© 2026 Journal of Pharmacy & Pharmacognosy Research, 14 (2), 2404, 2026
ISSN 0719-4250
https://jppres.com

DOI: https://doi.org/10.56499/jppres_14.2.2404

Original Article

Differentially expressed genes analysis for diagnostics and therapy of diarrhea due to pathogenicity and antibiotic resistance genes in children

[Análisis de genes expresados diferencialmente para el diagnóstico y tratamiento de la diarrea debida a genes de patogenicidad y resistencia a antibióticos en niños]

Linda Erlina[1,2,3], Fadilah Fadilah[2,3], Asmarinah[4], Badriul Hegar[5], Aryo Tedjo[2*]**

[1]Doctoral Program in Biomedical Sciences, Faculty of Medicine, Universitas Indonesia, Jakarta, 10430, Indonesia.
[2]Department of Medical Chemistry, Faculty of Medicine, Universitas Indonesia, Jakarta, 10430, Indonesia.
[3]Bioinformatics Core Facilities, Indonesian Medical Education and Research Institute (IMERI), Faculty of Medicine, Universitas Indonesia, Jakarta, 10430, Indonesia.
[4]Department of Medical Biology, Faculty of Medicine, Universitas Indonesia, Jakarta, 10430, Indonesia.
[5]Department of Child Health, Faculty of Medicine, Universitas Indonesia, Jakarta, 10430, Indonesia.
E-mail: *aryo.tedjo@ui.ac.id; **fadilah.msi@ui.ac.id

Abstract

Context: Pediatric diarrhea is often associated with dysbiosis and pathogenic microorganisms, necessitating further molecular investigation. Differential gene expression (DEG) analysis offers potential insights into biomarkers for the molecular pathogenic mechanisms of diarrhea and antibiotic resistance in children.

Aims: To identify differentially expressed genes (DEGs) associated with pathogenicity and antimicrobial resistance by analyzing publicly available transcriptomic two datasets (GSE69529 and GSE54070) in children with diarrhea.

Methods: DEG analysis was performed using GEO2R on GSE69529 and GSE54070 datasets. Multiple machine learning models (SVM, RF, NN, LR, XGBoost) were applied via Orange 3.10.2 for cross-validation. Cytoscape and databases such as STRINGdb, IntAct, and CTD were utilized for protein interaction and upstream regulatory analysis. To confirm ARGs, five fecal cohort samples were collected from the primary public health in Jakarta, Indonesia, and analysed using the Vitek-2 compact system.

Results: Key DEGs included *RYK*, *CREBL2*, and *ST14*, associated with intestinal integrity, inflammation, and immune regulation. Seven resistance genes were commonly identified in different age groups. Pathogen mapping revealed *Escherichia coli*, *Pseudomonas aeruginosa*, *Salmonella enterica*, *Staphylococcus aureus*, and *Salmonella typhimurium* as dominant. Resistance-related genes such as *mexA/acrA*, *arnA*, *tolC*, and *mexB/acrB* were implicated in β-lactam and CAMP resistance pathways. Resistance genes common to different age groups included *tetO*, *tetQ*, and *SUL2*. To validate those findings, Support Vector Machines and Neural Networks exhibited the best classification performance for machine learning models.

Conclusions: Several genes identified among DEGs related to pathogenicity and antimicrobial resistance markers may serve as potential biomarkers for developing diagnostic and therapeutic targets for pediatric diarrhea.

Keywords: antimicrobial resistance gene; childhood diarrhea; differentially expressed gene; pathogenic bacteria; upstream regulatory protein.

Resumen

Contexto: La diarrea pediátrica se asocia frecuentemente con disbiosis y microorganismos patógenos, lo que requiere mayor investigación molecular. El análisis de expresión génica diferencial (GED) ofrece información potencial sobre biomarcadores de los mecanismos patogénicos moleculares de la diarrea y la resistencia a los antibióticos en niños.

Objetivos: Identificar genes de expresión génica diferencial (GED) asociados con la patogenicidad y la resistencia a los antimicrobianos mediante el análisis de dos conjuntos de datos transcriptómicos disponibles públicamente (GSE69529 y GSE54070) en niños con diarrea.

Métodos: El análisis de GED se realizó utilizando GEO2R en los conjuntos de datos GSE69529 y GSE54070. Se aplicaron múltiples modelos de aprendizaje automático (SVM, RF, NN, LR, XGBoost) a través de Orange 3.10.2 para la validación cruzada. Se utilizaron Cytoscape y bases de datos como STRINGdb, IntAct y CTD para la interacción de proteínas y el análisis regulatorio ascendente. Para confirmar los ARG, se recogieron cinco muestras de cohorte fecal del centro de salud pública primaria de Yakarta (Indonesia) y se analizaron con el sistema compacto Vitek-2.

Resultados: Los DEG clave incluyeron *RYK*, *CREBL2* y *ST14*, asociados con la integridad intestinal, la inflamación y la regulación inmunitaria. Se identificaron siete genes de resistencia comunes en diferentes grupos de edad. El mapeo de patógenos reveló que *Escherichia coli*, *Pseudomonas aeruginosa*, *Salmonella enterica*, *Staphylococcus aureus* y *Salmonella typhimurium* eran dominantes. Los genes relacionados con la resistencia, como *mexA/acrA*, *arnA*, *tolC* y *mexB/acrB*, estuvieron implicados en las vías de resistencia a β-lactámicos y AMPc. Los genes de resistencia comunes a diferentes grupos de edad incluyeron *tetO*, *tetQ* y *SUL2*. Para validar estos hallazgos, las máquinas de vectores de soporte y las redes neuronales mostraron el mejor rendimiento de clasificación para los modelos de aprendizaje automático.

Conclusiones: Varios genes identificados entre los DEG relacionados con la patogenicidad y los marcadores de resistencia a los antimicrobianos podrían servir como posibles biomarcadores para el desarrollo de dianas diagnósticas y terapéuticas para la diarrea pediátrica.

Palabras Clave: bacterias patógenas; diarrea infantil; gen de expresión diferencial; gen de resistencia a los antimicrobianos; proteína reguladora ascendente.

ARTICLE INFO
Received: February 6, 2025.
Accepted: February 17, 2026.
Available Online: March 21, 2026.

AUTHOR INFO
ORCID: 0000-0003-1108-0600 (LE) 0000-0002-5924-1664 (BH)
0000-0002-8120-3138 (FF) 0000-0001-8998-3418 (AT)
0000-0001-5204-4054 (A)

Abbreviations: ARG: Antibiotic resistance genes; AMP: Ampicillin; AUC: Area under the curve; CA: Classification accuracy; CI: Confidence interval; CIP: Ciprofloxacin; CRO: Ceftriaxone; CARD: Comprehensive antibiotic resistance database; CTD: Comparative toxicogenomics database; DAEC: Diffusely adherent *E. coli*; DEG: Differentially expressed gene; EAEC: Enteroaggregative *E. coli*; EPEC: Enteropathogenic *E. coli*; ERY: Erythromycin; GEO: Gene expression omnibus database; GO: Gene ontology; KEGG: Kyoto encyclopedia of genes and genomes; Log2FC: Log 2 fold change; MCC: Matthew's correlation coefficient; NA: Nalidixic acid; PCA: Principal component analysis; PPI: Protein-protein interaction; ROC: Receiver operating characteristic; SXT: Sulfamethoxazole/trimethoprim; SVM: Support vector machine; TGFB1: Transforming growth factor beta 1; URP: Upstream regulatory proteins.

INTRODUCTION

Diarrhea disease remains a leading cause of morbidity in children, frequently driven by enteric infections and imbalances in gut microbial communities. Among pediatric populations, the presence of multiple pathogens can lead to complex interactions that complicate both diagnosis and clinical outcomes. Commonly implicated organisms include *Escherichia coli*, *Salmonella spp.*, *Shigella spp.*, *Campylobacter*, and *Rotavirus*, each with several pathotypes, such as enteroaggregative *E. coli* (EAEC), enteropathogenic *E. coli* (EPEC), and diffusely adherent *E. coli* (DAEC), that exhibit diverse pathogenic mechanisms. Despite their clinical significance, these pathogens are often underreported due to limitations in standard diagnostic protocols (Malik et al., 2019; Pelkonen et al., 2018; Tobias et al., 2015; Varela et al., 2015). Recent advances have revealed that microbial dysbiosis and antibiotic resistance significantly influence disease outcomes.

The gut microbiome not only harbours potential pathogens but also serves as a reservoir for antibiotic resistance genes (ARGs), which are transmitted through vertical or horizontal gene transfer. These ARGs have been identified even in healthy infants with no prior antibiotic exposure, suggesting early-life colonization and dissemination. Mobile genetic elements, including plasmids and bacteriophages, facilitate the spread of resistance determinants across microbial populations (Davies & Davies, 2010; Martínez & Baquero, 2014). Reports indicate the presence of antibiotic-resistant bacteria and genes in the intestines of healthy individuals, including infants who have not been exposed to antibiotics (Carvalho et al., 2022; Chong et al., 2020; Feng et al., 2018; Gasparrini et al., 2019).

In line with the findings of Natarajan et al. (2018), the detection of enteropathogenic *Escherichia coli* (EPEC) in stool samples from children at a tertiary hospital in Puducherry, India, showed a significant association with virulence and antimicrobial resistance (AMR) genes. Several AMR genes identified include *sul2*, *tetA*, *blaCTX*, *blaTEM*, *tetD*, *qnrs*, *dhfr1*, *aac(3)-IV*, *tetB*, *blaSHV*, and *tetD*. Okumu et al. (2025) summarized 64 articles published between January 2005 and September 2024, some clinically antibiotics such as sulfamethoxazole/trimethoprim (SXT) 71% [95%CI: 57–82%]; ampicillin (AMP) 56% [95%CI: 44–67%]; ciprofloxacin (CIP) 10% [95%CI: 5–20%]; and ceftriaxone (CRO) 8% [95%CI: 2–31%] were resistant to *E.coli*. For *Shigella* spp. AMR, some antibiotics were AMP 76% [95%CI: 60–87%]; nalidixic acid (NA) 9% [95%CI: 2–31%]; CIP 3% [95%CI: 0–15%]; and CRO 2% [95%CI: 0–19%]. The finding also stated for *Salmonella* spp. AMR were AMP 55% [95%CI: 35–73%] and SXT 25% [95%CI: 15–38%], the proportions of *Campylobacter* spp. AMR were erythromycin (ERY) 33% [95%CI: 12–64%] and CIP 27% [95%CI: 8–61%].

The genetic mechanisms underlying pathogenesis and resistance are increasingly studied through transcriptomic analyses. Differential gene expression (DEG) analysis enables the identification of potential biomarkers and therapeutic targets by elucidating host–pathogen interactions and antibiotic resistance pathways. Public repositories such as the gene expression omnibus (GEO) provide access to transcriptomic datasets that can be reanalyzed using bioinformatics pipelines. Machine learning has been employed to validate DEGs and enhance predictive accuracy. In parallel, protein–protein interaction networks allow for deeper exploration of functional relevance (Huang et al., 2025).

When integrated with bioinformatics and machine learning, DEG data can elucidate both pathogen-specific and shared molecular pathways, aiding in the development of diagnostic and therapeutic interventions (DeBerg et al., 2018; Lu et al., 2014). This study hypothesizes that distinct DEGs exist between diarrhea and healthy children across two public datasets (GSE69529 and GSE54070) and can serve as biomarkers for diagnostics and therapy. The aim was to highlight candidate genes with diagnostic or therapeutic potential across age groups and infectious agents.

MATERIAL AND METHODS

Differential Expressed Genes (DEG) analysis

Transcriptomic datasets GSE69529 and GSE54070 were obtained from the Gene Expression Omnibus (GEO) database. Differential expression analysis was performed using GEO2R, with sample groups defined by infection type or age. The analysis employed the Limma algorithm, and p-values were adjusted using

the Benjamini-Hochberg method, with a threshold of <0.05 and |Log2FC| > 1. A log2 transformation was used to normalize gene expression data.

In GSE69529, RNA sequencing data from whole blood samples were categorized into: (1) healthy children with bacterial colonization (12 samples), (2) healthy children without bacterial colonization (12 samples), and (3) children with diarrhea caused by DAEC (23 samples), EAEC (18 samples), EPEC (13 samples), *Rotavirus* (55 samples), *Salmonella* (36 samples), and *Shigella* (37 samples). For GSE54070, which contains microarray data on fecal samples, the groups included preschool (2–5 years) (23 samples: male nine and female 14) and school-aged children (10–11 years) (23 samples: male 11 and female 12) (DeBerg et al., 2018; Lu et al., 2014).

Machine learning optimization and validation

Significant DEGs (GSE69529) were further validated using supervised machine learning algorithms, including neural networks, support vector machines (SVMs), and naïve bayes. For GSE54070, several machine learning models were used, including SVM, random forest, neural network, XGBoost, and logistic regression. Analyses were performed in Orange v3.37.0. Performance metrics included Area under the receiver operating characteristic (ROC) curve (AUC), classification accuracy (CA), F1-score, precision, recall, and Matthew's correlation coefficient (MCC). Models were optimized with standard parameters and evaluated using cross-validation (Bisong, 2019; Chen & Guestrin, 2016; Pisner & Schnyer, 2020; Qi, 2012; Schmidhuber, 2015).

Protein-protein interaction to find the biomarker genes related to the pathogenicity

For pathogen-related DEG analysis, protein-protein interaction (PPI) networks were constructed in Cytoscape v3.10.2 using the STRING and IntAct databases. Shared DEGs genes consistently regulated across multiple pathogen types were mapped, and upstream regulatory proteins (URPs) were identified via the Comparative Toxicogenomics Database (CTD) (Davis et al., 2023; Shannon et al., 2003).

Molecular interaction to find the hub genes related to antimicrobial resistance

In GSE54070, upregulated and downregulated genes were filtered using a log2 fold-change threshold of |log2FC| ≥ 1. Overlapping resistance genes were identified using Venn diagram tools. Molecular interaction networks were generated in Cytoscape, linked to the IntAct database, and limited to *Escherichia coli* and *Pseudomonas aeruginosa* (del Toro et al., 2022). Functional enrichment analyses, including KEGG pathways and Gene Ontology (GO) terms, were conducted using ShinyGO, with p-value and P-Adj < 0.05 (Chen et al., 2015; Ge et al., 2020).

Confirmation of ARGs with antibiotic resistance class in cohort fecal samples in children with acute diarrhea

Five cohort fecal samples were collected from children under five years old with acute diarrhea at a primary healthcare facility in Jakarta, Indonesia. Ethical clearance has been obtained from the Health Research Ethics Committee, Faculty of Medicine, University of Indonesia (FKUI) - RSCM, with the number KET-1235/UN2.F1/ETIK/PPM.00.02/2023. All the samples were cultured and incubated for 24 hours in MacConkey agar, a selective medium for *E. coli*. ARGs and antibiotic resistance class were detected using the VITEK® 2 software version 10.

Statistical analysis and data visualization

In the DEG analysis, lmfit was used to fit linear models to the count data from GSE69529 and GSE54070. To set up contrasts of interest and recalibrate model coefficients, the makeContrasts and contrasts.fit packages, and empirical bayes moderation (ebayes) were also used to compute statistics and a table of the top significant genes. For visualization and quality control test results, topTable with p-adjust:fdr based on p-value, then summarized to test results as up, down, or not expressed using the decideTests function. To filter out the bad probes and create a Q-Q plot for the t-statistic, the qqt package was used. For the DEG visualization, the volcano plot was generated using the volcanoplot package (log p-value vs log fold change). All significantly upregulated and downregulated genes were listed based on Bonferroni and Hochberg (FDR) <0.05.

ANOVA was used in machine learning optimization and validation to identify the five most significant genes based on DEG profiles in DAEC, EAEC, *Salmonella*, *Shigella*, and *Rotavirus* vs. healthy samples. Principal Component Analysis (PCA) was performed with the following parameters: components PC1 and PC2, an explained variance of 84%, and normalized variables and a significance level (p<0.05). Various machine learning models were compared using the area under the ROC curve and a confusion matrix to assess actual and predicted values. Regulation of biomarker genes was tested using the Forest Effect Size package based on Log2FC (mean and standard deviation).

All workflow analysis details were provided in https://github.com/lindaerlina/DEG-Patterns-in-Pediatric-Diarrheal-Infections-and-Antimicrobial-Resistance-Genes.

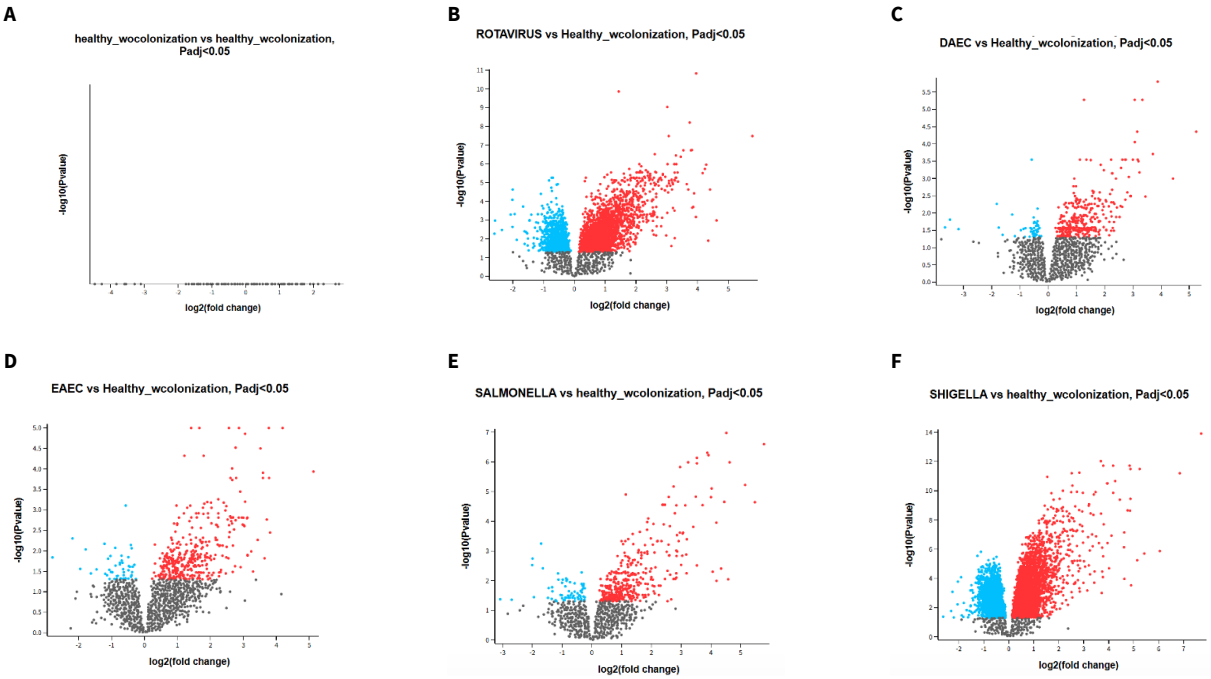


Figure 1. Volcano plot of DEG analysis results for GSE69529 comparing **(A)** the group of healthy children with blood containing bacterial colonies (colonization) to **(B)** the group of children with diarrhea and blood infected with *Rotavirus*, **(C)** DAEC, **(D)** EAEC, **(E)** *Salmonella*, and **(F)** *Shigella*.

Table 1. Five DEGs based on the highest ANOVA score (GSE69259) and cross-validation results.

Group	DEG_*Rotavirus*	DEG_*Shigella*	DEG_*Salmonella*	DEG_DAEC	DEG_EAEC
healthy_ wcolonization	*IFI27, CREBL2, RYK, SIGLEC5, ST14*	*NLRC4, LTBR4, RYK, C19orf38, BST1*	*RYK, CREBL2, ST14, C19orf38, OSCAR*	*SIGLEC5, DDAH2, LTB4R, ST14, SYN2*	*RYK, CREBL2, DDAH2, SYN2, S100A9*
	*Sensitivity: 0.945 *Specificity: 0.916	*Sensitivity: 0.861 *Specificity: 0.833	*Sensitivity: 0.833 *Specificity: 0.667	*Sensitivity: 0.909 *Specificity: 1.000	*Sensitivity: 0.833 *Specificity: 0.750

Table 2. Regulation of DEGs *RYK*, *ST14*, and *CREBL2* [Log2(fold change, FC)] on the GSE69259.

Symbol	log2(FC)_ *Rotavirus*	log2(FC)_ *Salmonella*	log2(FC)_ *Shigella*	log2(FC)_ DAEC	log2(FC)_ EAEC
RYK	-0.339	-0.316	-0.391	-0.349	-0.412
ST14	1.441	1.141	1.543	1.271	1.211
CREBL2	-0.501	-0.335	-0.397	-0.409	-0.489

RESULTS

DEG patterns in pediatric diarrheal infections

The analysis of the GSE69529 dataset identified differentially expressed genes (DEGs) that distinguished children with diarrhea caused by DAEC, EAEC, *Rotavirus*, *Salmonella*, or *Shigella* from healthy controls with bacterial colonization. No significant DEGs were observed between healthy children with and without colonization. Volcano plots demonstrated clear separation based on adjusted ($p<0.05$) (Fig. 1).

Top-ranking DEGs based on ANOVA scores, such as *RYK*, *CREBL2*, and *ST14*, showed consistent regulation across multiple infection types. *RYK* and *CREBL2* were downregulated, while *ST14* was upregulated in most pathogen-associated groups. These genes formed central nodes in a PPI network and were identified as potential therapeutic targets due to their cross-pathogen significance (Tables 1-2 and Fig. 2).

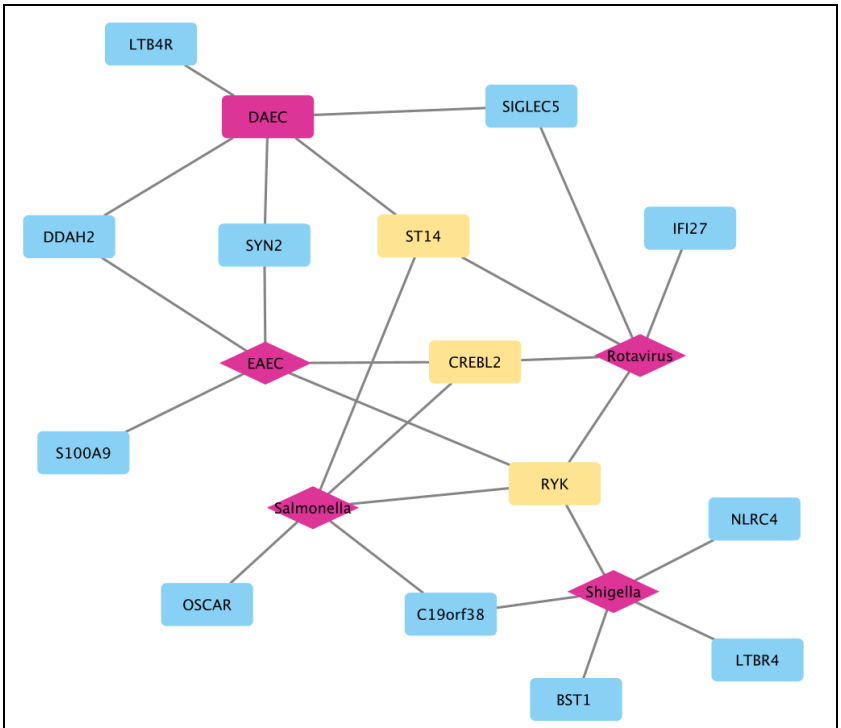


Figure 2. DEG network with the top five ANOVA scores on the GSE69259 dataset.

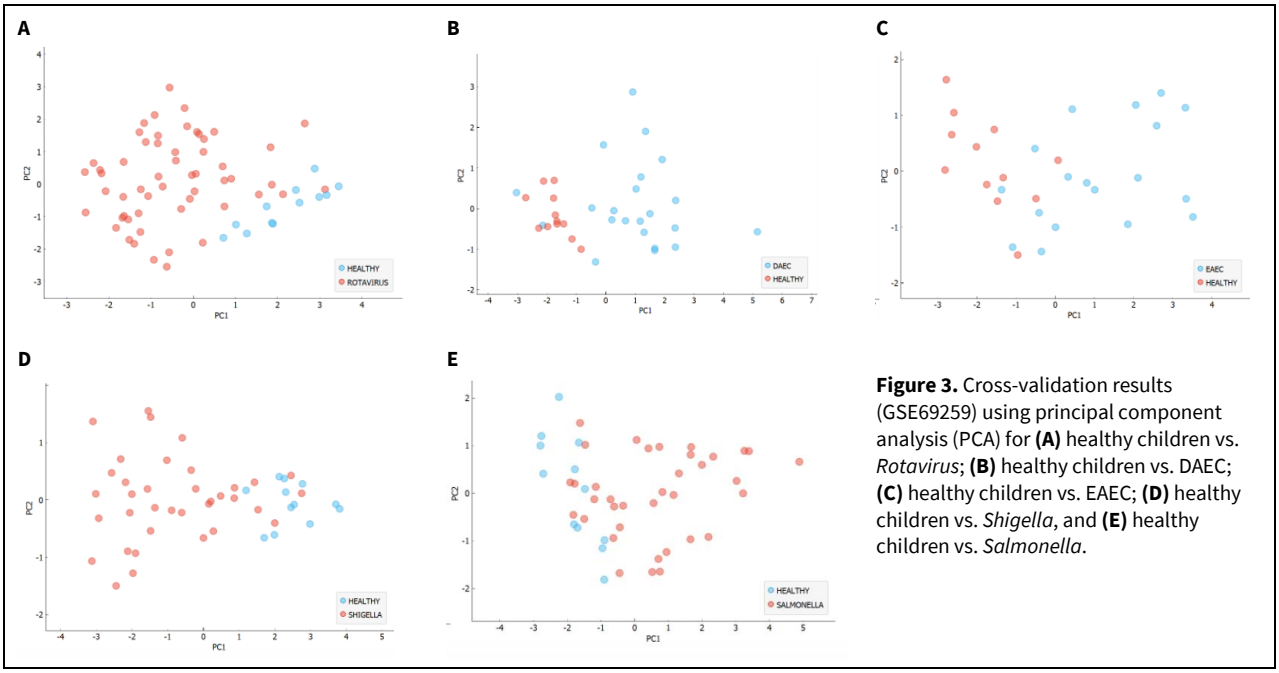


Figure 3. Cross-validation results (GSE69259) using principal component analysis (PCA) for **(A)** healthy children vs. *Rotavirus*; **(B)** healthy children vs. DAEC; **(C)** healthy children vs. EAEC; **(D)** healthy children vs. *Shigella*, and **(E)** healthy children vs. *Salmonella*.

Validation through machine learning

Cross-validation using SVM confirmed high sensitivity and specificity for the selected DEGs, particularly in distinguishing rotavirus infection (sensitivity 0.945; specificity 0.916) (Table 1). The top five genes were identified from ANOVA scores (Table S4). PCA further validated the discriminatory power of these gene signatures, using normalized variables to obtain cumulative and component variances (Fig. 3 and Table S5).

DGE network, regulation of DEGs *RYK*, *ST14*, and *CREBL2*, and forest effect size plot

As seen in the DEG network in Fig. 2, a total of 14 nodes (light blue and yellow) were found as the top 5 based on ANOVA scores in *Rotavirus*, *Shigella*,

Salmonella, DAEC, and EAEC (pink), and three nodes (light yellow), such as *ST14* (*Rotavirus*, *Salmonella*, and DAEC), *CREBL2* (*Rotavirus*, *Salmonella*, and EAEC), and *RYK* (*Rotavirus*, *Shigella*, *Salmonella*, and EAEC), were found as intersections in several organisms.

From Table 2, *RYK* and *CREBL2* were found downregulated with a range of Log2FC from -0.412 to -0.316 for *RYK* and -0.501 to -0.335 for *CREBL2*, and *ST14* was found upregulated with a range of Log2FC from 1.141 to 1.543. DEG profiles (mean Log2FC) for *RYK*, *CREBL2*, and *ST14* were calculated and plotted using Forest Effect Size with 95% confidence intervals (CIs) (Fig. 4).

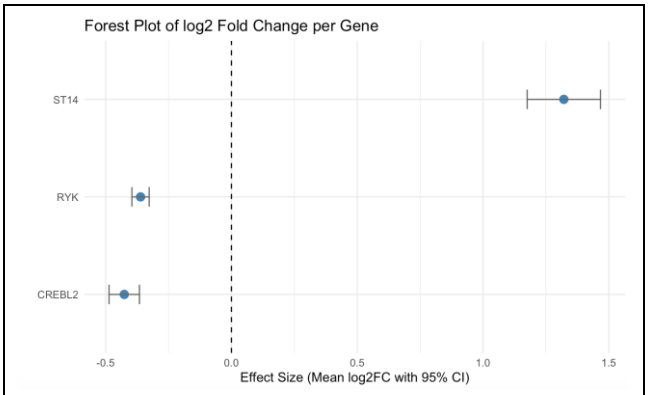


Figure 4. Forest effect size plot.

Identification of upstream regulatory proteins (URPs)

URP analysis via CTD revealed that regulatory proteins, including NFE2L2, TGFB1, NOG, and PPARA, modulate *CREBL2* and *ST14*. Activation or inhibition of these regulators could potentially normalize DEG expression associated with diarrhea infections (Table 3).

Resistance gene expression in age-stratified cohorts

The GSE54070 analysis revealed 5884 upregulated and 254 downregulated genes between preschool and school-aged children. Seven overlapping resistance genes (*tetO*, *tetW*, *tetQ*, *tet32*, *ermB*, *sul2*, *bl2e*) were identified. Among the top gut bacteria detected were *E. coli*, *P. aeruginosa*, *S. enterica*, and *S. aureus*. Molecular interaction analysis using Cytoscape identified 19 protein-coding DEGs involved in resistance pathways. Overexpressed proteins such as *acrA*, *acrB*, *macB*, and *tolC* were implicated in multidrug efflux mechanisms.

The preschool and school groups were compared for differences in gene expression, and the DEG results were visualized as a volcano plot using the GEO2R program (Fig. 5). A total of 5884 genes were upregulated with a Log2FC value of 1.00 - 2.61, and 254 genes were downregulated with a Log2FC value of -3.074 to -1.039 (Fig. 5A).

The upregulated and downregulated gene entities were reinspected because many were from different bacteria with the same gene name. The duplications were removed, and 11 downregulated and 267 upregulated genes were obtained. From the Venn diagram, seven genes overlapped between the two, including *tetO*, *ERMB*, *tetW*, *tet32*, *BL2E*, *tetQ*, and *SUL2* (Fig. 5B).

As shown in Fig. 6, more than 60 bacterial species were identified in the intestines of children in the preschool and school-age groups. The most abundant species include *Escherichia coli*, *Pseudomonas aeruginosa*, *Salmonella enterica*, *Staphylococcus aureus*, *Enterococcus faecalis*, *Acinetobacter baumannii*, *Salmonella typhimurium*, *Enterobacter cloacae*, *Citrobacter freundii*, and *Stenotrophomonas sp*.

Evaluation of all machine learning models using cross-validation with ten folds in a stratified manner. The logistic regression model selected Lasso (L1) regularization with C = 1. In the Random Forest model, the replicable training parameter was set to a maximum tree depth of 3, and it does not split subsets smaller than 5.

In the support vector machine (SVM) model, the SVM cost was set to 1.00, the regression loss epsilon to 0.10, the kernel to Linear, the numerical tolerance to 0.0010, and the iteration limit to 100.

Table 3. Relationship between DEG *RYK*, *ST14*, and *CREBL2* and URP and predicted activity targets.

Protein	Interaction	DEG (mRNA)	PubMed ID	Target of DEG regulation	Predicted target of activity protein
NFE2L2/NRF2	increased expression	*CREBL2*	30529165	Up	Activation
NOG	Increased expression	*ST14*	27188386	Down	Inhibition
IL6	Affects expression	*ST14*	30171875	Down	Unknown
TGFB1	Decreased expression	*ST14*	30165855	Down	Activation
PPARA	Increased expression	*ST14*	19710929	Down	Inhibition

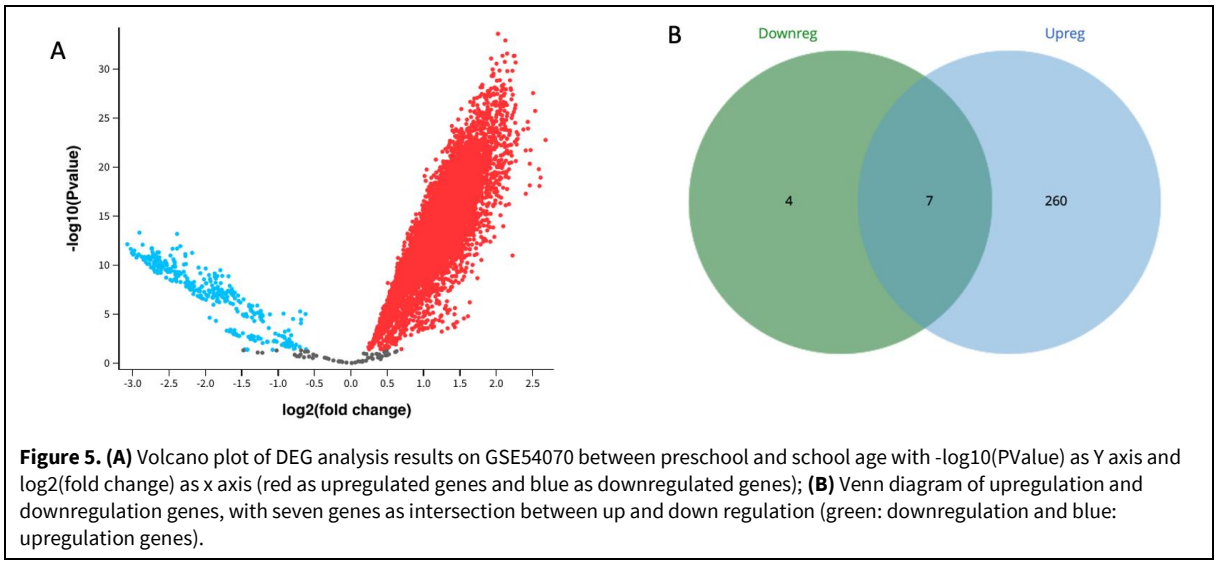


Figure 5. (A) Volcano plot of DEG analysis results on GSE54070 between preschool and school age with -log10(PValue) as Y axis and log2(fold change) as x axis (red as upregulated genes and blue as downregulated genes); **(B)** Venn diagram of upregulation and downregulation genes, with seven genes as intersection between up and down regulation (green: downregulation and blue: upregulation genes).

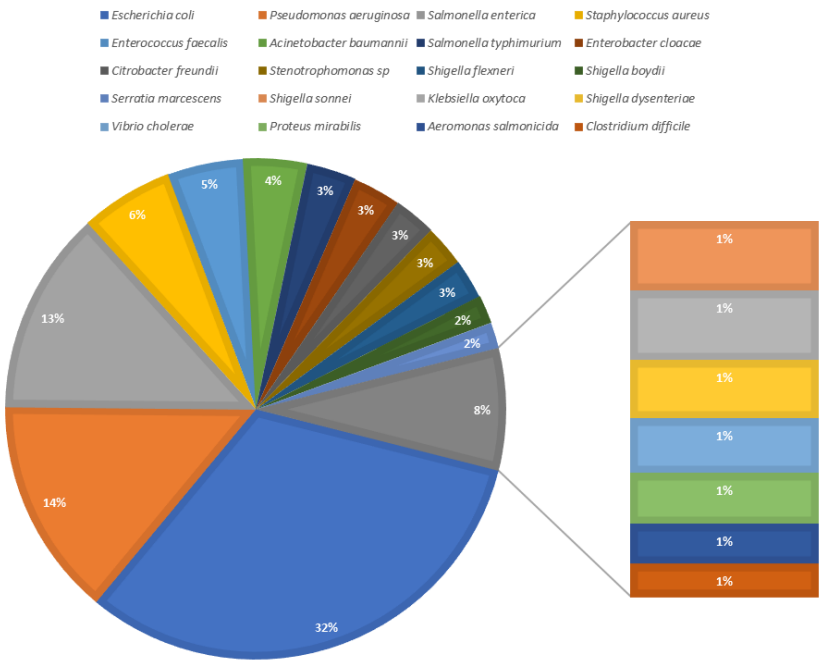


Figure 6. Top 20 bacteria species in the gut of children (preschool and school-age) includes *Escherichia coli* 32% (blue), *Pseudomonas aeruginosa* 13% (bright orange), *Salmonella enterica* 14% (light grey), *Staphylococcus aureus* 6% (light yellow), *Enterococcus faecalis* 5% (light blue), *Acinetobacter baumannii* 4% (light green), *Salmonella typhimurium* 3% (dark blue), *Enterobacter cloacae* 3% (dark brown), *Citrobacter freundii* 3% (dark grey), *Stenotrophomonas sp* 3% (brown), *Shigella flexneri* 3% (blue), *Shigella boydii* 2% (green), *Serratia marcescens* 2% (light blue) and for part grey 8% are divided into *Shigella sonnei, Klebsiella oxytoca, Shigella dysenteriae, Vibrio cholerae, Proteus mirabilis, Aeromonas salmonicida, Clostridium difficile* (all of species are 1%).

In the Extreme Gradient Boosting (XGBoost) model, the basic properties were set to a number of trees 100; learning rate 0.300 with replicable training, regularization lambda = 5; limit the depth of individual trees = 6; the fraction of training instances, features for each tree, features for each level, and features each split were set at 1.00. For validation results, all models have the same AUC score of 0.997; CA, F1, Precision, and Recall scores of 0.995; and MCC value of 0.943, except for the random forest model with a score of 0.941 (Table 4). The benchmarking study for selecting ML validation parameters was compared with a previous study on learning models to identify DEG biomarkers in pediatric patients with sepsis (Abbas & EL-Manzalawy, 2020). Similar parameters were used, such as Random Forest with RF100 and XGBoost with 100 weak learners, as the optimal settings for achieving higher AUC, ROC, precision, accuracy, F1, MCC, and recall.

Cytoscape version 3.10.2, integrated with the IntAct database, generated molecular interactions from 267 upregulated genes. As shown in Fig. 7, 19 protein-coding nodes formed 91 edges, and several proteins were identified in *IntAct* mutations, including *acrB*, *acrZ*, *macB*, *bacA*, *mexA*, and *mexB*. All 19 nodes (19 proteins) were submitted to ShinyGO 0.81 to analyze the pathway (KEGG) and Gene Ontology (GO) enrichment.

Table 4. Supervised machine learning model optimization and validation of GSE54070 DEG results.

Model	AUC	CA	F1	Precision	Recall	MCC
Logistic regression	0.997	0.995	0.995	0.995	0.995	0.943
Random forest	0.997	0.995	0.995	0.995	0.995	0.941
SVM	0.997	0.995	0.995	0.995	0.995	0.943
XGBoost	0.997	0.995	0.995	0.995	0.995	0.943
Neural network	0.997	0.995	0.995	0.995	0.995	0.943

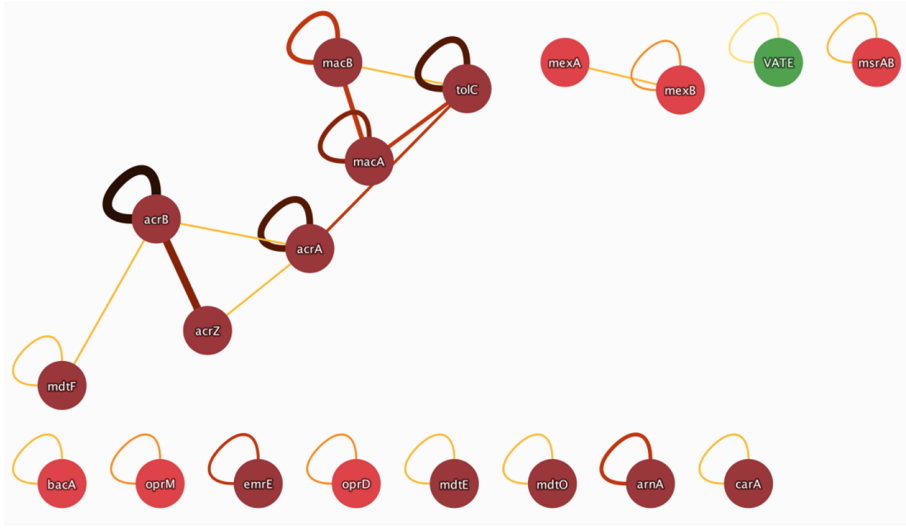


Figure 7. Molecular interactions of 19 nodes of protein-coding from 267 upregulation genes.

Functional enrichment of resistance mechanisms

ShinyGO-based enrichment analysis linked the upregulated resistance genes to key antimicrobial pathways: beta-lactam resistance, CAMP resistance, and teichoic acid biosynthesis. GO terms emphasized xenobiotic transport, efflux transmembrane activity, and membrane-bound protein complexes as central themes in resistance development.

In ShinyGO 0.81, species were set to *Escherichia coli* and three KEGG pathways were identified related to the antimicrobial resistance mechanism, such as beta-lactam resistance through *OMP*, *RND*, *MFP*, *mexA/acrA*, and *mexB/acrB*, which became biomarkers (enrichment FDR value 7.4E-05) (Fig. 8); Cationic antimicrobial peptide (CAMP) resistance with *tolC*, *acrA*, *acrB*, and *arnA* as biomarkers (enrichment FDR value 2.7E-05) (Fig. 9), and teichoic acid biosynthesis with *BacA* as a biomarker (enrichment FDR value 3.6E-02) and teichoic acid biosynthesis with *BacA* as a biomarker (enrichment FDR value 3.6E-02) (Fig. 10).

GO was evaluated in three terms: biological process, molecular function, and cellular components. Some terms in the GO biological process were associated with response to antibiotics and drugs, drug transport, and export. Some terms were highlighted in efflux, xenobiotic, and bile acid transmembrane transporter in the GO molecular function. In the GO cellular component, the most concerning were the efflux pump complex, the macAB-*tolC* complex, transmembrane components, and membrane components (intrinsic and integral) (Table 5).

ARGs and antibiotic resistance class were detected in five cohort fecal samples in children with acute diarrhea

From fecal cultures on MacConkey agar, all samples were confirmed as *E. coli* by pinkish colonies and tested with the Vitek-2 Compact System to detect ARGs and the antibiotic resistance class.

As shown in Table 6, Samples 1 and 2 have aminoglycoside resistance with the same ARGs *aac(3)-IId* and *aadA2*, but different ARGs for beta-lactam resistance *blaCTX-M-27* and *blaCTX-M-15* (Sample 1) and *blaOXA-10* and *blaOXA-1* (Sample 2). Samples 1, 3, and 4 have the same beta-lactam ARGs *blaCTX-M-15*. Samples 1 and 5 have the same ARGs *aadA5*. Samples 2 and 5 have the same *ARGs aac(6')-Ib-cr*.

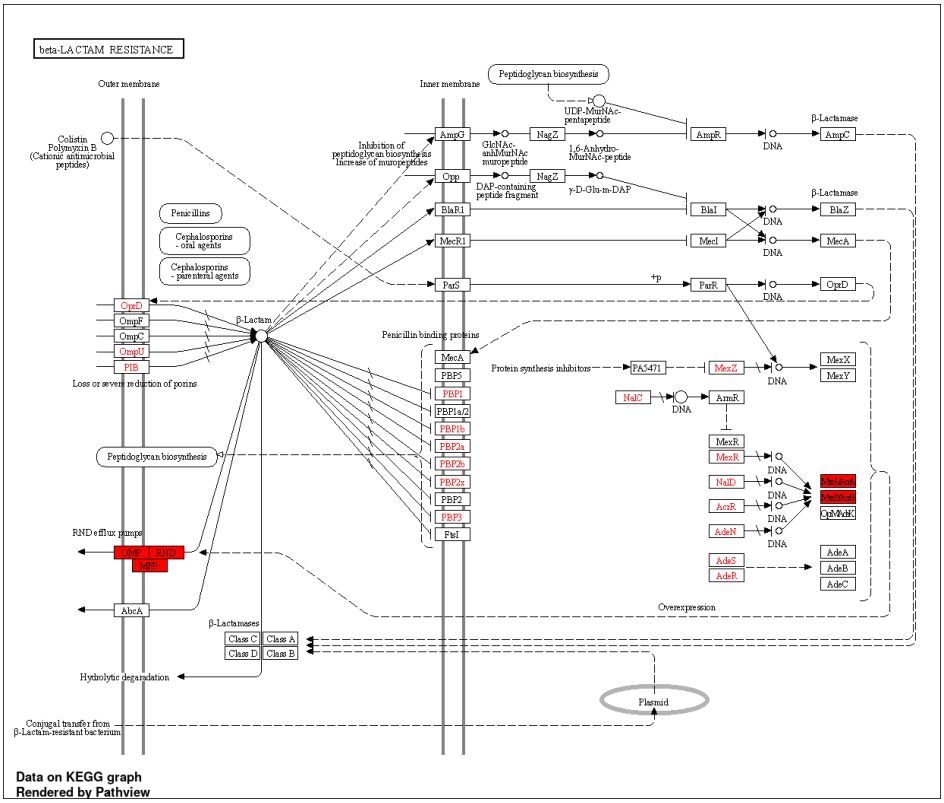


Figure 8. Some genes were found to be overexpressed (*mexA/acrA* and *mexB/acrB*) related to *OMP*, *RND*, and *MFP*, and were identified as markers in the beta-lactam resistance pathway (all genes are highlighted in a red box).

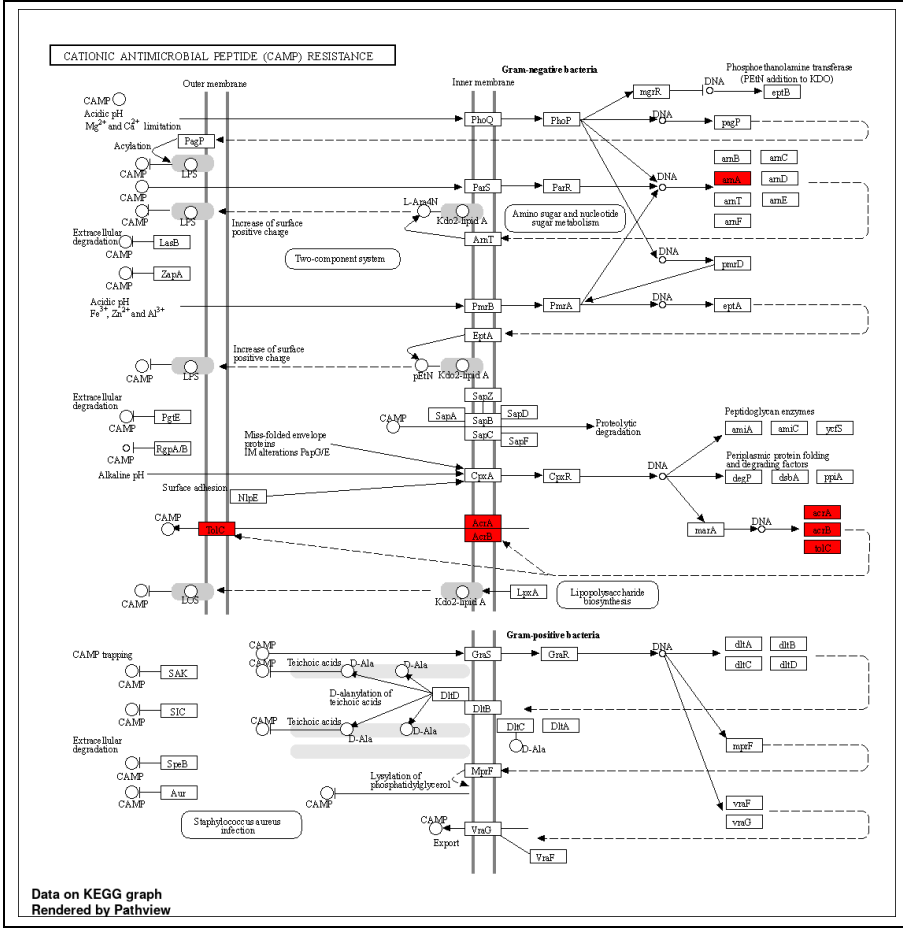


Figure 9. Some genes were found to be overexpressed (*mexA/acrA*, *arnA*, *tolC*, and *mexB/acrB*), which are markers of the cationic antimicrobial peptide (CAMP) pathway (all genes are highlighted in a red box).

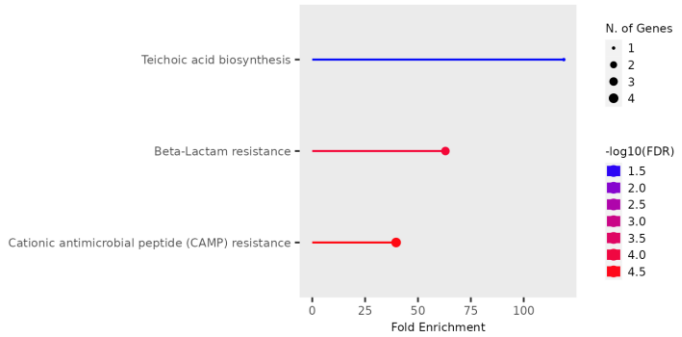


Figure 10. KEGG pathway analysis of 19 proteins related to antimicrobial resistance.

Table 5. GO enrichment from 19 proteins related to antimicrobial resistance.

GO term	Description	GO term (ID)	FDR
Biological process	Xenobiotic transport	GO:0042908	2.38E-11
	Response to the antibiotic	GO:0046677	5.16E-11
	Xenobiotic detoxification	GO:1990961	5.16E-11
	Response to the chemical	GO:0042221	8.56E-08
	Response to the drug	GO:0042493	1.55E-05
	Bile acid and bile salt transport	GO:0015721	2.39E-05
	Drug transport	GO:0015893	0.0012
	Drug export	GO:004618	0.006
Molecular function	Xenobiotic transmembrane transporter	GO:0042910	3.17E-08
	Efflux transmembrane transporter	GO:0015562	2.9E-05
	Transmembrane transporter	GO:0022857	0.00011
	Identical protein binding	GO:0042802	0.00019
	Bile acid transmembrane transporter	GO:0015125	0.00079
Cellular component	Efflux pump complex	GO:1990281	1.45E-06
	macAB-*tolC* complex	GO:1990196	4.7E-05
	Transmembrane transporter complex	GO:1902495	0.00024
	Intrinsic component of the membrane	GO:0031224	0.00049
	Integral component of the membrane	GO:0016021	0.003
	Cell periphery	GO:0071944	0.003
	Plasma membrane	GO:0005886	0.0085
	Intrinsic component of the plasma membrane	GO:0031226	0.0119
	Protein-containing complex	GO:0032991	0.0323

Table 6. AMR genes and antibiotic resistance group confirmation using VITEK® 2.

Sample code	Antibiotic resistance class	Antibiotic resistance genes (ARGs)
Sample 1	Aminoglycoside	*aac(3)-IId, aadA2 and aadA5, and aph(3'')-Ib*
	Beta-lactam	*blaCTX-M-27 and blaCTX-M-15*
Sample 2	Aminoglycoside	*aac(3)-IId, aac(6')-Ib-cr, and aadA2*
	Beta-lactam	*blaOXA-10 and blaOXA-1*
Sample 3	Beta-lactam	*blaCTX-M-15*
Sample 4	Beta-lactam	*blaCTX-M-15*
Sample 5	Aminoglycoside	*aac(6')-Ib-cr and aadA5*

DISCUSSION

The findings of this study demonstrate that distinct gene expression profiles are associated with specific diarrhea pathogens in pediatric subjects. Notably, non-pathogenic bacterial colonization did not elicit significant transcriptional changes, supporting the hypothesis that only pathogenic organisms trigger measurable systemic immune responses. The use of machine learning algorithms, particularly SVM and neural networks, validated the predictive capacity of selected DEGs, confirming their diagnostic potential.

Analysis of shared DEGs was conducted to support clinical data on diarrhea caused by microbial infection, followed by a search for URPs. Shared DEGs, or overlapping DEGs, refer to genes that show different (significant) expression levels across more than one comparison group or condition; in this case, DEGs are related to groups for each infectious microorganism that causes diarrhea (Zhao et al., 2018). In principle, these upstream regulators can include genes, proteins (URPs), miRNAs, drugs, chemicals, or toxins (Nguyen et al., 2023). The NOG (noggin) protein is a URP for shared DEGs with *ST14*. This protein is essential in regulating cell growth and differentiation during embryonic development. NOG is mainly known as an antagonist of BMP (bone morphogenetic protein) signaling, which plays a role in various essential signaling pathways for tissue development, including intestinal epithelial development. The imbalance between BMP and noggin was also predicted to compromise intestinal epithelial integrity, increasing susceptibility to infection and intestinal inflammation, which can lead to diarrhea. Like BMP, matriptase, encoded by *ST14*, is a protein that plays a role in the signaling pathway regulating the integrity of the intestinal epithelial layer and is also thought to act as an antagonist of NOG.

In addition to NOG, TGFB1, and PPARA, *ST14* is a URP for these genes. TGFB1 is a cytokine that regulates various biological processes, including inflammation and immune responses. TGFB1 can modulate immune cell activity, contributing to intestinal inflammation. Increased levels of TGFB1 are often associated with inflammatory conditions, which can worsen intestinal mucosal damage and interfere with nutrient absorption, thus contributing to diarrhea. Although TGFB1 can trigger inflammation, it also plays a role in tissue healing after damage. During infection, TGFB1 helps restore intestinal epithelial integrity after inflammation (Boumessid et al., 2023; Chen et al., 2023). PPARA regulates lipid metabolism and inflammatory responses in the gut, which are relevant to gastrointestinal health and infection (Guo et al., 2022). However, several PPARA agonists, such as pioglitazone, ciprofibrate, and pravastatin, are known to cause diarrhea (Xi et al., 2020). Based on the explanation above, TGFB1 and PPARA have essential roles in the pathogenesis of diarrhea caused by infection. Changes in both activities can be evaluated from changes in *ST14* expression in pediatric subjects with diarrhea-causing microorganisms.

Escherichia coli (ETEC) infection and inhibition of nuclear factor erythroid 2–related factor (2NFE2L2, also known as NRF2) activity can lead to increased oxidative stress and impaired intestinal integrity. In addition, rotavirus infection causes significant damage to intestinal epithelial cells, leading to increased permeability and fluid loss, which are major contributors to diarrhea. The NRF2 pathway is essential for protecting against inflammatory responses that compromise the intestinal barrier, triggered by subsequent infections (Lai et al., 2024). In addition, NRF2 activation reduces levels of pro-inflammatory cytokines, which are crucial for the immune response to pathogens (Chu et al., 2024). By inhibiting cytokines such as TNF-α and IL-6, NRF2 helps maintain a balanced immune response and prevents excessive inflammation (Pant et al., 2024). Changes in NRF2 activity appear to be evaluable from changes in *CREBL2* expression, as shown in this study.

Among the DEGs, *RYK*, *CREBL2*, and *ST14* emerged as central candidates due to their consistent regulation across multiple infections. These genes are associated with epithelial integrity and immune modulation. For instance, *ST14* encodes matriptase, a protease involved in epithelial barrier maintenance and inflammatory signalling. Its regulation by TGFB1 and NOG suggests that upstream signalling pathways are viable therapeutic targets (Bernatik et al., 2017; Danielsen et al., 2018; Hatamzade Esfahani & Day, 2023; List et al., 2009; Mulloy & Rider, 2015; Nikolovska et al., 2023). Similarly, *CREBL2* appears to be regulated by the antioxidant response factor NFE2L2, consistent with evidence that oxidative stress contributes to intestinal damage in infections such as rotavirus and *E. coli* (Jin et al., 2024).

The continuous evolution of bacterial resistance mechanisms is driven by changes in cell wall structure and various cellular characteristics. Resistance mechanisms include efflux pumps, uptake of resistant drugs, mutations at target sites, and alterations in target sites, and efflux pumps (Devi et al., 2024). All antimicrobial resistance genes (ARGs) derived from *E. coli* (*acrB*, *acrD*, *acrF*, *acrS*, *bacA*, *baeR*, *baeS*, *cpxA*, *CRP*, *emrA*, *emrB*, *emrK*, *emrR*, *emrY*, *eptA*) are listed in the comprehensive antibiotic resistance database (CARD). *E. coli* acrA and *E. coli acrR* exhibit mutations that confer multidrug antibiotic resistance. *E. coli mdfA*, *E. coli ampC1* beta-lactamase, *E. coli ampH* beta-lactamase, *E. coli emrE*, and *E. coli GlpT* exhibit mutations that confer resistance to fosfomycin. *E. coli marR* mutants associated

with antibiotic resistance, *E. coli ampC* beta-lactamase, *E. coli soxR* mutations linked to antibiotic resistance, *E. coli soxS* mutations related to antibiotic resistance, *evgS*, *evgA*, *gadW*, *gadX*, *kdpE*, *marA*, *mdtA*, *mdtB*, *mdtC*, *mdtE*, *mdtF*, *mdtG*, *mdtH*, *mdtM*, *mdtN*, *mdtO*, *mdtP*, *msbA*, *PmrF*, *TEM-116*, *TolC*, *ugd*, and *YojI* have been documented in this species. Several are indeed specific to this species (Jia et al., 2017; McArthur et al., 2013).

The findings of this study, as detailed in Table 5, demonstrate that overexpression of the *acrAB* and *tolC* genes affects bile production across GO biological processes and molecular functions. Furthermore, increased bile concentrations stimulate *RamR* interaction, leading to the overexpression of *ramA* and an enhancement in the production of the *AcrAB-TolC* system. Bile consists of a complex mixture of organic and inorganic components, including fatty acids and bile acids or salts. Prior research suggests that bile may significantly enhance the expression of RND efflux pumps (Novelli & Bolla, 2024). Bile components in the intestinal tract induce the expression of the *AcrAB-TolC* pump in enterobacteria, including opportunistic pathogens such as *S. enterica* and *E. coli*. Bile salts in *E. coli* result in the overexpression of *acrAB* and the inhibition of *ompF*, an outer membrane porin (Baucheron et al., 2014; Prouty et al., 2004; Yamasaki et al., 2019).

The RND efflux pump constitutes a substantial tripartite complex that traverses the Gram-negative membrane, extending from the inner to the outer membrane. The structure generally consists of a homomeric assembly of three components: the inner membrane pump (IMP), the membrane fusion protein (MFP), and the outer membrane porin (OMP) (Klenotic et al., 2021). The outer membrane of gram-negative bacteria is composed of various outer membrane proteins that form a β-barrel transmembrane structure. Most outer membrane proteins (OMPs) are incorporated into the outer membrane (OM) through the β-barrel assembly machinery (BAM) complex. The BAM complex in *E. coli* consists of two essential proteins, BamA and BamD, and three nonessential accessory proteins: BamB, BamC, and BamE (Thewasano et al., 2023). Investigating the overexpression of antimicrobial resistance genes can enhance drug discovery and development by targeting specific molecular pathways (Fig. 8). Although machine learning models showed high predictive accuracy, experimental validation through RT-qPCR or Western blotting is essential to confirm these transcriptomic findings. *In vitro* studies assessing gene function in intestinal epithelial models could further elucidate mechanistic pathways.

The confirmation study with five cohort fecal samples (Table 6) revealed that some samples (Sample 1-4) showed multidrug resistance to third-generation cephalosporins, such as cefotaxime, ceftazidime, and ceftriaxone, with ARGs *blaCTX-M-15*, *blaCTX-M-27*, *blaOXA-10*, and *blaOXA-1*. These genes play a role in the beta-lactam resistance pathway via extended-spectrum beta-lactamase genes (Yassara et al., 2025).

Regarding AMR, key efflux system genes (*acrB*, *macB*, *tolC*) were overexpressed in age-stratified analyses. These findings correspond with previous reports from the comprehensive antibiotic resistance database and suggest that even in the absence of antibiotic exposure, young children may harbor clinically relevant resistance mechanisms. Functional enrichment highlighted pathways central to multidrug resistance, including beta-lactam and CAMP resistance, underscoring the adaptive nature of bacterial populations under selective pressures, such as bile acids (Figs. 8 and 9). Some of the ARGs detected in five fecal samples in children with acute diarrhea consist of aminoglycosides and beta-lactam resistance, such as *aac(3)-Iid*, *aadA2*, *aadA5*, *aac(6')-Ib-cr*, and *blaCTX-M-15*. Collectively, the DEG profiles and associated regulatory networks offer insights into both the pathogenesis and potential intervention strategies for pediatric diarrhea disease.

CONCLUSION

This study identified several differentially expressed genes and antimicrobial resistance markers associated with pediatric diarrhea infections. Genes such as *RYK*, *CREBL2*, and *ST14* were consistently dysregulated across multiple pathogen-specific groups and were associated with key pathways involved in mucosal integrity and immune response. The activity of upstream regulatory proteins, including NFE2L2, TGFB1, PPARA, and NOG, was predicted to modulate the expression of these targets, providing potential avenues for host-directed therapies. Additionally, genes involved in efflux mechanisms and membrane integrity, such as *acrB*, *tolC*, and *macB*, were overexpressed in age-stratified cohorts, highlighting the persistence of resistance traits within the pediatric gut microbiome. Functional pathway analyses reinforced the relevance of multidrug efflux systems in resistance phenotypes. The integration of DEG analysis, machine learning validation, and protein interaction mapping provides a comprehensive framework for future diagnostic tool development and therapeutic targeting in diarrhea. Experimental validation is strongly recommended to confirm the functional relevance of the identified genes.

CONFLICT OF INTEREST

The research was conducted entirely free from any involvement in commercial, financial, or personal matters that might be perceived as a possible conflict of interest or that could have influenced the outcomes or interpretation of the study.

ACKNOWLEDGMENTS

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

GENERATIVE ARTIFICIAL INTELLIGENCE (AI)

The authors used Grammarly (Superhuman Platform) to improve the language and readability of the manuscript. The authors reviewed and edited the content and take full responsibility for its accuracy. No images were generated or manipulated using AI tools.

REFERENCES

Abbas M, El-Manzalawy Y (2020) Machine learning based refined differential gene expression analysis of pediatric sepsis. BMC Med Genomics 13: 122. https://doi.org/10.1186/s12920-020-00771-4

Baucheron S, Nishino K, Monchaux I, Canepa S, Maurel MC, Coste F, Roussel A, Cloeckaert A, Giraud E (2014) Bile-mediated activation of the *acrAB* and *tolC* multidrug efflux genes occurs mainly through transcriptional derepression of ramA in *Salmonella enterica* serovar Typhimurium. J Antimicrob Chemother 69(9): 2400–2406. https://doi.org/10.1093/jac/dku140

Bernatik O, Radaszkiewicz T, Behal M, Dave Z, Witte F, Mahl A, Cernohorsky NH, Krejci P, Stricker S, Bryja V (2017) A Novel Role for the BMP Antagonist Noggin in Sensitizing Cells to Non-canonical Wnt-5a/Ror2/Disheveled Pathway Activation. Front Cell Dev Biol 5: 47. https://doi.org/10.3389/fcell.2017.00047

Bisong E (2019) Logistic Regression. In: Building Machine Learning and Deep Learning Models on Google Cloud Platform. Apress, Berkeley, CA, pp. 243–250. https://doi.org/10.1007/978-1-4842-4470-8_20

Boumessid K, Lacroix V, Ovtchinnikova E, Quaranta Nicaise M, Roy M, Dumay A, Thenet S, Carriere M, Mas E, Barreau F (2023) Promoting intestinal healing by exclusive enteral nutrition with TGF-β in a mouse model of colitis. bioRxiv 2023.08.31.555691. https://doi.org/10.1101/2023.08.31.555691

Carvalho MJ, Sands K, Thomson K, Portal E, Mathias J, Milton R, Gillespie D, Dyer C, Akpulu C, Boostrom I, Hogan P, Saif H, Ferreira A, Nieto M, Hender T, Hood K, Andrews R, Watkins WJ, Hassan B, Walsh TR (2022) Antibiotic resistance genes in the gut microbiota of mothers and linked neonates with or without sepsis from low- and middle-income countries. Nat Microbiol 7(9): 1337–1347. https://doi.org/10.1038/s41564-022-01184-y

Chen L, Chu C, Lu J, Kong X, Huang T, Cai YD (2015) Gene ontology and KEGG pathway enrichment analysis of a drug target-based classification system. PLoS ONE 10(5): e0126492. https://doi.org/10.1371/journal.pone.0126492

Chen L, Qiu X, Dupre A, Pellon-Cardenas O, Fan X, Xu X, Rout P, Walton KD, Burclaff J, Zhang R, Fang W, Ofer R, Logerfo A, Vemuri K, Bandyopadhyay S, Wang J, Barbet G, Wang Y, Gao N, Verzi MP (2023) TGFB1 induces fetal reprogramming and enhances intestinal regeneration. Cell Stem Cell 30(11): 1520-1537.e8. https://doi.org/10.1016/j.stem.2023.09.015

Chen T, Guestrin C (2016) KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785–794. https://doi.org/10.1145/2939672.2939785

Chong CW, Alkatheeri AHS, Ali N, Tay ZH, Lee YL, Paramasivam SJ, Jeevaratnam K, Low WY, Lim SHE (2020) Association of antimicrobial resistance and gut microbiota composition in human and non-human primates at an urban ecotourism site. Gut Pathog 12: 14. https://doi.org/10.1186/s13099-020-00352-x

Chu Z, Zhu L, Zhou Y, Yang F, Hu Z, Luo Y, Li W, Luo F (2024) Targeting Nrf2 by bioactive peptides alleviate inflammation: Expanding the role of gut microbiota and metabolites. Crit Rev Food Sci Nutr 65(17): 3314–3333. https://doi.org/10.1080/10408398.2024.2367570

Danielsen ET, Olsen AK, Coskun M, Nonboe AW, Larsen S, Dahlgaard K, Bennett EP, Mitchelmore C, Vogel LK, Troelsen JT (2018) Intestinal regulation of suppression of tumorigenicity 14 (*ST14*) and serine peptidase inhibitor, Kunitz type -1 (SPINT1) by transcription factor CDX2. Sci Rep 8: 11813. https://doi.org/10.1038/s41598-018-30216-z

Davies J, Davies D (2010) Origins and evolution of antibiotic resistance. Microbiol Mol Biol Rev 74(3): 417–433. https://doi.org/10.1128/MMBR.00016-10

Davis AP, Wiegers TC, Johnson RJ, Sciaky D, Wiegers J, Mattingly CJ (2023) Comparative toxicogenomics database (CTD): Update 2023. Nucleic Acids Res 51(D1): D1257–D1262. https://doi.org/10.1093/nar/gkac833

DeBerg HA, Zaidi MB, Altman MC, Khaenam P, Gersuk VH, Campos FD, Perez-Martinez I, Meza-Segura M, Chaussabel D, Banchereau J, Estrada-Garcia T, Linsley PS (2018) Shared and organism-specific host responses to childhood diarrheal diseases revealed by whole blood transcript profiling. PLoS ONE 13(1): e0192082. https://doi.org/10.1371/journal.pone.0192082

del Toro N, Shrivastava A, Ragueneau E, Meldal B, Combe C, Barrera E, Perfetto L, How K, Ratan P, Shirodkar G, Lu O, Mészáros B, Watkins X, Pundir S, Licata L, Iannuccelli M, Pellegrini M, Martin MJ, Panni S, Hermjakob H (2022) The IntAct database: efficient access to fine-grained molecular interaction data. Nucleic Acids Res 50(D1): D648–D653. https://doi.org/10.1093/nar/gkab1006

Devi NS, Mythili R, Cherian T, Dineshkumar R, Sivaraman GK, Jayakumar R, Prathaban M, Duraimurugan M, Chandrasekar V, Peijnenburg WJGM (2024) Overview of antimicrobial resistance and mechanisms: The relative status of the past and current. The Microbe 3: 100083. https://doi.org/10.1016/j.microb.2024.100083

Feng J, Li B, Jiang X, Yang Y, Wells GF, Zhang T, Li X (2018) Antibiotic resistome in a large-scale healthy human gut microbiota deciphered by metagenomic and network analyses. Environ Microbiol 20(1): 355–368. https://doi.org/10.1111/1462-2920.14009

Gasparrini AJ, Wang B, Sun X, Kennedy EA, Hernandez-Leyva A, Ndao IM, Tarr PI, Warner BB, Dantas G (2019) Persistent metagenomic signatures of early-life hospitalization and antibiotic treatment in the infant gut microbiota and resistome. Nat Microbiol 4(12): 2285–2297. https://doi.org/10.1038/s41564-019-0550-2

Ge SX, Jung D, Yao R (2020) ShinyGO: a graphical gene-set enrichment tool for animals and plants. Bioinformatics 36(8): 2628–2629. https://doi.org/10.1093/bioinformatics/btz931

Guo YX, Wang BY, Gao H, Hua RX, Gao L, He CW, Wang Y, Xu JD (2022) Peroxisome proliferator–activated receptor-α: A pivotal regulator of the gastrointestinal tract. Front Mol Biosci 9: 864039. https://doi.org/10.3389/fmolb.2022.864039

Hatamzade Esfahani N, Day AS (2023) The role of TGF-β, Activin and follistatin in inflammatory bowel disease. Gastrointest Disord 5(2): 167–186. https://doi.org/10.3390/gidisord5020015

Huang YN, Jaiswal PV, Rajes A, Yadav A, Yu D, Liu F, Scheg G, Shih E, Boldirev G, Nakashidze I, Sarkar A, Mehta JH, Wang K, Patel KK, Mirza MAB, Hapani KC, Peng Q, Ayyala R, Guo R, Kapur

S, Ramesh T, Ciorbă D, Munteanu V, Bostan V, Dimian M, Abedalthagafi MS, Mangul S (2025) The systematic assessment of completeness of public metadata accompanying omics studies in the Gene Expression Omnibus data repository. Genome Biology 26(1): 274. https://doi.org/10.1186/s13059-025-03725-0

Jia B, Raphenya AR, Alcock B, Waglechner N, Guo P, Tsang KK, Lago BA, Dave BM, Pereira S, Sharma AN, Doshi S, Courtot M, Lo R, Williams LE, Frye JG, Elsayegh T, Sardar D, Westman EL, Pawlowski AC, McArthur AG (2017) CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. Nucleic Acids Res 45(D1): D566–D573. https://doi.org/10.1093/nar/gkw1004

Jin S, Xu H, Yang C, Karmin O (2024) Regulation of oxidative stress in the intestine of piglets after enterotoxigenic *Escherichia coli* (ETEC) infection. Biochim Biophys Acta Mol Cell Res 1871(5): 119711. https://doi.org/10.1016/j.bbamcr.2024.119711

Klenotic PA, Moseng MA, Morgan CE, Yu EW (2021) Structural and functional diversity of resistance–nodulation–cell division transporters. Chem Rev 121(9): 5378–5416. https://doi.org/10.1021/acs.chemrev.0c00621

Lai X, Wu A, Yu B, Yan H, Luo J, Zheng P, Yu J, Chen D (2024) Retinoic acid alleviates rotavirus-induced intestinal damage by regulating redox homeostasis and autophagic flux in piglets. Anim Nutr 16: 409–421. https://doi.org/10.1016/j.aninu.2023.12.007

List K, Kosa P, Szabo R, Bey AL, Wang CB, Molinolo A, Bugge TH (2009) Epithelial integrity is maintained by a matriptase-dependent proteolytic pathway. Am J Pathol 175(4): 1453–1463. https://doi.org/10.2353/ajpath.2009.090240

Lu N, Hu Y, Zhu L, Yang X, Yin Y, Lei F, Zhu Y, Du Q, Wang X, Meng Z, Zhu B (2014) DNA microarray analysis reveals that antibiotic resistance-gene diversity in human gut microbiota is age related. Sci Rep 4: 4302. https://doi.org/10.1038/srep04302

Malik YS, Verma AK, Kumar N, Touil N, Karthik K, Tiwari R, Bora DP, Dhama K, Ghosh S, Hemida MG, Abdel-Moneim AS, Bányai K, Vlasova AN, Kobayashi N, Singh RK (2019) Advances in diagnostic approaches for viral etiologies of diarrhea: From the lab to the field. Front Microbiol 10: 1957. https://doi.org/10.3389/fmicb.2019.01957

Martínez JL, Baquero F (2014) Emergence and spread of antibiotic resistance: Setting a parameter space. Ups J Med Sci 119(2): 68–77. https://doi.org/10.3109/03009734.2014.901444

McArthur AG, Waglechner N, Nizam F, Yan A, Azad MA, Baylay AJ, Bhullar K, Canova MJ, De Pascale G, Ejim L, Kalan L, King AM, Koteva K, Morar M, Mulvey MR, O'Brien JS, Pawlowski AC, Piddock LJV, Spanogiannopoulos P, Wright GD (2013) The comprehensive antibiotic resistance database. Antimicrob Agents Chemother 57(7): 3348–3357. https://doi.org/10.1128/AAC.00419-13

Mulloy B, Rider CC (2015) The bone morphogenetic proteins and their antagonists. In Vitam Horm 99: 63–90. https://doi.org/10.1016/bs.vh.2015.06.004

Natarajan M, Kumar D, Mandal J, Biswal N, Stephen S (2018) A study of virulence and antimicrobial resistance pattern in diarrhoeagenic *Escherichia coli* isolated from diarrhoeal stool specimens from children and adults in a tertiary hospital, Puducherry, India. J Health Popul Nutr 37: 17. https://doi.org/10.1186/s41043-018-0147-z

Nguyen TM, Craig DB, Tran D, Nguyen T, Draghici S (2023) A novel approach for predicting upstream regulators (PURE) that affect gene expression. Sci Rep 13: 18571. https://doi.org/10.1038/s41598-023-41374-0

Nikolovska K, Tytmonas E, Seidler A, di Stefano G, Voigtländer T, Hellms T, Lenzen H, Seidler U (2023) Altered colonocyte differentiation may lead to decreased electrolyte and fluid absorption in the ileocolon of IBD patients by affecting the expression pattern of the responsible ion transporters. Physiology 38: S1. https://doi.org/10.1152/physiol.2023.38.S1.5732076

Novelli M, Bolla JM (2024) RND Efflux pump induction: A crucial network unveiling adaptive antibiotic resistance mechanisms of gram-negative bacteria. Antibiotics 13(6): 501. https://doi.org/10.3390/antibiotics13060501

Okumu NO, Muloi DM, Moodley A, Watson J, Kiarie A, Ochieng L, Wasonga JO, Mutisya C, Alumasa L, Ngeranwa JJN, Cumming O, Cook EAJ (2025) Antimicrobial resistance in community-acquired enteric pathogens among children aged ≤ 10-years in low-and middle-income countries: a systematic review and meta-analysis. Front Microbiol 16: 1539160. https://doi.org/10.3389/fmicb.2025.1539160

Pant T, Uche N, Juric M, Zielonka J, Bai X (2024) Regulation of immunomodulatory networks by Nrf2-activation in immune cells: Redox control and therapeutic potential in inflammatory diseases. Redox Biol 70: 103077. https://doi.org/10.1016/j.redox.2024.103077

Pisner DA, Schnyer DM (2020) Support vector machine. In Machine Learning, Methods and Applications to Brain Disorders, pp. 101–121. https://doi.org/10.1016/B978-0-12-815739-8.00006-7

Pelkonen T, Dos Santos MD, Roine I, Dos Anjos E, Freitas C, Peltola H, Laakso S, Kirveskari J (2018) Potential diarrheal pathogens common also in healthy children in Angola. Pediatr Infect Dis J 37(5): 424-428. https://doi.org/10.1097/INF.0000000000001781

Prouty AM, Brodsky IE, Falkow S, Gunn JS (2004) Bile-salt-mediated induction of antimicrobial and bile resistance in *Salmonella typhimurium*. Microbiology 150(4): 775–783. https://doi.org/10.1099/mic.0.26769-0

Qi Y (2012) Random Forest for Bioinformatics. In: Zhang C, Ma Y (eds.) Ensemble Machine Learning. New York, NY: Springer, pp. 307–323. https://doi.org/10.1007/978-1-4419-9326-7_11

Schmidhuber J (2015) Deep learning in neural networks: An overview. Neural Netw 61(1): 85–117. https://doi.org/10.1016/j.neunet.2014.09.003

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T (2003) Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. Genome Res 13(11): 2498–2504. https://doi.org/10.1101/gr.1239303

Thewasano N, Germany EM, Maruno Y, Nakajima Y, Shiota T (2023) Categorization of *Escherichia coli* outer membrane proteins by dependence on accessory proteins of the β-barrel assembly machinery complex. J Biol Chem 299(7): 104821. https://doi.org/10.1016/j.jbc.2023.104821

Tobias J, Kassem E, Rubinstein U, Bialik A, Vutukuru SR, Navaro A, Rokney A, Valinsky L, Ephros M, Cohen D, Muhsen K (2015) Involvement of main diarrheagenic *Escherichia coli*, with emphasis on enteroaggregative E. coli, in severe non-epidemic pediatric diarrhea in a high-income country. BMC Infect Dis 15(1): 79. https://doi.org/10.1186/s12879-015-0804-4

Varela G, Batthyány L, Bianco MN, Pérez W, Pardo L, Algorta G, Robino L, Suárez R, Navarro A, Pírez MC, Schelotto F (2015) Enteropathogens associated with acute diarrhea in children from households with high socioeconomic level in Uruguay. Int J Microbiol 2015: 592953. https://doi.org/10.1155/2015/592953

Xi Y, Zhang Y, Zhu S, Luo Y, Xu P, Huang Z (2020) PPAR-mediated toxicology and applied pharmacology. Cells 9(2): 352. https://doi.org/10.3390/cells9020352

Yamasaki S, Nakashima R, Sakurai K, Baucheron S, Giraud E, Doublet B, Cloeckaert A, Nishino K (2019) Crystal structure of the multidrug resistance regulator RamR complexed with bile acids. Sci Rep 9: 177. https://doi.org/10.1038/s41598-018-36025-8

Yassara S, Zeouk I, Jaouhar S, Sbiti M, Bekhti K (2025) Extended-spectrum beta-lactamases: definition, history, an update on their genetic environment and detection methods. J Med Microbiol 74(6): 002033. https://doi.org/10.1099/jmm.0.002033

Zhao B, Erwin A, Xue B (2018) How many differentially expressed genes: A perspective from the comparison of genotypic and phenotypic distances. Genomics 110(1): 67–73. https://doi.org/10.1016/j.ygeno.2017.08.007

AUTHOR CONTRIBUTION:

Contribution	Erlina L	Fadilah F	Asmarinah	Hegar B	Tedjo A
Concepts or ideas	x	x			x
Design	x	x			x
Definition of intellectual content	x	x			x
Literature search	x	x	x	x	x
Experimental studies	x	x			
Data acquisition	x	x			x
Data analysis	x	x			x
Statistical analysis	x				x
Manuscript preparation	x	x	x	x	x
Manuscript editing	x	x			x
Manuscript review	x	x	x	x	x

Citation Format: Erlina L, Fadilah F, Asmarinah, Hegar B, Tedjo A (2026) Differentially expressed genes analysis for diagnostics and therapy of diarrhea due to pathogenicity and antibiotic resistance genes in children. J Pharm Pharmacogn Res 14(2): 2404. https://doi.org/10.56499/jppres_14.2.2404

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.

Open Access: This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/ licenses/by/4.0/), which permits use, duplication, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

Supplementary data

Table S1. Machine Learning Model Validation for DEG data (GSE69529).

Model	AUC	CA	F1	Precision	Recall	MCC
Neural network	0.917	0.941	0.942	0.950	0.941	0.883
SVM	0.943	0.912	0.913	0.929	0.912	0.831
Naïve Bayes	0.913	0.912	0.913	0.929	0.912	0.831

Table S2. Comparison Area Under ROC Curve for DEG data (GSE69529).

Machine learning model	Neural network	SVM	Naïve Bayes
Neural Network		0.169	0.271
SVM	0.831		0.813
Naïve Bayes	0.729	0.187	

Table S3. Confusion matrix for machine learning validation for DEG data (GSE69529).

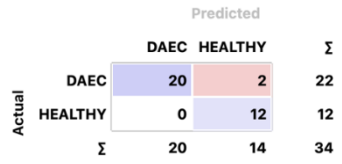
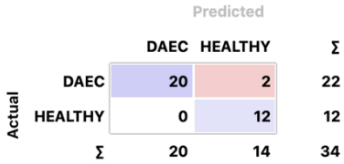
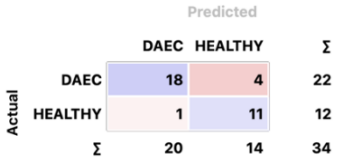
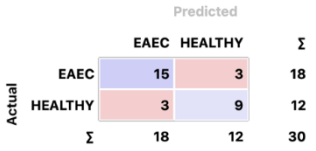
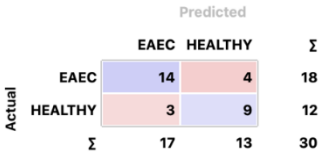
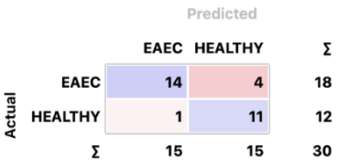
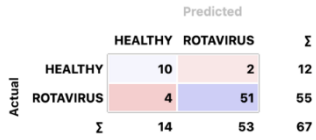
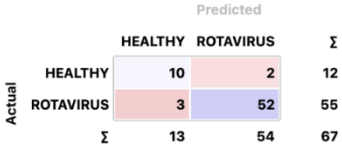
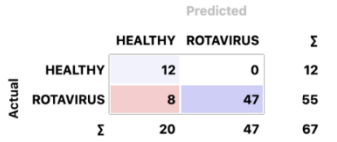
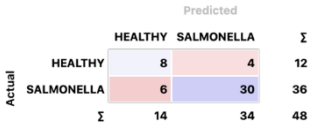
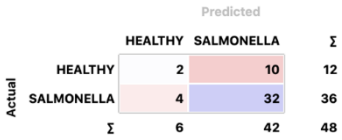
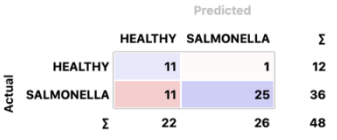
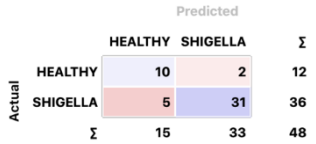
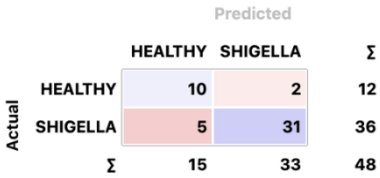
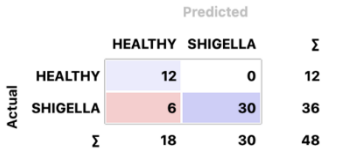
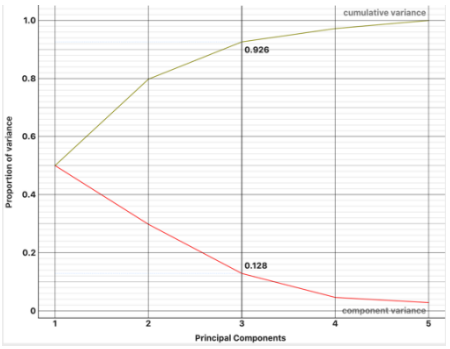
Neural network	SVM	Naïve Bayes
		
		
		
		
		

Table S4. The top five best-ranked ANOVA scores between healthy vs. *Rotavirus*, *Shigella*, *Salmonella*, DAEC, and EAEC.

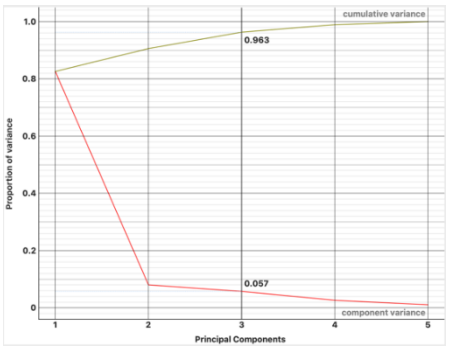
	Gene	ANOVA		Gene	ANOVA		Gene	ANOVA
Rotavirus	*IFI27*	23.242	***Shigella***	*NLRC4*	33.966	***Salmonella***	*RYK*	16.959
	CREBL2	21.565		*LTB4R*	33.491		*CREBL2*	15.006
	RYK	19.626		*RYK*	32.137		*ST14*	13.087
	SIGLEC5	17.364		*C19orf38*	31.402		*C19orf38*	11.190
	ST14	16.348		*BST1*	30.808		*OSCAR*	10.323
DAEC	*SIGLEC5*	19.307	**EAEC**	*RYK*	25.510			
	DDAH2	17.544		*CREBL2*	20.150			
	LTB4R	17.496		*DDAH2*	16.003			
	ST14	16.168		*SYN2*	14.565			
	SYN2	12.732		*S100A9*	13.795			

Table S5. Principal component analysis (PCA) based on the top five best-ranked ANOVA scores between healthy vs. *Rotavirus*, *Shigella*, *Salmonella*, DAEC, and EAEC.

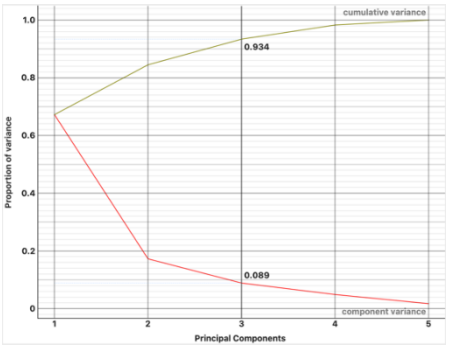
Rotavirus



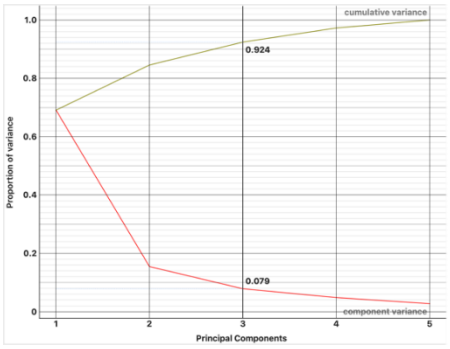
Shigella



Salmonella



DAEC



EAEC

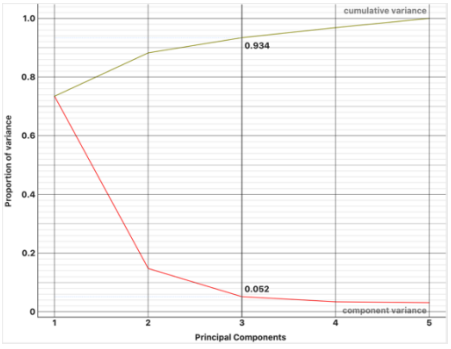


Table S6. Comparison Area Under ROC Curve for DEG data (GSE54070).

Machine learning model	Logistic regression	Random forest	SVM	XGBoost	Neural network
Logistic Regression		0.551	0.682	0.685	0.560
Random Forest	0.449		0.558	0.651	0.475
SVM	0.318	0.442		0.579	0.376
XGBoost	0.315	0.349	0.421		0.350
Neural Network	0.440	0.525	0.624	0.650	

Table S7. Cross-validation variance folds from 2 to 10 (non-stratified and stratified).

Model	AUC	CA	F1	Precision	Recall	MCC
Cross-validation folds: 2, stratified						
Logistic regression	0.997	0.995	0.995	0.995	0.995	0.943
Random forest	0.998	0.995	0.995	0.995	0.995	0.939
SVM	0.997	0.995	0.995	0.995	0.995	0.943
XGBoost	0.997	0.995	0.995	0.995	0.995	0.939
Neural network	0.997	0.995	0.995	0.995	0.995	0.943
Cross-validation folds: 2, non-stratified						
Logistic regression	0.997	0.995	0.995	0.995	0.995	0.943
Random forest	0.997	0.995	0.995	0.995	0.995	0.943
SVM	0.997	0.995	0.995	0.995	0.995	0.943
XGBoost	0.997	0.994	0.994	0.995	0.994	0.932
Neural network	0.997	0.995	0.995	0.995	0.995	0.943
Cross-validation folds: 3, stratified						
Logistic regression	0.997	0.995	0.995	0.995	0.995	0.943
Random forest	0.997	0.995	0.995	0.995	0.995	0.941
SVM	0.997	0.995	0.995	0.995	0.995	0.943
XGBoost	0.998	0.994	0.994	0.994	0.994	0.928
Neural network	0.997	0.995	0.995	0.995	0.995	0.943
Cross-validation folds: 3, non-stratified						
Logistic regression	0.997	0.995	0.995	0.995	0.995	0.943
Random forest	0.997	0.995	0.995	0.995	0.995	0.943
SVM	0.997	0.995	0.995	0.995	0.995	0.943
XGBoost	0.998	0.995	0.995	0.995	0.995	0.939
Neural network	0.997	0.995	0.995	0.995	0.995	0.943
Cross-validation folds: 5, stratified						
Logistic regression	0.997	0.995	0.995	0.995	0.995	0.943
Random forest	0.997	0.995	0.995	0.995	0.995	0.943
SVM	0.997	0.995	0.995	0.995	0.995	0.941
XGBoost	0.997	0.994	0.994	0.994	0.994	0.928
Neural network	0.997	0.995	0.995	0.995	0.995	0.943

Table S7. Cross-validation variance folds from 2 to 10 (non-stratified and stratified) (continued...)

Model	AUC	CA	F1	Precision	Recall	MCC
Cross-validation folds: 5, non-stratified						
Logistic regression	0.997	0.995	0.995	0.995	0.995	0.943
Random forest	0.997	0.995	0.995	0.995	0.995	0.941
SVM	0.997	0.982	0.983	0.987	0.982	0.830
XGBoost	0.997	0.993	0.993	0.993	0.993	0.917
Neural network	0.997	0.995	0.995	0.995	0.995	0.943
Cross-validation folds: 10, stratified						
Logistic regression	0.997	0.995	0.995	0.995	0.995	0.943
Random forest	0.997	0.995	0.995	0.995	0.995	0.941
SVM	0.997	0.995	0.995	0.995	0.995	0.943
XGBoost	0.997	0.995	0.995	0.995	0.995	0.943
Neural network	0.997	0.995	0.995	0.995	0.995	0.943
Cross-validation folds: 10, non-stratified						
Logistic regression	0.997	0.995	0.995	0.995	0.995	0.943
Random forest	0.998	0.995	0.995	0.995	0.995	0.943
SVM	0.997	0.994	0.994	0.995	0.994	0.936
XGBoost	0.997	0.995	0.995	0.995	0.995	0.939
Neural network	0.997	0.995	0.995	0.995	0.995	0.943

Table S8. Confusion matrix for machine learning validation for DEG data (GSE54070).

Logistic regression

Actual \ Predicted	Downregulation	Upregulation	Σ
Downregulation	254	0	254
Upregulation	30	5597	5627
Σ	284	5597	5881

Random forest

Actual \ Predicted	Downregulation	Upregulation	Σ
Downregulation	253	1	254
Upregulation	30	5597	5627
Σ	283	5598	5881

SVM

Actual \ Predicted	Downregulation	Upregulation	Σ
Downregulation	254	0	254
Upregulation	30	5597	5627
Σ	284	5597	5881

XGBoost

Actual \ Predicted	Downregulation	Upregulation	Σ
Downregulation	254	0	254
Upregulation	30	5597	5627
Σ	284	5597	5881

Neural network

Actual \ Predicted	EAEC	HEALTHY	Σ
EAEC	14	4	18
HEALTHY	3	9	12
Σ	17	13	30