BIOINFORMATICS AND BIOMEDICAL RESEARCH JOURNAL
2021, Vol. 4, No. 2, 57 - 69
http://dx.doi.org/10.11594/bbrj.04.02.03

Research Article

A Review: Analysis of Microbiome Diversity in Cancer

Norma Nur Azizah[1*], Ade Arsianti[2], Linda Erlina[2]

[1] Bioinformatics, Biomedical Science Magister Program, Faculty of Medicine Universitas Indonesia, Jakarta, 10430, Indonesia
[2] Department of Medical Chemistry, Faculty of Medicine, Universitas Indonesia, Jakarta, 10430, Indonesia

Article history:
Submission November 2021
Revised December 2021
Accepted December 2021

**Corresponding author:*
E-mail: norma.nur01@ui.ac.id

ABSTRACT

There have been many studies linking the microbiome to disease, including cancer. The microbiome is all the genetic material in the form of bacteria, fungi, or viruses that exist in parts of the human body. The microbiome is reported to trigger cancer due to the presence of certain pathogenic microbes or the presence of dysbiosis. Therefore, it is necessary to analyze the diversity of the microbiome to see the relationship. Metagenomics is the latest technology that can be used to analyze the diversity of the microbiome in cancer. There are two approaches in metagenomic analysis, namely targeted sequencing and sequencing of the entire metagenome. This paper will discuss a targeted sequencing metagenomic approach for the identification of microbial diversity in cancer.

Keywords: Microbiome, cancer, targeted sequencing, metagenomics, diversity

Introduction

Cancer is a disease that causes the second most deaths globally and contributed to the number of deaths of around 10 million in 2020. In the same year, cancer in Indonesia was ranked the third-highest death rate after heart disease and stroke, with the number of cancer cases in Indonesia reaching 4.8 million with a death toll was 207,210. Globally, cancers with the highest number of cases are lung, breast, and colorectal cancer. Meanwhile, in Indonesia, the highest number of cancer sufferers is breast and colorectal cancer [1-2].

The microbiome is all genetic material of bacteria, fungi, and viruses that are existing in an organism or environment. The microbiome is also known as microbial organisms, is usually limited to bacteria, which is located in certain body parts, such as the intestines, skin, and lungs. Currently, the existence of the microbiome is one of the interesting things in the health sector. Many studies report the role of the microbiome in the body related to an imbalance (dysbiosis) in the number of microbial populations, the presence of infection by specific pathogens (specific bacteria), and bacteria that modulate carcinogens that can cause various health problems [3].

Because infectious agents such as bacteria and viruses can create toxins (poisons), carcinogenic metabolites, and induce chronic inflammation, they have been linked to cancer. Infectious-agent-associated cancers account for approximately 20% of all cancer types [4]. Due to the presence of microbiome dysbiosis in

How to cite:
Azizah, N, N, Arsianti A, E & Erlina, L. (2021). A Review: Analysis of Microbiome Diversity in Cancer. *Bioinformatics and Biomedical Research Journal 4*(2): 57 - 69. doi: 10.11594/bbrj.04.02.03

the body, it was recently suggested that the microbiome is one of the risk factors for the development of cancer. The ability to analyze the complete microbiome of numerous organs as a functioning unit is now possible because of advances in technology. The identification of new microbiomes that perform diverse activities in the body might be indicated by data indicating the diversity of the microbiome; it is not unusual to identify microbiota related to cancer.

Knowing the diversity of the body's microbiome can be done using traditional culture technology, but this method can only capture a small part of the microbiota in the body, which is around 30%. In microbiome research, the most recent technology that has been created and widely employed, particularly in genomic sequencing technology, is known as metagenomic analysis. The metagenomic analysis is a useful application of modern genomic techniques for the study of microbial microorganism communities directly in their natural environment [5]. One of the uses of this analysis is to know the sequence of the microbial genome in a healthy or sick individual which can be used for various purposes such as identifying markers of microbial dysbiosis that cause diseases such as cancer, and it also can identify certain dominant species. Furthermore, the results of a metagenomic analysis can be applied for early diagnostic methods that are fast, inexpensive, non-invasive, and inclusive to the target cancer healing therapies.

This paper will discuss the relationship between the microbiome and cancer, as well as discuss the further methods of metagenomic analysis that are often used to determine the microbiome diversity associated with cancer.

Relationship of Microbiome and Cancer

The Role of The Microbiome and Cancer

The microbiome is all genetic material of bacteria, fungi, and viruses that exist in an organism, including humans, and its environment. The microbiome of an individual organism is unique, and it is first exposed from infancy. The microbiome of the baby depends on the microbiome of the mother during labor and breastfeeding. The microbiome and the human body have a mutually beneficial relationship, but it is possible to harm humans if the balance (homeostasis) of the microbiome is disturbed by diet disturbances and the presence of an infectious disease. In addition, the microbiome in a person's body can change due to environmental influences that can lead to health benefits or even the risk of disease.

Each organ's microbiome is unique, indicating that inflammation and carcinogenesis have organ-specific consequences. The microbiome acts in an organ-specific manner due to the existence of significant and functionally relevant inter-individual microbiome variability [6], which makes it a possible determinant of illness development (including cancer). Furthermore, the microbiome's diversity and abundance in various places inside the organ can be different. Because the density of the large intestine microbiome is larger than the small intestine microbiome, the rate of colorectal cancer in the colon is higher than in the small intestine [7]. Bacterial populations differ between luminal and mucosal-associated communities in the gastrointestinal system. Even though many organs, such as the liver, lack a known microbiome, they are possibly exposed to microorganism-associated molecular patterns (MAMPs) and bacterial metabolites due to their anatomical link with the gut. Many investigations in germ-free animals have demonstrated the effect of genetically produced microbiota and carcinogens on spontaneous tumor growth in many organs, including the skin, colon, liver, breast, and lung [8-10].

In general, the microbiome in the body can induce cancer through several mechanisms that modulate carcinogenesis, including inflammation-mediated by MAMPs, DNA damage, and the production of microbiome metabolites that can induce genotoxicity (Figure 1) [11].

By activating Toll-Like Receptors (TLRs) in macrophages, myofibroblasts, epithelial cells, and tumor cells, changes in the microbiome and host defenses may promote a greater bacterial translocation and MAMPs-mediated inflammation. This impact can occur in other organs either locally or remotely (metastasis).

Bacterial genotoxins such as colibactin and cytolethal distending toxins mediate the effects of genotoxins (CDT). Genotoxins that come into direct touch with cells and enter the nucleus of host cells can cause DNA damage in organs near

the microbiome. In addition to genotoxins released by bacteria, other genotoxins that can cause DNA damage include reactive oxygen and nitrogen species (ROS and RNS) released by inflammatory cells like macrophages and hydrogen sulfide (H2S) released by the bacterial microbiota. Activation of genotoxins such as acetaldehyde, intake of nitrosamines and other carcinogens, hormone metabolisms such as estrogen and testosterone, bile acid metabolism, and changes in energy harvest are all possible effects of microbiome metabolites.

Through the inactivation of carcinogens, the synthesis of short-chain fatty acids such as butyrate, and the biological activity of cancer-preventing phytochemicals, the microbiome also promotes tumor-suppressive effects (shown in green in Figure 1). AREG (amphiregulin), DCA (deoxycholic acid), EREG (epiregulin), IL (interleukin), NFB (Nuclear factor B), NLR (NOD-B.like receptor), STAT3 (signal transducer and activator of transcription 3), TH17 (T helper 17), and TNF (tumor necrosis factor) are just a few of the tumorigenic and tumor suppressor mediators[11].

Microbiome Diversity in Cancer

Microbiome diversity in cancer varies depending on the type of cancer itself. In this section, we will discuss examples of microbiome diversity in colorectal cancer and breast cancer.

Microbiome in the body is reported to induce and cause the development of disease including cancer. The occurrence of cancer due to the microbiome can occur due to several conditions namely, cancer triggered by certain pathogenic bacteria that induce carcinogenesis and cancer promoted by microbiome dysbiosis. Microbiome dysbiosis is related to the excessive use of antibiotics that can reduce the body's microbiome population.

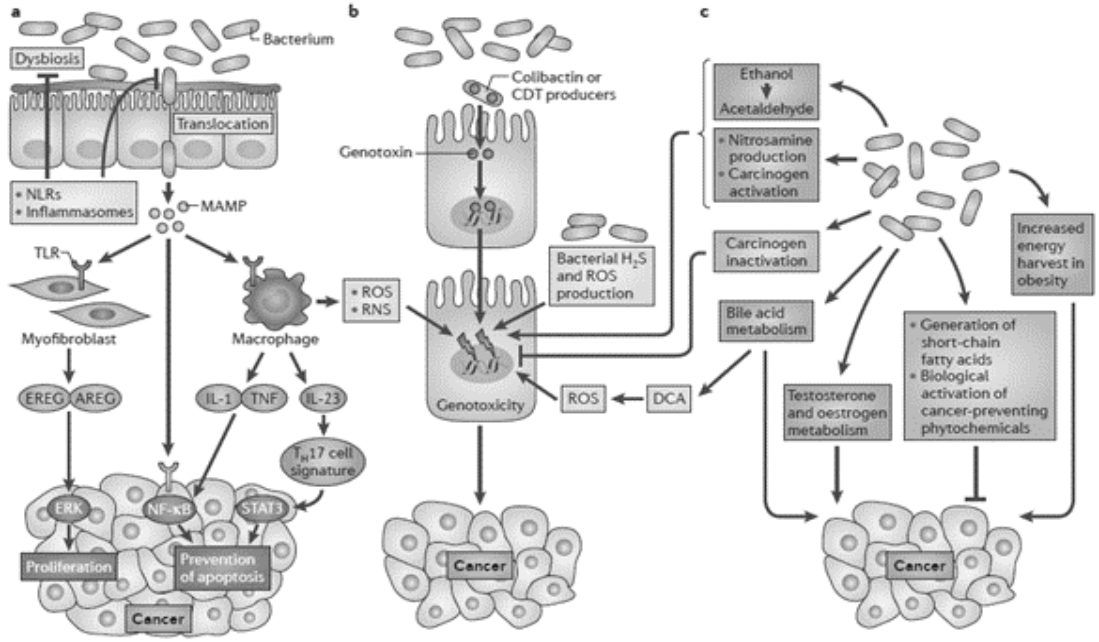

Figure 1. Bacterial Microbial Mechanisms Modulating Carcinogens (a) Inflammatory Mechanisms (b) DNA Damage Mechanisms (c) Microbiome Metabolic Mechanisms [11]

For example, in the case of gastric cancer is caused by infection with the pathogenic bacteria of H. pylori which cause sequential development starting from gastritis, gastric ulcers, atrophy, and finally gastric cancer [12].

In contrast to gastric cancer, which is caused by specific pathogenic bacteria, cancer is driven by an altered host-microbiome and dysbiosis, for example, in colorectal cancer and breast cancer. In stool and intestinal mucosal

samples from colorectal cancer patients, the diversity of the bacterial microbiome was shown to be lower than in healthy people. Furthermore, certain microorganisms in the bacterial microbiome of colorectal cancer patients changed significantly. *Bacteroides fragilis, Fusobacterium nucleatum, Enterococcaceae* or *Campylobacter, Peptostreptococcus, Enterococcus faecalis, Escherichia coli, Shigella,* and *Streptococcus gallolyticus* were found to be significantly higher in colorectal cancer patients, while *Faecalibacterium, Blautia, Clostridium,* and *Roseburia* were significantly lower [13].

Furthermore, based on 16S rRNA analysis, the diversity of the bacterial microbiome in breast cancer patients was observed to be decreased or reduced in some taxa, with some substantially larger abundances, such as *Enterobacteriaceae, Bacillus,* and *Staphylococcus spp*. In breast cancer cells, *Escherichia coli* bacteria from the *Enterobacteriaceae* family, and *Staphylococcus epidermidis* both cause double-stranded DNA cleavage [14]. Another study found that the diversity of the bacterial microbiome differed between benign and aggressive breast cancer.

In his study, Hiken et al. 2016 found that five genera, *Fusobacterium, Atopobium, Gluconacetobacter, Hydrogenophaga,* and *Lactobacillus*, showed changes in the relative abundance of bacteria in benign and malignant breast cancer cells. In contrast to malignant tumor cells, benign tumor cells have a reduced relative abundance or decrease [15-16].

Method of Microbiome Diversity Analysis

The study of genetic material extracted from environmental samples is known as metagenomics. The genetic material studied in this study is usually bacteria, fungi, archaea, and even viruses. This method is useful when we want to understand what microbes exist and what they do in their environment. Metagenomic analysis has a wide range of applications, including not just identifying the variety of the microbiome in cancer patients, but also in fields like marine and soil research, which play an essential role in environmental ecosystems.

Determining the diversity of the microbiome has been carried out conventionally with traditional procedures for a long time, namely collecting samples, coating them on a plate, and sequencing the growing microbes. This conventional method has a weakness, namely that most microbes cannot be cultured by coating because many factors affect their growth such as contamination and others. With the metagenomic method, it is possible to skip this step of microbial propagation (culture) by extracting genetic material (nucleic acid) directly from the sample and providing access to 100% theoretical genetic material [17]. So, with this method, the entire microbial population can be determined.

There are 2 types of metagenomic approaches namely, metagenome targeted sequencing and metagenomic shotgun sequencing or also known as Whole Metagenome Shotgun (WMS) sequencing.

Metagenome Targeted Sequencing

Targeted metagenome sequencing is a method of sequencing that focuses on genes (amplicons) in conserved areas. Protected Areas 16S ribosomal RNA (16S rRNA), 18S ribosomal RNA (18S rRNA), and Internal Transcribed Spacer (ITS) are the genes targeted in this technique [18]. 16S rRNA refers to 16S ribosomal ribonucleic acid (rRNA), with S (Svedberg) as the measurement unit (sedimentation rate). This rRNA is a key component of bacterial ribosomes' small subunit (SSU), as well as mitochondria and chloroplasts. Conserved regions have variable regions that can identify various groups of organisms. The 16S variable area is known to have 9 hypervariable regions named V1-V9, this variable area is also known to be not evenly distributed. In this application, conserved regions were amplified with PCR primers and sequenced. Primers 27F and 1492R were used to amplify the full-length 16S gene, which was then sequenced using Sanger DNA [19].

Regarding the purpose of using this method, the main purpose of sequencing is only to describe the microbial community taxonomically. This method can only determine the species or, more accurately, the genera in a sample without genetic information, although managing the data is simple enough for microbiological investigation in some cases. The steps

commonly used in the analysis of microbiome diversity using the 16S rRNA metagenome targeted sequencing method are as follows (Figure S1).

The sample used for microbiome analysis in cancer is usually stool or cancer tissue taken from a cancer patient (biopsy). The sample is extracted to take its DNA with a certain kit, then amplified with rRNA genes in certain areas using PCR according to the primer. Furthermore, DNA sequencing was carried out using NGS (Next Generations Sequencing) technology to obtain sequence groups to be determined and identified taxonomy with operational taxonomic units (OTU).

The application of amplicon sequencing to metagenomics is distinctive, and many bioinformatics methodologies, tools, and pipelines have been presented. Qiime 2 and Mortur are two pipelines that have received a lot of attention. This analysis can make use of a variety of pipelines. As illustrated in Figure 2, the bioinformatics technique for microbiological studies can be separated into three primary elements.

The entire technique begins with typical sequence amplicon pre-processing. The elimination of undesired sections of the sequence, such as the sequencing adaptor or barcode sequence required to differentiate certain samples, is the hallmark of this stage. After that, the sequences are demultiplexed, which categorizes them [20].

A particular clustering phase is necessary after the low-quality sequences have been filtered out. Operational Taxonomic Units (OTUs), which represent groups of similar organisms, are used to sort sequences. Following that, a diversity analysis may be performed, and the microbiome can be visualized in a lot of formats after dimensionality reduction.

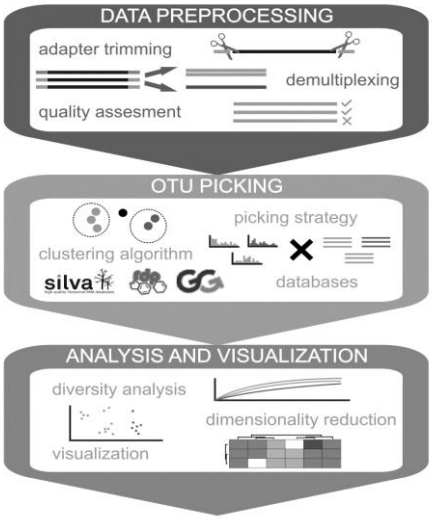

Figure 2. Bioinformatics Strategy for Targeted Sequencing [20]

Data Pre-processing

The first step in metagenome targeted sequencing data analysis is data pre-processing which includes adapter trimming and demultiplexing. Short oligonucleotides bonded to DNA sequenced during library preparation, a laboratory sample preparation step that permits separated DNA to be read by a sequencing machine, are known as adapters. Adapters tied to the 5' end of the DNA fragment may precede the length of the reads, but if the DNA fragment is shorter, the adapter at the 3' end will appear in the reads. The adapter must be trimmed from the reads so that only the amplicons are left. The presence of adapters can complicate the identification of taxonomic readings because adapters are synthetic sequences. Tools that can be used in trimming adapters include Trimmomatic which is a special platform for handling Illumina data for descriptive microbial studies [21].

The general sequencing of multiple samples is called multiplexing, which is often used in High Throughput Sequencing (HTS). Several samples were mixed and sequenced simultaneously. Each DNA fragment is labeled with a barcode sequence that allows one sample sequence to be identified from another. Then, between the adaptor and the sequenced fragment, a short synthetic oligonucleotide sequence known as a Tag or multiplex identifier (MID) is bonded [22]. As a result, each amplicon has a tag at the start and end.

Demultiplexing is the process of sorting reads into multiple files in such a way that each file only contains reads from the same sample. Demultiplexing does not necessitate the use of specific instruments; instead, the same tools that are used to trim adapters can be used. Tags and primers are removed after readings have been separated into distinct files, in the same way, that adapters are removed.

OTU Picking

An important step after the demultiplexing process is OTU Picking or OTU selection. An OTU is a group of organisms classified together based on DNA similarities. The identification of existing species using reference databases, the identification of new organisms, data quantification, and the application of various statistical and analytical tools for microbiome description are all possible in this category.

The goal of establishing a specific OTU is to identify all known and unknown organisms in a sample. De novo closed reference clustering, and open reference clustering is the three basic techniques for choosing OTUs [23]. De novo clustering is accomplished by comparing readings and grouping them into OTUs based on their similarity. Reads are compared to reference databases to form closed reference groups. Unassigned reads are excluded from further processing, and new OTUs are generated from reads assigned by the same sequence in the database. Combining both de novo and closed reference clustering, open reference clustering is achieved. Reads are first compared to the database and then reads that cannot be allocated are grouped from scratch [23].

The selection of a reference database is critical in data analysis. The database is utilized not only to categorize sequences in the selection of OTUs but also to determine the taxonomy of OTUs that are produced from scratch. Green genes (http://greengenes.lbl.gov/), SILVA (http://www.arb-silva.de/), VAMPS (http://www.vamps.mbl.edu/), and the Ribosomal Database Project (http://rdp.cme.msu.edu/) are some of the publicly accessible databases for microbial study.

Analysis and Visualization

After the data is converted into an OTU table, further information can be obtained using an analysis technique of microbiome diversity in the sample and phylogenetic relationships in a community. The diversity analysis is the initial step in interpreting amplicon metagenomic data. Alpha (α) diversity and beta (ß) diversity are the two most used types of diversity analysis. The term "alpha diversity" refers to the species in the sample and indicates the relationship between their occurrences. Beta diversity, on the other hand, tracks variations in species composition between samples [24].

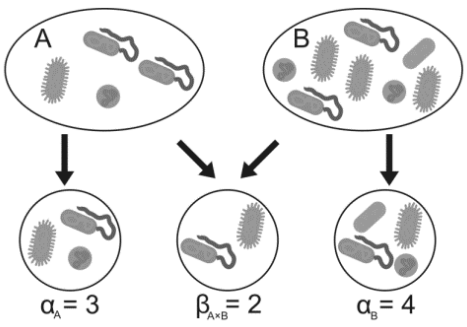

Figure 3. Principle of Diversity Analysis

Figure 3 describes the principle of microbial diversity analysis in samples. Sample A has four sequences from three different species, so diversity is three, while Sample B has eight sequences from four different species, thus diversity is four. Because the two samples have two distinct species in common, the diversity is two.

Analysis of microbial species diversity includes three components, namely species richness, species evenness, and taxonomy or diversity phylogeny [24]. The number of species in a microbial community is measured by species richness. Species evenness refers to the number of species in a community and how evenly they are distributed. Meanwhile, diversity phylogeny shows the relationship between species in a community.

The final step in microbiome studies after diversity analysis is to visualize the results in various ways. Efficient visualization methods can help discover hidden relationships between the microbiome and are very important in the interpretation of metagenomic data. There are many ways to visualize metagenomic data using various libraries and tools, including data visualization in scripting languages like Python, R, Matlab, and others. [20]

Visualization can also use software applications for statistical computing such as SPSS and Orange 3. Even metagenomic data visualization can be processed through certain web servers such as View Qiime (https://view.qiime2.org/) and Emperor (https://biocore.github. io/emperor/).

In addition, tools for analyzing microbiome diversity can be performed with web-based servers such as Galaxy for microbiome /ASaiM (https://galaxyproject.eu/index-metagenomics.html).

The following is an example of the results of visualization of metagenomic data from the analysis of microbial diversity in breast tumor tissue compared to normal tissue [25]. Figure S2. A shows a bar graph of the analysis of the abundance of microbiome alpha from normal, normal pair, and breast tumor tissue samples. The diversity index is the simplest method to describe the OTU table found in diversity. Index diversity can be analogized with descriptive statistics such as mean, median, standard deviation, and others. There are several diversity indices in this study, namely Chao1, to determine estimates of differences in species richness, and the Shannon index to determine differences in species evenness. Figure S2. B shows the principal coordinate plot, where this visualization is useful for seeing differences in beta diversity.

Based on the research conducted by Smith *et al*. (2019), the microbiome diversity between normal tissue, normal pair, and breast tumor tissue was significantly different with $p = 0.026$ based on the Shannon index. In addition, normal breast tissue and normal pair had significantly higher alpha diversity than breast tumor tissue based on Richness index ($p = 0.017$), Chao1 ($p = 0.0021$), and Fisher's alpha ($p = 0.00087$). While the beta diversity was shown by principal coordinate analysis of the unweighted UniFrac distance, the normal tissue samples clustered significantly differently from the tumor tissue samples ($p = 0.002$). Normal tissue samples and normal pairs showed a significant difference along PC2 (5.54%) compared to breast tumor tissue.

Figure S3 is a visualization that shows the proportional abundance of the microbiome in normal tissue and breast tumor at the genus level, where there are differences in abundance at the genus level in normal tissue and breast tumor tissue. Genus *paucibater* and *pseudomonas* were more dominant in tumor tissue than normal tissue. In addition, in tumor tissue, the genus *alcaligenes* was also detected which was not detected in normal tissue.

Figure S4 illustrates the phylum levels in breast tumor tissue and normal tissue. The blue color represents fewer or no detectable phyla while red represents abundant phyla, phylum hierarchical groupings, and sample types are shown in the dendrogram. Smith et al. 2019 succeeded in analyzing the abundance of microbes at the phylum level described on the Spearman Heatmap (Figure S4). There were fewer phyla *Thermi* and *Actinobacteria* identified in tumor tissue, while *Fusobacteria* and *Spirochaetes* were more abundant in tumor tissue compared to normal and paired normal tissue.

In addition, visualization of microbial diversity analysis can be displayed in the form of a circular cladogram that describes differences in taxonomic abundance. A taxonomic level of class, order, or family is represented by each letter in a circular ring of vertices. The taxa detected in normal tissue were much higher than in breast cancers, as indicated by the black arrows. In Figure S4 the class *Clostridia*, *Bacterodia*, and the family *Ruminococcaceae* are the most abundant in the tumor samples. Meanwhile, the cladogram with black arrows in the order *Actinomycetales* increased significantly in normal tissue compared to breast tumor tissue [25].

According to another study, environmental, host genetic, and socioeconomic factors all influence the breast cancer prevalence landscape, with extensive implications on racial differences for breast cancer subtypes. Thyagarajan *et al.* (2020) published the findings of their study to see if race-specific bacteria may be found in breast tissue. They found normal tissue samples adjacent to the tumor (NAT) in a group of Black non-Hispanic (BNH) and White non-Hispanic (WNH) female patients during their research. TNBC (triple-negative breast cancer) and TNBC (triple-positive breast cancer) tissue were collected (TPBC). A retrospective tumor sequencing approach based on the 16S rRNA gene was used in their study. It was discovered that the makeup of the microbiota differed significantly depending on the relative abundance of particular taxa.

The differences in microbial diversity between tumors and normal tissues are shown in Figure 5 in Supplementary Data 5 (S5), with different patterns in the WNH and BNH populations. Shannon diversity ($p = 0.05$) and evenness ($p = 0.04$) in tumor tissue were considerably lower in the BNH population with TNBC than in NAT tissue. The Shannon diversity ($p = 0.05$) and evenness ($p = 0.04$) of tumor tissue were considerably higher in the WNH population with TNBC than in the NAT group. In tumors, the WNH TNBC population displayed higher richness (ACE $p = 0.06$; Chao1 $p = 0.06$) than normal tissues (ACE $p = 0.06$; Chao1 $p = 0.06$). We sequenced 10 more breast tumor tissues from the WNH TNBC tissue with NAT to confirm the inverted pattern results from the alpha diversity metric. The microbiological diversity index exhibited a similar pattern in the WNH tumor samples, demonstrating increased Shannon diversity ($p = 0.04$) and richness (ACE $p = 0.004$; Chao1 $p = 0.006$). Tumor tissue had significantly higher alpha diversity (ACE ($p = 0.04$) and Chao1 richness ($p = 0.05$) than normal tissue in TPBC patients. There were no significant changes in Shannon diversity and evenness between tumors and NAT. significant, most likely as a result of the considerable inter-individual heterogeneity [26].

Microorganisms can be studied through metagenomic investigations in a variety of ways, including starting in the lab with DNA sequencing. Using breast sample data from an ENA database, Jain & Yadav (2022) recently reported on the diversity of bacteria in breast cancer (European Nucleotide Database). The complete GALAXY server metagenomics analysis approach was used to evaluate four data sets in the form of FASTQ received with access number PRJEB25419. In addition, numerous technologies are utilized to classify and identify microbes in samples, including FastQC, Trim galore, KRAKEN2, and Krona pie charts. They found several microorganisms such as *Actinobacteria*, *Bifidobacterium cuniculi*, *Murnine leukemia virus*, *Human endogenous retrovirus*, and *Pezizomycotina* through taxonomy and metagenomic examination of breast cancer. In contrast to bacteria found in breast cancer samples, the percentages expected for *Nematocera*, *Braychycera*, *Galegeae*, and *Nelumbo nucifera* were 4 percent, 2 percent, 3 percent, and 2 percent, respectively [27].

Conclusion

The microbiome has a relationship with the incidence or growth of the disease, including cancer. The relationship between microbiome diversity and cancer incidence is due to the modulation of carcinogens through 3 mechanisms, namely inflammation mechanisms, DNA damage mechanisms, and the presence of genotoxin-like microbiome metabolites. In addition, the microbiome can induce cancer due to being triggered by the presence of pathogenic bacteria or dysbiosis. The analysis of microbial diversity in cancer has developed using the latest technology, namely metagenomic

analysis. Metagenomic analysis on microbial studies has two approaches, metagenomic analysis targeted sequencing and whole metagenomic sequencing. The most widely used approach is metagenomic targeted sequencing by targeting variable regions of 16S rRNA. The steps of the metagenomic analysis of targeted sequencing include DNA extraction, DNA amplification by PCR, DNA sequencing with NGS, pre-processing data, OTU retrieval, and analysis and visualization. Research by Smith et al. (2019) revealed that alpha microbiome diversity in breast tumor tissue was lower than normal and paired with normal breast tissue, with relative abundance at the genus level dominated by *paucibater, pseudomonas*, and *alcaligenes*. At the phylum level, there were fewer phyla *Thermi* and *Actinobacteria*, while *Fusobacteria* and *Spirochaetes* were more abundant in tumor tissue than normal and paired tissues. In addition, based on cladogram analysis, the class *Clostridia*, *Bacterodia*, and family *Ruminococcaceae* had the highest abundance in breast tumor tissue**.** These observations provide clinically relevant information on the metagenome of breast cancer. To establish the correlation between these bacteria and the development and progression of breast cancer, more research is needed.

References

1. WHO. Cancer [Internet]. Geneva: WHO. 2017. [dikunjungi pada 2021 Feb 17]. Tersedia dari: http://www.who.int/mediacentre/factsheets/fs297/en/.
2. KEMENKES.Jenis Kanker Rentan Menyerang Manusia. [Internet]. Jakarta: Kemenkes.2020. [dikunjungi pada 2021 Feb 17]. Tersedia dari https://www.kemkes.go.id/article/view/20011400002/jenis-kanker-ini-rentan-menyerang-manusia.html
3. Lee KA, Luong MK, Shaw H, Nathan P, Bataille V, Spector TD. The gut microbiome: what the oncologist ought to know. British Journal of Cancer [Internet]. 2021; Available from: https://doi.org/10.1038/s41416-021-01467-x
4. Mukesh Verma. Mechanistic and Technical Challenges in Studying the Human Microbiome and Cancer Epidemiology. SAGE Journal [Internet].2016: Available from: https://doi.org: 10.1177/1533034616645219
5. Riesenfeld CS, Schloss P, Handelsman J (2004) Metagenomics: Genomic analysis of microbial communities. Annu Rev Genet 38: 525–552.
6. Consortium, H. M. P. Structure, function and diversity of the healthy human microbiome. Nature 486, 207–214 (2012).
7. O'Hara, A. M. & Shanahan, F. The gut flora as a forgotten organ. EMBO Rep. 7, 688–693 (2006).
8. Dapito, D. H. et al. Promotion of hepatocellular carcinoma by the intestinal microbiota and TLR4. Cancer Cell 21, 504–516 (2012)
9. Yoshimoto, S. et al. Obesity-induced gut microbial metabolite promotes liver cancer through senescence secretome. Nature 499, 97–101 (2013).
10. Li, Y. et al. Gut microbiota accelerate tumor growth via c-jun and STAT3 phosphorylation in APCMin/+ mice. Carcinogenesis 33, 1231–1238 (2012)
11. Robert F. Schwabe and Christian Jobin. The Microbiome and Cancer. Nature Reviews Cancer.[Internet].2013. available from: doi:10.1038/nrc3610
12. Fox, J. G. & Wang, T. C. Inflammation, atrophy, and gastric cancer. J. Clin. Invest. 117, 60–69 (2007).
13. Wang T., Cai G., Qiu Y., Fei N., Zhang M., Pang X., Jia W., Cai S., Zhao L. Structural segregation of gut microbiota between colorectal cancer patients and healthy volunteers. ISME J. 2011; 6:320–329. DOI: 10.1038/ismej.2011.109.
14. Urbaniak C, Gloor GB, Brackstone M, Scott L, Tangney M, Reid G. The Microbiota of Breast Tissue and Its Association with Breast Cancer. Environmental Microbiology Appl. (2016) 82:5039–48. DOI: 10.1128/AEM.01235-16
15. Zahra Eslami, Keivan MAzidjadeh, Sina Halvaei, Fatemeh Babapirali, Rezvan Esmaeili. Microbiome and Breast Cancer: New Role and for Ancient Population. Fron Oncol.2020.10(120).
16. Hieken TJ, Chen J, Hoskin TL, Walther-Antonio M, Johnson S, Ramaker S, et al. The microbiome of aseptically collected human breast tissue in benign and malignant disease. Sci Rep. (2016) 6:30751. DOI: 10.1038/srep30751
17. Gencore. Metagenomic.[Internet].[dikunjungi pada 2021 Nov 15]. Dapat di akses pada: https://learn.gencore.bio.nyu.edu/metgenomics/
18. Unterseher, Martin, Jumpponen, Ari, Öpik, Maarja, Tedersoo, Leho, Moora, Mari, Dormann, Carsten F., et al. Species abundance distributions and richness estimations in fungal metagenomics - Lessons learned from community ecology. Molecular Ecology. 2011. 20(2), p. 275–285.
19. EzbioCloud. 16S RRNA and 16S RRNA.[Internet].2017. [dikunjungi pada 2021 Nov 15]. Dapat diakses pada: https://help.ezbiocloud.net/16s-rrna-and-16s-rrna-gene/
20. Karel Sedlar. Methods For Comparative Analysis Of Metagenomic Data. Research gate. 2019. DOI: 10.13140/RG.2.2.12388.42882
21. Bolger, Anthony M, Lohse, Marc and Usadel, Bjoern. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics. 2014. 30(15), p. 2114–2120
22. Esling, Philippe, Lejzerowicz, Franck and Pawlowski, Jan. Accurate multiplexing and filtering for high-throughput amplicon-sequencing. Nucleic Acids Research. 2015. 43(5), p. 2513– 2524
23. Navas-molina, José A., Peralta-Sánchez, Juan M., González, Antonio, Mcmurdie, Paul J., Vázquez-Baeza, Yoshiki, XU, Zhenjiang, et al. Advancing our understanding of the human microbiome using QIIME. In :

Methods in Enzymology. 2013. P. 371–444. ISBN 9780124078635.
24. BIOMCARE. Key term in microbiome projects [Internet]. 2019. [dikunjungi pada 2021 Nov 22]. dapat diakses pada: https://biomcare.com/info/key-terms-in-microbiome-projects/
25. Alan Smith, at al. Distinct microbial communities that differ by race, stage, or breast tumor subtype in breast tissues of non-Hispanic Black and nonHispanic White women. Nature. 2019. 9:11940.
26. Thyagarajan, S., Zhang, Y., Thapa, S. et al. Comparative analysis of racial differences in breast tumor microbiome. Sci Rep 10, 14116 (2020). https://doi.org/10.1038/s41598-020-71102-x.
27. Jain, M., & Yadav, R. (2022). Metagenomics Analysis of Breast Cancer to Study Bacterial Diversity. *Journal of Drug Delivery and Therapeutics*, *12*(2-S), 119–126. https://doi.org/10.22270/JDDT.V12I2-S.5433

Supplementary 1

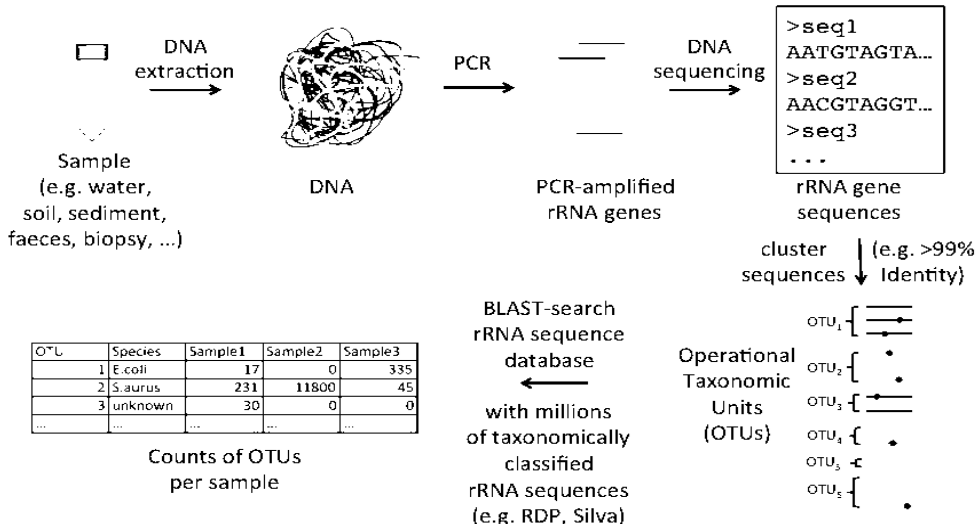

Figure S1. Workflow Metagenome Targeted Sequencing 16S rRNA[17]

Supplementary 2

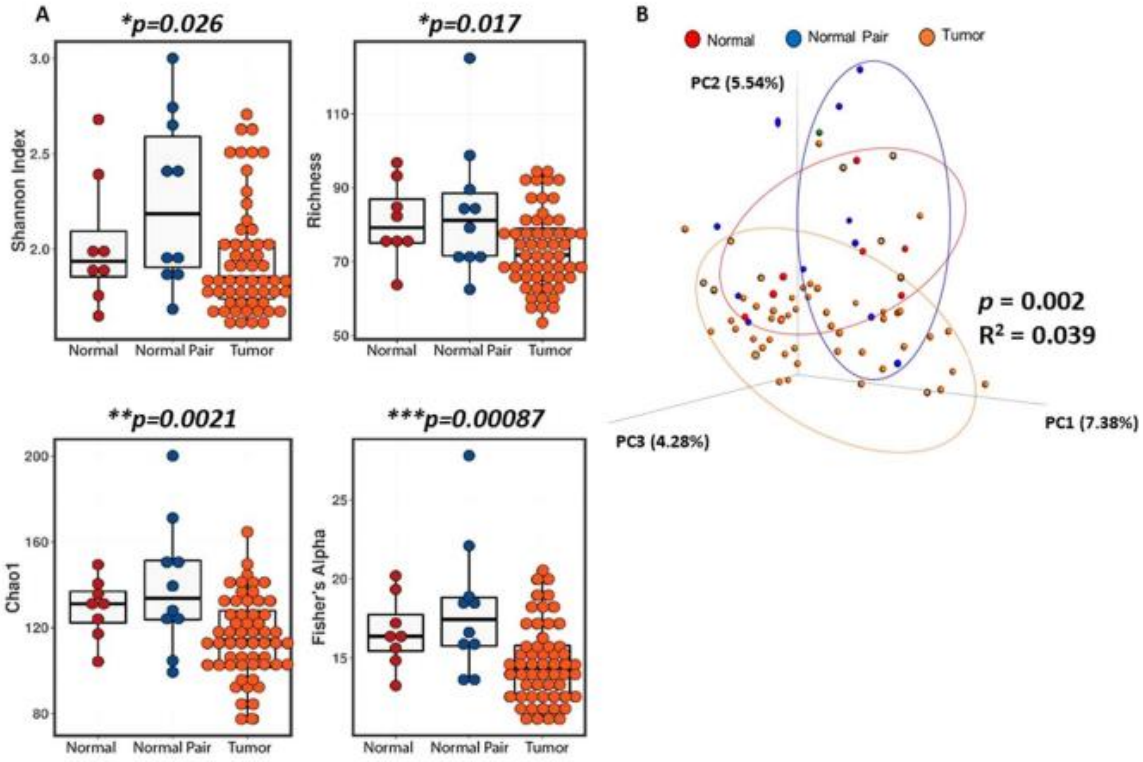

Figure S2. Visualization Analysis diversity A) Bar graph diversity; B) Principle Coordinates (PCs) plots

Supplementary 3

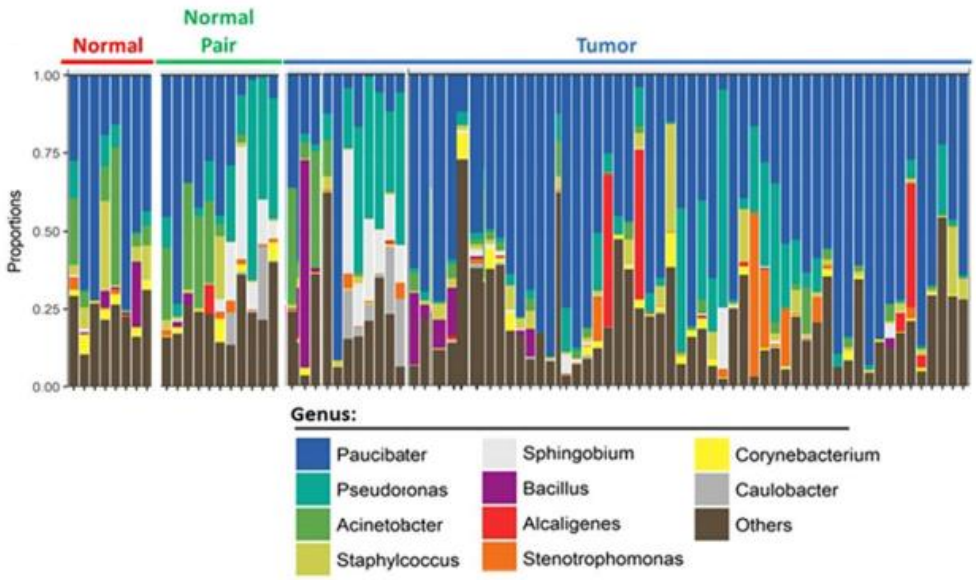

Figure S3 Bar Plots Relative Abundance of Genus Level Microbiota in Normal Tissue and Breast Tumors

Supplementary 4

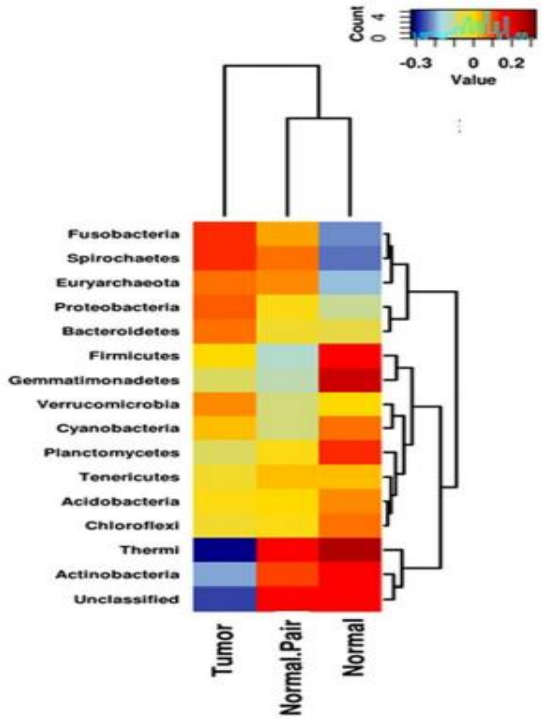

Figure S4 Spearman Heat maps

Supplementary 5

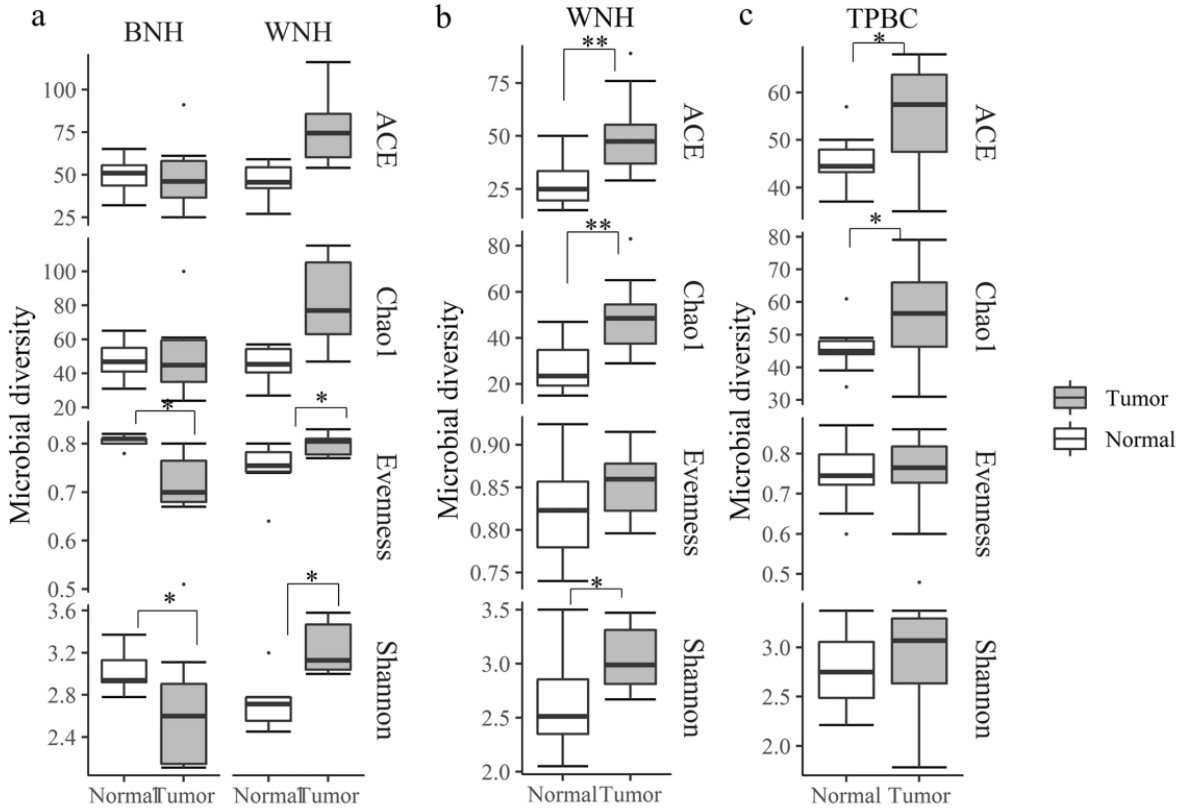

*Figure S5 Bacterial diversity in tumor and adjacent normal tissue of (**a**) TNBC from WNH (n = 6) and BNH (n = 7); (**b**) TNBC WNH 2nd set (n = 10) and (**c**) TPBC pool (n = 10) cancer patient samples. Dots represent outliers; *p < 0.05; **p < 0.01, Wilcoxon Signed Rank Test [26]*