



## OPEN ACCESS

## EDITED BY

Muhammad Tahir Ul Qamar,  
Government College University,  
Faisalabad, Pakistan

## REVIEWED BY

Nahum Puebla-Osorio,  
University of Texas MD Anderson Cancer  
Center, United States  
Fatima Noor,  
University of Lahore, Pakistan

## \*CORRESPONDENCE

Jessica Galloway-Peña  
✉ jgallowaypena@cvm.tamu.edu

RECEIVED 15 May 2025

ACCEPTED 31 July 2025

PUBLISHED 21 August 2025

## CITATION

McMahon S, Franklin S and Galloway-Peña J  
(2025) Utilization of machine learning to  
predict antibiotic resistant event outcomes  
in acute myeloid leukemia patients  
undergoing induction chemotherapy.  
*Front. Cell. Infect. Microbiol.* 15:1629422.  
doi: 10.3389/fcimb.2025.1629422

## COPYRIGHT

© 2025 McMahon, Franklin and  
Galloway-Peña. This is an open-access article  
distributed under the terms of the [Creative  
Commons Attribution License \(CC BY\)](#). The  
use, distribution or reproduction in other  
forums is permitted, provided the original  
author(s) and the copyright owner(s) are  
credited and that the original publication in  
this journal is cited, in accordance with  
accepted academic practice. No use,  
distribution or reproduction is permitted  
which does not comply with these terms.

# Utilization of machine learning to predict antibiotic resistant event outcomes in acute myeloid leukemia patients undergoing induction chemotherapy

Stephanie McMahon, Samantha Franklin  
and Jessica Galloway-Peña\*

Laboratory of Jessica Galloway-Peña, Texas A&M University, Department of Veterinary Pathobiology,  
Interdisciplinary Graduate Program in Genetics and Genomics, College Station, TX, United States

**Introduction:** Acute myeloid leukemia (AML) patients are highly susceptible to infection. Moreover, prophylactic and empirical antibiotic treatment during chemotherapy disrupts the gut microbiome, raising the risk for antibiotic-resistant (AR) opportunistic pathogens. There is limited data on risk factors for AR infections or colonization events in treated cancer patients, and no predictive models exist. This study aims to combine metagenomic and antibiotic administration data to develop a model predicting AR event outcomes.

**Methods:** Baseline stool microbiome, antibiotic administration, resistome, and clinical metadata from 95 patients were utilized to build a Random Forest model to predict AR infection and colonization events by serious AR threats. Additionally, sparse canonical correlation analysis assessed correlations between microbiome and resistome data, while Spearman correlation networks identified direct associations with AR event outcomes and secondary variables.

**Results:** AR-events were identified in 14 of the 95 included patients, with 8 developing AR infections and 9 identified as AR colonized. A Random Forest model predicted AR event outcomes (AUC = 0.73), identifying bacterial taxa and antibiotic resistance gene (ARG) classes as key variables of importance. *Methanobrevibacter smithii*, *Clostridium leptum*, and *Bacteroides dorei* were identified as key taxa associated with reduced risk of AR events, suggesting the potential roles of commensals in maintaining gut microbial resilience during chemotherapy. ARG classes, particularly those conferring resistance to lincosamides, macrolides, and streptogramins, were negatively associated with AR events.

**Conclusion:** These results underscore the value of integrating microbiome and resistome features to reveal potential protective mechanisms and improve risk prediction for AR outcomes in vulnerable patients.

## KEYWORDS

antibiotic resistance, acute myeloid leukemia, random-forest, microbiome, resistome

# 1 Introduction

Infections caused by multidrug-resistant organisms (MDROs) represent a major public health challenge, emphasizing the urgent need to understand the underlying factors contributing to antimicrobial resistance (Caniça et al., 2015; Castro-Sánchez et al., 2016; Palmore and Henderson, 2013). This understanding is essential for reducing infection rates and preserving the effectiveness of existing antimicrobial therapies (Caniça et al., 2015; Castro-Sánchez et al., 2016; Palmore and Henderson, 2013). The growing prevalence of antibiotic-resistant bacteria jeopardizes the effectiveness of commonly used antibiotics, complicating the treatment of an increasing number of infections and necessitating immediate action (Antimicrobial resistance, 2023; Walsh et al., 2023). The Centers for Disease Control and Prevention (CDC) has identified several pathogens as “urgent,” “serious,” and “concerning” antimicrobial-resistant (AR) threats in their 2019 Antibiotic Resistance Threats report (CDC, 2019). Among these, five pathogens are categorized as urgent threats, including methicillin-resistant *Staphylococcus aureus* (MRSA), multidrug-resistant *Pseudomonas aeruginosa*, vancomycin-resistant *Enterococcus* (VRE), extended-spectrum beta-lactamase-producing *Enterobacteriaceae* (ESBL), and carbapenem-resistant *Enterobacteriaceae* (CRE) (CDC, 2019). If left unrestrained, AR pathogens are projected to become the leading cause of death by 2050 (O’Neill, 2016).

Increasing evidence has underscored the essential role of the microbiome in defending against colonization and infection by antibiotic-resistant pathogens (Buffie and Pamer, 2013; Lewis et al., 2015; Ubeda et al., 2013). In addition to the microbiome defending against AR pathogens through colonization resistance, it also plays a role in immunomodulatory functions that influence infections in distant body sites (Belkaid and Hand, 2014; Latorre et al., 2015; Lewis et al., 2015; Panwar et al., 2021; Pickard et al., 2017; Thaiss et al., 2016). The widespread use of broad-spectrum antibiotics exerts selective pressure on microbial populations, facilitating the emergence of resistant strains (Chinemerem Nwobodo et al., 2022). Specifically, antibiotic treatment can disrupt the gut microbiota, leading to reduced microbial diversity, depletion of beneficial species, and the promotion of antibiotic-resistant gene (ARG) proliferation (Becattini et al., 2016; Gibson et al., 2015; Korpela et al., 2016; Modi et al., 2014). Moreover, overuse of antibiotics is known to not only facilitate the growth of MDROs, but also promote the horizontal transfer of resistance genes among organisms within the microbiota (Korrry et al., 2020). This transfer can lead to the emergence of new multidrug-resistant pathogens, potentially restricting treatment options for bacterial infections. As a result, studies that explore the protective and harmful roles of the microbiota, the resistome (the collection of antibiotic resistance genes within the microbiota), and the specific impacts of antibiotic treatments on microbial communities contributing to the rise of antibiotic resistance are vital for developing new strategies to combat antimicrobial resistance (Caniça et al., 2015).

Patients with hematological malignancies are especially vulnerable to a wide range of infections, particularly those caused

by AR pathogens (Gedik et al., 2014). Life-threatening infections are common, particularly in individuals who are immunocompromised with chemotherapy-induced cytopenia (Rapoport et al., 2021). Febrile neutropenia is especially prevalent in patients with acute myeloid leukemia (AML), affecting roughly 80–90% of individuals (Hansen et al., 2019; Taur and Pamer, 2016). To manage these risks, fluoroquinolone prophylaxis is routinely used before and during chemotherapy, and broad-spectrum empirical therapy is administered upon the onset of neutropenic fever following established standards of care (Boccia et al., 2022; Chemotherapy-Induced Neutropenia and Febrile Neutropenia in the US: A Beast of Burden That Needs to Be Tamed? | The Oncologist | Oxford Academic, n.d.; X. Wang et al., 2024; Zimmer and Freifeld, 2019). However, the prolonged use of antibiotics in these patients has been shown to disrupt their gut microbiota, leading to the accumulation of antimicrobial resistance genes (ARGs) (Aitken et al., 2021; Doan et al., 2020; Galloway-Peña et al., 2016; J. Galloway-Peña et al., 2017; Galloway-Peña et al., 2020; Iwan et al., 2024; Jutkina et al., 2016; Nobel et al., 2015). Chemotherapy also exacerbates the risk of AR infection by damaging the gut mucosa, increasing intestinal permeability, and facilitating the translocation of potentially resistant pathogens from the gut into the circulation, thereby heightening the risk of difficult-to-treat systemic infections (Sougiannis et al., 2021; Touchefeu et al., 2014).

The gut resistome is of particular significance in hospitalized cancer patients, as it serves as a key reservoir of ARGs that should no longer be overlooked when examining AR infectious complications (Gibson et al., 2014; Shono et al., 2016). The gut microbiome contains a large pool of ARGs, and these genes can be transferred among bacterial species within the microbiota via horizontal gene transfer, which presents potentially serious consequences for infections originating from the microbiome as well as transmission in the hospital environment (Forslund et al., 2013; Gibson et al., 2015; Pehrsson et al., 2013). Bacteria may carry ARGs that confer resistance to a single antibiotic or mobile genetic elements (MGEs) that provide resistance to multiple antibiotics. As the selection pressure for these bacteria increases, so does the number of bacteria harboring ARGs or MGEs, amplifying the resistome and resulting in more difficult-to-treat infections (Nielsen et al., 2021). Common bacterial culprits in infections among AML patients include the same key AR-threat pathogens identified by the CDC (VRE, MRSA, ESBL-producing *Enterobacteriaceae*, CRE, and multidrug-resistant *P. aeruginosa*) (On et al., 2022; Rolston, 2014). Although it is well-documented that cancer patients are frequently colonized and infected by AR pathogens, there is still a paucity of comprehensive data regarding the risk factors associated with AR infections and toxicities during chemotherapy (McMahon et al., 2023; Nanayakkara et al., 2021). This gap in knowledge underscores the need for further research to better understand the complex dynamics of the microbiome and AR infections in this vulnerable patient population.

In this study, we developed a machine learning model integrating patient antibiotic administration records with baseline fecal microbiome and resistome data. This model allowed us to identify and rank predictors associated with AR events in AML

patients undergoing remission induction chemotherapy (IC). Additionally, sparse canonical correlation analysis was used to determine correlations between baseline ARGs and microbial taxa. A final network analysis was performed to identify variable connectivity with development of an AR event, and the directionality in which those variables relate to the event.

## 2 Methods

### 2.1 Study design and participants

Longitudinal stool samples and clinical data were collected from two cohorts of adult AML patients undergoing IC at MD Anderson Cancer Center (MDACC) between September 2013 and February 2020, for a total of 154 patients. The first cohort, PA13-0339, comprised 98 AML patients enrolled between September 2013 and August 2015, providing 566 stool samples collected as previously described (Galloway-Peña et al., 2016, 2020; J. R. Galloway-Peña et al., 2017). The second cohort, PA15-0780, included 56 adult AML patients enrolled from January 2015 to February 2020, contributing 216 stool samples. For this cohort, the stool sample taken within approximately one week of the start of IC was considered the baseline sample. Longitudinal samples were collected twice a week for the first four weeks, weekly for weeks four to eight, every other week for weeks eight to twelve, and then every two weeks after, continuing until either 24 weeks or loss of follow-up. Patients with missing baseline stool samples were excluded from the study. Between both cohorts, the average and median time of collection for a baseline sample was one day prior to chemotherapy initiation. A histogram depicting the time from baseline stool sample collection to initiation of chemotherapy is shown in [Supplementary Figure 1](#). Patients from these two cohorts were only included in the analyses and model if they had complete metadata and sequencing data, which included baseline stool metagenomics, baseline gut resistome data, complete antibiotic administration data for the entirety of the study period, and the clinical factors of gender, chemotherapy type, and chemotherapy intensity available. This left 95 total patients with all data available to be included in the study.

### 2.2 16S rRNA sequencing and analyses of the stool samples

Genomic DNA was extracted from longitudinal stool samples using the QIAamp Fast DNA Stool Mini Kit (Qiagen), with modifications to the standard protocol that included an additional bead-beating lysis step. Each stool sample was placed in a tube containing a 3.2-mm steel bead, approximately 150 mg of zirconium beads, and lysis buffer. The samples were then homogenized using a bead-beater at 3800 rpm for 8 minutes (BioSpec) to facilitate DNA isolation. Amplicon libraries targeting the 16S V4 region were generated, and Illumina MiSeq sequencing was conducted on the fecal microbial DNA using a 2 × 250 bp

paired-end protocol (Galloway-Peña et al., 2016, 2020; J. R. Galloway-Peña et al., 2017). A no-template control was used during the PCR, and a no-sample control for the extraction. The resulting reads were merged, dereplicated, and length-filtered using VSEARCH. Denoising and chimera detection were performed with the UNOISE3 commands, and the unique sequences, also known as zero-radius Amplicon Sequence Variants (ASVs), were taxonomically classified using Mothur with the SILVA database (version 138) (Edgar, 2016; Schloss et al., 2009). Alpha and beta diversity metrics were calculated in QIIME 2. The 16S rRNA sequences from the PA13 cohort have been previously published and are available in the NCBI Sequence Read Archive under Bioproject IDs PRJNA352060 and PRJNA526551 (Galloway-Peña et al., 2016, 2020; J. R. Galloway-Peña et al., 2017). The 16S rRNA sequences from the PA15 cohort are deposited in the NCBI Sequence Read Archive under Bioproject number PRJNA1124986.

### 2.3 AR event identification

All longitudinal stool samples underwent 16S rRNA sequencing to identify those with ≥ 3% of their 16S rRNA reads mapping to genera associated with the CDC urgent threat antibiotic-resistant (AR) pathogens, including vancomycin-resistant Enterococci (VRE), carbapenem-resistant Enterobacteriaceae (CRE), extended-spectrum beta-lactamase-producing Enterobacteriaceae (ESBL), multidrug-resistant *Pseudomonas aeruginosa* (MDRP), and methicillin-resistant *Staphylococcus aureus* (MRSA). The 3% threshold was chosen as it was previously determined that patients with less than 3% of 16S reads mapping to an AR-threat-associated genera had very little likelihood of obtaining a positive culture of an AR pathogen on selective and differential media (McMahon et al., 2023). Stool samples with >3% of 16S rRNA reads mapping to Enterobacteriaceae, *Escherichia*, *Enterobacter*, *Acinetobacter*, *Klebsiella*, or *Pseudomonas* were then cultured on CRE or ESBL selective media (Hardy Diagnostics Cat. No G323 and G321). Samples with >3% reads of *Enterococcus* were plated on VRE media (Chromagar Cat. No VR952), while those with *Staphylococcus* reads were streaked on MRSA selective media (Hardy Diagnostics Cat. No G307BX). Any colonies that grew on the selective media were then sub-cultured onto BBL Trypticase Soy Agar with 5% Sheep Blood (BD Biosciences) for isolation of individual colonies, which were stored at -80°C.

Matrix-Assisted Laser Desorption/Ionization Time-Of-Flight Mass Spectrometry (MALDI-TOF) was used to identify bacterial species from the purified colonies (Bruker MALDI Biotyper). After species identification, antibiotic susceptibility testing (AST) was performed using the VITEK2 system (Biomérieux). The AST-GN69 and AST-XN06 cards were used for Gram-negative bacteria, while AST-GP75 was used for Gram-positive isolates. If any bacterial isolate from a stool sample was confirmed to be CRE, ESBL-producing Enterobacteriaceae, VRE, MDRP, or MRSA, the patient was classified as having confirmed AR colonization (ARC).

In addition, any infectious bacterial isolates identified by the clinical microbiology lab at MDACC during the study, while AML

patients were enrolled, were stored at  $-80^{\circ}\text{C}$ . The clinical microbiology lab was responsible for determining the bacterial species and antimicrobial susceptibilities of these isolates. If a patient was diagnosed by the lab with a microbially defined infection caused by CRE, ESBL-producing Enterobacteriaceae, VRE, MDRP, or MRSA, the patient was classified as having a confirmed AR infection (ARI). These two groups—patients with confirmed ARC and those with ARI—were combined and classified as having an “AR event” for this study.

## 2.4 Whole genome sequencing of AR-event bacterial isolates and shotgun metagenomic sequencing of baseline samples

Following species identification, DNA was extracted from the ARI or ARC bacterial isolates using the MasterPure Gram-positive DNA purification kit (Lucigen). The extracted DNA was then used to prepare sequencing libraries with the Illumina DNA Tagmentation Library Prep kit (Illumina, San Diego, CA, USA). To evaluate the quality of the prepared libraries, the Qubit bioanalyzer and the Qubit dsDNA HS Assay Kit (Invitrogen, Waltham, MA, USA) were utilized. Once assessed, the libraries were pooled and sent to the North Texas Genome Center for sequencing on the Illumina NovaSeq 6000 S4 flow cell, using a 150-base pair paired-end read protocol. Subsequent sequencing data analysis was conducted on the Grace computing cluster at Texas A&M University. The sequencing reads were down-sampled to six million read pairs per sample using Seqtk (v1.3) before being assembled with SPAdes (v3.14.1) under the “—isolate” parameter. Annotation of the assembled sequences was performed using the RAST toolkit-based pipeline on BV-BRC (<https://www.patricbrc.org/>). The bacterial isolate sequencing data have been archived in the NCBI Sequence Read Archive under Bioproject ID PRJNA1129516.

For baseline fecal DNA extraction for metagenomic purposes, a modified version of the Qiagen Blood and Tissue kit (Qiagen, Valencia, CA, USA) was employed. Each sample, consisting of 150 mg of frozen fecal material, was combined with 500  $\mu\text{L}$  of sterile InhibitEx Buffer (Qiagen, Valencia, CA, USA), 150 mg of silicon beads (Lysing Matrix B, MP Biomedical), and an appropriate amount of 2.4 mm metal beads from a hard tissue grinding mix (VWR, Radnor, PA, USA). The samples were placed in sterile tubes and subjected to bead beating at 4.5 m/s for four minutes using the MP Biomedical FastPrep-24 Classic system. The resulting suspension was vortexed, heated to  $95^{\circ}\text{C}$  for seven minutes, and centrifuged at 15,000 rpm for three minutes. DNA was extracted from the supernatant by adding 20  $\mu\text{L}$  of Proteinase K and Buffer AL, followed by vortexing. After incubation at  $70^{\circ}\text{C}$  for 30 minutes, 200  $\mu\text{L}$  of 100% ethanol was introduced, and the mixture was inverted multiple times before being transferred to a DNeasy spin column for purification. The extracted DNA was stored at  $-20^{\circ}\text{C}$  for further analysis.

For shotgun metagenomic sequencing of baseline stool samples, DNA libraries were prepared using the Illumina Nextera XT DNA Library Prep kit (Illumina, San Diego, CA, USA). The quality of the libraries was confirmed using the Qubit bioanalyzer (Invitrogen, Waltham, MA, USA). Once validated, the libraries were pooled and sequenced at the North Texas Genome Center using the Illumina NovaSeq 6000 S4 platform, following the same protocol as the bacterial isolates. Samples were down-sampled to 6 million read pairs to normalize sequencing depth across subjects and reduce potential batch-related artifacts. This depth is sufficient for reliable class-level detection of ARGs in gut metagenomes, particularly when using curated HMM-based classifiers (BugSeq). A no-sample control was utilized for the extraction. Spike-in standards were not used in this study. Batch effects were minimized using consistent extraction protocols, library preparation methods, and sequencing conditions. Assembly of shotgun metagenomic sequences was performed using MEGAHIT (v1.2.8) and metaSPAdes (v3.14.1). The relative abundance of bacterial taxa was determined using MetaPhlAn2 (v2.8.1). The assembled shotgun sequences were further binned and annotated through the RAST Binning Service (RBS) on PATRIC. The finalized shotgun metagenomic data have been deposited in the NCBI Sequence Read Archive under Bioproject IDs PRJNA1129514 and PRJNA1128111.

## 2.5 Resistome analyses

Taxonomic binning of assembled sequences from shotgun metagenomic data was performed using BugSeq (v4.0), following previously established methods (Chandrasekhar et al., 2022; Fan et al., 2021; Gauthier et al., 2022). Contigs were aligned against a curated reference sequence database using minimap2, and alignment results were evaluated based on query coverage and average nucleotide identity (ANI) thresholds to determine taxonomic classification. Antimicrobial resistance (AMR) determinants were identified by screening taxonomic bins with BugSeqs AMR analysis (v4.0). Contigs were analyzed against a curated protein database containing over 6,500 sequences associated with AMR, using both threshold alignment and Hidden Markov Models to accurately classify gene alleles and families. Additionally, taxon-specific models for phenotypic AMR prediction, incorporating single-nucleotide variants, insertions, deletions, and other genetic markers of resistance, were applied to over 50 bacterial species. The resulting taxonomic binning data and AMR predictions were compiled into comprehensive reports for further analysis. Antibiotic resistance genes (ARGs) in this study were examined at the class level.

## 2.6 Machine learning/random forest model

A machine learning model was developed by incorporating baseline microbiome species abundances via shotgun metagenomic sequencing, antibiotic resistance gene (ARG) data, antibiotic administration data, for time from baseline sample to chemotherapy



start date, and binary clinical variables such as sex, chemotherapy type, and chemotherapy intensity to predict AR-threat outcomes. The model employed a Random Forest algorithm, which leverages multiple decision trees to reach an outcome. This approach, an extension of the bagging method, incorporates both bagging and feature randomness to create a “forest” of decision trees. Several patients were excluded from the study. Reasons for exclusion included: lack of baseline stool sample, lack of antibiotic administration data, lack of complete clinical metadata, and lack of resistome data. This left 95 patients with complete metadata to be included in the study (Table 1). Of these, 14 patients experienced an AR event during induction chemotherapy, defined either ARC or ARI events. Antibiotic administration variables were only included in the model if they were received by at least 10% of patients (9 or more patients). Similarly, ARG classes from resistome data were only included if they were present in at least 10% of patients. Baseline microbiome data was filtered to an abundance of 0.001 and a prevalence of 10%.

To minimize dataset variability, overfitting, bias and to enhance robustness, we ran four different models. For each model, analysis was performed across 100 independent 80–20 stratified training-testing splits. Model 1 contained 89 variables and featured ARG class by presence/absence, presence/absence of antibiotic administrations received  $\geq 72$  hours, and baseline species abundance from shotgun metagenomic sequencing. Model 2 also contained 89 variables but featured unique ARG counts by class, presence/absence of antibiotic administrations received  $\geq 72$  hours, and baseline species abundance. Model 3 contained 92 variables with ARG class by presence/absence, presence/absence of any administration (regardless of duration), and baseline species abundance. Finally, Model 4 also contained 92 variables, with unique ARG counts by class, presence/absence of any antibiotic administration, and baseline species abundance. Antibiotic administration was evaluated at a threshold of  $\geq 72$  hours of administration or as the presence/absence of any administration, regardless of duration, as both are biologically relevant for infectious implications in this patient cohort. AML patients receiving IC may receive empirical broad-spectrum antibiotic treatment, which might be de-escalated upon return of negative cultures ( $\sim 48$ hrs) or switched to a separate antibiotic with the correct spectrum of activity if the pathogen is found resistant to the prescribed antibiotic. Thus, presence/absence would account for any antibiotic given empirically or otherwise, whereas a model that considers antibiotic administration  $\geq 72$  hours is relevant for antibiotics that were maintained/continued for an infectious implication (AR or not) after return of culture or continuation of symptomology.

The resulting outcomes of each iteration were then aggregated, either as ROC curves or box plots for each model. Variable importance scores for each of the 100 iterations were calculated and averaged. To better amend the model's performance, the top 20, 15, and 10 most influential variables were selected for inclusion for each of the four models. The model with the 15 most influential variables demonstrated the best performance in comparison to 10 or 20 in all four models. Each model using the top 15 most influential variables was then executed 100 times, and the aggregated results were used to produce a final AUC-ROC curve. The variable importance scores were calculated as the raw sum of decreases in Gini impurity across all trees in the

Random Forest and were not normalized. The Youden's index was calculated as sensitivity plus specificity minus one to summarize the overall accuracy of the diagnostic test. A final model (Model 2) was chosen based on the 1) highest AUC, 2) best balance of sensitivity and specificity, and 3) a balanced Youden's index. The final model was optimized by increasing the value of *mtry* to increase the number of variables randomly sampled in each try and increasing the node size to increase the depth of each tree. We then used Shapely Additive exPlanations (SHAP) to interpret the optimized machine learning model.

## 2.7 Sparse canonical correlation analysis

Sparse canonical correlation analysis (sCCA) was utilized to detect ARG class-microbial species associations by identifying highly correlated linear combinations of variables while ensuring sparsity through variable-specific weight constraints. A centered log-ratio (CLR) transformation was applied to the microbial abundance data, with zero values substituted by a pseudocount calculated as  $\min(\text{relative abundance})/2$ . Sparse Canonical Correlation Analysis (sCCA) was performed using the ‘PMA’ package (v. 1.2-2) in R, correlating the CLR-transformed gut microbiome taxonomic composition with the ARG classes. Hyperparameters were optimized using the CCA.permute function (*nperms* = 100, *niter* = 5) prior to model fitting. To enforce sparsity, a Lasso penalty was applied, with the “*typex*” and “*typez*” parameters set to “standard” for the corresponding canonical vector (Rashidi et al., 2022).

## 2.8 Network development and selection of variables

To investigate the interactions between classes of ARGs by count of unique genes within each class, antibiotic administration by presence/absence, microbial taxa abundance at the species level, and AR events, a network was constructed based on a correlation matrix derived from the sample data. The matrix was created using Spearman's rank correlation test, with positive values indicating a positive correlation to the AR event and negative values indicating a negative correlation (Spearman, 1904). The resulting correlation matrix was imported into Gephi, where the network was visualized using its built-in functions (Bastian et al., 2009). Following the initial analysis, edges with a weight below 0.2 were removed, resulting in the identification of 11 variables associated with AR events. The nodes and edges were then uploaded to generate the network.

# 3 Results

## 3.1 Patient characteristics

During their inpatient treatment, 14 patients experienced an AR event. Of these, 8 patients were identified with antibiotic-resistant

TABLE 1 Characteristics of the patients included for AR event analysis.

Patient characteristics	AR-event			
	Yes	No	P-value	$\chi^2$
Patient count, N (%)	14 (14.7)	81 (85.3)		
Infection	8			
Colonization	9			
Cohort, N (%)				
PA13-0339	6 (6.3)	56 (58.9)		
PA15-0780	8 (8.4)	25 (26.3)		
Sex, N (%)			0.592	0.288
Female	8 (57.1)	40 (49.4)		
Male	6 (42.9)	41 (50.6)		
Chemotherapy intensity, N (%)			0.66	0.193
High	9 (64.3)	47 (58.0)		
Low	5 (35.7)	34 (42.0)		
Chemotherapy type, N (%)			0.83	0.882
Fludarabine <sup>2</sup>	4 (28.6)	18 (22.2)	0.732	–
Non-Fludarabine (High) <sup>2</sup>	5 (35.8)	29 (35.8)	>.99	–
Hypomethylators <sup>2</sup>	4 (28.6)	21 (25.9)	>.99	–
Other (Low) <sup>2</sup>	1 (7.0)	13 (16.1)	0.685	–
Antibiotic administration <sup>1</sup>			0.843	8.815
Levofloxacin	8 (57.1)	63 (77.8)	0.101	2.692
Metronidazole <sup>2</sup>	3 (21.4)	10 (12.3)	0.4	–
Linezolid <sup>2</sup>	13 (92.9)	57 (70.4)	0.104	–
Meropenem	7 (50)	34 (42)	0.576	0.313
Cefepime	9 (64.3)	49 (60.5)	0.788	0.072
Amikacin	5 (35.7)	11 (13.6)	0.041	4.176
Trimethoprim-Sulfamethoxazole <sup>2</sup>	0 (0)	9 (11.1)	0.3475	–
Daptomycin	5 (35.7)	22 (27.2)	0.512	0.429
Piperacillin-Tazobactam	6 (42.9)	22 (27.2)	0.234	1.415
Cefpodoxime	5 (35.7)	23 (28.4)	0.579	0.308
Ciprofloxacin <sup>2</sup>	2 (14.3)	16 (19.8)	>.99	–
Tigecycline <sup>2</sup>	4 (28.6)	13 (16.0)	0.269	–
Minocycline <sup>2</sup>	3 (21.4)	9 (11.1)	0.377	–
Ertapenem <sup>2</sup>	2 (14.3)	16 (19.8)	>.99	–

<sup>1</sup>Antibiotic Administration is defined as any administration of the antibiotic.

<sup>2</sup>Fisher's Exact test was used.

infections, with one patient exhibiting two distinct infections. Among the infections, 37.5% were attributed to carbapenem-resistant *Pseudomonas aeruginosa*, 12.5% to MRSA, and 62.5% to ESBL *Escherichia coli*. Furthermore, 9 patients were identified as being GI-colonized, with 22.2% of those colonized with VRE, 22.2% with carbapenem-resistant and ESBL-producing *E. cloacae*, 22.2%

with ESBL-producing *Klebsiella pneumoniae*, 33.3% with MRSA, and 22.2% ESBL-producing *E. coli*. Of these 14 total AR event patients, 3 patients had overlapping calls, where they had both ARC and ARI (Supplementary Table 1). Analysis of clinical variables demonstrated that sex, chemotherapy type, and chemotherapy intensity did not show any significant differences between AR

event groups (Table 1). The only variable that showed a statistical difference between the two groups was amikacin administration ( $p=0.041$ ) (Table 1). These results ultimately show that there are no clinical confounding variables between the two AR event groups.

## 3.2 Development of model

To develop a model capable of predicting AR event outcomes, a Random Forest-based approach was employed to determine the optimal model configuration and identify the variables that most significantly contributed to the model's performance. The patient data were randomly partitioned into a training set (80% of the samples) and a testing set (20% of the samples). Each model was trained over 100 independent iterations, each with 500 trees. Variable importance scores were calculated for each feature, and each model's performance was assessed using the area under the curve (AUC) and Youden's index, which calculates the maximum vertical distance between the curve of true positives and false positive rates on an ROC curve. After running each model with all variables (89–92 variables depending on the model), the top 15 variables for each model were selected based on their importance scores and the models subsequently re-run with only the 15 retained variables (80/20 split, 100 iterations, 500 trees each). The aggregated results were used to produce a final AUC-ROC curve for each model (Figure 1). Based on antibiotic administration data and antibiotic-resistant gene (ARG) data from BUGSEQ, we constructed four distinct model variations to assess which data configuration produced the most robust results. Model 1 incorporated ARG class by presence/absence, presence/absence of antibiotics administered  $\geq 72$  hours, and baseline taxa abundance by species (AUC=0.667, Youden's=0.879) (Figure 1A). Model 2 included ARG data by count of unique genes by class, presence/absence of antibiotics administered for  $\geq 72$  hours, and baseline species abundances (AUC=0.689, Youden's=0.835) (Figure 1B). Model 3 utilized ARG class presence/absence data, presence/absence of any antibiotic administered (regardless of duration), and baseline species abundance (AUC=0.666, Youden's =0.871) (Figure 1C). Model 4 comprised ARG data by count of unique genes by class, presence/absence of any antibiotic administered, and baseline abundance data by species (AUC=0.687, Youden's=0.855) (Figure 1D). Further clinical metrics and confusion matrices for each model can be found in Supplementary Tables 2 and 3, respectively.

We determined that Model 2 was the best choice among the four models as it offers the strongest overall performance balance. It has the highest AUC (0.689), indicating the best discrimination between positive and negative cases. While its sensitivity (0.630) is slightly lower than the others, it achieves the highest specificity (0.643) among the models. Moreover, although models 1, 3, and 4 had higher sensitivity, their specificity was not much above 50%. Having determined that Model 2 was the most balanced model, the model was optimized, yielding an aggregate AUC value of 0.730 (CI: 0.0251), and a Youden's index of 0.863 (Figure 2A). The aggregate optimized model had a mean AUC value of 0.742, a median AUC

value of 0.745, and a range from 0.505 to 0.969 for the 100 iterations (Supplementary Figure 2).

Of the 15 variables utilized in this model, 13 were bacterial taxa, with *Methanobrevibacter smithii* (mean variable importance score (mVIS) of 0.717) and *Blautia producta* (mVIS of 0.630), being the most contributory to the model. Two classes of ARGs were also contributory to the model, those being genes predicted to encode resistance to macrolides (mVIS of 0.573) and genes classified as those predicted to encode resistance to lincosamides, macrolides, and streptogramin (mVIS of 0.542). The additional 11 variables consisted of bacterial taxa, with mean VIS scores ranging from 0.522 to 0.622 (Figure 2B). The SHAP method was used to explain the contribution or importance of each of the 15 features on the prediction of the model for AR events (Figure 2C). The predominant clustering of SHAP values below zero for each variable indicates that these features tend to have a negative impact on the prediction of an AR event, supporting a non-event classification. Specifically, higher values of ARG counts of genes belonging to macrolide and lincosamide/macrolides/streptogramin classes contributed negatively to the model. It also appeared that the higher feature values of *M. smithii*, *Clostridium leptum*, and *Bacteroides dorei* contributed negatively towards the model, which were the top three features with the greatest effect on prediction, according to the mean SHAP value. Conversely, it appeared that higher feature values of baseline abundances of *Bifidobacterium longum*, *Ruminococcus gnavus*, *Eubacterium rectale*, and *Parabacteriodes* (unclassified), contributed positively to the model.

## 3.3 Sparse canonical correlation analysis

To better understand the relationship between baseline species abundances and antibiotic resistance gene (ARGs) classes in patients at baseline, a sCCA analysis was conducted to explore the associations between these two groups. A heatmap was generated using Pearson correlation coefficients (Supplementary Figure 3). Notably, only 2 ARG classes and 10 taxa were found to be correlated between baseline gut microbiota species and classes of ARGs. Hierarchical clustering revealed four bacterial clusters, with two that were particularly noteworthy. The first cluster included *Streptococcus thermophilus* and *Bacteroides thetaiotaomicron*, which were negatively correlated with genes encoding resistance to glycopeptides and lincosamides/macrolide/streptogramins. In contrast, the second cluster revealed that *C. leptum* and *Ruminococcus torques* exhibited a positive correlation with genes conferring resistance to glycopeptides and lincosamide/macrolides/streptogramins.

## 3.4 Network development

Thus, to further comprehend how the variables relate to the development of an AR event, a network analysis was conducted to explore these correlations. After constructing a correlation matrix with a minimum correlation threshold of 0.2, 11 variables were

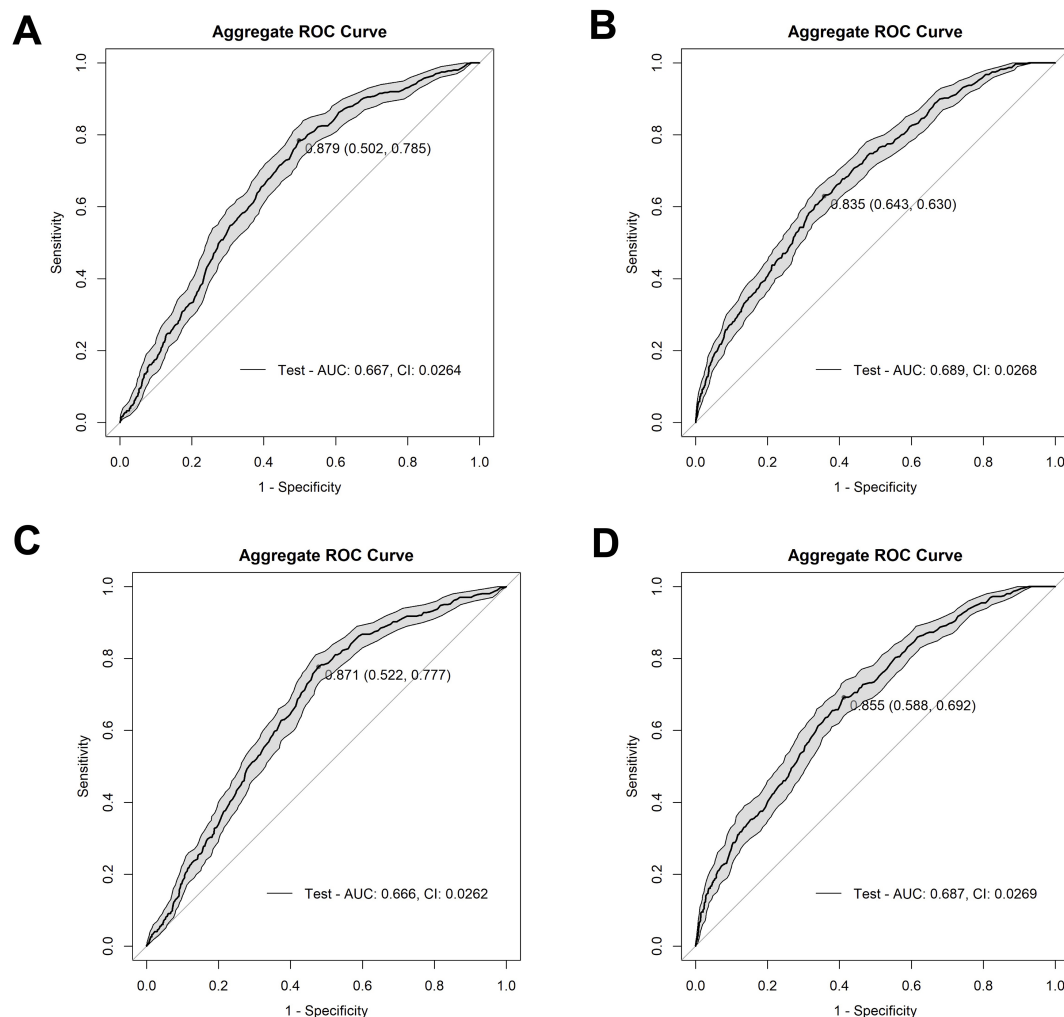


FIGURE 1

Receiver operating characteristic (ROC) curves depicting the performance of our four model methods. (A) depicts Model 1 (AUC=0.667) with ARG class by presence vs absence, presence/absence of antibiotics at  $\geq 72$  hours' worth of administrations, and baseline species abundance, (B) depicts Model 2 (AUC=0.689) with unique ARG counts by class, presence/absence of antibiotics  $\geq 72$  hours of administrations, and baseline species abundance, (C) depicts Model 3 (AUC=0.666) with ARG class by presence vs absence, antibiotic administration presence/absence of any administration, and baseline species abundance, and (D) depicts Model 4 (AUC=0.687) with unique ARG counts by class, antibiotic presence/absence of any administration, and baseline species abundance. The shaded grey area on the graph shows confidence intervals for all models. The Youden's Index, a metric that identifies the optimal threshold on a ROC curve, is indicated by the dot on the curve, followed by the specificity and sensitivity values, respectively.

identified as linked to the development of an AR event and visualized in a network (Figure 3; Supplementary Table 4). The network analysis suggests that a patient's baseline taxa might play a critical role in protecting against an AR event, as all 11 variables were negatively correlated with AR event occurrence.

## 4 Discussion

In this study, we utilized a novel machine learning approach to predict antibiotic-resistant event outcomes in patients undergoing induction chemotherapy for acute myeloid leukemia. The developed Random Forest model integrated baseline fecal microbiome composition, antibiotic administration data, and resistome profiles, successfully identifying critical microbial taxa

and antibiotic resistance gene classes associated with AR events. Our best performing model, which incorporated ARG class count data and antibiotics by presence or absence at  $\geq 72$  hours, yielded a promising predictive performance with an AUC of 0.730.

Across our integrated analyses, *M. smithii*, *C. leptum*, and *B. dorei* consistently emerged as key taxa associated with AR events during induction chemotherapy, underscoring their potential roles in shaping the gut ecosystem's vulnerability to resistant pathogen colonization or infection. *M. smithii* was the top-ranked feature in the Random Forest model by both mVIS and SHAP values and determined to be negatively associated with AR events in the correlation network. This species commonly contributes to gut microbial homeostasis through hydrogen consumption and syntrophic interactions with fermentative bacteria, which can help maintain community stability during antibiotic-induced



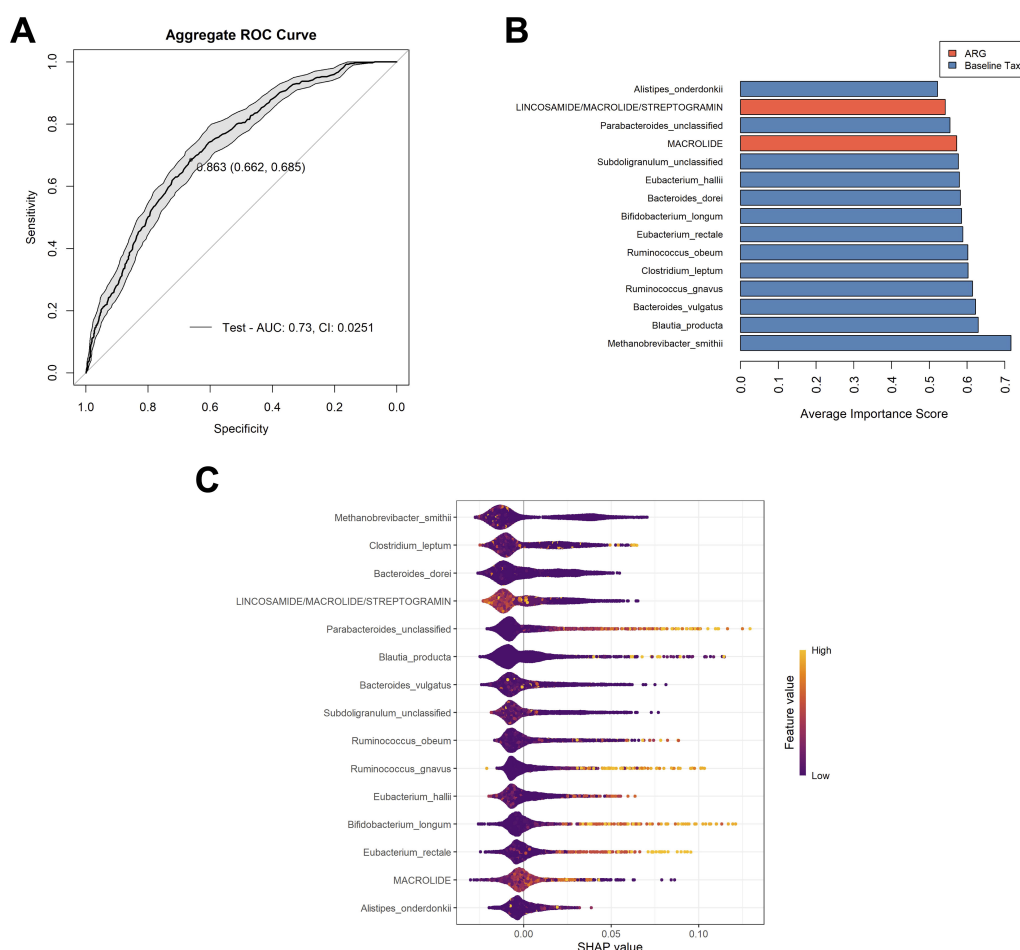


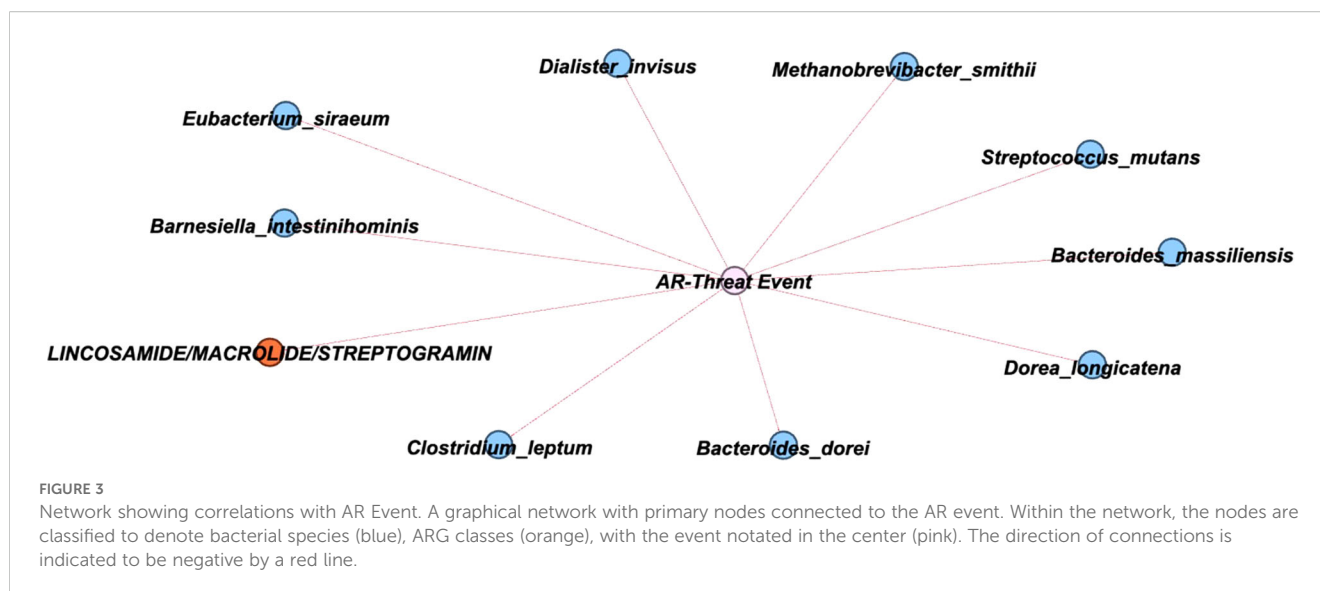
FIGURE 2

Analysis of the chosen model. **(A)** Aggregate ROC curve for Model 2 following optimization. The optimal cutoff point, determined by maximizing the Youden's index is indicated on the curve, followed by the specificity and sensitivity, respectively. The shaded grey area on the graph shows the confidence interval for the model. **(B)** The top 15 variables used in the optimized model that had the highest contributions to model performance. Variable importance scores are calculated by the mean decrease in node impurity (Gini importance). Variables are color coordinated with bacterial taxa in blue and ARG classes in orange. Most contributory variables are listed from most (bottom) to least contributory variables (top). **(C)** The SHAP (Shapely Additive exPlanations) beeswarm plot summarized the impact of each feature of the model's predictions across all samples and aggregated runs. Each point represents a single sample's SHAP value for a given feature, with the x-axis indicating the SHAP value (the effect of the feature on the model output for that sample). Features are ranked on the y-axis by their overall importance (mean absolute SHAP value). Color represents the original value of the feature for each sample (with lighter colors indicating higher values and darker colors indicating lower values). Points to the right indicate that the feature increases the predicted probability of the outcome for those samples, while points to the left indicate a decrease. The distribution and color gradient of the points reveal how high or low values of each feature influence the model's predictions. The variables are listed in order of most contributory to the model's prediction across all samples to least from top to bottom.

perturbation (Adrian et al., 2023; Ghavami et al., 2018; Malat et al., 2024). Prior studies have linked *M. smithii* to a more resilient gut environment and a reduced risk of inflammatory or infectious conditions, which are consistent with our findings of this species having a negative correlation with antibiotic-resistant event development (Chen et al., 2024; Cisek et al., 2024; Ghavami et al., 2018). Interestingly, the sCCA, a method designed to reveal coordinated variation between taxa and resistome features, depicted *M. smithii* showing weak positive correlations with several ARG classes, including those conferring resistance to lincosamides, macrolides, and streptogramins, which were also found to be negatively contributory to AR event prediction and

correlated with AR events in the correlation network. This may suggest that *M. smithii* persists in communities where ARGs are present but well-regulated by stable microbial networks, preventing overgrowth of pathogenic species despite the elevated resistance gene abundance (Dongre et al., 2025; Nhu and Young, 2023).

*C. leptum* was also positively correlated with the ARG abundance of genes conferring resistance to glycopeptides and the lincosamides/macrolides/streptogramins via the sCCA, indicating this species tends to co-occur with these resistance elements in the gut ecosystem. Previous studies have similarly linked *C. leptum* to dysbiotic gut environments and long-term ARG retention following antibiotic exposure (Korpela et al., 2016; Nielsen et al., 2021). Yet



despite these associations with elevated ARG content, *C. leptum* was shown to be negatively correlated with the development of an AR event in the correlation network, and negatively contributory to the Random Forest model via SHAP analyses. Rather than serving as a direct risk factor, its presence may reflect a type of microbiome that, while enriched for resistance genes, is still ecologically balanced enough to prevent extensive growth or translocation of pathogenic organisms (Grenda et al., 2022; Guo et al., 2020). This underscores the ecological nuance in the gut microbiome, where a species might contribute to resistance gene maintenance without necessarily increasing infection risk. Moreover, the presence of ARGs does not necessarily equate to virulence/pathogenesis of a pathogen, nor does it mean those genes are expressed. This also highlights the value of using multiple analytical approaches to uncover different layers of interaction between microbes, resistance elements, and clinical outcomes.

*B. dorei* was also identified as an important feature negatively contributory in the Random Forest model and demonstrated a negative correlation with AR event outcomes in the network analysis, suggesting that its presence may help guard against resistant pathogen colonization or infection. *B. dorei* is a prominent commensal species that contributes to gut health by reinforcing epithelial barrier integrity, producing immunomodulatory metabolites, and participating in the exclusion of pathogenic bacteria (Panwar et al., 2021; Pickard et al., 2017). These attributes become particularly relevant during chemotherapy, when the gut barrier is compromised, and antibiotic disruption leaves the microbiome vulnerable. The fact that *B. dorei* was shown to be negatively associated with the development of AR events suggests that it may help maintain gastrointestinal integrity or colonization resistance against pathogens under stress, possibly acting as part of a protective microbial buffer against the emergence or expansion of resistant organisms.

Among all the antibiotic resistance gene classes we examined, the group conferring resistance to lincosamides, macrolides, and streptogramins stood out for their consistent and unexpected association with reduced AR event risk. These genes showed up across each different analysis, 1) they were among the top predictors in the Random Forest model, suggesting a strong relationship with AR event risk; 2) they showed covariation with specific microbial taxa in the sCCA, indicating possible ecological linkage; and 3) they demonstrated a negative correlation with AR events in the network analysis. This last point is particularly unusual as resistance genes from this class are typically associated with broad-spectrum antibiotic use and worse clinical outcomes, especially in settings involving high-risk pathogens (Khodabandeh et al., 2019). One possibility as to why we are seeing these ARG classes negatively associated with AR events is that they might represent intrinsic resistances present among taxa that are part of stable microbial communities that help resist pathogenic invasion (Crits-Christoph et al., 2022; Moradi et al., 2022). Rather than signaling a harmful shift, the presence of these genes might reflect a microbiome that's been shaped by past antibiotic exposure but remains functionally resilient (Bhattarai et al., 2024; Fishbein et al., 2023). This highlights an important point: the presence of resistance genes doesn't always equate to increased infection risk; it depends heavily on the microbial community they're part of and the ecological context in which they persist.

Our predictive model represents an advancement beyond existing approaches by simultaneously assessing the risk of both antibiotic-resistant infection and gastrointestinal colonization. This dual focus is relatively uncommon in current models, which often concentrate solely on infection outcomes, specific types of infections, or specific pathogens (Rich et al., 2022; Z. Wang et al., 2024). Our model's design encompasses pathogens deemed as serious threats according to the CDC and various infection types,

moving beyond the typical focus on single-species predictions. While our model yields a moderate AUC of 0.730, this performance is comparable to or exceeds that of other recent predictive models for ARI, where AUCs range from 0.57 to 0.75 (Liao et al., 2023; Rich et al., 2022). Additionally, the model achieved a Youden Index of 0.863, reflecting a strong overall balance between sensitivity and specificity. However, certain limitations warrant discussion. First, it is possible that utilizing a 16S rRNA sequencing-based abundance threshold at the genus level to determine which longitudinal stools to perform selective and differential culturing on might have led to missed ARC events. Moreover, it is also important to note that although combining colonization and infection events may obscure biologically distinct processes, this choice was made to not only better power analyses, but because it is commonly believed that bloodstream infections may stem from gastrointestinal colonization and translocation, making it biologically relevant to include these categories together. Second, while aggregating SHAP values across multiple independently trained models enhances the robustness of feature importance estimates, this approach can also amplify ambiguity, particularly when the underlying data relationships are complex. In this study, the variables utilized were microbial, which are known to exhibit extensive feature interactions and nonlinear relationships. Additionally, microbial taxa data are inherently zero-inflated, which contributes to some of the ambiguity in the SHAP interpretation, as reflected in the SHAP beeswarm plot where most data points are colored purple, indicating low or absent abundance for most taxa. These characteristics likely contribute to the observed ambiguity in SHAP results, as the effect of individual taxa on model predictions may vary depending on the presence and absence of other taxa and may not follow simple monotonic patterns. Nevertheless, the results in the SHAP plot are corroborated by those in the network plot, supporting the overall validity of our findings. Lastly, while our model demonstrates a notable improvement over existing tools, we recognize that our small cohort size, number of events, and single institution analyses are limiting the efficacy of our model. This model is in its developmental stages and further studies will be necessary to validate and refine this model in larger, multicenter cohorts and extend the analysis to different high-risk populations to enhance clinical applicability. Unfortunately, we have not been able to validate our model on an independent cohort, due to the inability to find a similar clinical cohort with the same sequencing and clinical metadata available.

In conclusion, this study provides an improvement in the predictive modeling of AR events in AML patients by integrating comprehensive microbiome, resistome, and clinical data. Although, in the early stages of model development, the performance of this model suggests a possibility for clinical utility and highlights its potential for future implementation in infection risk stratification once the limitations are addressed. These findings not only offer valuable predictors for clinical decision-making but also emphasize the critical importance of microbiome preservation during antibiotic therapy. Ongoing research and validation in diverse clinical populations will be essential to refine and implement these predictive tools effectively in clinical practice. Moreover,

functional validation via metatranscriptomics, metabolomics, or gnotobiotic models would be valuable to determine the role of the predicted protective taxa, their metabolic products, and contributory ARGs on AR event risk.

## Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/[Supplementary Material](#).

## Ethics statement

The studies involving humans were approved by MD Anderson Cancer Center Institutional Review Board (PA13-0339 and PA15-0780). The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

## Author contributions

SM: Data curation, Formal analysis, Investigation, Methodology, Visualization, Writing – original draft, Writing – review & editing. SF: Formal Analysis, Methodology, Writing – review & editing. JG: Conceptualization, Funding acquisition, Investigation, Supervision, Writing – original draft, Writing – review & editing.

## Funding

The author(s) declare financial support was received for the research and/or publication of this article. These studies were supported by K01AI143881 (NIAID).

## Acknowledgments

We acknowledge the staff at MD Anderson Cancer Center for their help in identification and collection of patient samples. We thank the MD Anderson Cancer Center CCSG Microbiome core facility (NIH/NCI no. P30CA016672) for 16S rRNA sequencing. We also thank the Texas Institute for Genome Sciences and Society for shotgun metagenomic sequencing and J.C. Blazier for assistance with sequence analyses.

## Conflict of interest

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

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

## Generative AI statement

The author(s) declare that Generative AI was used in the creation of this manuscript. Grammarly was utilized to edit the abstract with the prompt “shorten”. CHATGPT 4.0 was utilized to edit small parts of the body of the manuscript with the prompt “make sound more professional or scientific”

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure

accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fcimb.2025.1629422/full#supplementary-material>

## References

- Adrian, M. A., Ayati, B. P., and Mangalam, A. K. (2023). A mathematical model of *Bacteroides thetaiotaomicron*, *Methanobrevibacter smithii*, and *Eubacterium rectale* interactions in the human gut. *Sci. Rep.* 13, 21192. doi: 10.1038/s41598-023-48524-4
- Aitken, S. L., Sahasrabhojane, P. V., Kontoyannis, D. P., Savidge, T. C., Arias, C. A., Ajami, N. J., et al. (2021). Alterations of the oral microbiome and cumulative carbapenem exposure are associated with *Stenotrophomonas maltophilia* infection in patients with acute myeloid leukemia receiving chemotherapy. *Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am.* 72, 1507–1513. doi: 10.1093/cid/ciaa778
- Antimicrobial resistance. (2023). Available online at: <https://www.who.int/news-room/fact-sheets/detail/antimicrobial-resistance> (Accessed January 8, 2025).
- Bastian, M., Heymann, S., and Jacomy, M. (2009). Gephi: an open source software for exploring and manipulating networks. *Proc. Int. AAAI Conf. Web Soc. Media* 3, 361–362. doi: 10.1609/icwsm.v3i1.13937
- Becattini, S., Taur, Y., and Pamer, E. G. (2016). Antibiotic-induced changes in the intestinal microbiota and disease. *Trends Mol. Med.* 22, 458–478. doi: 10.1016/j.molmed.2016.04.003
- Belkaid, Y., and Hand, T. W. (2014). Role of the microbiota in immunity and inflammation. *Cell* 157, 121–141. doi: 10.1016/j.cell.2014.03.011
- Bhattarai, S. K., Du, M., Zeamer, A. L., Morzfeld, B., Kellogg, T. D., Firat, K., et al. (2024). Commensal antimicrobial resistance mediates microbiome resilience to antibiotic disruption. *Sci. Transl. Med.* 16, eadi9711. doi: 10.1126/scitranslmed.adi9711
- Boccia, R., Glaspy, J., Crawford, J., and Apro, M. (2022). Chemotherapy-Induced Neutropenia and Febrile Neutropenia in the US: A Beast of Burden That Needs to Be Tamed? *Oncologist* 27 (8), 625–636. doi: 10.1093/oncolo/oyac074
- Buffie, C. G., and Pamer, E. G. (2013). Microbiota-mediated colonization resistance against intestinal pathogens. *Nat. Rev. Immunol.* 13 (11), 790–801. doi: 10.1038/nri3535
- Canica, M., Manageiro, V., Jones-Dias, D., Clemente, L., Gomes-Neves, E., Poeta, P., et al. (2015). Current perspectives on the dynamics of antibiotic resistance in different reservoirs. *Res. Microbiol.* 166, 594–600. doi: 10.1016/j.resmic.2015.07.009. Special Issue: Trends in Environmental Microbiology for Public Health.
- Castro-Sánchez, E., Moore, L. S. P., Husson, F., and Holmes, A. H. (2016). What are the factors driving antimicrobial resistance? Perspectives from a public event in London, England. *BMC Infect. Dis.* 16, 465. doi: 10.1186/s12879-016-1810-x
- CDC (2019). *Antibiotic Resistance Threats in the United States, 2019* (Atlanta, GA: U.S. Department of Health and Human Services, CDC). doi: 10.15620/cdc:82532.
- Chandrakumar, I., Gauthier, N. P. G., Nelson, C., Bonsall, M. B., Locher, K., Charles, M., et al. (2022). BugSplit enables genome-resolved metagenomics through highly accurate taxonomic binning of metagenomic assemblies. *Commun. Biol.* 5, 151. doi: 10.1038/s42003-022-03114-4
- Chen, S., Zhang, Z., Liu, S., Chen, T., Lu, Z., Zhao, W., et al. (2024). Consistent signatures in the human gut microbiome of longevous populations. *Gut Microbes* 16, 2393756. doi: 10.1080/19490976.2024.2393756
- Chinemere Nwobodo, D., Ugwu, M. C., Oiselo Anie, C., Al-Ouqaili, M. T. S., Chinedu Ikem, J., Victor Chigozie, U., et al. (2022). Antibiotic resistance: the challenges and some emerging strategies for tackling a global menace. *J. Clin. Lab. Anal.* 36 (9), e24655. doi: 10.1002/jcla.24655
- Cisek, A. A., Szymańska, E., Aleksandrak-Piekarczyk, T., and Cukrowska, B. (2024). The role of methanogenic archaea in inflammatory bowel disease—A review. *J. Pers. Med.* 14, 196. doi: 10.3390/jpm14020196
- Crits-Christoph, A., Hallowell, H. A., Koutouvalis, K., and Suez, J. (2022). Good microbes, bad genes? The dissemination of antimicrobial resistance in the human microbiome. *Gut Microbes* 14, 2055944. doi: 10.1080/19490976.2022.2055944
- Doan, T., Worden, L., Hinterwirth, A., Arzika, A. M., Maliki, R., Abdou, A., et al. (2020). Macrolide and nonmacrolide resistance with mass azithromycin distribution. *N. Engl. J. Med.* 383, 1941–1950. doi: 10.1056/NEJMoa2002606
- Dongre, D. S., Saha, U. B., and Saroj, S. D. (2025). Exploring the role of gut microbiota in antibiotic resistance and prevention. *Ann. Med.* 57, 2478317. doi: 10.1080/07853890.2025.2478317
- Edgar, R. C. (2016). UNOISE2: improved error-correction for Illumina 16S and ITS amplicon sequencing. doi: 10.1101/081257
- Fan, J., Huang, S., and Chorlton, S. D. (2021). BugSeq: a highly accurate cloud platform for long-read metagenomic analyses. *BMC Bioinf.* 22, 160. doi: 10.1186/s12859-021-04089-5
- Fishbein, S. R. S., Mahmud, B., and Dantas, G. (2023). Antibiotic perturbations to the gut microbiome. *Nat. Rev. Microbiol.* 21, 772–788. doi: 10.1038/s41579-023-00933-y
- Forslund, K., Sunagawa, S., Kultima, J. R., Mende, D. R., Arumugam, M., Typas, A., et al. (2013). Country-specific antibiotic use practices impact the human gut resistome. *Genome Res.* 23, 1163–1169. doi: 10.1101/gr.155465.113
- Galloway-Peña, J., Brumlow, C., and Shelburne, S. (2017). Impact of the microbiota on bacterial infections during cancer treatment. *Trends Microbiol.* 25, 992–1004. doi: 10.1016/j.tim.2017.06.006
- Galloway-Peña, J. R., Shi, Y., Peterson, C. B., Sahasrabhojane, P., Gopalakrishnan, V., Brumlow, C. E., et al. (2020). Gut microbiome signatures are predictive of infectious risk following induction therapy for acute myeloid leukemia. *Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am.* 71, 63–71. doi: 10.1093/cid/ciz777
- Galloway-Peña, J., Smith, D. P., Sahasrabhojane, P., Ajami, N. J., Wadsworth, W. D., Dayer, N. G., et al. (2016). The role of the gastrointestinal microbiome in infectious complications during induction chemotherapy for acute myeloid leukemia. *Cancer* 122, 2186–2196. doi: 10.1002/cncr.30039
- Galloway-Peña, J. R., Smith, D. P., Sahasrabhojane, P., Wadsworth, W. D., Fellman, B. M., Ajami, N. J., et al. (2017). Characterization of oral and gut microbiome temporal variability in hospitalized cancer patients. *Genome Med.* 9 (1), 21. doi: 10.1186/s13073-017-0409-1
- Gauthier, N. P. G., Locher, K., MacDonald, C., Chorlton, S. D., Charles, M., and Manges, A. R. (2022). Alterations in the nasopharyngeal microbiome associated with SARS-CoV-2 infection status and disease severity. *PloS One* 17, e0275815. doi: 10.1371/journal.pone.0275815
- Gedik, H., Şimşek, F., Kantürk, A., Yildirmak, T., Arica, D., Aydin, D., et al. (2014). Bloodstream infections in patients with hematological Malignancies: which is more fatal – cancer or resistant pathogens? *Ther. Clin. Risk Manage.* 10, 743–752. doi: 10.2147/TCRM.S68450



- Ghavamian, S. B., Rostami, E., Sephay, A. A., Shahrokh, S., Balaii, H., Aghdaei, H. A., et al. (2018). Alterations of the human gut *Methanobrevibacter smithii* as a biomarker for inflammatory bowel diseases. *Microb. Pathog.* 117, 285–289. doi: 10.1016/j.micpath.2018.01.029
- Gibson, M. K., Forsberg, K. J., and Dantas, G. (2015). Improved annotation of antibiotic resistance determinants reveals microbial resistomes cluster by ecology. *ISME J.* 9, 207–216. doi: 10.1038/ismej.2014.106
- Gibson, M. K., Pesesky, M. W., and Dantas, G. (2014). The yin and yang of bacterial resilience in the human gut microbiota. *J. Mol. Biol.* 426, 3866–3876. doi: 10.1016/j.jmb.2014.05.029
- Grenda, T., Grenda, A., Domaradzki, P., Krawczyk, P., and Kwiatek, K. (2022). Probiotic potential of clostridium spp.—Advantages and doubts. *Curr. Issues Mol. Biol.* 44, 3118–3130. doi: 10.3390/cimb44070215
- Guo, P., Zhang, K., Ma, X., and He, P. (2020). Clostridium species as probiotics: potentials and challenges. *J. Anim. Sci. Biotechnol.* 11, 24. doi: 10.1186/s40104-019-0402-1
- Hansen, B.-A., Wendelbo, Ø., Bruserud, Ø., Hemsing, A. L., Mosevoll, K. A., and Reikvam, H. (2019). Febrile neutropenia in acute leukemia. *Epidemiology, etiology, pathophysiology and treatment. Mediterr. J. Hematol. Infect. Dis.* 12, e2020009. doi: 10.4084/MJHID.2020.009
- Iwan, E., Grenda, A., Bomba, A., Bielińska, K., Wasyl, D., Kiesko, R., et al. (2024). Gut resistome of NSCLC patients treated with immunotherapy. *Front. Genet.* 15. doi: 10.3389/fgene.2024.1378900
- Jutkina, J., Rutgersson, C., Flach, C.-F., and Joakim Larsson, D. G. (2016). An assay for determining minimal concentrations of antibiotics that drive horizontal transfer of resistance. *Sci. Total Environ.* 548–549, 131–138. doi: 10.1016/j.scitotenv.2016.01.044
- Khodabandeh, M., Mohammadi, M., Abdolsalehi, M. R., Alvandimanes, A., Gholami, M., Bibalan, M. H., et al. (2019). Analysis of resistance to macrolide-lincosamide-streptogramin B among mecA-positive staphylococcus aureus isolates. *Osong Public Health Res. Perspect.* 10, 25–31. doi: 10.24171/j.phrp.2019.10.1.06
- Korpela, K., Salonen, A., Virta, L. J., Kekkonen, R. A., Forslund, K., Bork, P., et al. (2016). Intestinal microbiome is related to lifetime antibiotic use in Finnish pre-school children. *Nat. Commun.* 7, 10410. doi: 10.1038/ncomms10410
- Korry, B. J., Cabral, D. J., and Belenky, P. (2020). Metatranscriptomics reveals antibiotic-induced resistance gene expression in the murine gut microbiota. *Front. Microbiol.* 11. doi: 10.3389/fmicb.2020.00322
- Latorre, M., Krishnareddy, S., and Freedberg, D. E. (2015). Microbiome as mediator: Do systemic infections start in the gut? *World J. Gastroenterol.* 21, 10487–10492. doi: 10.3748/wjg.v21.i37.10487
- Lewis, B. B., Buffie, C. G., Carter, R. A., Leiner, I., Toussaint, N. C., Miller, L. C., et al. (2015). Loss of microbiota-mediated colonization resistance to clostridium difficile infection with oral vancomycin compared with metronidazole. *J. Infect. Dis.* 212, 1656–1665. doi: 10.1093/infdis/jiv256
- Liao, Q., Feng, Z., Lin, H., Zhou, Y., Lin, J., Zhuo, H., et al. (2023). Carbapenem-resistant gram-negative bacterial infection in intensive care unit patients: Antibiotic resistance analysis and predictive model development. *Front. Cell. Infect. Microbiol.* 13. doi: 10.3389/fcimb.2023.1109418
- Malat, I., Drancourt, M., and Grine, G. (2024). Methanobrevibacter smithii cell variants in human physiology and pathology: A review. *Heliyon* 10 (18), e36742. doi: 10.1016/j.heliyon.2024.e36742
- McMahon, S., Sahasrabhojane, P., Kim, J., Franklin, S., Chang, C.-C., Jenq, R. R., et al. (2023). Contribution of the oral and gastrointestinal microbiomes to bloodstream infections in leukemia patients. *Microbiol. Spectr.* 11, e0041523. doi: 10.1128/spectrum.00415-23
- Modi, S. R., Collins, J. J., and Relman, D. A. (2014). Antibiotics and the gut microbiota. *J. Clin. Invest.* 124, 4212–4218. doi: 10.1172/JCI72333
- Moradi, J., Fathollahi, M., Halimi, S., Alvandi, A., Abiri, R., Vaziri, S., et al. (2022). Characterization of the resistome in *Lactobacillus* genomic sequences from the human gut. *J. Glob. Antimicrob. Resist.* 30, 451–458. doi: 10.1016/j.jgar.2022.05.014
- Nanayakkara, A. K., Boucher, H. W., Fowler, J. V. G., Jezek, A., Outtersen, K., and Greenberg, D. E. (2021). Antibiotic resistance in the patient with cancer: Escalating challenges and paths forward. *CA. Cancer J. Clin.* 71, 488–504. doi: 10.3322/caac.21697
- Nhu, N. T. Q., and Young, V. B. (2023). The relationship between the microbiome and antimicrobial resistance. *Clin. Infect. Dis.* 77, S479–S486. doi: 10.1093/cid/ciad641
- Nielsen, K. L., Olsen, M. H., Pallegá, A., Ebdrup, S. R., Sørensen, N., Lukjancenko, O., et al. (2021). Microbiome compositions and resistome levels after antibiotic treatment of critically ill patients: an observational cohort study. *Microorganisms* 9, 2542. doi: 10.3390/microorganisms9122542
- Nobel, Y. R., Cox, L. M., Kirigin, F. F., Bokulich, N. A., Yamanishi, S., Teitler, I., et al. (2015). Metabolic and metagenomic outcomes from early-life pulsed antibiotic treatment. *Nat. Commun.* 6, 7486. doi: 10.1038/ncomms8486
- O'Neill, J. (2016). *Tackling drug-resistant infections globally: final report and recommendations. Review on Antimicrobial Resistance.* Wellcome Trust and HM Government.
- On, S., Rath, C. G., Lan, M., Wu, B., Lau, K. M., Cheung, E., et al. (2022). Characterization of infections in patients with acute myeloid leukemia receiving venetoclax and a hypomethylating agent. *Br. J. Hematol.* 197, 63–70. doi: 10.1111/bjh.18051
- Palmore, T. N., and Henderson, D. K. (2013). Managing transmission of carbapenem-resistant enterobacteriaceae in healthcare settings: A view from the trenches. *Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am.* 57, 1593–1599. doi: 10.1093/cid/cit531
- Panwar, R. B., Sequeira, R. P., and Clarke, T. B. (2021). Microbiota-mediated protection against antibiotic-resistant pathogens. *Genes Immun.* 22, 255–267. doi: 10.1038/s41435-021-00129-5
- Pehrsson, E. C., Forsberg, K. J., Gibson, M. K., Ahmadi, S., and Dantas, G. (2013). Novel resistance functions uncovered using functional metagenomic investigations of resistance reservoirs. *Front. Microbiol.* 4. doi: 10.3389/fmicb.2013.00145
- Pickard, J. M., Zeng, M. Y., Caruso, R., and Núñez, G. (2017). Gut microbiota: role in pathogen colonization, immune responses and inflammatory disease. *Immunol. Rev.* 279, 70–89. doi: 10.1111/immr.12567
- Rapoport, B. L., Cooksley, T., Johnson, D. B., Anderson, R., and Shannon, V. R. (2021). Treatment of infections in cancer patients: an update from the neutropenia, infection and myelosuppression study group of the Multinational Association for Supportive Care in Cancer (MASCC). *Expert Rev. Clin. Pharmacol.* 14, 295–313. doi: 10.1080/17512433.2021.1884067
- Rashidi, A., Ebadi, M., Rehman, T. U., Elhusseini, H., Halaweish, H., Kaiser, T., et al. (2022). Compilation of longitudinal gut microbiome, serum metabolome, and clinical data in acute myeloid leukemia. *Sci. Data* 9, 468. doi: 10.1038/s41597-022-01600-2
- Rich, S. N., Jun, I., Bian, J., Boucher, C., Cherabuddi, K., Morris, J. G., et al. (2022). Development of a prediction model for antibiotic-resistant urinary tract infections using integrated electronic health records from multiple clinics in north-central florida. *Infect. Dis. Ther.* 11, 1869–1882. doi: 10.1007/s40121-022-00677-x
- Rolston, K. V. I. (2015). Infections in patients with acute leukemia. In: *Infections in Hematology*. G. Maschmeyer and K. Rolston (eds). (Berlin, Heidelberg: Springer). doi: 10.1007/978-3-662-44000-1\_1
- Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., et al. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* 75, 7537–7541. doi: 10.1128/AEM.01541-09
- Shono, Y., Docampo, M. D., Peled, J. U., Perobelli, S. M., Velardi, E., Tsai, J. J., et al. (2016). Increased GVHD-related mortality with broad-spectrum antibiotic use after allogeneic hematopoietic stem cell transplantation in human patients and mice. *Sci. Transl. Med.* 8, 339ra71. doi: 10.1126/scitranslmed.aaf2311
- Sougiannis, A. T., VanderVeen, B. N., Davis, J. M., Fan, D., and Murphy, E. A. (2021). Understanding chemotherapy-induced intestinal mucositis and strategies to improve gut resilience. *Am. J. Physiol.-Gastrointest. Liver Physiol.* 320, G712–G719. doi: 10.1152/ajpgi.00380.2020
- Spearman, C. (1904). The proof and measurement of association between two things. *Am. J. Psychol.* 15, 72–101. doi: 10.2307/1412159
- Taur, Y., and Pamer, E. G. (2016). Microbiome mediation of infections in the cancer setting. *Genome Med.* 8, 40. doi: 10.1186/s13073-016-0306-z
- Thaiss, C. A., Zmora, N., Levy, M., and Elinav, E. (2016). The microbiome and innate immunity. *Nature* 535, 65–74. doi: 10.1038/nature18847
- Toucheffeu, Y., Montassier, E., Nieman, K., Gastinne, T., Potel, G., Bruley des Varannes, S., et al. (2014). Systematic review: the role of the gut microbiota in chemotherapy- or radiation-induced gastrointestinal mucositis – current evidence and potential clinical applications. *Aliment. Pharmacol. Ther.* 40, 409–421. doi: 10.1111/apt.12878
- Ubeda, C., Bucci, V., Caballero, S., Djukovic, A., Toussaint, N. C., Equinda, M., et al. (2013). Intestinal microbiota containing *Barnesiella* species cures vancomycin-resistant *Enterococcus faecium* colonization. *Infect Immun.* 81 (3), 965–973. doi: 10.1128/IAI.01197-12
- Walsh, T. R., Gales, A. C., Laxminarayan, R., and Dodd, P. C. (2023). Antimicrobial resistance: addressing a global threat to humanity. *PLoS Med.* 20, e1004264. doi: 10.1371/journal.pmed.1004264
- Wang, X., Arya, S. V., Patel, S., Saw, S., Decena, M. A., Hirsh, R., et al. (2024). Antibiotic practice and stewardship in the management of neutropenic fever: a survey of US institutions. *Infect. Control Hosp. Epidemiol.* 45, 1168–1175. doi: 10.1017/ice.2024.103
- Wang, Z., Wu, H., Guo, Y., Zhu, L., Dai, Z., Zhang, H., et al. (2024). Development and validation of a novel prediction model for Carbapenem-resistant organism infection in a large-scale hospitalized patients. *Diagn. Microbiol. Infect. Dis.* 110, 116415. doi: 10.1016/j.diagmicrobio.2024.116415
- Zimmer, A. J., and Freifeld, A. G. (2019). Optimal management of neutropenic fever in patients with cancer. *J. Oncol. Pract.* 15, 19–24. doi: 10.1200/JOP.18.00269