

RESEARCH

Open Access



Relationship between tumor microbiota transcriptional activity and gene expression in breast cancer

Enuo Liu^{1,2}, Fan Zhang⁴, Tiansheng Xu⁴, Luyi Ye¹, Sean Si Qian Ma¹ and Zai-Si Ji^{1,3*} 

Abstract

Background A few studies have reported the distribution of the microbiota in breast cancer tissues, but few reports have compared the microbiota in different subtypes of breast cancer tissue. Moreover, no study has reported on the relationship between the microbiota and gene expression in breast tumor.

Methods Sections of formalin-fixed paraffin-embedded (FFPE) tissue were prepared from the breast tumors of 70 patients and were subjected to microarray analysis to identify gene expression profiles. The same total RNA samples were also used to analyze the microbiota activity in tumor tissues by performing 16 S rRNA sequencing and internal transcribed spacer (ITS) sequencing of reverse transcript cDNA with Illumina Miseq. Pearson's correlation coefficient was used for calculating the correlation between microbial relative activity and gene expression.

Results The microbiota transcriptional activity of 70 FFPE samples mainly consisted of the phyla *Bacteroidetes*, *Firmicutes* and *Proteobacteria*. *Prevotella_9*, *Bacteroides* and *Alloprevotella* were the most active genera in ER+/HER2-, ER+/HER2+ and ER-/HER2+ tumors, while triple-negative samples exhibited a higher activity of *Lactobacillus*. In ER-negative samples (triple-negative and ER-/HER2+), 479 genes, including the breast carcinogenesis genes phospholipase A2, histone cluster 2, Crk-like, and cyclin D1, were significantly positive associated with the activity of *Lactobacillus*.

Conclusion This was the first study to clarify an association between the breast tumor microbiota transcriptional activity and the expression of carcinogenesis genes in ER-negative breast cancer. Changes in the microbiota of breast tissue induced by external factors might be one of the key causes of ER negative breast cancer.

Keywords Tumor microbiota, ER-negative breast cancer, Carcinogenesis, *Lactobacillus*, *Bacteroides*

*Correspondence:

Zai-Si Ji

jizaisi@sibpt.com

¹NHC Key Laboratory of Reproduction Regulation, Shanghai Institute for Biomedical and Pharmaceutical Technologies, Fudan University, 2140 Xietu road, 200032 Shanghai, China

²College of Food Sciences and Technology, Shanghai Ocean University, 999 Hucheng Road, Shanghai, China

³Division of Research and Development, Meiji Co., Ltd, 1-29-1 Nanakuni, 192-0919, TokyoHachioji, Japan

⁴Shanghai OE Biotech Co., Ltd, 1188 Lianhang road, 201114 Shanghai, China



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Introduction

Breast cancer is a hormonally driven cancer especially for estrogen receptor (ER) positive breast cancer [1]. For patients with ER-positive breast cancer, hormonal therapy such as treated with tamoxifen for 5 years reduces the annual breast cancer death rate by 31% [2], and tamoxifen therapy for 10 years can halve breast cancer mortality during the second decade after diagnosis [3]. Conversely, such hormonal therapy has little or no effect on recurrence or mortality in patients with ER-negative breast cancer [4]. The majority of prognostic markers, which are overexpressed in patients with good prognosis in ER-negative breast cancer including HER2-overexpressing breast cancer and triple-negative breast cancer [5], associated with the activation of complement and immune response pathways [6]. Immunotherapies have been used to treat ER-negative breast cancer [7].

Several studies reported unique microbial signatures in tissue in most major types of cancer (breast, prostate, lung, ovary, pancreas, bone, melanoma and brain tumors), and breast cancer has a particularly rich and diverse microbiota [8, 9]. In samples collected from outside of the marginal zone positioned approximately 5 cm away from the tumor's edge in Italian subjects, researchers identified more similarities than differences between breast tumors and adjacent normal tissues [10]. The most active taxa in breast tissue samples obtained from Canadian subjects, which were collected outside the marginal zone 5 cm from the tumor, were *Bacillus* (11.4%), *Acinetobacter* (10.0%), *Enterobacteriaceae* (8.3%), *Pseudomonas* (6.5%), *Staphylococcus* (6.5%), *Propionibacterium* (5.8%), *Comamonadaceae* (5.7%), *Gammaproteobacteria* (5.0%), and *Prevotella* (5.0%) [11]. In breast tissue samples collected in Ireland (taken at least 5 cm from the primary tumor site), the most active taxa were *Enterobacteriaceae* (30.8%), *Staphylococcus* (12.7%), *Listeria welshimeri* (12.1%), *Propionibacterium* (10.1%), and *Pseudomonas* (5.3%) [11], indicated that there might be significant differences among ethnic groups. A higher relative abundance of *Bacillus* (*Firmicutes*), *Enterobacteriaceae* (*Proteobacteria*) and *Staphylococcus* (*Firmicutes*) was reported in Canadian women with breast cancer than in those without breast cancer [12]. Conversely, an increased relative abundance of gram-positive organisms including *Corynebacterium* (*Actinobacteria*), *Staphylococcus* (*Firmicutes*), *Actinomyces* (*Actinobacteria*), and *Propionibacteriaceae* (*Actinobacteria*), and a decreased relative abundance of *Methylobacterium* (*Proteobacteria*) were reported in American women with breast cancer compared with those in American women without breast cancer [13].

Although the distinct microbial characteristics of breast tumor tissues have been compared with those of normal adjacent tissue, breast skin tissue, breast skin

swabs, and buccal swabs [14], the association of the microbial community with gene expression has not been reported, and few analysis based on the ER, progesterone receptor (PR) and human epidermal growth factor receptor-2 (HER2) status has been described. In this study, we focused on the microbiota activity of subtype breast cancer tissues by 16s RNA / ITS (Internal transcribed spacer) sequencing of cDNA reverse transcribed from RNA, and its association with gene expression obtained by RNA microarray hybridization. The results could provide not only the microbiota of breast cancer subtype, but also the relationship between microbiota and gene expression which might contribute new insights into the clinical treatment of breast cancer.

Materials and methods

Study design and sample collection

Seventy formalin-fixed paraffin-embedded (FFPE) samples were collected in Huangpu District Central Hospital of Shanghai, China. The samples were prepared from fresh ductal tumors in women with breast cancer (age, 43–76 years) undergoing breast surgery at the operation room of the same hospital. The tumor tissues were preserved in formalin and transferred to the laboratory immediately. FFPE processing was performed in the sterile clean room using sterilized materials and new open reagents. FFPE samples were divided into four groups based on the ER, PR and HER2 positive (+) or negative (-) status, including 7 samples from ER+/HER2- patients, 48 samples from ER+/HER2+ types, 8 samples from ER-/HER2+ patients and 7 samples from triple-negative patients confirmed by immunohistochemistry.

RNA extraction, microarray hybridization, and data analysis

A RecoverAll™ Total Nucleic Acid Isolation Kit (Ambion, AM1975) was used to extract total RNA from 10- μ m FFPE tumor tissues and two environmental controls according to the manufacturer's protocol. As environmental controls, the blank digestion buffer added protease was processed in parallel with the tumor tissue samples. The process was performed in clean room mainly involved deparaffinization, protease digestion, nucleic acid isolation, DNase digestion, and final purification. RNA purity and quantification were evaluated using a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, USA). RNA integrity was assessed using an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA).

Gene expression profiles were investigated using an Affymetrix Human Clariom™D Assay (OE Biotechnology Co., Ltd., Shanghai, China). Sample labeling, microarray hybridization, and washing were performed according to the manufacturer's standard protocols. Briefly, total RNA

was transcribed into double-stranded cDNA and purified. Next, second-cycle cDNAs were used for fragmentation and biotin labeling. Labeled cDNA samples were hybridized to microarrays. After washing and staining, the microarrays were scanned using an Affymetrix Scanner 3000 (Affymetrix).

For microarray analysis, Affymetrix GeneChip Command Console (version 4.0, Affymetrix) software was used to extract raw data. Next, Expression Console (version 1.3.1, Affymetrix) software was used for RMA normalization for gene analysis. Differentially expressed genes were then identified using Genespring software (version 13.1; Agilent Technologies). The difference in expression between each group of microarray samples was expressed as the fold-change. A heatmap generated using the clustermap function of the Seaborn package in Python was used to illustrate gene expression among samples. GO analysis and KEGG analysis were applied to determine the roles of these differentially expressed mRNAs played in these GO terms or pathways. The microarray data have been uploaded to the GEO database under the accession codes GSE183231 and GSE185439.

16 S rRNA sequencing, bacterial diversity, and taxonomic analysis

After extraction and quality control, 0.5 µg of RNA from each sample was used for reverse transcription (10 µl), and the cDNA was diluted to 100 µl for subsequent 16 S library construction. PCR amplification of the V3-V4 hypervariable regions of the bacterial 16 S rRNA was performed in a 25 µl reaction volume using universal primer pairs (343 F: 5'-TACGGRAGGCAGCAG-3'; 798R: 5'-AGGGTATCTAATCCT-3'). The reverse primer contained a sample barcode, and both primers were linked to an Illumina sequencing adapter (Illumina, San Diego, CA, USA). The amplicon products were purified using Agencourt AMPure XP beads (Beckman Coulter Co., USA) and quantified using a Qubit dsDNA assay kit. Sequencing was performed on an Illumina Miseq with two paired-end read cycles of 300 bases each. The 16 S rRNA sequencing data have been uploaded to the GenBank Sequence Read Archive under accession code PRJNA769523.

Paired-end reads were preprocessed using Trimmomatic software [15] to detect and remove ambiguous bases (N) in the data quality control step. Clean reads were subjected to primer sequence removal and clustering to generate operational taxonomic units (OTUs) using VSEARCH software with a similarity cutoff of 97% [16]. The representative read of each OTU was selected using the QIIME package, and then all representative reads were annotated using the Silva database (Version 123) and the RDP classifier [17]. The microbial richness and diversity of breast cancer samples was estimated

using indices of alpha diversity including Good's coverage, the Chao1 index (community richness) [18] and Shannon index (diversity) [19]. The UniFrac distance matrix generated using QIIME software was used for unweighted UniFrac nonmetric multidimensional scaling (NMDS) to display bacterial beta diversity.

Internal transcribed spacer (ITS) sequencing, fungal diversity, and taxonomic analysis

For fungal diversity analysis, the ITS1 variable regions were amplified using the universal primer pairs (ITS1F: 5'-CTTGGTCATTTAGAGGAAGTA-3'; ITS2: 5'-GCT-GCGTTCTTCATCGATGC-3') after reverse transcription. After PCR product purification and library quantification, sequencing was performed on an Illumina MiSeq with two paired-end read cycles of 300 bases each. (Illumina Inc., San Diego, CA). The ITS sequencing data have been uploaded to the GenBank Sequence Read Archive under accession code PRJNA769523.

Data quality control, OTU generation, and representative sequence selection were performed using the same methods described for 16 S rRNA sequencing. All representative reads were annotated and searched against the UNITE and NCBI databases (ITS rDNA) using BLAST [20]. Then, the fungal diversity including alpha diversity (Good's coverage, Chao1 index and Shannon index) and beta diversity (UniFrac distance for NMDS) was used to assess the distribution of fungi in each sample and the differences between groups. R package was used to analyze the significant differences among four groups using Analysis of Variance (ANOVA) statistical test.

Correlation analysis

The correlation between gene expression and the relative activity of bacteria was calculated using Pearson's correlation coefficient. The threshold set for a significant correlation was $r > 0.7$ and $p < 0.05$.

Results

Data of surgically treated patients

The clinical and pathological features of the 70 patients are summarized in Table 1. Compared to the triple-negative group, the other three groups had no significant difference for age at surgery ($p < 0.05$). There also was no significant difference among the four groups regarding tumor size, lymph nodes, and tumor grade by chi-square test.

Bacteria in four types of breast tumor

16 S rRNA gene amplicon sequencing of the V3-V4 hypervariable regions was performed using cDNA samples reverse transcript from total RNA extracted from 70 FFPE samples. In total, 4,846,572 high-quality sequences of the 16 S rRNA gene in 70 samples were obtained by

Table 1 Clinical and pathological features of 70 breast cancer patients in surgery

	ER+/HER2-	ER+/HER2+	ER-/HER2+	Triple -negative	P value
subjects (n)	7	48	8	7	
age at surgery (year)	48.9±10.1	54.0±8.4	53.9±10.6	59.0±11.5	
<i>P</i> value	0.105 *	0.167 *	0.386 *		
ER status (IHC)	+	+	-	-	
PR status (IHC)	+	+	-	-	
HER2 status (IHC)	-	+	+	-	
lymph node					
Positive (n)	5 (71%)	24 (50%)	5 (63%)	5 (71%)	0.497#
Negative (n)	2 (29%)	24 (50%)	3 (38%)	2 (29%)	
Tumor size					0.943#
> 20 mm (n)	3 (43%)	21 (44%)	4 (50%)	4 (57%)	
≤ 20 mm (n)	4 (57%)	27 (56%)	4 (50%)	3 (43%)	
Histological grade					
grade 1	0 (0%)	7 (15%)	1 (13%)	1 (14%)	0.153#
grade 2	4 (57%)	30 (63%)	2 (25%)	3 (43%)	
grade 3	3 (43%)	11 (23%)	5 (63%)	3 (43%)	

ER+: ER-positive > 10% tumor cells; ER-: ER-positive < 10% tumor cells. PR+: PR positive > 10% tumor cells; PR-: PR-positive < 10% tumor cells. HER2+: HER2 positive > 10% tumor cells; HER2-: HER2 positive < 10% tumor cells; Data are means ± standard deviations or presented as number of patients (%). Percentages are calculated from the total number of patients with known values. * Student's *t* test; *p* value: ER+/HER2- versus triple-negative; or ER+/HER2+ versus triple-negative; or ER-/HER2+ versus triple-negative; # chi-square test

high-throughput DNA sequencing, and 23,032 OTUs were obtained. Good's coverage exceeded 95.5% for the four groups, indicating that the sequencing depth was sufficient for tissue microbiota investigation in patients with breast cancer (Fig. 1A). According to the alpha diversity, triple-negative tumors exhibited lower richness (Chao index, Fig. 1B) for the lower 8018 OTUs detected than other three groups 12,658 OTUs (ER-/HER2+), 12,721 OTUs (ER+/HER2-), 19,974 OTUs (ER+/HER2+), respectively. However, there was no significant difference in the diversity indicated by the Shannon index among the four groups (Fig. 1C). The NMDS analysis (Fig. 1D) revealed differences in the tissue microbiota composition of triple-negative and other three groups tumors. On the other hand, the two environmental controls contained completely different microbiota abundance compared to the 70 tumor samples (supplementary file 1).

A total of 35 phyla were found and the relative activity exceeded 85% for the three major bacterial phyla *Bacteroidetes*, *Firmicutes*, and *Proteobacteria* (Table 2). The triple-negative group displayed a significantly lower relative activity of *Bacteroidetes* than other three groups ($p < 0.01$ or $p < 0.001$), and a significantly higher relative activity of *Firmicutes* than ER+/HER2- ($p < 0.05$) and ER+/HER2+ group ($p < 0.001$) (Fig. 2A). At the genus level, the triple-negative group exhibited a significantly higher relative activity of *Lactobacillus* ($p < 0.01$ or $p < 0.001$), and significantly lower relative activity of *Prevotella_9*, *Allorivotella* and *Bacteroides* ($p < 0.01$ or $p < 0.001$) than other three groups (Fig. 2B). *Prevotella_9*, *Allorivotella* and *Bacteroides* are anaerobes belongings to Bacteroidetes,

whereas *Lactobacillus* consists of facultative anaerobes belongings to Firmicutes.

Fungal diversity analysis in four types of breast tumor

ITS1 variable region sequencing was performed on cDNA samples reverse transcript from total RNA extracted from 70 FFPE samples. In total, 3,966,334 high-quality sequences of ITS1 genes were obtained by high-throughput DNA sequencing, and 3,441 OTUs were obtained. Good's coverage exceeded 99.6% for the four groups (Supplementary file 2 A), indicating that the sequencing depth was sufficient for tissue microbiota investigation in patients with breast cancer. According to the alpha diversity (Supplementary file 2B-C), four groups displayed similar richness (Chao1 index) and ER-/HER2+ group displayed higher diversity (Shannon index, $p < 0.01$), indicating the better homogeneity of OUT sequences in ER-/HER2+ group. NMDS analysis (Supplementary file 2D) revealed that the fungal composition was similar among triple-negative, ER-/HER2+, ER+/HER2- and ER+/HER2+ tumors.

Regarding the fungal community, the relative activity of phylum *Ascomycota* exceeded 30% in all of the groups (Supplementary file 3). The groups exhibited a similar relative activity (FDR $p > 0.05$) at the top 4 phyla (mean relative activity > 0.01) including *Ascomycota*, *Basidiomycota*, *Rozellomycota*, and *Mortierellomycota* (Supplementary file 3), and also exhibited a similar relative activity (FDR $p > 0.05$) at class, order, family and genus using ANOVA statistical test. On the other hand, the two environmental

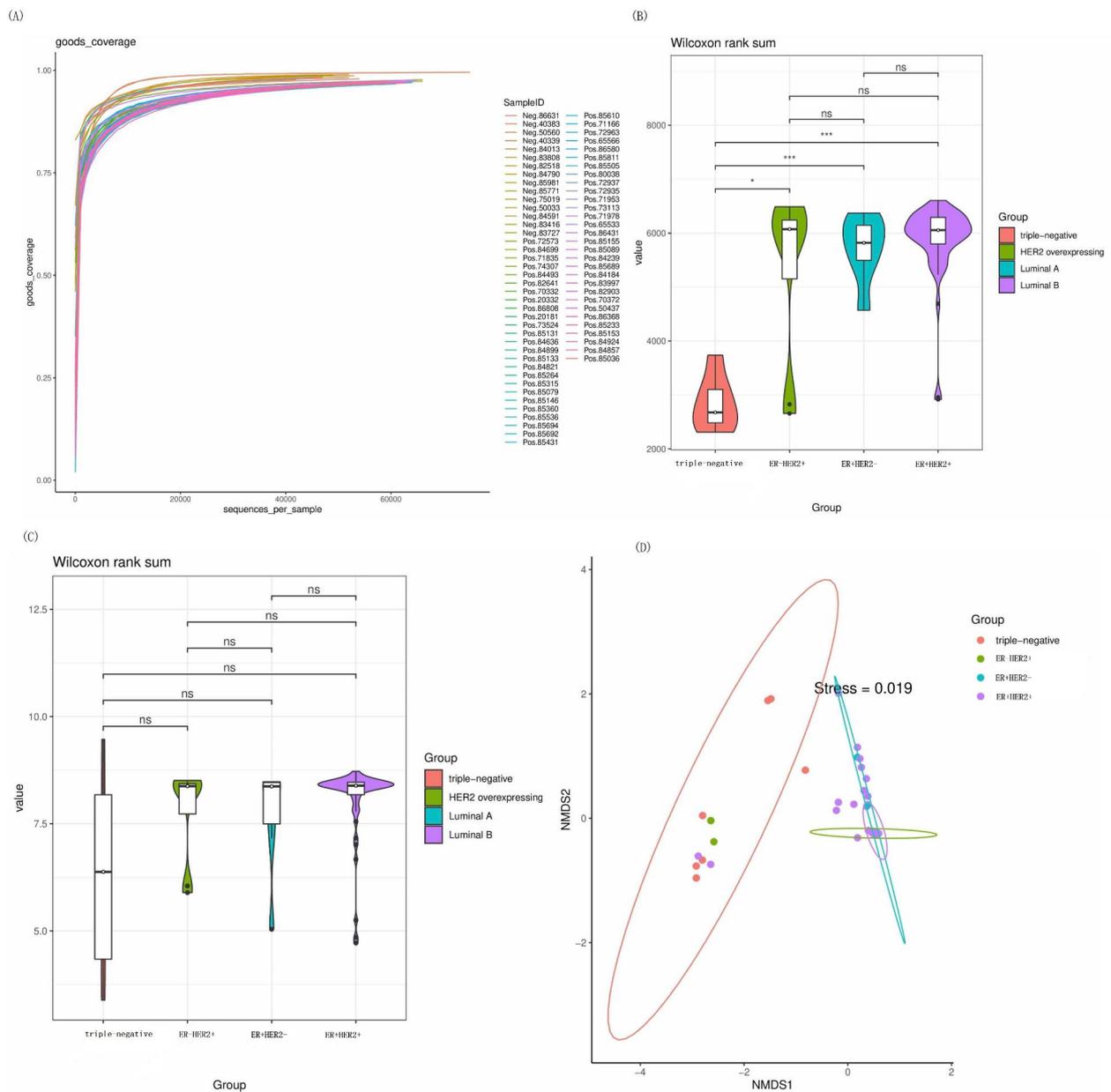


Fig. 1 Analysis of bacterial profiles in four groups. (A) Good's coverage for the four groups; (B) comparison of bacterial richness (Chao 1) among the four groups; (C) comparison of bacterial diversity (Shannon) among the four groups; (D) analysis of beta diversity (NMDS) among the four groups. Wilcoxon rank sum test was used for comparisons. * p value < 0.05, ** p value < 0.01, *** p value < 0.001

controls failed in amplification step, indicating no fungi existed (supplementary file 1).

Microarray analysis of gene expression in ER-negative and ER-positive breast tumors

Gene expression profiling was performed on RNA samples obtained from 70 breast tumors using the Affymetrix Human Clariom™ Assay containing 135,750 probe sets including 18,858 Entrez gene RNAs and 66,845 lincRNAs. Changes in the expression of all the RNAs corresponding to the probe sets on the microarray were

depicted in Supplementary file 4 A. Using the arbitrarily chosen criteria of >1.5-fold change in expression and $p < 0.05$, 478 transcripts were determined to be differentially expressed between ER-negative (triple-negative and ER-/HER2+) and ER-positive (ER+/HER2- and ER+/HER2+) tumors. In total, 440 genes were upregulated in ER-positive tumors including 149 coding genes, 121 noncoding genes, and 170 other genes (Supplementary file 4B). In particular, the expression of arginine and glutamate rich 1, a MED1-interacting protein required for estrogen-dependent gene transcription and breast

Table 2 Mean relative activity of the top 5 bacteria phyla and top 10 genera in ER-/HER2+ and other three groups

Taxa-phylum	Mean of relative activity (%)			
	ER+/HER2-	ER+/HER2+	ER-/HER2+	tri-ple-negative
<i>Firmicutes</i>	25.21	30.25	37.69	52.20
<i>Bacteroidetes</i>	57.21	60.53	53.55	18.65
<i>Proteobacteria</i>	17.04	8.07	4.10	17.96
<i>Actinobacteria</i>	0.21	0.57	3.88	6.61
<i>Epsilonbacteraeota</i>	0.06	0.20	0.13	1.52
Taxa-genus				
<i>Prevotella_9</i>	14.40	14.72	12.42	0.54
<i>Lactobacillus</i>	0.20	2.89	12.38	38.62
<i>Bacteroides</i>	12.88	14.24	12.11	5.54
<i>Alloprevotella</i>	8.87	9.25	8.21	1.12
<i>Blautia</i>	2.99	3.19	2.98	0.87
<i>Christensenellaceae_R-7_group</i>	1.88	1.94	1.78	0.08
<i>Ruminococcus_1</i>	1.93	1.95	1.70	0.09
<i>Prevotellaceae_Ga6A1_group</i>	1.78	1.78	1.56	0.05
<i>Ruminococcaceae_UCG-014</i>	1.50	1.50	1.33	0.45
<i>[Eubacterium]_coprostanoligenes_group</i>	1.38	1.40	1.22	0.29

cancer cell growth [21], was 4.9-fold higher in ER-negative tumors than ER-positive tumors ($p < 0.001$, FDR $p = 0.0267$). Moreover, 38 genes had lower expression in ER-negative tumors, including 8 coding genes, 21 non-coding genes, and 9 other genes (Supplementary file 4B).

As revealed by Gene Ontology (GO) enrichment analysis between ER-negative and ER-positive groups, the DEGs were associated with 160 GO terms including 102 in the category of “biological process”, 32 in the category of “cellular component” and 26 in the category of “molecular function”. The top 20 GO terms are presented in Supplementary file 4 C. Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis revealed differentially expressed genes in six pathways such as “One carbon pool by folate” (two genes, $p = 0.003$, Supplementary file 5) [22].

Correlation analysis between gene expression and the activity of *Lactobacillus* in ER-negative tumors

A strong correlation between gene expression and the relative activity of *Lactobacillus* was identified ($r > 0.7$, $p < 0.05$) in the ER-negative group. Significant correlations were observed for phospholipase A2 group IIE (PLA2G2E, Table 3), and the expression of 219 other genes including 6 keratin-associated protein genes was correlated with the activity of *Lactobacillus*. Moreover, the expression of 258 genes was negatively correlated with the activity of *Lactobacillus* ($r < -0.7$, $p < 0.05$).

Seven enriched KEGG pathways were correlated with the activity of *Lactobacillus* in the ER-negative group ($p < 0.05$), including “renal cell carcinoma” (CDC42, CRKL, PRCC), “viral carcinogenesis” (LTBR, CCND1, HIST2H2BE, CDC42, HIST2H2BF) and 94 GO terms were correlated with the activity of *Lactobacillus* in this group ($p < 0.05$) as presented in Supplementary file 6 A-B [22].

Correlation analysis between gene expression and the activity of *Bacteroides* in the ER-negative tumors

A significant correlation between gene expression and the relative activity of *Bacteroides* was not found among 55 ER-positive samples. However, significantly lower expression of several genes, including PLA2G2E, HIST2H2BE, SHC1, CRKL, CCND1, PRCC, CDC42 and LTBR, was correlated with the relative activity of *Bacteroides* in 15 ER-negative samples (Table 3). The expression of CASP2 positively correlated with the relative activity of *Bacteroides* and *Alloprevotella* (correlation coefficient score = 0.79, $p < 0.01$).

Discussion

Microbiota characteristics of ER-negative tumor tissue and the possible reasons

Whether the microorganisms in FFPE tissues may be arise from environmental contamination rather than originate from tissue was clear after comparison with the environmental control samples. The analysis of 16 S RNA sequencing of the solvent used in this study showed completely different microbiota content than that in FFPE samples. The analysis of ITS RNA sequencing of the environmental control failed in amplification indicating no fungi was detected. In addition, paraffin can preserve microorganisms, but microorganisms will not proliferate in paraffin. Therefore, the results regarding bacteria and fungi obtained from our artificial results are convincing.

The characteristics of the microbiota of breast tumor tissue have been documented in European, Canadian and American studies. The principal phylum in healthy breast tissue and tissue adjacent to breast tumors is *Proteobacteria* which exceeds 50% abundance, followed by *Firmicutes* (35%), *Actinobacteria* (10%), and *Bacteroidetes* (5%), as revealed by Canadian and Irish studies [11]. The breast tumor tissues have higher relative abundance of *Proteobacteria* than healthy tissues shown by Italian study [23]. In non-Hispanic White women and non-Hispanic Black women, *Proteobacteria* is most abundant in normal, normal adjacent to tumor, and breast tumor tissue, with fewer *Firmicutes*, *Bacteroidetes* and *Actinobacteria* [24]. Similar results were shown for USA women with more than 50% relative abundance of *Proteobacteria* followed by *Firmicutes*, *Actinobacteria* and *Bacteroidetes* in healthy, high risk, tumor adjacent and tumor tissues [25].

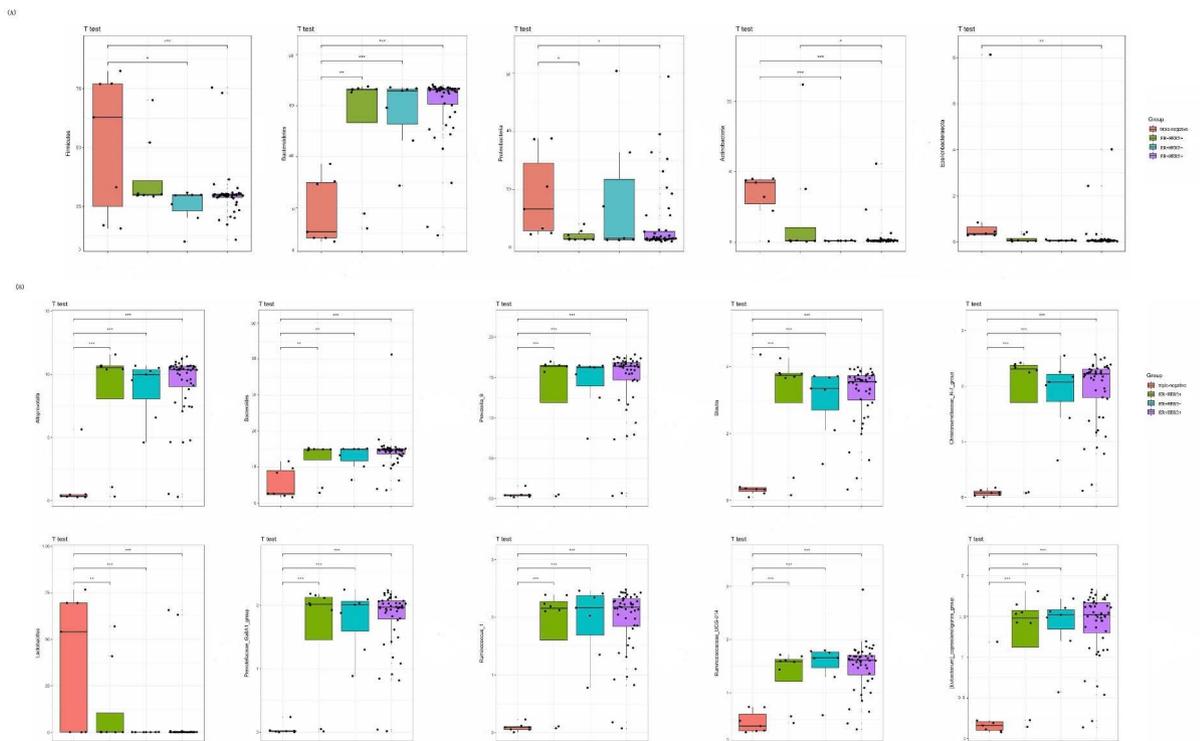


Fig. 2 Comparison of the microbial relative activity (%) among the four groups (A) Box plots of the top 5 phyla; (B) box plots of the top 10 genera; The black line inside the box represents the median. The vertical line represents the lowest datum still within 1.5 interquartile range (IQR) of the lower quartile and the highest datum still within 1.5 IQR of the upper quartile. Student's t test was used for comparisons. * p value < 0.05, ** p value < 0.01, *** p value < 0.001

These studies indicate similar results for Canadian, Irish, Italian, American, non-Hispanic White women and non-Hispanic Black women in microbiota phyla distribution, but there are significant differences in microbiota genera distribution as described in introduction. There are might be significant differences of breast tissue microbiota among ethnic groups. On the other hand, distinct microbial signatures associate with different breast cancer types, including higher abundance of genus *Lactobacillus* in ER+/HER2- ER+/HER2+ and ER-/HER2+ breast tumor than healthy breast samples, and in triple negative tumors several genera significantly correlate with severe (dead) clinical outcomes [26]. In this study, the microbial characteristics of 70 breast tumors of Asian women were analyzed. Although the top phyla were *Bacteroidetes*, *Firmicutes*, *Actinobacteria* and *Proteobacteria* as same as previous studies, the distribution was unique in this study especially for ER negative tumors in which there were higher activity of genus *Lactobacillus* (phylum *Firmicutes*) and positively correlated with several carcinogenesis genes expression. There are might be relationships between breast tissue microbes and cancer development. Because of the limited FFPE samples, in this study, total RNA was reverse transcribed into cDNA and used for both of PCR amplification of the bacterial

16 S rRNA and human gene microarray hybridization. Higher *Bacteroidetes* and *Firmicutes* activity might indicate the higher transcriptional activity but not always represent higher abundance in tumor tissues.

Lactobacillus was not found in the milk of healthy human [27], indicating that infection by this genus was more likely to occur from the outside environment. The hypothesis for the microbiota characteristics of triple-negative breast tumor tissue is that during lactation, *Lactobacillus* invades the mother's nipple and breast from baby's mouth because infants may inhale the bacteria in the birth canal [28]. In nipple aspirate fluid collected from patients with breast cancer, there is a relatively higher activity of the genus *Alistipes* (*Bacteroidetes*) than in the fluid from healthy control women [29]. This finding is consistent with the results in this study.

The tumor microbiota correlated with tumorigenesis gene expression in ER-negative breast cancer

It has been reported that the bacterial DNA load in breast tumors correlates inversely with advanced cancer [30], but few studies have examined the relationship between the tumor microbiota and cellular gene expression. To determine whether the change in the microbiota in breast cancer is related to the occurrence

Table 3 Gene expression positively correlated to *Lactobacillus* abundance and negatively correlated to *Bacteroides* abundance in ER-negative tumor tissues ($r > 0.7$ and $p < 0.05$)

Gene Symbol	Taxonomy 1	Correlation Coefficient	P value	Taxonomy 2	Correlation Coefficient	P value	Description
IL5RA				<i>Lactobacillus</i>	0.92	0.000001	interleukin 5 receptor, alpha
VN1R66P				<i>Lactobacillus</i>	0.92	0.000001	vomer nasal 1 receptor 66 pseudogene
KRTAP10-2				<i>Lactobacillus</i>	0.91	0.000002	keratin associated protein 10–2
CCDC34	<i>Bacteroides</i>	-0.92	0.000001	<i>Lactobacillus</i>	0.89	0.000009	coiled-coil domain containing 34
C1orf68				<i>Lactobacillus</i>	0.89	0.000012	chromosome 1 open reading frame 68
BRD1	<i>Bacteroides</i>	-0.85	0.000073	<i>Lactobacillus</i>	0.85	0.00005	bromodomain containing 1
TRGVB	<i>Bacteroides</i>	-0.89	0.000009	<i>Lactobacillus</i>	0.85	0.000057	T-cell receptor gamma variable B (pseudogene)
OGFOD2				<i>Lactobacillus</i>	0.84	0.000076	2-oxoglutarate and iron-dependent oxygenase domain containing 2
PRCC	<i>Bacteroides</i>	-0.84	0.000096	<i>Lactobacillus</i>	0.84	0.000082	papillary renal cell carcinoma (translocation-associated)
TBL2				<i>Lactobacillus</i>	0.84	0.000086	actin, beta-like 2
ANAPC4				<i>Lactobacillus</i>	0.84	0.000104	anaphase promoting complex subunit 4
SIRPD				<i>Lactobacillus</i>	0.83	0.000121	signal-regulatory protein delta
PLA2G2E	<i>Bacteroides</i>	-0.77	0.000821	<i>Lactobacillus</i>	0.83	0.000135	phospholipase A2, group IIE
MALL				<i>Lactobacillus</i>	0.83	0.000143	mal, T-cell differentiation protein-like
ERC1				<i>Lactobacillus</i>	0.83	0.000147	ELKS/RAB6-interacting/CAST family member 1
PYCR1				<i>Lactobacillus</i>	0.82	0.000163	pyrroline-5-carboxylate reductase 1
CGB1				<i>Lactobacillus</i>	0.82	0.000205	chorionic gonadotropin, beta polypeptide 1
KRTAP5-8				<i>Lactobacillus</i>	0.81	0.000228	keratin associated protein 5–8
HIST2H2BE	<i>Bacteroides</i>	-0.94	0.0000002	<i>Lactobacillus</i>	0.80	0.000357	histone cluster 2, H2be
CRKL	<i>Bacteroides</i>	-0.85	0.000050	<i>Lactobacillus</i>	0.76	0.000989	v-crk avian sarcoma virus CT10 oncogene homolog-like
CCND1	<i>Bacteroides</i>	-0.84	0.000080	<i>Lactobacillus</i>	0.76	0.000935	cyclin D1
CDC42	<i>Bacteroides</i>	-0.72	0.002428	<i>Lactobacillus</i>	0.72	0.002575	cell division control protein 42 homolog
LTBR	<i>Bacteroides</i>	-0.71	0.002741	<i>Lactobacillus</i>	0.74	0.001720	lymphotoxin beta receptor (TNFR superfamily, member 3)
HIST2H2BF	<i>Bacteroides</i>	-0.71	0.002930	<i>Lactobacillus</i>	0.70	0.003442	histone cluster 2, H2bf
SHC1	<i>Bacteroides</i>	-0.87	0.00026	<i>Lactobacillus</i>	0.71	0.003305	SHC (Src homology 2 domain containing) transforming protein 1

and development of cancer, we analyzed the correlation between the microbiota and gene expression using the same total RNA extracted from FFPE samples. There was no correlation between the bacterium and gene expression in ER-positive (ER+/HER2- and ER+/HER2+ types) tumor tissue. The expression of 221 genes was positively associated with the higher activity of *Lactobacillus* in ER-negative (ER-/HER2+ and triple-negative types) samples ($r > 0.7$, $p < 0.05$) including breast cancer biomarker genes PLA2G2E, HIST2H2BE, CRKL, and CCND1 (Table 3) and the expression of same genes was negatively correlated with the relative activity of *Bacteroides* (Table 3). Fungal activity was not related to any gene expression in the same ER-negative samples.

PLA2G2E (group IIE sPLA2) expression was significantly associated with the activity of *Lactobacillus* in the ER-negative group, and its expression was 1.21-fold higher in the ER-negative group than in the ER-positive

group ($p < 0.001$). PLA2s comprise a superfamily that is generally divided into six subfamilies: cytosolic PLA2s (cPLA2s), calcium-independent PLA2s (iPLA2s), secreted PLA2s (sPLA2s), lysosomal PLA2s, platelet-activating factor acetylhydrolases, and adipose-specific PLA2s [31]. PLA2 is an esterase that cleaves glycerophospholipids to release fatty acids and lysophospholipids and may be associated with tumorigenesis in human tissues [32]. An analysis of receiver operating characteristic curves revealed that plasma PLA2 (sPLA2s) activities were higher in patients with breast cancer than in healthy controls. Plasma PLA2 activity may serve as a biomarker for patients with breast cancer [33].

CRKL (Crk-like) is an adapter protein that has crucial roles in multiple biological processes, including cell proliferation, adhesion, and migration. CRKL induces cyclin D1 and phosphorylated extracellular signal-regulated kinase expression, overexpression of CRKL correlates

with progression and malignant proliferation of human breast cancers [34]. CCND1 (cyclin D1) has been solidly established as an oncogene with an important pathogenetic role in breast cancer [35]. CCND1 overexpression is found in more than 50% of human breast cancers and causes mammary cancer in transgenic mice. Dysregulation of CCND1 gene expression or function contributes to the loss of normal cell cycle control during tumorigenesis [36]. HIST2H2BE (histone cluster 2) promotes the progression of invasive ductal carcinoma [37].

Specific microbes such as *Bacteroides*, *Streptococcus*, *Bacteroides massiliensis*, *Faecalibacterium prausnitzii*, *Eubacterium rectale*, and *Mycoplasma genitalium* have been associated with differing risks of prostate cancer development or the extensiveness of prostate cancer disease [38]. Moreover, the present study found an association of the microbiota in ER-negative breast tumor with breast cancer biomarker gene expression which may be associated with tumorigenesis.

The tumor microbiota disorder (higher *Lactobacillus* activity) in triple-negative breast cancer, possibly caused by external factors, associated with the carcinogenesis gene expression. Although the mechanism by which the tumor microbiota relating to the gene expression remains unknown, we speculate that the breast tumor microenvironment might be involved in some way. However, these results have limitations for the smaller sample size especially of the triple-negative group including only 7 samples, and need to be confirmed in a larger cohort. Whether the disorder of the microbiota in breast tissue is related to the lactation history of spontaneously delivered infants and mothers' vaginitis also remains to be confirmed.

This was the first study to identify a significant association of the breast tumor tissue microbiota with the expression of carcinogenesis genes in ER-negative breast cancer. This finding indicated that changes in the microbiota of breast tissue induced by external factors might be one of the key cause of breast carcinogenesis. We believe that the results in this study provide new targets for breast cancer treatment.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12885-023-10726-4>.

Supplementary Material 1

Acknowledgements

We thank Drs. Ming Xu and Jianyu Pei of the Central Hospital of Huangpu District, Shanghai, and Dr. Xiaolin Wang of Shanghai Meiji Health Science and Technology Co., Ltd for their assistance with FFPE sample preparation.

Author Contribution

ZSJ designed the study and helped prepare the manuscript. FZ and TX performed the experiments. LY and SSM performed the data analysis and

prepared Figs. 1 and 2. EL wrote the manuscript. All authors reviewed the manuscript.

Funding

There is no relevant funding.

Data Availability

The microarray data were deposited in the Gene Expression Omnibus (GEO) database under accession code GSE183231 and GSE185439, which were obtained from the same cohort experiment. The 16 S rRNA and ITS sequencing data have been uploaded to GenBank Sequence Read Archive under accession number PRJNA769523. The codes used during the study are available from the corresponding author by request, including program codes for Figs. 1 and 2.

Declarations

Ethics approval and consent to participate

The Medical Ethics Committee, Shanghai Institute of Planned Parenthood Research approved the protocols for RNA extraction, microarray hybridization, 16s rRNA gene amplicon sequencing, ITS sequencing, and data analysis (approval number PJ2019-18). The study was performed in accordance with the guidelines of the Declaration of Helsinki. All patients or their legal guardians provided informed consent to participate in this study.

Consent for publication

The study protocol was explained to each of the 70 patients or their legal guardians. All of the subjects or their legal guardians gave their informed consent to publish the information in an online open-access publication.

Competing interests

The authors declare no competing interest.

Received: 6 July 2022 / Accepted: 9 March 2023

Published online: 16 March 2023

References

- Zengul AG, Demark-Wahnefried W, Barnes S, Morrow CD, Bertrand B, Berryhill TF, Fruge AD. Associations between Dietary Fiber, the fecal microbiota and estrogen metabolism in Postmenopausal women with breast Cancer. *Nutr Cancer*. 2021;73(7):1108–17.
- Early Breast Cancer Trialists', Collaborative G. Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials. *Lancet*. 2005;365(9472):1687–717.
- Davies C, Pan H, Godwin J, Gray R, Arriagada R, Raina V, Abraham M, Medeiros Alencar VH, Badran A, Bonfill X, et al. Long-term effects of continuing adjuvant tamoxifen to 10 years versus stopping at 5 years after diagnosis of oestrogen receptor-positive breast cancer: ATLAS, a randomised trial. *Lancet*. 2013;381(9869):805–16.
- Early Breast Cancer Trialists', Collaborative G, Davies C, Godwin J, Gray R, Clarke M, Cutter D, Darby S, McGale P, Pan HC, Taylor C, et al. Relevance of breast cancer hormone receptors and other factors to the efficacy of adjuvant tamoxifen: patient-level meta-analysis of randomised trials. *Lancet*. 2011;378(9793):771–84.
- Lehmann BD, Bauer JA, Chen X, Sanders ME, Chakravarthy AB, Shyr Y, Pietenpol JA. Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. *J Clin Invest*. 2011;121(7):2750–67.
- Teschendorff AE, Miremadi A, Pinder SE, Ellis IO, Caldas C. An immune response gene expression module identifies a good prognosis subtype in estrogen receptor negative breast cancer. *Genome Biol*. 2007;8(8):R157.
- Law AMK, Lim E, Ormandy CJ, Gallego-Ortega D. The innate and adaptive infiltrating immune systems as targets for breast cancer immunotherapy. *Endocr Relat Cancer*. 2017;24(7):X1.
- Poore GD, Kopylova E, Zhu Q, Carpenter C, Fraccacio S, Wandro S, Kosciolk T, Janssen S, Metcalf J, Song SJ, et al. Microbiome analyses of blood and tissues suggest cancer diagnostic approach. *Nature*. 2020;579(7800):567–74.

9. Nejman D, Liviyatan I, Fuks G, Gavert N, Zwang Y, Geller LT, Rotter-Maskowitz A, Weiser R, Mallel G, Gigi E, et al. The human tumor microbiome is composed of tumor type-specific intracellular bacteria. *Science*. 2020;368(6494):973–80.
10. Costantini L, Magno S, Albanese D, Donati C, Molinari R, Filippone A, Masetti R, Merendino N. Characterization of human breast tissue microbiota from core needle biopsies through the analysis of multi hypervariable 16S-rRNA gene regions. *Sci Rep*. 2018;8(1):16893.
11. Urbaniak C, Cummins J, Brackstone M, Macklaim JM, Gloor GB, Baban CK, Scott L, O'Hanlon DM, Burton JP, Francis KP, et al. Microbiota of human breast tissue. *Appl Environ Microbiol*. 2014;80(10):3007–14.
12. Urbaniak C, Gloor GB, Brackstone M, Scott L, Tangney M, Reid G. The microbiota of breast tissue and its association with breast Cancer. *Appl Environ Microbiol*. 2016;82(16):5039–48.
13. Wang H, Altemus J, Niazi F, Green H, Calhoun BC, Sturgis C, Grobmyer SR, Eng C. Breast tissue, oral and urinary microbiomes in breast cancer. *Oncotarget*. 2017;8(50):88122–38.
14. Hieken TJ, Chen J, Hoskin TL, Walther-Antonio M, Johnson S, Ramaker S, Xiao J, Radisky DC, Knutson KL, Kalari KR, et al. The Microbiome of aseptically collected human breast tissue in Benign and Malignant Disease. *Sci Rep*. 2016;6:30751.
15. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30(15):2114–20.
16. Rognes T, Flouri T, Nichols B, Quince C, Mahe F. VSEARCH: a versatile open source tool for metagenomics. *PeerJ*. 2016;4:e2584.
17. Wang Q, Garrity GM, Tiedje JM, Cole JR. Naive bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol*. 2007;73(16):5261–7.
18. Chao A, Bunge J. Estimating the number of species in a stochastic abundance model. *Biometrics*. 2002;58(3):531–9.
19. Hill TC, Walsh KA, Harris JA, Moffett BF. Using ecological diversity measures with bacterial communities. *FEMS Microbiol Ecol*. 2003;43(1):1–11.
20. Mount DW. Using the Basic Local Alignment Search Tool (BLAST). *CSH Protoc* 2007, 2007:pdb top17.
21. Zhang D, Jiang P, Xu Q, Zhang X. Arginine and glutamate-rich 1 (ARGLU1) interacts with mediator subunit 1 (MED1) and is required for estrogen receptor-mediated gene transcription and breast cancer cell growth. *J Biol Chem*. 2011;286(20):17746–54.
22. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. 2000;28(1):27–30.
23. Esposito MV, Fosso B, Nunziato M, Casaburi G, D'Argenio V, Calabrese A, D'Aiuto M, Botti G, Pesole G, Salvatore F. Microbiome composition indicate dysbiosis and lower richness in tumor breast tissues compared to healthy adjacent paired tissue, within the same women. *BMC Cancer*. 2022;22(1):30.
24. Smith A, Pierre JF, Makowski L, Tolley E, Lyn-Cook B, Lu L, Vidal G, Starlard-Davenport A. Distinct microbial communities that differ by race, stage, or breast-tumor subtype in breast tissues of non-hispanic black and non-hispanic white women. *Sci Rep*. 2019;9(1):11940.
25. Tzeng A, Sangwan N, Jia M, Liu CC, Keslar KS, Downs-Kelly E, Fairchild RL, Al-Hilli Z, Grobmyer SR, Eng C. Human breast microbiome correlates with prognostic features and immunological signatures in breast cancer. *Genome Med*. 2021;13(1):60.
26. Banerjee S, Tian T, Wei Z, Shih N, Feldman MD, Peck KN, DeMichele AM, Alwine JC, Robertson ES. Distinct Microbial Signatures Associated with different breast Cancer types. *Front Microbiol*. 2018;9:951.
27. Hunt KM, Foster JA, Forney LJ, Schutte UM, Beck DL, Abdo Z, Fox LK, Williams JE, McGuire MK, McGuire MA. Characterization of the diversity and temporal stability of bacterial communities in human milk. *PLoS ONE*. 2011;6(6):e21313.
28. Arakawa A, Ichikawa H, Kubo T, Motoi N, Kumamoto T, Nakajima M, Yonemori K, Noguchi E, Sunami K, Shiraishi K, et al. Vaginal transmission of Cancer from Mothers with Cervical Cancer to Infants. *N Engl J Med*. 2021;384(1):42–50.
29. Chan AA, Bashir M, Rivas MN, Duvall K, Sieling PA, Pieber TR, Vaishampayan PA, Love SM, Lee DJ. Characterization of the microbiome of nipple aspirate fluid of breast cancer survivors. *Sci Rep*. 2016;6:28061.
30. Xuan C, Shamonki JM, Chung A, Dinome ML, Chung M, Sieling PA, Lee DJ. Microbial dysbiosis is associated with human breast cancer. *PLoS ONE*. 2014;9(1):e83744.
31. Peng Z, Chang Y, Fan J, Ji W, Su C. Phospholipase A2 superfamily in cancer. *Cancer Lett*. 2021;497:165–77.
32. Cummings BS. Phospholipase A2 as targets for anti-cancer drugs. *Biochem Pharmacol*. 2007;74(7):949–59.
33. Qu J, Zhao X, Wang J, Liu C, Sun Y, Cai H, Liu J. Plasma phospholipase A2 activity may serve as a novel diagnostic biomarker for the diagnosis of breast cancer. *Oncol Lett*. 2018;15(4):5236–42.
34. Zhao T, Miao Z, Wang Z, Xu Y, Wu J, Liu X, You Y, Li J. Overexpression of CRKL correlates with malignant cell proliferation in breast cancer. *Tumour Biol*. 2013;34(5):2891–7.
35. Arnold A, Papanikolaou A. Cyclin D1 in breast cancer pathogenesis. *J Clin Oncol*. 2005;23(18):4215–24.
36. Velasco-Velazquez MA, Li Z, Casimiro M, Loro E, Homsy N, Pestell RG. Examining the role of cyclin D1 in breast cancer. *Future Oncol*. 2011;7(6):753–65.
37. He B, Chen J, Song W, Bai Y. miR-646/TET1 mediated demethylation of IRX1 promoter upregulates HIST2H2BE and promotes the progression of invasive ductal carcinoma. *Genomics*. 2021;113(3):1469–81.
38. Sha S, Ni L, Stefil M, Dixon M, Mouraviev V. The human gastrointestinal microbiota and prostate cancer development and treatment. *Investig Clin Urol*. 2020;61(Suppl 1):43–550.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.