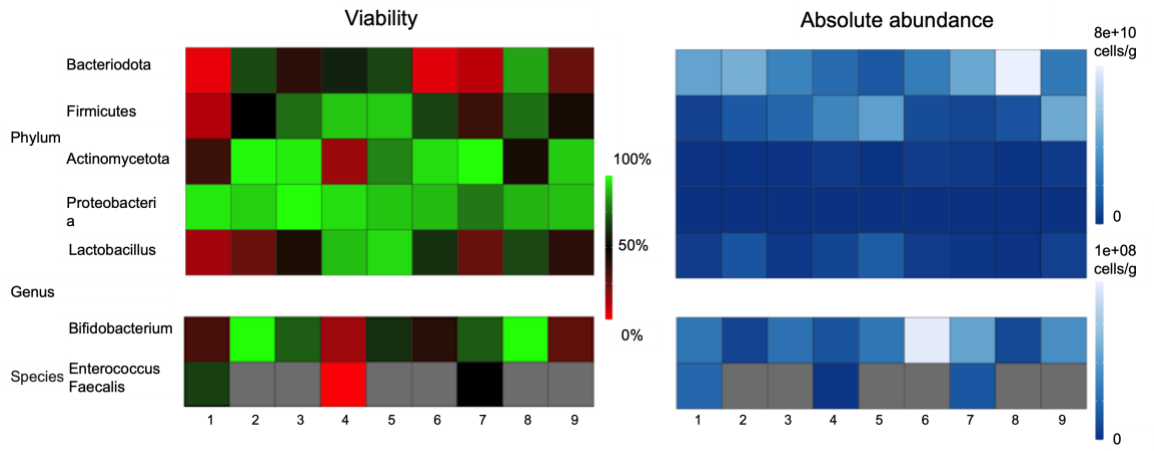


Figure 3-7 Viability of different taxonomic group. (A) The work flow of single cell viability staining and biosensor hybridization. (B) Image processing of bright field to determine the ROIs. (C) PI staining (red) of dead bacteria. (D) The phylum (green) and (E) genus (blue) biosensors signals under fluorescent microscope.

The Figure 3-8 displays the results of viability and absolute abundance. A side-by-side comparison allows us to discern the quantity of active bacteria along with their taxonomic



information. Additionally, this comparison enables us to evaluate the abundance of active bacteria within specific taxonomic groups across different samples.

Figure 3-8 A side-by-side comparison reveals the viability of various taxonomic groups and their absolute abundance.

GadB gene expression in single cell

We demonstrated the analysis of gene expression in bacteria at the single-cell level using the biosensor platform. Gut bacteria, such as *Lactobacillu plantarum*, are crucial regulators of γ-aminobutyric acid (GABA), a neurotransmitter that regulates gut motility, immune function, mood, and behavior. These bacteria produce GABA through the activity of enzymes including glutamate decarboxylase B (gadB), which converts L-glutamate into GABA by removing the carboxyl group from glutamate. We designed an mRNA probe to identify bacterial communities that express gadB. Figure 3-9 shows gadB upregulation in *Lactobacillus* in response to 30 mg/mL of glutamate. The results demonstrate the biosensor's ability to detect time-dependent gadB upregulation upon glutamate stimulation.

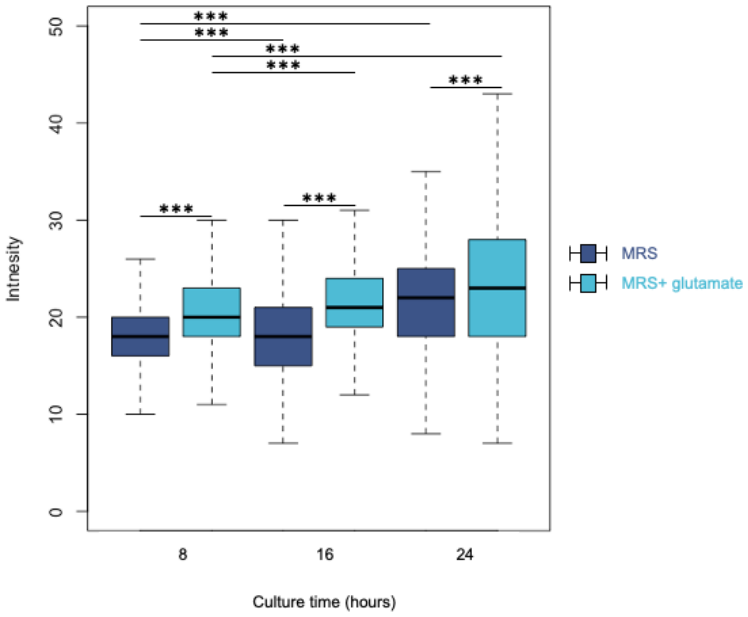


*Figure 3-9 The intensity of the gadB gene biosensors at single-cell level. The level of significance of two-way ANOVA was ***p < 0.001.*

Differences in gut microbiome profiles among mouse models of familial alzheimer's disease

The 9 healthy mice were used as mouse models of familial Alzheimer's disease. Samples numbered 1, 3, 4, 5, 6 were developed using 5XFAD mice[141], recapitulating major features of Alzheimer's Disease amyloid pathology. Samples numbered 2, 7, 8 are wild type (WT). Sample number 9 was discarded for other reasons, and genotyping was not recorded. The

Figure 3-10 displays the absolute abundance of different taxonomic groups in both 5XFAD and WT. Figure 3-11 illustrates the viability of various taxonomic groups in both 5XFAD and WT. This allows us to compare not only the differences in absolute abundance but also the viability of different taxonomic groups in both 5XFAD and WT.

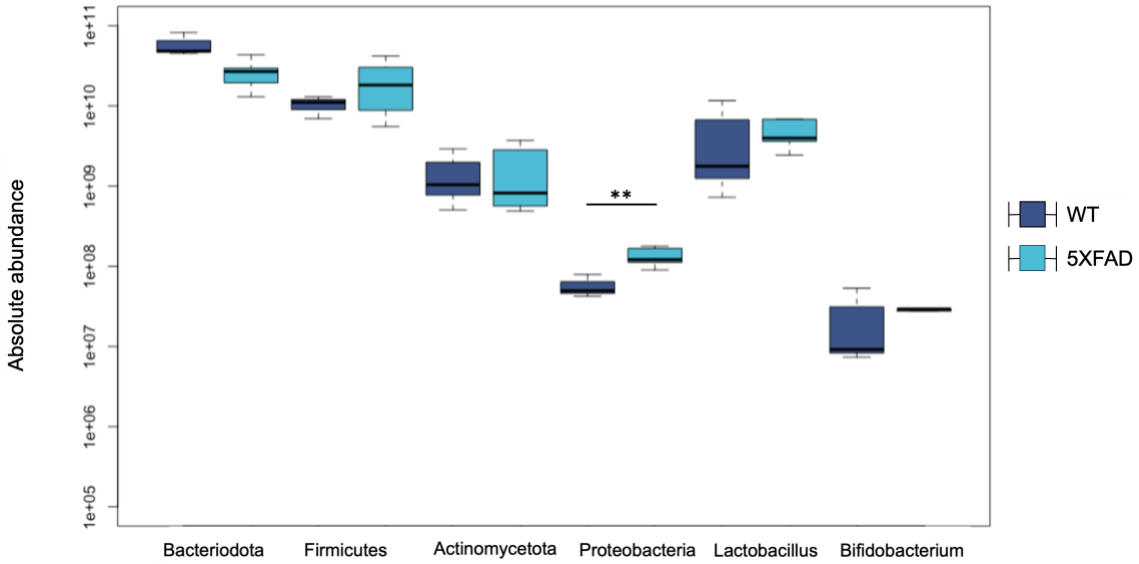


Figure 3-10 Absolute abundance of different taxonomic groups in both 5XFAD and WT.

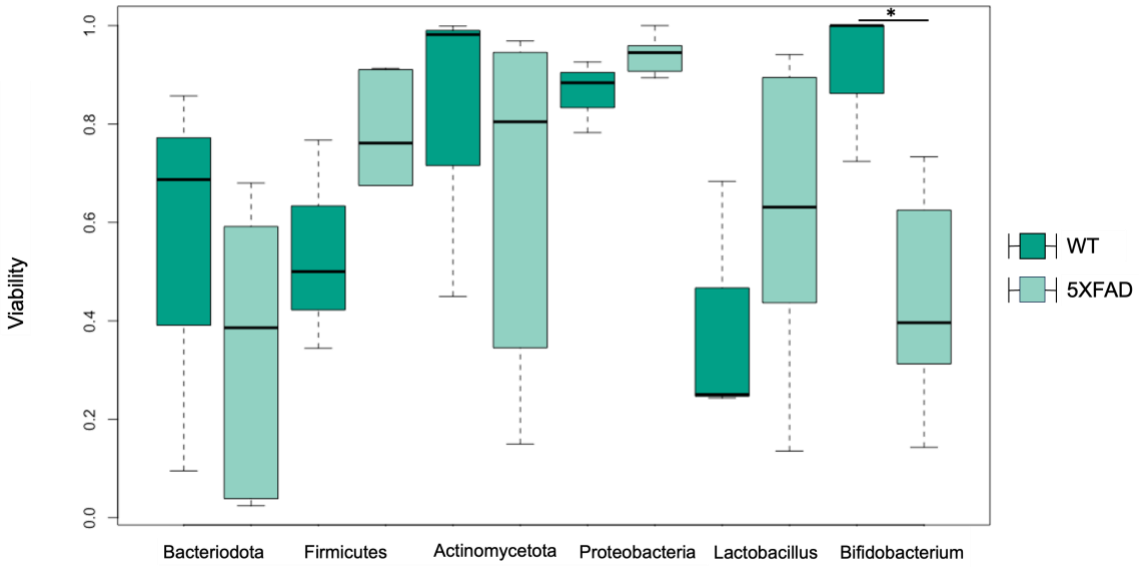


Figure 3-11 Viability of different taxonomic groups in both 5XFAD and WT.

We observe a notable increase in the numbers of *Proteobacteria*, aligning with findings from other study[142]. In addition to the rise in absolute abundance, there is a noticeable trend towards higher viability of *Proteobacteria* in 5XFAD mice. By performing transversal analysis, our results also show patterns of increase in the ratios of viable *Firmicutes/Bacteroidot*a and *Bifidobacterium/Actinomycetota*. Despite the absolute abundance of *Lactobacillus* being similar between WT and 5xFAD groups, there was an increase in the amount of viable *Lactobacillus* in the 5xFAD group.

Comparative analysis of gadB gene expression in *Lactobacillus* versus the entire taxonomic spectrum among mouse models of familial alzheimer's disease

To investigate gadB gene expression in *Lactobacillus* within 5XFAD mice, we employed phylogenetic biosensors targeting both the *Firmicutes* phylum and *Lactobacillus* genus, along with gadB gene biosensors, in stool samples. Figure 3-12 illustrates that, in the absence of phylogenetic biosensors, the gadB gene biosensors (full taxonomy) yield a p-value of <0.05. However, the small effect size observed between the two groups (Cohen's d of 0.2) implies that the statistical analysis does not unveil a significant difference. This suggests that the observed variation may lack practical significance or substantial impact on the outcome.

On the other hand, concurrent application of phylogenetic biosensors allows for the detection of *Lactobacillus* gadB gene expression (Figure 3-13). The results demonstrate a high effect size (Cohen's d of 0.65), significantly higher than that observed in wild-type mice. Our

combined use of phylogenetic and gene biosensors enables the capture of information not discernible by bulk technologies such as qPCR or sequencing.

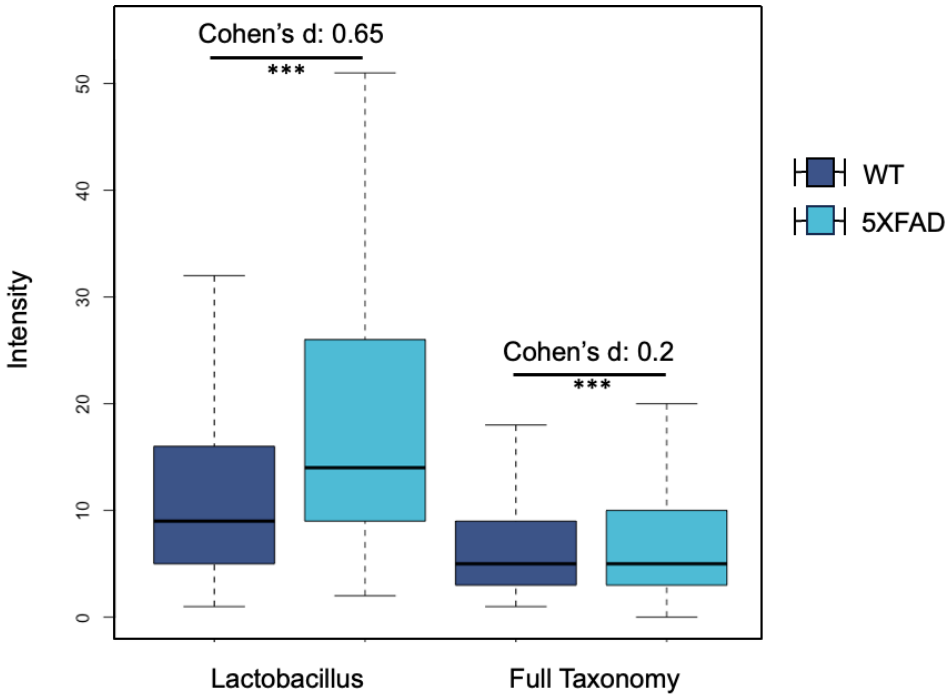


Figure 3-12 GadB gene expression in Lactobacillus versus the entire taxonomic spectrum.

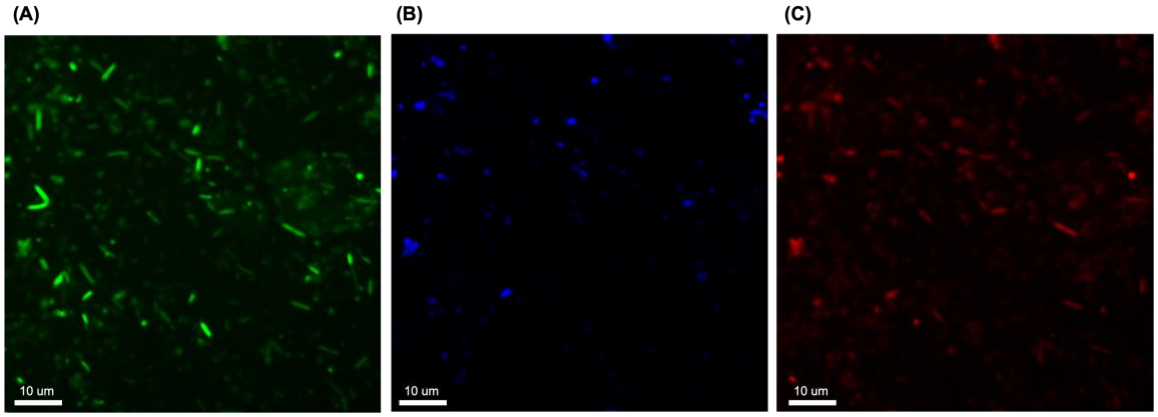


Figure 3-13 Concurrent application of phylogenetic biosensors allows for the detection of Lactobacillus gadB gene expression (A) The phylum (green) and (B) genus (blue) biosensors (C) gadB gene biosensors (red) signals under fluorescent microscope.

Centrifugal microwell array for high-throughput sample preparation

While microfluidic devices have demonstrated the capability to handle multiple conditions, enable parallel drug screening, and even facilitate single-cell analysis, their widespread adoption is hindered by the need for specialized pumps to achieve precise flow control. This requirement imposes a significant entry barrier. However, the utilization of a centrifugal microwell array offers a solution to overcome such limitations without the necessity for high-end equipment. The device comprises multiple layers, including acrylic, double-sided tape, and MAS-coated glass slides, as depicted in Figure 3-14 (A). Fabrication of the acrylic and double-sided tape was accomplished using a laser cutting machine. Figure 3-14 (B) and (C) illustrate the assembly of the microwell array to the adapter, allowing for centrifugation in individual wells. The bacteria can be immobilized on the adhesive surface by placing the device into a well plate centrifuge. Subsequent steps such as pipetting, washing, incubating, and imaging can be performed seamlessly on the same device.

For processes like enzyme treatment and probe transformation, over 10 centrifuge steps are required. The proposed microfluidic device enables parallel sample processing, a crucial aspect for efficient diagnosis. The multiwell design facilitates high-throughput sample preparation. The results, as shown in the Figure 3-15, indicate that even after 10 rounds of pipetting and washing, 99.7% of bacteria remain adhered to the adhesive surface. This underscores the effectiveness and reliability of the device in sample retention and processing. We use this microwell array to conduct the experiment of GadB gene expression in single cell.

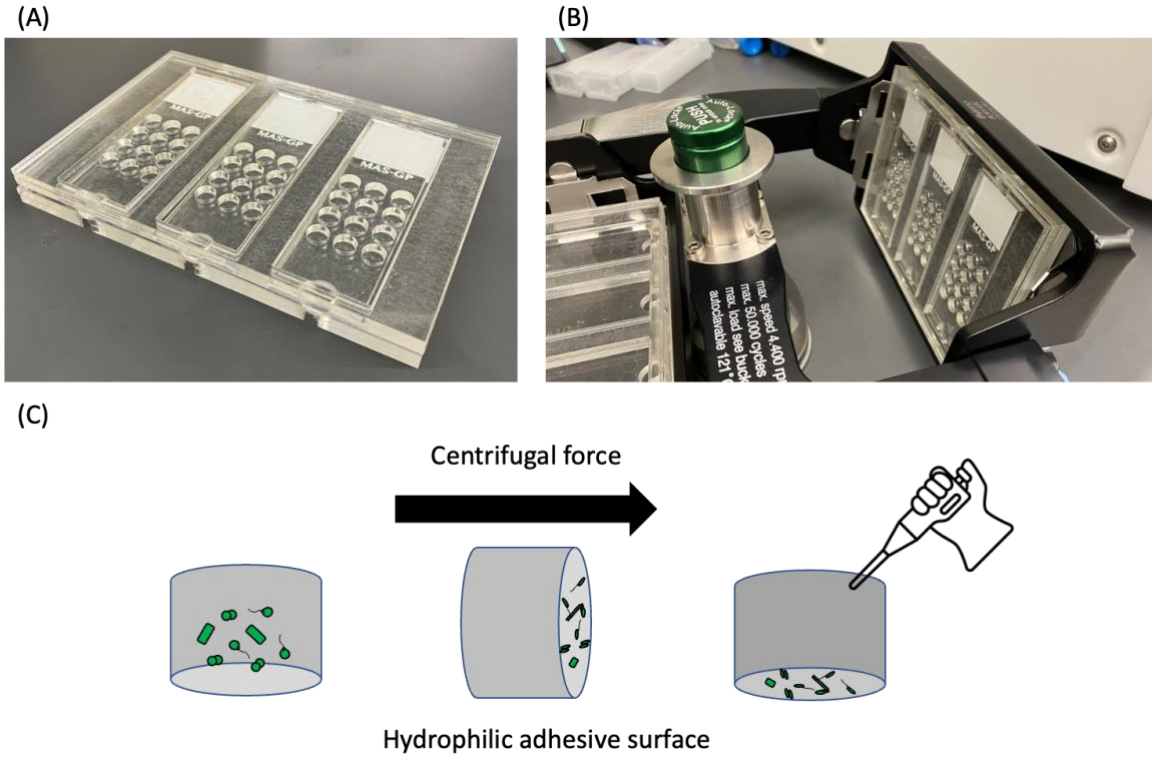


Figure 3-14 Centrifugal microwell array for high-throughput sample preparation (A)multiwell devices and adapter. (B)Integrated with bench-top centrifuge.(C)The illustration of spinning process.

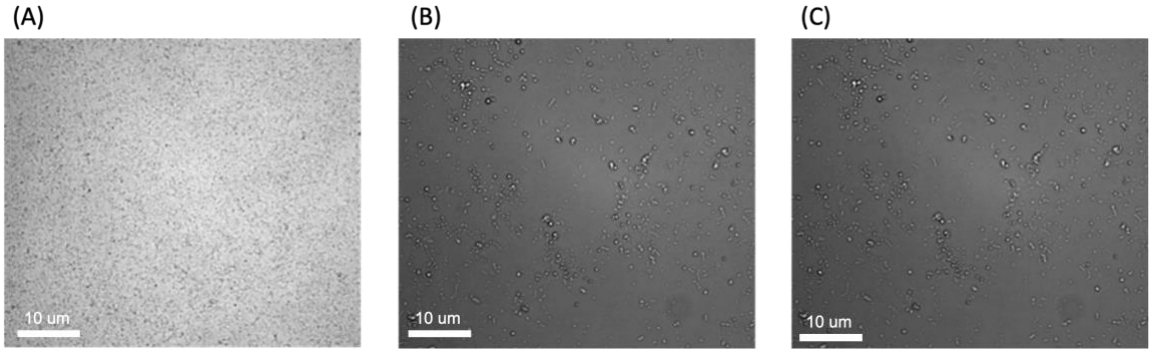


Figure 3-15 Bright field images of bacteria inside microwell. (A) Before spinning bacteria are out of focus. (B) After spinning bacteria are sitting on the same focus. (C) After 10 rounds of pipetting and washing, 99.7% of bacteria remain adhered to the adhesive surface.

Chapter 4 Conclusion and Future Work

Conclusion

We provide an in-depth exploration of multimodal biosensor design. It delves into the complex methodologies for creating these biosensors, focusing on the determination of consensus sequences, which are crucial for sensor accuracy. We also discuss the statistical analysis of binding energies, essential for understanding sensor interactions at a molecular level. Additionally, we examine the prediction of secondary structures, a key factor in sensor stability and function. We also comprehensively detail the multimodal biosensor design, focusing on enhancing detection capabilities with multiple probe combinations, streamlining the design process through an integrated R package, and applying these technologies in specific contexts. It highlights designing phylogenetic biosensors for analyzing the mouse fecal microbiome and developing gene biosensors for single-cell gene expression studies. This comprehensive approach contribution to advancing the field of biosensor technology.

We address microbial community profiling via multimodal biosensor and delve into various methodologies and applications. We focus on the application of phylogenetic biosensors to study fecal microbiomes, the viability of bacteria using Live/Dead kits, and the development of mRNA biosensors for assessing gene expression levels at the single-cell level. Techniques such as fluorescence in situ hybridization (FISH) are optimized for bacterial cells, and we outline detailed protocols for sample collection, preparation, and imaging analysis.

We also discuss the challenges of sampling from large populations of bacteria within stool samples and the use of 16S rRNA sequencing for microbial community profiling. We introduce concepts such as the limit of detection for biosensors and the importance of both relative and absolute abundance in understanding microbial communities. The viability of different taxonomic

groups is assessed, and there's a focus on the gadB gene, which is vital for GABA synthesis in bacteria.

In conclusion, this study validated the application of phylogenetic biosensors with culturable strains, including *Bacteroides fragilis, Lactobacillus plantarum, Bifidobacterium longum*, and *Escherichia coli*. *Enterococcus faecalis* was introduced to assess the system's limit of detection, yielding a calculated LoD of 0.0075%. The high R-squared value (0.9997) and a closely aligned coefficient with the transformation efficiency (0.9369) indicated strong correlation and efficacy. Acknowledging potential errors originating from factors like sampling, transformation efficiency, and probe coverage, these findings collectively underscore the robustness of the phylogenetic biosensor approach in detecting and quantifying target bacterial strains within complex microbiome samples.

In summary, our taxonomic analysis focused on key phyla, including *Proteobacteria, Bacteroidota, Actinomycetota,* and *Firmicutes*, as well as specific genera like *Lactobacillus*, *Bifidobacterium,* and *Enterococcus faecalis,* known for their physiological significance. Collaboration with CD Genomics for 16S rRNA sequencing validated the microbial composition of mouse fecal samples, and the biosensor platform effectively mirrored these findings. Notably, it detected significant changes in *Bacteroidota* and *Firmicutes* at the phylum level in samples 6, 7, and 8, demonstrating strong concordance with sequencing data. At the genus and species levels, distinct patterns emerged for *Lactobacillus, Bifidobacterium*, and *Enterococcus faecalis*. Absolute abundance analysis, based on actual cell numbers, provided valuable insights into bacterial quantities, revealing variations not captured by 16S rRNA sequencing read numbers. Incorporating viability stains like Propidium iodide allowed simultaneous assessment of bacterial viability and taxonomic information, enabling a comprehensive understanding of active bacteria within specific taxonomic groups across different samples. Nonetheless, it is generally observed that

Proteobacteria exhibit the greatest aerotolerance among these groups, leading to their increased viability in our findings. This may indicate a potential issue of oxygen exposure during the processing stage. This integrated approach enhances our ability to characterize the microbial composition and viability in complex samples, offering a more nuanced perspective on microbial dynamics.

This study successfully showcased the analysis of gene expression in bacteria at the single-cell level using a biosensor platform. The focus was on gut bacteria, particularly *Lactobacillus plantarum*, known as crucial regulators of γ-aminobutyric acid (GABA). GABA plays a vital role in regulating various physiological functions, including gut motility, immune function, mood, and behavior. The study highlighted the significance of enzymes such as glutamate decarboxylase B (gadB) in the production of GABA by these bacteria.

A specific mRNA probe designed for gadB enabled the identification of bacterial communities expressing this gene. The findings revealed gadB upregulation in *Lactobacillus* in response to 30 mg/mL of glutamate. This result demonstrated the biosensor's capability to detect time-dependent gadB upregulation upon glutamate stimulation. The study contributes valuable insights into the dynamics of gene expression in bacteria, particularly in response to external stimuli, paving the way for a deeper understanding of the regulatory mechanisms involved in neurotransmitter production by gut bacteria.

We also explore the differences in gut microbiome profiles among mouse models, particularly focusing on those with familial Alzheimer's disease, to understand the changes in microbial populations and viability. In conclusion, this study utilized 9 healthy mice as mouse models for familial Alzheimer's disease, with samples derived from 5XFAD mice and wild-type (WT) mice. The analysis focused on the absolute abundance and viability of different taxonomic groups in both 5XFAD and WT mice. Our study provides interesting insights into the changes in

the microbiota associated with the 5xFAD model. In agreement with previous reports[142,143], 5xFAD exhibited an increase in the absolute abundance of *Proteobacteria.* Moreover, there was a distinct trend towards higher viability of *Proteobacteria* in 5XFAD mice. By performing transversal analysis, our results also show patterns of increase in the ratios of viable *Firmicutes/Bacteroidota* and *Bifidobacterium/Actinomycetota*. Despite the absolute abundance of *Lactobacillus* being similar between WT and 5xFAD groups, there was an increase in the amount of viable *Lactobacillus* in the 5xFAD group. Our data also show an increase in gadB mRNA expression in *Lactobacillus* in 5XFAD fecal microbiota. Further investigation will be required to confirm these observations and elucidate the mechanistic origin of the alteration of the fecal microbiota in familial Alzheimer's disease. We anticipate that the platform will provide a valuable tool for elucidating the complex, bidirectional relationship between the gut microbiota and neurodegenerative disorders and other medical conditions.

Discussion

The human microbiota, comprising trillions of microorganisms residing in various regions of our bodies, holds a pivotal position in numerous physiological processes like digestion, immune response, and hormone regulation. The presence of microbial imbalances, known as dysbiosis, has been associated with a wide range of health conditions. Leveraging the potential of the microbiota has the potential to bring about a revolutionary shift in the development of diagnostic and therapeutic strategies, with a focus on improving treatment outcomes, minimizing complications, and preventing the recurrence of diseases.

A notable challenge in translating these advancements into clinical practice is the absence of rapid microbiota analysis methods capable of generating clinically relevant data. This thesis introduces innovative multimodal biosensors meticulously designed for precise single-cell microbiome profiling. These biosensors not only facilitate the detection but also the functional assessment of microbial populations at the single-cell level, providing a holistic understanding of the microbiome's role in health and disease. The proposed biosensors offer a comprehensive analysis toolkit, capable of determining absolute microbial abundance, viability, spatial distribution, and gene expression profiles. Furthermore, the application of these multimodal biosensors extends to conducting microbial community profiling on clinical samples. There is a specific emphasis on exploring the variations in the gut microbiome within mouse models of familial Alzheimer's disease.

However, while the biosensors represent a significant advancement, they fall short of providing the comprehensive information offered by sequencing technology. To delve deeper into functional ability and taxonomic details, there is a need for additional gene and phylogenic probes. Handling closely related species requires the development of more sophisticated barcodes, multiple rounds of hybridizations, and species-specific probes. The execution of FISH, fluorescent imaging, and imaging analysis also demands the expertise of an experienced operator. The path to process

automation and user-friendliness leans towards the integration of microfluidic systems. Moreover, the challenges extend to spatial mapping on tissue, where addressing autofluorescent issues and minimizing background noise are crucial aspects yet to be fully resolved. Despite these hurdles, this research remains dedicated to filling a critical gap in microbial diagnostics, setting the stage for the seamless integration of microbiome analysis into personalized medicine and clinical care.

Appendix A Sequence of Biosensors

	donor	quencher
Bac_1	CGACACCTCACGGCACGAGC/3Alex488N/	GCTCGTGCCGTGAGG
Bac_2	TTAAGCCGACACCTCACGGC/3Alex488N/	GCCGTGAGGTGTCGG
Firm_1	CGTAGTTAGCCGTGACTTTC/3Alex488N/	GAAAGTCACGGCTAA
Firm_2	AGGCCGGCTACTGATCGTCG/3Alex488N/	CGACGATCAGTAGCC
Firm_3	CCCCGACACCTAGTATTCAT/3Alex488N/	ATGAATACTAGGTGT
Actino_1	CGGCCATTGTAGCATGCGTG/3Alex488N/	CACGCATGCTACAAT
Actino_2	TGGCAACATGCGACGGGGGT/3Alex488N/	ACCCCCGTCGCATGT
Proteo_1	GCCCGTAAGGGCCATGAGGA/3Alex488N/	TCCTCATGGCCCTTA
Proteo_2	GCCCAGTAATTCCGATTAAC/3Alex488N/	GTTAATCGGAATTAC
Lacto	GCAGGTTCGCTTCTCGTTGT/3AlexF647N/	ACAACGAGAAGCGAA
Bifido	ACCGTTAAGCGATGGACTTT/3AlexF647N/	AAAGTCCATCGCTTA
Enterococcus faecalis	GAAAGCGCCTTTCACTCTTATGC/3AlexF647N/	

Table 7 Sequence of taxonomic probes

	donor	quencher
GadB_1	CGACACCTCACGGCACGAGC/3Alex488N/	GCTCGTGCCGTGAGG
GadB_2	TTAAGCCGACACCTCACGGC/3Alex488N/	GCCGTGAGGTGTCGG

Table 8 Sequence of gene probes

Appendix B R package

```
library("foreach")
library("parallel")
library("doParallel")
cores=detectCores()
registerDoParallel(cores)
library("seqinr")
library("seqRFLP")

target_path<-"/Users/jyong-hueilee/desktop/gadb.fas"
nontarget_path<-"/storage/home/j/jkl6028/Firm_N_20148seqs.fas"
probe_length=18
threshold_energy= 0.6
threshold_gap=0.9
threshold_match=0
quencher_length=25
RNAcofold <- '"/usr/local/bin/RNAcofold"'
RNAfold <- '"/usr/local/bin/RNAfold"'

run_RNAfold <- function(Sequences, RNAfold.path = "RNAfold",parallel.cores = 4){

 Seqs.validate <- Sequences[which(lengths(Sequences) < 30000)]

 if(length(Seqs.validate) < length(Sequences)){
  message("Due to the limitation of RNAfold,")
  message("Sequences with length more than 30000 nt will be omitted.")
  message(length(Sequences) - length(Seqs.validate), " sequences have been removed.", "\n")
  Sequences <- Seqs.validate
 }

 if (parallel.cores == 2) message("Users can try to set parallel.cores = -1 to use all cores!", "\n")

 parallel.cores <- ifelse(parallel.cores == -1, parallel::detectCores(), parallel.cores)
 cl <- parallel::makeCluster(parallel.cores, outfile = "")

 message("\n", "Sequences Number: ", length(Sequences), "\n")
 message("Processing...", "\n")
 index <- 1
 info <- paste(ceiling(length(Sequences) / parallel.cores), ",", sep = "")
 parallel::clusterExport(cl, varlist = c("info", "index"), envir = environment())
 sec.seq <- parallel::parSapply(cl, Sequences, secondary_seq, info = info,
                    RNAfold.path = RNAfold.path)
 parallel::stopCluster(cl)
 sec.seq <- data.frame(sec.seq, stringsAsFactors = FALSE)
 message("\n", "Completed.", "\n")
 sec.seq
```

```
}

secondary_seq <- function(OneSeq, info, RNAfold.path){
 seq.string <- unlist(seqinr::getSequence(OneSeq, TRUE))

 # assign("index", index + 1, inherits = TRUE)
 index <- get("index")
 showMessage <- paste(index, "of", info, "length:", nchar(seq.string), "nt", "\n")
 cat(showMessage)

 RNAfold.command <- paste(RNAfold.path, "--noPS")

 seq.ss <- system(RNAfold.command, intern = TRUE, input = seq.string)
 index <<- index + 1

 seq.ss[3] <- as.numeric(substr(seq.ss[2], nchar(seq.string) + 3, nchar(seq.ss[2]) - 1))
 seq.ss[2] <- substr(seq.ss[2], 1, nchar(seq.string))
 seq.ss
}



########################################################Probe
candidate

x  <- read.alignment(target_path, "fasta")
xx<-as.matrix(x)
y<-as.matrix(table(xx[,1]))
for(i in 1: ncol(xx)-1)
{
 y=merge(y, as.matrix(table(xx[,i+1])), by=c("row.names"),all=T, sort=T)
 rownames(y) <- y$Row.names #reset rownames
 y$Row.names <- NULL #remove added rownames col
}
y<-y[,-1]
colnames(y)<-c(1: ncol(y))
y<-y/nrow(xx)
RWNM<-row.names(y)

consensus =vector()
score =vector()
entropy =vector()
y[is.na(y)] <- 0

# y=y+(1/nrow(xx))#pseudocount (Laplace'rule of succession)

####################################
```

```
for(i in 1: ncol(y))
{

 consensus[i]<-RWNM[which(y[,i] == max(y[,i]))]
 score[i]<-max(y[,i])
 entropy[i]<-sum(na.omit(-y[,i]*log2(y[,i])))

}
two_minus_entropy<-2-entropy
merge_score_entropy<-rbind(consensus,score,entropy)


for (i in 1: ncol(merge_score_entropy)){
 if (merge_score_entropy[1,i]=="-" & as.numeric(merge_score_entropy[2,i])<threshold_gap){
  merge_score_entropy[1,i]<-NA
 }

}

space=which(merge_score_entropy[1,]=="-")

merge_score_entropy=merge_score_entropy[,-space]


probe_set=0
q=0
probe_matrix=matrix(ncol= probe_length)
entropy_m=matrix(ncol= probe_length)
score_product=matrix(ncol= 1)
entropy_sum=matrix(ncol= 1)
k= length(merge_score_entropy[1,])- probe_length+1
for(i in 1:k){
 j=i+probe_length-1
 if(sum(is.na(merge_score_entropy[1,i:j]))==0){
  probe_matrix <-rbind(probe_matrix,merge_score_entropy[1,i:j])
  score_product<-rbind(score_product,prod(as.numeric(merge_score_entropy[2,i:j])))
  entropy_sum<-rbind(entropy_sum,sum(as.numeric(merge_score_entropy[3,i:j])))
  entropy_m<-rbind(entropy_m,merge_score_entropy[3,i:j])
  if(q==0){
   probe_set=probe_set+1
  }
  q=1
 }
 else{
  q=0
 }
}
probe_matrix <-probe_matrix [-1,]
```

```
score_product <-score_product [-1,]
entropy_sum <-entropy_sum [-1,]
entropy_m <-entropy_m [-1,]
#########
#########plot(as.numeric(score_product),xlab = "",ylab = "Score_product",pch='.')
#########lines(as.numeric(score_product))
#########plot(as.numeric(entropy_sum),xlab = "",ylab = "Entropy_sum",pch='.')
#########lines(as.numeric(entropy_sum))

#########space_score_product<-which(score_product>=threshold_match)
#########probe_matrix<-probe_matrix[space_score_product,]

##########################################################Secondary
matrix number and length

ss<-list()
sslist<-list()
for (i in 1:nrow(xx)) {
 space=which(xx[i,]=="-")
 ss2=xx[i,]
 ss2=ss2[-space]
 ss[[i]]<-ss2

}

#################### Change R version
RNAcofold <- "'/usr/local/bin/RNAcofold"'
RNAfold <- "'/usr/local/bin/RNAfold"'
sss <- run_RNAfold(ss[1:nrow(xx)], RNAfold.path = RNAfold, parallel.cores = 40)

for (i in 1:nrow(xx)) {
 ss_v<-strsplit(sss[2,i],"")
 sslist[[i]]<-ss_v[[1]]
}
#####################################################
cores=detectCores()
registerDoParallel(cores)
ssloop<-list()
for (i in 1:nrow(xx)) {
 ss3=sslist[[i]]
 count_0<-0
 for (v in 1:length(ss3)){
  if (ss3[v]=="."){
   count_0=count_0+1
   b=v-count_0+1
   ss3[b:v]=count_0
  }
  else{
```

```
    ss3[v]<-0
    count_0=0
   }
   ssloop[[i]]<-ss3
  }
}
registerDoParallel(cores)
secondary_matrix<-vector()
secondary_matrix<-foreach(l = 1:nrow(probe_matrix), .combine = rbind) %dopar% {
 final_v_matrix<-vector()
 final_r_matrix<-0
 temp_loop_matrix_f<-vector()
 for (i in 1: nrow(xx)){
  final_matrix<-vector()
  ss2=ss[[i]]
  b=length(ss2)-probe_length+1
  resvoir=vector() #mismatch of probes in the whole sequence

  for (v in 1:b) {
   mismatch=0
   for (j in 1:probe_length){

    if(ss2[v+j-1]!= probe_matrix[l,j]){
     mismatch=mismatch+1
    }
   }
   resvoir[v]=mismatch
  }
  seq_min_po=which.min(resvoir) #min position of mismatch in one sequence
  ssv=sslist[[i]]
  temp_ss_matrix=ssv[seq_min_po:(seq_min_po+probe_length-1)]
  tem<-as.matrix(table(temp_ss_matrix))

  if ("." %in% row.names(tem)){
   final_v_matrix<-cbind(final_v_matrix,tem['.',])
  }
  else{
   final_v_matrix<-cbind(final_v_matrix,0)
  }
  ssl=as.numeric(ssloop[[i]])
  temp_loop_matrix=ssl[seq_min_po:(seq_min_po+probe_length-1)]
  temp_loop_matrix_f<-rbind(temp_loop_matrix_f,temp_loop_matrix)
 }
 final_l_matrix<-vector()
 for (i in 1:probe_length){
  final_l_matrix[i]<-mean(temp_loop_matrix_f[,i])
 }
 final_r_matrix<-mean(final_v_matrix[1,]) #unpaired bases
```

```
 final_r_matrix<-cbind(final_r_matrix,t(final_l_matrix))
 final_r_matrix
}


#####################                    Change                    R
version##############################quencher energy
RNAcofold <- "'/usr/local/bin/RNAcofold"'
RNAfold <- "'/usr/local/bin/RNAfold"'

probe_revcom<-list()
probe_combine_q<-0
for(l in 1:nrow(probe_matrix)){
 probe_revcom[[l]]=tolower(revComp(paste(probe_matrix[l,],collapse = "")))
 temp_g=paste(probe_matrix[l,1:quencher_length],collapse = "")
 probe_combine_q=cbind(probe_combine_q,paste(probe_revcom[[l]],"&",temp_g,collapse = ",
sep = ""))
}
probe_combine_q<-probe_combine_q[1,-1]
energy_quencher<-run_RNAfold(probe_combine_q[1:nrow(probe_matrix)], RNAfold.path =
RNAcofold, parallel.cores = 4)
energy_quencher_f=energy_quencher[3,]

#########################################################hairpin;  self
dimer; GC content



probe_combine_self<-0
for(l in 1:nrow(probe_matrix)){

probe_combine_self=cbind(probe_combine_self,paste(probe_revcom[[l]],"&",probe_revcom[[l]],collap
se = '', sep = ''))
}
probe_combine_self<-probe_combine_self[1,-1]
self_dimer<-run_RNAfold(probe_combine_self[1:nrow(probe_matrix)],  RNAfold.path  =
RNAcofold, parallel.cores = 4)

hairpin<-run_RNAfold(probe_revcom[1:nrow(probe_matrix)], RNAfold.path = RNAfold,
parallel.cores = 4)

gc_result<-vector()
for (i in 1:nrow(probe_matrix)){
 gc= table(probe_matrix[i,])
 gc_f=as.matrix(gc)
 gc_rowpo_c=which(rownames(gc_f)=="c")
```

```
 gc_rowpo_g=which(rownames(gc_f)=="g")
 gc_rowpo=cbind(gc_rowpo_c,gc_rowpo_g)

 gc_result[i]=sum(gc_f[gc_rowpo,1])/probe_length
}
#########################################################


registerDoParallel(cores)
mismatch_energy=vector()
mismatch_energy<-foreach(l = 1:nrow(probe_matrix), .combine = rbind) %dopar% {
 seq_min=vector()
 probe_combine=vector()
 probe_revcom<-list()
 for(i in 1:nrow(xx)){
  space_non=which(xx[i,]== "-")  #remove gaps in target sequence
  non_sequence=xx[i,-space_non]  #remove gaps in target sequence
  b=length(non_sequence)-probe_length+1
  resvoir=vector() #mismatch of probes in the whole sequence
  for (v in 1:b) {
   mismatch=0
   for (j in 1:probe_length){

    if(non_sequence[v+j-1]!= probe_matrix[l,j]){
     mismatch=mismatch+1
    }
   }
   resvoir[v]=mismatch
  }

  seq_min_po=which.min(resvoir) #min position of mismatch in one sequence
  ssg=ss[[i]]
  temp_g_matrix=ssg[seq_min_po:(seq_min_po+probe_length-1)]
  temp_g=paste(temp_g_matrix,collapse = "")
  probe_revcom[[i]]=tolower(revComp(paste(probe_matrix[l,],collapse = "")))
  probe_combine[i]=paste(probe_revcom[[i]],"&",temp_g,collapse = '', sep = '')
 }
 probe_combine
}
#################### Change R version
RNAcofold <- "'/usr/local/bin/RNAcofold'"
RNAfold <- "'/usr/local/bin/RNAfold'"
energy_f<-0
for(l in 1:nrow(probe_matrix)){
 energy=matrix()
 energy<-run_RNAfold(mismatch_energy[l,1:nrow(xx)], RNAfold.path = RNAcofold,
parallel.cores = 4)
 energy_f=cbind(energy_f,as.numeric(energy[3,]))
```

```
}
energy_f<-energy_f[,-1]

##################################################################
####
energy_result<-vector()

for(l in 1:nrow(probe_matrix)){
 count_q=0
 for (i in 1:nrow(xx)){
  if (as.numeric(energy_f[i,l])<as.numeric(energy_quencher_f[l])){
   count_q=count_q+1
  }
 }
 energy_result[l]=count_q/nrow(xx)
}


##################################################################
####

nx  <- read.alignment(nontarget_path, "fasta")
nxx<-as.matrix(nx)
ssn<-list()
for (i in 1:nrow(nxx)) {
 space=which(nxx[i,]=="-")
 ss2=nxx[i,]
 ss2=ss2[-space]
 ssn[[i]]<-ss2

}

registerDoParallel(cores)
mismatch_energy_non=vector()
mismatch_energy_non<-foreach(l = 1:nrow(probe_matrix), .combine = rbind) %dopar% {
 seq_min=vector()
 probe_combine=vector()
 probe_revcom<-list()
 for(i in 1:nrow(nxx)){
  space_non=which(nxx[i,]== "-")  #remove gaps in non-target sequence
  non_sequence=nxx[i,-space_non]  #remove gaps in non-target sequence
  b=length(non_sequence)-probe_length+1
  resvoir=vector() #mismatch of probes in the whole sequence
  for (v in 1:b) {
   mismatch=0
   for (j in 1:probe_length){

    if(non_sequence[v+j-1]!= probe_matrix[l,j]){
```

```
            mismatch=mismatch+1
           }
          }
          resvoir[v]=mismatch
         }

         seq_min_po=which.min(resvoir) #min position of mismatch in one sequence
         ssg=ssn[[i]]
         temp_g_matrix=ssg[seq_min_po:(seq_min_po+probe_length-1)]
         temp_g=paste(temp_g_matrix,collapse = "")
         probe_revcom[[i]]=tolower(revComp(paste(probe_matrix[l,],collapse = "")))
         probe_combine[i]=paste(probe_revcom[[i]],"&",temp_g,collapse = '', sep = '')
        }
        probe_combine
       }
       ###################### Change R version
       energy_f_non<-0
       for(l in 1:nrow(probe_matrix)){
        energy=matrix()
        energy<-run_RNAfold(mismatch_energy_non[l,1:nrow(nxx)], RNAfold.path = RNAcofold,
parallel.cores = 4)
        energy_f_non=cbind(energy_f_non,as.numeric(energy[3,]))
       }
       energy_f_non<-energy_f_non[,-1]

       ##################################################################
####
       energy_result_non<-vector()
       for(l in 1:nrow(probe_matrix)){
        count_q=0
        for (i in 1:nrow(nxx)){
         if (as.numeric(energy_f_non[i,l])<as.numeric(energy_quencher_f[l])){
          count_q=count_q+1
         }
        }
        energy_result_non[l]=count_q/nrow(nxx)
       }
       energy_result_non=1-energy_result_non
       ################################################################
       library("lattice")
       library("gridExtra")
       secondary_number=secondary_matrix[,1]
       secondary_matrix=secondary_matrix[,-1]
       ######################

       qualified_probe<-which(energy_result>=threshold_energy                    &
energy_result_non>=threshold_energy)
       probe_matrix_qualified<-probe_matrix[qualified_probe,]
```

```
        energy_result_qualified<-energy_result[qualified_probe]
        energy_result_non_qualified<-energy_result_non[qualified_probe]
        energy_result_qualified_m<-
data.frame(x=energy_result_qualified,y=1:length(energy_result_qualified))
        energy_result_qualified_non_m<-
data.frame(x=energy_result_non_qualified,y=1:length(energy_result_non_qualified))
        plot1 <-barchart(x~y,data=energy_result_qualified_m,xlab = "",ylab = "percentage of target
phylum",horiz = FALSE)
        plot2 <-barchart(x~y,data=energy_result_qualified_non_m,xlab = "",ylab = "percentage of non-
target phylum",horiz = FALSE)
        df_all<-
data.frame(x=1:length(energy_result_qualified),secondary_number=secondary_number[qualified_pro
be],self_dimer=as.numeric(self_dimer[3,qualified_probe]),hairpin=as.numeric(hairpin[3,qualified_pro
be]))
        plot3 <-barchart(secondary_number~x,data=df_all,xlab = "",ylab = "number of unpaired
bases",horiz = FALSE)
        plot4 <-barchart(-self_dimer~x,data=df_all,xlab = "",ylab = "self-dimer deltaG",horiz = FALSE)
        plot5 <-barchart(-hairpin~x,data=df_all,xlab = "",ylab = "hairpin deltaG",horiz = FALSE)
        plot6 <- levelplot(secondary_matrix[qualified_probe,],xlab = "",ylab = "Unpaired
bases",col.regions=heat.colors(100))
        grid.arrange(plot1,plot2,plot3,plot4,plot5, nrow=5)

        plot7 <- levelplot(entropy_m[qualified_probe,],xlab = "",ylab =
"Entropy",col.regions=heat.colors(100))
        grid.arrange(plot6,plot7,nrow=2,widths=1:1)
```

Bibliography

1 Sender, R., Fuchs, S. & Milo, R. Are We Really Vastly Outnumbered? Revisiting the Ratio of Bacterial to Host Cells in Humans. *Cell* **164**, 337-340 (2016). https://doi.org/10.1016/j.cell.2016.01.013

2 Turnbaugh, P. J. *et al.* The human microbiome project. *Nature* **449**, 804-810 (2007). https://doi.org/10.1038/nature06244

3 Gill, S. R. *et al.* Metagenomic analysis of the human distal gut microbiome. *Science* **312**, 1355-1359 (2006). https://doi.org/10.1126/science.1124234

4 Turnbaugh, P. J. *et al.* An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444**, 1027-1031 (2006). https://doi.org/10.1038/nature05414

5 Round, J. L. & Mazmanian, S. K. The gut microbiota shapes intestinal immune responses during health and disease. *Nat Rev Immunol* **9**, 313-323 (2009). https://doi.org/10.1038/nri2515

6 Nicholson, J. K. *et al.* Host-gut microbiota metabolic interactions. *Science* **336**, 1262-1267 (2012). https://doi.org/10.1126/science.1223813

7 Wolfe, A. J. & Brubaker, L. Urobiome updates: advances in urinary microbiome research. *Nat Rev Urol* **16**, 73-74 (2019). https://doi.org/10.1038/s41585-018-0127-5

8 Chen, J. W. *et al.* The Lung Microbiome: A New Frontier for Lung and Brain Disease. *International Journal of Molecular Sciences* **24** (2023). https://doi.org/ARTN 2170

10.3390/ijms24032170

9 Pflughoeft, K. J. & Versalovic, J. Human microbiome in health and disease. *Annu Rev Pathol* **7**, 99-122 (2012). https://doi.org/10.1146/annurev-pathol-011811-132421

10 Cho, I. & Blaser, M. J. The human microbiome: at the interface of health and disease. *Nat Rev Genet* **13**, 260-270 (2012). https://doi.org/10.1038/nrg3182

11 Seekatz, A. M., Rao, K., Santhosh, K. & Young, V. B. Dynamics of the fecal microbiome in patients with recurrent and nonrecurrent Clostridium difficile infection. *Genome Med* **8**, 47 (2016). https://doi.org/10.1186/s13073-016-0298-8

12 Gubatan, J. *et al.* Gut Microbiome in Inflammatory Bowel Disease: Role in Pathogenesis, Dietary Modulation, and Colitis-Associated Colon Cancer. *Microorganisms* **10** (2022). https://doi.org/10.3390/microorganisms10071371

13 Wong, S. H. & Yu, J. Gut microbiota in colorectal cancer: mechanisms of action and clinical applications. *Nat Rev Gastroenterol Hepatol* **16**, 690-704 (2019). https://doi.org/10.1038/s41575-019-0209-8

14 Ramirez-Labrada, A. G. *et al.* The Influence of Lung Microbiota on Lung Carcinogenesis, Immunity, and Immunotherapy. *Trends Cancer* **6**, 86-97 (2020). https://doi.org/10.1016/j.trecan.2019.12.007

15 Whiteside, S. A., Razvi, H., Dave, S., Reid, G. & Burton, J. P. The microbiome of the urinary tract--a role beyond infection. *Nat Rev Urol* **12**, 81-90 (2015). https://doi.org/10.1038/nrurol.2014.361

16 Gebrayel, P. *et al.* Microbiota medicine: towards clinical revolution. *J Transl Med* **20**, 111 (2022). https://doi.org/10.1186/s12967-022-03296-9

17 Zmora, N., Soffer, E. & Elinav, E. Transforming medicine with the microbiome. *Sci Transl Med* **11** (2019). https://doi.org/10.1126/scitranslmed.aaw1815

18 Halsey, T. M. *et al.* Microbiome alteration via fecal microbiota transplantation is effective for refractory immune checkpoint inhibitor-induced colitis. *Sci Transl Med* **15**, eabq4006 (2023). https://doi.org/10.1126/scitranslmed.abq4006

19 Wang, Y. *et al.* Fecal microbiota transplantation for refractory immune checkpoint inhibitor-associated colitis. *Nat Med* **24**, 1804-1808 (2018). https://doi.org/10.1038/s41591-018-0238-9

20 Davar, D. *et al.* Fecal microbiota transplant overcomes resistance to anti-PD-1 therapy in melanoma patients. *Science* **371**, 595-602 (2021). https://doi.org/10.1126/science.abf3363

21 Pettenati, C. & Ingersoll, M. A. Mechanisms of BCG immunotherapy and its outlook for bladder cancer. *Nat Rev Urol* **15**, 615-625 (2018). https://doi.org/10.1038/s41585-018-0055-4

22 Patangia, D. V., Anthony Ryan, C., Dempsey, E., Paul Ross, R. & Stanton, C. Impact of antibiotics on the human microbiome and consequences for host health. *Microbiologyopen* **11**, e1260 (2022). https://doi.org/10.1002/mbo3.1260

23 Hou, K. *et al.* Microbiota in health and diseases. *Signal Transduct Target Ther* **7**, 135 (2022). https://doi.org/10.1038/s41392-022-00974-4

24 David, L. A. *et al.* Diet rapidly and reproducibly alters the human gut microbiome. *Nature* **505**, 559-563 (2014). https://doi.org/10.1038/nature12820

25 Selway, C. A., Eisenhofer, R. & Weyrich, L. S. Microbiome applications for pathology: challenges of low microbial biomass samples during diagnostic testing. *J Pathol Clin Res* **6**, 97-106 (2020). https://doi.org/10.1002/cjp2.151

26 Li, H. *et al.* Single-cell pathogen diagnostics for combating antibiotic resistance. *Nature Reviews Methods Primers* **3**, 6 (2023). https://doi.org/10.1038/s43586-022-00190-y

27 Davenport, M. *et al.* New and developing diagnostic technologies for urinary tract infections. *Nat Rev Urol* **14**, 296-310 (2017). https://doi.org/10.1038/nrurol.2017.20

28 Tjandra, K. C. *et al.* Diagnosis of Bloodstream Infections: An Evolution of Technologies towards Accurate and Rapid Identification and Antibiotic Susceptibility Testing. *Antibiotics (Basel)* **11** (2022). https://doi.org/10.3390/antibiotics11040511

29 Kuczynski, J. *et al.* Experimental and analytical tools for studying the human microbiome. *Nat Rev Genet* **13**, 47-58 (2012). https://doi.org/10.1038/nrg3129

30 Cole, J. R. *et al.* The Ribosomal Database Project (RDP-II): previewing a new autoaligner that allows regular updates and the new prokaryotic taxonomy. *Nucleic Acids Research* **31**, 442-443 (2003). https://doi.org/10.1093/nar/gkg039

31 O'Callaghan, J. L., Willner, D., Buttini, M., Huygens, F. & Pelzer, E. S. Limitations of 16S rRNA Gene Sequencing to Characterize Species in the Upper Genital Tract. *Frontiers in Cell and Developmental Biology* **9** (2021). https://doi.org/ARTN 641921

10.3389/fcell.2021.641921

32 Janda, J. M. & Abbott, S. L. 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: Pluses, perils, and pitfalls. *Journal of Clinical Microbiology* **45**, 2761-2764 (2007). https://doi.org/10.1128/Jcm.01228-07

33 Chiu, C. Y. & Miller, S. A. Clinical metagenomics. *Nat Rev Genet* **20**, 341-355 (2019). https://doi.org/10.1038/s41576-019-0113-7

34 McArdle, A. J. & Kaforou, M. Sensitivity of shotgun metagenomics to host DNA: abundance estimates depend on bioinformatic tools and contamination is the main issue. *Access Microbiol* **2**, acmi000104 (2020). https://doi.org/10.1099/acmi.0.000104

35 Blauwkamp, T. A. *et al.* Analytical and clinical validation of a microbial cell-free DNA sequencing test for infectious disease. *Nat Microbiol* **4**, 663-674 (2019). https://doi.org/10.1038/s41564-018-0349-6

36 Lan, F., Demaree, B., Ahmed, N. & Abate, A. R. Single-cell genome sequencing at ultra-high-throughput with microfluidic droplet barcoding. *Nat Biotechnol* **35**, 640-646 (2017). https://doi.org/10.1038/nbt.3880

37 Köffer, J., Fronztek, A. & Eigner, U. Development and validation of a bacterial gastrointestinal multiplex RT-PCR assay for use on a fully automated molecular system. *J Microbiol Meth* **210** (2023). https://doi.org/ARTN 106754

10.1016/j.mimet.2023.106754

38 Yang, S. Y. *et al.* Multiplex Tests for Respiratory Tract Infections: The Direct Utility of the FilmArray Respiratory Panel in Emergency Department. *Can Respir J* **2020** (2020). https://doi.org/Artn 6014563

10.1155/2020/6014563

39 Paliy, O., Kenche, H., Abernathy, F. & Michail, S. High-throughput quantitative analysis of the human intestinal microbiota with a phylogenetic microarray. *Appl Environ Microbiol* **75**, 3572-3579 (2009). https://doi.org/10.1128/AEM.02764-08

40 Kang, S. *et al.* Dysbiosis of Fecal Microbiota in Crohn's Disease Patients as Revealed by a Custom Phylogenetic Microarray. *Inflamm Bowel Dis* **16**, 2034-2042 (2010). https://doi.org/10.1002/ibd.21319

41 Huang, Y. J. *et al.* Airway microbiota and bronchial hyperresponsiveness in patients with suboptimally controlled asthma. *J Allergy Clin Immunol* **127**, 372-381 e371-373 (2011). https://doi.org/10.1016/j.jaci.2010.10.048

42 Talebian, S., Dehghani, F., Weiss, P. S. & Conde, J. Evolution of CRISPR-enabled biosensors for amplification-free nucleic acid detection. *Trends Biotechnol* (2023). https://doi.org/10.1016/j.tibtech.2023.07.004

43 Weng, Z. *et al.* CRISPR-Cas Biochemistry and CRISPR-Based Molecular Diagnostics. *Angew Chem Int Ed Engl* **62**, e202214987 (2023). https://doi.org/10.1002/anie.202214987

44 Li, S. T., Zhu, L. J., Lin, S. H. & Xu, W. T. Toehold-mediated biosensors: Types, mechanisms and biosensing strategies. *Biosens Bioelectron* **220** (2023). https://doi.org/ARTN 114922

10.1016/j.bios.2022.114922

45 Takahashi, M. K. *et al.* A low-cost paper-based synthetic biology platform for analyzing gut microbiota and host biomarkers. *Nat Commun* **9**, 3347 (2018). https://doi.org/10.1038/s41467-018-05864-4

46 Croxatto, A., Prod'hom, G. & Greub, G. Applications of MALDI-TOF mass spectrometry in clinical diagnostic microbiology. *FEMS Microbiol Rev* **36**, 380-407 (2012). https://doi.org/10.1111/j.1574-6976.2011.00298.x

47 Clark, A. E., Kaleta, E. J., Arora, A. & Wolk, D. M. Matrix-assisted laser desorption ionization-time of flight mass spectrometry: a fundamental shift in the routine practice of clinical microbiology. *Clin Microbiol Rev* **26**, 547-603 (2013). https://doi.org/10.1128/CMR.00072-12

48 Yang, Y., Lin, Y. & Qiao, L. Direct MALDI-TOF MS Identification of Bacterial Mixtures. *Anal Chem* **90**, 10400-10408 (2018). https://doi.org/10.1021/acs.analchem.8b02258

49 Cheng, W. *et al.* Rapid identification of bacterial mixtures in urine using MALDI-TOF MS-based algorithm profiling coupled with magnetic enrichment. *Analyst* **147**, 443-449 (2022). https://doi.org/10.1039/d1an02098f

50 Tajik, M., Baharfar, M. & Donald, W. A. Single-cell mass spectrometry. *Trends Biotechnol* **40**, 1374-1392 (2022). https://doi.org/10.1016/j.tibtech.2022.04.004

51 Schirmer, M. & Dusny, C. Microbial single-cell mass spectrometry: status, challenges, and prospects. *Curr Opin Biotechnol* **83**, 102977 (2023). https://doi.org/10.1016/j.copbio.2023.102977

52 Bauermeister, A., Mannochio-Russo, H., Costa-Lotufo, L. V., Jarmusch, A. K. & Dorrestein, P. C. Mass spectrometry-based metabolomics in microbiome investigations. *Nat Rev Microbiol* **20**, 143-160 (2022). https://doi.org/10.1038/s41579-021-00621-9

53 Kirwan, J. A. Translating metabolomics into clinical practice. *Nature Reviews Bioengineering* **1**, 228-229 (2023). https://doi.org/10.1038/s44222-023-00023-x

54 Van Den Bossche, T. *et al.* The Metaproteomics Initiative: a coordinated approach for propelling the functional characterization of microbiomes. *Microbiome* **9**, 243 (2021). https://doi.org/10.1186/s40168-021-01176-w

55 Peters, D. L. *et al.* Metaproteomic and Metabolomic Approaches for Characterizing the Gut Microbiome. *Proteomics* **19**, e1800363 (2019). https://doi.org/10.1002/pmic.201800363

56 Yu, Y. *et al.* Comprehensive Metaproteomic Analyses of Urine in the Presence and Absence of Neutrophil-Associated Inflammation in the Urinary Tract. *Theranostics* **7**, 238-252 (2017). https://doi.org/10.7150/thno.16086

57 Long, S. *et al.* Metaproteomics characterizes human gut microbiome function in colorectal cancer. *NPJ Biofilms Microbiomes* **6**, 14 (2020). https://doi.org/10.1038/s41522-020-0123-4

58 Jagtap, P. D. *et al.* BAL Fluid Metaproteome in Acute Respiratory Failure. *Am J Respir Cell Mol Biol* **59**, 648-652 (2018). https://doi.org/10.1165/rcmb.2018-0068LE

59 Klatt, N. R. *et al.* Vaginal bacteria modify HIV tenofovir microbicide efficacy in African women. *Science* **356**, 938-945 (2017). https://doi.org/10.1126/science.aai9383

60 Frickmann, H. *et al.* Fluorescence in situ hybridization (FISH) in the microbiological diagnostic routine laboratory: a review. *Crit Rev Microbiol* **43**, 263-293 (2017). https://doi.org/10.3109/1040841X.2016.1169990

61 Alvarez-Barrientos, A., Arroyo, J., Cantón, R., Nombela, C. & Sánchez-Pérez, M. Applications of flow cytometry to clinical microbiology. *Clinical Microbiology Reviews* **13**, 167-+ (2000). https://doi.org/Doi 10.1128/Cmr.13.2.167-195.2000

62 Zimmermann, J. *et al.* High-resolution microbiota flow cytometry reveals dynamic colitis-associated changes in fecal bacterial composition. *Eur J Immunol* **46**, 1300-1303 (2016). https://doi.org/10.1002/eji.201646297

63 Rubbens, P., Props, R., Kerckhof, F. M., Boon, N. & Waegeman, W. Cytometric fingerprints of gut microbiota predict Crohn's disease state. *ISME J* **15**, 354-358 (2021). https://doi.org/10.1038/s41396-020-00762-4

64 Xu, Y. *et al.* Culture-dependent and -independent investigations of microbial diversity on urinary catheters. *J Clin Microbiol* **50**, 3901-3908 (2012). https://doi.org/10.1128/JCM.01237-12

65 Scheuermann-Poley, C. *et al.* Fluorescence In Situ Hybridization as Diagnostic Tool for Implant-associated Infections: A Pilot Study on Added Value. *Prs-Glob Open* **11** (2023). https://doi.org/ARTN e4994

10.1097/GOX.0000000000004994

66 Gao, J. *et al.* Nanotube assisted microwave electroporation for single cell pathogen identification and antimicrobial susceptibility testing. *Nanomedicine : nanotechnology, biology, and medicine* **17**, 246-253 (2019). https://doi.org/10.1016/j.nano.2019.01.015

67 Clarke, R. G. & Pinder, A. C. Improved detection of bacteria by flow cytometry using a combination of antibody and viability markers. *J Appl Microbiol* **84**, 577-584 (1998). https://doi.org/10.1046/j.1365-2672.1998.00384.x

68 Shi, H. *et al.* Highly multiplexed spatial mapping of microbial communities. *Nature* **588**, 676-681 (2020). https://doi.org/10.1038/s41586-020-2983-4

69 Dar, D., Dar, N., Cai, L. & Newman, D. K. Spatial transcriptomics of planktonic and sessile bacterial populations at single-cell resolution. *Science* **373**, 758-+ (2021). https://doi.org/ARTN eabi4882

10.1126/science.abi4882

70 Cunningham, M. *et al.* Shaping the Future of Probiotics and Prebiotics. *Trends Microbiol* **29**, 667-685 (2021). https://doi.org/10.1016/j.tim.2021.01.003

71 Falagas, M. E., Betsi, G. I., Tokas, T. & Athanasiou, S. Probiotics for prevention of recurrent urinary tract infections in women: a review of the evidence from microbiological and clinical studies. *Drugs* **66**, 1253-1261 (2006). https://doi.org/10.2165/00003495-200666090-00007

72 Al Sharaby, A., Abugoukh, T. M., Ahmed, W., Ahmed, S. & Elshaikh, A. O. Do Probiotics Prevent Clostridium difficile-Associated Diarrhea? *Cureus* **14**, e27624 (2022). https://doi.org/10.7759/cureus.27624

73 Fernandez-Barat, L., Lopez-Aladid, R. & Torres, A. Reconsidering ventilator-associated pneumonia from a new dimension of the lung microbiome. *Ebiomedicine* **60**, 102995 (2020). https://doi.org/10.1016/j.ebiom.2020.102995

74 Chu, N. D. *et al.* Dynamic Colonization of Microbes and Their Functions after Fecal Microbiota Transplantation for Inflammatory Bowel Disease. *Mbio* **12** (2021). https://doi.org/ARTN e00975-21

10.1128/mBio.00975-21

75 Goodwin, S., McPherson, J. D. & McCombie, W. R. Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet* **17**, 333-351 (2016). https://doi.org/10.1038/nrg.2016.49

76 Wang, Y. H., Zhao, Y., Bollas, A., Wang, Y. R. & Au, K. F. Nanopore sequencing technology, bioinformatics and applications. *Nature Biotechnology* **39**, 1348-1365 (2021). https://doi.org/10.1038/s41587-021-01108-x

77 Loit, K. *et al.* Relative Performance of MinION (Oxford Nanopore Technologies) versus Sequel (Pacific Biosciences) Third-Generation Sequencing Instruments in Identification of Agricultural and Forest Fungal Pathogens. *Appl Environ Microb* **85** (2019). https://doi.org/ARTN e01368-19

10.1128/AEM.01368-19

78 Hess, J. F. *et al.* Library preparation for next generation sequencing: A review of automation strategies. *Biotechnol Adv* **41** (2020). https://doi.org/ARTN 107537

10.1016/j.biotechadv.2020.107537

79 Peiffer-Smadja, N. *et al.* Performance and impact of a multiplex PCR in ICU patients with ventilator-associated pneumonia or ventilated hospital-acquired pneumonia. *Crit Care* **24**, 366 (2020). https://doi.org/10.1186/s13054-020-03067-2

80 Lemee, L. *et al.* Multiplex PCR targeting tpi (triose phosphate isomerase), tcdA (Toxin A), and tcdB (Toxin B) genes for toxigenic culture of Clostridium difficile. *J Clin Microbiol* **42**, 5710-5714 (2004). https://doi.org/10.1128/JCM.42.12.5710-5714.2004

81 Peri, A. M., Stewart, A., Hume, A., Irwin, A. & Harris, P. N. A. New Microbiological Techniques for the Diagnosis of Bacterial Infections and Sepsis in ICU Including Point of Care. *Curr Infect Dis Rep* **23**, 12 (2021). https://doi.org/10.1007/s11908-021-00755-0

82 Sogaard, K. K. *et al.* Evaluation of the clinical relevance of the Biofire FilmArray pneumonia panel among hospitalized patients. *Infection* (2023). https://doi.org/10.1007/s15010-023-02080-1

83 Li, H., Morowitz, M., Thomas, N. & Wong, P. K. Rapid Single-Cell Microbiological Analysis: Toward Precision Management of Infections and Dysbiosis. *SLAS Technol* **24**, 603-605 (2019). https://doi.org/10.1177/2472630319858922

84 Wang, X. F., Howe, S., Deng, F. L. & Zhao, J. C. Current Applications of Absolute Bacterial Quantification in Microbiome Studies and Decision-Making Regarding Different Biological Questions. *Microorganisms* **9** (2021). https://doi.org/ARTN 1797

10.3390/microorganisms9091797

85 Kau, A. L., Ahern, P. P., Griffin, N. W., Goodman, A. L. & Gordon, J. I. Human nutrition, the gut microbiome and the immune system. *Nature* **474**, 327-336 (2011). https://doi.org/10.1038/nature10213

86 Cani, P. D. Human gut microbiome: hopes, threats and promises. *Gut* **67**, 1716-1725 (2018). https://doi.org/10.1136/gutjnl-2018-316723

87 Cryan, J. F. *et al.* The Microbiota-Gut-Brain Axis. *Physiol Rev* **99**, 1877-2013 (2019). https://doi.org/10.1152/physrev.00018.2018

88 Lee, J., Kim, H. S., Jo, H. Y. & Kwon, M. J. Revisiting soil bacterial counting methods: Optimal soil storage and pretreatment methods and comparison of culture-dependent and -independent methods. *PLoS One* **16**, e0246142 (2021). https://doi.org/10.1371/journal.pone.0246142

89 Vandeputte, D. *et al.* Quantitative microbiome profiling links gut community variation to microbial load. *Nature* **551**, 507-511 (2017). https://doi.org/10.1038/nature24460

90 Stojanov, S., Berlec, A. & Strukelj, B. The Influence of Probiotics on the Firmicutes/Bacteroidetes Ratio in the Treatment of Obesity and Inflammatory Bowel disease. *Microorganisms* **8** (2020). https://doi.org/10.3390/microorganisms8111715

91 Taub, M. A., Corrada Bravo, H. & Irizarry, R. A. Overcoming bias and systematic errors in next generation sequencing data. *Genome Med* **2**, 87 (2010). https://doi.org/10.1186/gm208

92 Bonk, F., Popp, D., Harms, H. & Centler, F. PCR-based quantification of taxa-specific abundances in microbial communities: Quantifying and avoiding common pitfalls. *J Microbiol Methods* **153**, 139-147 (2018). https://doi.org/10.1016/j.mimet.2018.09.015

93 Acinas, S. G., Sarma-Rupavtarm, R., Klepac-Ceraj, V. & Polz, M. F. PCR-induced sequence artifacts and bias: insights from comparison of two 16S rRNA clone libraries constructed from the same sample. *Appl Environ Microbiol* **71**, 8966-8969 (2005). https://doi.org/10.1128/AEM.71.12.8966-8969.2005

94 Nearing, J. T. *et al.* Microbiome differential abundance methods produce different results across 38 datasets. *Nat Commun* **13**, 342 (2022). https://doi.org/10.1038/s41467-022-28034-z

95 Hansen, S. J. Z. *et al.* Droplet Digital PCR Is an Improved Alternative Method for High-Quality Enumeration of Viable Probiotic Strains. *Front Microbiol* **10**, 3025 (2019). https://doi.org/10.3389/fmicb.2019.03025

96 Luo, Y. *et al.* SAMBA: A Multicolor Digital Melting PCR Platform for Rapid Microbiome Profiling. *Small Methods* **6**, e2200185 (2022). https://doi.org/10.1002/smtd.202200185

97 Hilty, M. *et al.* Disordered microbial communities in asthmatic airways. *PLoS One* **5**, e8578 (2010). https://doi.org/10.1371/journal.pone.0008578

98 Gottschick, C. *et al.* The urinary microbiota of men and women and its changes in women during bacterial vaginosis and antibiotic treatment. *Microbiome* **5**, 99 (2017). https://doi.org/10.1186/s40168-017-0305-3

99 Kennedy, K. M. *et al.* Questioning the fetal microbiome illustrates pitfalls of low-biomass microbial studies. *Nature* **613**, 639-649 (2023). https://doi.org/10.1038/s41586-022-05546-8

100 Biesbroek, G. *et al.* Deep sequencing analyses of low density microbial communities: working at the boundary of accurate microbiota detection. *PLoS One* **7**, e32942 (2012). https://doi.org/10.1371/journal.pone.0032942

101 Eisenhofer, R. *et al.* Contamination in Low Microbial Biomass Microbiome Studies: Issues and Recommendations. *Trends Microbiol* **27**, 105-117 (2019). https://doi.org/10.1016/j.tim.2018.11.003

102 Sun, Z. *et al.* Species-resolved sequencing of low-biomass or degraded microbiomes using 2bRAD-M. *Genome Biol* **23**, 36 (2022). https://doi.org/10.1186/s13059-021-02576-9

103 Abellan-Schneyder, I., Schusser, A. J. & Neuhaus, K. ddPCR allows 16S rRNA gene amplicon sequencing of very small DNA amounts from low-biomass samples. *Bmc Microbiology* **21** (2021). https://doi.org/ARTN 349

10.1186/s12866-021-02391-z

104 Sekirov, I., Russell, S. L., Antunes, L. C. & Finlay, B. B. Gut microbiota in health and disease. *Physiol Rev* **90**, 859-904 (2010). https://doi.org/10.1152/physrev.00045.2009

105 Kim, D. *et al.* Comparison of sampling methods in assessing the microbiome from patients with ulcerative colitis. *BMC Gastroenterol* **21**, 396 (2021). https://doi.org/10.1186/s12876-021-01975-3

106 Levitan, O. *et al.* The gut microbiome-Does stool represent right? *Heliyon* **9** (2023). https://doi.org/ARTN e13602

10.1016/j.heliyon.2023.e13602

107 Norsworthy, A. N. & Pearson, M. M. From Catheter to Kidney Stone: The Uropathogenic Lifestyle of Proteus mirabilis. *Trends Microbiol* **25**, 304-315 (2017). https://doi.org/10.1016/j.tim.2016.11.015

108 Pelling, H. *et al.* Bacterial biofilm formation on indwelling urethral catheters. *Letters in Applied Microbiology* **68**, 277-293 (2019). https://doi.org/10.1111/lam.13144

109 Szollosi, J., Lockett, S. J., Balazs, M. & Waldman, F. M. Autofluorescence Correction for Fluorescence in-Situ Hybridization. *Cytometry* **20**, 356-361 (1995). https://doi.org/DOI 10.1002/cyto.990200412

110 Watrous, J. D. & Dorrestein, P. C. Imaging mass spectrometry in microbiology. *Nature Reviews Microbiology* **9**, 683-694 (2011). https://doi.org/10.1038/nrmicro2634

111 Zhang, H., Delafield, D. G. & Li, L. J. Mass spectrometry imaging: the rise of spatially resolved single-cell omics. *Nature Methods* **20**, 327-330 (2023). https://doi.org/10.1038/s41592-023-01774-6

112 Suryavanshi, M., Bhute, S., Kajale, S., Gune, R. & Shouche, Y. Molecular Imprints of Bacteria in Oxalate Containing Kidney Stone: A Preliminary Report. *JOURNAL OF CLINICAL AND DIAGNOSTIC RESEARCH* **12** (2018). https://doi.org/10.7860/JCDR/2018/32561.11341

113 Mariappan, P. & Loong, C. W. Midstream urine culture and sensitivity test is a poor predictor of infected urine proximal to the obstructing ureteral stone or infected stones: a prospective clinical study. *J Urol* **171**, 2142-2145 (2004). https://doi.org/10.1097/01.ju.0000125116.62631.d2

114 Charton, M., Vallancien, G., Veillon, B. & Brisset, J. M. Urinary tract infection in percutaneous surgery for renal calculi. *J Urol* **135**, 15-17 (1986). https://doi.org/10.1016/s0022-5347(17)45500-5

115 Gault, M. H. *et al.* Bacteriology of urinary tract stones. *J Urol* **153**, 1164-1170 (1995).

116 Dornbier, R. A. *et al.* The microbiome of calcium-based urinary stones. *Urolithiasis* **48**, 191-199 (2020). https://doi.org/10.1007/s00240-019-01146-w

117 Ben-Amor, K. *et al.* Genetic diversity of viable, injured, and dead fecal bacteria assessed by fluorescence-activated cell sorting and 16S rRNA gene analysis. *Appl Environ Microbiol* **71**, 4679-4689 (2005). https://doi.org/10.1128/AEM.71.8.4679-4689.2005

118 Ben-Amor, K. *et al.* Genetic diversity of viable, injured, and dead fecal bacteria assessed by fluorescence-activated cell sorting and 16S rRNA gene analysis. *Appl Environ Microb* **71**, 4679-4689 (2005). https://doi.org/10.1128/Aem.71.8.4679-4689.2005

119 Trinh, K. T. L. & Lee, N. Y. Recent Methods for the Viability Assessment of Bacterial Pathogens: Advances, Challenges, and Future Perspectives. *Pathogens* **11** (2022). https://doi.org/10.3390/pathogens11091057

120 Emerson, J. B. *et al.* Schrodinger's microbes: Tools for distinguishing the living from the dead in microbial ecosystems. *Microbiome* **5**, 86 (2017). https://doi.org/10.1186/s40168-017-0285-3

121 Berney, M., Hammes, F., Bosshard, F., Weilenmann, H. U. & Egli, T. Assessment and interpretation of bacterial viability by using the LIVE/DEAD BacLight Kit in combination with flow cytometry. *Appl Environ Microbiol* **73**, 3283-3290 (2007). https://doi.org/10.1128/AEM.02750-06

122 Nocker, A., Sossa-Fernandez, P., Burr, M. D. & Camper, A. K. Use of propidium monoazide for live/dead distinction in microbial ecology. *Appl Environ Microbiol* **73**, 5111-5117 (2007). https://doi.org/10.1128/AEM.02987-06

123 Wang, Y. *et al.* Whole microbial community viability is not quantitatively reflected by propidium monoazide sequencing approach. *Microbiome* **9**, 17 (2021). https://doi.org/10.1186/s40168-020-00961-3

124 Zhang, X., Li, L., Butcher, J., Stintzi, A. & Figeys, D. Advancing functional and translational microbiome research using meta-omics approaches. *Microbiome* **7**, 154 (2019). https://doi.org/10.1186/s40168-019-0767-6

125 Ackermann, M. A functional perspective on phenotypic heterogeneity in microorganisms. *Nat Rev Microbiol* **13**, 497-508 (2015). https://doi.org/10.1038/nrmicro3491

126 Arsene-Ploetze, F., Bertin, P. N. & Carapito, C. Proteomic tools to decipher microbial community structure and functioning. *Environ Sci Pollut Res Int* **22**, 13599-13612 (2015). https://doi.org/10.1007/s11356-014-3898-0

127 Smirnov, K. S. *et al.* Challenges of metabolomics in human gut microbiota research. *Int J Med Microbiol* **306**, 266-279 (2016). https://doi.org/10.1016/j.ijmm.2016.03.006

128 Skinner, S. O., Sepulveda, L. A., Xu, H. & Golding, I. Measuring mRNA copy number in individual Escherichia coli cells using single-molecule fluorescent in situ hybridization. *Nat Protoc* **8**, 1100-1113 (2013). https://doi.org/10.1038/nprot.2013.066

129 Maidak, B. L. *et al.* The RDP (ribosomal database project) continues. *Nucleic acids research* **28**, 173-174 (2000).

130 Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids research* **28**, 27-30 (2000).

131 Yang, J., Park, J., Park, S., Baek, I. & Chun, J. Introducing murine microbiome database (MMDB): a curated database with taxonomic profiling of the healthy mouse gastrointestinal microbiome. *Microorganisms* **7**, 480 (2019).

132 Mcwilliam, H. *et al.* Analysis Tool Web Services from the EMBL-EBI. *Nucleic Acids Research* **41**, W597-W600 (2013). https://doi.org/10.1093/nar/gkt376

133 Ashelford, K. E. PRIMROSE: a computer program for generating and estimating the phylogenetic range of 16S rRNA oligonucleotide probes and primers in conjunction with the RDP-II database. *Nucleic Acids Research* **30**, 3481-3489 (2002). https://doi.org/10.1093/nar/gkf450

134 Ludwig, W. ARB: a software environment for sequence data. *Nucleic Acids Research* **32**, 1363-1371 (2004). https://doi.org/10.1093/nar/gkh293

135 Lorenz, R. *et al.* ViennaRNA Package 2.0. *Algorithms for Molecular Biology 2011 6:1* **6** (2011-11-24). https://doi.org/10.1186/1748-7188-6-26

136 EKIN, İ. H. *et al.* The presence and prevalence of Enterococcus faecalis and Enterococcus faecium strains in urine and stool samples. *Turkish Journal of Veterinary Research* **2**, 14-18 (2018).

137 Siegel, J. D., Rhinehart, E., Jackson, M. & Chiarello, L. 2007 guideline for isolation precautions: preventing transmission of infectious agents in health care settings. *American journal of infection control* **35**, S65-S164 (2007).

138 Zhu, B. *et al.* Rapid and specific detection of Enterococcus faecalis with a visualized isothermal amplification method. *Frontiers in Cellular and Infection Microbiology*, 1339 (2022).

139 Dillman, D. A. *Mail and Internet surveys: The tailored design method--2007 Update with new Internet, visual, and mixed-mode guide*. (John Wiley & Sons, 2011).

140 Armbruster, D. A. & Pry, T. Limit of blank, limit of detection and limit of quantitation. *The clinical biochemist reviews* **29**, S49 (2008).

141 Oakley, H. *et al.* Intraneuronal β-amyloid aggregates, neurodegeneration, and neuron loss in transgenic mice with five familial Alzheimer's disease mutations: potential factors in amyloid plaque formation. *Journal of Neuroscience* **26**, 10129-10140 (2006).

142 Chen, C. *et al.* Gut dysbiosis contributes to amyloid pathology, associated with C/EBPβ/AEP signaling activation in Alzheimer's disease mouse model. *Science advances* **6**, eaba0466 (2020).

143 Brandscheid, C. *et al.* Altered gut microbiome composition and tryptic activity of the 5xFAD Alzheimer's mouse model. *Journal of Alzheimer's Disease* **56**, 775-788 (2017).

VITA

Jyong-Huei Lee

EDUCATIONS

Pennsylvania State University	Ph.D., Biomedical Engineering	2019-2024
National Taiwan University	Master, Mechanical Engineering	2015-2017
National Taiwan University	Bachelor, Engineering Science and Ocean Engineering	2011-2014

PUBLICATIONS

J.-H. Lee, S. M. Chin, K. E. Mach, A. Bobenchik, J. C. Liao, S. Yang, P. K. Wong, "Translating Microbiota Analysis for Clinical Applications", " Nature Reviews Bioengineering", 2024.

J.-H. Lee, C. V. D. Linden, F. J. Diaz, P. K. Wong, " A Reconfigurable Microfluidic Building Block Platform for High-Throughput Nonhormonal Contraceptive Screening," Lab on a Chip, 2022.

J.-H. Lee, S. M. Chin, D. Chan, J. C. Liao, S. Yang, N. Zhang, P. K. Wong, " Rapid Microbial Profiling via Multimodal Biosensors for Transversal Analysis (MBioTA)", In Preparation.

J.-H. Lee, S. M. Chin, J. C. Liao, S. Yang, P. K. Wong, "Microwell Array for Studying Single Cell Killing Kinetics of Resistant Bacteria", In Preparation.

K. C. Tjandra, N. Ram-Mohan, R. Abe, M. M. Hashemi, **J.-H. Lee**, S. M. Chin, M. Roshardt, J. C. Liao, P. K. Wong, S. Yang, "Diagnosis of Bloodstream Infections: An Evolution of Technologies towards Accurate and Rapid Identification and Susceptibility Testing," Antibiotics, 2022.

B. Forsyth, P. Torab, **J.-H. Lee**, T. Malcom, T.-H. Wang, J. C. Liao, S. Yang, E. Kvam, C. Puleo, P. K. Wong, "A Rapid Single-Cell Antimicrobial Susceptibility Testing Workflow for Bloodstream Infections," Biosensors, 2021.