npj | breast cancer

Article

Published in partnership with the Breast Cancer Research Foundation

https://doi.org/10.1038/s41523-025-00806-y

Race-related host and microbe transcriptomic signatures in triple-negative breast cancer

Check for updates

Roshan Kumar[1,2,3], Susan Duyar-Ayerdi[1,2], Aishwarya Sundaresan[4], Vinodh Srinivasasainagendra[4], Chandra Sekhar Pedamallu[5], Michael Behring[6], Darshan Shimoga Chandrashekar[6], Isam-Eldin Eltoum[6], Sooryanarayana Varambally[6], Hemant K. Tiwari[4], Sadeep Shrestha[7], Paul L. Auer[2,8], Lubna N. Chaudhary[9], John R. Kirby[10], Clayton Yates[11,12], Upender Manne[6] ✉ & Akinyemi I. Ojesina[1,2,10] ✉

Triple-negative breast cancer (TNBC) shows racial disparities, with higher incidence in women of African ancestry (AA) compared to European ancestry (EA). Meta-transcriptomic analysis of TNBC tumor tissues from AA (n = 17) and EA (n = 19) subjects revealed distinct microbial landscapes. Hierarchical clustering based on microbial transcripts separated samples into two groups predominantly defined by racial ancestry. Bacterial genera including *Hafnia* and *Cedecea* were more abundant in AA tumors, while *Erwinia* was higher in EA tumors. Cellular composition analysis by xCell revealed differences in immune cell populations, with AA tumors having higher Th1 cell abundance and EA tumors containing higher macrophage M2 cell abundance. Nonetheless, AA women with high M2 abundance experienced poorer disease-free survival (DFS) than EA women. Integrative analyses revealed that high expression of human *SPDYE2B* gene was associated with *Hafnia* abundance and decreased DFS, highlighting complex host-microbe interactions in TNBC patients.

Breast cancer represents a preeminent health burden, leading to approximately 15% of cancer-related deaths in women worldwide. In the United States about 13 of every 100 women are expected to develop breast cancer during their lifetime[1]. Triple-negative breast cancer (TNBC), a highly aggressive form of breast cancer is characterized by the absence of estrogen receptor (ER), progesterone receptor (PgR), and human epidermal growth factor receptor-2 (HER2)[2] as determined by immunohistochemistry[3]. Patients diagnosed with TNBC face an increased risk of distant metastasis and mortality within the first 5 years of diagnosis[4].

Various factors, both genetic and non-genetic, are associated with an elevated risk for TNBC[5–8]. TNBCs exhibit racial disparities, with higher incidence and mortality rates in women of African ancestry (AA) compared to those of European ancestry (EA)[9–11]. AA women tend to be diagnosed at younger ages and more advanced disease stages[12,13]. Although socioeconomic differences contribute to these disparities[14,15], biological factors are also involved in TNBC risk and outcomes for AA patients[8]. Well-documented pathogenic variants in the *BRCA1* and *BRCA2* genes partially determine breast cancer susceptibility, and many new therapies target known genetic drivers[16]. Our previous studies show gene expression differences between AAs and EAs[17,18]. However, the contribution of environmental factors remains poorly understood. Recent studies reveal diverse microbial communities in normal and malignant breast tissue[19–29], with tumor type-specific bacteria present in both cancer and immune cells[29]. In the present study, we aimed to provide insights into the potential interplay between the tumor microbes and host genes in AA and EA patients and their associations with prognostic outcomes.

[1]Department of Obstetrics and Gynecology, Medical College of Wisconsin, Milwaukee, WI, USA. [2]Medical College of Wisconsin Cancer Center, Milwaukee, WI, USA. [3]Post-Graduate Department of Zoology, Magadh University, Bodh-Gaya, Bihar, India. [4]Department of Biostatistics, University of Alabama at Birmingham, Birmingham, AL, USA. [5]Excelra, Boston, MA, USA. [6]Department of Pathology, University of Alabama at Birmingham, Birmingham, AL, USA. [7]Department of Epidemiology, University of Alabama at Birmingham, Birmingham, AL, USA. [8]Division of Biostatistics, Data Science Institute, Medical College of Wisconsin, Milwaukee, WI, USA. [9]Department of Medicine, Division of Hematology and Oncology, Medical College of Wisconsin, Milwaukee, WI, USA. [10]Department of Microbiology and Immunology, Medical College of Wisconsin, Milwaukee, WI, USA. [11]Department of Pathology, Johns Hopkins School of Medicine, Baltimore, MD, USA. [12]Department of Oncology, Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins University School of Medicine, Baltimore, MD, USA. ✉e-mail: upendermanne@uabmc.edu; aojesina@mcw.edu

Results

Patient cohort

We investigated 17 AAs and 19 EA samples, a subset of tumors defined by genetic ancestry models in larger, recently described RNASeq studies[17,18,30] of women with TNBCs, diagnosed at the University of Alabama at Birmingham (UAB) between 2001 and 2012 (Table 1). The samples used were approved by the Institutional Review Board at UAB (IRB number: 060911009). When available, clinical information such as age, gender, date of diagnosis, and pathological stage (according to the American Joint Committee on Cancer [AJCC] staging system) of the tumor, determined after surgical resection, was obtained from the patients' medical records. Metastatic staging was not performed unless clinically indicated. Therefore, all patients were considered by their treating surgeons to have localized, non-metastatic cancer.

The survival time ranged between 1.7 months after diagnosis to over 14.8 years (Fig. 1a). By the end of the follow-up period (median = 84.7 months), 16 of the 36 patients had died due to cancer. Of the 20 patients alive, 14 were considered disease-free (Fig. 1a). The clinical stage at diagnosis spanned between stages II to IV, with most being stage II, followed by stage III (Fig. 1a). There were significantly more EA patients with stage II than AA patients ($p = 0.006$) (Table 1). Concordant with what was expected among TNBCs, most patients had grade III tumors regardless of race. Patients who received radiation in addition to chemotherapy (neoadjuvant and/or adjuvant) had better DFS than those without radiation treatment, but there was no statistically significant difference in DFS between AA and EA patients (Supplementary Fig. 1).

Table 1 | Race and clinical attributes of the selected patients

	Level	AA	CA	*p* value
Sample Size		17	19	
Stage (%)	II	10 (58.8)	18 (94.7)	0.006
	III	7 (41.2)	0 (0.0)	
	IV	0 (0.0)	1 (5.3)	
Grade (%)	II/III	2 (11.8)	1 (5.3)	0.92
	III/III	15 (88.2)	18 (94.7)	
Neoadjuvant Chemotherapy (%)	No	11 (64.7)	16 (88.9)	0.194
	Yes	6 (35.3)	2 (11.1)	
Adjuvant Chemotherapy (%)	No	3 (23.1)	1 (7.1)	0.534
	Yes	10 (76.9)	13 (92.9)	
Radiation (%)	No	6 (40.0)	6 (37.5)	1
	Yes	9 (60.0)	10 (62.5)	

Microbial abundance and cluster analysis

RNA sequencing of macro-dissected breast tumor samples was performed using standard methods[17,18,30]. The identity and abundance of microbial transcripts in tumor samples were determined using the PathSeq computational subtraction pipeline[31,32]. Unsupervised hierarchical clustering of relative microbial abundance data using Spearman rank correlation revealed two major microbe-derived (MD) clusters, MD-Cluster 1 and MD-Cluster 2 (Fig. 1a). Interestingly, these MD-Clusters exhibited race-related separation (Fisher's exact p = 0.044). MD-Cluster I was predominantly comprised of AA patients (ntotal = 15, nAA = 10); MD-Cluster II was dominated by EA patients (ntotal = 20, nEA = 14). Within each MD-Cluster, samples formed sub-clusters by racial group. In the AA cohort, patients in MD-Cluster 1 had better DFS than those in MD-Cluster 2 (Fig. 1b), whereas, in the EA cohort, patients in MD-Cluster 2 had better DFS than those in MD-Cluster 1 (Fig. 1c).

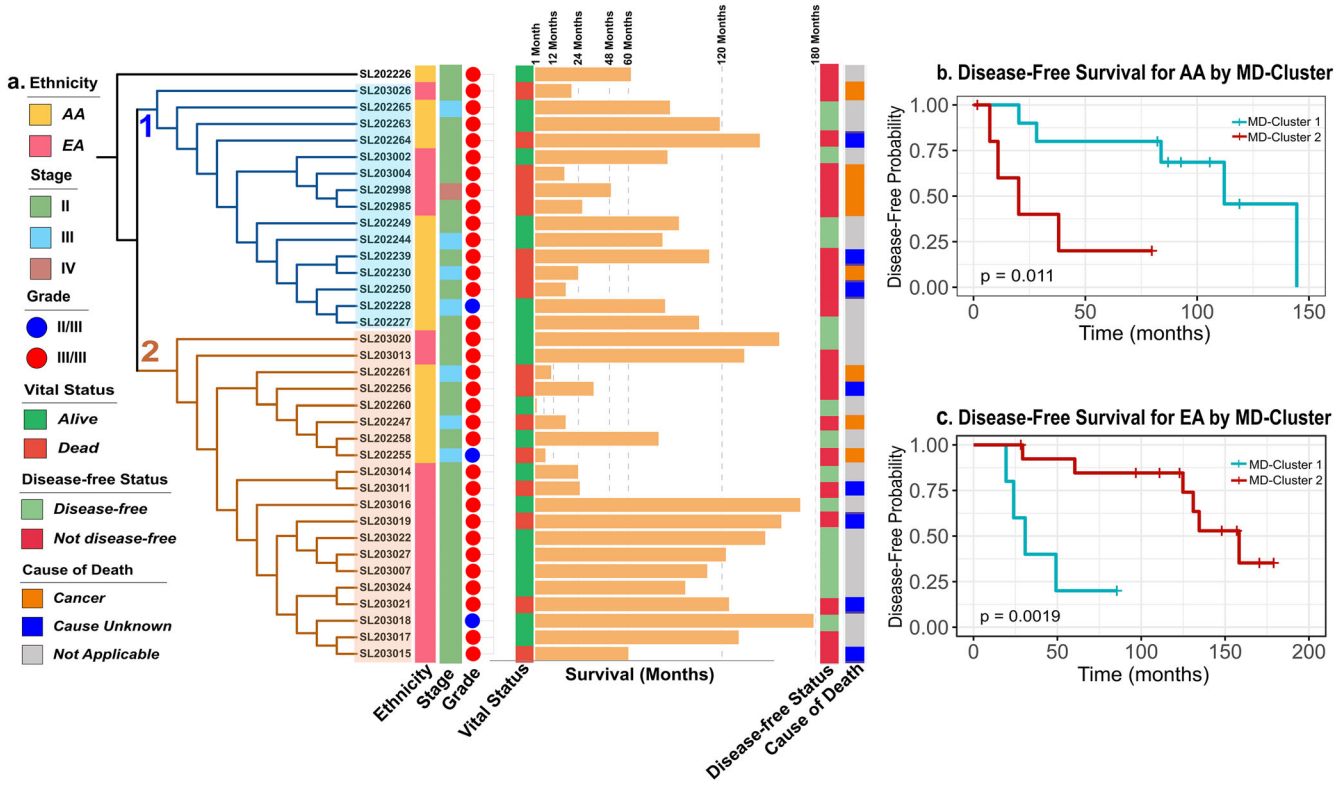

Fig. 1 | Clinicopathological features of microbial transcript-based tumor MD-Clusters. a Hierarchical clustering separated tumors into two major MD-Clusters by microbial transcript profiles. MD-Cluster 1 predominantly contained AA samples (nAA = 10), while MD-Cluster 2 was enriched for EA samples (nEA = 14). Within each MD-Cluster, sub-clusters further segregated by race. Metadata associated with each sample, including stage, grade, vital status, disease-free status (Disease-free: alive without disease or unknown status; Not disease-free: dead or alive with confirmed disease), and cause of death, is displayed. **b**, **c** Kaplan–Meier analysis of disease-free survival (DFS) for AA and EA, respectively, based on MD-Clusters.

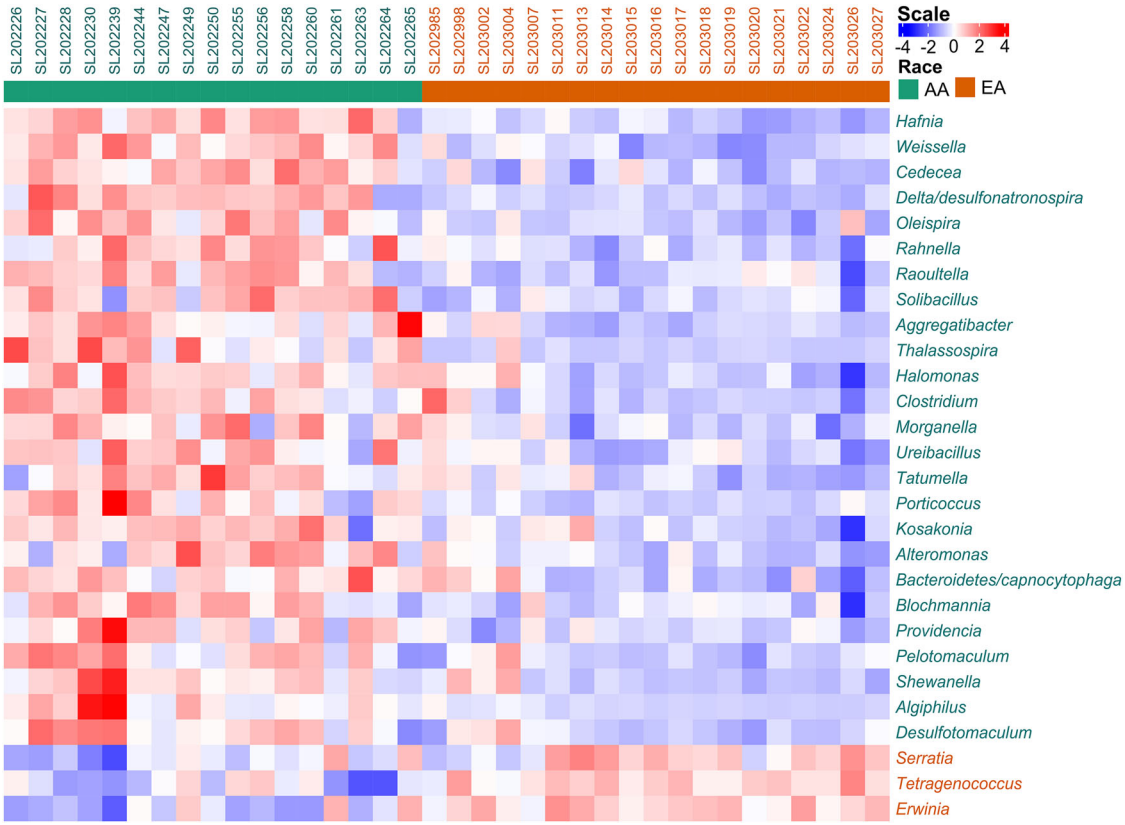

Fig. 2 | Differential abundance of tumor microbiome in tnbc patients by race. Comparative marker selection analysis was performed using t-tests (FDR < 0.05) to identify taxa with differential abundance between AA and EA patients.

The most abundant genera in both AA and EA tumors were similar: *Bacillus, Escherichia, Serratia, Shigella*, and *Cronobacter* (Supplementary Data 1). However, there were significant differences in relative microbial abundance in tumors across the two racial groups (Fig. 2). *Hafnia, Weissella, Cedecea, Delta/Desulfonatronospira, Olespira, Rahnella, Raoultella, Solibacillus, Kosakonia, Aggregatibacter, Thalassospira*, *Halomonas*, and *Clostridium* species exhibited significantly higher abundance (FDR ≤ 0.05) in the AA population. In contrast, *Erwinia*, *Tetragenococcus*, and *Serratia* showed significantly higher abundance (FDR ≤ 0.05) in the EA patients (Fig. 2). Additionally, the differential abundance of microbes based on MD-Clusters (Supplementary Fig. 2) identified *Serratia* was notably abundant in MD-Cluster 2, which was predominantly comprised of EA patients. Interrogation of the Microbe Atlas Project database (https://microbeatlas.org/) revealed that the above-listed bacterial genera have been previously detected in human samples.

Differential expression of genes based on race and MD-Cluster

Differential expression analysis of human RNASeq data by race revealed 40 differentially upregulated ($L_2FC \geq 1$ and FDR ≤ 0.05), and 129 differentially downregulated genes in AA patients compared with EA patients ($L_2FC \leq -1$ and FDR ≤ 0.05), including protein-coding and non-coding genes (Fig. 3a, Supplementary Data 2). Of these, the top 5 differentially upregulated (by L_2FC) coding genes in AAs were *SPDYE2B, HIST1H4F, NES, SPDYE2*, and *HIST1H3I*, while the top 5 non-coding upregulated genes in AAs were *CSPG4P10*, *RNU6-1266P*, *MTND4P12*, *TCEB3CL2*, and *CSPG4P8*. In contrast, the top 5 downregulated coding genes in AAs were *ISM1, IGFBP3, PYGL, BCAS4*, and *SYTL2*, while the top 5 non-coding downregulated genes in AAs were *RN7SKP48*, *MTRNR2L10*, *MTRNR2L11*, *RP11-777B9.5*, and *RN7SL499P*.

Ingenuity Pathway Analysis (IPA; Qiagen) revealed several significant canonical pathways ($p \leq 0.05$) with differential expression between AA and EA patients. DNA replication pre-initiation and DNA damage/Telomere stress-induced senescence were upregulated in AAs compared with EAs (Fig. 3b), with higher expression of histone H2A and H2B family genes, such as *H2AC14, H2AC7, H2BC12, H2BC14, H2BC15*, and *H2BC9* (Supplementary Data 3). In contrast, the stathmin1 pathway, the Th2 pathway, the S100 family signaling pathway, and IL-4 signaling were downregulated in AAs (Fig. 3b). A list of genes involved in these pathways is provided in Supplementary Data 3.

Differential gene expression analysis of the MD-Clusters revealed 31 upregulated ($L2FC \geq 1$ and FDR ≤ 0.05) and 64 downregulated genes ($L2FC \leq -1$ and FDR ≤ 0.05) (Supplementary Fig. 3) in MD-Cluster 1. The top 5 upregulated coding genes were *PRRG3, MT1H, RNY5, LRRD1* and *WDR93* while the top 5 non-coding genes were *SNORD21*, *MT-TF*, *CTD-2651B20.7*, *SNORD55* and *SNORD23*. The top 5 downregulated coding genes were *SPDEF*, *CXCL17*, *CHRDL1*, *IGF1* and *FCRL5*, while the top 5 downregulated non-coding genes were *CTD-2545G14.7*, *TRAJ33*, *TRAJ39*, *TRAJ30* and *SSXP10* (Supplementary Data 4).

IPA revealed upregulation of genes engaged in tRNA processing in mitochondria and RNA polymerase II transcription of snRNA genes in MD-Cluster 1 (Fig. 3c). Conversely, triglyceride metabolism and docosahexaenoic acid (DHA) pathways were downregulated in MD-Cluster 1. A list of the genes involved in these pathways is provided in Supplementary Data 5.

Next, we investigated the cell type enrichment in the tumors using xCell[33]. The analysis revealed significant enrichment (p-value ≤ 0.05) of preadipocytes in MD-Cluster 2 (Fig. 3d). In addition, Th1 cells, megakaryocyte-erythroid progenitor (MEP) cells, and pro B-cells in AA patients and M2 macrophages in EA patients (Fig. 3e, Supplementary Fig. 4A). Interestingly, a linear relationship was observed between M1 and M2 abundance in AA but not in EA women (Supplementary Fig. 4B), and survival analysis with patient stratification by race and M2 abundance revealed that AA women with high M2 abundance had poorer DFS than other subgroups (Supplementary Fig. 4C).

Fig. 3 | Differentially expressed genes and related pathways in TNBC patients. (**a**) Volcano plot showing the log2 fold change and adjusted p-values for all differentially expressed genes between AA and EA cohorts (| log2FoldChange | ≥ 1, padj < 0.05). **b**, **c** Bar charts depicting canonical pathways predicted to be differentially activated in AA (red) versus EA (cyan), and MD-Cluster 1 (violet) vs MD-Cluster 2 (bright blue), respectively. Pathways were identified based on a Z-score cutoff of ±2 and a p-value of <0.05. Detailed lists of genes and pathways are available in Supplementary Data 2, 3. **d**, **e** Box plots of the xCell scores for preadipocytes by MD-Cluster and Th1 cells by race, respectively.

Integrated analysis of microbial and host gene abundance and their relationship to disease-free survival

To identify relationships between the microbiome and the host transcriptome, we combined microbial transcript abundance and host gene expression data into a single matrix. Hierarchical clustering using Spearman rank correlation was performed on the combined data matrix. Of the 1000 clusters generated, there were 828 homologous clusters (823 with host genes only, 5 with bacteria only) while 172 heterologous clusters contained at least one host gene and one bacterium (Fig. 4, Supplementary Data 6). There were 7 "high-confidence" heterologous clusters, which displayed strong positive correlative relationships (Spearman rank R > 0.7) between microbial abundance and host gene expression (Supplementary Data 7).

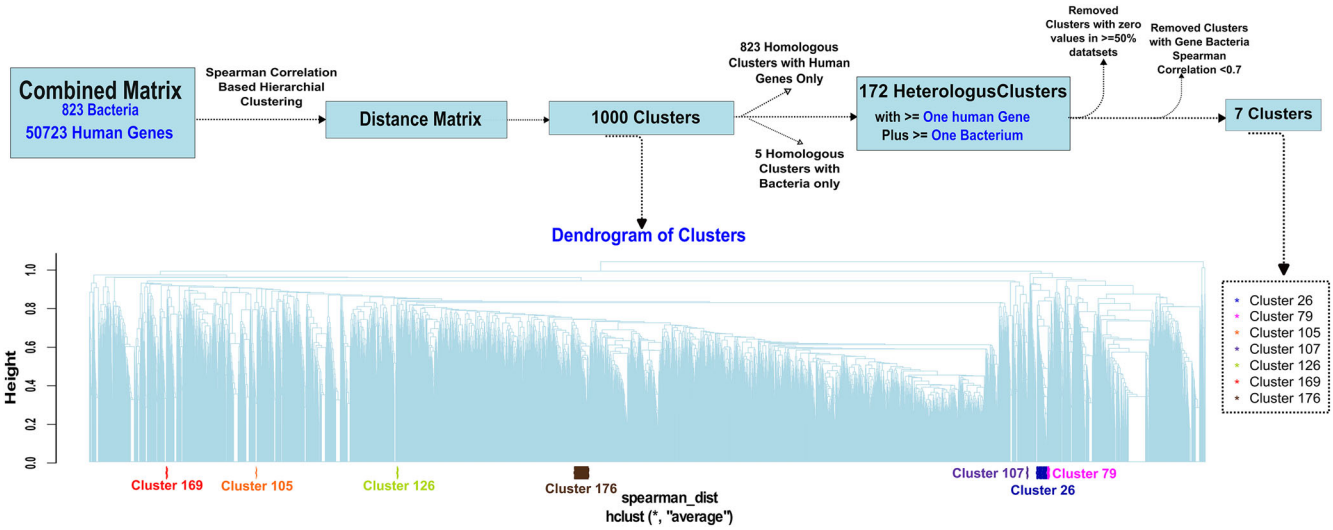

Fig. 4 | Hierarchical clustering-based relationships between host genes and bacteria. Host genes and microbial transcripts were hierarchically clustered based on co-abundance patterns, with homologous clusters containing either host genes only or bacteria only filtered out to retain high-confidence heterologous clusters containing both host genes and bacteria with a Spearman correlation >0.7.

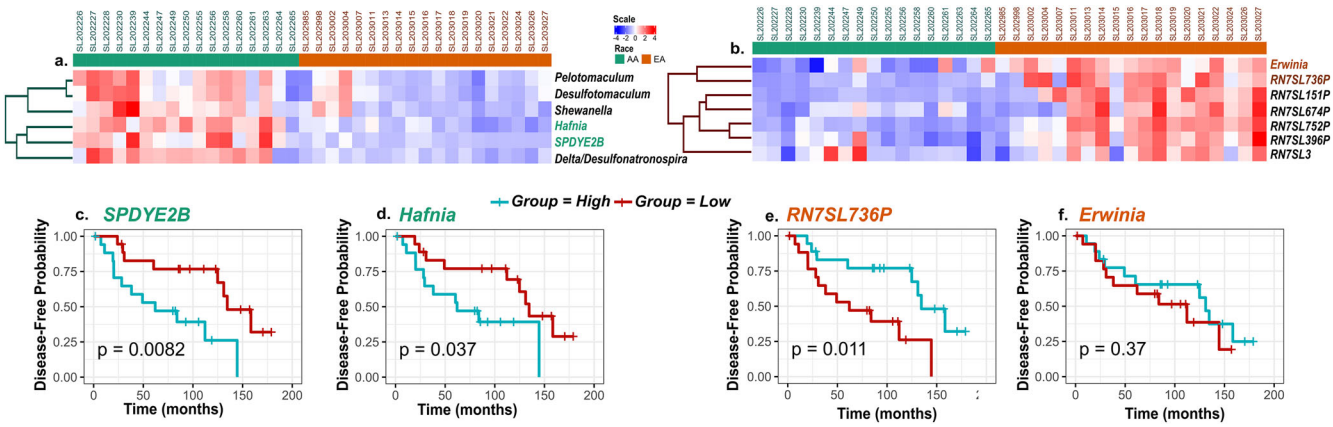

Fig. 5 | Correlative and survival relationships between host gene expression and microbial transcript abundance. a Sub-cluster from cluster 26 showing clustering of *SPDYE2B* gene with *Hafnia*, *Desulfotomaculum*, *Pelotomaculum*, *Shewanella*, and *Delta/Desulfonatronospira*. **b** Sub-cluster from cluster 176 showing clustering of *RN7SL736P* pseudogene with *Erwinia*. **c** Kaplan–Meier (KM) analysis of DFS by *SPDYE2B* gene expression. **d** KM analysis of DFS by *Hafnia* abundance. **e** KM analysis of DFS by *RN7SL736P* gene expression. **f** KM analysis of DFS by *Erwinia* abundance.

Cluster 26 revealed two notable relationships. In one sub-cluster, the host gene *SPDYE2B* closely clustered with *Hafnia* (R = 0.77), *Shewanella* (R = 0.74), *Delta/Desulfonatronospira* (R = 0.69), *Desulfotomaculum* (R = 0.66), and *Pelotomaculum* (R = 0.66). These features were more abundant in AA patients (Fig. 5a). Another sub-cluster of cluster 26 showed close clustering of the *SCARNA4* gene with *Shinella* (R = 0.79). Notably, *Erwinia* abundance was highly correlated in a sub-cluster of cluster 176 with levels of the *RN7SL736P* pseudogene (R = 0.756), which was more abundant in EA patients (Fig. 5b). Other highly correlated host-microbe associations (Supplementary Data 7) included *SNORA18* and *Halothiobacillus* in cluster 79 (R = 0.704), *BLOC1S1* and *Desulforudis* in cluster 105 (R = 0.732), *IGLVI-70* and *Delta/Arcobacter* in cluster 107 (R = 0.723), *KRT5* and *Heliobacterium* in cluster 126 (R = 0.73), *TMEM203* and *Idiomarina* in cluster 169 (R = 0.72), and *CCNYL1* and *Oceanobacillus* in cluster 174 (R = 0.72).

We evaluated the relationship(s) between these features (microbe(s)/gene(s)) and DFS (Fig. 5c–f, Supplementary Fig. 4). For each host gene or microbe, the cohort was stratified into high and low groups determined by the median values for each feature. High expression of the host gene *SPDYE2B* was associated with reduced DFS ($p = 0.0082$) (Fig. 5c). Further, high abundance of *Hafnia* (Fig. 5d) or *Shewanella* (Supplementary Fig. 4) were also associated with worse DFS (p = 0.037 and $p = 0.012$, respectively). However, although high *RN7SL736P* levels were associated with better DFS ($p = 0.011$), there was no association between its correlated microbe *Erwinia* and DFS (Fig. 5e, f).

In a Cox proportional hazards model adjusting for race, MD-Cluster, stage, and treatment type, only a high abundance of *SPDYE2B* and race remained associated with worse DFS (Supplementary Data 8).

Discussion

Host genes affect the pathogenesis and survival outcomes of patients with TNBCs[5,6,34]. The variability in the tumor microenvironment (TME), is involved in shaping the transcriptomic landscape and driving intratumor heterogeneity across various malignancies[35,36] and these fluctuations can alter the growth of microbiota within the TME[37]. Given this dynamic interplay, we investigated the relationships between the tumor microbiome, host transcriptome and outcomes in TNBC patients. For this analysis, we extracted RNA from formalin-fixed paraffin-embedded (FFPE) tissue samples and performed microbial analysis using PathSeq. The use of FFPE samples for microbiome research has gained increased

traction in recent[34–36]. A comparative study between FFPE and fresh frozen samples demonstrate that 78.55% of operational taxonomic units remain consistent across preservation methods, with no significant differences observed in core bacterial composition and alpha diversity[37]. Other studies[34–36], have established the feasibility and reliability of retrospective microbiome studies using archived clinical specimens- which represent the most widespread sample type available in cancer research.

Microbial transcript-based hierarchical clustering of TNBC patients (n = 36) revealed two distinct microbiome-derived clusters (MD-Clusters), each dominated by samples from different racial cohorts. We identified relatively higher abundance of *Hafnia* and *Shewanella* in AA tumors, and *Erwinia* and *Serratia* in EA tumors. The differential abundance of bacterial species by race suggests a role for these microbes in modulating tumor biology and influencing clinical outcomes in a race-dependent manner. Notably, *Hafnia* species are frequently isolated from clinical specimens and are often associated with extra-intestinal infections, particularly in immunocompromised individuals[38,39]. Although *Hafnia* is currently being marketed as a probiotic with anorexigenic properties[40], *Hafnia* infections have also been associated with fatal outcomes[41]. *Hafnia* produces a cytolytic toxin active against Vero cells[42] and has been shown to survive within HeLa cells[43]. These contrasting observations suggest the need for further studies on the pathogenic potential of this bacterial genus. Although our understanding of the factors that may potentially contribute to *Hafnia* pathogenesis within a host is limited, suggested mechanisms include biofilm formation, motility, and adherence to eukaryotic cells[44]. Similarly, *Shewanella*, though a rare pathogen in humans, has been associated with pulmonary and blood infections in cancer patients, indicating its relevance in the context of immunocompromised individuals and its capacity to affect TME[45,46]. The microbes significantly abundant in EA patients, such as *Serratia*, cause infections in cancer patients[47]. The substance known as Coley's toxin, formed by the toxins produced by *Serratia marcescens* and *Streptococcus pyogenes*, induces tumor regression and has been successfully used to treat a variety of cancers, including sarcomas, carcinomas, lymphomas, myelomas, and melanomas[48].

We observed race-associated differential expression in both protein-coding and non-coding RNAs (ncRNAs) genes. Two of the most differentially expressed coding genes in AA patients (*SPDYE2* and *SPDYE2B)* encode proteins that belong to the Speedy/Ringo family of non-canonical activators of the cyclin-dependent kinases, CDK1 and CDK2, which control cell cycle progression[49,50]. Overexpression of the prototypical family member RingoA/Spy1 has been observed in breast cancer tissue microarrays, and siRNA knockdown of RingoA/Spy1 results in decreased proliferation of the TNBC cell line MDA-MB-231[51,52]. Therefore, it is possible that high expression of these Speedy/Ringo genes may differentially modulate cell cycle regulation in AA TNBC patients.

Some of the top upregulated genes in the AA cohort included pseudogenes and ncRNAs, such as *CSPG4P10*, *RNU6-1266P*, *RP11-17J14.2*, and *MTND4P12*, suggesting potential regulatory roles for race-associated pseudogenes in TNBC pathogenesis. For example, the pseudogene *MTND4P12* functions as a competing endogenous RNA (ceRNA), regulating the expression of the well-established oncogene AURKB[53]. By sequestering microRNAs that would otherwise target and repress AURKB mRNA, *MTND4P12* may facilitate the aberrant expression of this kinase, leading to tumor growth and metastasis. Interestingly, *MTND4P12* expression is negatively correlated with overall survival in cutaneous melanoma patients[53]. Although *MTND4P12* was significantly upregulated in AAs, in the present study, *AURKB* itself did not exhibit significant differential expression, suggesting that additional mechanisms may be involved in regulating its activity.

Pathway enrichment analyses provided further insights into the biological processes and signaling cascades dysregulated in a race- and MD-Cluster-related manner. The downregulation of DNA replication and damage response pathways in AAs and upregulation of histone genes *H2AC7*, *H2AC14*, *H2BC9*, *H2BC12*, *H2BC14*, and *H2BC15* in AAs is notable because downregulation of histone pathways (H2A and H2B) is associated with anthracycline sensitivity in breast cancers[54]. In addition, pathways involved in breast cancer regulation by stathmin1 are downregulated in AA patients. Stathmin-1 is a cytoplasmic phosphoprotein involved in mitotic spindle formation and cell division[55]. Its downregulation in AA TNBCs indicates a reduced microtubule formation, which could affect cell division, leading to a differential response to treatments targeting microtubule dynamics, such as taxane-based chemotherapies[56,57].

Comparison of gene expression based on MD-Clusters revealed significant upregulation of genes involved in mitochondrial tRNA processing in MD-Cluster 1 (enriched for AA tumors), and genes related to triglyceride metabolism and docosahexaenoic acid (DHA) pathways in MD-Cluster 2 (predominantly EA tumors). The DHA pathway induces apoptosis in cancer cells by activating the MAPK signaling pathway and suppressing Akt phosphorylation[58,59], and may potentially promote tumor growth, progression, and metastasis in breast cancers[60]. It is therefore possible that upregulation of DHA pathways in MD-Cluster 2 may induce apoptotic effects in EA patients, potentially underlying better disease-free survival.

We performed xCell-based cell enrichment analyses to identify the various cell types associated with race and MD-Clusters. We observed that Th1 cells, MEPs, and pro B-cells were enriched in AA patients; macrophage M2 cells were more abundant in EA patients, with preadipocytes enriched in MD-Cluster 2, dominated by CAs. At first glance, the immune cell findings appear to be at variance with previous studies suggesting the poorer outcomes in breast cancer in general among AA women may be linked in part to higher M2 abundance in AA breast tumors than EA tumors[61,62]. However, similar to our findings, other TNBC-focused studies[63] have observed lower levels of M2 macrophages in AA women, and nuanced interpretation of our data underscores the complexity of these relationships. Our observation of a linear relationship between M1 and M2 abundance in AA but not in EA women is in keeping with the link between mixed M1/M2 phenotype and tumor aggressiveness[64]. Very interestingly, we also observed that the subset of AA women with the highest tumor M2 abundance had poorer DFS compared with EA women regardless of their tumor M2 abundance. This suggests that there is still a relationship between high M2 abundance and poor outcomes in AA women in spite of the overall lower M2 abundance in AA women compared with EA women.

Our integrated analysis of microbial and transcriptional clusters identified significant host-microbe associations that potentially influence DFS in patients. Cluster 26 emerged as particularly notable, with strong positive correlations between *SPDYE2B* and *Hafnia*. Both of these features were significantly abundant in AA patients. A high abundance of *Hafnia* and *SPDYE2B* was associated with worse DFS, suggesting that, for subsets of TNBC patients, these features may synergize in influencing tumor biology and patient outcomes.

The identification of race-specific host-microbe interactions in TNBC offers new insight into how tumor biology and health disparities may differ across populations. The strong positive correlation between *SPDYE2B* expression and *Hafnia* abundance in AA patients, coupled with their shared association with poor DFS, suggests a synergistic relationship that may contribute to the observed racial disparities in TNBC outcomes. This host-microbe axis potentially creates a feedback loop where microbial metabolites or virulence factors influence cell cycle regulation through the Speedy/Ringo pathway, while altered host gene expression may reciprocally modify the tumor microenvironment to favor specific microbial populations. The differential immune cell compositions observed—with AA tumors showing higher Th1 cell abundance and EA tumors exhibiting greater M2 macrophage infiltration—may serve as both drivers and consequences of these distinct microbial communities. Importantly, our observation that AA women with high M2 abundance experienced the poorest survival outcomes suggests that the interplay between immune phenotype and race may create unique vulnerabilities that could be targeted as therapeutic avenues. These findings highlight the necessity of incorporating both host genetics and microbial ecology into precision medicine approaches for TNBC, as traditional biomarkers may inadequately capture the complexity of tumor biology across diverse populations.

Our current study serves as an important proof-of-concept that demonstrates the feasibility of identifying race-related host-microbe interactions in TNBC. While our sample size is small, it is comparable to other exploratory microbiome studies on breast cancer tissues[65–67]. In addition, no two studies have produced the exact same microbial associations. This can be explained in part by the samples being collected from different geographical sites. We also hypothesize that the role of microbes in breast cancer is primarily facultative rather than causative and that the effects are subserved by microbial products and/or metabolites which may be more similar in function than bacterial identity and taxonomy may suggest. Bacteria identified in different studies will ultimately help to narrow down the culprit metabolites in future work.

We have presented a novel integrative analytical framework for identifying host-microbe associations that can be applied to larger cohorts to validate race-associated interactions and their clinical implications in TNBC. Larger cohort studies incorporating multi-omics data and comprehensive clinical annotations would facilitate the identification of robust microbiome-host signatures predictive of disease outcomes, paving the way for the development of novel prognostic biomarkers and therapeutic strategies tailored to distinct patient populations.

Methods

Host transcriptome analysis

The samples were collected as macro-dissected breast tumor tissues during routine surgical procedures at UAB between 2001 and 2012[17]. A retrospective cohort of formalin-fixed, paraffin-embedded (FFPE) archival tissues from the Division of Anatomic Pathology at UAB was subjected to RNA extraction using standard protocol. The extraction of RNA from the macro-dissected samples was subjected to RNA extraction using standard protocol. Raw sequencing reads were assessed for quality using FastQC v0.11.8 (https://github.com/s-andrews/FastQC). Adapters and low-quality bases were trimmed from the raw reads with Trimmomatic v0.36[68] to retain only high-quality sequences for downstream analysis. The trimmed reads were aligned to the human reference genome (GRCh37) using the HISAT2 aligner v2.0.4[69]. Gene type annotations were performed with GENCODE version 46. Following read alignment and rRNA removal, the RNASeq data were assembled into transcript models using Stringtie v1.3.3[69]. Differential expression (DE) analysis between race groups was performed using the Bioconductor package DESeq2 version 1.36.0[70] to compute log2FoldChange (L2FC) and Benjamini-Hochberg (BH)-based False Discovery Rate (FDR) adjusted p-values. Statistical significance was determined at an FDR threshold of 0.05.

Microbial transcriptome analysis

To identify microbial sequences from tumor RNASeq data (n = 36), we used the PathSeq pipeline[71,72]. PathSeq first removes low-quality and low-complexity reads. Host reads are then filtered out through a two-step process. First, k-mer matching (default, $k = 31$) rapidly identifies short sequences of human origin. Second, the Burrows-Wheeler Aligner (BWA-MEM)[73] iteratively aligns reads to the human reference genome and removes those with defined identity and coverage. Unaligned non-host reads are subsequently aligned to a microbial genome database (RefSeq: archaea, bacteria, fungi, protozoa, and virus) using BWA-MEM[73]. The mapped reads were then used to derive the microbial read counts at each taxonomic level in each sample.

Using the microbial data, we performed Spearman rank correlation-based hierarchical clustering of tumors with average linking to determine microbe-derived clusters (MD-Clusters). To identify microbes with robust race-specific abundance differences between AA and EA cohorts, we used the comparative marker selection tool in Morpheus (https://software.broadinstitute.org/morpheus/). Microbial transcripts count data were $\log_2$(counts+1) transformed and analyzed by t-tests between AA and EA groups. BH-based false discovery rate (FDR) adjustment was applied to select significant markers at an FDR threshold of 0.05.

Differential expression of host genes based on MD-Cluster profiles

Similar to the race differential expression analysis, MD-Cluster-based differential expression was performed using DESeq2 version 1.36.0[67]. Log_2(fold changes) and FDR-adjusted p-values were computed to identify genes with significant expression differences between MD-Cluster groups. An FDR threshold of 0.05 was used to determine statistical significance.

Pathway analysis

Pathway analysis was conducted using the IPA software (QIAGEN) by using the gene expression data derived from both race and MD-Cluster stratifications. Coding genes with an absolute $\log_2$ fold change ($|L_2(FC)| \geq 1$ and a $p < 0.05$ were subjected to pathway analysis and construction of a protein-protein interaction (PPI) network. Pathways or networks with absolute Z-scores ≥ 2 and $p < 0.05$ were considered significantly enriched.

Integration of microbiome and transcriptomic data

To identify associations between the microbiome composition and host gene expression, microbial transcript abundance and gene expression data were combined into a single matrix. This matrix was subjected to hierarchical clustering using Spearman correlation with the average linkage method to generate 1000 clusters. Of these, clusters with only bacteria or only host genes (homologous clusters) were filtered out, but those containing at least one bacterium and at least one host gene (heterologous clusters), with data representation (non-zero) in ≥50% of samples, were retained. Subsequently, high-confidence heterologous clusters were identified using a cutoff of Spearman correlation >0.7 between host genes and bacteria (Supplementary Fig. 5).

Cell-type enrichment analysis

To investigate cellular heterogeneity in TNBC tumors, we performed gene-signature-based, cell-type enrichment analysis using xCell version 1.1.0 (https://github.com/dviraran/xCell/tree/master). This analysis generated xCell scores for various cell types. Further, we assessed the association of these xCell scores with race and MD-Cluster by using the Wilcoxon rank-sum test.

Disease-free survival

Kaplan-Meier (KM) curves were generated to determine disease-free survival (DFS) using 'survival', 'survminer', and 'ggplot2' R packages. Curves were stratified by race, MD-Cluster, treatment types, or microbe/gene-based abundance (high/low groups split by median). Survival distributions were compared using the log-rank test, with $p < 0.05$ considered statistically significant. The Cox proportional hazards model was used to analyze the impact of *Hafnia* and *SPDYE2B* abundance on DFS, adjusting for race, MD-Cluster, stage, and treatment type.

Data availability

The datasets are available in the BioProject database under Accession No. PRJNA598161. All analyses were performed using publicly available software and standard protocols. No custom code was developed for this study.

Code availability

The specific software versions and parameters used are detailed in the Methods section.

Received: 17 December 2024; Accepted: 22 July 2025;
Published online: 08 August 2025

References

1. *Key Statistics for Breast Cancer*, https://www.cancer.org/cancer/types/breast-cancer/about/how-common-is-breast-cancer.html.

2. Wang, D. Y., Jiang, Z., Ben-David, Y., Woodgett, J. R. & Zacksenhaus, E. Molecular stratification within triple-negative breast cancer subtypes. *Sci. Rep.* **9**, 19107 (2019).
3. Yin, L., Duan, J. J., Bian, X. W. & Yu, S. C. Triple-negative breast cancer molecular subtyping and treatment progress. *Breast Cancer Res.* **22**, 61 (2020).
4. Dent, R. et al. Triple-negative breast cancer: clinical features and patterns of recurrence. *Clin. Cancer Re.s* **13**, 4429–4434 (2007).
5. Stevens, K. N., Vachon, C. M. & Couch, F. J. Genetic susceptibility to triple-negative breast cancer. *Cancer Res.* **73**, 2025–2030 (2013).
6. Stevens, K. N. et al. Common breast cancer susceptibility loci are associated with triple-negative breast cancer. *Cancer Res.* **71**, 6240–6249 (2011).
7. So, J. Y., Ohm, J., Lipkowitz, S. & Yang, L. Triple negative breast cancer (TNBC): Non-genetic tumor heterogeneity and immune microenvironment: emerging treatment options. *Pharm. Ther.* **237**, 108253 (2022).
8. Masi, C. M. & Olopade, O. I. Racial and ethnic disparities in breast cancer: a multilevel perspective. *Med Clin. North Am.* **89**, 753–770 (2005).
9. Ademuyiwa, F. O. & Olopade, O. I. Racial differences in genetic factors associated with breast cancer. *Cancer Metastasis Rev.* **22**, 47–53 (2003).
10. Daly, B. & Olopade, O. I. A perfect storm: how tumor biology, genomics, and health care delivery patterns collide to create a racial survival disparity in breast cancer and proposed interventions for change. *CA Cancer J. Clin.* **65**, 221–238 (2015).
11. Ellsworth, D. L., Turner, C. E. & Ellsworth, R. E. A review of the hereditary component of triple negative breast cancer: high- and moderate-penetrance breast cancer genes, low-penetrance loci, and the role of nontraditional genetic elements. *J. Oncol.* **2019**, 4382606 (2019).
12. Mehrotra, J. et al. Estrogen receptor/progesterone receptor-negative breast cancers of young African-American women have a higher frequency of methylation of multiple genes than those of Caucasian women. *Clin. Cancer Res.* **10**, 2052–2057 (2004).
13. Deshpande, A. D. et al. Racial disparities in breast cancer survival: an analysis by age and stage. *J. Surg. Res.* **153**, 105–113 (2009).
14. Eom, K. Y. et al. Neighborhood and racial influences on triple negative breast cancer: evidence from Northeast Ohio. *Breast Cancer Res. Treat.* **198**, 369–381 (2023).
15. Aoki, R. F. et al. Individual- and neighborhood-level socioeconomic status and risk of aggressive breast cancer subtypes in a pooled cohort of women from Kaiser Permanente Northern California. *Cancer* **127**, 4602–4612 (2021).
16. Fackenthal, J. D. & Olopade, O. I. Breast cancer risk associated with BRCA1 and BRCA2 in diverse populations. *Nat. Rev. Cancer* **7**, 937–948 (2007).
17. Saleh, M. et al. Comparative analysis of triple-negative breast cancer transcriptomics of Kenyan, African American and Caucasian Women. *Transl. Oncol.* **14**, 101086 (2021).
18. Martini, R. et al. African ancestry-associated gene expression profiles in triple-negative breast cancer underlie altered tumor biology and clinical outcome in women of African Descent. *Cancer Discov.* **12**, 2530–2551 (2022).
19. Xuan, C. et al. Microbial dysbiosis is associated with human breast cancer. *PLoS ONE***9**, e83744 (2014).
20. Urbaniak, C. et al. Microbiota of human breast tissue. *Appl. Environ. Microbiol.* **80**, 3007–3014 (2014).
21. Banerjee, S. et al. Distinct microbiological signatures associated with triple negative breast cancer. *Sci. Rep.* **5**, 15162 (2015).
22. Thompson, K. J. et al. A comprehensive analysis of breast cancer microbiota and host gene expression. *PLoS ONE***12**, e0188873 (2017).
23. Banerjee, S. et al. Distinct microbial signatures associated with different breast cancer types. *Front. Microbiol.* **9**, 951 (2018).
24. Dieleman, S. et al. Exploring the potential of breast microbiota as biomarker for breast cancer and therapeutic response. *Am. J. Pathol.* **191**, 968–982 (2021).
25. Tzeng, A. et al. Human breast microbiome correlates with prognostic features and immunological signatures in breast cancer. *Genome Med.* **13**, 60 (2021).
26. Parida, S. & Drewes, J. L. Unwanted passengers: Microbes hitchiking in breast cancer metastases. *Cell Host Microbe* **30**, 875–877 (2022).
27. Parida, S., Siddharth, S., Xia, Y. & Sharma, D. Concomitant analyses of intratumoral microbiota and genomic features reveal distinct racial differences in breast cancer. *NPJ Breast Cancer* **9**, 4 (2023).
28. Desalegn, Z. et al. Human breast tissue microbiota reveals unique microbial signatures that correlate with prognostic features in adult ethiopian women with breast cancer. *Cancers* **15**, 4893 (2023).
29. Nejman, D. et al. The human tumor microbiome is composed of tumor type-specific intracellular bacteria. *Science* **368**, 973–980 (2020).
30. Derakhshan, F. & Reis-Filho, J. S. Pathogenesis of triple-negative breast cancer. *Annu. Rev. Pathol.* **17**, 181–204 (2022).
31. Natua, S., Dhamdhere, S. G., Mutnuru, S. A. & Shukla, S. Interplay within tumor microenvironment orchestrates neoplastic RNA metabolism and transcriptome diversity. *Wiley Interdiscip. Rev. RNA* **13**, e1676 (2022).
32. Hoogstrate, Y. et al. Transcriptome analysis reveals tumor microenvironment changes in glioblastoma. *Cancer Cell* **41**, 678–692 e677 (2023).
33. Li, X. et al. Host-microbiota interactions contributing to the heterogeneous tumor microenvironment in colorectal cancer. *Physiol. Genom.* **56**, 221–234 (2024).
34. Apopa, P. L. et al. PARP1 is up-regulated in non-small cell lung cancer tissues in the presence of the cyanobacterial toxin microcystin. *Front. Microbiol.* **9**, 1757 (2018).
35. Fernandez, M. F. et al. Breast cancer and its relationship with the microbiota. *Int. J. Environ. Res. Public Health* **15**, https://doi.org/10.3390/ijerph15081747 (2018).
36. Bundgaard-Nielsen, C., Baandrup, U. T., Nielsen, L. P. & Sorensen, S. The presence of bacteria varies between colorectal adenocarcinomas, precursor lesions and non-malignant tissue. *BMC Cancer* **19**, 399 (2019).
37. Chang, J. et al. Potential values of formalin-fixed paraffin-embedded tissues for intratumoral microbiome analysis in breast cancer. *Heliyon* **9**, e16267 (2023).
38. Cooney, S., O'Brien, S., Iversen, C. & Fanning, S. in *Encyclopedia of Food Safety* (ed Yasmine, M) 433-441 (Academic Press, 2014).
39. Lang, H. et al. Specific strains of honeybee gut lactobacillus stimulate host immune system to protect against pathogenic Hafnia alvei. *Microbiol. Spectr.* **10**, e0189621 (2022).
40. Dechelotte, P. et al. The Probiotic Strain H. alvei HA4597((R)) improves weight loss in overweight subjects under moderate hypocaloric diet: a proof-of-concept, multicenter randomized, double-blind placebo-controlled study. *Nutrients* **13**, https://doi.org/10.3390/nu13061902 (2021).
41. Ionescu, M. I., Neagoe, D. S., Craciun, A. M. & Moldovan, O. T. The gram-negative bacilli isolated from caves-sphingomonas paucimobilis and hafnia alvei and a review of their involvement in human infections. *Int. J. Environ. Res. Public Health* **19**, https://doi.org/10.3390/ijerph19042324 (2022).
42. Abbott, S. L. et al. Clinical and laboratory diagnostic characteristics and cytotoxigenic potential of Hafnia alvei and Hafnia paralvei strains. *J. Clin. Microbiol.* **49**, 3122–3126 (2011).
43. Padilla, D. et al. Invasion and intracellular survival of Hafnia alvei strains in human epithelial cells. *J. Appl Microbiol.* **105**, 1614–1622 (2008).
44. Ramos-Vivas, J. et al. The molecular weaponry produced by the bacterium hafnia alvei in foods. *Molecules* **27**, https://doi.org/10.3390/molecules27175585 (2022).

45. Zhang, F. et al. Rare Shewanella spp. associated with pulmonary and bloodstream infections of cancer patients, China: a case report. *BMC Infect. Dis.* **18**, 454 (2018).
46. Jeffery, S. Shewanella dysentery in a patient with underlying malignancy. *Med. J. Malays.* **69**, 284–285 (2014).
47. Bodey, G. P., Rodriguez, V. & Smith, J. P. Serratia sp. infections in cancer patients. *Cancer* **25**, 199–205 (1970).
48. Kucerova, P. & Cervinkova, M. Spontaneous regression of tumour and the role of microbial infection-possibilities for cancer treatment. *Anticancer Drugs* **27**, 269–277 (2016).
49. Gastwirt, R. F., McAndrew, C. W. & Donoghue, D. J. Speedy/RINGO regulation of CDKs in cell cycle, checkpoint activation and apoptosis. *Cell Cycle* **6**, 1188–1193 (2007).
50. Gonzalez, L. & Nebreda, A. R. RINGO/Speedy proteins, a family of non-canonical activators of CDK1 and CDK2. *Semin. Cell Dev. Biol.* **107**, 21–27 (2020).
51. Gutierrez, G. J. et al. Meiotic regulation of the CDK activator RINGO/Speedy by ubiquitin-proteasome-mediated processing and degradation. *Nat. Cell Biol.* **8**, 1084–1094 (2006).
52. Al Sorkhy, M. et al. The cyclin-like protein Spy1/RINGO promotes mammary transformation and is elevated in human breast cancer. *BMC Cancer* **12**, 45 (2012).
53. Guo, Y. et al. Inhibition of AURKB, regulated by pseudogene MTND4P12, confers synthetic lethality to PARP inhibition in skin cutaneous melanoma. *Am. J. Cancer Res* **10**, 3458–3474 (2020).
54. Braunstein, M. et al. Downregulation of histone H2A and H2B pathways is associated with anthracycline sensitivity in breast cancer. *Breast Cancer Res.* **18**, 16 (2016).
55. Mistry, S. J. & Atweh, G. F. Role of stathmin in the regulation of the mitotic spindle: potential applications in cancer therapy. *Mt Sinai J. Med.* **69**, 299–304 (2002).
56. Huang, Q., Mei, Z. & Han, X. Efficacy and safety of taxanes combined with chemotherapy drugs in advanced triple negative breast cancer: a meta-analysis of 26 randomized controlled trials. *Front. Oncol.* **12**, 972767 (2022).
57. Chou, C. W., Huang, Y. M., Chang, Y. J., Huang, C. Y. & Hung, C. S. Identified the novel resistant biomarkers for taxane-based therapy for triple-negative breast cancer. *Int. J. Med. Sci.* **18**, 2521–2531 (2021).
58. Yin, Y., Sui, C., Meng, F., Ma, P. & Jiang, Y. The omega-3 polyunsaturated fatty acid docosahexaenoic acid inhibits proliferation and progression of non-small cell lung cancer cells through the reactive oxygen species-mediated inactivation of the PI3K /Akt pathway. *Lipids Health Dis.* **16**, 87 (2017).
59. Toit-Kohn, J. L., Louw, L. & Engelbrecht, A. M. Docosahexaenoic acid induces apoptosis in colorectal carcinoma cells by modulating the PI3 kinase and p38 MAPK pathways. *J. Nutr. Biochem.* **20**, 106–114 (2009).
60. Zipinotti Dos Santos, D. et al. The impact of lipid metabolism on breast cancer: a review about its role in tumorigenesis and immune escape. *Cell Commun. Signal* **21**, 161 (2023).
61. Koru-Sengul, T. et al. Breast cancers from black women exhibit higher numbers of immunosuppressive macrophages with proliferative activity and of crown-like structures associated with lower survival compared to non-black Latinas and Caucasians. *Breast Cancer Res. Treat.* **158**, 113–126 (2016).
62. Charan, M. et al. Molecular and Cellular Factors Associated with Racial Disparity in Breast Cancer. *Int. J. Mol. Sci.* **21**, https://doi.org/10.3390/ijms21165936 (2020).
63. Marczyk, M. et al. Tumor immune microenvironment of self-identified African American and non-African American triple negative breast cancer. *NPJ Breast Cancer* **8**, 88 (2022).
64. Pe, K. C. S. et al. Triple-negative breast cancer influences a mixed M1/M2 macrophage phenotype associated with tumor aggressiveness. *PLoS ONE***17**, e0273044 (2022).
65. Vernaci, G. et al. Characterization of gut microbiome composition in patients with triple-negative breast cancer treated with neoadjuvant chemotherapy. *Oncologist* **28**, e703–e711 (2023).
66. Sohail, S. & Burns, M. B. Integrating current analyses of the breast cancer microbiome. *PLoS One* **18**, e0291320 (2023).
67. Davis, M. et al. Identification of distinct heterogenic subtypes and molecular signatures associated with african ancestry in triple negative breast cancer using quantified genetic ancestry models in admixed race populations. *Cancers* **12**, https://doi.org/10.3390/cancers12051220 (2020).
68. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
69. Pertea, M., Kim, D., Pertea, G. M., Leek, J. T. & Salzberg, S. L. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat. Protoc.* **11**, 1650–1667 (2016).
70. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
71. Kostic, A. D. et al. PathSeq: software to identify or discover microbes by deep sequencing of human tissue. *Nat. Biotechnol.* **29**, 393–396 (2011).
72. Walker, M. A. et al. GATK PathSeq: a customizable computational tool for the discovery and identification of microbial sequences in libraries from eukaryotic hosts. *Bioinformatics* **34**, 4287–4289 (2018).
73. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).

Acknowledgements

This work was funded, in part, by National Institutes of Health (NIH) grants, including a Career Development Supplement (3U54CA118948-14S2) and an R01 award from the National Cancer Institute (NCI) (R01CA279323-02) as well as by the Avon Foundation for Women, the Siegfried and Marianne Endlichhofer Trust, and Medical College of Wisconsin start-up funds provided to A.I.O.; U.M. (5U54CA118948) and C.Y. (5U54CA118623) are partly supported by NIH/NCI awards.

Author contributions

Conceptualization, R.K., S.D.A., U.M. and A.I.O.; methodology, R.K., S.D.A., A.S., P.L.A., J.R.K. and A.I.O.; software, R.K., A.S., V.S. and A.I.O.; validation, R.K., A.S. and A.I.O.; formal analysis, R.K., A.S. and A.I.O.; investigation, R.K., S.D.A., A.S. and A.I.O.; resources, R.K., C.S.P., M.B., D.S.C., I.E.E., S.V., H.K.T., S.S., C.Y., U.M., and A.I.O.; writing-original draft preparation, R.K., S.D.A., and A.I.O.; writing, review and editing, R.K., S.D.A., A.S., V.S., C.S.P., M.B., D.S.C., I.E.E., S.V., H.K.T., S.S., P.L.A., L.N.C., J.R.K., C.Y., U.M. and A.I.O.; visualization, R.K.; supervision, U.M. and A.I.O.; All authors have read and agreed to the published version of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41523-025-00806-y.

Correspondence and requests for materials should be addressed to Upender Manne or Akinyemi I. Ojesina.