

Integrating intestinal microbiome and urinary metabolome data to predict secondary infection in critically ill patients

Received: 10 October 2025

Accepted: 22 December 2025

Published online: 13 March 2026

Cite this article as: Linz C., Tsenova K., Dettmer K. *et al.* Integrating intestinal microbiome and urinary metabolome data to predict secondary infection in critically ill patients. *Crit Care* (2026). <https://doi.org/10.1186/s13054-025-05818-5>

Charlotte Linz, Kristiyana Tsenova, Katja Dettmer, Lisa Ellmann, Peter J. Oefner, Wolfram Gronwald, Fedja Farowski, Alina M. Rüb, Daniel E. Freedberg, Philipp Koehler, Jorge Garcia Borrega, Jan-Hendrik Naendrup, Maria J. G. T. Vehreschild & Boris Böll

We are providing an unedited version of this manuscript to give early access to its findings. Before final publication, the manuscript will undergo further editing. Please note there may be errors present which affect the content, and all legal disclaimers apply.

If this paper is publishing under a Transparent Peer Review model then Peer Review reports will publish with the final article.

Integrating intestinal microbiome and urinary metabolome data to predict secondary infection in critically ill patients

Charlotte Linz¹, Kristiyana Tsenova², Katja Dettmer³, Lisa Ellmann³, Peter J. Oefner³

Wolfram Gronwald³, Fedja Farowski^{1,2}, Alina M. Rüb^{1,2}, Daniel E. Freedberg⁴, Philipp Koehler^{1,5,6}

Jorge Garcia Borrega¹, Jan-Hendrik Naendrup¹, Maria J.G.T. Vehreschild^{1,2} * and Boris Böll¹⁺ *

* Contributed equally

¹ University of Cologne, Cologne, Germany, Faculty of Medicine and University Hospital Cologne, Department I of Internal Medicine, Division of Hematology-Oncology/Critical Care Medicine/Infectious Diseases, Center for Integrated Oncology Aachen Bonn Cologne Düsseldorf (CIO ABCD)

² Goethe University Frankfurt, Frankfurt am Main, Germany, University Hospital Frankfurt, Department II of Internal Medicine, Infectious Diseases

³ University of Regensburg, Regensburg, Germany, Institute of Functional Genomics

⁴ Columbia University, New York, United States, Division of Digestive and Liver Diseases, Mailman School of Public Health, Department of Epidemiology

⁵ University of Cologne, Cologne, Germany, Faculty of Medicine and University Hospital Cologne, Department I of Internal Medicine, Division of Clinical Immunology

⁶ University of Cologne, Cologne, Germany, Faculty of Medicine and University Hospital Cologne, Institute of Translational Research, Cologne Excellence Cluster on Cellular Stress Responses in Aging-Associated Diseases (CECAD)

+ Correspondence: Boris Böll, University Hospital Cologne, Kerpener Strasse 62, Cologne, Germany; email: boris.boell@uk-koeln.de

Abstract

Background: Secondary infection (SI), including ventilator-associated pneumonia (VAP) and bloodstream infection (BSI), represents a major complication in critically ill patients. Current clinical risk stratification approaches

prove inadequate for timely and precise identification of at-risk patients. This study identifies intestinal microbiome and urinary metabolome characteristics (“multi-omics data”) associated with SI occurrence, investigates convergence of the respiratory microbiome with the intestinal microbiome, and determines whether multi-omics integration enhances prognostic discrimination for patients at risk of developing SI.

Methods: We analyzed data from mechanically ventilated patients from two cohorts: University Hospital Cologne (UHC), Germany, and Columbia University Medical Center (CUMC), New York, United States. The core dataset (n = 88; 64 UHC and 24 CUMC) assessed multi-omics integration for SI prediction, with an UHC subset (n = 55) providing more comprehensive clinical and microbiome characterization. Baseline intestinal and respiratory microbiome, as well as urinary metabolome data were collected within 48 hours of intensive care unit admission or intubation using 16S ribosomal ribonucleic acid (rRNA) sequencing and nuclear magnetic resonance (NMR) spectroscopy. SI was defined as new-onset BSI or VAP occurring ≥ 48 hours after enrollment. Regression and classification models compared clinical-only approaches with integrated multi-omics models using model selection criteria, area under the curve (AUC), and Matthews correlation coefficients.

Results: SI occurred in 28% of patients, with prior antibiotic exposure associated with SI (84% vs. 41%, $q < 0.01$; odds ratio 2.57, $p = 0.17$). SI patients exhibited significantly lower baseline intestinal microbial diversity (Shannon diversity, 1.96 vs. 3.47, $p < 0.01$) and greater *Enterococcus* abundance (46% vs. 11%, $q = 0.02$), with similar patterns observed in the respiratory microbiome. Urinary NMR analysis identified metabolites mapping to features at 0.935 ppm (2-oxoisocaproate, isoleucine) in the core dataset, and at 8.025 ppm (quinolinate) in the UHC subset as elevated in SI patients. Multi-omics models demonstrated modest but consistent improvement over clinical-only models (AUC: 0.75 vs. 0.64).

Conclusions: SI susceptibility in critically ill patients associates with underlying clinical severity, prior antibiotic exposure, and microbiota disruption. Multi-omics integration yielded consistent predictive improvement, supporting prospective validation as a proof-of-concept approach for early SI risk stratification.

Keywords

Secondary infection, ventilator-associated pneumonia (VAP), bloodstream infection (BSI), intestinal microbiome, urinary metabolites

Background

Secondary infection (SI), notably ventilator-associated pneumonia (VAP) and bloodstream infection (BSI), significantly increases morbidity, mortality, and healthcare costs in critically ill patients (1, 2). Over half of intensive care unit (ICU) patients experience at least one infection during their stay, with approximately one-fifth classified as ICU-acquired. Associated mortality rates reach 28–34% in hospitalized populations (1). The emergence of multidrug-resistant pathogens further constrains therapeutic options, emphasizing the need for enhanced early risk identification and preventive strategies (3).

The intestinal microbiota plays a central role in modulating host susceptibility to SI through mechanisms of colonization resistance. Many nosocomial pathogens originate from this reservoir; however, not all colonized patients develop active infections. This protective function relies on commensal anaerobes that restrict pathogen expansion via nutrient competition, antimicrobial compound production, and bacteriophage activity (4, 5). Fecal biodiversity is often considered a surrogate measure of this resistance. Reduced microbial diversity and pathogen overgrowth have been associated with increased infection risk, disease severity, and mortality in ICU populations (6-9). Antibiotic exposure, a nearly universal feature in critical illness, further disrupts these protective communities and favors pathogen proliferation (6).

Despite these well-established associations, baseline microbiome features have not been systematically evaluated for their independent predictive value in SI development. This uncertainty reflects methodological variability, limited longitudinal data establishing causality, and inter-individual heterogeneity in dysbiotic responses. The predictive value of baseline microbiome features for infection remains unclear (10, 11). Additionally, current sequencing technologies remain unsuitable for real-time clinical deployment. Consequently, non-invasive urinary metabolites warrant investigation as potential surrogate markers of microbiota integrity. Previous work has identified associations between depleted urinary 3-indoxyl sulfate (3-IS), a tryptophan-derived metabolite produced by intestinal commensal bacteria, and adverse outcomes, including gut microbiota disruption and increased one-year mortality, both following allogeneic stem cell transplantation (12) and ICU admission for mainly sepsis, post-surgery or cardiac causes (13).

Therefore, this study addresses three interconnected objectives: (1) to identify intestinal microbiome and urinary metabolome signatures (“multi-omics data”) associated with SI development, (2) to investigate the convergence of the respiratory microbiome with the intestinal microbiome, and (3) to evaluate whether integrating multi-omics data with clinical variables enhances predictive discrimination compared to clinical assessment alone.

Patients and methods

Study framework and patient cohorts

This analysis draws on data from two independent tertiary-care academic cohorts: the medical ICU at the University Hospital Cologne (UHC), Germany, and five medical or surgical ICUs at Columbia University Medical Center (CUMC), New York, United States. A stepwise analytical approach was implemented to systematically evaluate multi-omics integration.

Initially, patients from both institutions were combined into a core dataset ($n = 88$; 64 UHC and 24 CUMC) to assess the baseline impact of incorporating intestinal microbiome and urinary metabolome data on SI prediction. Subsequently, a more detailed investigation was performed on an UHC-only subset comprising 80 patients total (the original 64 core patients plus 16 additional unique UHC patients). From this expanded cohort, 55 patients with complete microbiome and urinary metabolome data were included in integrated predictive modeling. This subset enabled more comprehensive microbiome profiling, including extended characterization of the intestinal microbiome, and respiratory microbiota profiling via bronchoalveolar lavage (BAL) fluid and endotracheal aspirate (ETA) sampling (additional details on dataset characterization are available in Supplementary Figure S1, Additional File 1).

Patient selection criteria

Eligibility criteria included age ≥ 18 years, ICU admission, and mechanical ventilation initiated within 3 days of enrollment. Exclusion criteria, applied during initial cohort recruitment, included COVID-19 at any point during sampling (laboratory limitations, UHC), prior *Clostridioides difficile* infection within 90 days (CUMC), recent bacterial BSI, or recent ICU admissions (both within 30 days, CUMC).

Data and specimen collection

Baseline demographic, clinical, and laboratory variables were obtained from medical records at enrollment. For the UHC cohort, biospecimens including BAL fluid, ETA, urine, fecal and deep rectal swabs were collected within 48 hours of intubation. Given strong concordance between fecal and rectal swab microbiota profiles, results were pooled accordingly (14). For the CUMC cohort, rectal swabs and urine were collected within 4 hours of ICU admission. In the core dataset, microbiome analysis included assessment of intestinal microbial diversity using the Shannon diversity index (15). For the subset of UHC patients, analyses were expanded to include respiratory specimens (BAL and ETA) and more detailed microbial profiling beyond standard diversity metrics.

Outcome definition

SI was defined as new-onset BSI and/or VAP, occurring at least 48 hours following enrollment. This threshold is consistent with the Robert Koch Institute (RKI) guideline on nosocomial pneumonia, which defines hospital-acquired pneumonia (including VAP) as pneumonia developing more than 48 hours after hospital admission (16). Diagnostic criteria were further informed by the United States Centers for Disease Control and Prevention (CDC) clinical guidelines for both VAP and BSI cases (17, 18). All bloodstream cultures were independently reviewed by experienced clinicians to differentiate true bacteremia from contamination; single-positive cultures without supporting clinical evidence were not classified as infections. For each patient, only the first ICU admission was examined, with all SIs captured within 30 days of enrollment.

Descriptive statistics

Statistical analyses were performed using R (v4.4.1 and v4.2.0; R Foundation, Vienna, Austria) (19). Continuous variables are presented as medians or means [interquartile range (IQR); full range], and categorical variables as frequencies [%]. Group comparisons used Wilcoxon rank-sum (exact), Fisher's exact, or Pearson's chi-squared test, with p values < 0.05 considered statistically significant. Multiple testing correction using the Benjamini–Hochberg procedure to control the false discovery rate (FDR) was applied to clinical, microbiome, and metabolite data where appropriate, using a q value threshold of 0.05. Random forest imputation addressed selected data missing at random (see Supplementary Table S5, Additional File 5).

Microbiome analysis

Microbiome profiling was performed independently at both study sites following established protocols. Methods used for the CUMC cohort have been described previously (13). For the UHC cohort, details are provided in the Additional File 3; briefly, genomic DNA was extracted using the ZymoBIOMICS™ DNA/RNA Miniprep Kit

(Zymo Research; Irvine, CA, USA), and the V3–V4 regions of the 16S rRNA gene were sequenced on an Illumina MiSeq platform (Illumina; San Diego, CA, USA). Sequencing data were processed via the DADA2 plugin (20) in QIIME2 (21). Taxonomic classification was performed via a naïve Bayes classifier (sklearn) (22) trained on the SILVA database release 138 (23). Microbial dominance was defined as at least 30% relative abundance of a single taxon. Alpha-diversity (α -diversity) was assessed using the Shannon (15) and Simpson's diversity indices (24), the abundance-based coverage estimator (ACE) (25), the Chao1 estimator (26), species richness (SR), and phylogenetic diversity (PD) (27). Beta-diversity (β -diversity) was evaluated via UniFrac (28) and Bray–Curtis dissimilarity (29) metrics. Differential abundance analysis was performed using linear discriminant analysis (LDA) effect size (LEfSe) (30), considering taxa with an LDA score above 2.

Urine analysis

Urine analysis procedures are detailed in the Additional File 4; briefly, analyses were performed using nuclear magnetic resonance (NMR) spectroscopy and ultra-high performance liquid chromatography-tandem mass spectrometry (UHPLC–MS/MS). The former was carried out on a Bruker Avance III HD 600 MHz spectrometer, with metabolite identification and quantification performed using the Chenomx NMR Suite 8.3 (Chenomx Inc., Edmonton, Alberta, Canada). Data were processed with TopSpin 4.14 (Bruker) and AMIX 3.9.13 (Bruker). For classification analyses, spectra were divided into 1029 evenly spaced buckets of 0.01 ppm width, followed by creatinine normalization. Spectral overlap in the NMR data has limited interpretability, with some features reflecting multiple compounds. Creatinine and 3-IS were quantified by UHPLC-MS/MS using an ExionLC-30AD HPLC system coupled with a Triple Quad 6500+ mass spectrometer (AB Sciex, Germany, Darmstadt). 3-IS concentration was normalized to creatinine; quantification was based on calibration curves using stable isotope-labeled standards.

Metabolomic bucket selection

To address dataset heterogeneity, NMR spectral bucket selection was performed using three complementary feature selection approaches: feature importance based on machine learning algorithms (Random Forest, xgbTree, glmnet), selection by filtering based on random forest, and Boruta. The resulting feature set was refined by combining outputs and removing redundancy. Details can be found in the Additional File 4.

Predictive modeling

To systematically assess the additive value of multi-omics integration, integrative modeling evaluated microbiome and metabolome features through four sequential layers: (1) clinical data alone, (2) clinical data combined with microbiome data, (3) clinical data combined with selected metabolomic buckets, and (4) all three combined. Candidate predictors included demographic, clinical, comorbidity and severity-of-illness variables (see Additional File 5 for details), along with Shannon diversity, *Enterococcus* dominance (UHC subset only), and selected urinary metabolomic buckets. Covariate selection was performed via stepwise logistic regression in forward, backward, and bidirectional modes, informed by model performance metrics including the Akaike information criterion (AIC) and Nagelkerke's pseudo- R^2 . The final model refinement was guided by manual review and expert consensus. Collinearity among candidate predictors was systematically assessed via variance inflation factor (VIF) analysis, with predictors exhibiting $VIF > 5$ excluded to ensure coefficient stability and interpretability.

Classification models

Machine learning classification models were developed by training (80%) and testing (20%) splits, with recursive feature elimination (RFE) identifying optimal feature subsets, ensuring class balance and avoiding reuse of data between training and testing. Again, four data configurations were tested: (1) clinical data alone, (2) clinical data combined with microbiome data, (3) clinical data combined with selected metabolomic buckets, and (4) all three combined (candidate predictors detailed in Additional File 5). The models were trained using 5-fold cross-validation repeated five times. Model performance was evaluated via the Matthews correlation coefficient (MCC) (31), and the receiver operating characteristic (ROC) curve with area under the curve (AUC) (32).

Survival analysis

Survival estimates were calculated using the Kaplan–Meier method with log-rank testing. Overall survival spans from enrollment to death from any cause, with survivors censored at the time of last contact.

Results

Patient characteristics

The core dataset comprised 88 patients, of whom 25 (28%) developed an SI at a median of 11 days following enrollment. The median age was 60 years; 45 patients (52%) were male. The UHC subset comprised 80 patients, with 23 (28%) developing an SI. The median age was 58 years; 46 patients (57%) were male. Demographics were

broadly similar across data sets and groups; however, prior antibiotic exposure significantly associated with SI development in both datasets (core dataset: 21 [84%] vs. 26 [41%], $q = 0.01$; UHC subset: 18 [78%] vs. 19 [33%], $q < 0.01$). In the UHC subset, SI patients demonstrated a higher prevalence of hemato-oncological comorbidities (19 [83%] vs. 28 [49%], $q = 0.045$), along with increased use of immunosuppressives (11 [48%] vs. 9 [16%], $q = 0.02$) and chemotherapy (11 [48%] vs. 8 [14%], $q = 0.01$). These patients further exhibited significantly lower platelet counts (75 vs. $180 \times 10^9/L$, $q = 0.046$), hemoglobin (8.7 vs. 11.1 g/dL, $q < 0.01$), and hematocrit levels (23% vs. 31%, $q < 0.01$) (see Table 1 and Supplementary Table S1, Additional File 1 for details on the UHC subset).

Table 1: Patient characteristics

Demographic and clinical characteristics of patients included in the core dataset. Comparative summary of the overall dataset and subgroups stratified by secondary infection (SI) status.

	Total dataset ($n = 88$) n (%), median (IQR)	Patients without SI ($n = 63$) n (%), median (IQR)	Patients with SI^a ($n = 25$) n (%), median (IQR)	q value^b
Age (years)	60 (47, 71)	62 (51, 72)	55 (42, 66)	0.56
Male gender	45 (52%)	34 (54%)	11 (46%)	1.00
ICU admission characteristics				
Origin				0.04
In-house hospital ward	28 (33%)	14 (23%)	14 (58%)	
Emergency department	39 (46%)	34 (56%)	5 (21%)	
External hospital	18 (21%)	13 (21%)	5 (21%)	
Admission diagnosis^c				
Respiratory failure	60 (68%)	43 (68%)	17 (68%)	1.00
Sepsis	22 (25%)	15 (24%)	7 (28%)	1.00
Shock	13 (15%)	11 (17%)	2 (8.0%)	0.91
Neurological condition	13 (15%)	9 (14%)	4 (16%)	1.00
Cardiovascular condition	12 (14%)	10 (16%)	2 (8.0%)	1.00
Other	12 (14%)	9 (14%)	3 (12%)	1.00
Antibiotic treatment preceding ICU admission^d	47 (53%)	26 (41%)	21 (84%)	0.01
Vasopressor administration at baseline	80 (92%)	58 (92%)	22 (92%)	1.00
Administration of anti-infective agents up to day 90	88 (100%)	63 (100%)	25 (100%)	-
Clinical and laboratory data at baseline				
Platelet count [x $10^9/L$]	168 (75, 269)	180 (91, 270)	109 (62, 243)	0.75

Creatinine [mg/dL]	1.2 (0.8, 2.1)	1.1 (0.8, 2.1)	1.3 (0.8, 2.0)	1.00
Bilirubin [mg/dL]	0.6 (0.3, 1.1)	0.6 (0.3, 1.0)	0.9 (0.3, 2.7)	0.56
MAP [mmHg]	73 (63, 80)	70 (61, 80)	75 (65, 80)	1.00
p _a O ₂ [mmHg]	98 (76, 122)	99 (79, 126)	91 (74, 116)	0.75
F _i O ₂ [%]	43 (36, 60)	45 (37, 60)	40 (36, 50)	1.00

^a UHC: 17 patients; CUMC: 8 patients

^b Wilcoxon rank sum test, Pearson's chi-squared test, Fisher's exact test; false discovery rate correction for multiple testing

^c This variable allowed for multiple answers.

^d Within six months preceding ICU admission, including treatment at the time of admission, coded as a binary variable (yes/no), encompassing both broad- and narrow-spectrum antibiotics

IQR interquartile range, *SI* secondary infection, *ICU* intensive care unit, *MAP* mean arterial pressure

Secondary infection characteristics

SIs occurred in 28% (25/88) of patients, 17% (15/88) experienced at least one episode of BSI and 16% (14/88) developed VAP; these categories were not mutually exclusive. The most common BSI pathogens were *Enterococcus* species, partly including vancomycin-resistant strains, and coagulase-negative *Staphylococcus* species. VAP was caused primarily by Gram-negative organisms, including *Klebsiella pneumoniae* (partly multidrug-resistant), *Pseudomonas aeruginosa*, and extended-spectrum beta-lactamase (ESBL)-producing *Escherichia coli* (see Supplementary Table S2, Additional File 2).

Survival outcomes

In the core dataset, median follow-up extended to 90 days; SI did not significantly impact overall mortality (34% mortality, $p = 0.79$). With more data available, the UHC subset (43% mortality, $p = 0.09$) showed significantly lower ICU and hospital survival among SI patients ($p = 0.048$ and $p < 0.01$, respectively), along with prolonged ICU (33 vs. 11 days) and hospital stays (69 vs. 18 days) (both $p < 0.01$; see Supplementary Figures S7 and S8, Tables S7 and S8, Additional File 6).

The intestinal and lower respiratory microbiome

In the core dataset, patients with subsequent SI demonstrated significantly reduced Shannon diversity at enrollment (1.96 vs. 3.47, $p < 0.01$; Figure 1). In the UHC subset with comprehensive profiling, SI patients exhibited substantially greater baseline *Enterococcus* abundance (46% vs. 11%, $q < 0.01$), with similar patterns at higher taxonomic levels, including the Bacilli class (61% vs. 19%, $q < 0.01$), and Firmicutes phylum domination (70% vs. 22%, $q < 0.01$). In contrast, patients not progressing to SI presented greater relative abundances of anaerobic taxa such as Clostridia (37% vs. 19%, $q = 0.04$), Bacteroidia (20% vs. 7%, $q < 0.01$) and Gammaproteobacteria

Figure 1: Differences in microbial composition and diversity relative to secondary infection (SI) status

(a) **Genus-level microbial abundance.** Bar plots display the relative abundance of the 15 most common bacterial genera identified in patient samples at baseline, comparing individuals without subsequent secondary infection (left panel, No SI) to those with subsequent secondary infection (right panel, SI). Each bar represents one patient, with colors indicating different genera. This visualization provides an overview of the dominant taxa, notably *Enterococcus*, and highlights compositional shifts associated with SI status. Statistical comparisons between groups are presented in Additional File 3, Supplementary Table S3a. (b) **LEfSe discriminant analysis of microbial taxa.** The cladogram summarizes taxa that best distinguish between SI and No SI groups, identified using Linear Discriminant Analysis Effect Size (LEfSe). The circular plot organizes bacteria by phylogenetic relationships, with branch colors marking taxa enriched in either SI (green) or No SI (red) patients. Taxa displayed here contribute most strongly to differences between groups. Corresponding LDA (discriminative effect size) scores are reported in Additional File 3, Supplementary Figure S2b. (c) **α -diversity comparison.** Boxplots show microbial α -diversity (Shannon diversity index) for both the core dataset (left) and the UHC subset (right), stratified by SI status. The Shannon index measures both the richness (number of types) and evenness (distribution among types) of bacteria present within each sample; higher values indicate greater diversity. Statistical results and additional diversity metrics can be found in Additional File 3, Supplementary Table S3c. (d) **Microbial community structure (β -diversity).** Principal Coordinate Analysis (PCoA) plots visualize overall differences in microbial community structure between SI and No SI groups, based on weighted UniFrac distances (which account for both the presence and evolutionary relationships of taxa). Each point represents a patient's sample; solid ellipses indicate the 95% confidence intervals of the group centroids, while dashed ellipses represent the area expected to contain 95% of the data points for each group, assuming a multivariate normal distribution. The percentage of variation explained by each axis is labeled. Results using alternative distance metrics are presented in Additional File 3, Supplementary Figure S2a.

Urinary features

NMR analysis identified several key spectral features associated with SI, as detailed in Additional File 4. Notably, spectral features at 0.925 and 0.935 ppm corresponded primarily to 2-oxoisocaproate and isoleucine, although partial overlap with protein signals was observed. As illustrated in Supplementary Figure S5 of Additional File 4, both 2-oxoisocaproate and, to a lesser degree, isoleucine contribute significantly to the NMR bucket, which comprises signals with chemical shifts ranging from 0.930 and 0.940 ppm. After creatinine normalization, urinary levels of both metabolites were elevated in SI patients relative to non-SI controls, with 2-oxoisocaproate as the

predominant component (Figure 2). Building on previous research (13), 3-IS concentrations and their influence on SI were additionally evaluated but did not differ significantly between groups in the core dataset (see Supplementary Figure S6, Additional File 4).

In the UHC subset, elevated quinolate (8.025 ppm) associated with SI, whereas 3-IS concentrations were significantly lower in patients with subsequent SI ($p = 0.01$). This latter association, however, did not retain significance in univariate logistic regression (see Figure 2 and Supplementary Figure S6, Additional File 4).

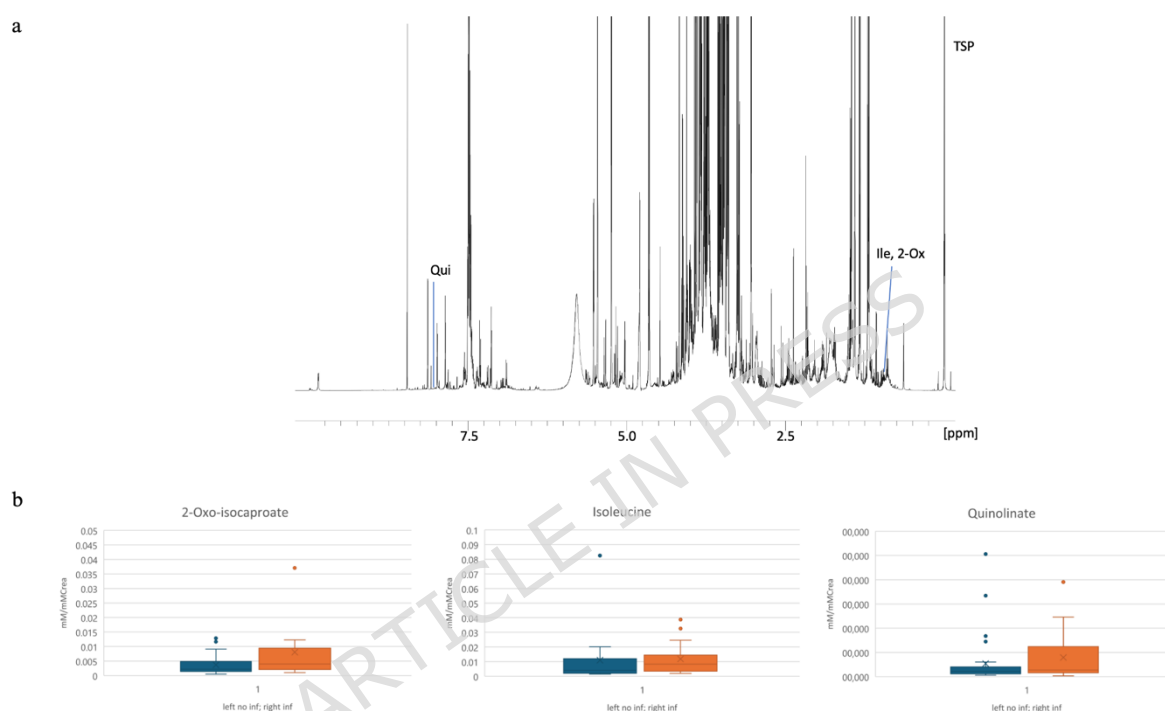


Figure 2: Urinary NMR analysis of metabolites and their creatinine-normalized levels in relation to subsequent secondary infection

- Selected urinary NMR buckets; all buckets have a width of 0.01 ppm; the bucket at 0.935 ppm corresponds to both isoleucine (Ile) and 2-oxoisocaproate (2-Ox), and the bucket at 8.025 ppm corresponds to quinolate (Qui).
- Creatinine-normalized absolutely quantified levels of 2-oxoisocaproate, isoleucine and quinolate in patients with and without subsequent secondary infection (inf)

Prediction of secondary infection through regression and classification analyses

Core dataset analysis

The integrated multi-omics model (clinical data + microbiome data + metabolome data) demonstrated improved performance ($AIC = 74.5$, $R^2 = 0.53$) compared to clinical-only models ($AIC = 104.71$, $R^2 = 0.10$). Key predictors

included the urinary NMR feature at 0.935 ppm (odds ratio (OR) = 4.41, 95% confidence interval (CI) 2.18–10.86, $p < 0.01$) and Shannon diversity (OR = 0.35, 95% CI 0.16–0.67, $p < 0.01$). Clinical variables including mean arterial pressure (MAP; OR = 2.12, 95% CI 1.06–4.56, $p = 0.04$) and bilirubin (OR = 1.78, 95% CI 1.05–4.27, $p = 0.07$) further contributed (Table 2).

Classification analysis showed that microbiome integration (Shannon diversity index) alone yielded marginal improvement (AUC 0.64 to 0.66, Δ AUC = 0.02; MCC 0.08 to 0.14). Metabolomic integration provided substantial gains (AUC 0.64 to 0.75, Δ AUC = 0.11; MCC 0.08 to 0.38). Addition of microbiome data to the metabolomic model conferred no additional benefit (AUC 0.75, MCC 0.36) (Figure 3).

UHC subset analysis

Multivariable modeling reinforced findings, with the comprehensive integrated model (AIC = 42.7, $R^2 = 0.61$) substantially outperforming clinical-only approaches (AIC = 96.3, $R^2 = 0.07$). Key predictors included the Acute Physiology and Chronic Health Evaluation II (APACHE II) score (OR = 3.21, 95% CI 1.26–10.5, $p = 0.01$), Shannon diversity (OR = 0.34, 95% CI 0.11–0.84, $p = 0.02$; Table 2), and *Enterococcus* dominance (OR = 4.96, 95% CI 1.32–19.77, $p = 0.02$ when modeled alone, OR = 2.3, 95% CI 0.4–12.99, $p = 0.34$ when modeled with clinical and metabolome data; Supplementary Table S7, Additional File 5).

Classification analysis demonstrated more substantial microbiome contribution in this dataset: microbiome integration (Shannon diversity index and *Enterococcus* dominance) alone achieved an AUC of 0.82 versus 0.67 for clinical data alone (Δ AUC = 0.15, MCC 0.52). Metabolomic integration similarly yielded an AUC of 0.83 and MCC of 0.52. Concurrent integration of microbiome and metabolome features with clinical variables achieved the highest predictive accuracy (AUC 0.88, MCC 0.52), indicating complementary information across omics layers (Figure 3).

Selected NMR features

The most informative urinary NMR features varied depending on the available clinical and microbiome data, with the 0.935 ppm feature (2-oxoisocaproate and, to a lesser degree, isoleucine) predominating in the core dataset and the 8.025 ppm feature (quinolinate) in the UHC subset (Table 2). However, alternative NMR feature selections yielded comparable predictive performance. Model fit indices were similar across different feature combinations, with AIC values ranging from 74.53 to 88.06 in the core dataset and 41.96 to 50.17 in the UHC subset, indicating that multiple metabolomic signatures could achieve similar discrimination. Detailed feature selection outcomes and model comparisons are presented in Table 2 and Supplementary Table S6 (Additional File 5).

Antibiotic exposure

When modeled in place of the other clinical parameters from the final models, prior antibiotic exposure emerged as a consistent predictor in the integrated model (core dataset: OR = 2.57, $p = 0.17$; UHC subset: OR = 2.10, $p = 0.42$). VIF analysis confirmed no substantial collinearity with Shannon diversity (VIF = 1.20) and *Enterococcus* dominance (VIF = 1.22), respectively. Nested model comparisons further demonstrate that microbiome features improve predictive performance over models with antibiotic exposure alone (Supplementary Table S6, Additional File 5).

Table 2: Multivariable regression analysis for secondary infection

Models were constructed using baseline clinical variables alone and in combination with microbiome and/or urinary metabolome data.

a) Core dataset

Characteristic (at baseline)	Clinical data (OR, 95% CI) ^{a,b}	Clinical + microbiome data (OR, 95% CI) ^{a,c}	Clinical + urinary metabolome data (OR, 95% CI) ^{a,d}	Clinical + microbi- ome + urinary metabolome data (OR, 95% CI) ^{a,e}
	AIC = 104.71	AIC = 94.3	AIC = 82.7	AIC = 74.5
Bilirubin	2.11 (1.13, 5.32; $p = 0.06$)	1.6 (0.96, 3.65; $p = 0.13$)	2.25 (1.2, 5.9; $p = 0.04$)	1.78 (1.05, 4.27; $p = 0.07$)
MAP	1.13 (0.68, 1.9; $p = 0.64$)	1.35 (0.21, 0.66; $p = 0.3$)	1.63 (0.89, 3.15; $p = 0.13$)	2.12 (1.06, 4.56; $p = 0.04$)
Shannon diversity		0.38 (0.21, 0.66; $p < 0.01$)		0.35 (0.16, 0.67; $p < 0.01$)
Urinary NMR fea- ture at 0.935 ppm^f			4.33 (2.21, 10.02; $p < 0.01$)	4.41 (2.18, 10.86; $p < 0.01$)

^a OR (95% confidence interval, p value)

^b $R^2 = 0.10$

^c $R^2 = 0.28$

^d $R^2 = 0.42$

^e $R^2 = 0.53$

^f Key feature identified on the basis of urinary NMR profiles in the core dataset ($n = 88$)

MAP mean arterial pressure, NMR nuclear magnetic resonance, ppm parts per million

b) UHC subset

Characteristic (at baseline)	Clinical data (OR, 95% CI) ^{a,b}	Clinical + microbiome data (OR, 95% CI) ^{a,c}	Clinical + urinary metabolome data (OR, 95% CI) ^{a,d}	Clinical + microbi- ome + urinary metabolome data
---------------------------------	--	--	--	---

	(OR, 95% CI) ^{a,e}			
	AIC = 96.3	AIC = 51.1	AIC = 59.2	AIC = 42.7
APACHE II score	1.66 (0.99, 2.93; $p = 0.06$)	2.88 (1.33, 7.51; $p = 0.02$)	1.89 (0.96, 3.97; $p = 0.07$)	3.21 (1.26, 10.5; $p = 0.01$)
Shannon diversity		0.27 (0.11, 0.58; $p < 0.01$)		0.34 (0.11, 0.84; $p = 0.02$)
Urinary NMR feature at 8.025 ppm^f			4.57 (2.07, 12.38; $p < 0.01$)	5.11 (1.8, 20.7; $p < 0.01$)

^a OR (95% confidence interval (CI), p value)

^b $R^2 = 0.07$

^c $R^2 = 0.43$

^d $R^2 = 0.41$

^e $R^2 = 0.61$

^f Key feature identified on the basis of urinary NMR profiles in the subset of UHC patients ($n = 64$)

APACHE II Acute Physiology-Age-Chronic Health Evaluation II, *spp.* species, *NMR* nuclear magnetic resonance, *ppm* parts per million

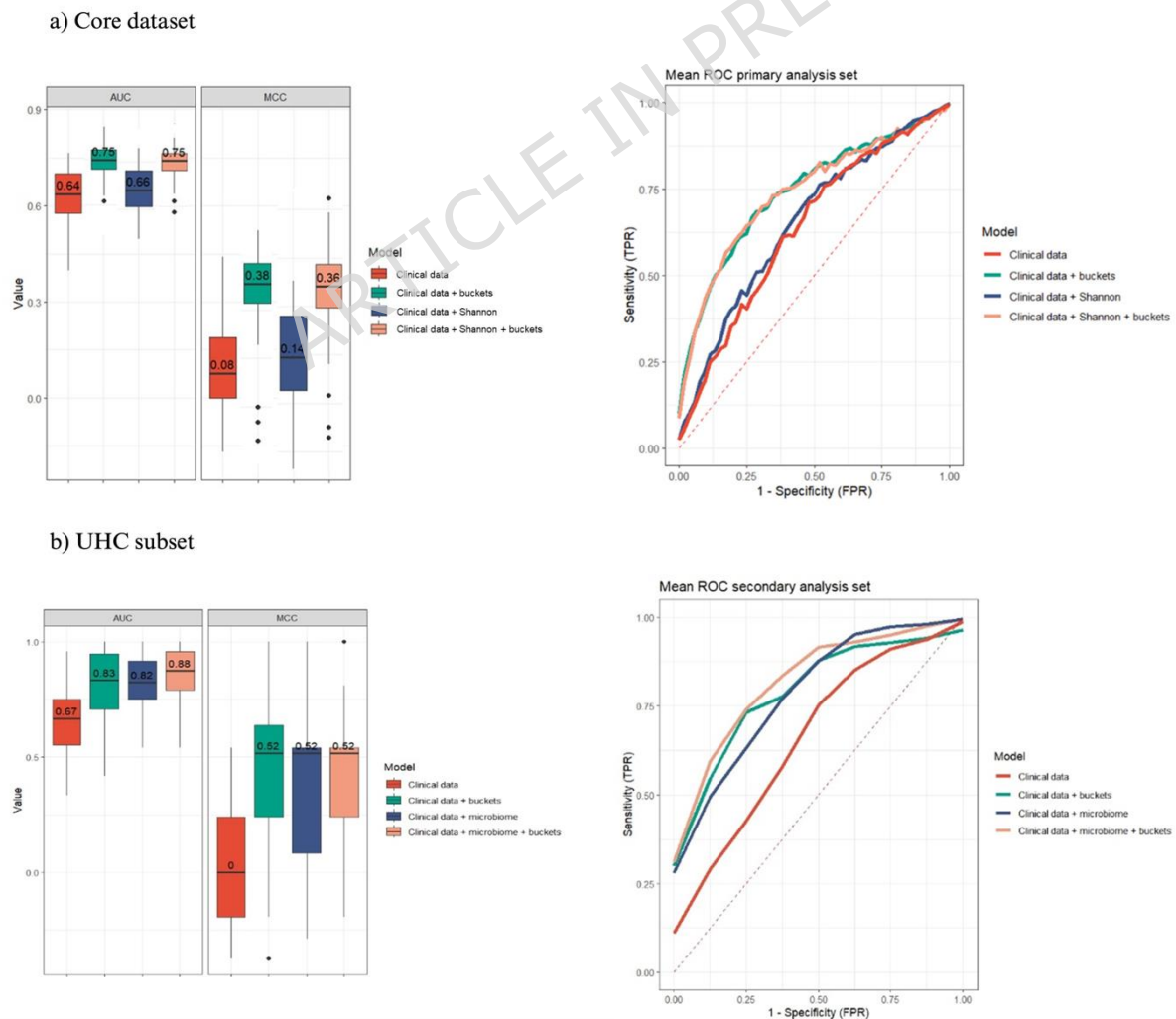


Figure 3: Risk factor analysis for secondary infection

Classification performance results for four data configurations, based on data from the 88 patients in the core dataset (a) and the 55 patients in the UHC subset (b). Left: the box plots illustrate the performance of 50 models across each data configuration on the basis of AUC and MCC scores. Right: the figure shows the mean ROC curves based on the 50 models for each data configuration

Discussion

SI remains a significant challenge in the ICU, prolonging critical illness, increasing mortality, and driving up healthcare costs (1, 2). In this study, SI associated with shorter median survival, reduced overall survival in both ICU and hospital settings, and longer lengths of stay across care environments. The clinical burden of these infections underscores an important gap: despite growing mechanistic insight into host- and pathogen-related determinants of SI, current preventive and diagnostic strategies remain inadequate for early and precise patient risk identification. This work addressed this gap by identifying intestinal microbiome and urinary metabolome features associated with SI development, investigating cross-site microbiota convergence patterns, and determining whether integration of multi-omics data with standard clinical assessment enhances predictive discrimination compared to clinical evaluation alone.

Five key observations emerged from this analysis. First, SI development associated with conventional clinical severity markers, including MAP, bilirubin, and APACHE II scores. Second, intestinal microbiota disruption — characterized by reduced diversity and *Enterococcus* dominance — represented an independent and significant risk factor for SI. Third, antibiotic-induced microbial disruption manifests coordinately across gastrointestinal and respiratory sites. Fourth, distinct urinary metabolomic features mapping to branched-chain amino acids (2-oxoisocaproate, isoleucine) and tryptophan metabolites (quinolinate) emerged as predictors of infection. And fifth, multi-omics integration enhanced predictive accuracy in a context-dependent manner: while metabolomic features consistently improved performance across both datasets, microbiome contributions demonstrated variability dependent upon the specificity of microbial metrics and the dataset employed.

Clinical risk factors

Established clinical severity indicators retained significant prognostic value for SI prediction in this analysis (33, 34). MAP and bilirubin, both central components of the Sequential Organ Failure Assessment (SOFA) score (35), as well as the APACHE II score in the UHC subset (36), emerged as consistent predictors. These findings align

with the well-established relationship between worsening organ dysfunction and infection susceptibility. In the UHC subset, SI patients presented with a higher prevalence of malignancy, consistent with recognized infection risks in immunocompromised populations (37). This association likely reflects both the disease burden inherent to underlying malignancy and the established predisposition to poor outcomes characteristic of cancer patients — features often reflected in elevated severity scores (38, 39).

Microbiome disruption as a key determinant of infection risk

Beyond established clinical factors, microbiome alterations are increasingly recognized as key determinants of adverse outcomes in ICU patients (4). In our study, baseline *Enterococcus* dominance ($\geq 30\%$ relative abundance) significantly associated with infection development, consistent with previous reports linking *Enterococcus* abundance to poor outcomes (6, 7, 40). This association likely reflects the increased prior antibiotic exposure observed in patients progressing to SI. The ecological mechanisms driving this relationship are well-characterized. Broad-spectrum antibiotics, especially those with anaerobic coverage, disrupt normal microbiome diversity and create ecological niches favorable for *Enterococcus* proliferation, frequently resulting in vancomycin-resistant *Enterococci* (VRE) colonization (40). Studies have shown that ICU patient microbiota can lose diversity within 72 hours of antibiotic exposure, facilitating *Enterococcus* dominance linked to VRE colonization and elevated mortality (6, 7). Notably, *Enterococcus* dominance at ICU admission has been associated with 19% and 22% increases in mortality and infection risk, respectively, with 77% of such patients exhibiting VRE positivity (7). In transplant populations, *Enterococcus* dominance increases VRE-derived BSI risk nine-fold (40).

Reduced microbial diversity independently associated with SI development across both study datasets. Higher baseline diversity conferred protective effects, a pattern well-documented in immunocompromised populations (8). However, traditional diversity metrics may fail to capture critical ecological shifts such as opportunistic pathogen blooms, and their predictive capacity varies with patient population and immune status (6). In this study, approximately two-thirds of the core dataset originated from a specialized medical ICU focused on hematology-oncology patients, in whom preservation of microbial diversity appeared particularly important. This likely reflects the heightened immune reconstitution needs characteristic of this population; a relationship potentially less pronounced in more heterogeneous ICU settings (7).

Emerging evidence supports the gut–lung axis as a pathway for microbial translocation and cross-site disruption. In this analysis, respiratory microbiota showed parallel patterns of increased *Enterococcus* abundance and reduced diversity in SI patients, paralleling intestinal patterns. These likely antibiotic-induced ecological shifts

contribute to adverse outcomes through loss of colonization resistance across multiple mucosal sites, reinforcing the interdependence of microbial communities in infection susceptibility (41).

Mechanistic insights into urinary metabolites

Realizing the clinical potential of colonization risk assessment requires reliable, rapidly measurable biomarkers. Urinary metabolites combine the advantages of non-invasive sample collection and compatibility with routine clinical workflows, coupled with its metabolic richness and relative freedom from protein and cellular interference (42-45). In this study, NMR spectroscopy identified multiple metabolomic features associated with subsequent SI development, with the bucket at 0.935 ppm and 8.025 ppm emerging as the most informative discriminators within final multivariable models. These features were selected in the context of combined clinical and microbiome data, yielding two distinct but biologically plausible signatures across the partially overlapping datasets and underscoring anticipated synergistic interactions among clinical, microbial, and metabolic factors. Importantly, these associations remain correlative rather than causal; shared confounders of critical illness pathophysiology may underlie the observed patterns.

The NMR feature at 0.935 ppm comprises 2-oxoisocaproate and isoleucine, both closely linked to branched-chain amino acid (BCAA) metabolism and identified as key discriminators of subsequent SI in the core dataset. Their urinary accumulation likely reflects broader disruptions in host–microbial BCAA metabolism, potentially resulting from depletion of BCAA-consuming taxa, systemic inflammation, or catabolic stress — conditions common in critical illness and frequently accompanied by renal dysfunction. Experimental data suggest that ketoacids such as 2-oxoisocaproate can modulate macrophage polarization toward pro-tumoral or pro-inflammatory macrophage states, providing a biologically plausible link between BCAA catabolism and immune function (46). Elevated isoleucine, in turn, has been associated with chronic metabolic and inflammatory states, including liver disease (47) and type 2 diabetes (48). It further acts as an immunonutrient that supports immune organ development and function, promotes host-defense peptide expression, and enhances antimicrobial responses in experimental models (49, 50). In this context, increased urinary isoleucine in patients who later develop SI may represent a compensatory host response to dysbiosis-associated or critical-illness–related immune dysfunction. Although vitamin deficiencies — thiamine (51) and pyridoxine (52) — could theoretically explain elevated levels of 2-oxoisocaproate and isoleucine, routine ICU supplementation and vitamin-enriched nutrition make these etiologies unlikely.

The signal at 8.025 ppm corresponds to quinolinate, a kynurenine pathway intermediate derived from tryptophan degradation, which consistently discriminated SI in the UHC subset. Urinary elevation primarily reflects inflammation-induced activation of host indoleamine 2,3-dioxygenase, the rate-limiting enzyme, driven by proinflammatory cytokines. Reduced microbial tryptophan catabolism by depleted commensals (*Clostridia*/*Bacteroidia*), supported by lower 3-IS in SI patients, may secondarily increase substrate availability for this pathway. Functionally, quinolinate serves as a central intermediate in the de novo synthesis of nicotinamide adenine dinucleotide (NAD⁺), essential for cellular energy metabolism, DNA repair, and stress response regulation. During immune activation, increased quinolinate may reflect a metabolic shift toward NAD⁺ production to support host defense, while reduced quinolinate phosphoribosyltransferase activity, impairing conversion to downstream precursors like nicotinic acid mononucleotide, likely promotes accumulation under inflammatory conditions (53-55). Clinically, elevated quinolinate has been linked to adverse outcomes (56) and neurotoxic, pro-inflammatory effects in other infectious and inflammatory settings, supporting its relevance as a host-response marker (57).

Notably, 2-oxoisocaproate, isoleucine, and quinolinate are largely host-derived metabolites, with microbial communities modulating their levels at varying degrees. This indicates that urinary metabolites capture a host-metabolic dimension of SI risk that is not fully reflected by microbiota composition alone, thereby complementing the information contained in microbiome profiles.

The complementary predictive value of multi-omics integration

Building on these metabolite insights, integration of multi-omics data into predictive modeling consistently demonstrated improvements over traditional clinical models, though the extent and nature of these gains varied substantially with dataset characteristics and omics modalities employed. In the core dataset, microbiome integration (Shannon diversity index alone) yielded only marginal predictive improvement ($\Delta\text{AUC} = 0.02$), whereas the more comprehensively characterized UHC subset demonstrated substantial enhancement ($\Delta\text{AUC} = 0.15$) when microbiome data were added (including both the Shannon diversity and *Enterococcus* dominance). This discrepancy likely reflects the differential discriminatory capacity of these microbial metrics: while Shannon diversity quantifies overall community richness and evenness (15), it provides limited independent prognostic information. In contrast, mechanistically informative features such as *Enterococcus* dominance confer substantially greater predictive value. These observations are consistent with systematic reviews of lower gut dysbiosis in critical illness, where reductions in α -diversity do not consistently associate with in-hospital mortality, while specific taxa abundance, including *Enterococcus* species, serve as stronger predictors of clinical outcomes (6).

Recent evidence further underscores the advantage of microbiome data integration: autoencoder-based integration of multi-compartmental microbiome data with clinical variables consistently yields superior prediction accuracy, achieving up to 98% accuracy for mortality in lung microbiome analyses, while microbiome taxa alone perform markedly weaker (53-65%) (58). Similarly, hierarchical machine learning incorporating time-series microbiome and clinical data for predicting growth faltering in preterm infants has demonstrated marked improvements in predictive precision and cost efficiency compared to clinical data alone (59).

Urinary metabolomic integration, by contrast, consistently improved predictive performance in both datasets. The UHC cohort achieved comparable accuracy from either metabolome or microbiome data alone. Crucially, concurrent integration of both omics layers with clinical variables achieved the highest predictive accuracy (AUC 0.88 in UHC subset), indicating complementary information across distinct biological domains. These findings underscore a fundamental principle: metabolomic signatures appear to capture generalizable immunometabolic states and microbiome-related disruptions relevant to infection susceptibility across heterogeneous clinical contexts. Microbiome features, conversely, exhibit context-dependent predictive value varying substantially based on the specificity and biological informativeness of metrics employed.

Urinary metabolomics emerging clinical applications in critical illness and infection risk

Accumulating evidence supports urinary metabolomics as a feasible and informative approach for stratifying critically ill populations and predicting adverse outcomes. Urinary metabolic profiles have been linked to cardiovascular risk in type 1 diabetes (60) and progression of chronic kidney disease (61). Further, urinary metabolites mapping to amino acid biosynthesis and glycolipid metabolism associated with immunoglobulin A nephropathy progression (AUCs exceeding 0.9) (62). In cirrhotic patients, serum and urinary metabolites involved in S-adenosyl methionine and tryptophan metabolism have shown prognostic value for predicting acute kidney injury and dialysis requirement (63). Similarly, combat-injured patients exhibited associations between urinary lactate, glycine, and 1-methylnicotinamide levels and mortality or renal replacement therapy need (64), supporting the feasibility of urine-based prognostic models in critically ill populations.

More specifically regarding infection risk and aligned with current findings on quinolinate and 3-IS, recent studies demonstrate urinary metabolomics' capacity to capture alterations in the tryptophan-kynurenine pathway. In hospitalized COVID-19 patients, elevated urinary kynurenine, 3-hydroxykynurenine, and 3-hydroxyanthranilate correlated with systemic inflammation and disease severity (65). In surgical ICU patients, plasma quinolinic acid demonstrated very good discrimination between non-septic and pre-septic patients, supporting our findings that tryptophan pathway activation predicts infection risk (66). Gut-derived uremic toxins such as 3-IS

have been proposed as candidate biomarkers of organ dysfunction and may reflect microbiome activity relevant to infection susceptibility (13). In allogeneic stem cell transplant recipients, low urinary 3-IS levels associated with significantly increased transplant-related mortality, reinforcing the protective role of this microbiome-derived metabolite (12). Emerging work further demonstrates urinary metabolomics' ability to discriminate pathogen-specific infection phenotypes — unique urinary metabolite patterns distinguish *Pseudomonas aeruginosa* VAP from both non-*Pseudomonas* VAP and uninfected controls (67). Lastly, urine profiling has identified agmatine and N6-methyladenine as markers accurately predicting urinary tract infection pathogens across large cohorts (68, 69).

Methodological considerations and limitations

Several constraints warrant explicit acknowledgment as they impact appropriate interpretation and generalizability of our findings. The modest patient numbers ($n = 88$ core dataset, $n = 55$ UHC subset) combined with clinical heterogeneity result in moderate statistical power and an increased risk of overfitting, particularly given the high dimensionality of the NMR metabolomic dataset (1,029 spectral buckets). The results therefore reflect within-cohort performance without formal external validation, limiting confidence in model transportability to other ICU settings.

Post hoc integration of two geographically and clinically distinct cohorts was undertaken to enhance statistical power and generalizability, yet inherently introduces heterogeneity through differences in patient demographics, healthcare systems, clinical practices, infection-control protocols, and specimen collection procedures. These differences may confound biological signals and bias feature selection toward cohort-specific associations, thereby limiting generalizability to other ICU contexts.

From a microbiological perspective, 16S rRNA gene sequencing provides only genus- to species-level taxonomic resolution, precluding strain-level discrimination. Cross-cohort variations in sequencing workflows introduce batch effects potentially confounding biological interpretation.

Spectral overlap in untargeted NMR metabolomics limits definitive metabolite annotation, such that bucket integrals function as proxies for underlying metabolic pathways. Consequently, the reported associations involving 2-oxoisocaproate, isoleucine, and quinolinate should be interpreted cautiously and independently validated through orthogonal quantitative assays to confirm metabolite identities and enable mechanistic interpretation.

Characterization of antibiotic exposure was limited to clinical documentation without stratification by drug class, dosing, or pharmacokinetics, precluding quantitative assessment of selective pressure and class-specific effects on microbiota composition.

Finally, single baseline sampling provides only a metabolic and microbial snapshot; dynamic alterations preceding sampling or emerging prior to infection onset remain uncharacterized. This design obscures temporal relationships between predictive marker detection and infection development, precluding determination of whether identified signatures represent primary drivers of infection susceptibility or downstream consequences of early dysbiosis. The dynamic trajectory of microbiota-metabolome alterations during the at-risk period remains unexplored.

Clinical translation requirements

Given these constraints, modeling results should be interpreted as proof-of-concept rather than a ready-for-deployment clinical prediction tool. External validation and further mechanistic investigation are essential prerequisites before clinical translation. Prospective studies must validate findings across larger, more homogeneous patient populations while examining interactions between underlying pathophysiology (e.g., sepsis, acute respiratory distress syndrome, post-surgical critical illness) and dysbiotic trajectories. Ideally, true external validation through training on one cohort and prospective testing on an independent population would rigorously establish model transportability.

Longitudinal profiling at frequent intervals during early ICU admission will enable trajectory analysis determining whether progressive microbiome disruption predicts infection development with greater sensitivity than baseline assessment. Emerging evidence suggests that biomarker kinetics over time confer superior prognostic value compared to single-timepoint measurements (13); our preliminary data reveal microbiome convergence patterns as early as enrollment, emphasizing the critical importance of temporal resolution in capturing dysbiotic trajectories. Extending this longitudinal approach to evaluate the influence of antibiotic exposure on microbiome evolution will be essential to differentiate treatment effects from intrinsic disease dynamics. Furthermore, comparing recovery trajectories in successfully treated versus secondarily infected patients may help identify dynamic microbiome and metabolome markers with higher predictive potential than static measurements.

Mechanistic validation remains essential for translational impact. Targeted quantitative assays should confirm associations with 2-oxoisocaproate, isoleucine, and quinolinate, establishing absolute concentrations and validating amino acid metabolism dysregulation as a microbiome-derived mechanism of infection susceptibility. The differential selection of NMR features between datasets further suggests context-dependence of metabolomic

signatures, warranting systematic exploration of geographic and clinical variations in biomarker selection across cohorts to clarify whether universal signatures exist or whether context-specific markers are required for optimal performance.

A critical recognition must frame all future work: both 16S rRNA sequencing and NMR metabolomics remain research tools unsuitable for real-time clinical decision-making. Sequencing requires a few hours up to days for sample processing and analysis; NMR similarly demands substantial analytical infrastructure and technical expertise. Bridging this technology gap demands development of rapid, point-of-care assays that preserve the mechanistic insights identified herein while meeting practical ICU requirements: minimal specimen handling, turnaround time of hours rather than days, cost-effectiveness permitting routine deployment, and interpretability for clinical personnel without specialist training.

Conclusion

This systematic evaluation of multi-omics integration provides modest but consistent evidence for the added value of incorporating intestinal microbiome and urinary metabolome data with standard clinical assessment for SI prediction in critically ill patients. SI susceptibility associates with clinical severity markers, microbiota disruption (particularly diminished microbial diversity and *Enterococcus* dominance), and dysregulated amino acid and tryptophan metabolism reflected in urinary metabolite alterations. Multi-omics integration enhanced predictive discrimination dependent upon dataset and metric specificity.

These findings establish a foundation for prospective validation studies and mechanistic investigation, bringing closer a future precision medicine approach to infection prevention in the ICU. However, substantial research and technical development remain necessary before these biomarkers and analytical approaches transition from research tools to clinically deployable interventions.

Declarations

- **Ethics approval and consent to participate:** Informed consent was obtained from participants or legal representatives. The study was approved by local institutional review boards (University Hospital Cologne #17-333, Columbia University Medical Center #AAAR5608), and was conducted in accordance

with the Declaration of Helsinki (Fortaleza, 2013) and the International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use – Good Clinical Practice guidelines (ICH-GCPs).

- **Consent for publication:** not applicable
- **Availability of data and materials:** The datasets used and/or analyzed during the current study are available from the corresponding author upon reasonable request.
- **Competing interests**
 - **KD** reports grants not related to this study from the DFG (TRR374), the Bavarian Ministry of Science and Art (Metabodefense), and the German Federal Ministry of Research and Education (ERA-NET Transcan-Neu II - NK-4-GBM).
 - **WG** reports grants not related to this study from the DFG (TRR374) and the German Federal Ministry of Research and Education (TAhRget).
 - **PJO** reports grants not related to this study from the DFG (TRR374) and the German Federal Ministry of Research and Education (TAhRget and ERA-NET Transcan-Neu II - NK-4-GBM).
 - **DEF** reports grants from the U.S. Department of Defense (PRP181960), CARB-X (with Seres Therapeutics), Otsuka, and Kyowa Kirin. He has consulted for KeepBio.
 - **PK** reports grants or contracts from the German Federal Ministry of Research and Education (BMBF) B-FAST (Bundesweites Forschungsnetz Angewandte Surveillance und Testung) and NAPKON (Nationales Pandemie Kohorten Netz, German National Pandemic Cohort Network) of the Network University Medicine (NUM), the State of North Rhine-Westphalia and the Dr.Heinz Lux-Stiftung; Consulting fees Ambu GmbH, Gilead Sciences, infill healthcare communication GmbH, Mundipharma Research Limited, Noxxon N.V. and Pfizer Pharma; Honoraria for lectures from Akademie für Infektionsmedizin e.V., Ambu GmbH, Astellas Pharma, BioRad Laboratories Inc., Datamed GmbH, European Confederation of Medical Mycology, Gilead Sciences, GPR Academy Ruesselsheim, HELIOS Kliniken GmbH, Jazz Pharmaceuticals Germany GmbH, Lahn-Dill-Kliniken GmbH, medupdate GmbH, MedMedia GmbH, MSD Sharp & Dohme GmbH, Pfizer Pharma GmbH, Sanofi-Aventis Deutschland GmbH, Scilink Comunicación Científica SC, streamedup! GmbH, University Hospital and LMU Munich and VITIS GmbH; Participation on an Advisory Board from Ambu GmbH, Gilead Sciences, Mundipharma Research Limited and Pfizer Pharma; A filed patent at the German Patent and Trade

Mark Office (DE 10 2021 113 007.7); Other non-financial interests from Elsevier, Wiley and Taylor & Francis online outside the submitted work.

- **JGB** reports grants and travel expenses from KITE/Gilead outside of the submitted work.
- **JHN** reports research funding from the Else Kröner-Fresenius Foundation (EKFS) for research on ECMO-induced coagulopathy, but reports no conflicts of interest in regard to the present study.
- **MJGTV** reports grants or contracts from MSD, Heel, Roche, and Tillotts; consulting fees from Gilead, Tillotts, Pfizer, Bioaster, GSK, and Ecraid; and honoraria for lectures, presentations, speakers bureaus, manuscript writing, or educational events from Akademie für Infektionsmedizin, AstraZeneca, bioMérieux, DGI, European Society of Neurogastroenterology, Falk Foundation, FomF GmbH, Gilead, GSK, Helios Kliniken, Hessisches Landessozialgericht, Infektio Forum, Janssen Cilag GmbH, Klinikum Kassel, Landesärztekammer Hessen, LMU Kliniken, MSD, Pfizer, Streamed Up, St. Vincent Hospital, and Tillotts.
- **BB** reports lecture and advisory fees from Amgen, Kite/Gilead, MSD, Miltenyi, Noscendo, Novartis, Pfizer, Roche and Takeda as well as financing for research projects from Astellas, Celgene, Kite/Gilead, MSD, and Takeda outside of the submitted work.
- **All other authors report no potential conflicts of interest.**
- **Funding:** This study was supported by an unrestricted grant from MSD, and by the Hessian Ministry of Science and Research, Arts and Culture within the LOEWE research-initiative ACLF-I.
- **Authors' contributions**
 - **CL:** data acquisition (chart review), microbial data acquisition and interpretation (specimen processing), comprehensive statistical analysis, data interpretation, visualization, original manuscript drafting, manuscript review and editing
 - **KT:** statistical analysis (metabolomic feature selection and classification models), visualization, manuscript review and editing
 - **KD:** metabolomic data acquisition and interpretation (mass spectrometry), manuscript review and editing
 - **LE:** metabolomic data acquisition and interpretation (specimen processing and NMR spectroscopy)
 - **PJO:** supervision, manuscript review and editing

- **WG**: metabolomic data acquisition and interpretation (NMR spectroscopy), statistical analysis (foundational metabolomics-based models), visualization, original manuscript drafting, manuscript review and editing
- **FF**: microbial data acquisition and interpretation (bioinformatics), visualization, manuscript review and editing
- **AMR**: microbial data acquisition (specimen processing, library preparation, sequencing), manuscript review and editing
- **DEF**: resources (Columbia University Medical Center cohort), manuscript review and editing
- **PK**: patient management, manuscript review and editing
- **JGB**: patient management, data acquisition (chart review), manuscript review and editing
- **JHN**: patient management, data acquisition (chart review), manuscript review and editing
- **MJGTV**: conceptualization, funding acquisition, project administration and supervision, data interpretation, manuscript review and editing
- **BB**: conceptualization, funding acquisition, project administration and supervision, patient management, comprehensive statistical analysis, data interpretation, original manuscript drafting, manuscript review and editing
- All authors approved the final draft of the manuscript and agreed on submitting it to *Critical Care*.
- **Acknowledgments**: The authors would like to thank all patients and their relatives for participating in the study, and sincerely acknowledge the commitment of ICU physicians and nurses to patient care. Additionally, we extend our gratitude to Matthias Kochanek, Alexander Shimabukuro-Vornhagen and Dennis Eichenauer for their contributions to this work.

Additional material

- Additional File 1 (Additional_file_1.pdf): patient enrollment, study design, and clinical characteristics of the UHC subset.
- Additional File 2 (Additional_file_2.pdf): secondary infection characteristics.
- Additional File 3 (Additional_file_3.pdf): microbiome analyses and corresponding extended findings.
- Additional File 4 (Additional_file_4.pdf): urine analyses and corresponding extended findings.
- Additional File 5 (Additional_file_5.pdf): classification analysis, missing data, and extended findings of the multivariable regression analysis.

- Additional File 6 (Additional_file_6.pdf): survival analysis.

List of abbreviations

- **3-IS** 3-indoxyl sulfate
- **ACE** Abundance-based coverage estimator
- **AIC** Akaike information criterion
- **APACHE II** Acute Physiology-Age-Chronic Health Evaluation II
- **AUC** Area under the curve
- **BAL** bronchoalveolar lavage
- **BCAA** branched-chain amino acid
- **BSI** bloodstream infection
- **CDC** Centers for Disease Control and Prevention
- **CUMC** Columbia University Medical Center
- **ESBL** extended-spectrum beta-lactamase
- **ETA** endotracheal aspirates
- **FDR** false discovery rate
- **ICU** intensive care unit
- **IQR** interquartile range
- **LDA** linear discriminant analysis
- **LEfSe** linear discriminant analysis effect size
- **MAP** mean arterial pressure
- **MCC** Matthews correlation coefficient
- **NAD⁺** nicotinamide adenine dinucleotide
- **NMR** nuclear magnetic resonance
- **OR** odds ratio
- **PCoA** Principal coordinate analysis
- **PD** phylogenetic diversity
- **RFE** recursive feature elimination
- **RKI** Robert Koch Institute
- **ROC** receiver operating characteristic

- **SI** secondary infection
- **SOFA** Sequential Organ Failure Assessment
- **SR** species richness
- **UHC** University Hospital Cologne
- **UHPLC–MS/MS** Ultra-High Performance Liquid Chromatography–Tandem Mass Spectrometry
- **VAP** ventilator-associated pneumonia
- **VIF** variance inflation factor
- **VRE** vancomycin-resistant *Enterococci*

References

1. Vincent J-L, Sakr Y, Singer M, Martin-Loeches I, Machado FR, Marshall JC, et al. Prevalence and Outcomes of Infection Among Patients in Intensive Care Units in 2017. *JAMA*. 2020;323:1478-87.
2. Zimlichman E, Henderson D, Tamir O, Franz C, Song P, Yamin CK, et al. Health care-associated infections: a meta-analysis of costs and financial impact on the US health care system. *JAMA Intern Med*. 2013;173:2039-46.
3. Gajic I, Tomic N, Lukovic B, Jovicevic M, Kekic D, Petrovic M, et al. A Comprehensive Overview of Antibacterial Agents for Combating Multidrug-Resistant Bacteria: The Current Landscape, Development, Future Opportunities, and Challenges. *Antibiotics*. 2025;14:221.
4. Hou K, Wu Z-X, Chen X-Y, Wang J-Q, Zhang D, Xiao C, et al. Microbiota in health and diseases. *Signal Transduct Target Ther*. 2022;7:135.
5. Maier L, Stein-Thoeringer C, Ley RE, Brötz-Oesterhelt H, Link H, Ziemert N, et al. Integrating research on bacterial pathogens and commensals to fight infections-an ecological perspective. *Lancet Microbe*. 2024;5:100843.
6. Evans T, Ali U, Anderton R, Raby E, Manning L, Litton E. Lower gut dysbiosis and mortality in acute critical illness: a systematic review and meta-analysis. *Intensive Care Med Exp*. 2023;11:6.
7. Freedberg DE, Zhou MJ, Cohen ME, Annavajhala MK, Khan S, Moscoso DI, et al. Pathogen colonization of the gastrointestinal microbiome at intensive care unit admission and risk for subsequent death or infection. *Intensive Care Med*. 2018;44:1203-11.
8. Taur Y, Jenq RR, Perales MA, Littmann ER, Morjaria S, Ling L, et al. The effects of intestinal tract bacterial diversity on mortality following allogeneic hematopoietic stem cell transplantation. *Blood*. 2014;124:1174-82.
9. Yeoh YK, Zuo T, Lui GC, Zhang F, Liu Q, Li AY, et al. Gut microbiota composition reflects disease severity and dysfunctional immune responses in patients with COVID-19. *Gut*. 2021;70:698-706.

10. Procházková N, Laursen MF, La Barbera G, Tsekitsidi E, Jørgensen MS, Rasmussen MA, et al. Gut physiology and environment explain variations in human gut microbiome composition and metabolism. *Nat Microbiol.* 2024;9:3210-25.
11. Fernández-Pato A, Sinha T, Gacesa R, Andreu-Sánchez S, Gois MFB, Gelderloos-Arends J, et al. Choice of DNA extraction method affects stool microbiome recovery and subsequent phenotypic association analyses. *Sci Rep.* 2024;14:3911.
12. Weber D, Oefner PJ, Hiergeist A, Koestler J, Gessner A, Weber M, et al. Low urinary indoxyl sulfate levels early after transplantation reflect a disrupted microbiome and are associated with poor outcome. *Blood.* 2015;126:1723-8.
13. Kuo SZ, Dettmer K, Annavajhala MK, Chong DH, Uhlemann AC, Abrams JA, et al. Associations between urinary 3-indoxyl sulfate, a gut microbiome-derived biomarker, and patient outcomes after intensive care unit admission. *J Crit Care.* 2021;63:15-21.
14. Bassis CM, Moore NM, Lolans K, Seekatz AM, Weinstein RA, Young VB, et al. Comparison of stool versus rectal swab samples and storage conditions on bacterial community profiles. *BMC Microbiol.* 2017;17:78.
15. Shannon CE. A mathematical theory of communication. *Bell Syst Tech J.* 1948;27:379-423.
16. Rademacher J, Ewig S, Grabein B, Nachtigall I, Pletz M, Abele-Horn M, et al. Update der S3-Leitlinie: Epidemiologie, Diagnostik und Therapie erwachsener Patienten mit nosokomialer Pneumonie [Update of the S3 guideline: Epidemiology, diagnosis and treatment of adult patients with nosocomial pneumonia]. *Pneumologie.* 2024 Jan;78(3):145-207. German AWMF-Register Nr. 020/013. Available from: https://register.awmf.org/assets/guidelines/020-013l_S3_Epidemiologie-Diagnostik-Therapie-erwachsener-Patienten-nosokomiale-Pneumonie_2024-03.pdf.
17. CDC/NHSN. Pneumonia (Ventilator-associated [VAP] and non-ventilator-associated Pneumonia [PNEU]) Event, Atlanta, Georgia, USA. 2025. <https://www.cdc.gov/nhsn/pdfs/pscmanual/6pscvapcurrent.pdf>. Accessed 24 May 2025.
18. CDC/NHSN. Bloodstream Infection Event (Central Line-Associated Bloodstream Infection and Non-central Line Associated Bloodstream Infection), Atlanta, Georgia, USA. 2025. https://www.cdc.gov/nhsn/pdfs/pscmanual/4psc_clabscurrent.pdf. Accessed 24 May 2025.
19. R Core Team (2024). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
20. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods.* 2016;13:581-3.
21. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol.* 2019;37:852-7.
22. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python. *J Mach Learn Res.* 2012;12.
23. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* 2013;41:590-6.
24. Simpson EH. Measurement of Diversity. *Nature.* 1949;163:688-.
25. Chao A, Lee SM. Estimating the Number of Classes via Sample Coverage. *J Am Stat Assoc.* 1992;87:210-7.

26. Chao A. Nonparametric Estimation of the Number of Classes in a Population. *Scand Statist.* 1984;11:265-70.
27. Faith DP. Conservation evaluation and phylogenetic diversity. *Biol Conserv.* 1992;61:1-10.
28. Lozupone C, Knight R. UniFrac: a new phylogenetic method for comparing microbial communities. *Appl Environ Microbiol.* 2005;71:8228-35.
29. Bray JR, Curtis JT. An Ordination of the Upland Forest Communities of Southern Wisconsin. *Ecol Monogr.* 1957;27:325-49.
30. Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS, et al. Metagenomic biomarker discovery and explanation. *Genome Biol.* 2011;12:R60.
31. Chicco D, Tötsch N, Jurman G. The Matthews correlation coefficient (MCC) is more reliable than balanced accuracy, bookmaker informedness, and markedness in two-class confusion matrix evaluation. *Biodata Min.* 2021;14:13.
32. Bewick V, Cheek L, Ball J. Statistics review 13: Receiver operating characteristic curves. *Crit Care.* 2004;8:508.
33. van Vught LA, Klein Klouwenberg PM, Spitoni C, Scicluna BP, Wiewel MA, Horn J, et al. Incidence, Risk Factors, and Attributable Mortality of Secondary Infections in the Intensive Care Unit After Admission for Sepsis. *JAMA.* 2016;315:1469-79.
34. Zhao GJ, Li D, Zhao Q, Song JX, Chen XR, Hong GL, et al. Incidence, risk factors and impact on outcomes of secondary infection in patients with septic shock: an 8-year retrospective study. *Sci Rep.* 2016;6:38361.
35. Vincent JL, Moreno R, Takala J, Willatts S, De Mendonça A, Bruining H, et al. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. *Intensive Care Med.* 1996;22:707-10.
36. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. *Crit Care Med.* 1985;13:818-29.
37. Wang Q, Berger NA, Xu R. Analyses of Risk, Racial Disparity, and Outcomes Among US Patients With Cancer and COVID-19 Infection. *JAMA Oncol.* 2021;7:220-7.
38. Cantón-Bulnes ML, Jiménez-Sánchez M, Alcántara-Carmona S, Gimeno-Costa R, Berezo-García JÁ, Beato C, et al. Determinants of mortality in cancer patients with unscheduled admission to the Intensive Care Unit: A prospective multicenter study. *Medicina Intensiva.* 2022;46:669-79.
39. Bosch-Compte R, Visa L, Rios A, Duran X, Fernández-Real M, Gomariz-Vilaldach G, et al. Prognostic factors in oncological patients with solid tumours requiring intensive care unit admission. *Oncol Lett.* 2023;26:525.
40. Taur Y, Xavier JB, Lipuma L, Ubeda C, Goldberg J, Gobourne A, et al. Intestinal domination and the risk of bacteremia in patients undergoing allogeneic hematopoietic stem cell transplantation. *Clin Infect Dis.* 2012;55:905-14.
41. Wheatley RM, Caballero JD, van der Schalk TE, De Winter FHR, Shaw LP, Kapel N, et al. Gut to lung translocation and antibiotic mediated selection shape the dynamics of *Pseudomonas aeruginosa* in an ICU patient. *Nat Commun.* 2022;13:6523.
42. Bouatra S, Aziat F, Mandal R, Guo AC, Wilson MR, Knox C, et al. The human urine metabolome. *PLoS One.* 2013;8:e73076.

43. Remer T, Montenegro-Bethancourt G, Shi L. Long-term urine biobanking: Storage stability of clinical chemical parameters under moderate freezing conditions without use of preservatives. *Clin Biochem.* 2014;47:307-11.
44. Li M. Urine Reflection of Changes in Blood. In: Gao Y, editor. *Urine Proteomics in Kidney Disease Biomarker Discovery*. Dordrecht: Springer Netherlands; 2015. p. 13-9.
45. Miller IJ, Peters SR, Overmyer KA, Paulson BR, Westphall MS, Coon JJ. Real-time health monitoring through urine metabolomics. *npj Digit Med.* 2019;2:109.
46. Cai Z, Li W, Brenner M, Bahiraii S, Heiss EH, Weckwerth W. Branched-chain ketoacids derived from cancer cells modulate macrophage polarization and metabolic reprogramming. *Front Immunol.* 2022;13:966158.
47. Lo EKK, Felicianna, Xu JH, Zhan Q, Zeng Z, El-Nezami H. The Emerging Role of Branched-Chain Amino Acids in Liver Diseases. *Biomedicines.* 2022;10.
48. Brial F, Le Lay A, Dumas M-E, Gauguier D. Implication of gut microbiota metabolites in cardiovascular and metabolic diseases. *Cell Mol Life Sci.* 2018;75:3977-90.
49. Ren M, Cai S, Zhou T, Zhang S, Li S, Jin E, et al. Isoleucine attenuates infection induced by *E. coli* challenge through the modulation of intestinal endogenous antimicrobial peptide expression and the inhibition of the increase in plasma endotoxin and IL-6 in weaned pigs. *Food Funct.* 2019;10:3535-42.
50. Zhao R, Jiang C, Yuan Y, Zhang S, Ghonaim AH, Che C, et al. Isoleucine Enhanced the Function of the Small Intestinal Mucosal Barrier in Weaned Piglets to Alleviate Rotavirus Infection. *Animals (Basel).* 2024;14.
51. Marrs C, Lonsdale D. Hiding in Plain Sight: Modern Thiamine Deficiency. *Cells.* 2021;10:2595.
52. Mann G, Mora S, Madu G, Adegoke OAJ. Branched-chain Amino Acids: Catabolism in Skeletal Muscle and Implications for Muscle and Whole-body Metabolism. *Front Physiol.* 2021;12:702826.
53. Saade MC, Clark AJ, Parikh SM. States of quinolinic acid excess in urine: A systematic review of human studies. *Front Nutr.* 2022;9:1070435.
54. Roager HM, Licht TR. Microbial tryptophan catabolites in health and disease. *Nat Commun.* 2018;9:3294.
55. Wang Z, Gao Y, Zhang C, Hu H, Guo D, Xu Y, et al. Quinolate Phosphoribosyltransferase is an Antiviral Host Factor Against Hepatitis C Virus Infection. *Sci Rep.* 2017;7:5876.
56. Michaelis S, Zelzer S, Schneider C, Schnedl WJ, Baranyi A, Meinitzer A, et al. The possible role of quinolinic acid as a predictive marker in patients with SARS-CoV-2. *Clin Chim Acta.* 2023;550:117583.
57. Guillemin GJ. Quinolinic acid, the inescapable neurotoxin. *Febs j.* 2012;279:1356-65.
58. Dhakal B, Savarapu LSK, Sayed K. Integrating multi-compartment microbiome data with clinical parameters enhances mortality prediction using autoencoder. *J Microbiol Methods.* 2025;238:107267.
59. Lugo-Martinez J, Xu S, Levesque J, Gallagher D, Parker LA, Neu J, et al. Integrating longitudinal clinical and microbiome data to predict growth faltering in preterm infants. *J Biomed Inform.* 2022;128:104031.
60. Antikainen AA, Mutter S, Harjutsalo V, Thorn LM, Groop PH, Sandholm N. Urinary metabolomics provide insights into coronary artery disease in individuals with type 1 diabetes. *Cardiovasc Diabetol.* 2024;23:425.
61. Kim Y, Lee J, Kang MS, Song J, Kim SG, Cho S, et al. Urinary Metabolite Profile Predicting the Progression of CKD. *Kidney360.* 2023;4:1048-57.
62. Jeon YH, Lee S, Kim DW, Kim S, Bae SS, Han M, et al. Serum and urine metabolomic biomarkers for predicting prognosis in patients with immunoglobulin A nephropathy. *Kidney Res Clin Pract.* 2023;42:591-605.

63. Bajaj JS, Garcia-Tsao G, Reddy KR, O'Leary JG, Vargas HE, Lai JC, et al. Admission Urinary and Serum Metabolites Predict Renal Outcomes in Hospitalized Patients With Cirrhosis. *Hepatology*. 2021;74:2699-713.
64. Gisewhite S, Stewart IJ, Beilman G, Luszczek E. Urinary metabolites predict mortality or need for renal replacement therapy after combat injury. *Crit Care*. 2021;25:119.
65. Dewulf JP, Martin M, Marie S, Oguz F, Belkhir L, De Greef J, et al. Urine metabolomics links dysregulation of the tryptophan-kynurenine pathway to inflammation and severity of COVID-19. *Sci Rep*. 2022;12:9959.
66. Zeden JP, Fusch G, Holtfreter B, Schefold JC, Reinke P, Domanska G, et al. Excessive tryptophan catabolism along the kynurenine pathway precedes ongoing sepsis in critically ill patients. *Anaesth Intensive Care*. 2010;38:307-16.
67. Jongers B, Hotterbeekx A, Bielen K, Vervliet P, Boddaert J, Lammens C, et al. Identification of Potential Urinary Metabolite Biomarkers of Pseudomonas aeruginosa Ventilator-Associated Pneumonia. *Biomark Insights*. 2022;17:11772719221099131.
68. Gregson DB, Wildman SD, Chan CCY, Bihan DG, Groves RA, Aburashed R, et al. Metabolomics strategy for diagnosing urinary tract infections. *medRxiv*. 2024:2021.04.07.21255028.
69. Chan CCY, Gregson DB, Wildman SD, Bihan DG, Groves RA, Aburashed R, et al. Metabolomics strategy for diagnosing urinary tract infections. *Nat Commun*. 2025;16:2658.