Gut Microbiome of Patients With Breast Cancer in Vietnam

Sang M. Nguyen, MD, PhD[1]; Huong T.T. Tran, MD, PhD[2,3]; Jirong Long, PhD[1]; Martha J. Shrubsole, PhD[1]; Hui Cai, MD, PhD[1]; Yaohua Yang, PhD[4]; Lan M. Nguyen, MD, PhD[5]; Giang H. Nguyen, MD, MPH[2]; Chu V. Nguyen, MD, PhD[2]; To V. Ta, MD, PhD[2]; Jie Wu, MD, MPH[1]; Qiuyin Cai, MD, PhD[1]; Wei Zheng, MD, MPH, PhD[1]; Thuan V. Tran, MD, PhD[3,6]; and Xiao-Ou Shu, MD, MPH, PhD[1]

DOI https://doi.org/10.1200/GO.23.00234

ABSTRACT

PURPOSE Gut microbiota play an important role in human health, including cancer. Cancer and its treatment, in turn, may alter the gut microbiome. To understand this complex relationship, we profiled the gut microbiome of 356 Vietnamese patients with breast cancer.

MATERIALS AND METHODS Stool samples were collected before chemotherapy, with 162 pre- and 194 postsurgery. The gut microbiome was measured by shotgun metagenomic sequencing. Associations of gut microbial diversity, taxa abundance, and gut microbiome health index (GMHI) with sociodemographic, clinical factors, and tumor characteristics were evaluated.

RESULTS Postsurgery samples were associated with significantly lower α- and β-diversities ($P < .001$) and showed significant differences in the abundance of 15% of 2,864 investigated taxa (false discovery rate [FDR] <0.1) compared with presurgery samples. An unhealthy gut microbiome was prevalent among patients with breast cancer, with a mean GMHI of −0.79 and −2.81 in pre- and postsurgery stool samples, respectively. In an analysis of 162 presurgery stool samples, diagnosis delay was significantly associated with lower α-diversity, variation in β-diversity, an increased abundance of species *Enorma massiliensis*, and a decreased abundance of *Faecalicoccus pleomorphus*. High intake of fiber was significantly associated with lower α-diversity and a higher abundance of species belonging to *Bifidobacterium*, *Prevotella*, and *Bacteroides* gena (FDR < 0.1). We did not find that cancer stage and subtype, menopausal status, comorbidity, antibiotic use during 3 months before stool collection, or physical activity was significantly associated with α- and β-diversities or GMHI although a few significant differences were observed in taxa abundance.

CONCLUSION Our study revealed that diagnosis delay, high fiber intake, and breast cancer surgery, which is always followed by antibiotic prophylaxis in Vietnam, led to a less diverse and unhealthy gut microbiome among patients with breast cancer.

ACCOMPANYING CONTENT

Appendix

Accepted November 15, 2023
Published February 15, 2024

JCO Global Oncol 10:e2300234

INTRODUCTION

Gut microbiota play a critical role in disease prevention by maintaining barrier homeostasis, providing protection against pathogen overgrowth and maturation, continuously educating the immune response, influencing epithelial hyperproliferation, and supporting vascularization in the GI tract.[1-5] Gut microbiota also play a vital role in regulating intestinal endocrine functions by providing a source of energy biogenesis, biosynthesizing vitamins, regulating neurologic signaling and neurotransmitters, metabolizing bile acids, reacting to or modifying specific drugs, and eliminating exogenous toxins.[6-10] It has been suggested that gut microbiota might contribute to cancer development and progression and differences in therapeutic responses, clinical characteristics, and clinical outcomes although epidemiologic evidence is limited.[11-13] Studies have shown that the gut microbiome varies according to BMI levels, menopausal status, clinical stages, and histologic grades among patients with breast cancer.[14-16] The gut microbiomes of patients with breast cancer have been shown to differ from those with benign breast tumors and vary by clinicopathologic characteristics, including estrogen receptor (ER), progesterone receptor (PR), Ki-67 protein levels, and human epidermal growth factor receptor 2 (HER2) status.[17] HER2 status and age at menarche have been linked to gut microbiome α-diversity, and specific microbes.[18] These studies, however, were limited by their small sample size.

In the present study, we comprehensively evaluated the associations between the gut microbiome and sociodemographic characteristics, lifestyle, and clinical factors using

shotgun metagenome data from 356 patients with breast cancer who were enrolled in the Vietnam Breast Cancer Study (VBCS).

MATERIALS AND METHODS

Study Population

The VBCS enrolled Vietnamese patients with newly diagnosed breast cancer from the Vietnam National Cancer Hospital and Hanoi Oncology Hospital in Hanoi, Vietnam. The design and method of the VBCS have been published elsewhere.[19,20] Briefly, 501 Vietnamese women with breast cancer were recruited from inpatient surgical units and chemotherapy inpatient or outpatient units of two major cancer hospitals in Northern Vietnam, the Vietnam National Cancer Hospital and Hanoi Oncology Hospital, between July 2017 and June 2018 (response rate of 93.1%). Women age 18-79 years with newly diagnosed breast cancer who were chemotherapy-naïve were eligible for the study. Women with a history of cancer or concurrent life-threatening illnesses (eg, stroke or heart failure) were excluded. Informed consent was obtained from all participants. Human participant research approval for this study was obtained from the Vietnam National Cancer Institute (VNCI) and the Vanderbilt University Medical Center.[19]

Trained interviewers conducted in-person interviews to collect information on demographic characteristics, reproductive factors, menstrual history, family history of cancer, and lifestyle factors at the time of enrollment. In addition, a validated frequency questionnaire consisting of 68 food items/groups was used to capture dietary intake during the past 5 years.[19] The intake of macronutrients including fiber, fat, carbohydrates, and protein was calculated with the sum of grams from individual food consumed per day and estimated on the basis of the 2007 and 2017 Vietnamese food composition tables.[21,22] Furthermore, participants were asked to provide information on exercise participation and daily activities during the 10-year period preceding their enrollment. A standard metabolic equivalent (MET) score was assigned to each activity, on the basis of the Compendium of Physical Activities developed by Ainsworth et al.[23,24] Total physical activity (MET hours/week), including leisure time and daily activities, was calculated.

Clinical information such as TNM stage, breast cancer subtype, and treatment was collected by reviewing patients' medical records. This study included participants who have provided a stool sample at baseline before neoadjuvant/adjuvant chemotherapy. We excluded participants whose tumors were benign on the basis of pathologic reviews (n = 9) and those diagnosed at stage 0 (n = 2). Stool samples from four participants were excluded because of low DNA yields. In addition, participants with incomplete medical chart reviews or missing treatment information were excluded (n = 34). Finally, 356 participants were included in this study (Appendix Fig A1).

Stool Sample Collection

Stool samples were collected using fecal occult blood test (FOBT) cards at study enrollment following a standard protocol. Two FOBT cards were provided to each participant, with clear instructions on how to collect stool samples in hospital or at home. The stool samples were transferred to the VNCI research laboratory within 24 hours of collection and then stored in a −80°C freezer. History of antibiotic use for 5 days or longer during the past 3 months before stool collection was obtained. We defined postsurgery stool samples as those collected after breast cancer surgery, but before receiving adjuvant chemotherapy, presurgery stool samples were defined as those collected before breast cancer surgery or from patients who did not undergo breast cancer surgery. In total, 194 samples were collected postoperatively and 162 samples were collected preoperatively. The median time interval from surgery to stool collection was approximately 18 days for postsurgery stool samples.

Microbiome Profiling

DNA Extraction and Shotgun Metagenomic Sequencing

The DNeasy PowerSoil Kit (Qiagen, Hilden, Germany) was used to isolate DNA from stool samples. The DNBSEQ Short-read library preparation kit was used to build sequencing libraries from DNA samples for shotgun metagenomic sequencing. Sequencing was conducted at the 150-bp paired end using MGISEQ-2000 at BGI Americas. DNA extraction, library preparation, and sequencing of all samples were performed in one batch.

Sequencing Data Processing

Raw reads were processed using Trimmomatic v0.39 to trim low-quality bases, after which reads with fewer than 105 nucleotides, that is, 70% of original read lengths, were discarded.[25] Then, Bowtie2 v2.3.0 was used to remove reads that could be mapped on the human genome (GRCh38).[26] Clean reads were subjected to taxonomic profiling and estimation of the absolute abundance using Kraken v2.1.1 and Bracken v2.6, with bacterial genomes from the Unified Human Gastrointestinal Genome (UHGG) as reference.[27-29] Within each sample, only taxa with a relative abundance of ≥0.001% were considered detected.[30,31] Details on metagenomic sequencing data are summarized in Appendix Tables A1 and A2. To measure the gut microbiome health index (GMHI), species taxonomic profiling was additionally performed using MetaPhlAn2 v2.7.0, which classifies metagenomic reads to taxonomies on the basis of a database (mpa_v20_m200) of clade-specific marker genes derived from approximately 17,000 microbial genomes (Appendix 1).[32]

Statistical Analysis

We rarefied the species-level absolute abundance of every sample to the minimum number of clean reads (n = 3,578,947)

among 356 samples, using the R function *vegan::rarefy*.[33] We then estimated α- and β-diversities using the R functions *vegan::diversity* and *vegan::vegdist*, respectively.[33]

We measured α-diversity using the Chao1 index, Shannon index, and inverse Simpson index. Differences across groups were compared using the Wilcoxon rank-sum test. The square of the Shannon index and the square root of the inverse Simpson index were used in the analyses. The α-diversity indices within selected sociodemographic and clinical factor strata were estimated using the mean difference (β coefficients) and 95% CIs in linear regression models. Covariates included in the models were age group (<40, 40-49, 50-59, and ≥60 years); income level (tertile distribution); residence (urban area/rural area); age at menarche (≤15/>15 years); regular menstrual cycle (regular/irregular); menopausal status (premenopausal/postmenopausal); number of live births (≤1, 2, and ≥3); BMI (underweight <18.5, normal weight 18.5-22.9, Asian overweight 23.0-24.9, and Asian obese ≥25 kg/m^2); comorbidity (yes/no); antibiotic use within the past 3 months (yes/no); diagnosis delay (no delay ≤3 months, moderate delay of 4-8 months, and serious delay ≥9 months[34]); molecular subtype (hormone receptor–positive [HR+]/HER2-negative, HR+/HER2-positive, HER2-enriched, and triple-negative/basal-like); TNM cancer stage (stage I, stage II, and stage III-IV); fiber, fat, and carbohydrate intake; and physical activity (tertile distributions).

β-Diversity was measured using the Bray-Curtis dissimilarity matrix, weighted UniFrac distance matrix, and unweighted UniFrac distance. The permutational multivariate analysis of variance (PERMANOVA) test was applied to assess whether there was a difference in β-diversity by sociodemographic and clinical factors.[35] *R* square and *P* values from PERMANOVA tests were derived from models adjusted for the aforementioned covariates and 999 permutations using the R functions *vegan*.[33]

We derived the GMHI following the method of Gupta et al,[36] that is, on the basis of the presence of 50 microbial species associated with a healthy gut ecosystem. A positive or negative GMHI indicates a healthy or unhealthy fecal microbiome, respectively. A GMHI of 0 represented an equal balance between health-prevalent and health-scarce species. We applied multivariable linear regression analysis to estimate the β coefficients and 95% CIs of the GMHI associated with sociodemographic and clinical characteristics. The statistical analyses were performed using two-sided tests, and associations with *P* value <.05 were considered statistically significant.

Nonrarified data were used to evaluate the associations of gut microbial taxa with sociodemographic and clinical factors. Common taxa were defined as present in >50% of samples, rare taxa were defined as present in <50% of samples, and we limited our analysis to rare taxa that were present in ≥10% of samples. We evaluated the associations of gut microbial taxa with sociodemographic and clinical factors via a linear regression framework for differential abundance analysis (LinDA), in which centered log-ratio (clr) transformation was used to normalize the absolute abundance of taxa at each taxonomic level.[37] LinDA using the R package *MicrobiomeStat*[38] was performed with models adjusted for the aforementioned covariates. Log_2 fold change (Log_2 FC), SE, and *P* values for individual taxa were produced. The false discovery rate (FDR) was calculated at each taxonomic level using common and rare taxa to control for multiple testing. An association with an FDR-corrected *P* value <.1 was considered statistically significant. In our study, almost all participants (approximately 94%) who underwent breast cancer surgery also received antibiotic treatment within 1 week postoperatively. To avoid the influence of surgery on the gut microbiome, we restricted our association analyses to 162 participants whose stool samples were collected preoperatively. The R package *MicrobiotaProces* was used to visualize the microbiome data.[39] All statistical analyses were performed using R version 4.2.2.

Ethics Statement

The study was conducted in accordance with the guidelines of the Declaration of Helsinki and approved by the Institutional Review Boards of VNCI (No. 160530; IRB approval issued May 30, 2016) and the Vanderbilt University Medical Center (No. 161039 Vietnam Center of Research Excellence [V-CORE]—IRB approval issued May 31, 2018).

RESULTS

Gut Microbiome Differed for Samples Collected Before and After Breast Cancer Surgery

Compared with patients with postsurgery stool samples (n = 194), patients with presurgery stool samples (n = 162) were similar in age at diagnosis (the mean age of 49.5 *v* 49.6 years) and educational attainment, but more likely to have low income and live in rural areas. No differences were observed between the two groups in menstrual and reproductive history, menopausal status, family history of cancer, BMI levels, comorbidity, dietary intake, or physical activity. However, compared with patients with presurgery stool samples, patients with postsurgery stool samples tended to be diagnosed at earlier stages, were less likely to experience moderate and serious diagnosis delays, and had a lower percentage of HER2-enriched and triple-negative/basal-like subtypes. Among 162 patients with presurgery stool samples, 30.9% received antibiotic treatment during the past 3 months before stool collection, and then 80.9% of them underwent breast cancer surgery after stool collection (Table 1).

Presurgery stool samples had higher α-diversity than postsurgery stool samples, with *P* values being 3.04×10^{-5}, 6.22×10^{-6}, and .0021 for the Chao1, Shannon, and inverse

TABLE 1. Demographic Characteristics and Clinical Factors of Participants by Stool Collection Time at Breast Cancer Diagnosis

Characteristic	Overall (N = 356), No.	Stool Collection Time		*P*[a]
		Presurgery Sample (n = 162), No. (%)	Postsurgery Sample (n = 194), No. (%)	
Age at diagnosis, years	49.5 ± 9.5	49.5 ± 9.6	49.6 ± 9.5	
<40	48	22 (13.6)	26 (13.4)	.86
40-49	135	58 (35.8)	77 (39.7)	
50-59	120	58 (35.8)	62 (32.0)	
≥60	53	24 (14.8)	29 (14.9)	
Education				
Primary school	55	28 (17.3)	27 (13.9)	.18
Middle school	160	76 (46.9)	84 (43.3)	
High school	78	37 (22.8)	41 (21.1)	
College or higher	63	21 (13.0)	42 (21.6)	
Income				
Low (T1)	129	73 (45.1)	56 (28.9)	.003
Middle (T2)	114	49 (30.2)	65 (33.5)	
High (T3)	113	40 (24.7)	73 (37.6)	
Residence				
Urban area	136	46 (28.4)	90 (46.4)	<.001
Rural area	220	116 (71.6)	104 (53.6)	
Age at menarche, years				
≤15	164	73 (45.1)	91 (46.9)	.73
>15	192	89 (54.9)	103 (53.1)	
Regular menstrual cycle				
Regular	272	125 (77.2)	147 (75.8)	.76
Irregular	84	37 (22.8)	47 (24.2)	
Menopausal status				
Premenopausal	196	82 (50.6)	114 (58.8)	.12
Postmenopausal	160	80 (49.4)	80 (41.2)	
No. of live births				
≤1	65	32 (19.8)	33 (17.0)	.18
2	186	76 (46.9)	110 (56.7)	
≥3	105	54 (33.3)	51 (26.3)	
Family history of breast cancer				
No	342	156 (96.3)	186 (95.9)	.84
Yes	14	6 (3.7)	8 (4.1)	
BMI levels, kg/m^2				
Underweight (<18.5)	32	16 (9.9)	16 (8.2)	.69
Normal weight (18.5-22.9)	226	98 (60.5)	128 (66.0)	
Asian overweight (23.0-24.9)	66	31 (19.1)	35 (18.0)	
Asian obese (≥25)	32	17 (10.5)	15 (7.7)	
Comorbidity[b]				
No	290	134 (82.7)	156 (80.4)	.60
Yes	66	28 (17.3)	38 (19.6)	
Antibiotic use within the past 3 months				
No	112	112 (69.1)	0 (0.0)	<.001
Yes	244	50 (30.9)	194 (100.0)	
Diagnosis delay[c]				
No delay (≤3 months)	179	65 (40.1)	114 (58.8)	.001
Moderate delay (4-8 months)	113	58 (35.8)	55 (28.3)	
Serious delay (≥9 months)	64	39 (24.1)	25 (12.9)	

(continued on following page)

TABLE 1. Demographic Characteristics and Clinical Factors of Participants by Stool Collection Time at Breast Cancer Diagnosis (continued)

Characteristic	Overall (N = 356), No.	Stool Collection Time		P^a
		Presurgery Sample (n = 162), No. (%)	Postsurgery Sample (n = 194), No. (%)	
Breast cancer subtypes				
HR+/HER2-negative	144	58 (35.8)	86 (44.3)	.03
HR+/HER2-positive	87	34 (21.0)	53 (27.3)	
HER2-enriched	73	40 (24.7)	33 (17.0)	
Triple-negative/basal-like	52	30 (18.5)	22 (11.3)	
TNM cancer stage				
I	74	14 (8.6)	60 (30.9)	<.001
II	197	76 (46.9)	121 (61.4)	
III-IV	85	72 (44.5)	13 (6.7)	
Breast cancer surgery				
No	31	31 (19.1)	0 (0.0)	<.001
Yes	325	131 (80.9)	194 (100.0)	
Fiber intake, g/d	7.9 ± 4.6	8.6 ± 5.6	7.3 ± 3.4	.053
Fat intake, g/d	36.1 ± 20.7	37.0 ± 23.2	35.4 ± 18.4	.96
Carbohydrate intake, g/d	321 ± 112	327 ± 112	317 ± 112	.27
Protein, g/d	66.9 ± 29.4	67.9 ± 32.8	66.2 ± 26.3	.86
Physical activity, MET hr/wk	116 ± 75	109 ± 63	122 ± 84	.67
Total energy intake, kcal/d	1,847 ± 612	1,878 ± 652	1,822 ± 577	.50

Abbreviations: HER2, human epidermal growth factor receptor 2; HR, hormone receptor; MET, metabolic equivalent.
[a]*P* value was calculated using the chi-square test for categorical variables or the Wilcoxon rank-sum test for continuous variables.
[b]Comorbidities were identified if patients were diagnosed with specific comorbidities, including diabetes mellitus, hypertension, hyperlipidemia, coronary heart disease, stroke, myocardial infarction, arthritis, lupus, and another chronic disease at enrollment.
[c]Diagnosis delay was defined as a total delay time of more than 3 months, measured from the first signs or noticeable breast cancer symptoms to the diagnosis at these two hospitals.

Simpson indices, respectively. The lower α-diversity indices were mainly evident among stool samples collected within 14 days after surgery (postoperative weeks 1-2; n = 95) or 15-21 days after surgery (postoperative week 3; n = 62) in comparison with the presurgery stool samples. No significant difference in α-diversity was found among stool samples that were collected >21 days after surgery (postoperative week 4+; n = 37) when compared with the presurgery stool samples (Fig 1A).

A significant difference in the Bray-Curtis dissimilarity matrix was also found in the stool collection time (P = .001; Fig 1B). Similarly, significant differences were consistently observed for the weighted UniFracs distance and unweighted UniFracs distance matrices (P = .001 for both). The stool collection time explained 1.1%-3.2% of the variations in β-diversity in the overall analysis.

The presurgery stool samples had higher GMHI than the postsurgery stool samples (mean of −0.79 *v* −2.81, $P = 1.3 \times 10^{-11}$). This difference was observed regardless of the time since surgery ($P = 2.0 \times 10^{-7}$, 1.1×10^{-7}, and 8.3×10^{-5}, respectively, for samples collected within postoperative weeks 1-2, 3, and 4+ compared with the presurgery samples; Fig 1C).

Among the 356 stool samples, a total of 21 phyla, 29 classes, 77 orders, 244 families, 1,278 genera, and 4,206 species were identified. At the phylum level, the gut microbiota had high proportions (ie, relative abundance) of *Bacteroidota*, *Proteobacteria*, and *Firmicutes* groups with a dominance of *Firmicutes A* and a smaller proportion of other phyla, including *Actinobacteriota*, *Fusobacteriota*, *Verrucomicrobiota*, *Desulfobacterota A*, *Cyanobacteria*, *Synergistota*, *Elusimicrobiota*, *Spirochaetota*, *Campylobacterota*, and *Patescibacteria* (Fig 1D). After excluding rare taxa with a prevalence of <10% in all samples, a total of 17 phyla, 23 classes, 52 orders, 137 families, 646 genera, and 1,989 species were included to evaluate the associations of individual taxa with selected sociodemographic and clinical factors.

Presurgery stool samples were positively associated with two classes (*Peptococcia* and *Vampirovibrionia*), and four orders (*Flavobacteriales*, *Mycobacteriales*, *Propionibacteriales*, and *Pseudomonadales*) and were inversely associated with the phylum *Firmicutes C*, class *Negativicutes*, and two orders (*Bacillales A* and *Opitutales*). In addition, 77 genera and 353 species showed significant differences between groups with different stool collection times in the differential analysis. Overall, the abundance of 17.7% of the 1,989 investigated species and 15.4% of the 2,864 investigated taxa differed

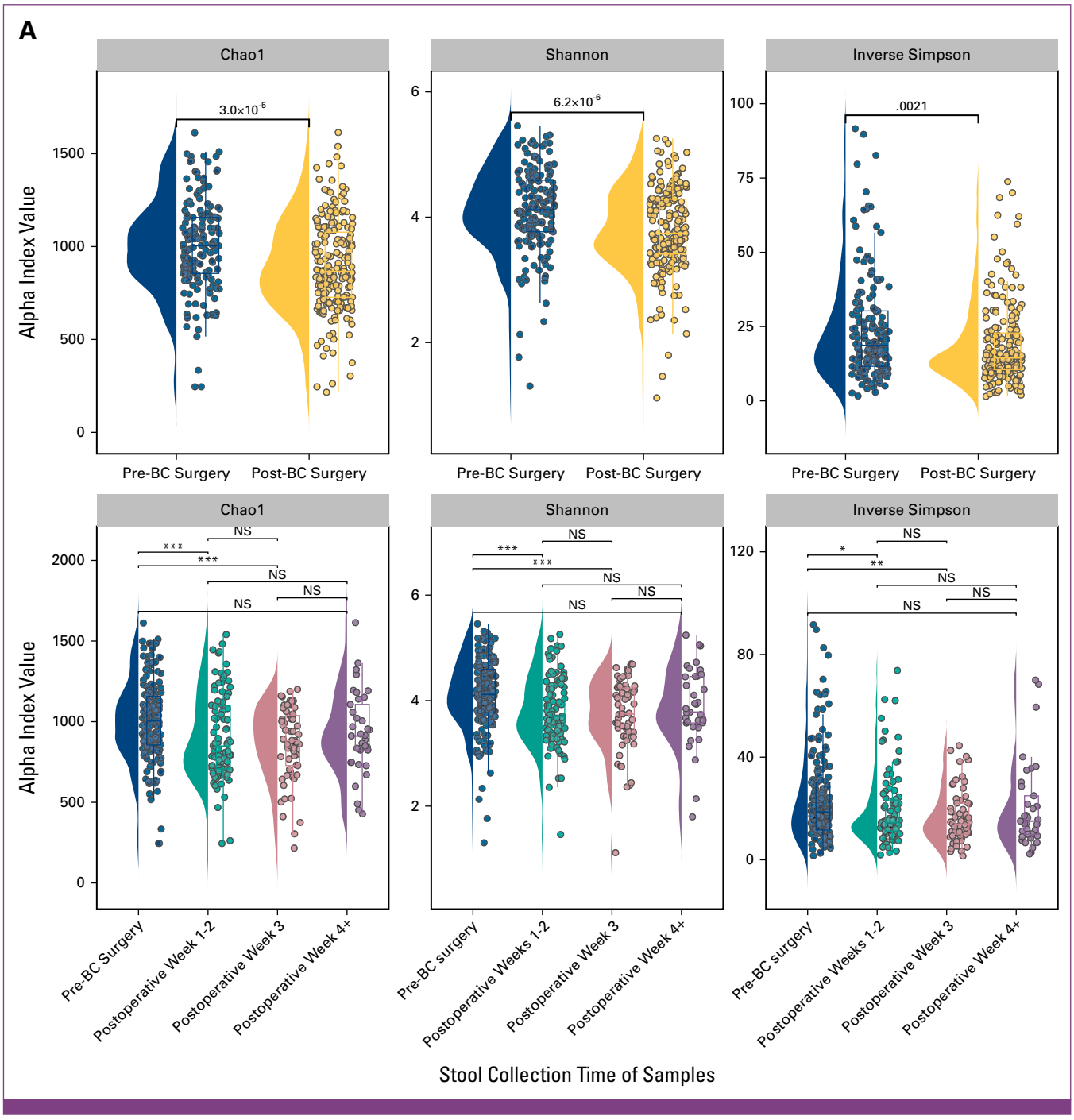

FIG 1. (A) Alpha diversity at the species level (measured in terms of the Chao1 richness, Shannon-Wiener diversity, and inverse Simpson index) by stool collection time. Differences across groups were compared using the Wilcoxon rank-sum test. (B) PCoA of Bray-Curtis diversity by stool collection time. (C) GMHI by stool collection time. Pre-BC surgery: collected stool samples before breast cancer surgery and chemotherapy (n = 162); post-BC surgery: collected stool samples after breast cancer surgery (n = 194). Postoperative Weeks 1-2: collected stool samples within 14 days after breast cancer surgery (n = 95); Postoperative Week 3: collected stool samples within 15-21 days after breast cancer surgery (n = 62); Postoperative Week 4+: collected stool samples after breast cancer surgery >21 days (n = 37). Differences across groups were compared using the Wilcoxon rank-sum test. **P* < .05; ***P* < .01; ****P* < .001. (D) Relative abundance by stool collection time. (E) Volcano plot of different abundance analyses for gut microbial taxa by stool collection time. A linear regression model was conducted for clr (centered log ratio)-transformed taxa abundance with adjustment for age group, income levels, residence, age at menarche, regular menstrual cycle, menopausal status, number of live births, BMI levels, comorbidity, antibiotic use, diagnosis delay, breast cancer subtypes, TNM cancer stage, fiber intake, fat intake, carbohydrate intake, protein intake, and physical activity. FDR for Log2 fold change (Log2 FC) was calculated at each taxonomic level and by common and rare taxa. Common taxa: prevalence ≥50% in the population; rare taxa: 10% ≤prevalence< 50% in the population. BC, breast cancer; FDR, false discovery rate; GMHI, gut microbiome health index; NS, nonsignificance; PCoA, Principal coordinate analysis; PERMANOVA, permutational multivariate analysis of variance. (continued on following page)

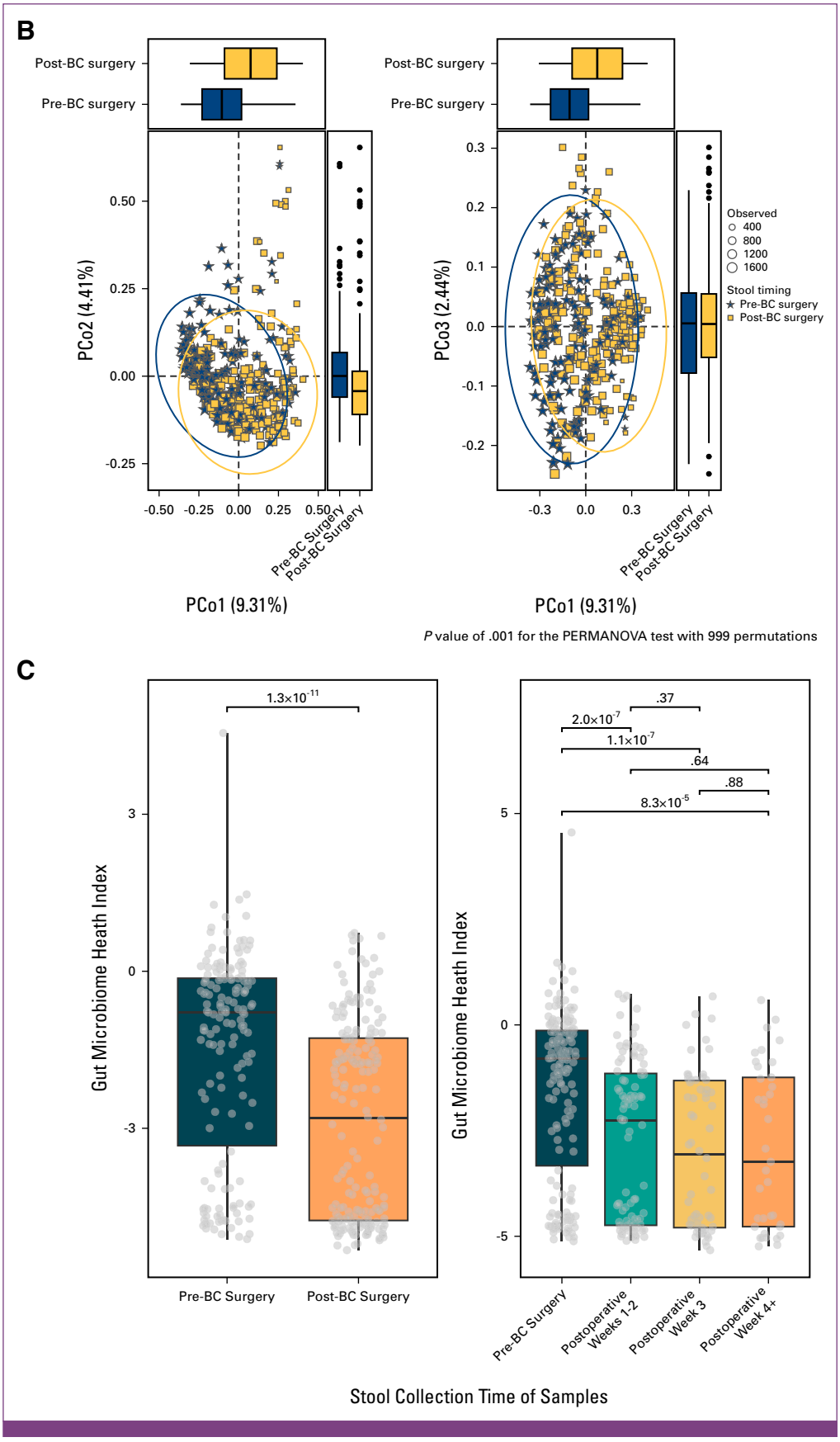

FIG 1. (Continued).

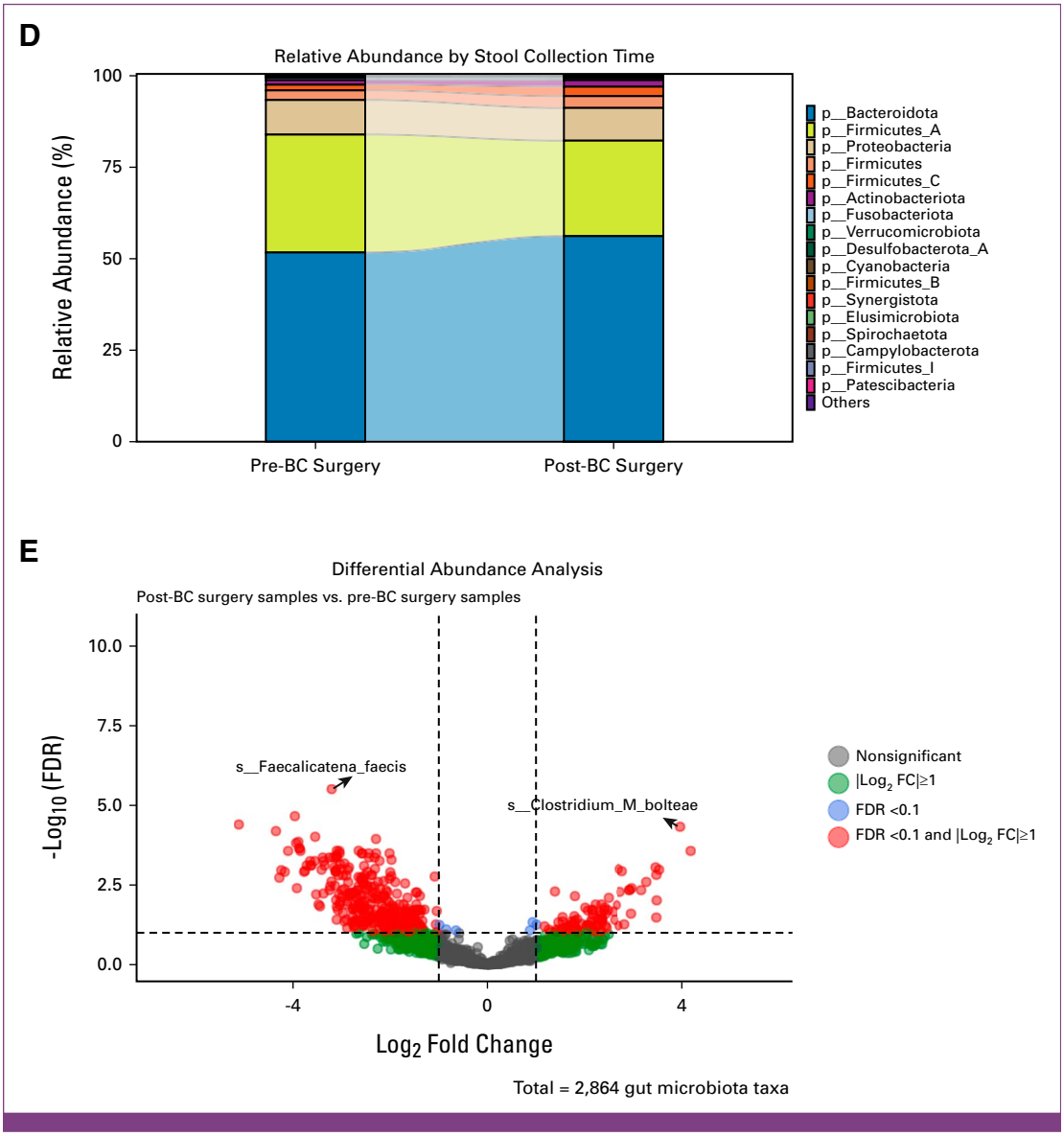

FIG 1. (Continued).

significantly by stool collection time after FDR correction (FDR < 0.1; Fig 1E).

Association of Gut Microbial Richness and Composition With Nonclinical and Clinical Factors

Given the influence of breast cancer surgery on the gut microbiome, we restricted our analyses to 162 participants who provided presurgery samples.

Multivariable linear regression analysis showed that all α-diversity indices were significantly decreased among patients who experienced a moderate or serious diagnosis delay, compared with the no-delay group ($P < .05$). In addition, the highest consumption of fiber (T3) was significantly associated with a decrease in all α-diversity indices compared with the lowest consumption of fiber (T1; $P < .05$). We did not find a significant association between α-diversity indices and other sociodemographic characteristics, lifestyle factors, or clinical features, including comorbidity, antibiotic use, breast cancer subtype, and TNM stage (Table 2).

PERMANOVA showed that income, diagnosis delay, and carbohydrate and protein intake were significant factors associated with β-diversity. Income level explained 3.3% of the variations in the Bray-Curtis dissimilarity matrix, whereas diagnosis delay significantly explained 3.3% of the variations in the unweighted UniFracs distance matrix. Carbohydrate intake significantly explained 1.3%-2.0% of the variations in the unweighted and weighted UniFracs

TABLE 2. Association of Selected Nonclinical and Clinical Factors With Alpha Diversity Indices Among Patients With Presurgery Stool Samples (n = 162)

	Chao1 Richness Index		Shannon-Wiener Diversity index[a]		Inverse Simpson Index[b]	
Characteristic	Model 1,[c] β (95% CI)	Model 2,[d] β (95% CI)	Model 1,[c] β (95% CI)	Model 2,[d] β (95% CI)	Model 1,[c] β (95% CI)	Model 2,[d] β (95% CI)
Age at diagnosis, years						
<40	0.0	0.0	0.00	0.00	0.00	0.00
40-49	−33.4 (−156.3 to 89.5)	−135.3 (−270.4 to −0.2)	−0.27 (−2.89 to 2.35)	−2.23 (−5.08 to 0.62)	0.09 (−0.73 to 0.91)	−0.34 (−1.24 to 0.55)
50-59	−24.2 (−147.1 to 98.7)	−137.9 (−303.4 to 27.6)	−0.47 (−3.09 to 2.15)	−2.82 (−6.31 to 0.67)	−0.11 (−0.93 to 0.71)	−0.71 (−1.81 to 0.39)
60+	−96.6 (−241.4 to 48.3)	−215.1 (−422.5 to −7.6)	−1.03 (−4.12 to 2.06)	−3.44 (−7.82 to 0.93)	0.03 (−0.94 to 0.99)	−0.56 (−1.94 to 0.82)
Income						
Low (T1)	0.0	0.0	0.00	0.00	0.00	0.00
Middle (T2)	−18.3 (−109.2 to 72.5)	−52.1 (−145.0 to 40.8)	0.03 (−1.90 to 1.96)	−0.82 (−2.78 to 1.14)	0.28 (−0.32 to 0.88)	0.01 (−0.61 to 0.63)
High (T3)	−18.7 (−115.5 to 78.1)	−33.1 (−134.0 to 67.8)	0.09 (−1.96 to 2.15)	−0.07 (−2.20 to 2.06)	0.27 (−0.37 to 0.91)	0.19 (−0.48 to 0.86)
Residence						
Urban area	0.0	0.0	0.00	0.00	0.00	0.00
Rural area	−18.7 (−104.2 to 66.8)	−35.2 (−126.8 to 56.4)	−0.13 (−1.95 to 1.68)	0.07 (−1.86 to 2.00)	−0.12 (−0.69 to 0.44)	0.11 (−0.50 to 0.72)
Age at menarche, years						
≤15	0.0	0.0	0.00	0.00	0.00	0.00
>15	65.0 (−11.9 to 141.8)	90.9 (5.6 to 176.2)	1.03 (−0.61 to 2.67)	1.46 (−0.34 to 3.26)	0.16 (−0.36 to 0.67)	0.24 (−0.32 to 0.81)
Regular menstrual cycle						
Regular	0.0	0.0	0.00	0.00	0.00	0.00
Irregular	28.5 (−63.2 to 120.3)	2.0 (−97.2 to 101.2)	0.11 (−1.84 to 2.06)	−0.13 (−2.22 to 1.96)	−0.14 (−0.74 to 0.47)	−0.03 (−0.69 to 0.63)
Menopausal status						
Premenopausal	0.0	0.0	0.00	0.00	0.00	0.00
Postmenopausal	3.4 (−73.7 to 80.5)	62.8 (−62.3 to 187.9)	0.37 (−1.26 to 2.01)	1.64 (−0.99 to 4.28)	0.22 (−0.29 to 0.73)	0.62 (−0.21 to 1.45)
No. of live births						
≤1	0.0	0.0	0.00	0.00	0.00	0.00
2	4.5 (−99.2 to 108.2)	5.9 (−105.9 to 117.8)	0.10 (−2.10 to 2.30)	0.15 (−2.21 to 2.51)	0.09 (−0.60 to 0.77)	0.07 (−0.67 to 0.82)
≥3	21.1 (−88.7 to 130.9)	−13.4 (−134.9 to 108.1)	0.37 (−1.96 to 2.70)	−0.43 (−2.99 to 2.13)	0.01 (−0.71 to 0.74)	−0.27 (−1.08 to 0.54)
BMI levels						
Normal weight	0.0	0.0	0.00	0.00	0.00	0.00
Underweight	−89.9 (−222.1 to 42.2)	−117.4 (−258.0 to 23.2)	−2.20 (−5.00 to 0.60)	−2.42 (−5.39 to 0.54)	−0.31 (−1.19 to 0.57)	−0.34 (−1.28 to 0.59)
Asian overweight	−47.3 (−148.3 to 53.7)	11.2 (−99.6 to 121.9)	−1.06 (−3.20 to 1.08)	−0.39 (−2.73 to 1.95)	−0.19 (−0.86 to 0.48)	−0.26 (−1.00 to 0.47)
Asian obese	5.8 (−123.0 to 134.5)	85.1 (−57.0 to 227.2)	−0.03 (−2.76 to 2.70)	0.94 (−2.06 to 3.94)	0.20 (−0.65 to 1.06)	0.41 (−0.53 to 1.36)
Comorbidity						
No	0.0	0.0	0.00	0.00	0.00	0.00
Yes	−34.6 (−136.4 to 67.3)	−62.4 (−184.3 to 59.5)	0.05 (−2.12 to 2.21)	−1.00 (−3.57 to 1.58)	0.08 (−0.59 to 0.75)	−0.46 (−1.27 to 0.35)
Antibiotic use within the past 3 months						
No	0.0	0.0	0.00	0.00	0.00	0.00
Yes	5.5 (−77.9 to 89.0)	24.5 (−61.1 to 110.2)	−0.11 (−1.88 to 1.66)	0.49 (−1.32 to 2.29)	−0.03 (−0.58 to 0.52)	0.16 (−0.41 to 0.73)

(continued on following page)

TABLE 2. Association of Selected Nonclinical and Clinical Factors With Alpha Diversity Indices Among Patients With Presurgery Stool Samples (n = 162) (continued)

	Chao1 Richness Index		Shannon-Wiener Diversity index[a]		Inverse Simpson Index[b]	
Characteristic	Model 1,[c] β (95% CI)	Model 2,[d] β (95% CI)	Model 1,[c] β (95% CI)	Model 2,[d] β (95% CI)	Model 1,[c] β (95% CI)	Model 2,[d] β (95% CI)
Diagnosis delay						
No delay	0.0	0.0	0.00	0.00	0.00	0.00
Moderate delay	−89.4 (−174.3 to −4.5)	−111.5 (−209.9 to −13.1)	−1.92 (−3.70 to −0.13)	−1.96 (−4.04 to 0.11)	−0.76 (−1.32 to −0.21)	−0.79 (−1.44 to −0.13)
Serious delay	−190.3 (−285.5 to −95.1)	−200.7 (−303.0 to −98.3)	−4.42 (−6.42 to −2.42)	−4.37 (−6.53 to −2.21)	−1.37 (−1.99 to −0.75)	−1.38 (−2.06 to −0.70)
Breast cancer subtypes						
HR+/HER2-negative	0.0	0.0	0.00	0.00	0.00	0.00
HR+/HER2-positive	−20.5 (−126.8 to 85.8)	−70.5 (−187.9 to 47.0)	−0.96 (−3.21 to 1.30)	−2.15 (−4.62 to 0.33)	−0.45 (−1.15 to 0.25)	−0.65 (−1.43 to 0.13)
HER2-enriched	−29.9 (−131.1 to 71.2)	9.2 (−100.6 to 118.9)	−0.95 (−3.10 to 1.20)	−0.27 (−2.58 to 2.05)	−0.26 (−0.92 to 0.41)	−0.12 (−0.85 to 0.61)
Triple-negative	30.1 (−80.6 to 140.8)	18.8 (−102.8 to 140.4)	−0.48 (−2.83 to 1.87)	−0.74 (−3.31 to 1.82)	−0.29 (−1.02 to 0.44)	−0.27 (−1.08 to 0.54)
TNM cancer stage						
I	0.0	0.0	0.00	0.00	0.00	0.00
II	−43.1 (−185.8 to 99.6)	−30.4 (−173.4 to 112.7)	−0.41 (−3.44 to 2.61)	−0.31 (−3.32 to 2.71)	−0.09 (−1.03 to 0.85)	−0.11 (−1.06 to 0.84)
III-IV	−70.4 (−213.7 to 73.0)	−11.2 (−157.8 to 135.5)	−1.28 (−4.32 to 1.76)	−0.16 (−3.26 to 2.93)	−0.37 (−1.31 to 0.58)	−0.04 (−1.01 to 0.94)
Fiber intake, g/d						
Low (T1)	0.0	0.0	0.00	0.00	0.00	0.00
Middle (T2)	−63.1 (−160.9 to 34.8)	−40.6 (−152.9 to 71.6)	−1.68 (−3.76 to 0.40)	−1.32 (−3.69 to 1.04)	−0.62 (−1.27 to 0.03)	−0.55 (−1.30 to 0.20)
High (T3)	−90.2 (−182.9 to 2.5)	−164.2 (−293.2 to −35.1)	−1.78 (−3.75 to 0.18)	−3.74 (−6.46 to −1.02)	−0.43 (−1.05 to 0.18)	−1.01 (−1.86 to −0.15)
Fat intake, g/d						
Low (T1)	0.0	0.0	0.00	0.00	0.00	0.00
Middle (T2)	−67.7 (−162.7 to 27.3)	−85.2 (−200.8 to 30.4)	−0.90 (−2.92 to 1.12)	−1.19 (−3.63 to 1.25)	0.04 (−0.59 to 0.67)	0.05 (−0.72 to 0.82)
High (T3)	−6.4 (−98.0 to 85.2)	−44.2 (−185.6 to 97.1)	0.49 (−1.45 to 2.44)	−0.29 (−3.27 to 2.69)	0.26 (−0.35 to 0.87)	0.07 (−0.87 to 1.01)
Carbohydrate intake, g/d						
Low (T1)	0.0	0.0	0.00	0.00	0.00	0.00
Middle (T2)	−109.4 (−206.1 to −12.6)	−106.4 (−222.3 to 9.5)	−2.65 (−4.69 to −0.61)	−2.60 (−5.04 to −0.15)	−0.93 (−1.56 to −0.30)	−0.94 (−1.71 to −0.17)
High (T3)	−18.7 (−110.6 to 73.3)	−33.8 (−176.4 to 108.8)	−0.35 (−2.29 to 1.58)	−1.35 (−4.36 to 1.66)	−0.19 (−0.80 to 0.41)	−0.59 (−1.54 to 0.36)
Protein intake, g/d						
Low (T1)	0.0	0.0	0.00	0.00	0.00	0.00
Middle (T2)	−37.0 (−132.3 to 58.4)	59.9 (−61.8 to 181.5)	−0.76 (−2.77 to 1.25)	1.20 (−1.37 to 3.76)	−0.17 (−0.79 to 0.46)	0.35 (−0.46 to 1.16)
High (T3)	40.2 (−52.6 to 132.9)	153.2 (−25.3 to 331.8)	1.44 (−0.51 to 3.40)	3.68 (−0.08 to 7.45)	0.50 (−0.10 to 1.11)	1.06 (−0.13 to 2.25)
Physical activity, MET hr/wk						
Low (T1)	0.0	0.0	0.00	0.00	0.00	0.00
Middle (T2)	−18.4 (−111.9 to 75.0)	−44 (−144.4 to 56.3)	−0.61 (−2.59 to 1.37)	−1.02 (−3.13 to 1.10)	−0.25 (−0.87 to 0.36)	−0.37 (−1.04 to 0.30)
High (T3)	80.0 (−12.5 to 172.6)	60.7 (−41.2 to 162.7)	1.67 (−0.29 to 3.63)	1.48 (−0.68 to 3.63)	0.38 (−0.23 to 0.99)	0.33 (−0.35 to 1.01)

Abbreviations: HER2, human epidermal growth factor receptor 2; HR, hormone receptor; MET, metabolic equivalent.

[a]Shannon index was transformed to the square of Shannon index.

[b]Inverse Simpson index was transformed to the square root of inverse Simpson index.

[c]Multivariable model 1: univariate linear regression model.

[d]Multivariable model 2 was multivariable model 1 with additional adjustment for the age group, income levels, residence, age at menarche, regular menstrual cycle, menopausal status, number of live births, BMI levels, comorbidity, antibiotic use, diagnosis delay, breast cancer subtypes, TNM cancer stage, fiber intake, fat intake, carbohydrate intake, protein intake, and physical activity.

TABLE 3. PERMANOVA Test Difference of Beta Diversity Between Demographic Characteristics and Clinical Factors Among Patients With Presurgery Stool Samples (n = 162)

	Bray-Curtis Dissimilarity Matrix		Unweighted UniFracs Distance Matrix		Weighted UniFracs Distance Matrix	
	R^2, % (*P* for the PERMANOVA test)		R^2, % (*P* for the PERMANOVA test)		R^2, % (*P* for the PERMANOVA test)	
Characteristic	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2
Age groups at diagnosis	2.2 (.18)	2.2 (.14)	1.9 (.42)	1.9 (.36)	2.3 (.24)	2.2 (.26)
Income levels	3.5 (.001)	3.3 (.001)	2.0 (.07)	1.9 (.07)	1.6 (.25)	1.6 (.24)
Residence	1.4 (.012)	1.0 (.06)	0.7 (0.22)	0.7 (.27)	0.9 (.19)	0.8 (.25)
Age at menarche	0.4 (.77)	0.5 (.67)	0.7 (.28)	0.7 (.27)	0.4 (.54)	0.4 (.64)
Regular menstrual cycle	0.5 (.59)	0.5 (.60)	0.3 (.96)	0.3 (.95)	0.5 (.44)	0.4 (.54)
Menopausal status	0.5 (.66)	0.5 (.60)	0.3 (.95)	0.5 (.61)	0.4 (.69)	0.4 (.57)
No. of live births	1.7 (.07)	1.3 (.30)	1.4 (.27)	1.4 (.24)	0.7 (.80)	0.4 (.98)
BMI levels	2.2 (.17)	2.1 (.20)	2.4 (.15)	2.4 (.12)	2.9 (.11)	2.4 (.22)
Comorbidity	0.7 (.23)	0.7 (.20)	0.6 (.31)	0.6 (.37)	0.3 (.82)	0.3 (.73)
Antibiotic use within the past 3 months	0.5 (.63)	0.5 (.58)	0.4 (.78)	0.4 (.70)	0.6 (.37)	0.7 (.37)
Diagnosis delay	1.5 (.15)	1.8 (.06)	3.0 (.002)	3.3 (.002)	1.1 (.48)	1.1 (.49)
Breast cancer subtypes	1.5 (.80)	1.2 (.98)	1.4 (.89)	1.3 (.91)	1.3 (.79)	1.0 (.88)
TNM cancer stags	1.0 (.72)	1.0 (.72)	1.0 (.65)	0.8 (.83)	1.4 (.32)	1.1 (.50)
Fiber intake	1.2 (.41)	1.2 (.33)	1.4 (.27)	1.5 (.15)	0.9 (.62)	0.7 (.79)
Fat intake	1.2 (.51)	1.2 (.41)	1.1 (.53)	1.3 (.33)	0.5 (.96)	0.7 (.79)
Carbohydrate intake	1.5 (.14)	1.7 (.06)	1.8 (.07)	2.0 (.039)	1.1 (.47)	1.3 (.42)
Protein intake	1.7 (.08)	1.9 (.036)	1.5 (.22)	1.4 (.24)	1.1 (.46)	2.3 (.09)
Physical activity	1.3 (.38)	1.1 (.65)	1.5 (.23)	1.2 (.43)	1.5 (.28)	1.2 (.40)

NOTE. Model 1 was the univariate PERMANOVA model. Model 2 was model 1 with additional adjustments for age group, income levels, residence, age at menarche, regular menstrual cycle, menopausal status, number of live births, BMI levels, comorbidity, antibiotic use, diagnosis delay, breast cancer subtypes, TNM cancer stage, fiber intake, fat intake, carbohydrate intake, protein intake, and physical activity.
Abbreviation: PERMANOVA, permutational multivariate analysis of variance.

distance matrix, whereas 1.9% of the variations in the Bray-Curtis dissimilarity matrix was associated with protein intake (all *P* for *PERMANOVA* <.05). No nonclinical and clinical factors were significant factors associated with all three β-diversities (Table 3).

Participants with high income had a significantly lower GMHI than participants with the lowest income ($\beta_{T2\ v\ T1}$ = −0.87 [95% CI, −1.64 to −0.09]; $\beta_{T3\ v\ T1}$ = −0.88 [95% CI, −1.73 to −0.04]). We did not find a significant association between GMHI and other sociodemographic characteristics, lifestyle factors, or clinical features (Table 4).

Association of Gut Microbial Taxa With Selected Sociodemographic and Clinical Factors

Stage II patients had a significantly lower abundance of the phylum *Synergistota* (Log_2 FC [SE] = −3.57 [1.35]; *P* = .009; FDR = 0.047), compared with stage I patients. The association was driven by the class *Synergistia* (Log_2 FC [SE] = −3.50 [1.33]; *P* = .009; FDR = 0.084). The abundance of the order *Haloplasmatales* was significantly higher in stage III-IV patients than in stage I patients (Log_2 FC [SE] = 5.37 [1.59]; $P = 9.87 \times 10^{-4}$; FDR = 0.034; Fig 2 and Appendix Table A3).

The genus *Acutalibacter* was more abundant in patients with HER2-enriched breast cancer than in those with HR+/HER2-negative breast cancer (Log_2 FC [SE] = 2.87 [0.77]; $P = 3.06 \times 10^{-4}$; FDR = 0.084). Patients with triple-negative breast cancer had a higher abundance of *Dorea spoo1185345* (Log_2 FC [SE] = 2.94 [0.64]; $P = 1.05 \times 10^{-5}$; FDR = 0.013) but a lower abundance of species *MGYG-HGUT-01722* (Log_2 FC [SE] = −3.80 [0.89]; $P = 4.04 \times 10^{-5}$; FDR = 0.025) than patients with HR+/HER2-negative breast cancer. In addition, five significant differences in the abundance of gut microbial taxa were found between HR+/HER2-positive breast cancer and HR+/HER2-negative breast cancers. Compared with those with HR+/HER2-negative breast cancer, the phylum *Cyanobacteria* was less abundant in patients with HR+/HER2-positive breast cancer (Log_2 FC [SE] = −0.99 [0.36]; *P* = .007; FDR = 0.089), whereas the phylum *Elusimicrobiota* was more abundant (Log_2 FC [SE] = 2.24 [0.85]; *P* = .009; FDR = 0.046), which was driven by the class *Elusimicrobia* (Log_2 FC [SE] = 2.32 [0.85]; *P* = .008; FDR = 0.067; Fig 2 and Appendix Table A4).

TABLE 4. Association of Selected Demographic Characteristics and Clinical Factors With the GMHI Among Patients With Presurgery Stool Samples (n = 162)

		GMHI	
Characteristic	Mean ± SD	Model 1,[a] β (95% CI)	Model 2,[b] β (95% CI)
Age at diagnosis, years			
<40	−0.85 ± 2.38	0.0	0.0
40-49	−0.74 ± 3.76	0.15 (−0.79 to 1.10)	−0.13 (−1.27 to 1.00)
50-59	−0.54 ± 1.90	0.47 (−0.47 to 1.42)	0.24 (−1.14 to 1.63)
60+	−3.10 ± 3.89	−1.19 (−2.30 to −0.07)	−1.58 (−3.32 to 0.15)
Income			
Low (T1)	−0.32 ± 2.25	0.0	0.0
Middle (T2)	−1.07 ± 3.12	−0.67 (−1.38 to 0.04)	−0.87 (−1.64 to −0.09)
High (T3)	−1.12 ± 3.90	−0.95 (−1.71 to −0.20)	−0.88 (−1.73 to −0.04)
Location			
Urban area	−0.85 ± 2.42	0.0	0.0
Rural area	−0.77 ± 3.45	0.02 (−0.66 to 0.70)	−0.41 (−1.18 to 0.36)
Age at menarche, years			
≤15	−0.99 ± 3.96	0.0	0.0
>15	−0.73 ± 2.12	0.39 (−0.22 to 1.00)	0.23 (−0.49 to 0.94)
Regular menstrual cycle			
Regular	−0.88 ± 2.85	0.0	0.0
Irregular	−0.73 ± 3.81	0.24 (−0.49 to 0.97)	−0.05 (−0.88 to 0.78)
Menopausal status			
Premenopausal	−0.73 ± 2.72	0.0	0.0
Postmenopausal	−1.06 ± 3.55	−0.28 (−0.89 to 0.34)	0.08 (−0.96 to 1.13)
No. of live births			
≤1	−1.36 ± 4.45	0.0	0.0
2	−0.78 ± 2.63	0.56 (−0.26 to 1.37)	0.33 (−0.61 to 1.27)
≥3	−0.74 ± 2.30	0.71 (−0.16 to 1.58)	0.49 (−0.52 to 1.51)
BMI levels			
Normal weight	−0.72 ± 2.05	0.0	0.0
Underweight	−1.90 ± 3.60	−0.86 (−1.90 to 0.18)	−0.97 (−2.15 to 0.21)
Asian overweight	−1.39 ± 4.33	−0.91 (−1.70 to −0.12)	−0.38 (−1.31 to 0.55)
Asian obese	−0.16 ± 3.09	−0.06 (−1.07 to 0.96)	0.46 (−0.73 to 1.65)
Comorbidity			
No	−0.77 ± 2.61	0.0	0.0
Yes	−1.40 ± 4.17	−0.75 (−1.55 to 0.05)	−0.45 (−1.47 to 0.57)
Antibiotic use within the past 3 months			
No	−0.75 ± 3.90	0.0	0.0
Yes	−1.02 ± 2.63	0.00 (−0.66 to 0.66)	0.16 (−0.56 to 0.88)
Diagnosis delay			
No delay	−0.82 ± 2.35	0.0	0.0
Moderate delay	−0.62 ± 3.83	0.02 (−0.68 to 0.73)	−0.13 (−0.95 to 0.70)
Serious delay	−1.13 ± 3.86	−0.34 (−1.13 to 0.45)	−0.53 (−1.38 to 0.33)
Breast cancer subtypes			
HR+/HER2-negative	−1.15 ± 3.34	0.0	0.0
HR+/HER2-positive	−0.53 ± 2.77	0.17 (−0.68 to 1.01)	−0.25 (−1.24 to 0.73)
HER2-enriched	−0.78 ± 2.98	0.00 (−0.80 to 0.81)	0.13 (−0.79 to 1.05)
Triple-negative	−0.33 ± 2.66	0.29 (−0.59 to 1.17)	0.17 (−0.84 to 1.19)

(continued on following page)

TABLE 4. Association of Selected Demographic Characteristics and Clinical Factors With the GMHI Among Patients With Presurgery Stool Samples (n = 162) (continued)

		GMHI	
Characteristic	Mean ± SD	Model 1,[a] β (95% CI)	Model 2,[b] β (95% CI)
TNM cancer stage			
I	−0.66 ± 0.74	0.0	0.0
II	−0.75 ± 4.12	−0.36 (−1.50 to 0.77)	−0.31 (−1.51 to 0.89)
III-IV	−1.12 ± 2.87	−0.47 (−1.61 to 0.67)	−0.39 (−1.62 to 0.84)
Fiber intake, g/d			
Low (T1)	−0.79 ± 2.23	0.0	0.0
Middle (T2)	−0.61 ± 1.89	0.18 (−0.61 to 0.96)	0.01 (−0.93 to 0.95)
High (T3)	−1.07 ± 4.02	−0.19 (−0.94 to 0.55)	−0.60 (−1.68 to 0.48)
Fat intake, g/d			
Low (T1)	−0.76 ± 3.13	0.0	0.0
Middle (T2)	−0.97 ± 3.97	−0.18 (−0.93 to 0.58)	−0.42 (−1.38 to 0.55)
High (T3)	−0.79 ± 2.17	0.24 (−0.5 to 0.97)	−0.12 (−1.31 to 1.06)
Carbohydrate intake, g/d			
Low (T1)	−0.80 ± 3.45	0.0	0.0
Middle (T2)	−1.10 ± 3.57	−0.17 (−0.95 to 0.61)	−0.40 (−1.37 to 0.57)
High (T3)	−0.75 ± 2.39	0.23 (−0.51 to 0.97)	0.04 (−1.16 to 1.23)
Protein, g/d			
Low (T1)	−1.09 ± 3.98	0.0	0.0
Middle (T2)	−0.77 ± 2.46	0.31 (−0.45 to 1.07)	0.68 (−0.33 to 1.70)
High (T3)	−0.79 ± 2.33	0.45 (−0.29 to 1.19)	0.77 (−0.72 to 2.27)
Physical activity, MET hr/wk			
Low (T1)	−0.91 ± 3.47	0.0	0.0
Middle (T2)	−0.96 ± 3.76	−0.03 (−0.79 to 0.72)	−0.09 (−0.93 to 0.75)
High (T3)	−0.50 ± 2.64	0.27 (−0.48 to 1.01)	−0.10 (−0.96 to 0.75)

Abbreviations: GMHI, gut microbiome health index; HER2, human epidermal growth factor receptor 2; HR, hormone receptor; MET, metabolic equivalent; SD, standard deviation.

[a]Multivariable model 1: univariate linear regression model.

[b]Multivariable model 2 was multivariable model 1 with additional adjustment for age group, income levels, residence, age at menarche, regular menstrual cycle, menopausal status, number of live births, BMI levels, comorbidity, antibiotic use, diagnosis delay, breast cancer subtypes, TNM cancer stage, fiber intake, fat intake, carbohydrate intake, protein intake, and physical activity.

DISCUSSION

In this study of 356 Vietnamese patients with breast cancer, we found significant differences in both α- and β-diversities and in the abundances of over 15% of the investigated gut microbial taxa between stool samples collected pre- and postsurgery. We also found that the GMHI was low among patients with breast cancer, particularly those with post-surgery stool samples. These results suggest that breast cancer surgery in Vietnam, which is almost always accompanied by prophylactic antibiotic treatment, might have a profound impact on the gut microbiome of patients with cancer.

Few studies have investigated the gut microbiome and breast cancer characteristics. A study of 31 European patients found that the percentage and absolute abundance of the *Bacteroidetes*, *Clostridium leptum cluster*, *Clostridium coccoides cluster*, *Facecalibacterium prausnitzii*, and *Blautia* species were significantly higher in patients with clinical stage II and III cancer than patients with clinical stage 0 and I.[14] Moreover, they reported that overweight and obese patients had a significant decline in the abundance of the *Firmicutes*, *Facecalibacterium prausnitzii*, *Eggerthella lenta*, and *Blautia* species, when compared with patients with a normal BMI.[14] In addition, a Taiwanese study involving 200 patients and 67 age-matched controls showed that premenopausal patients had significantly higher levels of *Anaerostipes* and *Bacteroides fragilis*, whereas postmenopausal patients had significantly higher *Proteobacteria* and *Klebsiella pneumoniae*.[16] Moreover, α-diversity was significantly reduced in premenopausal patients with breast cancer and β-diversity differed significantly only between patients with breast cancer and controls.[16] However, in analyses restricted to the 162 patients whose stool samples were collected preoperatively, we observed lower α-diversity indices among stage II and stage III-IV patients than among patients

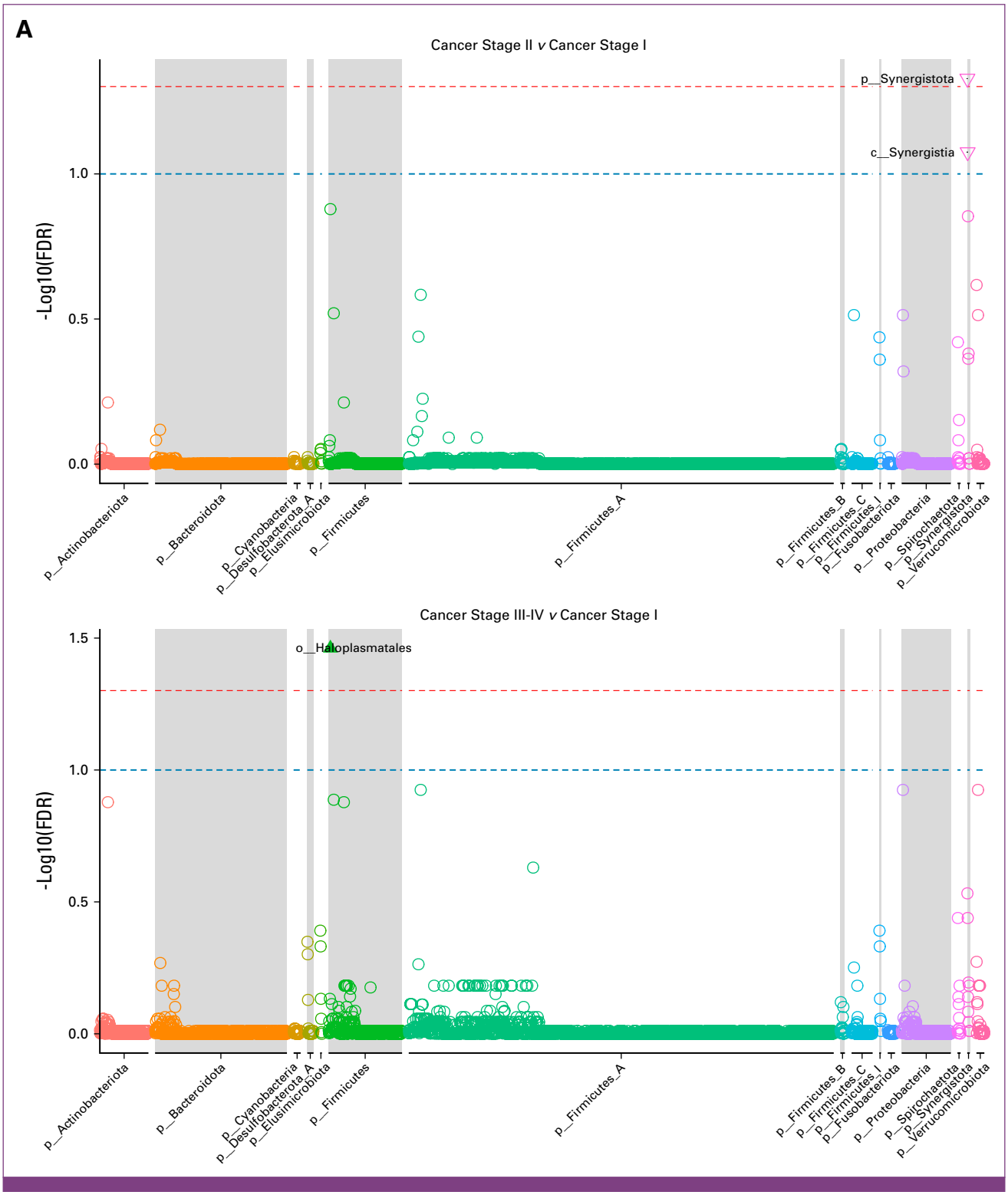

FIG 2. (A) Manhattan plots of different abundances of gut microbial taxa between cancer stages among participants with presurgery stool samples (n = 162). (B) Manhattan plots of different abundances of gut microbial taxa between breast cancer subtypes among participants with presurgery stool samples (n = 162). FDR for Log_2 FC was calculated at each taxonomic level by common and rare taxa. Red horizontal line, FDR = 0.05; blue horizontal line, FDR = 0.10. FDR, false discovery rate. (continued on following page)

with stage I although no significant association was found. Moreover, we found a significantly lower abundance of the phylum *Synergistota* and the class *Synergistia* among stage II patients and a higher abundance of the order *Haloplasmatales* among stage III-IV patients than in stage I patients. In our study, no significant association with α-diversity indices as β-diversity was observed for menopausal status, menstrual history, reproductive factors, or BMI levels. Furthermore, no

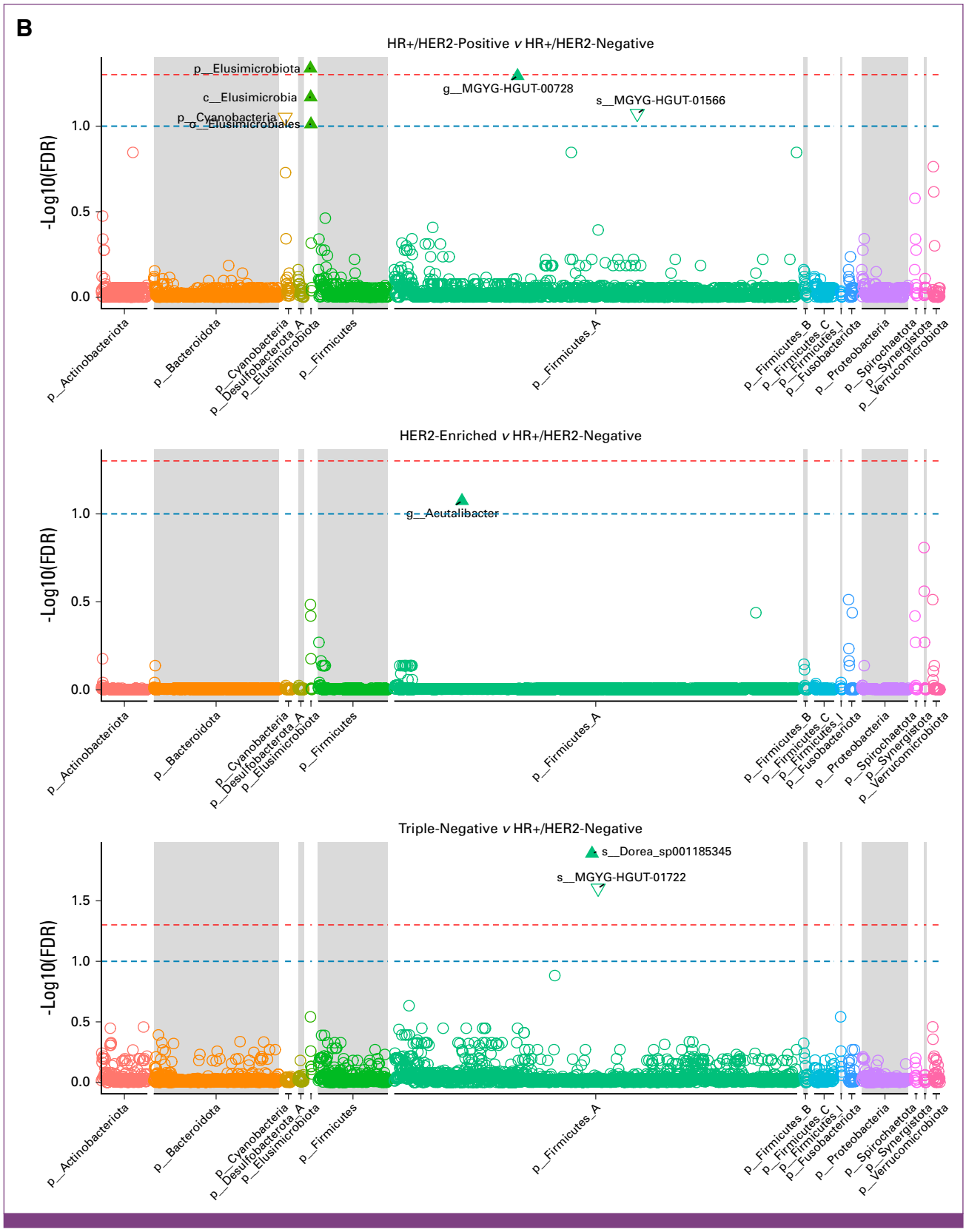

FIG 2. (Continued).

gut microbial taxa were associated with menopausal status (all FDR > 0.1) However, underweight patients had a significantly lower abundance of orders *Acidaminococcales* along with its family *Acidaminococcaceae* and a higher abundance of orders *Lactobacillales*, *Peptostreptococcales*, and some species such as *Clostridioides difficile A*, *Enterococcus A raffinosus*, *Enterococcus B durans*, and *Terrisporobacter othiniensis*. *Clostridioides difficile* is an intestinal pathogen that causes severe diarrhea.[40] In addition, we found that the genus *Agathobaculum* is significantly less abundant among obese patients. *Agathobaculum butyriciproducenes*, a butyrate- and lactate-producing bacterium linked to healthy plant-based foods, is a member of the genus *Agathobaculum* (Appendix Table A5).

Early studies reported that patients with different clinicopathologic factors showed different gut microbiome profiles.[17,18] In a study of 83 patients with invasive ductal breast carcinoma, Yang et al[17] found that members of the family *Prevotellaceae* were more abundant in patients with PR+ or ER+ tumors, whereas some bacteria, including *Hydrogenophilus*, *Lactobacillus*, and *Acinetobacter*, were more abundant in patients with breast cancer with PR− and ER− tumors. Furthermore, *Megasphaera* was enriched in patients with ER+ and HER2-positive tumors. *Lactobacillus*, *Clostridium*, and *Clostridiaceae* were enriched in patients with low Ki-67 expression (Ki-67 <30%), whereas *Ruminiclostridium*, *Tenericutes*, and *Mollicutes* were enriched in patients with high Ki-67 expression (Ki-67 ≥30%). A $P < .05$ was considered statistically significant for discriminant analysis in this study.[17] In another study of 37 patients with incident breast cancer, Wu et al[18] found that HER2 status and age at menarche were significantly associated with α-diversity and specific microbial taxa. However, no significant association was found between α-diversity or β-diversity and ER/PR status, tumor grade, or cancer stage.[18] Both studies applied 16S ribosomal RNA gene-based sequencing to measure gut microbiome. While Yang's study collected stool samples before any breast cancer treatment, the stool samples in Wu's study were collected before neoadjuvant/adjuvant chemotherapy but could have been either pre- or postsurgery, similar to our study. The influence of surgery on the gut microbiome was not considered in the study by Wu.

In our study, we did not find that α- and β-diversities were associated with breast cancer subtypes or clinicopathologic factors. We found significant associations between HER2 status and Ki-67 levels (<20% *v* ≥20%) and several gut microbial taxa; however, no significant associations were found for ER, PR, and menopausal status. A significantly lower abundance of the class *Brachyspirae* was observed in patients with HER2-positive breast cancer; high Ki-67 expression (Ki-67 ≥20%) was associated with a decreased abundance of the genus *CAG-724*, class *Verrucomicrobiae*, and its order *Verrucomicrobiales*, but a higher abundance of genus *Dorea* (all FDR < 0.1, data not shown). In terms of molecular subtypes, *Acutalibacter* was more abundant among patients with HER2-enriched breast cancer than among patients with HR+/HER2-negative breast cancer. Compared with patients with HR+/HER2-negative breast cancer, a higher abundance of *Dorea sp001185345* and a lower abundance of species *MGYG-HGUT-01722* were observed among patients with triple-negative breast cancer. A higher abundance of the phylum *Elusimicrobiota* and its class *Elusimicrobia* and a lower abundance of the phylum *Cyanobacteria* were observed among patients with HR+/HER2-positive breast cancer. However, the mechanisms underlying these associations remain unclear. The lack of consistent findings between our study and previous studies may reflect differences in patient selection, stool sample collection time, sequencing methods, and multiple comparisons.

Notably, our study found that those who experienced a diagnosis delay had significantly lower α-diversity indices and variation in β-diversity (ie, unweighted UniFracs distance matrix). Compared with patients with no diagnosis delay, we found an increased abundance of the species *Enorma massiliensis* among patients with a serious diagnosis delay, whereas the species *Faecalicoccus pleomorphus*, the phylum *Elusimicrobiota*, and its class *Elusimicrobia* showed significantly decreased abundance among patients who experienced a moderate diagnosis delay (Appendix Table A6). Our previous study reported that diagnosis delay was common among Vietnamese patients with breast cancer.[34] In our study, 60% of patients with presurgery stool samples had a diagnosis delay and they are more likely to have a late stage of breast cancer (Appendix Table A7). Lifestyle changes, cancer progression, and medication use, such as herbal or alternative medical treatments, during the delayed period might have altered the gut microbiota, causing reduced diversity and a nonhealthy gut microbiome. Moreover, we speculate that physical and psychological stresses experienced by patients during the delayed period may alter their dietary intake, leading to changes in the gut microbial community.[19] Of note, approximately 31% of patients with presurgery stool samples who had a diagnosis delay reported that they received antibiotics for 5 days or longer during the past 3 months before stool collection. However, history of antibiotic use was not significantly associated with α- and β-diversities. Nevertheless, antibiotic use was significantly associated with a lower abundance of seven gut microbial taxa including the families *Eggerthellaceae*, *Erysipelatoclostridiaceae*, *Lachnospiraceae*, *Butyricicoccaceae*, and *Pasteurellaceae*, as well as a higher abundance of the phylum *Eremiobacterota* (driven by class *UBP9*), phylum *Synergistota* (driven by class *Synergistia* and order *Synergistales*), and class *Verrucomicrobiae* (Appendix Table A8). Our finding suggests that a history of antibiotic use during the past 3 months may alter the abundance of selected specific gut microbes but has little impact on overall gut microbiome diversity.

Last but at least, we found a significantly reduced α-diversity index among patients with breast cancer who had higher levels of fiber intake. We did not find that

α-diversities were associated with fat, carbohydrate, or protein intake. In agreement with our finding, decreased α-diversity was observed among several target healthy populations in randomized trials after administration of dietary fiber although evidence is not entirely consistent.[41] In our previous study among Chinese adults, we found that a long-term diet with more fruits and vegetables was positively associated with α-diversity.[42] Conversely, frequent habitual intake of whole grains and vegetables did not increase α-diversity in other Chinese population.[43] In addition, reduced α-diversity was observed among patients with diabetes who consumed a higher level of dietary fiber.[44] In terms of gut microbial taxa analysis, we found a significantly higher abundance of species belonging to *Bifidobacterium*, *Prevotella*, and *Bacteroides*, such as the species *Bifidobacterium gallinarum*, *Bifidobacterium pullorum*, *Bifidobacterium scardovii*, *Bacteroides clarus*, *Bacteroides graminisolvens*, *Bacteroides stercorirosoris*, *Bacteroides thetaiotaomicron*, *Prevotella melaninogenica*, and *Prevotella oris* among patients with breast cancer who had high fiber intake (T3), compared with those with low fiber intake (T1). We also observed a lower abundance of the phylum *Firmicutes A* and its two classes (*Clostridia* and *Mahella*) and family *Megasphaeraceae*, a member of the phylum *Firmicutes C*, among patients who had a high consumption of fiber (Appendix Table A9). Our findings support the hypothesis that high-fiber diets can enrich specific fiber-digested strains, which may inhibit the residence or growth of detrimental species, resulting in a reduction of α-diversity.[45]

To our knowledge, this is the first study evaluating the association between the gut microbiome and clinical and demographic factors among Vietnamese patients with breast cancer and one of the largest studies ever on the gut microbiome and breast cancer characteristics to date. In addition, shotgun metagenomic sequencing was performed in our study, which provided enhanced taxonomic resolution. We used a human bacterial genome from the UHGG collection as a reference, a massive sequence catalog containing information on approximately 4,600 species, with 71% lacking a cultured representative.[27-29] This allowed us to estimate the prevalence and abundance of species or genes with enhanced resolution and accuracy.

Several limitations should be considered when interpreting the findings. First, the study was based on a cross-sectional analysis. Thus, the causality of these associations cannot be established. Second, the statistical power of the study was compromised when the analyses were restricted to participants with preoperative stool samples. Finally, our findings may not be generalizable to the gut microbiome profile of all Vietnamese patients with breast cancer, particularly those treated in other settings in Vietnam.

In conclusion, diagnosis delay, high fiber intake, and breast cancer surgery substantially affect the gut microbiome profile among Vietnamese patients with breast cancer, leading to a less diverse and unhealthy gut microbial community. We also found that diagnosis delay, BMI levels, fiber intake, history of antibiotic use, cancer stage, and molecular subtypes were associated with specific gut microbes. Further research is needed to investigate how these gut microbiome differences influence the efficacy and effectiveness of cancer treatment.

AFFILIATIONS

[1]Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, TN
[2]Vietnam National Cancer Institute, National Cancer Hospital, Hanoi, Vietnam
[3]Hanoi Medical University, Hanoi, Vietnam
[4]Department of Public Health Sciences, School of Medicine, Center for Public Health Genomics, UVA Comprehensive Cancer Center, University of Virginia, Charlottesville, VA
[5]Hanoi Oncology Hospital, Hanoi, Vietnam
[6]Ministry of Health, Hanoi, Vietnam

CORRESPONDING AUTHOR

Huong T.T. Tran, MD, PhD, Hanoi Medical University & Vietnam National Cancer Institute, 43 Quan Su St, Hoan Kiem District, Hanoi, 10000 Vietnam; e-mail: huongtranthanh@hmu.edu.vn.

SUPPORT

Supported by NIH/NCI Grant No. P20 CA210300 and OISE-19-66185-1 (PIs: X.-O.S. and T.V.T.). In addition, S.M.N. was supported by a VECD Global Health Fellowship, funded by the National Cancer Institute (NCI) and the Fogarty International Center (FIC) of the NIH (D43 TW009337), and Ingram Cancer Professorship Fund (to X.-O.S.).

DATA SHARING STATEMENT

Data are available on request. The data underlying this article will be shared upon reasonable request to the corresponding author.

AUTHOR CONTRIBUTIONS

Conception and design: Sang M. Nguyen, Huong T.T. Tran, Martha J. Shrubsole, To V. Ta, Thuan V. Tran, Xiao-Ou Shu
Financial support: Sang M. Nguyen, Xiao-Ou Shu
Administrative support: Sang M. Nguyen, Huong T.T. Tran, Jirong Long, Giang H. Nguyen, Thuan V. Tran, Xiao-Ou Shu
Provision of study materials or patients: Sang M. Nguyen, Huong T.T. Tran, Xiao-Ou Shu
Collection and assembly of data: Sang M. Nguyen, Huong T.T. Tran, Lan M. Nguyen, Giang H. Nguyen, Chu V. Nguyen, Jie Wu, Xiao-Ou Shu
Data analysis and interpretation: Sang M. Nguyen, Huong T.T. Tran, Jirong Long, Hui Cai, Yaohua Yang, Qiuyin Cai, Wei Zheng, Xiao-Ou Shu
Manuscript writing: All authors
Final approval of manuscript: All authors
Accountable for all aspects of the work: All authors

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

The following represents disclosure information provided by authors of this manuscript. All relationships are considered compensated unless otherwise noted. Relationships are self-held unless noted. I = Immediate Family Member, Inst = My Institution. Relationships may not relate to the subject matter of this manuscript. For more information about ASCO's conflict of interest policy, please refer to www.asco.org/rwc or ascopubs.org/go/authors/author-center.

Open Payments is a public database containing information reported by companies about payments made to US-licensed physicians (Open Payments).

Martha J. Shrubsole
Research Funding: Pfizer (Inst)

No other potential conflicts of interest were reported.

ACKNOWLEDGMENT

The authors would like to give special thanks to the project participants and research staff members, without whom this study would not have been possible. This work was conducted in part using the resources of the Advanced Computing Center for Research and Education at Vanderbilt University, Nashville, TN.

REFERENCES

1. Dzutsev A, Goldszmid RS, Viaud S, et al: The role of the microbiota in inflammation, carcinogenesis, and cancer therapy. Eur J Immunol 45:17-31, 2015
2. Kamada N, Chen GY, Inohara N, et al: Control of pathogens and pathobionts by the gut microbiota. Nat Immunol 14:685-690, 2013
3. Fulde M, Hornef MW: Maturation of the enteric mucosal innate immune system during the postnatal period. Immunol Rev 260:21-34, 2014
4. Ijssennagger N, Belzer C, Hooiveld GJ, et al: Gut microbiota facilitates dietary heme-induced epithelial hyperproliferation by opening the mucus barrier in colon. Proc Natl Acad Sci USA 112: 10038-10043, 2015
5. Reinhardt C, Bergentall M, Greiner TU, et al: Tissue factor and PAR1 promote microbiota-induced intestinal vascular remodelling. Nature 483:627-631, 2012
6. Neuman H, Debelius JW, Knight R, et al: Microbial endocrinology: The interplay between the microbiota and the endocrine system. FEMS Microbiol Rev 39:509-521, 2015
7. Canfora EE, Jocken JW, Blaak EE: Short-chain fatty acids in control of body weight and insulin sensitivity. Nat Rev Endocrinol 11:577-591, 2015
8. Yano JM, Yu K, Donaldson GP, et al: Indigenous bacteria from the gut microbiota regulate host serotonin biosynthesis. Cell 161:264-276, 2015
9. Devlin AS, Fischbach MA: A biosynthetic pathway for a prominent class of microbiota-derived bile acids. Nat Chem Biol 11:685-690, 2015
10. Haiser HJ, Gootenberg DB, Chatman K, et al: Predicting and manipulating cardiac drug inactivation by the human gut bacterium Eggerthella lenta. Science 341:295-298, 2013
11. Zhang YJ, Li S, Gan RY, et al: Impacts of gut bacteria on human health and diseases. Int J Mol Sci 16:7493-7519, 2015
12. Fernández MF, Reina-Pérez I, Astorga JM, et al: Breast cancer and its relationship with the microbiota. Int J Environ Res Public Health 15:1747, 2018
13. Wu AH, Vigen C, Tseng C, et al: Effect of chemotherapy on the gut microbiome of breast cancer patients during the first year of treatment. Breast Cancer (Dove Med Press) 14:433-451, 2022
14. Luu TH, Michel C, Bard JM, et al: Intestinal proportion of Blautia sp. is associated with clinical stage and histoprognostic grade in patients with early-stage breast cancer. Nutr Cancer 69:267-275, 2017
15. Fruge AD, Van der Pol W, Rogers LQ, et al: Fecal *Akkermansia muciniphila* is associated with body composition and microbiota diversity in overweight and obese women with breast cancer participating in a presurgical weight loss trial. J Acad Nutr Diet 120:650-659, 2020
16. Hou MF, Ou-Yang F, Li CL, et al: Comprehensive profiles and diagnostic value of menopausal-specific gut microbiota in premenopausal breast cancer. Exp Mol Med 53:1636-1646, 2021
17. Yang P, Wang Z, Peng Q, et al: Comparison of the gut microbiota in patients with benign and malignant breast tumors: A pilot study. Evol Bioinform Online 17:11769343211057573, 2021
18. Wu AH, Tseng C, Vigen C, et al: Gut microbiome associations with breast cancer risk factors and tumor characteristics: A pilot study. Breast Cancer Res Treat 182:451-463, 2020
19. Nguyen SM, Tran HT, Nguyen LM, et al: Association of fruit, vegetable and animal food intakes with breast cancer risk overall and by molecular subtype among Vietnamese women. Cancer Epidemiol Biomarkers Prev 31:1026-1035, 2022
20. Nguyen SM, Pham AT, Nguyen LM, et al: Chemotherapy-Induced toxicities and their associations with clinical and non-clinical factors among breast cancer patients in Vietnam. Curr Oncol 29: 8269-8284, 2022
21. National Institute of Nutrition: Vietnamese Food Composition Table. Hanoi, Vietnam, Medical Publisher, 2007
22. National Institute of Nutrition: Vietnamese Food Composition Table. Hanoi, Vietnam, Medical Publisher, 2017
23. Ainsworth BE, Haskell WL, Whitt MC, et al: Compendium of physical activities: An update of activity codes and MET intensities. Med Sci Sports Exerc 32:S498-S504, 2000 (suppl 9)
24. Ainsworth BE, Haskell WL, Herrmann SD, et al: 2011 Compendium of physical activities: A second update of codes and MET values. Med Sci Sports Exerc 43:1575-1581, 2011
25. Bolger AM, Lohse M, Usadel B: Trimmomatic: A flexible trimmer for illumina sequence data. Bioinformatics 30:2114-2120, 2014
26. Langmead B, Salzberg SL: Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357-359, 2012
27. Wood DE, Lu J, Langmead B: Improved metagenomic analysis with Kraken 2. Genome Biol 20:257, 2019
28. Lu J, Breitwieser FP, Thielen P, et al: Bracken: Estimating species abundance in metagenomics data. PeerJ Comput Sci 3:e104, 2017
29. Almeida A, Nayfach S, Boland M, et al: A unified catalog of 204,938 reference genomes from the human gut microbiome. Nat Biotechnol 39:105-114, 2021
30. Shao Y, Forster SC, Tsaliki E, et al: Stunted microbiota and opportunistic pathogen colonization in caesarean-section birth. Nature 574:117-121, 2019
31. Ye SH, Siddle KJ, Park DJ, et al: Benchmarking metagenomics tools for taxonomic classification. Cell 178:779-794, 2019
32. Truong DT, Franzosa EA, Tickle TL, et al: MetaPhlAn2 for enhanced metagenomic taxonomic profiling. Nat Methods 12:902-903, 2015
33. Oksanen J, Blanchet FG, Kindt R, et al: Package 'vegan'. Community Ecology Package. R Package Version 2:1-295, 2013
34. Nguyen SM, Nguyen QT, Nguyen LM, et al: Delay in the diagnosis and treatment of breast cancer in Vietnam. Cancer Med 10:7683-7691, 2021
35. Tang Z-Z, Chen G, Alekseyenko AV: PERMANOVA-S: Association test for microbial community composition that accommodates confounders and multiple distances. Bioinformatics 32:2618-2625, 2016
36. Gupta VK, Kim M, Bakshi U, et al: A predictive index for health status using species-level gut microbiome profiling. Nat Commun 11:4635, 2020
37. Gloor GB, Wu JR, Pawlowsky-Glahn V, et al: It's all relative: Analyzing microbiome data as compositions. Ann Epidemiol 26:322-329, 2016
38. Zhou H, He K, Chen J, et al: LinDA: Linear models for differential abundance analysis of microbiome compositional data. Genome Biol 23:95, 2022
39. Xu S, Zhan L, Tang W, et al: MicrobiotaProcess: A comprehensive R package for deep mining microbiome. Innovation (Camb) 4:100388, 2023
40. Sandhu BK, McBride SM: Clostridioides difficile. Trends Microbiol 26:1049-1050, 2018
41. Fu J, Zheng Y, Gao Y, et al: Dietary fiber intake and gut microbiota in human health. Microorganisms 10:2507, 2022
42. Yu D, Nguyen SM, Yang Y, et al: Long-term diet quality is associated with gut microbiome diversity and composition among urban Chinese adults. Am J Clin Nutr 113:684-694, 2021
43. Zhang Y, Chen H, Lu M, et al: Habitual diet pattern associations with gut microbiome diversity and composition: Results from a Chinese adult cohort. Nutrients 14:2639, 2022
44. Fu J, Xu K, Ni X, et al: Habitual dietary fiber intake, fecal microbiota, and hemoglobin A1c level in Chinese patients with type 2 diabetes. Nutrients 14:1003, 2022
45. Valdes AM, Walter J, Segal E, et al: Role of the gut microbiota in nutrition and health. BMJ 361:k2179, 2018

APPENDIX 1. SEQUENCING SUMMARY

In our study, we obtained an average of 10.19 million (minimum-maximum: 4.05-10.90) raw sequencing reads per sample. After quality-trimming and human reads removal steps, an average of 10.14 million (minimum-maximum: 4.05-10.89) clean reads per sample was retained for taxonomic profiling. On average, 87.9% (min-max: 71.0%-92.7%) of clean reads were classified by Kraken2 against the UHGG reference database (Appendix Table A2).

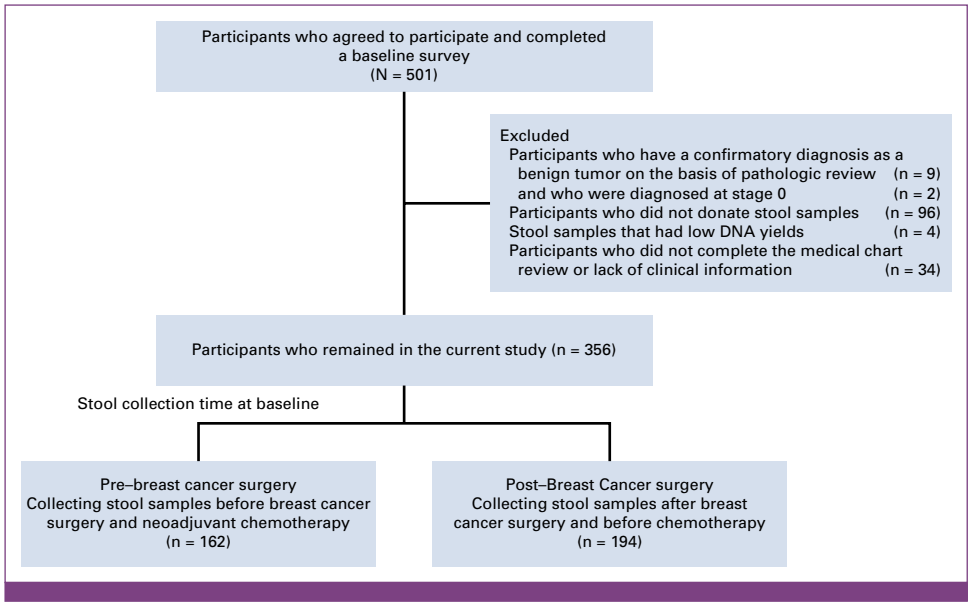

FIG A1. Flow diagram of study participants.

TABLE A1. Sequencing Coverage and Quality Statistics for Each Sample

	No. of Reads				
Participant ID	Raw Read	Clean Read[a]	Sequence Unclassified[b]	Sequence Classified[b]	% Classified
VNBC00005	10,196,555	10,196,085	968,036	9,228,049	90.51
VNBC00014	10,431,482	10,416,457	1,168,516	9,247,941	88.78
VNBC00015	10,336,251	10,331,851	1,187,867	9,143,984	88.50
VNBC00016	10,390,046	10,387,051	1,054,452	9,332,599	89.85
VNBC00025	10,523,566	10,522,856	1,384,562	9,138,294	86.84
VNBC00030	10,505,007	10,467,354	1,362,153	9,105,201	86.99
VNBC00035	10,060,275	10,056,691	1,089,064	8,967,627	89.17
VNBC00036	10,401,836	10,400,113	1,340,040	9,060,073	87.12
VNBC00039	10,338,377	10,336,461	1,173,521	9,162,940	88.65
VNBC00040	10,035,985	10,035,193	1,088,974	8,946,219	89.15
VNBC00043	10,076,320	10,041,034	1,181,855	8,859,179	88.23
VNBC00045	9,512,719	9,512,017	839,416	8,672,601	91.18
VNBC00046	10,009,135	10,008,826	985,701	9,023,125	90.15
VNBC00048	10,784,239	10,757,128	1,238,600	9,518,528	88.49
VNBC00050	10,390,736	10,390,151	1,394,315	8,995,836	86.58
VNBC00052	10,363,295	10,361,235	1,224,872	9,136,363	88.18
VNBC00053	10,345,906	10,344,985	1,218,814	9,126,171	88.22
VNBC00058	10,663,055	10,650,826	1,291,855	9,358,971	87.87
VNBC00059	10,717,299	10,658,782	1,005,547	9,653,235	90.57
VNBC00060	10,775,237	10,772,182	1,248,852	9,523,330	88.41
VNBC00062	10,015,133	9,982,886	864,861	9,118,025	91.34
VNBC00063	10,439,861	10,430,974	1,032,810	9,398,164	90.10
VNBC00064	10,476,223	10,423,808	1,342,158	9,081,650	87.12
VNBC00065	10,099,024	10,097,842	1,277,480	8,820,362	87.35
VNBC00067	10,511,274	10,504,936	1,281,237	9,223,699	87.80
VNBC00070	10,114,505	10,113,720	1,284,237	8,829,483	87.30
VNBC00072	10,569,690	10,563,509	1,227,303	9,336,206	88.38
VNBC00073	10,115,987	10,115,733	1,165,094	8,950,639	88.48
VNBC00076	10,666,443	10,665,910	1,188,921	9,476,989	88.85
VNBC00078	10,538,176	10,537,237	1,011,759	9,525,478	90.40
VNBC00079	10,471,150	10,468,093	1,466,771	9,001,322	85.99
VNBC00093	7,139,441	7,138,457	857,321	6,281,136	87.99
VNBC00095	10,763,116	10,761,990	1,116,864	9,645,126	89.62
VNBC00098	10,035,106	10,034,604	1,481,604	8,553,000	85.24
VNBC00101	10,299,177	10,283,760	1,060,080	9,223,680	89.69
VNBC00103	10,527,216	10,507,158	1,335,549	9,171,609	87.29
VNBC00104	10,276,669	10,221,977	970,549	9,251,428	90.51
VNBC00107	10,362,619	10,320,547	1,884,960	8,435,587	81.74
VNBC00108	6,471,737	6,459,940	775,937	5,684,003	87.99
VNBC00109	10,434,319	10,433,964	1,480,014	8,953,950	85.82
VNBC00113	10,730,642	10,714,770	1,375,816	9,338,954	87.16
VNBC00114	10,665,678	10,663,778	1,210,842	9,452,936	88.65
VNBC00115	10,222,929	10,221,537	1,172,113	9,049,424	88.53
VNBC00116	10,308,035	10,298,323	2,343,748	7,954,575	77.24
VNBC00117	10,709,531	10,697,576	1,161,455	9,536,121	89.14
VNBC00118	10,384,513	10,383,101	1,598,369	8,784,732	84.61
VNBC00121	10,265,079	10,264,844	1,163,264	9,101,580	88.67
VNBC00122	10,517,828	10,485,120	1,676,334	8,808,786	84.01

(continued on following page)

TABLE A1. Sequencing Coverage and Quality Statistics for Each Sample (continued)

	No. of Reads				
Participant ID	Raw Read	Clean Read[a]	Sequence Unclassified[b]	Sequence Classified[b]	% Classified
VNBC00123	10,565,034	10,557,807	1,226,779	9,331,028	88.38
VNBC00129	10,444,046	10,443,471	1,408,864	9,034,607	86.51
VNBC00130	10,378,450	10,378,232	1,490,810	8,887,422	85.64
VNBC00131	10,327,730	10,326,658	1,192,500	9,134,158	88.45
VNBC00134	9,998,980	9,990,049	2,226,771	7,763,278	77.71
VNBC00135	10,473,866	10,464,963	1,142,415	9,322,548	89.08
VNBC00136	10,640,688	10,442,361	1,197,109	9,245,252	88.54
VNBC00143	10,651,051	8,429,915	882,325	7,547,590	89.53
VNBC00146	10,231,395	10,231,259	1,167,567	9,063,692	88.59
VNBC00152	9,964,620	9,962,766	1,024,028	8,938,738	89.72
VNBC00159	10,440,571	10,235,413	1,049,858	9,185,555	89.74
VNBC00160	10,370,030	10,365,658	1,291,625	9,074,033	87.54
VNBC00163	10,348,401	10,341,734	1,141,593	9,200,141	88.96
VNBC00165	9,043,353	9,039,461	860,302	8,179,159	90.48
VNBC00167	10,292,917	10,291,560	1,038,458	9,253,102	89.91
VNBC00168	10,709,137	10,695,639	1,364,895	9,330,744	87.24
VNBC00169	9,966,567	9,964,148	1,273,866	8,690,282	87.22
VNBC00171	10,359,986	10,355,134	1,012,511	9,342,623	90.22
VNBC00174	10,198,890	10,198,784	1,018,750	9,180,034	90.01
VNBC00176	10,586,901	10,585,568	1,194,022	9,391,546	88.72
VNBC00178	10,397,272	10,376,115	1,389,318	8,986,797	86.61
VNBC00181	9,855,650	9,854,616	981,041	8,873,575	90.04
VNBC00182	10,768,731	10,720,788	1,043,769	9,677,019	90.26
VNBC00184	10,109,681	10,109,222	1,026,895	9,082,327	89.84
VNBC00199	10,148,092	10,139,790	879,751	9,260,039	91.32
VNBC00201	10,672,167	10,662,791	1,119,536	9,543,255	89.50
VNBC00202	10,776,076	10,739,473	1,027,712	9,711,761	90.43
VNBC00203	10,025,030	10,024,626	1,095,098	8,929,528	89.08
VNBC00204	10,516,497	10,497,447	1,081,457	9,415,990	89.70
VNBC00211	10,632,460	10,624,610	1,168,116	9,456,494	89.01
VNBC00212	10,619,547	10,616,067	1,233,709	9,382,358	88.38
VNBC00213	10,358,748	10,358,156	1,302,832	9,055,324	87.42
VNBC00216	10,526,593	10,525,581	1,330,958	9,194,623	87.36
VNBC00217	7,505,129	7,504,642	864,149	6,640,493	88.49
VNBC00219	10,721,419	10,699,530	1,497,498	9,202,032	86.00
VNBC00220	10,730,955	10,708,141	1,369,461	9,338,680	87.21
VNBC00228	10,467,460	10,462,586	1,264,419	9,198,167	87.91
VNBC00229	10,416,103	10,414,604	1,470,718	8,943,886	85.88
VNBC00235	10,532,145	10,526,787	1,509,236	9,017,551	85.66
VNBC00237	10,192,925	10,189,737	1,117,909	9,071,828	89.03
VNBC00246	10,272,729	10,272,244	1,196,734	9,075,510	88.35
VNBC00248	9,964,673	9,950,781	1,061,286	8,889,495	89.33
VNBC00249	10,620,756	10,614,082	1,325,131	9,288,951	87.52
VNBC00252	10,567,364	10,563,287	1,140,934	9,422,353	89.20
VNBC00253	10,106,301	10,101,400	1,322,551	8,778,849	86.91
VNBC00255	10,455,607	10,455,106	1,250,934	9,204,172	88.04
VNBC00256	10,493,916	10,478,867	1,112,340	9,366,527	89.38
VNBC00265	10,174,195	10,160,592	1,101,766	9,058,826	89.16

(continued on following page)

TABLE A1. Sequencing Coverage and Quality Statistics for Each Sample (continued)

	No. of Reads				
Participant ID	Raw Read	Clean Read[a]	Sequence Unclassified[b]	Sequence Classified[b]	% Classified
VNBC00266	10,780,042	10,777,593	1,167,940	9,609,653	89.16
VNBC00268	10,549,857	10,546,268	1,039,151	9,507,117	90.15
VNBC00271	10,693,238	10,688,103	1,296,549	9,391,554	87.87
VNBC00272	10,567,671	10,567,324	1,436,033	9,131,291	86.41
VNBC00282	9,974,087	9,973,535	1,020,701	8,952,834	89.77
VNBC00301	9,589,979	9,587,258	1,068,850	8,518,408	88.85
VNBC00302	10,576,925	10,571,488	1,233,555	9,337,933	88.33
VNBC00303	10,491,466	10,409,150	1,151,312	9,257,838	88.94
VNBC00304	10,029,921	10,007,574	1,632,897	8,374,677	83.68
VNBC00306	10,079,668	10,069,728	1,151,279	8,918,449	88.57
VNBC00307	10,670,979	10,661,396	1,279,863	9,381,533	88.00
VNBC00309	10,617,851	10,617,544	1,242,831	9,374,713	88.29
VNBC00310	10,314,803	10,291,690	1,350,812	8,940,878	86.87
VNBC00311	10,698,966	10,660,846	1,319,972	9,340,874	87.62
VNBC00312	10,218,573	10,178,628	967,070	9,211,558	90.50
VNBC00313	10,470,184	10,465,223	1,451,708	9,013,515	86.13
VNBC00314	10,618,685	10,606,833	1,126,847	9,479,986	89.38
VNBC00315	10,591,729	10,584,201	1,510,648	9,073,553	85.73
VNBC00316	10,022,559	10,021,945	1,166,908	8,855,037	88.36
VNBC00317	10,124,130	10,118,393	1,318,787	8,799,606	86.97
VNBC00318	10,899,139	10,889,310	1,350,910	9,538,400	87.59
VNBC00320	10,292,655	10,277,915	1,231,615	9,046,300	88.02
VNBC00321	10,407,996	10,407,327	1,006,002	9,401,325	90.33
VNBC00323	10,691,856	10,546,953	913,855	9,633,098	91.34
VNBC00326	10,604,744	10,549,878	1,112,000	9,437,878	89.46
VNBC00327	10,615,712	10,172,279	1,015,379	9,156,900	90.02
VNBC00330	10,668,425	10,633,531	1,364,350	9,269,181	87.17
VNBC00331	10,528,454	10,517,088	1,370,290	9,146,798	86.97
VNBC00334	10,699,434	10,698,103	997,691	9,700,412	90.67
VNBC00336	4,053,098	4,052,270	461,058	3,591,212	88.62
VNBC00337	10,757,327	10,756,928	1,209,165	9,547,763	88.76
VNBC00338	10,390,137	10,385,117	888,905	9,496,212	91.44
VNBC00339	10,716,597	10,711,770	1,270,243	9,441,527	88.14
VNBC00340	10,361,750	10,326,922	1,055,759	9,271,163	89.78
VNBC00341	10,511,657	10,508,270	1,174,284	9,333,986	88.83
VNBC00342	9,968,587	9,955,892	801,349	9,154,543	91.95
VNBC00343	10,305,025	10,304,259	1,338,925	8,965,334	87.01
VNBC00345	10,716,677	10,709,733	1,532,614	9,177,119	85.69
VNBC00346	10,021,242	10,018,963	1,869,842	8,149,121	81.34
VNBC00347	9,963,990	9,963,625	1,040,544	8,923,081	89.56
VNBC00348	10,219,368	10,214,864	1,223,426	8,991,438	88.02
VNBC00351	10,572,200	10,567,447	1,173,170	9,394,277	88.90
VNBC00356	7,374,590	7,365,326	1,028,187	6,337,139	86.04
VNBC00357	10,621,679	10,595,347	1,321,540	9,273,807	87.53
VNBC00358	10,702,381	10,696,985	1,238,748	9,458,237	88.42
VNBC00359	10,334,842	10,334,689	1,212,493	9,122,196	88.27
VNBC00360	9,999,211	9,995,963	1,358,104	8,637,859	86.41
VNBC00361	10,252,893	10,239,170	1,091,401	9,147,769	89.34

(continued on following page)

TABLE A1. Sequencing Coverage and Quality Statistics for Each Sample (continued)

	No. of Reads				
Participant ID	Raw Read	Clean Read[a]	Sequence Unclassified[b]	Sequence Classified[b]	% Classified
VNBC00362	10,155,472	10,119,611	1,118,904	9,000,707	88.94
VNBC00363	10,056,856	10,053,980	1,067,777	8,986,203	89.38
VNBC00366	10,232,921	10,225,465	1,381,410	8,844,055	86.49
VNBC00367	10,740,908	10,740,140	1,394,144	9,345,996	87.02
VNBC00368	10,365,674	10,364,421	1,289,723	9,074,698	87.56
VNBC00369	10,235,449	10,235,268	882,861	9,352,407	91.37
VNBC00370	10,248,001	10,226,941	1,880,627	8,346,314	81.61
VNBC00373	10,747,476	10,743,734	3,110,270	7,633,464	71.05
VNBC00374	10,736,702	9,591,760	1,074,725	8,517,035	88.80
VNBC00375	10,296,607	10,247,791	1,095,142	9,152,649	89.31
VNBC00377	10,390,355	10,389,976	986,500	9,403,476	90.51
VNBC00378	9,618,506	9,446,724	1,077,262	8,369,462	88.60
VNBC00379	10,722,623	10,713,644	1,032,131	9,681,513	90.37
VNBC00380	10,364,904	10,361,717	1,142,600	9,219,117	88.97
VNBC00383	10,298,563	10,292,387	1,214,021	9,078,366	88.20
VNBC00384	10,703,989	9,747,190	923,861	8,823,329	90.52
VNBC00385	10,724,986	10,719,372	1,204,805	9,514,567	88.76
VNBC00386	10,008,094	9,927,364	1,125,088	8,802,276	88.67
VNBC00389	9,972,467	9,899,655	966,561	8,933,094	90.24
VNBC00390	10,585,527	10,574,213	1,159,894	9,414,319	89.03
VNBC00392	10,420,861	10,404,863	1,815,165	8,589,698	82.55
VNBC00393	9,997,203	9,996,598	1,214,859	8,781,739	87.85
VNBC00394	10,492,033	10,487,103	1,262,274	9,224,829	87.96
VNBC00395	10,539,655	10,535,290	1,372,054	9,163,236	86.98
VNBC00398	10,639,801	10,637,509	1,467,355	9,170,154	86.21
VNBC00400	10,693,543	10,690,662	1,625,812	9,064,850	84.79
VNBC00402	10,373,433	10,370,093	1,132,335	9,237,758	89.08
VNBC00403	10,569,505	10,569,183	1,080,930	9,488,253	89.77
VNBC00404	10,110,076	10,109,675	1,268,280	8,841,395	87.45
VNBC00406	10,786,770	10,777,831	1,334,110	9,443,721	87.62
VNBC00409	10,313,490	10,308,324	1,138,179	9,170,145	88.96
VNBC00412	10,452,490	10,450,971	1,510,815	8,940,156	85.54
VNBC00413	10,183,698	10,179,877	1,697,412	8,482,465	83.33
VNBC00414	10,729,050	10,725,527	1,543,466	9,182,061	85.61
VNBC00415	10,248,889	10,247,188	1,761,031	8,486,157	82.81
VNBC00418	10,420,710	10,419,382	1,670,513	8,748,869	83.97
VNBC00424	5,764,234	5,761,291	776,621	4,984,670	86.52
VNBC00425	10,754,455	10,751,341	1,188,866	9,562,475	88.94
VNBC00426	10,694,186	10,688,917	1,242,340	9,446,577	88.38
VNBC00427	10,543,210	10,541,269	1,218,498	9,322,771	88.44
VNBC00428	4,631,415	4,611,968	742,439	3,869,529	83.90
VNBC00429	10,629,477	10,290,167	1,150,492	9,139,675	88.82
VNBC00430	9,969,054	9,954,630	1,231,758	8,722,872	87.63
VNBC00431	9,983,061	9,921,387	1,712,976	8,208,411	82.73
VNBC00432	10,582,099	10,580,133	1,237,373	9,342,760	88.30
VNBC00433	10,487,064	10,485,849	1,116,712	9,369,137	89.35
VNBC00434	10,133,048	10,132,660	1,324,459	8,808,201	86.93
VNBC00435	10,598,942	10,572,061	1,385,976	9,186,085	86.89

(continued on following page)

TABLE A1. Sequencing Coverage and Quality Statistics for Each Sample (continued)

	No. of Reads				
Participant ID	Raw Read	Clean Read[a]	Sequence Unclassified[b]	Sequence Classified[b]	% Classified
VNBC00436	10,550,078	10,345,377	1,144,784	9,200,593	88.93
VNBC00437	10,482,051	10,472,438	1,283,004	9,189,434	87.75
VNBC00438	7,962,107	7,956,831	894,921	7,061,910	88.75
VNBC00439	10,310,229	10,308,273	908,941	9,399,332	91.18
VNBC00441	10,594,594	10,530,661	1,357,609	9,173,052	87.11
VNBC00442	10,340,388	10,193,710	1,520,228	8,673,482	85.09
VNBC00443	10,680,574	10,680,115	1,281,135	9,398,980	88.00
VNBC00445	10,529,689	10,526,709	1,572,507	8,954,202	85.06
VNBC00448	10,047,917	10,045,495	1,194,022	8,851,473	88.11
VNBC00450	10,544,406	10,541,244	1,116,043	9,425,201	89.41
VNBC00451	10,023,887	10,022,946	1,874,698	8,148,248	81.30
VNBC00452	10,746,891	10,744,595	1,174,408	9,570,187	89.07
VNBC00453	10,017,115	9,993,321	1,799,658	8,193,663	81.99
VNBC00454	10,618,598	10,595,381	1,000,487	9,594,894	90.56
VNBC00455	10,183,337	10,180,905	1,045,443	9,135,462	89.73
VNBC00457	10,491,107	10,466,244	1,215,960	9,250,284	88.38
VNBC00458	10,411,979	10,410,042	973,180	9,436,862	90.65
VNBC00459	10,632,368	10,627,326	1,025,908	9,601,418	90.35
VNBC00460	10,283,277	9,617,462	1,077,292	8,540,170	88.80
VNBC00461	10,220,320	10,219,420	1,534,640	8,684,780	84.98
VNBC00463	10,167,250	10,166,437	1,135,800	9,030,637	88.83
VNBC00466	10,160,863	10,126,436	1,555,670	8,570,766	84.64
VNBC00468	10,679,255	10,655,317	1,309,830	9,345,487	87.71
VNBC00470	10,554,389	10,539,526	1,527,506	9,012,020	85.51
VNBC00474	10,739,643	10,695,357	1,147,179	9,548,178	89.27
VNBC00475	10,598,046	10,596,714	1,594,178	9,002,536	84.96
VNBC00476	10,446,678	10,442,571	1,262,461	9,180,110	87.91
VNBC00478	10,627,682	10,606,391	1,169,183	9,437,208	88.98
VNBC00480	10,810,343	10,789,880	1,326,515	9,463,365	87.71
VNBC00482	10,573,450	10,573,297	1,064,969	9,508,328	89.93
VNBC00484	10,658,689	10,550,606	1,350,540	9,200,066	87.20
VNBC00486	10,198,545	10,195,335	915,843	9,279,492	91.02
VNBC00490	10,670,022	10,624,185	1,029,820	9,594,365	90.31
VNBC00491	10,589,193	10,587,939	1,437,357	9,150,582	86.42
VNBC00492	10,688,776	10,687,810	1,998,343	8,689,467	81.30
VNBC00494	9,989,716	9,987,540	1,927,892	8,059,648	80.70
VNBC00497	10,754,201	10,750,731	1,125,414	9,625,317	89.53
VNBC00498	9,930,615	9,917,614	1,010,921	8,906,693	89.81
VNBC00500	10,762,250	10,756,382	1,807,440	8,948,942	83.20
VNBC00502	10,619,096	10,595,615	772,036	9,823,579	92.71
VNBC00503	10,694,930	10,692,212	1,534,169	9,158,043	85.65
VNBC00509	10,756,015	10,741,902	1,044,086	9,697,816	90.28
VNBC00511	10,437,692	10,436,045	1,182,664	9,253,381	88.67
VNBC00518	10,504,562	10,504,043	1,231,724	9,272,319	88.27
VNBC00523	10,458,677	10,452,794	891,260	9,561,534	91.47
VNBC00525	10,001,837	9,977,422	1,385,436	8,591,986	86.11
VNBC00527	10,216,478	10,216,220	1,339,168	8,877,052	86.89
VNBC00528	10,490,523	10,366,867	900,469	9,466,398	91.31

(continued on following page)

TABLE A1. Sequencing Coverage and Quality Statistics for Each Sample (continued)

	No. of Reads				
Participant ID	Raw Read	Clean Read[a]	Sequence Unclassified[b]	Sequence Classified[b]	% Classified
VNBC00530	10,326,204	10,325,489	1,312,936	9,012,553	87.28
VNBC00532	10,615,580	10,612,086	1,129,435	9,482,651	89.36
VNBC00534	10,026,485	10,001,945	1,020,947	8,980,998	89.79
VNBC00535	10,303,151	10,302,884	1,489,582	8,813,302	85.54
VNBC00537	10,182,512	10,180,619	1,493,527	8,687,092	85.33
VNBC00538	6,609,245	6,609,173	638,541	5,970,632	90.34
VNBC00539	10,070,958	10,070,616	1,250,797	8,819,819	87.58
VNBC00543	9,551,193	9,525,217	1,104,204	8,421,013	88.41
VNBC00545	9,975,177	9,973,398	1,036,444	8,936,954	89.61
VNBC00548	10,083,591	10,080,787	1,798,079	8,282,708	82.16
VNBC00550	10,555,838	10,524,301	1,538,574	8,985,727	85.38
VNBC00551	10,409,132	10,408,988	1,257,049	9,151,939	87.92
VNBC00552	10,632,470	10,624,830	1,195,436	9,429,394	88.75
VNBC00553	10,453,597	10,449,661	1,048,977	9,400,684	89.96
VNBC00554	10,232,655	10,228,191	1,504,455	8,723,736	85.29
VNBC00555	10,378,506	10,378,210	1,264,340	9,113,870	87.82
VNBC00558	10,217,838	10,216,332	1,132,560	9,083,772	88.91
VNBC00559	10,318,825	10,317,886	1,071,127	9,246,759	89.62
VNBC00560	10,175,955	10,175,561	1,130,118	9,045,443	88.89
VNBC00561	10,715,163	10,710,747	1,081,598	9,629,149	89.90
VNBC00562	10,052,926	10,052,076	1,153,824	8,898,252	88.52
VNBC00563	10,354,869	10,350,575	1,364,962	8,985,613	86.81
VNBC00565	9,910,016	9,898,716	1,226,916	8,671,800	87.61
VNBC00566	10,042,639	10,042,404	1,000,446	9,041,958	90.04
VNBC00571	10,762,944	10,736,855	1,276,975	9,459,880	88.11
VNBC00577	10,585,344	10,575,007	1,126,498	9,448,509	89.35
VNBC00579	9,840,248	9,816,767	1,107,662	8,709,105	88.72
VNBC00584	10,293,267	10,284,307	1,105,838	9,178,469	89.25
VNBC00586	10,505,227	10,501,949	1,215,220	9,286,729	88.43
VNBC00589	10,513,368	10,511,069	1,296,464	9,214,605	87.67
VNBC00599	9,977,075	9,966,442	1,523,384	8,443,058	84.71
VNBC00603	10,345,444	10,343,943	996,669	9,347,274	90.36
VNBC00610	10,565,243	10,564,630	1,309,634	9,254,996	87.60
VNBC00627	10,720,850	10,719,254	1,157,137	9,562,117	89.21
VNBC00633	10,503,725	10,502,694	1,575,839	8,926,855	85.00
VNBC00635	10,666,710	10,661,082	1,133,328	9,527,754	89.37
VNBC00636	10,391,556	10,341,446	1,226,486	9,114,960	88.14
VNBC00643	10,617,708	10,613,908	1,492,053	9,121,855	85.94
VNBC00644	9,963,797	9,953,873	1,086,847	8,867,026	89.08
VNBC00657	9,990,149	9,982,803	1,454,030	8,528,773	85.43
VNBC00662	10,186,516	10,184,819	1,276,782	8,908,037	87.46
VNBC00672	10,514,181	10,510,846	1,359,190	9,151,656	87.07
VNBC00677	10,446,064	10,421,764	1,439,675	8,982,089	86.19
VNBC00679	10,267,523	10,267,192	1,250,700	9,016,492	87.82
VNBC00687	10,565,662	10,565,241	1,175,603	9,389,638	88.87
VNBC00688	10,189,669	10,141,324	1,259,083	8,882,241	87.58
VNBC00689	10,742,948	10,739,498	1,381,689	9,357,809	87.13
VNBC00695	10,715,601	10,713,273	1,545,811	9,167,462	85.57

(continued on following page)

TABLE A1. Sequencing Coverage and Quality Statistics for Each Sample (continued)

	No. of Reads				
Participant ID	Raw Read	Clean Read[a]	Sequence Unclassified[b]	Sequence Classified[b]	% Classified
VNBC00703	10,410,381	10,398,498	1,230,928	9,167,570	88.16
VNBC00707	10,024,049	10,022,688	1,085,119	8,937,569	89.17
VNBC00710	10,627,361	10,626,477	1,209,876	9,416,601	88.61
VNBC00711	10,437,600	10,432,149	1,140,100	9,292,049	89.07
VNBC00716	10,360,755	10,343,646	1,273,213	9,070,433	87.69
VNBC00719	10,607,707	10,604,984	1,140,739	9,464,245	89.24
VNBC00724	9,099,505	9,046,168	916,814	8,129,354	89.87
VNBC00730	10,606,019	10,576,872	1,435,797	9,141,075	86.43
VNBC00732	10,605,405	10,531,579	1,174,122	9,357,457	88.85
VNBC00735	9,561,161	9,560,404	968,132	8,592,272	89.87
VNBC00747	10,743,152	10,742,085	1,191,380	9,550,705	88.91
VNBC00754	10,481,819	10,480,994	1,049,903	9,431,091	89.98
VNBC00757	10,305,533	10,284,524	1,337,569	8,946,955	86.99
VNBC00767	10,596,166	10,593,260	1,088,404	9,504,856	89.73
VNBC00778	10,653,371	10,598,265	1,099,704	9,498,561	89.62
VNBC00787	10,279,252	10,278,460	1,253,424	9,025,036	87.81
VNBC00796	10,300,498	10,297,568	931,352	9,366,216	90.96
VNBC00801	10,097,485	10,043,307	1,008,944	9,034,363	89.95
VNBC00804	10,727,243	10,722,335	1,088,614	9,633,721	89.85
VNBC00809	10,745,988	10,742,830	1,113,884	9,628,946	89.63
VNBC00810	10,176,620	10,160,315	1,754,990	8,405,325	82.73
VNBC00814	10,528,894	10,495,314	1,341,477	9,153,837	87.22
VNBC00816	10,717,972	10,716,910	1,237,376	9,479,534	88.45
VNBC00818	7,871,017	7,860,720	922,394	6,938,326	88.27
VNBC00820	9,997,960	9,997,738	1,013,623	8,984,115	89.86
VNBC00821	10,017,212	10,016,689	1,345,880	8,670,809	86.56
VNBC00822	10,037,129	8,830,921	1,414,341	7,416,580	83.98
VNBC00835	10,362,805	10,361,189	1,254,879	9,106,310	87.89
VNBC00836	10,447,703	10,439,500	1,300,954	9,138,546	87.54
VNBC00838	10,656,474	10,656,179	1,478,225	9,177,954	86.13
VNBC00844	10,723,799	10,702,673	1,435,970	9,266,703	86.58
VNBC00846	10,451,046	10,293,968	1,222,521	9,071,447	88.12
VNBC00847	10,075,810	10,073,398	1,094,572	8,978,826	89.13
VNBC00856	10,471,380	10,395,237	1,172,672	9,222,565	88.72
VNBC00858	10,689,286	10,665,150	1,009,993	9,655,157	90.53
VNBC00865	10,657,500	10,551,063	1,245,474	9,305,589	88.20
VNBC00870	8,289,722	8,275,432	908,083	7,367,349	89.03
VNBC00871	10,440,554	10,422,635	1,341,772	9,080,863	87.13
VNBC00873	9,999,442	9,974,514	1,125,382	8,849,132	88.72
VNBC00875	10,640,705	10,635,917	1,707,764	8,928,153	83.94
VNBC00876	10,418,155	10,417,591	1,417,873	8,999,718	86.39
VNBC00878	10,188,388	10,171,248	1,296,776	8,874,472	87.25
VNBC00911	10,303,915	10,291,562	999,073	9,292,489	90.29
VNBC00923	5,937,485	5,936,915	742,360	5,194,555	87.50
VNBC00925	10,416,727	10,412,810	1,120,081	9,292,729	89.24
VNBC00926	10,116,250	10,113,669	1,302,329	8,811,340	87.12
VNBC00927	10,720,528	10,713,342	1,262,874	9,450,468	88.21
VNBC00928	10,730,632	10,729,698	1,148,817	9,580,881	89.29

(continued on following page)

TABLE A1. Sequencing Coverage and Quality Statistics for Each Sample (continued)

	No. of Reads				
Participant ID	Raw Read	Clean Read[a]	Sequence Unclassified[b]	Sequence Classified[b]	% Classified
VNBC00930	10,316,119	10,315,603	1,368,830	8,946,773	86.73
VNBC00933	10,107,218	10,106,669	1,166,744	8,939,925	88.46
VNBC00937	10,516,793	10,513,668	1,182,294	9,331,374	88.75
VNBC00943	10,096,679	10,094,392	1,203,845	8,890,547	88.07
VNBC00951	10,176,933	10,169,739	1,153,391	9,016,348	88.66
VNBC00953	10,552,667	10,549,008	1,056,971	9,492,037	89.98
VNBC00960	10,048,423	10,032,488	1,062,246	8,970,242	89.41
VNBC00963	10,774,809	10,769,727	1,450,429	9,319,298	86.53
VNBC00965	10,419,314	10,403,493	1,371,313	9,032,180	86.82
VNBC00966	10,272,600	10,270,400	1,103,679	9,166,721	89.25
VNBC00970	10,021,049	10,019,502	1,502,423	8,517,079	85.01
VNBC00971	8,117,202	8,116,952	790,200	7,326,752	90.26
VNBC00974	10,083,620	10,066,206	1,172,494	8,893,712	88.35
VNBC00975	8,219,701	8,208,641	1,365,321	6,843,320	83.37
VNBC00976	8,300,646	8,277,091	1,131,238	7,145,853	86.33
VNBC00977	10,612,643	10,600,466	1,464,377	9,136,089	86.19
VNBC00982	10,664,886	10,661,439	1,295,512	9,365,927	87.85
VNBC00985	10,539,432	10,539,038	889,913	9,649,125	91.56
VNBC00989	10,516,948	10,516,871	857,798	9,659,073	91.84
VNBC01041	6,704,170	6,703,838	827,840	5,875,998	87.65

[a]Reads after trimming of low-quality reads using Trimmomatic (v0.39) and removing human reads using Bowtie2 (v_2.3.0).
[b]Kraken2 (v_ 2.1.1) was used for microbial profiling against bacterial genomes from the UHGG reference database.

TABLE A2. Summary of Metagenomic Sequencing Data of 356 Stool Samples

Parameter	Mean (range)
Raw reads	10,189,904 (4,053,098-10,899,139)
Clean reads[a]	10,155,213 (4,052,270-10,889,310)
Classified reads[b]	8,925,094 (3,591,212-9,823,579)
% classified	87.9 (71.0-92.7)

[a]Reads after trimming of low-quality reads using Trimmomatic (v0.39) and removing human reads using Bowtie2 (v_2.3.0).
[b]Kraken2 (v_ 2.1.1) was used for microbial profiling against bacterial genomes from the UHGG reference database.

TABLE A3. Association of TNM Cancer Stage With Gut Microbial Taxa Among Participants With Pre–Breast Cancer Surgery Stool Samples (n = 162)

Microbial Taxa	Pre (%)[a] / RA, Median (%)[b]								
	Stage I (n = 14)	Stage II (n = 76)	Stage III-IV (n = 72)	Log_2 FC (SE) for Stage II *v* Stage I	*P*[c]	FDR	Log_2 FC (SE) for Stage III-IV *v* Stage I	*P*[c]	FDR
Phylum *Synergistota*	47.9|0.004	43.5|0.0025	48.2|0.0039	−3.57 (1.35)	.009	0.047	−2.65 (1.39)	.058	0.293
Class *Synergistia*	47.9|0.004	43.1|0.0025	48.2|0.0039	−3.50 (1.33)	.009	0.084	−2.62 (1.36)	.056	0.363
Phylum *Firmicutes*									
Order *Haloplasmatales*	51.4|0.0087	62.9|0.0042	63.5|0.008	4.58 (1.55)	.004	0.132	5.37 (1.59)	9.87×10^{-4}	0.034

NOTE. A linear regression model was conducted for clr (centered log-ratio)-transformed taxa abundance with adjustment for age group, income levels, residence, age at menarche, regular menstrual cycle, menopausal status, number of live births, BMI levels, comorbidity, antibiotic use, diagnosis delay, TNM cancer stage, fiber intake, fat intake, carbohydrate intake, protein intake, and physical activity.

Abbreviations: FC, fold change; FDR, false discovery rates; Pre, prevalence; RA, relative abundance.

[a]Common taxa: prevalence ≥50% in the population; rare taxa: 10% ≤prevalence< 50% in the subpopulation.

[b]Median relative abundance for rare taxa was calculated among carriers.

[c]FDR was calculated at each taxonomic level by common and rare taxa.

TABLE A4. Association of Breast Cancer Subtypes With Gut Microbial Taxa Among Participants With Pre–Breast Cancer Surgery Stool Samples (n = 162)

Microbial Taxa	Pre (%)[a] | RA, Median (%) [b]				Log_2 FC (SE) for HR+/HER2-Positive *v* HR+/HER2-Negative	*P*	FDR[c]	Log_2 FC (SE) for HER2-Enriched *v* HR+/HER2-Negative	*P*	FDR[c]	Log_2 FC (SE) for Triple-Negative *v* HR+/HER2-Negative	*P*	FDR[c]
	HR+/HER2-Negative (n = 58)	HR+/HER2-Positive (n = 34)	HER2-Enriched (n = 40)	Triple-Negative (n = 30)									
Phylum *cyanobacteria*	100|0.0374	97.7|0.0357	100|0.0403	96.2|0.0475	−0.99 (0.36)	.007	0.089	0.09 (0.34)	.801	0.968	0.12 (0.38)	.751	0.963
Phylum *Elusimicrobiota*	15.3|0.0044	10.3| 0.0033	9.6|0.0051	13.5|0.0068	2.24 (0.85)	.009	0.046	1.06 (0.79)	.183	0.328	1.10 (0.88)	.212	0.289
Class *Elusimicrobia*	15.3|0.0044	10.3|0.0033	9.6|0.0051	13.5|0.0068	2.32 (0.85)	.008	0.067	1.02 (0.80)	.203	0.381	1.10 (0.89)	.218	0.552
Phylum *firmicutes A*													
Family *Acutalibacteraceae*													
Genus *Acutalibacter*	77.8|0.0041	83.9|0.0053	87.7|0.0056	78.8|0.0059	0.85 (0.83)	.309	0.870	2.87 (0.77)	3.06×10^{-4}	0.084	2.52 (0.86)	.004	0.539
Family *UBA1255*													
Genus *MGYG-HGUT-00728*	9.0|0.0016	16.1|0.0017	11.0|0.0018	15.4|0.0021	2.34 (0.59)	1.37×10^{-4}	0.051	−0.11 (0.56)	.834	0.992	1.67 (0.62)	.008	0.357
Family *Lachnospiraceae*													
Species *Dorea sp001185345*	11.8|0.0019	14.9|0.0028	12.3|0.0021	26.9|0.0015	0.11 (0.62)	.854	0.991	0.38 (0.58)	.515	0.997	2.94 (0.64)	1.05×10^{-5}	0.013
Species *MGYG-HGUT-01566*	59.0|0.0028	47.1|0.0023	57.5|0.0031	50.0|0.0019	−4.06 (1.02)	1.11×10^{-4}	0.084	−1.05 (0.95)	.272	0.979	−1.49 (1.05)	.159	0.950
Species *MGYG-HGUT-01722*	31.9|0.0017	21.8|0.0017	21.9|0.0014	13.5|0.0022	−2.83 (0.86)	.001	0.404	−1.13 (0.81)	.163	0.997	−3.80 (0.89)	4.04×10^{-5}	0.025

NOTE. A linear regression model was conducted for clr (centered log ratio)-transformed taxa abundance with adjustment for age group, income levels, residence, age at menarche, regular menstrual cycle, menopausal status, number of live births, BMI levels, comorbidity, antibiotic use, diagnosis delay, TNM cancer stage, fiber intake, fat intake, carbohydrate intake, protein intake, and physical activity.

Abbreviations: FC, fold change; FDR, false discovery rates; HER2, human epidermal growth factor receptor 2; HR, hormone receptor; Pre, prevalence; RA, relative abundance.

[a]Common taxa: prevalence ≥50% in the population; rare taxa: 10% ≤prevalence< 50% in the subpopulation.

[b]Median relative abundance for rare taxa was calculated among carriers.

[c]FDR was calculated at each taxonomic level by common and rare taxa.

TABLE A5. Association of BMI Levels With Gut Microbial Taxa Among Participants With Pre–Breast Cancer Surgery Stool Samples (n = 162)

Microbial Taxa	Pre (%)[a] I RA, Median (%)[b] Normal Weight (n = 98)	Under- Weight (n = 16)	Asian Overweight (n = 31)	Asian Obese (n = 17)	Log_2 FC (SE) for Underweight *v* Normal Weight	*P*	FDR[c]	Log_2 FC (SE) for Asian Overweight *v* Normal Weight	*P*	FDR[c]	Log_2 FC (SE) for Asian Obese *v* Normal Weight	*P*	FDR[c]
Phylum *Actinobacteriota*													
Species *actinomyces sp900323545*	17.7|0.0016	25.0|0.0018	9.1|0.0014	15.6|0.0015	3.31 (0.10)	.001	0.099	−0.84 (0.79)	.286	0.999	0.56 (1.01)	.581	0.964
Species *MGYG-HGUT-00928*	15.5|0.0091	25.0|0.0125	13.6|0.0092	15.6|0.0201	3.97 (1.21)	.001	0.099	0.17 (0.95)	.861	0.999	0.42 (1.21)	.728	0.970
Phylum *Bacteroidota*													
Species *MGYG-HGUT-00956*	20.3|0.0016	31.3|0.0016	24.2|0.0013	15.6|0.0013	3.69 (1.04)	5.40×10^{-4}	0.087	0.64 (0.82)	.433	0.999	1.41 (1.05)	.182	0.845
Species *MGYG-HGUT-00322*	35.4|0.0024	46.9|0.003	30.3|0.0024	31.2|0.0031	4.39 (1.21)	4.16×10^{-4}	0.086	0.88 (0.95)	.361	0.999	0.27 (1.22)	.82	0.991
Phylum *firmicutes*													
Order *Lactobacillales*	100|0.1988	100|0.6546	100|0.2485	100|0.1389	2.20 (0.70)	.002	0.368	0.11 (0.550)	.845	0.917	−0.54 (0.71)	.445	0.943
Species *Enterococcus A raffinosus*	14.2|0.0024	37.5|0.0134	13.6|0.005	12.5|0.0066	3.33 (0.97)	8.62×10^{-4}	0.087	−0.13 (0.77)	.862	0.999	0.67 (0.99)	.502	0.964
Species *Enterococcus B durans*	21.2|0.0034	43.8|0.0189	19.7|0.006	21.9|0.0033	4.30 (1.23)	6.77×10^{-4}	0.087	0.11 (0.97)	.907	0.999	0.70 (1.25)	.576	0.964
Species *Lactobacillus H mucosae*	24.3|0.0061	37.5|0.0133	18.2|0.0118	25.0|0.0253	4.69 (1.36)	7.85×10^{-4}	0.087	0.33 (1.07)	.756	0.999	0.03 (1.38)	.985	0.998
Species *MGYG-HGUT-00974*	15.0|0.0026	18.8|0.0031	15.2|0.0044	12.5|0.0027	3.69 (0.98)	2.44×10^{-4}	0.074	1.10 (0.77)	.154	0.999	−1.28 (0.99)	.199	0.844
Phylum *Firmicutes A*													
Species *MGYG-HGUT-02946*	13.3|0.0014	25.0|0.0023	12.2|0.0018	3.1|0.0015	2.50 (0.72)	7.77×10^{-4}	0.087	0.36 (0.56)	.527	0.999	0.05 (0.72)	.946	0.993
Species *MGYG-HGUT-02170*	30.5|0.003	37.5|0.0019	31.8|0.0022	28.1|0.0029	4.48 (1.21)	3.00×10^{-4}	0.074	2.56 (0.95)	.008	0.999	−1.29 (1.22)	.291	0.895
Species *MGYG-HGUT-01063*	11.9|0.015	18.7|0.0106	9.1|0.0069	0|0	2.18 (0.63)	7.64×10^{-4}	0.087	0.17 (0.50)	.727	0.999	−0.48 (0.64)	.450	0.948
Species *An200 sp003268275*	22.6|0.0031	25.0|0.0025	21.2|0.0015	34.4|0.0027	3.60 (1.09)	.001	0.099	−0.12 (0.86)	.887	0.999	1.70 (1.10)	.126	0.837
Species *MGYG-HGUT-02944*	22.1|0.0022	40.6|0.0021	24.2|0.0022	6.3|0.0019	3.77 (0.99)	2.18×10^{-4}	0.074	0.54 (0.78)	.493	0.999	−2.00 (1.00)	.048	0.837
Genus *Agathobaculum*	99.1|0.0855	93.8|0.0634	95.5|0.0788	96.9|0.0611	−1.58 (0.60)	.009	0.166	0.07 (0.470)	.883	0.974	−2.32 (0.61)	2.12×10^{-4}	0.058
Order *Peptostreptococcales*	100|0.3136	100|0.3433	100|0.2302	100|0.1999	1.00 (0.35)	.004	0.051	0.07 (90.27)	.788	0.918	0.14 (0.35)	.696	0.943
Species *Clostridioides difficile A*	10.6|0.0033	18.8|0.0055	10.6|0.0028	6.3|0.0013	2.38 (0.70)	9.21×10^{-4}	0.087	−0.29 (0.55)	.603	0.999	−0.32 (0.71)	.653	0.966
Species *MGYG-HGUT-00102*	17.7|0.0033	34.4|0.0034	22.7|0.0028	9.4|0.0019	4.04 (1.07)	2.54×10^{-4}	0.074	1.27 (0.84)	.134	0.999	0.12 (1.08)	.913	0.992
Species *Terrisporobacter othiniensis*	26.5|0.0044	43.8|0.0039	22.7|0.004	12.5|0.002	4.87 (1.29)	2.64×10^{-4}	0.074	0.53 (1.02)	.602	0.999	−1.41 (1.31)	.281	0.893
Phylum firmicutes C													
Order *Acidaminococcales*	98.2|0.5723	81.2|0.493	95.5|0.5044	96.9|0.4254	−4.04 (1.11)	4.09×10^{-4}	0.014	−0.42 (0.88)	.632	0.917	−2.27 (1.13)	.045	0.934
Family *Acidaminococcaceae*	98.2|0.5723	81.2|0.493	95.5|0.5044	96.9|0.4254	−3.97 (1.13)	6.03×10^{-4}	0.042	−0.56 (0.89)	.529	0.981	−2.55 (1.14)	.027	0.627

NOTE. A linear regression model was conducted for clr (centered log-ratio)-transformed taxa abundance with adjustment for age group, income levels, residence, age at menarche, regular menstrual cycle, menopausal status, number of live births, BMI levels, comorbidity, antibiotic use, diagnosis delay, TNM cancer stage, fiber intake, fat intake, carbohydrate intake, protein intake, and physical activity.

Abbreviations: FC, fold change; FDR, false discovery rates; Pre, prevalence; RA, relative abundance.

[a]Common taxa: prevalence ≥50% in the population; rare taxa: 10% ≤prevalence< 50% in the subpopulation.

[b]Median relative abundance for rare taxa was calculated among carriers.

[c]FDR was calculated at each taxonomic level by common and rare taxa.

TABLE A6. Association of Diagnosis Delay With Gut Microbial Taxa Among Participants With Pre–Breast Cancer Surgery Stool Samples (n = 162)

Microbial Taxa	Pre (%)[a] / RA, Median (%)[b]			Log_2 FC (SE) for Moderate *v* No Delay	*P*[c]	FDR	Log_2 FC (SE) for Stage III-IV *v* Stage I	*P*[c]	FDR
	No Delay (n = 65)	Moderate Delay (n = 58)	Serious Delay (n = 39)						
Phylum *Actinobacteriota*									
Species *MGYG-HGUT-02211*	8.4|0.0025	11.5|0.0023	14.1|0.0067	0.32 (0.48)	.507	0.873	1.88 (0.50)	2.43×10^{-4}	0.074
Species *Enorma massiliensis*	16.8|0.0072	8.0|0.0021	18.8|0.0032	−0.14 (0.56)	.805	0.970	2.33 (0.58)	1.10×10^{-4}	0.045
Phylum *Bacteroidota*									
Species *MGYG-HGUT-01644*	34.1|0.0024	25.7|0.0027	26.6|0.002	−2.88 (0.78)	3.61×10^{-4}	0.074	−1.53 (0.82)	.063	0.427
Phylum *Elusimicrobiota*	14.5|0.0024	10.6|0.0059	10.9|0.0225	−2.00 (0.71)	.006	0.028	−0.45 (0.74)	.544	0.961
Class *Elusimicrobia*	14.5|0.0024	10.6|0.0059	10.9|0.0225	−2.01 (0.72)	.006	0.052	−0.29 (0.75)	.695	0.944
Phylum *Firmicutes*									
Species *Faecalicoccus pleomorphus*	17.9|0.0056	4.4|0.0014	6.3|0.0028	−3.06 (0.61)	2.17×10^{-6}	0.003	−1.54 (0.64)	.018	0.370
Phylum *Firmicutes A*									
Order *4C28d-15*	95.5|0.0351	92.0|0.0249	92.2|0.0204	−2.39 (0.69)	7.76×10^{-6}	0.027	−1.54 (0.72)	.035	0.608
Genus *UBA1685*	37.4|0.0027	38.1|0.0026	20.3|0.0023	−1.39 (0.78)	.077	0.531	−3.36 (0.81)	5.85×10^{-5}	0.022
Species *UBA1685 sp002320595*	37.4|0.0027	38.1|0.0024	20.3|0.0023	−1.67 (0.80)	.040	0.384	−3.39 (0.83)	8.34×10^{-5}	0.045
Species *MGYG-HGUT-04052*	12.8|0.0128	8.8|0.0036	4.7|0.0013	−2.46 (0.67)	3.24×10^{-4}	0.074	−2.31 (0.69)	.001	0.231
Species *MGYG-HGUT-02835*	15.1|0.0016	8.0|0.0043	17.2|0.0021	−2.31 (0.61)	2.55×10^{-4}	0.074	−0.06 (0.64)	.930	0.971
Species *MGYG-HGUT-02946*	11.7|0.0014	11.5|0.0023	20.3|0.0015	0.12 (0.23)	.818	0.971	2.23 (0.52)	3.77×10^{-5}	0.045
Species *MGYG-HGUT-04098*	15.1|0.0019	9.7|0.0018	10.9|0.0012	−1.89 (0.51)	2.82×10^{-4}	0.074	−0.80 (0.53)	.130	0.489
Species *MGYG-HGUT-00746*	22.3|0.0024	11.5|0.002	10.9|0.0023	−2.61 (0.58)	1.33×10^{-5}	0.008	−1.48 (0.60)	.015	0.370

NOTE. A linear regression model was conducted for clr- (centered log-ratio)-transformed taxa abundance with adjustment for age group, income levels, residence, age at menarche, regular menstrual cycle, menopausal status, number of live births, BMI levels, comorbidity, antibiotic use, diagnosis delay, TNM cancer stage, fiber intake, fat intake, carbohydrate intake, protein intake, and physical activity.

Abbreviations: FC, fold change; FDR, false discovery rates; Prem prevalence; RA, relative abundance.

[a]Common taxa: prevalence ≥50% in the population; rare taxa: 10% ≤prevalence< 50% in the subpopulation.

[b]Median relative abundance for rare taxa was calculated among carriers.

[c]FDR was calculated at each taxonomic level by common and rare taxa.

TABLE A7. Participant's Diagnosis Delay by Cancer Stage Separately for Presurgery and Postsurgery (N = 356)

		TNM Cancer Stage			
Diagnosis Delay	Overall	I	II	III-IV	*P*
Patients with presurgery stool samples					
No delay	65 (40.1)	7 (50.0)	36 (47.4)	22 (30.6)	.046
Moderate delay	58 (35.8)	4 (28.6)	29 (38.2)	25 (34.7)	
Serious delay	39 (24.1)	3 (21.4)	11 (14.5)	25 (34.7)	
Patients with postsurgery stool samples					
No delay	114 (58.8)	41 (68.3)	64 (52.9)	9 (69.2)	.297
Moderate delay	55 (28.3)	12 (20.0)	40 (33.0)	3 (23.1)	
Serious delay	25 (12.9)	7 (11.7)	17 (14.1)	1 (7.7)	
Overall					
No delay	179 (50.1)	48 (64.9)	100 (50.8)	31 (36.5)	.001
Moderate delay	113 (31.7)	16 (21.6)	69 (35.0)	28 (32.9)	
Serious delay	64 (18.0)	10 (13.5)	28 (14.2)	26 (30.6)	

TABLE A8. Association of Antibiotic Use With Gut Microbial Taxa Among Participants With Pre–Breast Cancer Surgery Stool Samples (n = 162)

	Pre (%)[a] | RA, Median (%)[b]				
Microbial Taxa	No (n = 112)	Yes (n = 50)	Log_2 FC (SE) for Yes *v* No	*P*[c]	FDR
Phylum *Actinobacteriota*					
Family *Eggerthellaceae*	100|0.1076	98.8|0.0978	−0.76 (0.26)	.004	0.053
Phylum *Eremiobacterota*	12.5|0.0016	18.9|0.0017	1.34 (0.55)	.017	0.043
Class *UBP9*	12.5|0.0016	18.9|0.0017	1.24 (0.58)	.033	0.088
Phylum *Firmicutes*					
Family *Erysipelatoclostridiaceae*	100|0.4456	99.6|0.4261	−1.10 (0.29)	1.78×10^{-4}	0.009
Phylum *Firmicutes A*					
Family *Lachnospiraceae*	100|19.5676	100|12.699	−0.57 (0.19)	.004	0.053
Family *Butyricicoccaceae*	100|0.176	97.9|0.0961	−1.10 (0.34)	.001	0.033
Family *CAG-508*	100|0.0715	98.4|0.0519	−1.11 (0.30)	2.59×10^{-4}	0.009
Phylum *Proteobacteria*					
Family *Pasteurellaceae*	73.2|0.0361	47.9|0.0206	−3.12 (1.16)	.008	0.095
Phylum *Spirochaetota*	/				
Class *Brachyspirae*	19.6|0.0022	7.8|0.0029	−1.31 (0.63)	.039	0.088
Phylum *Synergistota*	41.1|0.0023	47.1|0.0033	2.58 (0.81)	.002	0.009
Class *Synergistia*	41.1|0.0023	47.1|0.0033	2.49 (0.80)	.002	0.020
Order *Synergistales*	41.1|0.0023	47.1|0.0033	2.48 (0.80)	.002	0.040
Phylum *Verrucomicrobiota*					
Class *Verrucomicrobiae*	38.4|0.0122	48.8|0.0140	2.08 (0.99)	.039	0.088

NOTE. A linear regression model was conducted for clr (centered log-ratio)-transformed taxa abundance with adjustment for age group, income levels, residence, age at menarche, regular menstrual cycle, menopausal status, number of live births, BMI levels, comorbidity, antibiotic use, diagnosis delay, TNM cancer stage, fiber intake, fat intake, carbohydrate intake, protein intake, and physical activity.

Abbreviations: FC, fold change; FDR, false discovery rates; Pre, prevalence; RA, relative abundance.

[a]Common taxa: prevalence ≥50% in the population; rare taxa: 10% ≤prevalence< 50% in the subpopulation.

[b]Median relative abundance for rare taxa was calculated among carriers.

[c]FDR was calculated at each taxonomic level by common and rare taxa.

TABLE A9. Association of Fiber Intake With Gut Microbial Taxa Among Participants With Pre–Breast Cancer Surgery Stool Samples (n = 162)

Microbial Taxa	Pre (%)[a] | RA, Median (%)[b] Low (T1) (n = 49)	High (T3) (n = 63)	Log_2 FC (SE) for High (T3) *v* Low (T1)	*P*[c]	FDR
Phylum *Actinobacteriota*					
Order *Actinomycetales*					
Species *MGYG-HGUT-03800*	8.4|0.0026	11.0|0.0039	2.35 (0.64)	3.24×10^{-4}	0.031
Species *Bifidobacterium gallinarum*	13.5|0.0027	17.8|0.0026	2.84 (0.76)	2.58×10^{-4}	0.031
Species *Bifidobacterium pullorum*	11.8|0.0037	16.9|0.0023	3.32 (0.70)	4.77×10^{-4}	0.005
Species *Bifidobacterium scardovii*	27.7|0.0076	38.1|0.007	4.01 (1.24)	.002	0.087
Species *MGYG-HGUT-00775*	7.6|0.0022	11.9|0.0017	2.92 (0.63)	9.13×10^{-4}	0.005
Order *Mycobacteriales*	17.6|0.0033	22.9|0.0041	2.55 (0.82)	.002	0.037
Family *Mycobacteriaceae*	17.6|0.0033	22.0|0.0042	2.60 (0.83)	.002	0.037
Species *MGYG-HGUT-00759*	13.4|0.0032	13.6|0.0035	2.63 (0.70)	2.47×10^{-4}	0.031
Order *Propionibacteriales*	17.6|0.0032	17.8|0.0045	2.07 (0.80)	.011	0.09
Species *MGYG-HGUT-04127*	17.6|0.0032	16.9|0.0045	3.02 (0.82)	3.30×10^{-4}	0.031
Order *Coriobacteriales*					
Species *MGYG-HGUT-00949*	10.9|0.0059	15.3|0.0118	3.27 (0.89)	3.32×10^{-4}	0.031
Species *MGYG-HGUT-00955*	7.6|0.0045	13.6|0.0049	3.03 (0.67)	1.51×10^{-5}	0.005
Species *MGYG-HGUT-01583*	16.8|0.0099	16.9|0.0159	3.42 (1.03)	.001	0.077
Species *MGYG-HGUT-01997*	9.2|0.0054	10.2|0.0101	3.19 (0.78)	7.5×10^{-5}	0.018
Species *MGYG-HGUT-02864*	12.6|0.005	16.1|0.0076	3.12 (0.81)	1.99×10^{-4}	0.031
Species *MGYG-HGUT-02968*	13.4|0.009	22.0|0.0103	3.28 (.102)	.002	0.087
Phylum Bacteroidota					
Species *Bacteroides clarus*	90.8|0.0253	94.9|0.0253	3.35 (0.96)	6.77×10^{-4}	0.064
Species *Bacteroides graminisolvens*	13.4|0.0021	11.9|0.0035	2.37 (0.74)	.002	0.092
Species *Bacteroides sp002491635*	95.8|0.0555	99.2|0.0679	2.53 (0.75)	.001	0.077
Species *Bacteroides stercorirosoris*	97.5|0.039	98.3|0.0296	1.96 (0.57)	8.28×10^{-4}	0.070
Species *Bacteroides thetaiotaomicron*	98.3|0.6124	99.2|0.606	3.24 (0.90)	4.37×10^{-4}	0.064
Species *MGYG-HGUT-00013*	96.6|0.0682	98.3|0.0948	2.82 (0.74)	2.12×10^{-4}	0.064
Species *MGYG-HGUT-01977*	98.3|0.0441	96.6|0.039	2.19 (0.59)	2.90×10^{-4}	0.064
Species *MGYG-HGUT-03351*	99.2|0.1535	99.2|0.1715	2.45 (0.68)	4.77×10^{-4}	0.064
Species *Prevotella melaninogenica*	92.4|0.0061	94.9|0.0065	2.16 (0.59)	4.13×10^{-4}	0.064
Species *Prevotella oris*	95.0|0.0115	94.9|0.0122	1.85 (0.53)	6.28×10^{-4}	0.064
Species *MGYG-HGUT-01608*	10.1|0.0027	12.7|0.0022	2.45 (0.75)	.001	0.084
Species *MGYG-HGUT-01951*	10.1|0.002	17.8|0.0021	2.87 (0.79)	3.91×10^{-4}	0.034
Species *MGYG-HGUT-02867*	18.5|0.0015	23.7|0.0016	2.65 (0.83)	.002	0.093
Phylum *Firmicutes A*	100|28.3520	100|27.7037	−0.80 (0.25)	.002	0.025
Class *Clostridia*	100|28.3447	100|27.7037	−0.69 (0.250	.007	0.046
Species *MGYG-HGUT-02623*	26.1|0.0018	31.4|0.0024	3.95(1.11)	4.96×10^{-4}	0.038
Species *MGYG-HGUT-02946*	11.8|0.002	11.9|0.0014	2.27 (0.65)	7.69×10^{-4}	0.056
Species *MGYG-HGUT-01617*	16.0|0.0018	17.8|0.0025	2.91 (0.72)	9.42×10^{-5}	0.019
Species *MGYG-HGUT-03316*	100|0.041	100|0.0403	2.19 (0.62)	5.89×10^{-4}	0.064
Family *MGYG-HGUT-04273*	27.7|0.0035	35.6|0.0032	3.17 (1.11)	.005	0.088
Species *MGYG-HGUT-04273*	27.7|0.0035	35.6|0.0032	3.72 (1.13)	.001	0.084
Family *MGYG-HGUT-00495*	19.3|0.0021	17.8|0.002	−2.28 (0.78)	.004	0.089
Species *MGYG-HGUT-04496*	33.6|0.0032	41.5|0.0039	4.50 (1.25)	4.34×10^{-4}	0.035
Genus *MGYG-HGUT-04035*	40.3|0.002	22.0|0.0018	−3.48 (0.93)	2.51×10^{-4}	0.092
Species *MGYG-HGUT-04449*	8.4|0.0025	15.3|0.0027	3.97 (0.89)	1.69×10^{-5}	0.005
Class *Mahella*	55.5|0.0022	46.6|0.0018	−3.06 (1.08)	.005	0.046

(continued on following page)

TABLE A9. Association of Fiber Intake With Gut Microbial Taxa Among Participants With Pre–Breast Cancer Surgery Stool Samples (n = 162) (continued)

Microbial Taxa	Pre (%)[a] | RA, Median (%)[b]		Log_2 FC (SE) for High (T3) *v* Low (T1)	*P*[c]	FDR
	Low (T1) (n = 49)	High (T3) (n = 63)			
Phylum *Firmicutes C*					
Family *Megasphaeraceae*	45.4|0.0062	43.2|0.0088	−4.20 (1.45)	.004	0.089
Species *MGYG-HGUT-03352*	29.4|0.0022	44.9|0.0039	4.51 (1.19)	2.23×10^{-4}	0.031

NOTE. A linear regression model was conducted for clr (centered log-ratio)-transformed taxa abundance with adjustment for age group, income levels, residence, age at menarche, regular menstrual cycle, menopausal status, number of live births, BMI levels, comorbidity, antibiotic use, diagnosis delay, TNM cancer stage, fiber intake, fat intake, carbohydrate intake, protein intake, and physical activity.
Abbreviations: FC, fold change; FDR, false discovery rates; Pre, prevalence; RA, relative abundance.
[a]Common taxa: prevalence ≥50% in the population; rare taxa: 10% ≤prevalence< 50% in the subpopulation.
[b]Median relative abundance for rare taxa was calculated among carriers.
[c]FDR was calculated at each taxonomic level by common and rare taxa.