



OPEN ACCESS

EDITED BY

Shi Huang,
The University of Hong Kong,
Hong Kong SAR, China

REVIEWED BY

Wei-Hua Chen,
Huazhong University of Science and
Technology, China
Gongchao Jing,
Qingdao Institute of Bioenergy and
Bioprocess Technology (CAS), China

*CORRESPONDENCE

Junnan Xu
15040292193@126.com

Tao Sun
jianong@126.com

†These authors have contributed
equally to this work

SPECIALTY SECTION

This article was submitted to
Extra-intestinal Microbiome,
a section of the journal
Frontiers in Cellular and
Infection Microbiology

RECEIVED 30 August 2022

ACCEPTED 16 November 2022

PUBLISHED 13 December 2022

CITATION

Wang N, Yang J, Han W, Han M, Liu X,
Jiang L, Cao H, Jing M, Sun T and Xu J
(2022) Identifying distinctive tissue and
fecal microbial signatures and the
tumor-promoting effects of
deoxycholic acid on breast cancer.
Front. Cell. Infect. Microbiol.
12:1029905.
doi: 10.3389/fcimb.2022.1029905

COPYRIGHT

© 2022 Wang, Yang, Han, Han, Liu,
Jiang, Cao, Jing, Sun and Xu. This is an
open-access article distributed under
the terms of the [Creative Commons
Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use,
distribution or reproduction in other
forums is permitted, provided the
original author(s) and the copyright
owner(s) are credited and that the
original publication in this journal is
cited, in accordance with accepted
academic practice. No use,
distribution or reproduction is
permitted which does not comply with
these terms.

Identifying distinctive tissue and fecal microbial signatures and the tumor-promoting effects of deoxycholic acid on breast cancer

Na Wang^{1,2}, Jun Yang¹, Wenjie Han^{1,2}, Mengzhen Han^{1,2},
Xiaolin Liu³, Lei Jiang^{1,4}, Hui Cao^{1,4}, Mingxi Jing^{1,4}, Tao Sun^{1,4*†}
and Junnan Xu^{1,2,4*†}

¹Department of Breast Medicine, Cancer Hospital of China Medical University, Liaoning Cancer Hospital, Shenyang, China, ²Department of Pharmacology, Cancer Hospital of China Medical University, Liaoning Cancer Hospital, Shenyang, China, ³Department of Medicine, Liaoning Kanghui Biotechnology Co., Ltd., Shenyang, China, ⁴Department of Breast Medicine, Cancer Hospital of Dalian University of Technology, Liaoning Cancer Hospital, Shenyang, China

Introduction: A growing body of evidence indicates that the dysbiosis of both mammary and intestinal microbiota is associated with the initiation and progression of breast tumors. However, the microbial characteristics of patients with breast tumors vary widely across studies, and replicable biomarkers for early-stage breast tumor diagnosis remain elusive.

Methods: We demonstrate a machine learning-based method for the analysis of breast tissue and gut microbial differences among patients with benign breast disease, patients with breast cancer (BC), and healthy individuals using 16S rRNA sequence data retrieved from eight studies. QIIME 2.0 and R software (version 3.6.1) were used for consistent processing. A naive Bayes classifier was trained on the RDP v16 reference database to assign taxonomy using the Vsearch software.

Results: After re-analyzing with a total of 768 breast tissue samples and 1,311 fecal samples, we confirmed that *Halomonas* and *Shewanella* were the most representative genera of BC tissue. *Bacteroides* are frequently and significantly enriched in the intestines of patients with breast tumor. The areas under the curve (AUCs) of random forest models were 74.27% and 68.08% for breast carcinoma tissues and stool samples, respectively. The model was validated for effectiveness *via* cohort-to-cohort transfer (average AUC = 0.65) and leave-one-cohort-out (average AUC = 0.66). The same BC-associated biomarker *Clostridium_XIVa* exists in the tissues and the gut. The results of the *in-vitro* experiments showed that the *Clostridium*-specific-related metabolite deoxycholic acid (DCA) promotes the proliferation of HER2-positive BC cells and stimulates G0/G1 phase cells to enter the S phase, which may be related to the activation of peptide-O-fucosyltransferase activity functions and the neuroactive ligand-receptor interaction pathway.

Discussion: The results of this study will improve our understanding of the microbial profile of breast tumors. Changes in the microbial population may be present in both the tissues and the gut of patients with BC, and specific markers could aid in the early diagnosis of BC. The findings from *in-vitro* experiments confirmed that *Clostridium*-specific metabolite DCA promotes the proliferation of BC cells. We propose the use of stool-based biomarkers in clinical application as a non-invasive and convenient diagnostic method.

KEYWORDS

breast cancer, 16S rRNA, microbiome, intestine, cancer diagnosis, random forest, *Clostridium*, deoxycholic acid

Introduction

Breast cancer (BC) is a common type of cancer and accounts for most cancer-related deaths in women worldwide. According to the latest cancer statistics report, an estimated 2,261,419 new cases (11.7% of total cancer cases) and 684,996 cancer deaths (6.9% of total cancer deaths) were reported in women worldwide in 2020 (Sung et al., 2021). In recent years, the BC incidence rates have continued to increase at an annual rate of 0.5% per year (Siegel et al., 2021). The findings of a hallmark population-based study on cancer indicated that early-stage detection can help reduce breast cancer mortality by one-third (Elmore et al., 2005; Jafari et al., 2018). As the availability of high-throughput sequencing technology has continued to expand and its cost has decreased, the identification and characterization of the tissue or gut microbiome in patients with BC have constituted an active area of research. In addition to the various imaging techniques and biochemical parameters, tumor-related microbes may be used as diagnostic biomarkers for breast diseases, and investigating the microbiome could be a novel cancer diagnostic method.

Our perception of sterility in breast tissue has undergone significant changes over time. The breast microbiome may originate from the intestine and is highly malleable (Shively et al., 2018; Chadha et al., 2021). Recent advances in metagenomics and next-generation sequencing technologies have promoted a more comprehensive understanding of the complex relationships between human microbiota and breast disease. We confirmed that, across different studies, mammary gland tissues were found to harbor disparate microbial communities, some even exhibiting contrasting microbial abundances. Hieken et al. confirmed that *Fusobacterium*, *Gluconacetobacter*, *Lactobacillus*, and *Atopobium* were significantly enriched in breast cancerous tissues compared with those in healthy individuals (Hieken et al., 2016). Malignant tumor tissues also showed an increase in the relative abundance of *Ralstonia*, and a relative decrease in the

abundance of *Methylobacterium* was also observed compared with that in healthy controls in another study (Costantini et al., 2018). However, Xuan et al. showed *Methylobacterium* to have the greatest prevalence in malignant tumor tissues (Xuan et al., 2014). The gut microbiome, as one of the major sources of the breast microbiome, contributes to physiological immune modulation and serves as a critical component in the occurrence and progression of mammary carcinoma (Zhang et al., 2020; Zhang et al., 2021). Similarly, previous studies have not reported the same findings on intestinal microorganisms in patients with breast diseases. Studies focused on population-based community metrics have shown that the populations of *Coprococcus*, *Butyrivimonas*, and *Odoribacter* in fecal samples were most inversely associated with BC (Bobin-Dubigeon et al., 2021). In other studies, there is a high abundance of *Bacteroides*, *Clostridiaceae*, *Ruminococcaceae*, and *Faecalibacterium* in stool samples collected from patients with BC, along with a low abundance of *Romboutsia*, *Coprococcus*, *Dorea*, and *Lachnospiraceae* (Goedert et al., 2015; Byrd et al., 2021). Substantial variability in the abundances of microbial communities has been reported in different studies, possibly owing to various biological factors, such as individual genetic variance, ethnicity, and dietary habits that influence microbiota composition. Furthermore, the presence of hypervariable regions, the sequencing platform and the bioinformatics pipelines used, and the method of data processing differ from study to study, thus making tissue and fecal microbiome more challenging to analyze uniformly and systematically.

Meta-analysis serves as a useful research tool owing to its informative and unbiased nature, which helps reduce the effect of biological and technological differences across multiple studies and provide stabilized and accurate results (Gurevitch et al., 2018). Several meta-analyses have been performed for the identification of microbial markers in colorectal cancer based on 16S rRNA or shotgun genome sequencing datasets (Sze and Schloss, 2018; Mo et al., 2020; Wu et al., 2021). An urgent need

exists to explore and precisely identify stool- and tissue-based microbial biomarkers. While findings from certain studies have confirmed that the microbial communities differ significantly between patients with breast diseases and healthy controls, the co-differences of microbes between these two body sites remain unclear.

In this study, we performed an integrated analysis using public 16S rRNA sequence data from both breast tissues ($n = 768$) and stool samples ($n = 1,311$) from healthy women and women with breast diseases across eight studies to improve our understanding of specific microbial community in breast diseases and specific microbe-related metabolites in breast carcinogenesis. *Clostridium_XIVa* had simultaneously high expression both in breast tissue and gut environment. DCA was the main and specific metabolic biotransformed by *Clostridium* with primary bile acids and reabsorbed in the colon and recycled to the liver, then conjugating to glycine or taurine, similar to primary bile acids inducing biliary lipid secretion, and solubilizing cholesterol in the bile, contributing approximately 20% to the bile acid pool (Chiang, 2013). Yoshimoto et al. revealed that the levels of DCA and *Clostridium* genus increased simultaneously in high-fat diet mice by liquid chromatography–mass spectrometry (LC-MS) and 16S rRNA analysis (Yoshimoto et al., 2013). Considering the specific 7 α -dehydroxylating activity carried by the *Clostridium* genus, the explanation for it is that the *Clostridium* genus contributes to an increase in the DCA level. DCA could promote M1 macrophage polarization and pro-inflammatory cytokine production (Wang et al., 2020) and be involved in gastrointestinal cancer dose-dependently (Nagathihalli et al., 2014). DCA exposure could inhibit farnesoid X receptor expression and enhance the Wnt– β -catenin signaling pathway, thereby contributing to colon carcinogenesis (Yao et al., 2022). In addition, the enterohepatic circulation of DCA provokes senescence-associated secretory phenotype in hepatic stellate cells, facilitating hepatocellular carcinoma development (Yoshimoto et al., 2013). These notions prompted us to examine if *Clostridium*-specific-related DCA has key roles in BC development. Therefore, DCA was used to demonstrate the effect of specific microbe-related metabolites on BC cell growth and apoptosis. Based on these results, we combined the method with proteomic analysis to investigate the mechanisms by which microbial metabolites contribute to BC development. In addition, after pooling data from different cohorts, we constructed a random forest (RF) model to distinguish patients with BC from non-cancer controls. Cohort-to-cohort transfer validation and leave-one-cohort-out (LOCO) validation across multiple datasets were employed to overcome the influence of technical discrepancies and cohort heterogeneity. Collectively, our study suggests that distinct microbiota populations are present in breast tissues or gut of patients with BC and non-cancer controls, the RF model constructed based on specific members of the microbial community can more

accurately classify patients with BC, and *Clostridium*-specific DCA could stimulate BC cell growth.

Materials and methods

Study selection

PubMed was searched for studies (published on or before up to 01 January 2022) that used breast tissue or stool microbiota obtained from patients with breast disease using the terms “(Breast Neoplasms or breast cancer) AND (microbiome or microbiota) AND (metagenomic or mNGS or 16S rRNA).” The search was limited to studies that fit the following criteria: 1) had original data, 2) used fecal or breast tissue samples, and 3) included breast disease and normal control populations. Systematic reviews, letters to the editors, conference papers, case reports, and meta-analyses data were excluded. Of the nine studies that fulfilled the criteria, in eight studies, the 16S rRNA gene was characterized, and in only one study, shotgun metagenomics was used. The eight studies that characterized the 16S rRNA gene were included in this meta-analysis.

Data retrieval and processing

Raw sequencing data were retrieved from the National Center for Biotechnology Information Sequence Read Archive (NCBI-SRA) database using the following accession numbers (BioProject ID): PRJNA624822 for Nejman et al. (2020), SRP076038 for Urbaniak et al. (2016), PRJNA335375 for Hieken et al. (2016), PRJNA637875 for Thyagarajan et al. (2020), PRJNA383849 for Goedert et al. (2018), PRJNA658160 for Byrd et al. (2021), PRJNA566060 for Yao et al. (2021), and PRJNA630839 for Guan et al. (2020). The downloaded FASTQ data of each study were reprocessed separately with consistent processing to avoid the biases arising from bioinformatics analyses. All 16S rRNA gene datasets containing forward and reverse read files were processed using a standardized pipeline in QIIME 2.0 after an initial visualization of read quality using the FastQC and MultiQC software. The DADA2 (V.2018.11) software, wrapped in QIIME2, was used to filter sequencing reads with quality score $Q > 25$ and denoise reads into amplicon sequence variants, which resulted in the formation of feature tables and representative sequences for each study.

Combined processing

Feature tables and representative sequences in each study were merged using QIIME2's merge and merge-seqs commands. R software (version 3.6.1) was used to conduct further analyses using the data exported from QIIME2. A naive Bayes classifier

was trained on the RDP v16 reference database to assign taxonomy using the Vsearch software. Sequences that were identified as “Chloroplast,” “Mitochondria,” and those that were unclassified at the kingdom level were removed from the datasets. Samples that could not be mapped to any species were omitted. ANOVA-like analysis was used to quantify the effect of potential confounding factors and disease status. The total variance within the abundance of a given ASV was compared to the variance explained by disease status (BC, benign tumor, and normal) and the variance explained by confounding factors including study, age, and BMI. Variance calculations were performed on ranks in order to account for the non-Gaussian distribution of microbiome abundance data. Potential confounders were transformed into categorical data (study) as quartiles (age) and for the case of BMI into lean (>25), obese (25–30), and overweight (>30). Beta diversity in microbial composition was calculated using the Bray–Curtis dissimilarity metrics, and principal coordinate analysis (PCoA) was performed based on the Bray–Curtis dissimilarity matrix using the R “vegan” package. Microbiome data were CLR-transformed using the “compositions” package in R software. The significance of differential abundance was tested on genera using the STAMP software after CLR transformation of microbiome data (Welch’s *t*-test and *P*-values <0.05 were considered significant). BC-related biomarkers were identified by the Wilcoxon rank-sum tests based on CLR-transformed data. Genera with *P* <0.05 and high abundance in the BC group were considered as BC-related biomarkers.

RF model construction and evaluation

Based on the sequence information, we prepared RF models using the “randomForest” package with default parameters to distinguish the cancer status from the non-cancer status. Stratified 10-fold cross-validation was used to configure the training and testing datasets. The minimum error was calculated using five-fold cross-validation with the “rfcv” function. Based on the Mean-Decrease-Accuracy values, the model was constructed using the top 30 most important genera, which are considered as important microbial characteristics. LOCO and cohort-to-cohort validation were performed to evaluate the generalizability of microbe-based BC classifiers, including geographic variation and technical differences in the 16S rRNA microbial data across multiple studies. In LOCO validation, data from one cohort were used as the testing set to assess the training set, which was set by pooled data from the remaining cohorts. In the cohort-to-cohort validation, we constructed the RF classifier models using a single cohort and used other cohorts as the testing data separately to evaluate classifiers. In addition, a nested cross-validation was used on the training study to calculate the within-study accuracy.

Cell line culture and the configuration of metabolite solution

The human breast cancer cell lines MDA-MB-231 and SK-BR-3 were purchased from the ATCC (ATCC, Rockville, MD, USA) and cultured in complete DMEM (12800-058; Thermo Fisher Scientific, Inc., China) supplemented with 10% fetal bovine serum (FBS; C04001-500; Biological Industries, China) and 1% penicillin/streptomycin (P1400, Solarbio, China) at 37°C in 5% CO₂. DCA (D103698; Aladdin, China) was dissolved in DMSO (DH105-2; Dingguo Biology, China), and a stock solution of 250 mM was stored at –20°C.

Cell proliferation assay

Growth assays were performed using a Cell Counting Kit-8 (CCK8) assay kit (CK04; Dojindo Laboratories, Japan). Briefly, MDA-MB-231 and SK-BR-3 cells were seeded in 96-well plates at 10³ cells/well and cultured for 24 h and then treated with DCA for 24 h. After treatment, CCK8 was added to each well (10 µl of CCK8 substrate and 90 µl of complete DMEM), and the cells were incubated at 37°C for 1 h. The optical density (OD) was determined using a microplate reader at 450 nm (Multiskan MK3; Pioneer Co-operative UK Ltd., Britain). After normalization, cell viability was calculated using the formula: Cell viability = (OD value of the experimental well – OD value of the blank well)/(OD value of the control well – OD value of the blank well) × 100%.

Apoptosis analysis

We detected apoptosis by Annexin V FITC/PI staining. SK-BR-3 cells were seeded in six-well plates (105 cells/well), cultured for 24 h, and then treated with DCA (0, 40, 60, and 80 µmol/L) for 24 h. Following this, SK-BR-3 cells were harvested and resuspended in 500 µl of binding buffer. Five microliters of Annexin V/FITC (WLA001c; Wanleibio, China) and 10 µl of PI solution were added and incubated in the dark for 15 min at 25°C. The cells were assessed using a flow cytometer (BD FACSAria II; BD Co., USA), and data were analyzed using the FlowJo software.

Cell cycle analysis

Cell cycle analysis was performed using flow cytometry. Briefly, SK-BR-3 cells were seeded in six-well plates (10⁵ cells/well) and incubated for 24 h in DMEM supplemented with DCA (0, 40, 60, and 80 µmol/L). The cells were then harvested, centrifuged (1,500 rpm, 5 min), and fixed overnight with 70%

precooled ethanol at 4°C and resuspended in ice-cold phosphate-buffered saline (PBS). Subsequently, 100 µl of RNase A solution (WLA010a; Wanleibio, China) was added to the cell suspension, and the cells were incubated for 30 min at 37°C in a water bath. Five hundred microliters of PI was added to the cells, which were incubated in the dark at 4°C for 30 min. A flow cytometer was used to determine the cell cycle of the cells, and data were analyzed using the FlowJo software.

Proteomics analysis

SDT buffer (4% SDS, 100 mM of Tris-HCl, pH 7.6) was added to the sample. The lysate was sonicated and boiled for 10 min. After centrifugation at 14,000g for 15 min, the protein content in the supernatant was quantified using the BCA Protein Assay Kit (P0012, Beyotime, China). The proteins were separated on a 12% SDS-PAGE gel. One hundred microliters of iodoacetamide (IAA) (100 mM of IAA in UA buffer) was added to block the reduced cysteine residues, and the samples were incubated for 30 min in the dark. Finally, the protein suspensions were digested using 4 µg of trypsin (V5117, Promega, China) in 50 mM of NH₄HCO₃ buffer overnight at 37°C. The peptide segment was desalted using a C18 column. The peptide content was estimated based on the UV light spectral density at 280 nm using an extinction coefficient of 1.1 in a 0.1% (g/L) solution.

Analysis of mass spectrometry data

The samples were analyzed on a nanoElute (Bruker, Bremen, Germany) coupled to a timsTOF Pro (Bruker, Bremen, Germany) equipped with a CaptiveSpray source. Peptides were separated on a 25-cm × 75-µm analytical column, with 1.6 µm of C18 beads with a packed emitter tip (IonOpticks, Australia). The column was equilibrated using 4 column volumes before loading the sample in 100% buffer A (99.9% Milli-Q water, 0.1% FA). The samples were separated at 300 nl/min using a linear gradient as follows: 1.5 h gradient: 2%–22% buffer B for 75 min, 22%–37% buffer B for 5 min, 37%–80% buffer B for 5 min, and hold in 80% buffer B for 5 min.

Data analysis

The MaxQuant software was used to analyze the MS data. A maximum of two missed cleavage sites and a mass tolerance of 40 ppm for fragment ions were allowed. The cutoff of the global false discovery rate (FDR) for peptide and protein identification was set to 0.01. Protein abundance was calculated based on the

normalized spectral protein intensity (LFQ intensity). Protein fold change >2 or <0.5 and *P*-value (Student's *t*-test) <0.05 were considered to indicate differentially expressed proteins.

Bioinformatics analysis

All protein sequences were aligned to those present in the NCBI BLAST+. The GO terms of the sequences were selected using the top Bit-Score by Blast2GO. Following this, the annotation from GO terms to proteins was completed. Fisher's exact test was used to enrich the GO terms by comparing the number of differentially expressed proteins and total proteins correlated with the GO terms. The KEGG database was used for pathway analysis. Fisher's exact test was used to identify the significantly enriched pathways.

Statistical analysis

Each experiment was performed at least three times to ensure reproducibility. Results are presented as mean ± SEM. Significance was calculated using one-way ANOVA followed by Bonferroni's multiple comparisons test. GraphPad Prism 8.0 was used to perform all statistical analyses. Statistical significance was considered at **P* <0.05 and ***P* <0.01.

Results

Characteristics of the datasets in the meta-analysis

Eight studies that employed the 16S rRNA sequencing method for microbiome analysis of tissue and fecal samples from patients with benign and malignant breast disease and healthy controls were selected in this meta-analysis to characterize the distinct microbial characteristics. Among the studies, four were conducted in the USA, two in China, one in Israel, one in Canada, and one in Ghana. Notably, we only included data obtained from breast tissue samples from studies conducted by Nejman et al. and Hieken et al., and data from non-breast and non-tissue samples were excluded. In total, we pooled data from eight studies, which included data from 768 tissue samples and 1,311 fecal samples. The sample sizes and clinicopathological data from the datasets for each cohort are presented in [Table 1](#). After quality control using the DADA2 method, data satisfying the follow-up requirements for subsequent analysis were included. A summary of sequences analyzed by DADA2 is listed in [Table 2](#).

TABLE 1 Details of the large-scale breast tissue or gut datasets included in this study.

Study	Type	Group (n)	Age (average \pm SD)	BMI (average \pm SD)	Gravidity Yes (%)	Menopausal Yes (%)	Platform	Region	Country
Nejman (2020)	Tissue	Normal (51)	46.3 \pm 14.5	–	–	–	Illumina_Hiseq, Illumina_Miseq, Illumina_NextSeq	V2, V3, V5, V6, V8	Israel
		Benign (29)	41.7 \pm 11.9	–	–	–			
		Cancer (35)	57.6 \pm 12.6	–	–	–			
		Cancer-adjacent (173)	57.6 \pm 12.1	–	–	–			
Urbaniak (2016)	Tissue	Normal (23)	48.9 \pm 12.1	–	17 (73.9%)	12 (52.2%)	Illumina_MiSeq	V6	Canada
		Benign (13)	55.6 \pm 16.7	–	6 (46.2%)	2 (15.4%)			
		Cancer (32)	54.7 \pm 17.0	–	28 (87.5%)	29 (90.6%)			
Hieken (2016)	Tissue	Benign (12)	60.8 \pm 13.2	29.5 \pm 7.0	11 (91.7%)	6 (50.0%)	Illumina_MiSeq	V3–V5	USA
		Cancer (16)	62.8 \pm 12.1	30.1 \pm 6.4	13 (81.3%)	14 (87.5%)			
Thyagarajan (2020)	Tissue	Cancer (33)	25.0–78.0	–	–	–	Illumina_MiSeq	V3–V4	USA
		Cancer-adjacent (31)	25.0–78.0	–	–	–			
Goedert (2018)	Stool	Normal (144)	62.0 \pm 5.2	28.2 \pm 5.5	–	144 (100.0%)	Illumina_MiSeq	V4	USA
		Cancer (144)	61.0 \pm 3.1	28.3 \pm 5.8	–	144 (100.0%)			
Byrd (2021)	Stool	Normal (442)	46.9 \pm 12.9	28.0 \pm 7.9	391 (88.5%)	189 (42.8%)	Illumina_MiSeq	V4	Ghana
		Benign (111)	38.8 \pm 12.8	27.7 \pm 6.5	77 (69.4%)	28 (25.2%)			
		Cancer (403)	50.8 \pm 12.3	27.1 \pm 5.9	348 (86.4%)	226 (56.1%)			
Yao (2021)	Stool	Cancer (36)	–	–	–	–	Illumina_HiSeq	V3–V4	China
Guan (2020)	Stool	Cancer (31)	–	–	–	7 (22.6%)	Illumina_MiSeq	V4	China

Confounder analysis of the microbiome associated with BC

As studies differed from one another in many biological aspects, we first investigated the effect of potential confounders (including sample type, study, patients' age, BMI). This analysis

revealed that “sample type” has a predominant impact on species composition (Figure 1A), and analyzing the microbiome in different sample types separately is essential. Additionally, we quantified the effect of age, BMI, and study on microbiome composition and contrasted this with disease status (Figures S1, S2). The variance explained by “study” was greater than that by

TABLE 2 The number of reads produced by each step during quality control.

Cohorts	No. of samples	Sequence lengths (bp)	No. of reads				
			Input	Filtered	Denoised	Merged	Non-chimeric
Nejman (2020)	608	150-PE	121,500,721	117,747,260	117,546,602	98,076,145	95,936,476
Urbaniak (2016)	68	100-SE	901,286	899,895	885,093	–	878,656
Hieken (2016)	28	300-SE	4,571,009	3,840,371	3,828,956	–	3,664,393
Thyagarajan (2020)	64	300-PE	1,800,520	1,009,714	977,573	906,868	352,481
Goedert (2018)	288	250-PE	8,580,928	6,294,134	6,204,855	5,616,720	5,527,051
Byrd (2021)	956	150-PE	25,036,506	24,081,617	23,828,549	21,995,066	19,955,146
Yao (2021)	36	250-PE	10,636,634	9,953,346	9,825,646	8,848,662	1,279,577
Guan (2020)	31	150-PE	1,959,200	1,916,553	1,905,425	1,875,691	1,805,792

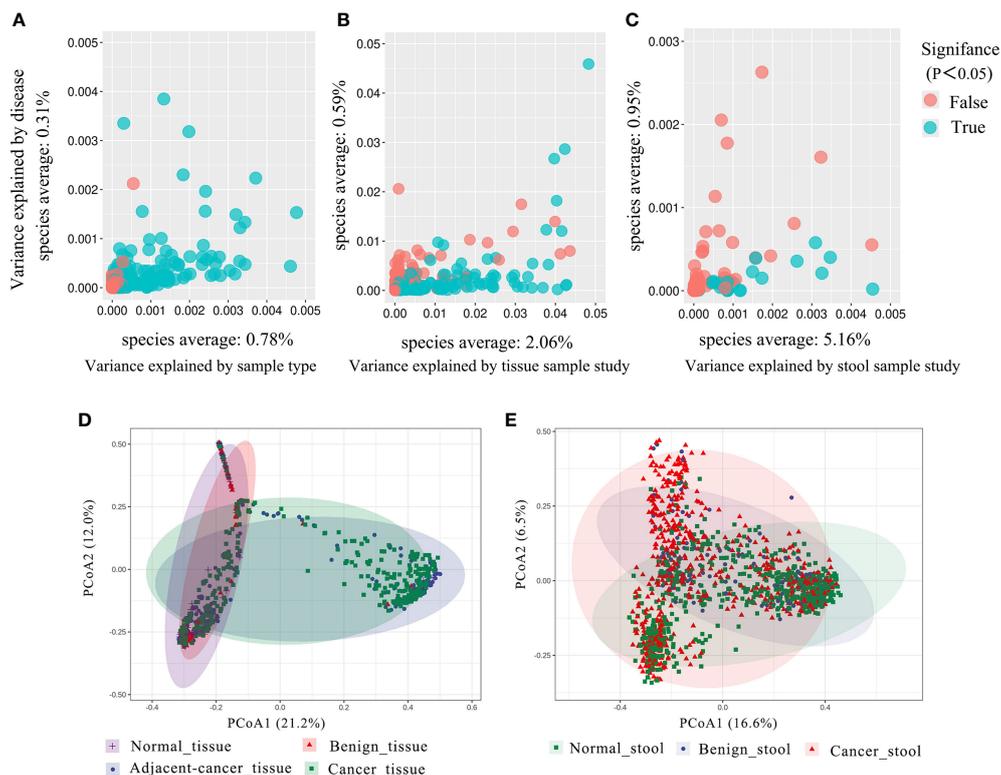


FIGURE 1

The total variance explained by disease status (BC, benign tumor, health) is plotted against total explained by sample type (A), study of tissue (B), and study of gut (C) for individual ASVs. The significantly differential ASVs are colored in blue and P -values were from the two-way ANOVA test. (D, E) Principal coordinate analysis (PCoA) of the breast tissue and gut microbiome based on the Bray–Curtis metric. Each ellipse in different color represents the 95% confidence level. (D) In breast tissue samples, the purple and red ellipses are similar, and the blue and green ellipses are almost overlapping, indicating insignificant differences between normal and benign tissues, adjacent-cancer, and cancer tissue. (E) The ellipses that represented normal, benign, and cancer stool microbiota are inconsistent, demonstrating significant differences among different groups.

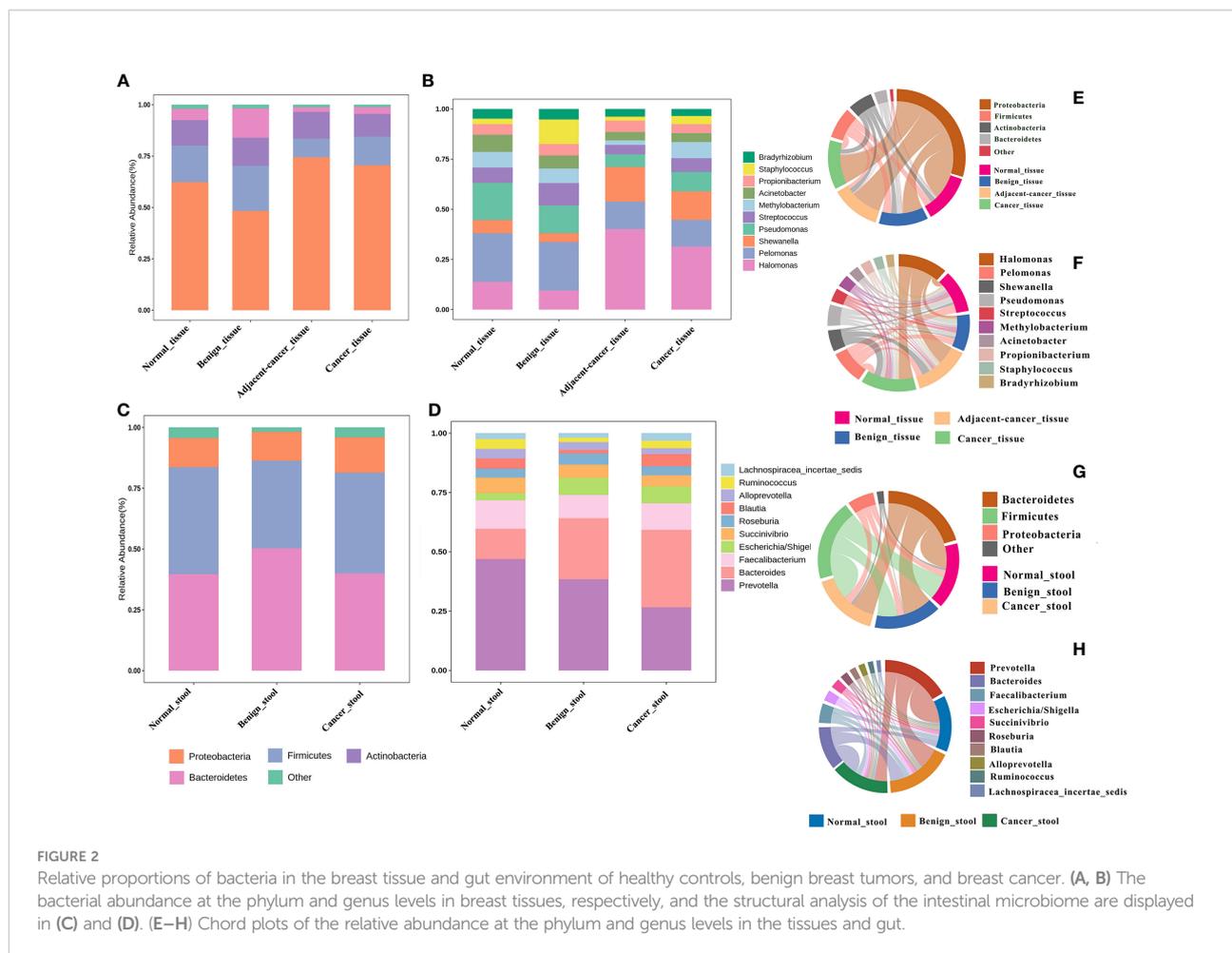
disease status and by other potential confounders. Study heterogeneity has a large effect on overall microbiome composition (Figures 1B, C).

Alterations in microbial composition in BC

Analysis based on Bray–Curtis dissimilarity indicated the extensive variations in the different sample groups. For the tissue microbiota, PCoA showed distinct distributions among normal breast tissue, benign breast tissue, breast carcinoma tissue, and lesion-adjacent cancer tissue (Figure 1D). Furthermore, this finding also indicated a clearer distinction between tissues from cancer (including cancerous and adjacent tissues) and non-cancer (including benign diseases and normal breast tissue) groups. For the stool microbiota, the ecological discrepancy among the gut microbiota from patients with

breast cancer and benign breast tumor and the healthy population was distinguishable (Figure 1E).

The idea of defining a core microbiota in breast tumor tissues or gut of patients with BC is intriguing because of its potential association with carcinogenesis. In the current meta-analysis, the four most abundant phyla in all groups were Proteobacteria, Firmicutes, Actinobacteria, and Bacteroidetes, with relative abundances depending on the tissue type. Compared with that in normal and benign tissues, the Proteobacteria phylum was slightly more abundant in cancer tissues and lesion-adjacent tissues, accounting for 70.5% and 74.4% of all bacteria. Meanwhile, the abundance of Bacteroidetes was lower (3.4% in the cancer group and 2.3% in the lesion-adjacent tissue group) (Figure 2A). Relative to normal and benign tissues, cancer and lesion-adjacent tissues exhibited a greater proportion of *Halomonas* and *Shewanella* and a lesser proportion of *Pelomonas* and *Pseudomonas* at the genus level (Figure 2B). Gut bacterial communities were distinct from the tissue microbial



communities, and Bacteroidetes, Firmicutes, and Proteobacteria were the major bacterial phyla (Figure 2C). In the gut samples obtained from patients with cancer, the abundance of bacteria from the *Prevotella* genus was found to decline, whereas the relative abundance of the *Bacteroides* genus increased, accounting for 19.5% of the total genera identified (Figure 2D). The microbiota from different types of breast tumors have distinct compositions in the tissue or intestinal environment (Figures 2E, H).

Differential abundance analysis identifies significantly enriched and depleted features

We deployed five differential abundance analysis groups, namely, cancer tissue vs. normal tissue, lesion-adjacent tissue vs. normal tissue, cancer tissue vs. benign tissue, cancer stool vs. normal stool, and cancer stool vs. benign stool, to investigate the potential differences in the microbiota. Compared with those in normal tissues, *Sphingomonas*, *Veillonella*, *Porphyromonas*, and

Flavobacterium were found to be significantly enriched in cancer tissues, whereas nine genera were found to be significantly depleted (Figures 3A, B). In the lesion-adjacent tissue vs. normal tissue group, 14 enriched and 7 depleted genera were identified in lesion-adjacent tissues (Figure S3). Eleven enriched and 15 depleted genera were identified in the cancer tissues compared with those in the benign tissues (Figure S4). Twelve enriched genera (including *Bacteroides*, *Escherichia/Shigella*, *Lachnospiraceae_incertae_sedis*, *Roseburia*, *Ruminococcus2*, *Clostridium_XIVa*, *Parabacteroides*, and *Phascolarctobacterium*) and *Prevotella*, *Succiniclasticum*, and *Anaerotruncus* were observed in the intestinal microbiome, analyzed using stool samples from patients with cancer compared with those of healthy controls (Figure 3C). Compared with those in the stool samples of patients with benign disease, 10 enriched genera and 6 depleted genera were identified in the stool samples of patients with cancer (Figure S5). This led us to identify *Prevotella*, which was depleted in patients with BC compared with that in the non-cancer group and constituted an overlapping differential characteristic in both tissue and stool sampling (Figure 3D).

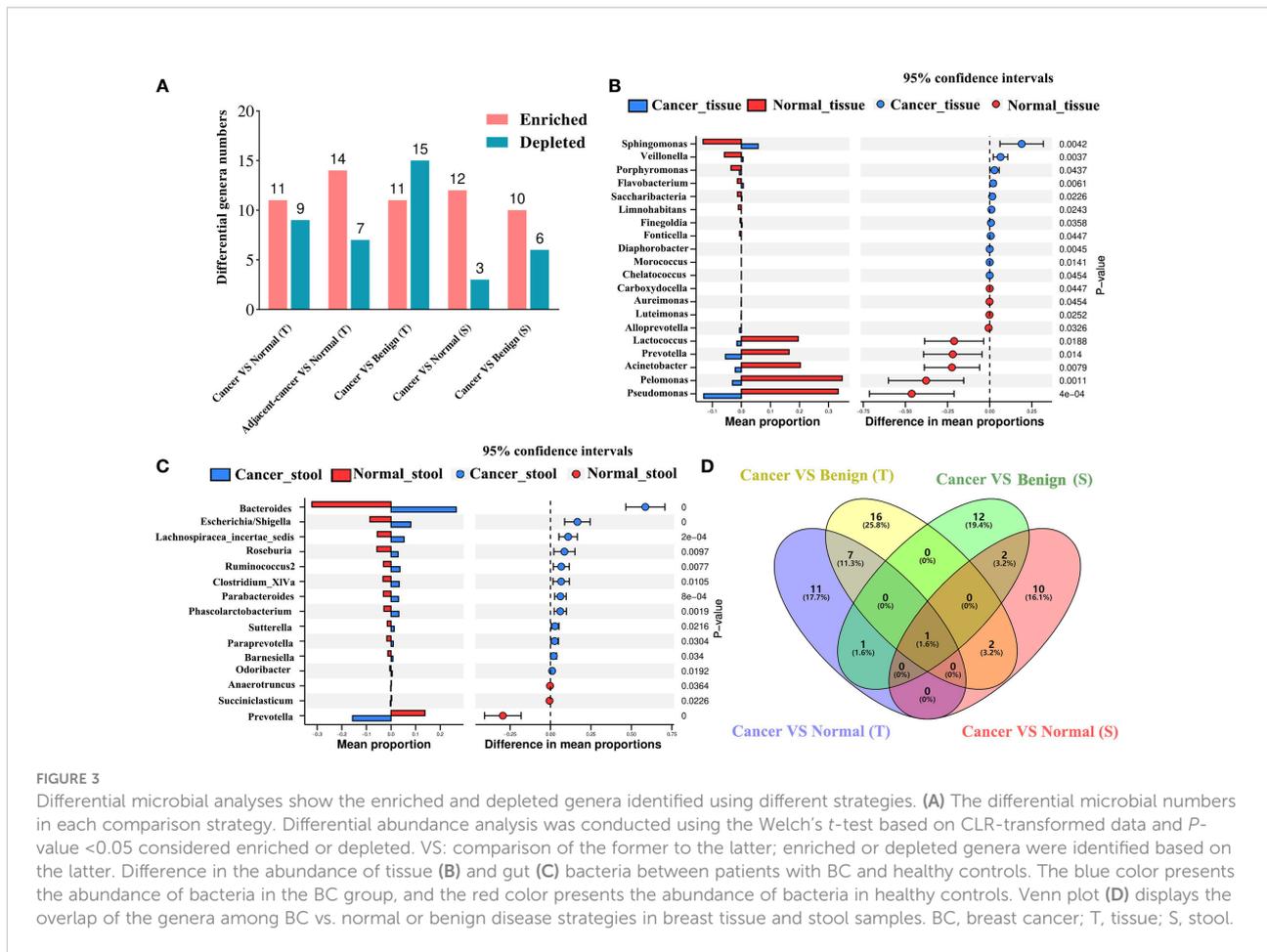


FIGURE 3 Differential microbial analyses show the enriched and depleted genera identified using different strategies. (A) The differential microbial numbers in each comparison strategy. Differential abundance analysis was conducted using the Welch’s *t*-test based on CLR-transformed data and *P*-value <0.05 considered enriched or depleted. VS: comparison of the former to the latter; enriched or depleted genera were identified based on the latter. Difference in the abundance of tissue (B) and gut (C) bacteria between patients with BC and healthy controls. The blue color presents the abundance of bacteria in the BC group, and the red color presents the abundance of bacteria in healthy controls. Venn plot (D) displays the overlap of the genera among BC vs. normal or benign disease strategies in breast tissue and stool samples. BC, breast cancer; T, tissue; S, stool.

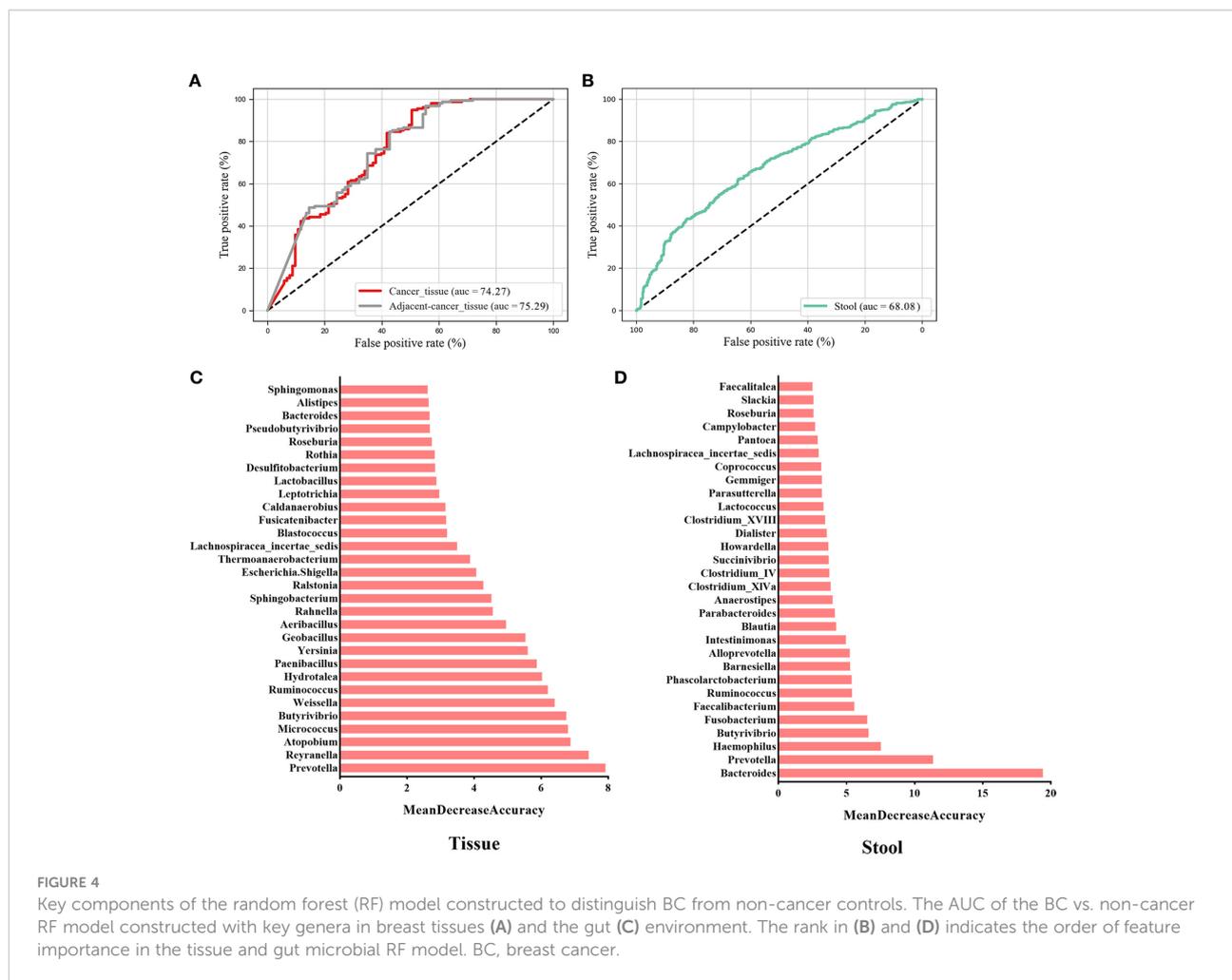
Microbial classification models for BC and non-cancer controls

To evaluate whether the microbial information can be used to distinguish cancer from non-cancer controls, we established 10-fold cross-validation RF classifiers by pooling tissue and stool samples, respectively. Here, benign and normal samples were classified into one category named “non-cancer.” The RF model (AUC = 74.27%) was constructed using the top 30 most important genera in the tissue samples (Figure 4A). The most important key microbial features were *Prevotella*, *Reyranella*, *Atopobium*, *Micrococcus*, *Butyrivibrio*, *Weissella*, *Ruminococcus*, *Hydrotalea*, *Paenibacillus*, and *Yersinia*, among others (Figure 4B). To determine whether cancer-adjacent tissues also play a pivotal role in distinguishing BC, we reconstructed the RF model using cancer-adjacent tissue and a non-cancerous microbiome, using the top 30 important genera in the cancerous vs. non-cancerous tissue model as the input for our new model. The AUC was 75.29% (Figure 4A). We noticed that the classifiers performed similarly in cancer vs. control and cancer-adjacent vs. control groups, likely because the cancer-adjacent tissue microbiome closely resembles that of cancerous tissues. This indicated that

adjacent tissues also possess the ability to diagnose BC. Additionally, when stool microbial characteristics were used to build the model for predicting cancer, the AUC was 68.08%, and the top 30 important features belong to *Bacteroides*, *Prevotella*, *Haemophilus*, *Butyrivibrio*, *Fusobacterium*, and *Faecalibacterium* (Figures 4C, D; Tables S1, S2).

Cohort-to-cohort and LOCO validation of BC classifiers

To evaluate whether the identified essential characteristics were reproducible across multiple studies, we conducted LOCO and cohort-to-cohort transfer validation by pooling data from cancerous and non-cancerous tissue samples. Cohort-to-cohort transfer validation was performed using the cancer stool vs. non-cancer stool strategy because only two cohorts were available. In the cancer tissue vs. non-cancer tissue RF models, the AUCs of cohort-to-cohort transfer validation ranged from 0.50 to 0.79, with an average of 0.65, whereas those in the LOCO analysis ranged from 0.60 to 0.69 (average AUC = 0.66) (Figure 5). Notably, the dataset provided by Nejman et al. was a better



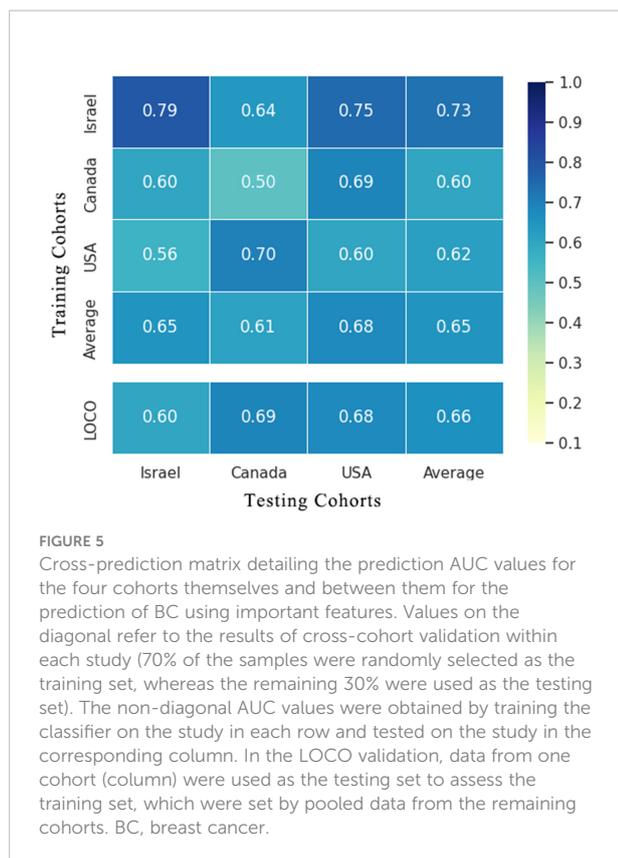
training set than those obtained from other studies and achieved a relatively higher testing AUC (average AUC = 0.73). When leaving the Neiman_Israel cohort out as the independent validation, the AUC decreased significantly compared with that of the other groups. This may be explained by the larger size of the dataset. Prediction models constructed using a larger sample size could help identify the signature signals of microorganisms associated with BC.

Models prepared using stool microbes could also be used to distinguish BC from healthy individuals. However, some weakness was noted in the cross-cohort validations. When the cohorts Goedert_USA and Byrd_Ghana were used as the validation datasets, respectively, and another cohort was used as the training dataset, the AUCs of the model were 0.54 and 0.59. This result suggests that low AUC values may be attributed to different geographical locations, and the assessment of characteristic gut microorganism communities requires accounting for geographical heterogeneity in dietary habits across different areas. Our results validated the predictive

ability and stability of the RF models prepared using important microbial genera as the feature set for BC diagnosis. Furthermore, geographic region, sample type, and sample size impact the diagnostic ability of the RF models.

Clostridium-specific-related DCA stimulates proliferation by promoting cell cycle entry into the S phase in HER2-positive BC cells

To quantify the distinctions corresponding to the breast tumor status and health, we performed the Wilcoxon rank-sum tests with CLR transformation data for discovering BC-related biomarkers in the tissue and gut microbiome, respectively. Among the more notable genera, *Clostridium_XIVa* is enriched in BC patients both in the tissue and gut environment and has the potential to be a biomarker of BC (Figures 6A, B).



We attempted to assess the proliferation role of *Clostridium*-specific DCA in the human breast cancer cells MDA-MB-231 and SK-BR-3 by performing the CCK8 assay. DCA increased the proliferation of HER2-positive SK-BR-3 cells compared with that of untreated cells (control), exerting the strongest effect at 80 μ M (Figure 6C). However, no significant role of DCA was determined in triple-negative MDA-MB-231 breast cancer cell proliferation (Figure 6D). As shown in Figure 6E, the number of colonies increased in cells treated for 36 h with 80 μ M of DCA compared with that in untreated cells. We subsequently verified, through flow cytometry analysis, whether the DCA-induced proliferation of breast cancer cells was associated with variations in the cell cycle. Treatment with DCA (0, 40, 60, 80 μ M) decreased the G0/G1 phase percentage from 36.0% to 27.1% and increased the S phase percentage from 31.4% to 43.3% in a dose-dependent manner (Figures 6F, G), thus facilitating G1/S progression. Furthermore, no significant effect of DCA was observed in the apoptosis of HER2-positive BC cells. The total apoptotic cell rates of SK-BR-3 cells were 3.3%, 3.0%, 2.5%, and 2.5% (Figures 6H, I). These data suggested that DCA stimulates the proliferation and growth of HER2-positive BC cells and promotes cell cycle progression into the S phase but does not contribute to HER2-positive BC cell apoptosis.

Quantitative proteomic analysis revealed changes in protein expression in DCA-treated SK-BR-3 BC cells

We performed quantitative tandem mass spectrometry experiments using 4D label-free labeling of DCA-treated and untreated SK-BR-3 BC cells. In total, we identified 5,331 proteins in both groups. Of these, 24 differentially expressed proteins (5 upregulated and 19 downregulated) were identified ($P < 0.05$ after t -test, fold change ≥ 2.0 , Figure 7A). The heatmap of the hierarchical cluster analysis showed that these proteins were well distinguished, which helped clearly visualize the changes in protein expression (Figure 7B). GO and KEGG pathway enrichment analyses were performed between DCA-treated and untreated SK-BR-3 BC cells. The most enriched GO terms of biological processes (BP), molecular functions (MF), and cellular components (CC) were annotated as protein O-linked fucosylation (two proteins, richFactor = 148.08), peptide-O-fucosyltransferase activity (two proteins, richFactor = 222.12), and endoplasmic reticulum lumen (seven proteins, richFactor = 12.54), respectively (Figures 7C, D, Table 3). The results of the KEGG pathway analysis reveal that neuroactive ligand-receptor interaction emerged as the most drastically enriched pathway and was linked to three upregulated proteins in DCA-treated BC cells, namely, nicotinic acetylcholine receptor alpha-9, P2X purinoceptor 5, and partitioning defective protein 3 (Figure 7E, Table 3). A complex network of proteins with high expression and significant differences in expression is shown in Figure 7F.

Discussion

Multiple studies worldwide have provided evidence on the close association between BC and microbial dysbiosis (Xuan et al., 2014; Goedert et al., 2015; Hieken et al., 2016; Costantini et al., 2018; Plaza-Diaz et al., 2019; Nejman et al., 2020; Bobin-Dubigeon et al., 2021; Byrd et al., 2021), and the abundances of specialized microorganisms are also correlated with an immune signature and prognostic characteristics (Gopalakrishnan et al., 2018; Terrisse et al., 2021; Tzeng et al., 2021). However, there exists a significant lack of agreement in the types of breast tumor-associated microorganisms identified in published studies. Here, we first presented a large-scale, integrative meta-analysis on public 16S rRNA datasets to characterize the breast tissue and gut microbiota signatures of patients with breast tumors and assess the potential of important microbial signatures for distinguishing BC from health statuses. Compared with datasets from a single study, pooled datasets of breast tissues and gut microbiota from multiple studies enabled the identification of microbiota signatures more accurately and comprehensively.

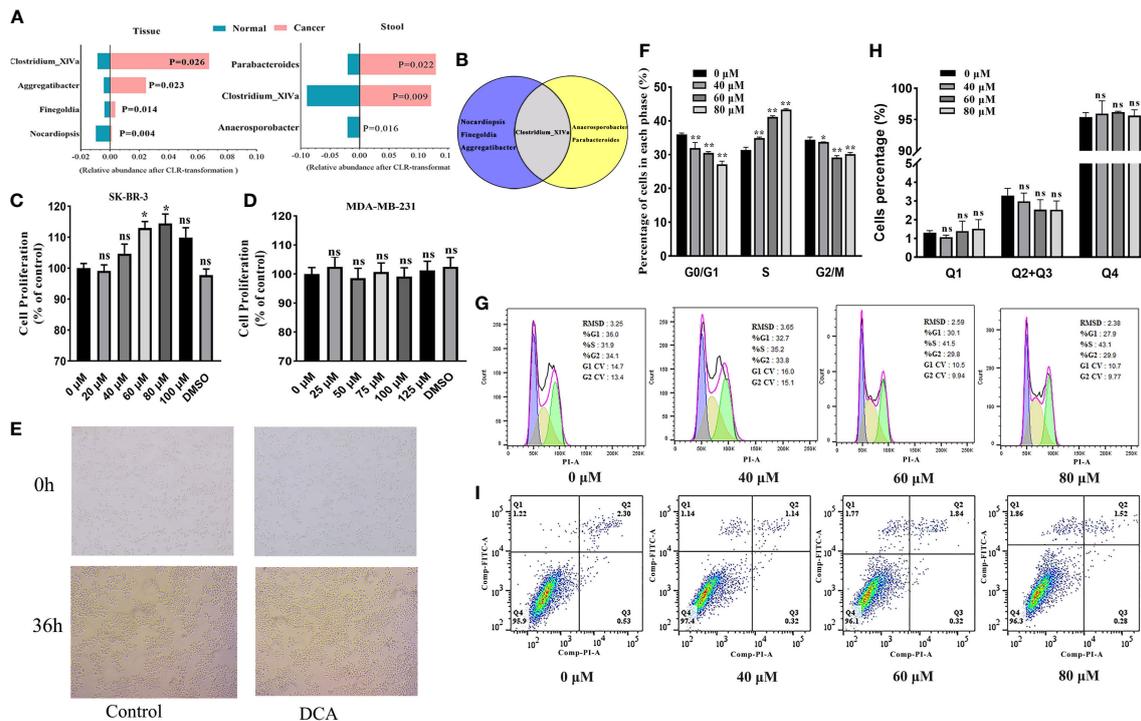


FIGURE 6

(A) The BC-related biomarkers in breast tissue and the gut environment (only BC and healthy controls are considered here). Venn plot (B) displays the *Clostridium_XIVa* as the overlap biomarker in both breast tissue and the gut environment. The SK-BR-3 and MDA-MB-231 cells were cultured in DMEM for 12 h and then treated with various concentrations of deoxycholic acid (DCA). The bar graph shows the percentage of viable SK-BR-3 (C) and MDA-MB-231 (D) cells after treatment with DCA for 24 h, as determined in the CCK8 assay. (E) Light microscopy showed proliferation-related changes in SK-BR-3 cells treated with 80 μM of DCA for 36 h. (F, G) Distribution of SK-BR-3 cells, treated for 24 h with 0, 40, 60, and 80 μM of DCA, in the G0/G1, S, and G2/M phases, determined using flow cytometry. (H, I) SK-BR-3 cells were stained with Annexin V/PI and analyzed by flow cytometry after being treated with 24 h with 0, 40, 60, and 80 μM of DCA. The percentage of apoptotic cells is shown in the histogram. The different quadrants Q1, Q2, Q3, and Q4 represent necrotic cells, early apoptotic cells, late apoptotic cells, and viable cells, respectively. Results are presented as means ± SEM from at least three independent experiments. * $P < 0.05$, ** $P < 0.01$ vs. control in each group. ns, no significance.

This meta-analysis revealed significant changes in the microbial flora of patients with BC, and we also found that critical microbial characteristics can be used to diagnose BC. In pooled RF models, specific microbes from malignant breast tissue could be used to discriminate between BC and non-cancer populations and yielded AUCs of 74.27% in the training module. Cohort heterogeneity is critical in trans-cohort generalization. In our work, important microbial features were validated for effectiveness *via* cohort-to-cohort transfer (average AUC = 0.65) as well as LOCO validation (average AUC = 0.66) while underscoring the adverse impact of technical and geographical discrepancies on the model's generalizability. Consistent with our hypothesis, microbial components are similar in malignant breast and lesion-adjacent tissues. The AUC was 75.29% in cancer-adjacent tissues vs. non-cancer control RF models, which was similar to that in cancer vs. non-cancer population models (AUC = 74.27%). The result indicates that the lesion-adjacent tissue

microbiome is sufficient for BC prediction, and lesion-adjacent tissues could be used for sampling during BC screening.

It is generally known that intestinal microbes actively participate in the development of various cancers, such as ovarian (Cheng et al., 2020; Laniewski et al., 2020), prostate (Fujita et al., 2022; O'Rourke, 2022), breast (Ingman, 2019; Hossain et al., 2021; Zhang et al., 2021), and colorectal cancers (Wong and Yu, 2019; Song et al., 2020; Rebersek, 2021). The known microbial mechanisms of action can modulate the tumor microenvironment by regulating the non-hematopoietic and hematopoietic components of the intestinal epithelial barrier and primary and secondary lymphoid organ activities, thereby affecting cancer occurrence and development (Sepich-Poore et al., 2021). Recent findings have revealed that mammary microbiota may be derived from the gut. Intestinal bacteria have easy access to the mammary gland *via* the entero-mammary pathway. This mechanism involves CD18⁺ cells and dendritic cells, which have the capacity to transmit intestinal microbes into the mammary

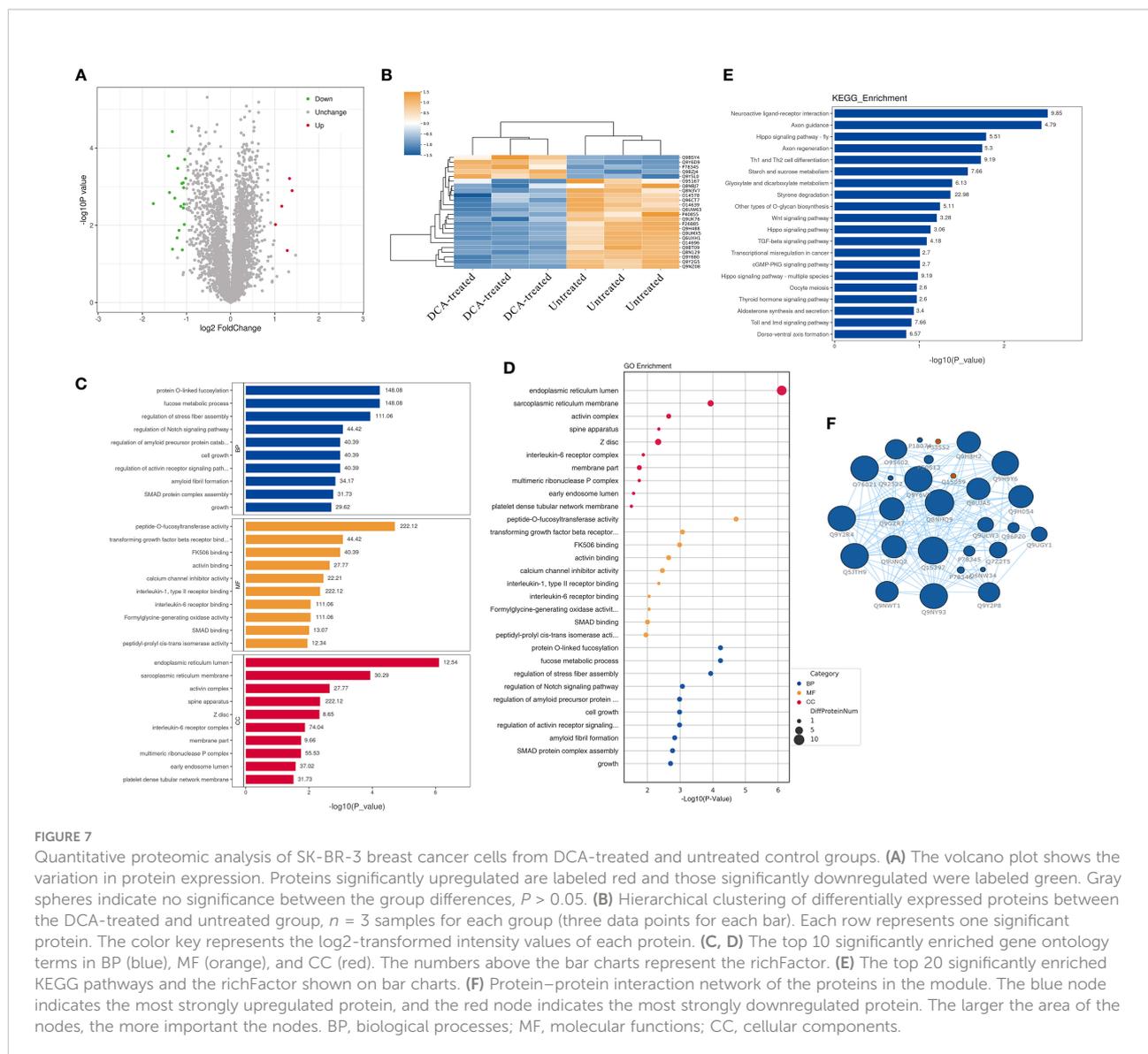


FIGURE 7

Quantitative proteomic analysis of SK-BR-3 breast cancer cells from DCA-treated and untreated control groups. (A) The volcano plot shows the variation in protein expression. Proteins significantly upregulated are labeled red and those significantly downregulated were labeled green. Gray spheres indicate no significance between the group differences, $P > 0.05$. (B) Hierarchical clustering of differentially expressed proteins between the DCA-treated and untreated group, $n = 3$ samples for each group (three data points for each bar). Each row represents one significant protein. The color key represents the \log_2 -transformed intensity values of each protein. (C, D) The top 10 significantly enriched gene ontology terms in BP (blue), MF (orange), and CC (red). The numbers above the bar charts represent the richFactor. (E) The top 20 significantly enriched KEGG pathways and the richFactor shown on bar charts. (F) Protein-protein interaction network of the proteins in the module. The blue node indicates the most strongly upregulated protein, and the red node indicates the most strongly downregulated protein. The larger the area of the nodes, the more important the nodes. BP, biological processes; MF, molecular functions; CC, cellular components.

TABLE 3 Distribution of proteins and signaling pathways after DCA treatment based on GO and KEGG analyses.

Database	Term	Count	P-value	FDR	richFactor	Protein names
GO	(BP) protein O-linked fucosylation	2	5.81203E-05	0.0041	148.08	POFUT1 (O-fucosyltransferase 1), POFUT2 (O-fucosyltransferase 2)
	(MF) Peptide-O-fucosyltransferase activity	2	1.94269E-05	0.0028	222.12	POFUT1 (O-fucosyltransferase 1), POFUT2 (O-fucosyltransferase 2)
	(CC) endoplasmic reticulum lumen	7	7.74755E-07	0.0002	12.54	FKBP7 (peptidyl-prolyl cis-trans isomerase FKBP7), ERAP1 (endoplasmic reticulum aminopeptidase 1), SUMF2 (inactive C-alpha-formylglycine-generating enzyme 2), POGlut2 (O-glucosyltransferase 2), CNPY3 (canopy homolog 3), CNPY4 (canopy homolog 4), FKBP2 (peptidyl-prolyl cis-trans isomerase FKBP2)
KEGG	Neuroactive ligand-receptor interaction	3	0.003067213	0.2688	9.85	CHRNA9 (nicotinic acetylcholine receptor alpha-9), P2RX5 (P2X purinoceptor 5), PARD3 (partitioning defective protein 3)

BP, biological processes; MF, molecular functions; CC, cellular components; FDR, false discovery rate.

gland (Angelopoulou et al., 2018; Chadha et al., 2021; Wang et al., 2021). In the present study, we showed that the abundance of *Bacteroides* in patients with BC was significantly higher than that in controls. This result is consistent with the findings from previous studies, which suggested *Bacteroides* associations with BC (Zhu et al., 2018; Byrd et al., 2021; Hou et al., 2021). Enterotoxigenic *Bacteroides fragilis* accelerated tumor cell growth in the mammary glands or intestines and metastatic progression in breast ducts using secretory *B. fragilis* toxins. Parida et al. (2021) showed that compared with control mice, mice harboring gut enterotoxigenic *B. fragilis* showed an almost 3.3-fold increase in the tumor volume over 7 weeks. It is promising that not all bacteria species are detrimental. *Prevotella* is considered a beneficial bacteria based on its low levels in patients with BC and is the only co-differential bacterium present in breast tissues and the intestinal environment in cancer and non-cancer populations. Evidence from existing literature suggests that *Prevotella* potentially exerts an antitumor effect through its metabolites. Acetic acid, the major metabolic product of *Prevotella*, significantly inhibited the synthesis of nitric oxide in AGS cells (Dong et al., 2017), which could enhance protooncogene expression and inhibit the apoptosis of tumor cells (Li et al., 1997; Ambs et al., 1998).

With respect to cancer diagnosis, critical intestinal microbial characteristics can help discriminate between cancer and non-cancer populations. In our study, RF models constructed using data on important gut microbiota (AUC = 68.08%) could provide insights on BC diagnosis. However, gut microbiota-related RF models yield a lower accuracy than tissue microbes in distinguishing cancer from non-cancer controls (AUC = 74.27%). This is likely because tissue-based samples are invasive and assess microbial structure at the lesion site more accurately than stool-based samples. Zhu et al. (2018) showed that the training set of fecal samples performed remarkably when patients with BC and healthy controls were compared, and the AUC value achieved was 85.52%, which was greater than the value obtained in the present study. One explanation is that in the stool samples from multiple studies, ethnic heterogeneity, varied dietary habits, and different exposure factors within groups could contribute to the discrepancies in the intestinal microbial communities, thus affecting model performance. Furthermore, the performance is potentially influenced by the different sequencing procedures, such as shotgun metagenomic sequencing and 16S rRNA sequencing. Shotgun metagenomic sequencing data are well-recognized to be advantageous for the identification of special microbes up to the species and even the strain level, which is more conducive to improving accuracy in BC diagnostic models. More importantly, the AUC value of the BC health model based on stool microbiota reported by Zhu et al. was greater than 80%, which further illustrates that the specific intestinal microbiota has BC diagnostic potential comparable to breast tissue microbiota, even if it is present far from the breast lesion site.

Currently, BC diagnosis is mainly based on imaging examinations, and mammography, magnetic resonance imaging

(MRI), ultrasound imaging, and computed tomography (CT) are the most common methods used in clinical applications, with a sensitivity of more than 80% (Jafari et al., 2018). However, despite the various benefits of imaging techniques, they have some limitations, such as costs and potential radiation risks. The most important caveat is that imaging modalities may yield false-positive and false-negative errors. For example, if lesions labeled as BI-RADS category $\geq 4b$ are considered to be malignant, the rates of misdiagnosis under mammography, ultrasound, and MRI would be 17.9%, 17.5%, and 35.3%, respectively (Tan et al., 2019). Although the AUCs are low owing to the influence of the sequencing platform, region, and other factors, BC and non-cancer states can be assessed using characteristic microorganisms. Particularly, gut microbes, as a non-intrusive and convenient method, are easily applicable in clinical settings.

As for the BC-associated biomarker *Clostridium_XIVa*, Huang et al. (2020) provided that *Clostridium XIVa* was enriched in non-small hepatocellular carcinoma. Based on the transcriptome and fecal microbiota data, the increased abundance of *Clostridium XIVa* accompanied by a decrease in CCL21, which facilitates tumor growth, indicates an unfavorable prognosis. *Clostridium XIVa* is highly potent in enhancing Treg cell abundance and inducing important anti-inflammatory molecules such as interleukin-10 (Atarashi et al., 2013), as well as affecting bile acid-controlled NKT cell accumulation. Changes in the *Clostridium XIVa* may result in alterations of the tumor microenvironment, which highlight the potential of *Clostridium XIVa* as a biomarker for breast cancer. Recent data suggest that the human microbiota contributes to several common types of cancer, not only by playing procarcinogenic roles of specific pathogens but also via the influence of their metabolome. For example, the intestinal microbiome-metabolite formate drives CRC tumor invasion via triggering AhR signaling while increasing cancer stemness (Ternes et al., 2022). Obesity induces changes in the gut microbiota, thereby increasing the levels of the gut bacterial metabolite DCA. The enterohepatic circulation of DCA stimulates the SASP phenotype in hepatic stellate cells, which in turn secretes various inflammatory and protumor factors in the liver (Yoshimoto et al., 2013). (Costarelli and Sanders, 2002) reported that plasma DCA concentration in patients with breast cancer was 52% higher than that in healthy women, supporting the hypothesis that DCA might be related to the occurrence and development of breast cancer. In this study, we demonstrated that *Clostridium*-specific DCA plays a molecule type-specific role in the proliferation of BC cells, significantly promoting the proliferation of HER2-positive BC cells but not affecting triple-negative BC cells. As a secondary metabolite of bile acids, DCA increases ROS production and DNA damage, which could cause K-RAS mutations, aneuploidy, and micronuclei formation (Si et al., 2021). Bile acids are also present in breast tissues and may originate from the gut (Rezen et al., 2022). Findings from proteomics analysis indicate that DCA promotes the activation of peptide-O-fucosyltransferase activity and triggers the neuroactive ligand-receptor interaction pathway, which

may play a role in cancer progression. Meanwhile, *POFUT1* and *POFUT2* genes, involved in cancer-related pathways, have been screened. The overexpression of *POFUT1* leads to Notch1 signaling dysregulation associated with poor prognosis (Wan et al., 2017; Deschuyter et al., 2020). Leng et al. (2018) stated that genes encoding proteins associated with fucosylation, including *POFUT1* and *POFUT2*, may serve as circulating biomarkers of lung cancer. Objectively, the findings suggest that *POFUT1* and *POFUT2* are worthy of further investigation for their functions in BC.

Unlike previous studies, ours is the first to use comprehensive datasets to investigate the microbial community in both breast and gut samples of patients with BC. Based on the findings from the meta-analysis of eight studies, the interference of biotechnology and regional differences were reduced, and we identified a consistent and reliable variation in the microbiome of patients with BC. Additionally, important microbial features of the tissue or intestinal environment were useful for the diagnosis of BC and could provide a new approach for cancer screening. We recognize that the study had limitations in validation, such as the lack of additional validation using actual clinical patient samples. To overcome this drawback, we strived to strengthen the evidence from other perspectives of the study design and used interstudy cross-validation and within-study validation simultaneously to verify the stability of important microbial features in BC diagnosis. Lastly, the interaction between microorganisms and the occurrence and progression of cancer is very complex, and a variety of bacteria or its metabolites may promote or inhibit the development of breast cancer. Herein, we only examined the effect of *Clostridium*-specific metabolite DCA on the proliferation, cycle, and apoptosis of breast cancer cells. More microorganisms and its metabolites are still worth exploring. In our future studies, external validation in large clinical populations will be necessary, and we would also like to integrate additional factors, such as lymph node metastasis, Ki67 expression, molecular subtypes of breast cancer, and progression-free survival to determine the microbiome associated with breast cancer prognosis.

By re-analyzing raw 16S rRNA gene sequence data from fecal and tissue samples, we suggested a strong association between the microbiota and breast tumorigenesis. The overlapping specific microbes of the breast tissue and gut environment may provide biomarkers of BC, and the similarities between the carcinoma and adjacent normal tissues indicate that microorganisms might pave the way for our understanding of microbial roles in carcinogenesis. We also found that the *Clostridium*-specific metabolite DCA could stimulate proliferation by promoting cell cycle progression into the S phase. From the perspective of disease diagnostics, important microbial features could aid BC screening at early stages. In conclusion, the findings of this study provide a novel perspective for further insights on the role of the microbiome in BC tumorigenesis and progression.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/supplementary material.

Author contributions

JX and TS conceived and designed the article. NW and JX wrote the draft of the paper. WH and MH collected and organized the published 16S rRNA sequencing data. NW and XL analyzed the metagenomics data. NW and JY completed the *in-vitro* cell experiment. LJ, HC, MJ, TS, and JX participated in the manuscript revision. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by Shenyang Breast Cancer Clinical Medical Research Center (2020-48-3-1, TS), Liaoning Cancer Hospital Yangtse River Scholars Project (TS, JX), LiaoNing Revitalization Talents Program (XLYC1907160, JX), Beijing Medical Award Foundation (YXJL-2020-0941-0752, TS; CORP-239-N27, HC), Wu Jieping Medical Foundation (320.6750.2020-12-21, 320.6750.2020-6-30, TS), and the Fundamental Research Funds for the Central Universities (2022029, TS; 2022030, JX).

Acknowledgments

We thank all the studies for sharing their data publicly.

Conflict of interest

XL was employed by the company Liaoning Kanghui Biotechnology Co., Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the

reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fcimb.2022.1029905/full#supplementary-material>

References

- Ambs, S., Ogunfusika, M. O., Merriam, W. G., Bennett, W. P., Billiar, T. R., and Harris, C. C. (1998). Up-regulation of inducible nitric oxide synthase expression in cancer-prone p53 knockout mice. *Proc. Natl. Acad. Sci. U.S.A.* 95 (15), 8823–8828. doi: 10.1073/pnas.95.15.8823
- Angelopoulou, A., Field, D., Ryan, C. A., Stanton, C., Hill, C., and Ross, R. P. (2018). The microbiology and treatment of human mastitis. *Med. Microbiol. Immunol.* 207 (2), 83–94. doi: 10.1007/s00430-017-0532-z
- Atarashi, K., Tanoue, T., Oshima, K., Suda, W., Nagano, Y., and Nishikawa, H. (2013). Treg induction by a rationally selected mixture of clostridia strains from the human microbiota. *Nature* 500, 232–236. doi: 10.1038/nature12331
- Bobin-Dubigeon, C., Luu, H. T., Leuillet, S., Laverigne, S. N., Carton, T., Le Vacon, F., et al. (2021). Faecal microbiota composition varies between patients with breast cancer and healthy women: A comparative case-control study. *Nutrients* 13 (8), 2705. doi: 10.3390/nu13082705
- Byrd, D. A., Vogtmann, E., Wu, Z., Han, Y., Wan, Y., Clegg-Lamprey, J. N., et al. (2021). Associations of fecal microbial profiles with breast cancer and nonmalignant breast disease in the Ghana breast health study. *Int. J. Cancer* 148 (11), 2712–2723. doi: 10.1002/ijc.33473
- Chadha, J., Nandi, D., Atri, Y., and Nag, A. (2021). Significance of human microbiome in breast cancer: Tale of an invisible and an invincible. *Semin. Cancer Biol.* 70, 112–127. doi: 10.1016/j.semcancer.2020.07.010
- Cheng, H., Wang, Z., Cui, L., Wen, Y., Chen, X., Gong, F., et al. (2020). Opportunities and challenges of the human microbiome in ovarian cancer. *Front. Oncol.* 10. doi: 10.3389/fonc.2020.00163
- Chiang, J. Y. (2013). Bile acid metabolism and signaling. *Compr. Physiol.* 3 (3), 1191–1212. doi: 10.1002/cphy.c120023
- Costantini, L., Magno, S., Albanese, D., Donati, C., Molinari, R., Filippone, A., et al. (2018). Characterization of human breast tissue microbiota from core needle biopsies through the analysis of multi hypervariable 16S-rRNA gene regions. *Sci. Rep.* 8 (1), 16893. doi: 10.1038/s41598-018-35329-z
- Costarelli, V., and Sanders, T. A. (2002). Plasma deoxycholic acid concentration is elevated in postmenopausal women with newly diagnosed breast cancer. *Eur. J. Clin. Nutr.* 56, 925–927. doi: 10.1038/sj.ejcn.1601396
- Deschuyter, M., Pennarubia, F., Pinault, E., Legardinier, S., and Maftah, A. (2020). Functional characterization of POFUT1 variants associated with colorectal cancer. *Cancers (Basel)* 12 (6), 1430. doi: 10.3390/cancers12061430
- Dong, T., Feng, Q., Liu, F., Chang, L. K., Zhou, X., Han, M., et al. (2017). Alteration of stomach microbiota compositions in the progression of gastritis induces nitric oxide in gastric cell. *Exp. Ther. Med.* 13 (6), 2793–2800. doi: 10.3892/etm.2017.4373
- Elmore, J. G., Armstrong, K., Lehman, C. D., and Fletcher, S. W. (2005). Screening for breast cancer. *JAMA* 293 (10), 1245–1256. doi: 10.1001/jama.293.10.1245
- Fujita, K., Matsushita, M., Banno, E., De Velasco, M. A., Hatano, K., Nonomura, N., et al. (2022). Gut microbiome and prostate cancer. *Int. J. Urol.* 29 (8), 793–798. doi: 10.1111/iju.14894
- Goedert, J. J., Hua, X., Bielecka, A., Okayasu, I., Milne, G. L., Jones, G. S., et al. (2018). Postmenopausal breast cancer and oestrogen associations with the IgA-coated and IgA-noncoated faecal microbiota. *Br. J. Cancer* 118 (4), 471–479. doi: 10.1038/bjc.2017.435
- Goedert, J. J., Jones, G., Hua, X., Xu, X., Yu, G., Flores, R., et al. (2015). Investigation of the association between the fecal microbiota and breast cancer in postmenopausal women: a population-based case-control pilot study. *J. Natl. Cancer Inst.* 107 (8), djv147. doi: 10.1093/jnci/djv147
- Gopalakrishnan, V., Spencer, C. N., Nezi, L., Reuben, A., Andrews, M. C., Karpinets, T. V., et al. (2018). Gut microbiome modulates response to anti-PD-1 immunotherapy in melanoma patients. *Science* 359 (6371), 97–103. doi: 10.1126/science.aan4236
- Guan, X., Ma, F., Sun, X., Li, C., Li, L., Liang, F., et al. (2020). Gut microbiota profiling in patients with HER2-negative metastatic breast cancer receiving metronomic chemotherapy of capecitabine compared to those under conventional dosage. *Front. Oncol.* 10. doi: 10.3389/fonc.2020.00902
- Gurevitch, J., Koricheva, J., Nakagawa, S., and Stewart, G. (2018). Meta-analysis and the science of research synthesis. *Nature* 555 (7695), 175–182. doi: 10.1038/nature25753
- Hieken, T. J., Chen, J., Hoskin, T. L., Walther-Antonio, M., Johnson, S., Ramaker, S., et al. (2016). The microbiome of aseptically collected human breast tissue in benign and malignant disease. *Sci. Rep.* 6, 30751. doi: 10.1038/srep30751
- Hossain, F., Majumder, S., David, J., Bunnell, B. A., and Miele, L. (2021). Obesity modulates the gut microbiome in triple-negative breast cancer. *Nutrients* 13 (10), 3656. doi: 10.3390/nu13103656
- Hou, M. F., Ou-Yang, F., Li, C. L., Chen, F. M., Chuang, C. H., Kan, J. Y., et al. (2021). Comprehensive profiles and diagnostic value of menopausal-specific gut microbiota in premenopausal breast cancer. *Exp. Mol. Med.* 53 (10), 1636–1646. doi: 10.1038/s12276-021-00686-9
- Huang, H., Ren, Z., Gao, X., Hu, X., Zhou, Y., Jiang, J., et al. (2020). Integrated analysis of microbiome and host transcriptome reveals correlations between gut microbiota and clinical outcomes in HBV-related hepatocellular carcinoma. *Genome Med.* 12, 102. doi: 10.1186/s13073-020-00796-5
- Ingman, W. V. (2019). The gut microbiome: A new player in breast cancer metastasis. *Cancer Res.* 79 (14), 3539–3541. doi: 10.1158/0008-5472.CAN-19-1698
- Jafari, S. H., Saadatpour, Z., Salmaninejad, A., Momeni, F., Mokhtari, M., Nahand, J. S., et al. (2018). Breast cancer diagnosis: Imaging techniques and biochemical markers. *J. Cell Physiol.* 233 (7), 5200–5213. doi: 10.1002/jcp.26379
- Laniewski, P., Ilhan, Z. E., and Herbst-Kralovetz, M. M. (2020). The microbiome and gynaecological cancer development, prevention and therapy. *Nat. Rev. Urol.* 17 (4), 232–250. doi: 10.1038/s41585-020-0286-z
- Leng, Q., Tsou, J. H., Zhan, M., and Jiang, F. (2018). Fucosylation genes as circulating biomarkers for lung cancer. *J. Cancer Res. Clin. Oncol.* 144 (11), 2109–2115. doi: 10.1007/s00432-018-2735-0
- Li, J., Billiar, T. R., Talanian, R. V., and Kim, Y. M. (1997). Nitric oxide reversibly inhibits seven members of the caspase family via s-nitrosylation. *Biochem. Biophys. Res. Commun.* 240 (2), 419–424. doi: 10.1006/bbrc.1997.7672
- Mo, Z., Huang, P., Yang, C., Xiao, S., Zhang, G., Ling, F., et al. (2020). Meta-analysis of 16S rRNA microbial data identified distinctive and predictive microbiota dysbiosis in colorectal carcinoma adjacent tissue. *mSystems* 5 (2), e00138-20. doi: 10.1128/mSystems.00138-20
- Nagathihalli, N. S., Beesetty, Y., Lee, W., Washington, M. K., Chen, X., Lockhart, A. C., et al. (2014). Novel mechanistic insights into ectodomain shedding of EGFR ligands amphiregulin and TGF- α : impact on gastrointestinal cancers driven by secondary bile acids. *Cancer Res.* 74 (7), 2062–2072. doi: 10.1158/0008-5472.CAN-13-2329
- Nejman, D., Livyatan, I., Fuks, G., Gavert, N., Zvang, Y., Geller, L. T., et al. (2020). The human tumor microbiome is composed of tumor type-specific intracellular bacteria. *Science* 368 (6494), 973–980. doi: 10.1126/science.aay9189
- O'Rourke, K. (2022). Gut microbiome linked to aggressive prostate cancer. *Cancer* 128 (5), 938. doi: 10.1002/cncr.34101

SUPPLEMENTARY FIGURE 1-2

The total variance explained by disease status (BC, benign tumor, health) are plotted against total explained by patient age, BMI in studies using breast tissue and fecal samples.

SUPPLEMENTARY FIGURE 3-5

Results of the for differential microbial features in lesion-adjacent tissue vs. normal tissue, cancer tissue vs. benign tissue, cancer stool vs. benign stool groups (P<0.05).

SUPPLEMENTARY TABLE 1-2

Key components of the random forest model constructed in this study.

- Parida, S., Wu, S., Siddharth, S., Wang, G., Muniraj, N., Nagalingam, et al. (2021). A procarcinogenic colon microbe promotes breast tumorigenesis and metastatic progression and concomitantly activates notch and beta-catenin axes. *Cancer Discovery* 11 (5), 1138–1157. doi: 10.1158/2159-8290.CD-20-0537
- Plaza-Diaz, J., Alvarez-Mercado, A. I., Ruiz-Marin, C. M., Reina-Perez, I., Perez-Alonso, A. J., Sanchez-Andujar, M. B., et al. (2019). Association of breast and gut microbiota dysbiosis and the risk of breast cancer: a case-control clinical study. *BMC Cancer* 19 (1), 495. doi: 10.1186/s12885-019-5660-y
- Rebersek, M. (2021). Gut microbiome and its role in colorectal cancer. *BMC Cancer* 21 (1), 1325. doi: 10.1186/s12885-021-09054-2
- Rezen, T., Rozman, D., Kovacs, T., Kovacs, P., Sipos, A., Bai, P., et al. (2022). The role of bile acids in carcinogenesis. *Cell Mol. Life Sci.* 79 (5), 243. doi: 10.1007/s00018-022-04278-2
- Sepich-Poore, G. D., Zitvogel, L., Straussman, R., Hastay, J., Wargo, J. A., and Knight, R. (2021). The microbiome and human cancer. *Science* 371 (6536), eabc4552. doi: 10.1126/science.abc4552
- Shively, C. A., Register, T. C., Appt, S. E., Clarkson, T. B., Uberseder, B., Clear, K. Y. J., et al. (2018). Consumption of Mediterranean versus Western diet leads to distinct mammary gland microbiome populations. *Cell Rep.* 25 (1), 47–56 e43. doi: 10.1016/j.celrep.2018.08.078
- Siegel, R. L., Miller, K. D., Fuchs, H. E., and Jemal, A. (2021). Cancer statistics 2021. *CA Cancer J. Clin.* 71 (1), 7–33. doi: 10.3322/caac.21654
- Si, H., Yang, Q., Hu, H., Ding, C., Wang, H., and Lin, X. (2021). Colorectal cancer occurrence and treatment based on changes in intestinal flora. *Semin. Cancer Biol.* 70, 3–10. doi: 10.1016/j.semcancer.2020.05.004
- Song, M., Chan, A. T., and Sun, J. (2020). Influence of the gut microbiome, diet, and environment on risk of colorectal cancer. *Gastroenterology* 158 (2), 322–340. doi: 10.1053/j.gastro.2019.06.048
- Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., et al. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* 71 (3), 209–249. doi: 10.3322/caac.21660
- Sze, M. A., and Schloss, P. D. (2018). Leveraging existing 16S rRNA gene surveys to identify reproducible biomarkers in individuals with colorectal tumors. *mBio* 9 (3), e00630-18. doi: 10.1128/mBio.00630-18
- Tan, H., Zhang, H., Lei, Z., Fu, F., and Wang, M. (2019). Radiological and clinical findings in sclerosing adenosis of the breast. *Med. (Baltimore)* 98 (39), e17061. doi: 10.1097/MD.00000000000017061
- Ternes, D., Tsenkova, M., Pozdeev, V. I., Meyers, M., Koncina, E., Atatri, S., et al. (2022). The gut microbial metabolite formate exacerbates colorectal cancer progression. *Nat. Metab.* 4 (4), 458–475. doi: 10.1038/s42255-022-00558-0
- Terrisse, S., Derosa, L., Iebba, V., Ghiringhelli, F., Vaz-Luis, I., Kroemer, G., et al. (2021). Intestinal microbiota influences clinical outcome and side effects of early breast cancer treatment. *Cell Death Differ* 28 (9), 2778–2796. doi: 10.1038/s41418-021-00784-1
- Thyagarajan, S., Zhang, Y., Thapa, S., Allen, M. S., Phillips, N., Chaudhary, P., et al. (2020). Comparative analysis of racial differences in breast tumor microbiome. *Sci. Rep.* 10 (1), 14116. doi: 10.1038/s41598-020-71102-x
- Tzeng, A., Sangwan, N., Jia, M., Liu, C. C., Keslar, K. S., Downs-Kelly, E., et al. (2021). Human breast microbiome correlates with prognostic features and immunological signatures in breast cancer. *Genome Med.* 13 (1), 60. doi: 10.1186/s13073-021-00874-2
- Urbaniak, C., Gloor, G. B., Brackstone, M., Scott, L., Tangney, M., and Reid, G. (2016). The microbiota of breast tissue and its association with breast cancer. *Appl. Environ. Microbiol.* 82 (16), 5039–5048. doi: 10.1128/AEM.01235-16
- Wang, L., Gong, Z., Zhang, X., Zhu, F., Liu, Y., Jin, C., et al. (2020). Gut microbial bile acid metabolite skews macrophage polarization and contributes to high-fat diet-induced colonic inflammation. *Gut Microbes* 12 (1), 1–20. doi: 10.1080/19490976.2020.1819155
- Wang, N., Sun, T., and Xu, J. (2021). Tumor-related microbiome in the breast microenvironment and breast cancer. *J. Cancer* 12 (16), 4841–4848. doi: 10.7150/jca.58986
- Wan, G., Tian, L., Yu, Y., Li, F., Wang, X., Li, C., et al. (2017). Overexpression of Pofut1 and activated Notch1 may be associated with poor prognosis in breast cancer. *Biochem. Biophys. Res. Commun.* 491 (1), 104–111. doi: 10.1016/j.bbrc.2017.07.053
- Wong, S. H., and Yu, J. (2019). Gut microbiota in colorectal cancer: mechanisms of action and clinical applications. *Nat. Rev. Gastroenterol. Hepatol.* 16 (11), 690–704. doi: 10.1038/s41575-019-0209-8
- Wu, Y., Jiao, N., Zhu, R., Zhang, Y., Wu, D., Wang, A. J., et al. (2021). Identification of microbial markers across populations in early detection of colorectal cancer. *Nat. Commun.* 12 (1), 3063. doi: 10.1038/s41467-021-23265-y
- Xuan, C., Shamonki, J. M., Chung, A., Dinome, M. L., Chung, M., Sieling, P. A., et al. (2014). Microbial dysbiosis is associated with human breast cancer. *PLoS One* 9 (1), e83744. doi: 10.1371/journal.pone.0083744
- Yao, Y., Li, X., Xu, B., Luo, L., Guo, Q., Wang, X., et al. (2022). Cholecystectomy promotes colon carcinogenesis by activating the wnt signaling pathway by increasing the deoxycholic acid level. *Cell Commun. Signal* 20 (1), 71. doi: 10.1186/s12964-022-00890-8
- Yao, Z. W., Zhao, B. C., Yang, X., Lei, S. H., Jiang, Y. M., and Liu, K. X. (2021). Relationships of sleep disturbance, intestinal microbiota, and postoperative pain in breast cancer patients: a prospective observational study. *Sleep Breath* 25 (3), 1655–1664. doi: 10.1007/s11325-020-02246-3
- Yoshimoto, S., Loo, T. M., Atarashi, K., Kanda, H., Sato, S., Oyadomari, S., et al. (2013). Obesity-induced gut microbial metabolite promotes liver cancer through senescence secretome. *Nature* 499 (7456), 97–101. doi: 10.1038/nature12347
- Zhang, J., Lu, R., Zhang, Y., Matuszek, Z., Zhang, W., Xia, Y., et al. (2020). tRNA queuosine modification enzyme modulates the growth and microbiome recruitment to breast tumors. *Cancers (Basel)* 12 (3), 628. doi: 10.3390/cancers12030628
- Zhang, J., Xia, Y., and Sun, J. (2021). Breast and gut microbiome in health and cancer. *Genes Dis.* 8 (5), 581–589. doi: 10.1016/j.gendis.2020.08.002
- Zhu, J., Liao, M., Yao, Z., Liang, W., Li, Q., Liu, J., et al. (2018). Breast cancer in postmenopausal women is associated with an altered gut metagenome. *Microbiome* 6 (1), 136. doi: 10.1186/s40168-018-0515-3