



OPEN

# Comparative analysis of racial differences in breast tumor microbiome

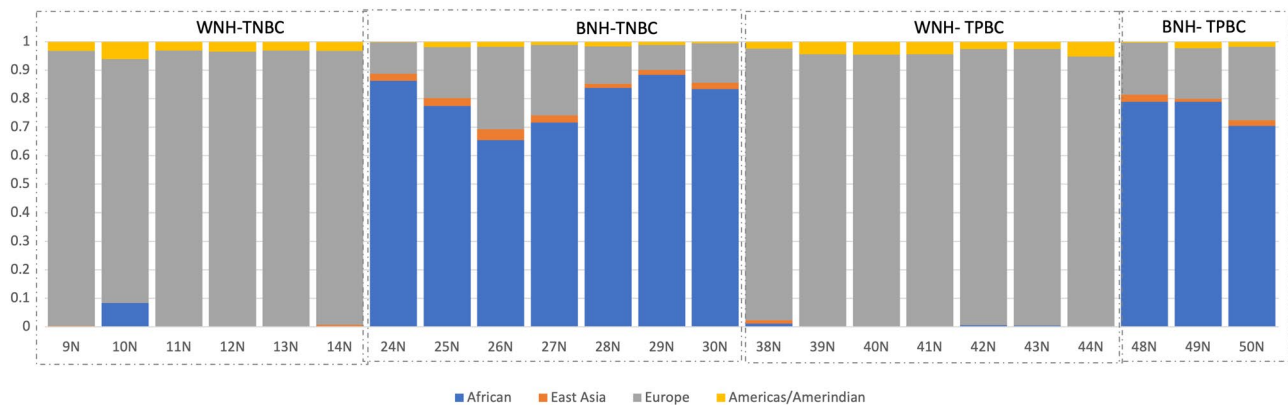
Srikantha Thyagarajan<sup>1,2</sup>, Yan Zhang<sup>1</sup>, Santosh Thapa<sup>3,4</sup>, Michael S. Allen<sup>1</sup>, Nicole Phillips<sup>1</sup>, Pankaj Chaudhary<sup>1,2</sup>, Meghana V. Kashyap<sup>5</sup> & Jamboor K. Vishwanatha<sup>1,2</sup>✉

Studies have demonstrated that environmental, host genetic, and socioeconomic factors influence the breast cancer prevalence landscape with a far-reaching influence on racial disparity to subtypes of breast cancer. To understand whether breast tissue harbors race-specific microbiota, we performed 16S rRNA gene-based sequencing of retrospective tumor and matched normal tissue adjacent to tumor (NAT) samples collected from Black non-Hispanic (BNH) and White non-Hispanic (WNH) women. Analysis of Triple Negative Breast cancer (TNBC) and Triple Positive Breast Cancer (TPBC) tissues for microbiota composition revealed significant differences in relative abundance of specific taxa at both phylum and genus levels between WNH and BNH women cohorts. Our main findings are that microbial diversity as measured by Shannon index was significantly lower in BNH TNBC tumor tissue as compared to matched NAT zone. In contrast, the WNH cohort had an inverse pattern for the Shannon index, when TNBC tumor tissue was compared to the matched NAT. Unweighted Principle Coordinates Analysis (PCoA) revealed a distinct clustering of tumor and NAT microbiota in both BNH and WNH cohorts.

TNBC, a breast cancer disparity with an especially aggressive subtype, is more prevalent in Black non-Hispanic (BNH) than in White non-Hispanic (WNH) women<sup>1,2</sup>. Breast cancer is highly heterogeneous with multiple subtypes where each subtype initiates and develops as a result of defective functioning of regulatory and signaling pathways<sup>3,4</sup>. BNH women manifest breast cancer disparity as higher mortality rate, poorer prognosis due to higher relapse, and higher incidence rates under age 45 than do WNH<sup>5,6</sup>. BNH women are approximately twice as likely to be diagnosed with highly aggressive estrogen receptor negative (ER-), estrogen and progesterone receptor negative and Her2 positive, or TNBC subtypes of basal-type breast cancer<sup>7-9</sup>. Several studies have shown that the paired normal tissue adjacent to tumor (NAT) as compared to tumor tissue did not differ in microbiota richness as reflected by the number of operational taxonomic units (OTUs), but differed in evenness as reflected by the abundance of specific species of bacteria in normal vs tumor zone<sup>10-12</sup>. Treatment of rats with *Lactobacillus casei* has been shown to reduce adjuvant-induced arthritis and increased abundance of other *Lactobacillus* species<sup>13</sup>. These findings are of interest as the antiproliferative activity of *Lactobacillus crispatus* and *Lactobacillus acidophilus* was found to be associated with health status of breast cancer cells<sup>14</sup>. The beneficial influence of breast feeding and breast milk in establishing the infant gut microbiota and health during adult life has been well documented in many studies<sup>15,16</sup>. A limited number of recent studies also demonstrated an association between compositional differences of microbial species between healthy and breast tumor tissues including unique bacterial microbiome signatures between different subtypes of breast cancer<sup>12,17,18</sup>.

Association of specific microbiome profiles with somatic genome SNP from colorectal tumor cells and matched normal tissue<sup>19</sup> suggest a role of host and tumor genetic backgrounds in microbiota structure. Among biological factors, epigenome-driven microbiota profile has been implicated in the etiology of several kinds of cancer including the association with breast cancer<sup>20</sup>. Several studies on gut and oral microbiota have demonstrated that association of distinct microbiota in different geographical and regional differences in human population cohorts adopting different lifestyle/occupation or socioeconomic status (SES)<sup>21,22</sup>. Bioinformatic

<sup>1</sup>Department of Microbiology, Immunology, and Genetics, University of North Texas Health Science Center, Fort Worth, TX 76107, USA. <sup>2</sup>Texas Center for Health Disparities, University of North Texas Health Science Center, Fort Worth, TX 76107, USA. <sup>3</sup>Department of Pathology, Texas Children's Microbiome Center, Texas Children's Hospital, Houston, TX 77030, USA. <sup>4</sup>Department of Pathology and Immunology, Baylor College of Medicine, Houston, TX, USA. <sup>5</sup>University of Nebraska Medical Center, Omaha, NE, USA. ✉email: jamboor.vishwanatha@unthsc.edu



**Figure 1.** Ancestry Analysis. The genetic ancestry of both tumor and normal tissue samples were analyzed using an Illumina Infinium Global Screening Array, and the data were analyzed using the methods as described in the text. The decoded Matched Normal samples are labelled as symbol N following the sample number in the study. WNH and BNH represents White non-Hispanic and Black non-Hispanic racial cohorts respectively.

analysis of published gut microbiota datasets showed association of distinct bacterial species among Caucasian, African American, Pacific islanders and Hispanic racial/ethnic groups<sup>23</sup>. A recent study demonstrated that racial differences existed in breast microbiota richness and specific bacterial taxa between WNH vs BNH women breast tumor tissue and adjacent normal tissue, which were derived from a biorepository representing the population cohorts encompassing the Black Belt Region of the southeastern U.S.<sup>24</sup>. In this study, we refer Caucasian as White non-Hispanic (WNH) and African American as Black non-Hispanic (BNH).

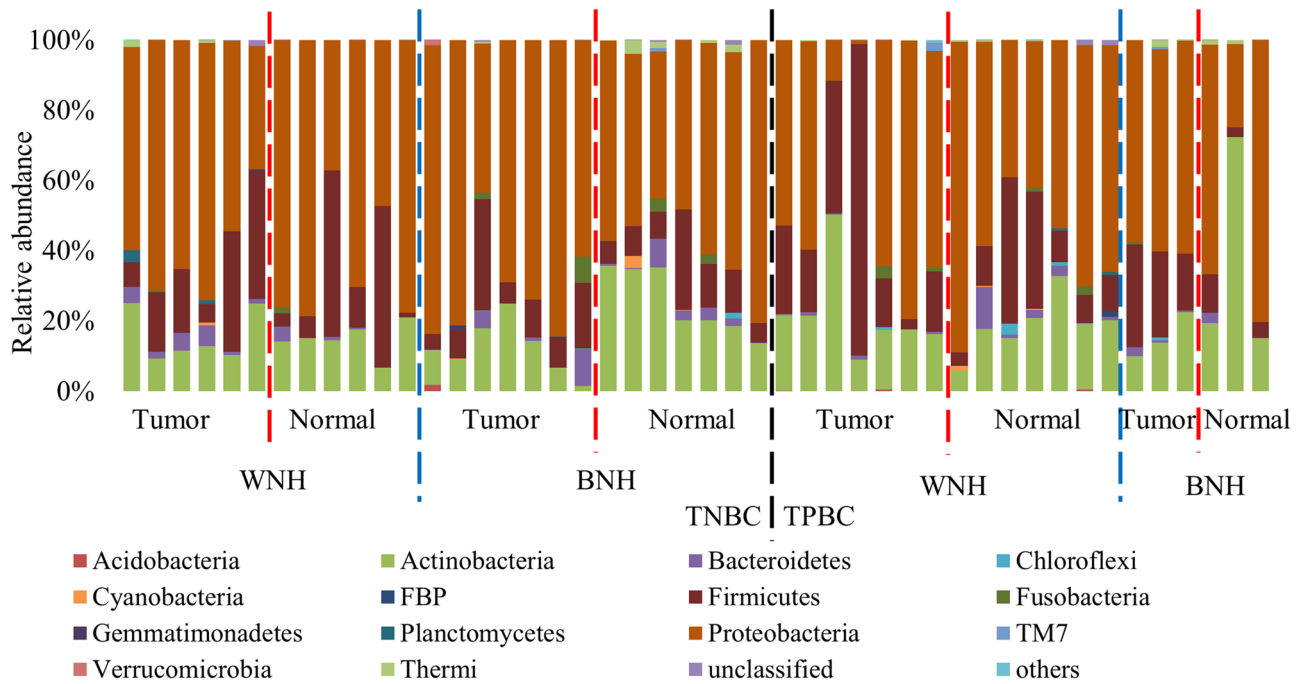
Therefore, the intent of this study was two-fold: first, to address whether racial differences in microbiota are associated with archived retrospective breast tissue of tumor and adjacent normal tissue among BNH and WNH women population cohorts representing New England, USA; second, to determine if racial ancestry is strongly associated with microbiome profiles irrespective of cancer subtypes. This information will lay the crucial foundation to understand the extent of influence of geography, race, and ancestry/genetic background on microbiota profiles and increased incidence of aggressive TNBC subtypes within African American (BNH) populations.

## Results

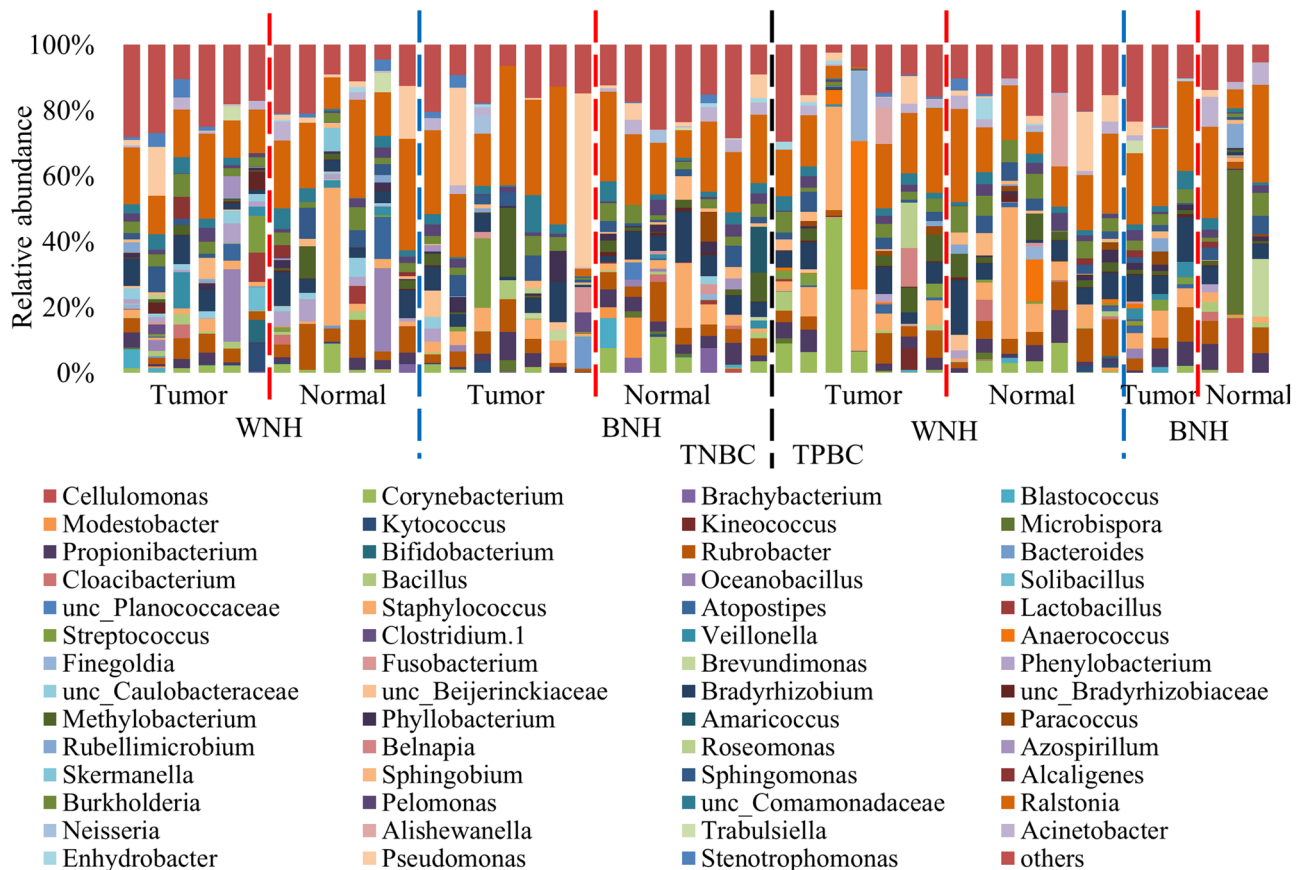
**Breast tissue sample collection traits and specimen category.** The deidentified frozen breast tissue samples were collected retrospectively from the biorepository of the Fox Chase Cancer Center, Philadelphia, PA (Table S1). All the breast cancer patients were residing in the New England region of USA. The age of the women at the time of breast tissue collection ranged between 27–78 years. The women were not subjected to antibiotic regimen prior to the collection of tissue samples. In addition, none of the patients had undergone radiation treatment. For the first set of microbiome analysis, we selected tumor and matched NAT. For the analysis, 6 TNBC WNH and 7 TNBC BNH, 7 TPBC WNH and 3 TPBC BNH were chosen. To verify the microbial diversity results, we analyzed a second set of 10 WNH TNBC samples of tumor and matched NAT.

**Ancestry of breast cancer subjects.** All tissue samples yielded reproducible ancestral assessment in individuals of primarily one ancestral background, confirming genetic ancestry concordance with self-described race/ethnicity (Fig. 1). As expected, admixture was detected in the BNH individuals, with the largest ancestral component deriving from African descent. WNH individuals showed a higher degree of homogenous ancestry, with the primary ancestral component indicated being European.

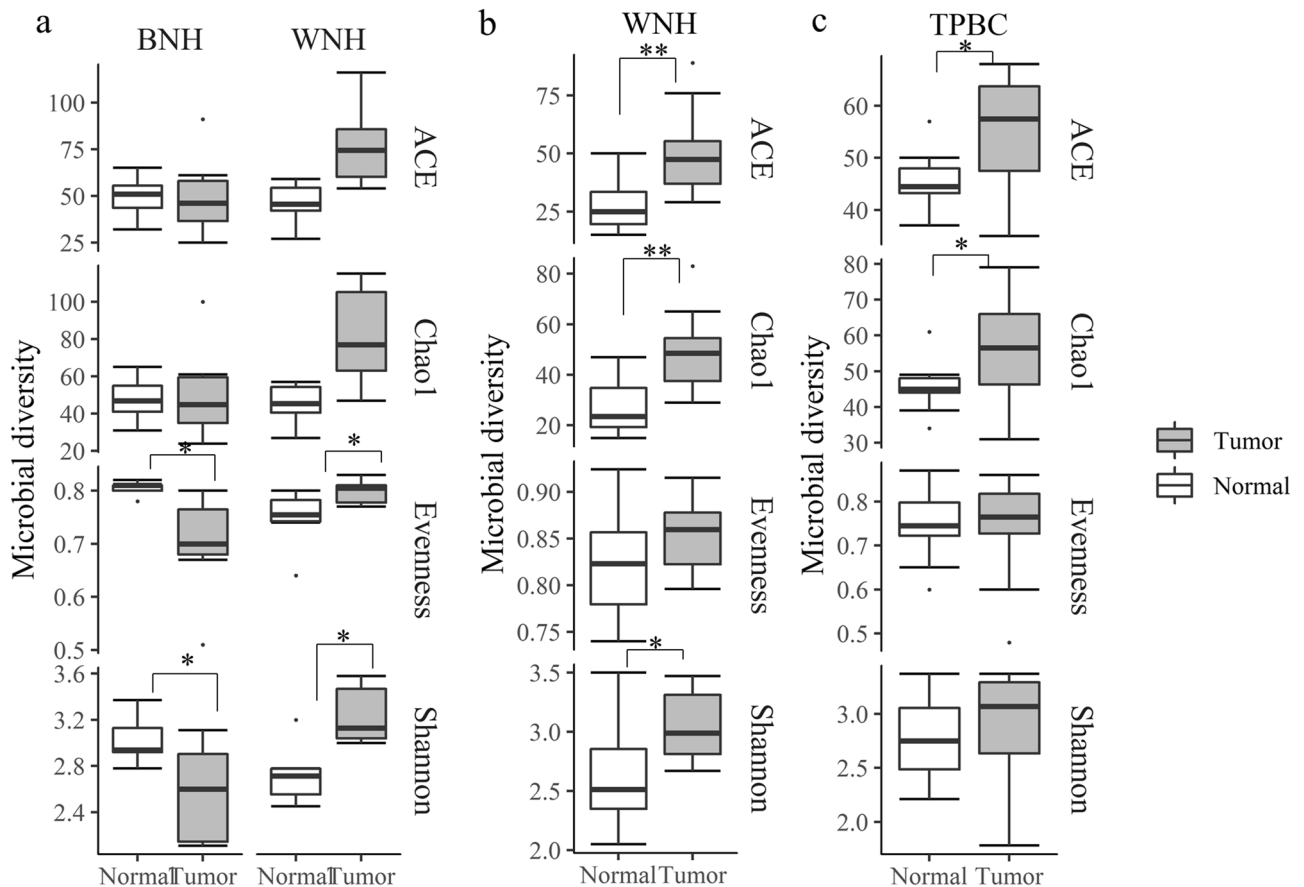
**Microbiota composition in breast tissue of cancer patients.** Comparison of the microbiota in tumor and adjacent normal breast tissues from 6 (excluding two admixed outliers based on ancestry analysis) WNH, 7 BNH TNBC patients, as well as 7 WNH, 3 BNH TPBC patients showed a total of 20 bacterial phyla and 419 genera were found in the breast tissue of cancer patients (Table S2). Microbial abundance at different phylogenetic levels of the above 46 samples (a set) and 20 samples from the second set (b set) are shown in Table S2. Across all the breast tissue samples, estimation of relative abundance revealed that the most dominant bacterial population was Proteobacteria (59.4% ± 18.6%), followed by Actinobacteria (19.1% ± 12.0%), Firmicutes (17.7% ± 16.5%), and Bacteroidetes (1.9% ± 2.8%) (Fig. 2). The top 5 most abundant bacterial genera were *Ralstonia*, *Staphylococcus*, unclassified Bradyrhizobiaceae, *Rubrobacter* and *Pseudomonas*. The bacterium *Ralstonia* (19.1% ± 9.0%) was the most dominant bacterial genus prevalent in all the tumor and normal tissues (Fig. 3). The bacterial genus *Staphylococcus* (6.4% ± 9.4%) was the second dominant bacterium detected in 21 out of 23 NAT and 22 out of 23 tumor breast tissues (Fig. 3). The bacterial genera unclassified Bradyrhizobiaceae (5.5% ± 4.0%) and *Rubrobacter* (5.4% ± 3.3%) were also observed in all the breast tissues including tumor and adjacent normal zone (Fig. 3). *Pseudomonas* (4.1% ± 9.5%) was prevalent in 17 out of 23 normal and 17 out of 23 tumor breast tissues (Fig. 3).



**Figure 2.** Normalized bar plot of bacterial community composition in tumor and normal tissues at the phylum level. Phyla with >1% relative abundance are shown. TNBC, triple negative breast cancer; TPBC, triple positive breast cancer; BNH, Black non-Hispanic; WNH, White non-Hispanic.



**Figure 3.** Normalized bar plot of bacterial community composition in tumor and normal tissues at the genus level. Genera with >1% relative abundance are shown. TNBC, triple negative breast cancer; TPBC, triple positive breast cancer; BNH, Black non-Hispanic; WNH, White non-Hispanic.



**Figure 4.** Bacterial diversity in tumor and adjacent normal tissue of (a) TNBC from WNH ( $n=6$ ) and BNH ( $n=7$ ); (b) TNBC WNH 2nd set ( $n=10$ ) and (c) TPBC pool ( $n=10$ ) cancer patient samples. Dots represent outliers; \* $p < 0.05$ ; \*\* $p < 0.01$ , Wilcoxon Signed Rank Test.

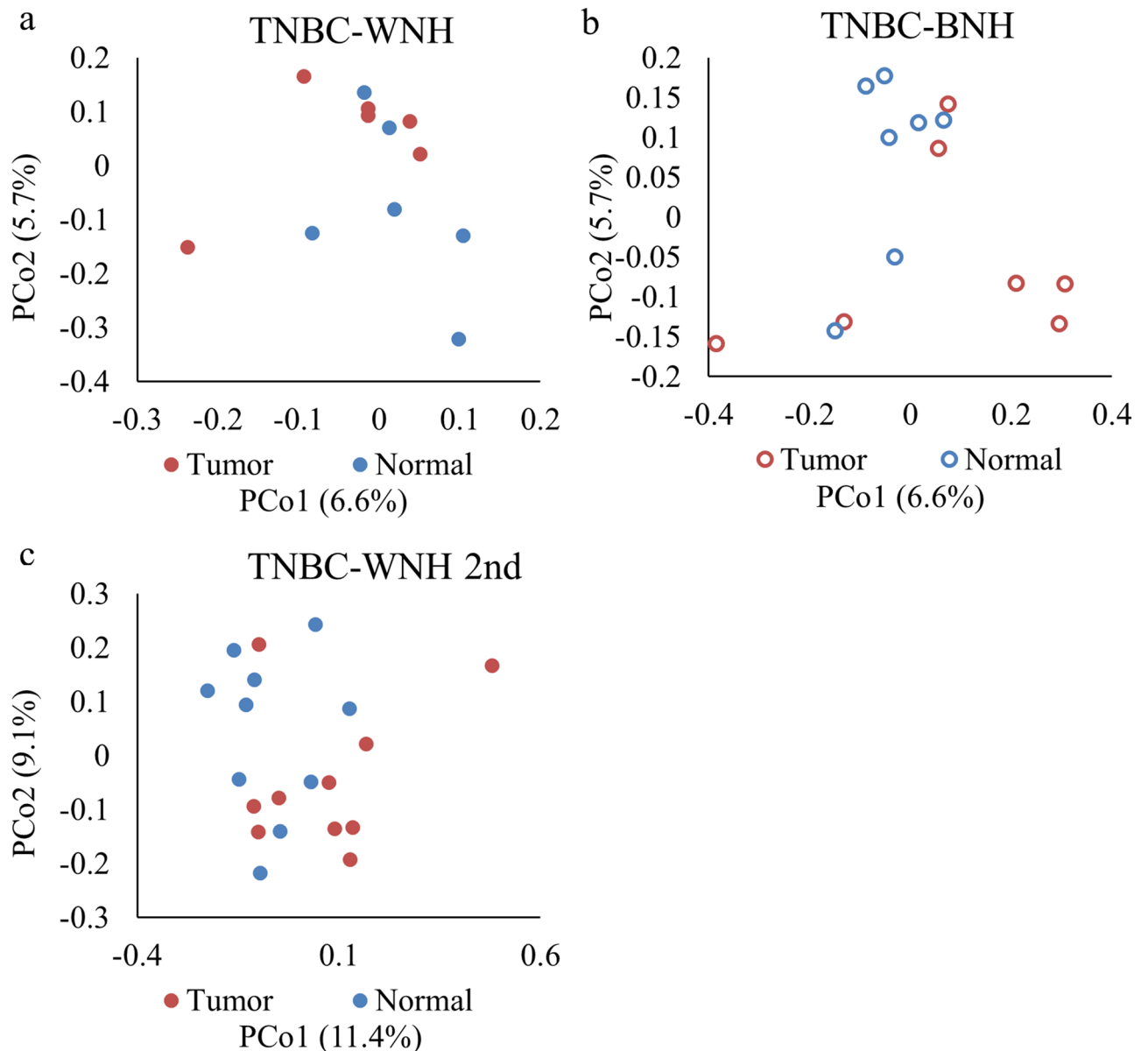
#### Microbiota diversity and community composition in tumor tissues was different from normal tissues.

The microbial diversity differences between tumor and normal tissue showed different patterns in WNH and BNH populations. For the BNH population with TNBC, the Shannon diversity ( $p=0.05$ ) and evenness ( $p=0.04$ ) in tumor tissues were significantly lower than the NAT tissue (Fig. 4a). In contrast, for the WNH population with TNBC, the Shannon diversity ( $p=0.05$ ) and evenness ( $p=0.04$ ) of tumor tissue were significantly higher than the NAT. TNBC WNH population also had higher richness (ACE  $p=0.06$ ; Chao1  $p=0.06$ ) in tumor than in normal tissue, although not statistically significant (likely due to sample size and sample heterogeneity). In order to confirm the results of the reversed pattern of alpha diversity metrics, we sequenced an additional 10 breast tumor tissues from WNH TNBC tissues along with NAT. The results revealed that microbial diversity indices showed the same pattern indicating higher Shannon diversity ( $p=0.04$ ) and richness (ACE  $p=0.004$ ; Chao1  $p=0.006$ ) in WNH tumor samples (Fig. 4b).

In TPBC patients, tumor tissue showed significantly higher alpha diversity metrics (ACE ( $p=0.04$ ) and Chao1 ( $p=0.05$ ) richness) than normal tissue (Fig. 4c). The Shannon diversity and evenness between tumor and NAT did not show significant differences, possibly due to high inter-individual heterogeneity.

Differences in microbial community structure between normal and tumor tissue were observed in TNBC patients using principal coordinate analysis (PCoA) ordination of unweighted UniFrac distances (Fig. 5). The unweighted PCoA plots showed different microbial communities between tumor and normal tissue in TNBC patients for both WNH (AMOVA  $p=0.02$ ) and BNH (AMOVA  $p=0.07$ ) groups (Fig. 5a and b). The PCoA based on weighted UniFrac distances did not reveal microbial community difference between tumor and normal tissue in TNBC patients. In the second set of TNBC WNH patients, the unweighted UniFrac PCoA analysis revealed some degree of distinct clustering of microbiota in tumor and NAT (AMOVA  $p=0.06$ ) (Fig. 5c).

To determine microbial differences between tumor and the normal tissue, we performed Wilcoxon Signed Rank Test with abundant bacterial phyla (>1%) and genera (>5%). The Wilcoxon Signed Rank test for microbiota composition revealed that in TNBC WNH patients, the phylum Bacteroidetes in tumor tissue was significantly higher than in NAT tissue ( $p=0.03$ ) (Fig. 6a). For TNBC BNH patients, the phylum Actinobacteria ( $p=0.03$ ) and unclassified genus of unclassified Bradyrhizobiaceae ( $p=0.03$ ) in tumor tissue were significantly lower than in normal tissue (Fig. 6b, c). For TPBC patients, at phylum level *Fusobacteria* ( $p=0.01$ ) showed higher abundance in the tumor tissue as compared to normal tissue. At genus level, *Streptococcus* ( $p=0.03$ ) showed the same trend (Fig. 6d, e).

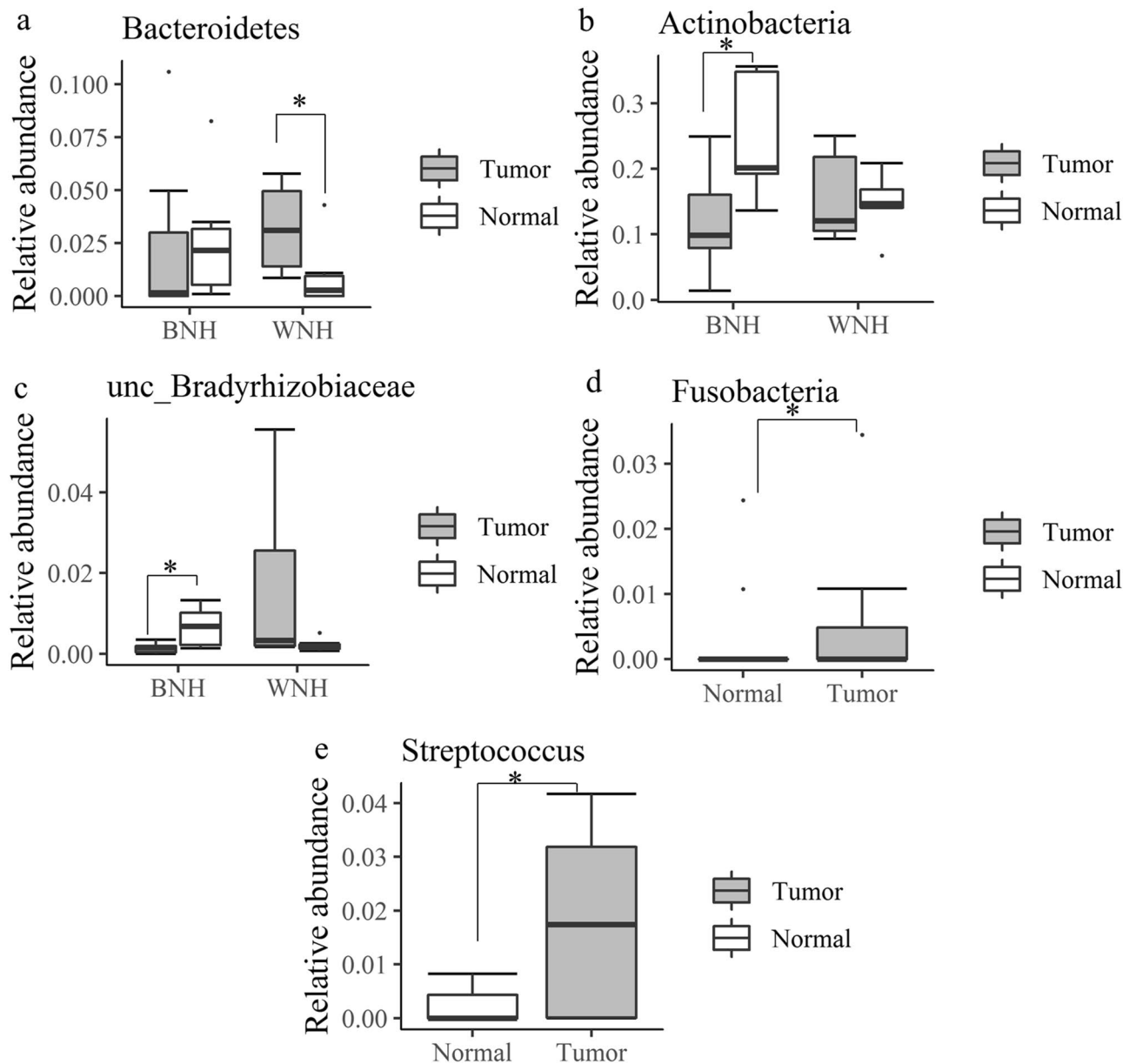


**Figure 5.** Principal coordinate analysis (PCoA) based on unweighted UniFrac distances showing clustering of tumor vs normal tissues for TNBC patients. TNBC-WNH (n=6); TNBC-BNH (n=7); TNBC-WNH 2nd (n=10) refers to second set of samples.

**WNH and BNH TNBC patients showed different microbiota diversity and community composition.** The relative abundance of Bacteroidetes, showed a variable pattern in tumor vs normal breast tissue in both WNH and BNH groups. There was inter-individual heterogeneity in relative abundance of Bacteroidetes in both tumor and normal tissue within BNH samples (Fig. 6a). In contrast, the WNH group showed significantly lower variation and relative abundance in normal tissue as compared to tumor tissue (Fig. 6a).

We found microbiota compositional differences between WNH and BNH racial groups. The matched normal tissue of TNBC in BNH cohort showed a significantly higher Shannon diversity ( $p=0.03$ ) and evenness ( $p=0.01$ ) than the WNH (Fig. 7a), but very similar microbial richness as measured by both ACE and Chao1 metrics. In contrast, in TNBC, WNH showed significantly higher microbial diversity (Shannon  $p=0.008$ , Chao1  $p=0.05$ ) and evenness ( $p=0.01$ ) than the BNH (Fig. 7a). PCoA based on unweighted UniFrac distances showed that TNBC WNH and BNH had significantly different microbial communities for both tumor (AMOVA  $p=0.05$ ) and normal tissues (AMOVA  $p=0.007$ ) (Fig. 7b, c).

To determine the microbial population differences between BNH and WNH samples, we performed a Wilcoxon Rank Sum Test with selected abundant bacterial phyla (>1%) and genera (>5%). At the genus level, several genera showed significant differences between BNH and WNH populations in both tumor and normal tissue. TNBC WNH patients had higher abundance of *Phenylobacterium* than BNH in both normal ( $p=0.003$ ) and tumor tissues ( $p=0.02$ ) (Fig. 8a). TNBC WNH patients had higher abundance of *Lactobacillus* ( $p=0.02$ ), in two

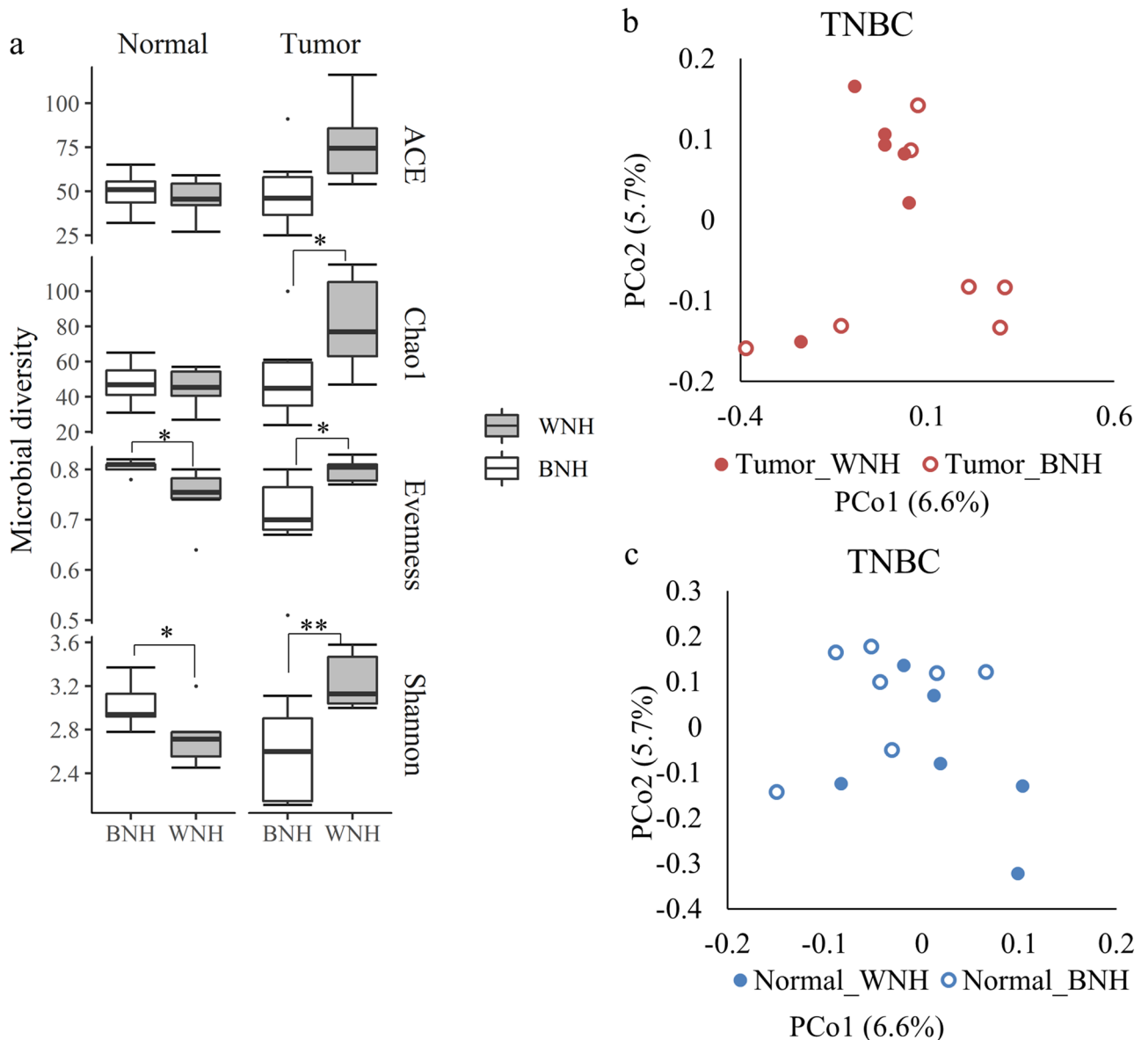


**Figure 6.** Microbiota relative abundance at phylum (a, b, d) and genus (c, e) levels showed differences between tumor and normal tissue in both TNBC and TPBC patients using Wilcoxon Signed Rank Test. (a–c) TNBC WNH (n = 6) and TNBC BNH (n = 7); (d, e) TPBC (n = 10). Dots indicate outliers; \*  $p < 0.05$ .

unclassified genera within the families Caulobacteraceae ( $p = 0.02$ ) and Bradyrhizobiaceae ( $p = 0.03$ ) than BNH in tumor tissue. The normal tissue from TNBC BNH patients had a higher abundance of unclassified genus of Planococcaceae than WNH ( $p = 0.04$ ) normal tissue, but not in normal tissue. At the phylum level, Thermi showed higher relative abundance in normal tissue and lower relative abundance in tumor tissue of TNBC BNH patients compared with WNH patients ( $p = 0.02$ ) (Fig. 8c). For TPBC patients, no significant taxon difference between WNH and BNH was detected in normal tissue. However, TPBC tumor tissue showed higher abundance of *Veillonella* ( $p = 0.01$ ) and *Phyllobacterium* ( $p = 0.02$ ) in BNH than WNH tissues (Fig. 8b). No statistically significant microbial diversity or community differences were detected between WNH and BNH with TPBC normal tissue.

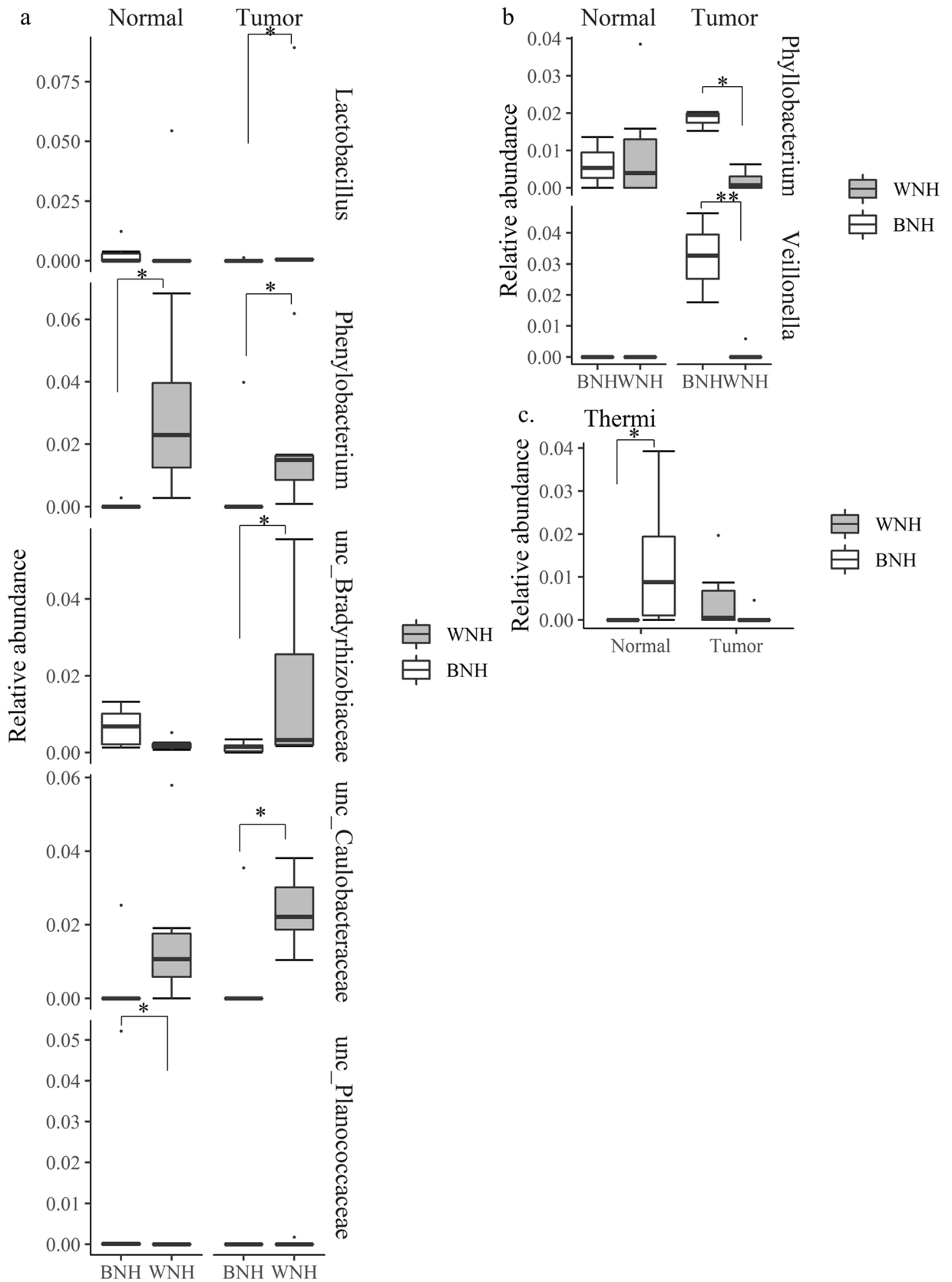
## Discussion

A significant health disparity exists among racial/ethnic populations in the US with African American ancestry in rates of TNBC<sup>1,2</sup>. Several studies have implicated a multitude of unique mechanisms such as apoptosis<sup>26</sup>, race-specific tumor environment<sup>27</sup>, somatic mutation landscape<sup>28</sup> and mitochondrial function<sup>29</sup> as mediators of health disparities and influenced by low socioeconomic status, psychosocial stress, life style and genetic ancestry. However, these mechanisms do not fully account for TNBC breast cancer disparity in African American populations. One of the biological factors suspected to play role in this health disparity is the microbiome<sup>30</sup>.



**Figure 7.** Comparison of microbiota of BNH and WNH in TNBC cancer patients. (a) bacterial community diversity and evenness; (b, c) Principal coordinate analysis (PCoA) based on unweighted UniFrac distances of (b) tumor and (c) normal adjacent tissues from WNH (n=6) and BNH (n=7) patient samples. Dots indicate outliers; \* $p < 0.05$ ; \*\* $p < 0.01$  determined by Wilcoxon Rank Sum Test.

The analysis of microbiota composition showed a significantly lower microbiota evenness and Shannon index in TNBC BNH tumor as compared to adjacent normal tissue, as well as alterations in the proportions of specific taxa. In contrast, WNH tumor showed the opposite scenario, with a higher evenness/Shannon index in tumor tissue compared to normal tissue. Smith et al. observed a higher microbial diversity and richness in the adjacent normal pair than in the normal, and tumor tissue<sup>24</sup>. However they did not perform paired analyses between tumor tissue and the paired adjacent normal tissue. In addition, their data include both WNH and BNH. The majority of their adjacent normal tissue were from WNH with 2 samples from BNH. To confirm our results we sequenced a second set of WNH samples and performed a paired comparison between the tumor and adjacent normal tissue. The microbiome analysis of the second set of WNH breast tumor/matched normal tissue showed a similar pattern to what we observed from the first set. The reason for reverse pattern in diversity could be due to the compositional differences of beneficial vs pathogenic species as demonstrated in a previous study between native African and migrated African American populations on microbiota profiles with beneficial vs CRC carcinogenic inducing metabolites<sup>31</sup>. This unique finding may be due to differential abundance of specific microbial taxa among ethnic groups, which in turn may influence TNBC onset/prevalence, prognosis and survival differences between BNH and WNH groups. Interestingly, TPBC microbiome analysis showed an increase in microbial richness component irrespective of race, although the evenness component remained very similar between tumor and adjacent normal tissue. The TPBC microbiota data also indicate that microbial metabolism



**Figure 8.** Relative abundance of taxa showed differences between BNH and WNH at the genus (**a**, **b**) and phylum (**c**) levels in TNBC (**a**, **c**) and TPBC (**b**) patients. BNH (n=7), WNH (n=6) Dots are outliers. \* $p < 0.05$ , \*\* $p < 0.01$ , Wilcoxon Rank Sum Test.



by estrogenic microbes may trigger tumorigenesis<sup>32</sup>. The breast tumor microbiota studies to date<sup>33,34</sup> have shown that the abundance ratio among specific microbial members could be detected among ethnicities. For example, Canadian and Irish derived breast tissues showed distinct microbiota signatures<sup>33</sup>. In another study a change in evenness alone, which is usually associated with a change in abundance of specific microbial members like *Methylobacterium* and *Sphingomonas* was associated with breast tumor state<sup>10</sup>.

The analysis of microbial composition revealed that the highly abundant phylum Actinobacteria showed lower abundance in tumor in comparison to normal BNH of TNBC. The lower abundant phylum Bacteroidetes showed similar abundance in both tumor and normal tissues of TNBC BNH, but higher abundance in WNH tumor associated tissue and with high interindividual variability. Bacteroidetes represents one of the key anaerobic commensals in the gut with flexibility to become pathogenic under unfavorable conditions in the host<sup>35</sup>. The Firmicutes to Bacteroidetes (F/B) ratio is regarded to be of importance in human gut microbiota composition and increase in the ratio has been directly correlated to obesity<sup>36,37</sup>. However, the similar F/B ratio in both BNH and WNH tumors and the wide difference in F/B ratio between BNH and WNH normal tissues observed in this study will require further study to fully understand the ramifications in breast cancer and health disparities. The phylum Actinobacteria with significantly lower abundance in BNH tumors, suggests alterations in evenness. This may be due to possible changes in ratios between classic beneficial microbes, *Bifidobacterium* and *Streptomyces*, and putatively pathogenic members such as *Propionibacterium*, *Corynebacterium*, etc.<sup>38</sup>. A recent meta-analysis of the published breast cancer microbiome databases using One Codex revealed the prevalence of significant evenness differences between the majority of microbial groups including Bacteroidetes and Actinobacteria rather than richness differences between normal and disease states<sup>34</sup>. Future studies should focus on both high resolution analysis using shot-gun metagenomics to identify microbial phenotypes and host genetic signatures to develop effective personalized therapeutic strategies. A large scale in silico analysis of gut microbiome data sets has shown the existence of microbiota profiles unique to four ethnic groups that included both BNH and WNH<sup>23</sup>. In that study, BNH showed the lowest gut microbiota evenness and richness as measured by Shannon index. The extent to which this impacts the breast microbiome in tumor and healthy tissues remains to be elucidated. In another recent study on the breast microbiome, *Ralstonia* was shown to be abundant in BNH tumors, while the family Xanthomonadaceae was abundant in WNH tumors<sup>24</sup>. In contrast, our study showed that *Ralstonia* was the dominant genus with no significant differences in abundance between WNH vs BNH or tumor vs normal tissues. One genus of unclassified Bradyrhizobiaceae showed a significant reduction in abundance and very low interindividual variability in BNH tumor as compared to high interindividual variability and higher abundance in normal tissue. Interestingly, the abundance trend was opposite in tumor and normal tissues in the WNH cohort. Since most studied members of this genus are nitrogen-fixing bacteria of the soil environment, their role in breast health is not yet clear. Interestingly, studies have reported the presence of *Bradyrhizobium* in colitis patients, and in lower respiratory airways<sup>39,40</sup>. The TPBC tumor samples showed a significant increase in abundance of *Fusobacteria* and *Streptococcus* in TPBC samples, which is consistent with earlier work on *Fusobacteria* where it was shown to be enriched in malignant breast and colorectal cancer<sup>12</sup>. The genus *Streptococcus* contains both beneficial and pathogenic species: *S. thermophilus* is a beneficial species shown to be enriched in normal breast tissue and protective against DNA damage by its unique SOD activity<sup>11</sup>. However, at least 12 species have been shown to be pathogenic, of which two well-known species *S. pyogenes* and *S. pneumoniae* have been found to be more abundant in breast tumor tissue than in the normal tissue<sup>10</sup>, and have also been found in human milk<sup>42</sup>.

One of the main objectives of this study was to identify if there are any race-specific microbes that could provide evidence for microbiome-host interactions in breast cancer pathogenesis as possible causes of health disparities, potentially leading to the development of more appropriate treatment options and possible preventive measures. The significance of higher abundance of the genus *Phenyllobacterium* in both WNH normal and tumor tissues is not clear. A recent study has shown that the higher abundance of this genus was observed in individuals with heavy alcohol drinkers with no sign of liver cirrhosis in comparison to non-alcohol drinkers<sup>43</sup>, suggesting a possible protective effect. A facultative bacteria, *Phenyllobacterium zucineum* had been identified from human erythroleukemia cell line<sup>44</sup>, although its stable association with cell line is unclear. The significance of higher abundance of the genus of unclassified Bradyrhizobiaceae identified in BNH normal as compared to tumor is not clear. In addition, the reversal of abundance levels in WNH NAT vs tumor cohort need further studies. These observations suggest a possible role of host genetics and microbiota in the etiology of cancer. Members of the genus *Bradyrhizobium* are also one of the more commonly found members in human milk, and in select human tissues, in addition to soil<sup>42</sup>. However, that study did not describe the racial identity of the sample. Our study showed the abundance of the unclassified Caulobactereaceae is very low in both the NAT and the tumor tissue of BNH TNBC cohort. In contrast, both tumor and normal tissue of WNH cohort showed a higher abundance of this bacterial signature. The importance of the low abundance of this taxon in BNH TNBC racial breast cancer disparity is currently unclear. An earlier study on breast microbiome using Pathochip showed that the genus *Caulobacter* signature in tumor tissues of three subclasses of breast cancer including TNBC, and TPBC<sup>15</sup>. The genus *Phyllobacterium* showed a significantly higher abundance of approximately fourfold in BNH TPBC tumor tissue as compared to TPBC WNH tumor tissue. On the other hand, the abundance in WNH normal tissue exhibited very high interindividual variability, while in tumor its abundance was lower overall between the samples, suggesting that this genus could have a mechanistic link with BNH TPBC tumorigenesis, but an unclear role in WNH tumorigenesis. The genus *Phyllobacterium* has also been shown to increase in abundance in gastric adenocarcinoma and in sub-populations of patient samples with gut necrotizing enterocolitis, although a direct functional role has not been established<sup>45,46</sup>. The genus *Veillonella*, one of the well-known core members of the human breast milk microbiota, exhibited medium abundance in TPBC BNH tumor tissue as compared to normal, or to tumor tissue of the WNH group. However, a high interindividual heterogeneity was also evident among BNH TPBC group. Interestingly, a recent study showed that distribution of this genus showed a higher abundance in infant fecal samples from African countries (30 to 50%) as compared to infant fecal samples from

the US (20–30%)<sup>47</sup>. *Veillonella*<sup>48</sup> has also been found to be one of the earliest saccharolytic colonizers of the developing infant as a minor core member (5%). It is also known to interact with other core microbial members including *Streptococcus* and *Bifidobacterium* to produce the key anti-inflammatory short chain fatty acid propionate, and is known to be one of the main members of healthy oral and gut microbiota<sup>48</sup>.

One of the limitations of this study is that only a limited number of TPBC-BNH samples were used. The significantly higher abundance in BNH TPBC as compared to WNH TPBC observed here should be verified with a larger scale sampling. The significantly higher abundance of the phylum Thermi was accompanied with higher interindividual variability in NAT as compared to tumor tissue of BNH TNBC cohort. In contrast, WNH TNBC showed an opposite abundance profile, with low in normal and high in tumor tissue. In another published study, Thermi was observed in lung adenocarcinoma patients with increased abundance in late stages of tumor<sup>49</sup>. Considering the environmental resistance traits of some members of the phylum, it is possible that this microbe acts to protect the host from insults, while the relative changes in abundance over threshold levels may indicate unfavorable host response to the disease state. On the other hand, the role played by the microbe may be race-specific as the WNH TNBC exhibits an opposite trend shown for lung tumors<sup>49</sup>.

A thorough understanding of the breast microbiome as impacted by host genetics, life-style, socioeconomic factors, and geography will be necessary to identify both the beneficial and harmful microbiota that may influence breast cancer development. Deeper understanding of these factors may lead to development of microbiome-based therapeutic strategy for breast cancer treatment.

## Materials and methods

**Study population and sample collection.** The frozen breast tissue samples collected from the biorepository were collected and processed in accordance with the UNTHSC IRB protocol (#2009–001). The deidentified, frozen breast tissue samples (n=50) were acquired from the biorepository at Fox Chase Cancer Center, Philadelphia, PA (additional information is described in results section). For the first set of microbiome analysis, we included 8 White non-Hispanic, of which 2 patient samples (n=4; n=2 tumor and n=2 NAT) were excluded from final analysis due to admixed ancestry, 7 Black non-Hispanic women with a triple negative breast cancer (TNBC), 7 White non-Hispanic triple positive breast cancer (TPBC), and 3 Black non-Hispanic women with TPBC. For the second set of microbiome analyses, we included, 10 additional White non-Hispanic WNH tumor and matched Normal Adjacent to Tumor (NAT) tissues (described in results section). The matched NAT from the same individuals were considered as normal. The patient's information on race, ethnicity, age at diagnosis, pathology report on stage of tumor as T1, T2 and T3 were collected on the samples (Table S1). All the tissue samples were derived from women subjects diagnosed with breast cancer. The biorepository information revealed that patients were not treated with antibiotics prior to the tissue collection although duration window of non-exposure to antibiotics prior to surgery was not included in their data (Table S1). The specimens are classified as BNH and WNH according to the self-reported data of patients from their respective geographical areas.

**Ancestry analysis of breast tissue.** The genomic DNA isolated for microbiome analysis was used for genome-wide SNP profiling to verify the racial identity of patient samples (Fig. 1). Admixture analysis for ancestry assessment was performed on genomic SNP data obtained from the Infinium Global Screening Array (Illumina). Data were analyzed by the method described previously<sup>50</sup>, and were pruned based on linkage disequilibrium. SNP genotypes downloaded from the Human Genome Diversity Project for the European, African, Americas/Amerindian and Asian super populations were merged with the sample dataset. Ancestry coefficients for the four parent ancestral populations (k=4) were estimated for based on LD-pruned nDNA SNPs using an unsupervised model-based approach as implemented in ADMIXTURE<sup>51</sup>.

**16S rRNA library preparation and Illumina MiSeq sequencing.** Generation of the 16S rDNA library, sequencing, raw read calling/processing, were performed by Zymo Research through their ZymoBIOMICS service (Zymo Research Corp. Irvine, CA). ZymoBIOMICS Microbial Community Standards (Zymo D6300) were used as positive controls. The details of the methods used is as follows: Genomic DNA was extracted from frozen breast tissue using the ZymoBIOMICS 96-bead DNA Extraction Kit (Zymo Research Corp.). 16S rRNA gene amplicon library was prepared using Quick-16 NGS Library Prep Kit (Zymo Research Corp.). Briefly, 16S rRNA gene was amplified with Illumina MiSeq compatible Quick-16 Primer Set 341F (CCTACGGGNGGC WGCAG) and 785R (GACTACHVGGGTATCTAATCC) targeting the V3-V4 hypervariable region. The PCR master mix containing: 10 µl of Quick-16S qPCR Premix, 4 µl of Quick-16S Primer Set V3-V4, 4 µl of ZymoBIOMICS DNase/RNase Free Water, and 2 µl DNA template. Both positive and negative controls for PCR quality assessment were included during amplification. The majority of samples were amplified for 18 cycles except low microbial DNA concentration samples were amplified for an additional 4 cycles. The thermocycler conditions were: 95 °C for 10 min, 20 cycles of 95 °C for 30 s, 55 °C for 30 s, 72 °C for 3 min and hold temperature, of 4 °C. Following amplification, estimation of DNA quality and concentration, followed by enzymatic clean-up was performed and samples were stored at 4 °C prior to sequencing.

Specific barcodes were added for each sample through index PCR amplification. The 20 µl PCR reaction recipe composed of 10 µl Quick-16S qPCR Premix, 4 µl ZymoBIOMICS DNase/RNase Free Water, 2 µl each of the Index Primer set A, and 2 µl purified DNA. The PCR condition was: 95 °C for 10 min, 5 cycles of 95 °C for 30 s, 55 °C for 30 s, 72 °C for 3 min followed by holding temperature at 4 °C. The amplified library was quantified and stored at 4 °C. A fluorescence standard curve was created by running in parallel real-time PCR reaction with the same thermocycler program as above using 20 µl with fluorescence standard, in triplicate configuration. Equal volumes of bar-coded samples were pooled prior to purification of the library. The pooled library was purified

using Select-a-Size Zymo MagBeads (Zymo Research Corp.) following manufacturer's specifications. The purified index bar-coded library was then denatured, diluted, and sequenced on Illumina MiSeq.

**Bioinformatics analysis.** The raw sequence data were analyzed using mothur (v.1.36.1) program following the MiSeq SOP as previously described<sup>52</sup>. Briefly, contigs were assembled from paired-end raw reads sequences with removal of ambiguous sequences from the dataset. Sequences were dereplicated and aligned against the SILVA 16S rRNA database (v132). Gaps were deleted from the aligned contigs following deletion of unaligned sequence reads with database reference sequences. The sequences were further dereplicated using Unique seqs and Precluster Commands and chimeras were identified and removed using UCHIME algorithm<sup>53</sup>. Operational taxonomic units (OTUs) were grouped based on 97% sequence similarity using the average neighbor clustering algorithm<sup>54</sup>. Taxonomic assignment was determined by comparing with the Green Genes database using RDP classifier with a minimum of 80% confidence<sup>55</sup>. Non-bacterial sequences including mitochondria, chloroplast, archaea and eukaryote, as well as undefinable sequences were excluded from the data set. Microbial community diversity (Shannon diversity), evenness and richness (Chao1 and abundance coverage-based estimator (ACE) were calculated at 97% cut-off after rarefied to the lowest sampling depths of 1,238 reads per sample<sup>56–58</sup>. Principle Coordinate Analysis (PCoA) based on Uni-Frac distances were performed to compare the microbial community differences among different samples<sup>59</sup>. The Molecular Variance analysis (AMOVA) was performed to assess the variability among and within different groups<sup>60</sup>. We also compared the microbial fingerprints differences between normal and tumor tissue as well as between different ethnic groups at phylum level and genus level. Abundant taxa (bacterial phyla > 1% and genera > 5%) were selected for this analysis. Wilcoxon Rank Sum Test or Wilcoxon Signed Rank Test were used to screen the significantly different bacterial taxa. Wilcoxon Rank Sum Test or Wilcoxon Signed Rank Test were performed using R program.

## Data availability

The data used in this study has been deposited in NCBI-SRA database. The raw sequence reads generated for this study are available in the NCBI Sequence Read Archive (SRA) under the BioProject ID PRJNA637875.

Received: 28 March 2020; Accepted: 10 August 2020

Published online: 24 August 2020

## References

- Gerend, M. A. & Pia, M. Social determinants of Black–White disparities in breast cancer mortality: A review. *Cancer Epide Miol. Biomark. Prev.* **17**, 2913–2923. <https://doi.org/10.1158/1055-9965.EPI-07-0633> (2008).
- Stark, A. *et al.* African ancestry and higher prevalence of triple-negative breast cancer: Findings from an international study. *Cancer* **116**, 4926–4932. <https://doi.org/10.1002/cncr.25276> (2010).
- Deshmukh, S. K. *et al.* Resistin and interleukin-6 exhibit racially-disparate expression in breast cancer patients, display molecular association and promote growth and aggressiveness of tumor cells through STAT3 activation. *Oncotarget* **6**, 11231–11241. <https://doi.org/10.18632/oncotarget.3591> (2015).
- Kang, J. H., Yu, B. Y. & Youn, D. S. Relationship of serum adiponectin and resistin levels with breast cancer risk. *J. Korean Med. Sci.* **22**, 117–121. <https://doi.org/10.3346/jkms> (2007).
- DeSantis, C. E., Ma, J., Goding, S. A., Newman, L. A. & Jemal, A. Breast cancer statistics, 2017, racial disparity in mortality by state. *CA A Cancer J. Clin.* **67**, 439–448. <https://doi.org/10.3322/caac.21412> (2017).
- American Cancer Society. *Cancer facts & figures for African Americans 2016–2018* (American Society, Atlanta, 2016).
- Sinshaw, H. M. *et al.* Association of race/ethnicity, socioeconomic status, and breast cancer subtypes in the National Cancer Data Base (2010–2011). *Breast Cancer Res. Treat.* **145**, 753–763. <https://doi.org/10.1007/s10549-014-2976-9> (2014).
- Kohler, B. A. *et al.* Annual Report to the Nation on the Status of Cancer, 1975–2011, featuring incidence of breast cancer subtypes by race/ethnicity, poverty, and state. *J. Natl. Cancer Inst.* <https://doi.org/10.1093/jnci/djv048> (2015).
- Walker, R. J., Strom, W. J. & Egede, L. E. Influence of race, ethnicity and social determinants of health on diabetes outcomes. *Am. J. Med. Sci.* **351**, 366–373. <https://doi.org/10.1016/j.amjms.2016.01.008> (2016).
- Xuan, J. M. *et al.* Microbial dysbiosis is associated with human breast cancer. *PLoS ONE* **9**, e83744. <https://doi.org/10.1371/journal.pone.0083744> (2014).
- Urbaniak, C. *et al.* The Microbiota of breast tissue and its association with breast cancer. *Appl. Environ. Microbiol.* **82**, 5039–5048. <https://doi.org/10.1128/aem.01235-16> (2016).
- Hieken, T. J. *et al.* The microbiome of aseptically collected human breast tissue in benign and malignant disease. *Sci. Rep.* **6**, 30751. <https://doi.org/10.1038/srep30751> (2016).
- Pan, H. *et al.* A single bacterium restores the microbiome dysbiosis to protect bones from destruction in a rat model of rheumatoid arthritis. *Microbiome* **7**(1), 107. <https://doi.org/10.1186/s40168-019-0719-1> (2019).
- Modarressi, R. E., Daneshvar, M., Beigom, M. & Mobasheri, E. M. Lactobacillus acidophilus and lactobacillus crispatus culture supernatants downregulate expression of cancer-testis genes in the MDA-MB-231 cell line. *Asian Pac. J. Cancer Prevent.* **15**, 4255–4259. <https://doi.org/10.7314/apjcp.2014.15.10.4255> (2014).
- Pannaraj, P. S. *et al.* Association between breast milk bacterial communities and establishment and development of the infant gut microbiome. *JAMA Pediatr.* **171**, 647–654. <https://doi.org/10.1001/jamapediatrics.2017.0378> (2017).
- Le Doare, K., Holder, B., Bassett, A. & Pannaraj, P. S. Mother's milk: A purposeful contribution to the development of the infant microbiota and immunity. *Front. Immunol.* **9**, 3610. <https://doi.org/10.3389/fimmu.2018.00361> (2018).
- Banerjee, S. *et al.* Distinct microbial signatures associated-with different breast cancer types. *Front. Microbiol.* **9**, 951. <https://doi.org/10.3389/fmicb.00951> (2018).
- Banerjee, S. *et al.* Distinct microbiological signatures associated with triple negative breast cancer. *Sci. Rep.* **5**, 15162. <https://doi.org/10.1038/srep15162> (2015).
- Burns, M. B. *et al.* Colorectal cancer mutational profiles correlate with defined microbial communities in the tumor microenvironment. *PLoS Genet.* **14**(6), e1007376. <https://doi.org/10.1371/journal.pgen.1007376> (2018).
- Qin, Y. & Wade, P. A. Crosstalk between the microbiome and epigenome: Messages from bugs. *J. Biochem.* **163**, 105–112. <https://doi.org/10.1093/jb/mvx080> (2018).
- Gupta, V. K., Paul, S. & Dutta, C. Geography, ethnicity or subsistence-specific variations in human microbiome composition and diversity. *Front. Microbiol.* **23**, 1162. <https://doi.org/10.3389/fmicb.2017.01162> (2017).

22. Pasoli, E. *et al.* Extensive unexplored human microbiome diversity revealed by over 150,000 genomes from metagenomes spanning age, geography, and lifestyle. *Cell* **176**, 649–662.e20. <https://doi.org/10.1016/j.cell.2019.01.001> (2019).
23. Brooks, A. W., Priya, S., Blekhan, R. & Bordenstein, S. R. Gut microbiota diversity across ethnicities in the United States. *PLoS Biol* **16**, e2006842. <https://doi.org/10.1371/journal.pbio.2006842> (2018).
24. Smith, A. *et al.* 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.* **9**, 11940. <https://doi.org/10.1038/s41598-019-48348-1> (2019).
25. Gilbert, K. J. *et al.* Recommendations for utilizing and reporting population genetic analyses: The reproducibility of genetic clustering using the program structure. *Mol. Ecol.* **21**, 4925–4930. <https://doi.org/10.1111/j.1365-294X.2012.05754> (2012).
26. Yedjou, C. G. *et al.* Assessing the racial and ethnic disparities in breast cancer mortality in the United States. *Int. J. Environ. Res. Public Health* **145**, 486. <https://doi.org/10.3390/ijerph14050486> (2017).
27. Martin, D. N. Differences in the tumor microenvironment between African-American and European-American breast cancer patients. *PLoS One* **4**, e4531. <https://doi.org/10.1371/journal.pone.0004531> (2009).
28. Jiagge, E., Chitale, D. & Newman, L. A. Triple-negative breast cancer, stem cells, and African ancestry. *Am. J. Pathol.* **188**, 271–279. <https://doi.org/10.1016/j.ajpath.2017.06.020> (2018).
29. Piyarathna, *et al.* 2019 ERR1 and PGC1 $\alpha$  associated mitochondrial alterations correlate with pan-cancer disparity in African Americans. *J. Clin. Invest.* **129**, 2351–2356. <https://doi.org/10.1172/JCI127579> (2019).
30. Findley, K., Williams, D. R., Grice, E. A. & Bonham, V. L. Health disparities and the microbiome. *Trends Microbiol.* **24**, 847–850. <https://doi.org/10.1016/j.tim> (2016).
31. Ou, J. *et al.* Diet, microbiota, and microbial metabolites in colon cancer risk in rural Africans and African Americans. *Am. J. Clin. Nutr.* **1**, 111–120. <https://doi.org/10.3945/ajcn.112.056689> (2013).
32. Rea, D. *et al.* Microbiota effects on cancer: From risks to therapies. *Oncotarget* **9**(25), 17915–17927. <https://doi.org/10.18632/oncotarget.24681> (2018).
33. Urbaniak, C. *et al.* Microbiota of human breast tissue. *Appl. Environ. Microbiol.* **80**, 3007–3014. <https://doi.org/10.1128/AEM.00242-14> (2014).
34. Parida, S. & Sharma, D. The power of small changes: Comprehensive analyses of microbial dysbiosis in breast cancer. *Biochim. Biophys. Acta Rev. Cancer* **1871**, 392–405. <https://doi.org/10.1016/j.bbcan.2019.04.001> (2019).
35. Wexler, H. M. Bacteroides: The good, the bad, and the nitty-gritty. *Clin. Microbiol. Rev.* **4**, 593–621. <https://doi.org/10.1128/CMR.00008-07> (2007).
36. Mariat, D. *et al.* The firmicutes/bacteroidetes ratio of the human microbiota changes with age. *BMC Microbiol.* **9**, 123. <https://doi.org/10.1186/1471-2180-9-123> (2009).
37. Rinninella, *et al.* What is the healthy gut microbiota composition? A changing ecosystem across age, environment, diet, and diseases. *Microorganisms* **7**, 8. <https://doi.org/10.3390/microorganisms7010014> (2019).
38. Barka, J. *et al.* Taxonomy, physiology, and natural products of actinobacteria. *Microbiol. Mol. Biol. Rev.* **80**, 1–43. <https://doi.org/10.1128/MMBR.00019-15> (2015) (Erratum in: *Microbiol Mol Biol Rev*; **80**(4)).
39. Bhatt, A. S. *et al.* Sequence-based discovery of *Bradyrhizobium enterica* in cord colitis syndrome. *N Engl J Med.* **369**, 517–528. <https://doi.org/10.1056/NEJMoa1211115> (2013).
40. Muhlebach, M. S. *et al.* Initial acquisition and succession of the cystic fibrosis lung microbiome is associated with disease progression in infants and preschool children. *PLoS Pathog* **14**, e1006798. <https://doi.org/10.1371/journal.ppat.1006798> (2018).
41. Ganesan, K., Guo, S., Fayyaz, S., Zhang, G. & Xu, B. Targeting programmed fusobacterium nucleatum Fap2 for colorectal cancer therapy. *Cancers (Basel)* **11**(10), 1592. <https://doi.org/10.3390/cancers11101592> (2019).
42. Hunt, K. M. *et al.* Characterization of the diversity and temporal stability of bacterial communities in human milk. *PLoS One* **6**, e21313. <https://doi.org/10.1371/journal.pone.0021313> (2011).
43. Puri, P. *et al.* TREAT Consortium. The circulating microbiome signature and inferred functional metagenomics in alcoholic hepatitis. *Hepatology* **67**, 1284–1302. <https://doi.org/10.1002/hep.29623> (2018).
44. Luo, Y. *et al.* Complete genome of *Phenylobacterium zucineum*—A novel facultative intracellular bacterium isolated from human erythroleukemia cell line K562. *BMC Genom.* **9**, 386. <https://doi.org/10.1186/1471-2164-9-386> (2008).
45. Ferreira, R. M. *et al.* Gastric microbial community profiling reveals a dysbiotic cancer-associated microbiota. *Gut* **67**(2), 226–236. <https://doi.org/10.1136/gutjnl-2017-314205> (2018).
46. Torrazza, R. M. *et al.* Intestinal microbial ecology and environmental factors affecting necrotizing enterocolitis. *PLoS ONE* **8**(12), e83304. <https://doi.org/10.1371/journal.pone.0083304> (2013).
47. Lackey, K. A. *et al.* What's normal? Microbiomes in human milk and infant feces are related to each other but vary geographically: The INSPIRE Study. *Front. Nutr.* **6**, 45. <https://doi.org/10.3389/fnut.2019.00045> (2019).
48. Milani, C. *et al.* The first microbial colonizers of the human gut: Composition, activities, and health implications of the infant gut microbiota. *Microbiol Mol Biol Rev* **81**, 36. <https://doi.org/10.1128/MMBR.00036-17> (2017).
49. Yu, G. *et al.* Characterizing human lung tissue microbiota and its relationship to epidemiological and clinical features. *Genome Biol.* **17**, 163. <https://doi.org/10.1186/s13059-016-1021-1> (2016).
50. Anderson, C. A. *et al.* Data quality control in genetic case-control association studies. *Nat. Protoc.* **9**, 1564. <https://doi.org/10.1038/nprot.2010.116> (2010).
51. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genomeresearch* **19**, 1655–1664. <https://doi.org/10.1101/gr.094052.109> (2009).
52. Schloss, P. D. *et al.* Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* **75**, 7537–7541. <https://doi.org/10.1128/aem.01541-09> (2009).
53. Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C. & Knight, R. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **27**, 2194–2200. <https://doi.org/10.1093/bioinformatics/btr381> (2011).
54. Schloss, P. D. & Handelsman, J. Toward a census of bacteria in soil. *PLOS Comput. Biol.* **2**, e92. <https://doi.org/10.1371/journal.pcbi.0020092> (2006).
55. Wang, Q., Garrity, G. M., Tiedje, J. M. & Cole, J. R. Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* **73**, 5261–5267. <https://doi.org/10.1128/aem.00062-07> (2007).
56. Shannon, C.E. & Weaver, W. *The Mathematical Theory of Communication* (University of Illinois Press Urbana, 1949).
57. Chao, A. Nonparametric estimation of the number of classes in a population. *Scand. J. Stat.* **11** 265–270 (1984).
58. Chazdon, R.L., Colwell, R.K., Denslow, J.S., & Guariguata, M.R. Statistical methods for estimating species richness of woody regeneration in primary and secondary rain forests of Northeastern Costa Rica. In *Forest Biodiversity Research, Monitoring and Modeling: Conceptual Background and Old World Case Studies* (eds. Dallmeier, F., & Comiskey, J.A.) 285–309 (eds. Dallmeier, F., & Comiskey, J.A.) 285–309 (Parthenon Publishing, Paris 1998).
59. Lozupone, C. & Knight, R. UniFrac: A new phylogenetic method for comparing microbial communities. *Appl. Environ. Microbiol.* **71**, 8228–8235 (2005). <https://doi.org/10.1128/aem.71.12.8228-8235.2005>.
60. Anderson, M. J. A new method for non-parametric multivariate analysis of variance. *Austral. Ecol.* **26**, 32–46. <https://doi.org/10.1111/j.1442-9993.2001.01070.pp.x> (2001).

## Acknowledgements

The research reported in this publication was supported by the National Institutes of Health under award numbers S21MD012472 and U54MD006882.

## Authors Contributions

Sr.T. and J.K.V. conceived and designed the study. J.K.V. supervised the study. Sr.T. procured tissue samples and organized samples to send to Zymo Corporation for DNA sequencing and generation of raw sequence reads, Sa.T. performed initial preliminary analysis of the data. Sa.T. and Sr.T. formatted and deposited the raw sequence files into the NCBI-SRA database. N.P. performed the ancestry analysis, P.C. assisted in tissue procurement and processing. Y.Z. processed the data using mothur, Y.Z. and M.S.A. analyzed data and performed statistical analyses. Sr.T. drafted introduction, methods, results and discussion sections; Y.Z. drafted the methods and results sections of the initial manuscript, M.V.K. assisted in manuscript revisions and proof reading. All authors provided feedback and insights into the manuscript. Sr.T., J.K.V., M.S.A., Y.Z., M.V.K. and Sa.T. edited and revised the manuscript, and all authors read, edited, and approved the final version of the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41598-020-71102-x>.

**Correspondence** and requests for materials should be addressed to J.K.V.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**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 Author(s) 2020