

RESEARCH ARTICLE

In-silico identification of bacterial key-genes directly or indirectly associated with the development and progression of colorectal cancer for exploring anti-bacterial agents

Nibas Kumar Pal¹, Md. Kaderi Kibria¹, Tasfia Noor², Md. Feroj Ahmed¹, Md. Shariful Islam¹, Md. Foysal Ahmed¹, Md. Abdul Latif¹, Mohammad Ali¹, Md. Al Noman³, Dipto Kundu¹, Ovi Sharma^{1,4}, Md. Nurul Haque Mollah^{1*}

1 Bioinformatics Lab (Dry), Department of Statistics, University of Rajshahi, Rajshahi, Bangladesh, **2** Department of Computer Science and Engineering, Rajshahi University of Engineering and Technology, Rajshahi, Bangladesh, **3** Department of Statistics and Data Science, Barishal University, Barishal, Bangladesh, **4** TMSS Medical College, Bogura, Bangladesh

* mollah.stat.bio@ru.ac.bd



OPEN ACCESS

Citation: Pal NK, Kibria MK, Noor T, Ahmed MF, Islam MS, Ahmed MF, et al. (2026) In-silico identification of bacterial key-genes directly or indirectly associated with the development and progression of colorectal cancer for exploring anti-bacterial agents. PLoS One 21(6): e0343565. <https://doi.org/10.1371/journal.pone.0343565>

Editor: Filomena de Nigris, Università degli Studi della Campania Luigi Vanvitelli, ITALY

Received: February 6, 2026

Accepted: June 4, 2026

Published: June 26, 2026

Copyright: © 2026 Pal et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data availability statement: The raw 16S rRNA sequence profile dataset analyzed in this study is publicly available. It can be freely downloaded from the online EBI database with accession

Abstract

Colorectal cancer (CRC), which includes malignancies of the colon and rectum, constitutes a major global health challenge. Though there are several drugs that targets CRC-related genes/proteins, but their performance is not yet reach to the satisfactory level. Moreover, their effectiveness gradually decreases over time with long-term use, a phenomenon known as drug resistance. Therefore, it is required to explore new alternative candidate drugs against CRC. Several studies recommended CRC-related dysregulated host-genes guided candidate drugs. However, microbiome guided drug discovery particularly targeting bacterial key genes (bKGs) within CRC-associated gut microbiota remains very limited. This study aims to identify bKGs as antibacterial targets within CRC-associated bacterial taxa for exploring anti-bacterial agents. At first, we analysed a 16S rRNA-seq profile dataset that contained 24 CRC and 50 healthy samples, where beta diversity analysis results showed significant differences in bacterial compositions between CRC and HC groups. Differential abundance analysis with threshold values at $|\log_2 FC| > 1.0$ and adjusted p -value < 0.05 identified 42 significantly altered bacterial taxa of which *Bacteroides fragilis*, *Bacteroides ovatus*, *Bacteroides uniformis*, and *Flavonifractor plautii* were prioritized based on effect size and published literature reporting their association with CRC. Further, an integrative subtractive genomics and protein-protein interaction (PPI) network analyses was used to identify top-ranked 10 essential bKGs (*ribD*, *ribBA*, *murA*, *alr*, *hisI*, *hisE*, *hisD*, *hisG*, *hisH*, and *hisB*) from these four CRC-associated bacterial taxa as putative antibacterial targets. Finally, three candidate drug molecules (Sulfasalazine, Aminoglutethimide, and Tipiracil) were recommended as the preliminary bKGs-guided candidate anti-bacterial agents for CRC through

number PRJEB77293. <https://www.ebi.ac.uk/ena/browser/view/PRJEB77293>.

Funding: The author(s) received no specific funding for this work.

Competing interests: The authors have declared that no competing interests exist.

molecular docking and ADME/T analyses. Further experimental and clinical validation is required to establish these compounds as the effective drugs targeting the bKGs for CRC. Thus, these findings may provide insights for developing innovative anti-bacterial treatment approach relevant to CRC.

1. Introduction

Colorectal cancer (CRC), defined as malignancy originating in the colon or rectum, is a major global health concern and remains one of the leading causes of cancer-related mortality worldwide [1,2]. Despite substantial advances in early detection and therapeutic interventions, the global burden of CRC continues to rise, particularly in low and middle-income countries [3]. In 2020, CRC accounted for more than 1.9 million new cases and approximately 930,000 deaths, ranking as the third most commonly diagnosed cancer and the second deadliest malignancy globally [4,5]. Its global incidence is projected to reach 3.2 million new cases by 2040 [3]. Current treatment strategies primarily focus on host-driven molecular pathways [6]. Many patients experience limited therapeutic response, disease recurrence or treatment failure over time due to drug resistance, tumor heterogeneity and adverse side effects [7,8]. Although cumulative genetic and epigenetic alterations may contribute to CRC development, host-derived mechanisms alone do not fully explain CRC initiation and progression. Recent studies increasingly highlight the gut microbiome as a potential factor associated with colorectal carcinogenesis [9]. Under physiological conditions, the intestinal microbiota plays an essential role in maintaining host metabolic balance, immune regulation and epithelial integrity [10]. Disruption of this balanced microbial ecosystem referred to as dysbiosis has been associated with the expansion of potentially harmful bacterial taxa and altered host-microbe interactions, which have been implicated in CRC development though the causal and directional nature of these relationships remains to be fully established [11,12]. High-throughput sequencing technologies, including 16S rRNA gene sequencing and shotgun metagenomics, have significantly enhanced our understanding of CRC-associated microbial communities [13]. Evidence suggests bidirectional association between the host genome and gut metagenome, where changes in microbial composition have been linked to alterations in host molecular pathways and, conversely, host genetic variation has been associated with differences in microbial ecology [14,15]. Pathogenic bacteria may further play a role in disease progression through the expression of virulence proteins that potentially interact with host proteins, subvert immune responses, disrupt cellular signaling, and promote bacterial survival and replication, though whether these interactions directly contribute to disease progression remains to be functionally established [16–18]. Several 16S rRNA-based studies have identified distinct bacterial taxa associated with CRC, supporting the investigation of their functional gene content as putative targets for microbiome-guided therapeutic strategies [19–23]. Despite these advances in microbial profiling, the functional genetic determinants particularly essential genes unique within CRC-associated bacterial taxa,

designated herein as bacterial key genes (bKGs), and their functionally encoded protein products designated as bacterial key proteins (bKPs), remain poorly characterized. While host genome-guided therapeutic strategies for CRC have been extensively explored [24,25], microbiome-guided drug discovery particularly targeting essential bKGs within CRC-associated bacteria remains largely unexplored. Considering the potential role of bacterial genes in mediating host-microbe interactions, identifying essential bKGs in CRC-associated bacteria as antibacterial targets represents a promising avenue for developing antibacterial therapeutic strategies against CRC. To address this gap, the present study employed a comprehensive in-silico framework including 16S rRNA-based microbial profiling, subtractive genomics, PPI network analysis, molecular docking, and ADME/T analysis to identify bKGs as putative antibacterial targets in CRC associated bacterial species, and to explore bKG-guided antibacterial drug candidates. By integrating 16S rRNA sequencing-based microbial profiling with functional genomics and computational drug screening, this study aims to bridge the gap between CRC-associated microbiome composition and microbiome-guided antibacterial drug discovery. The findings are expected to provide novel insights into putative antibacterial intervention strategies and contribute to the development of innovative therapeutic candidates for CRC targeting essential bKGs. The entire workflow of the study depicted in Fig 1.

2. Materials and methods

2.1. Data source and description

In this study, publicly available 16S rRNA gene sequencing data from human stools, along with their corresponding metadata were analyzed. Sequence reads were retrieved from the European Nucleotide Archive (ENA) under BioProject accession number PRJEB77293 (<https://www.ebi.ac.uk/ena/browser/view/PRJEB77293>). The original dataset comprised 148 samples with 36 colonic tissue specimens and 112 stool samples. Clinical metadata (e.g., age, BMI, diet, medication use, and antibiotic exposure) were not available for the publicly accessible dataset used in this study. For the present analysis, only stool-derived microbiome profiles from clearly defined CRC case and healthy control (HC) were included. Specifically, 24 stool samples from CRC patients and 50 samples from HC representing the Sri Lankan population were selected [26].

2.2. Raw 16S rRNA sequence data processing

Raw 16S rRNA sequencing reads were considered to proper quality control and preprocessing prior to downstream analysis. As the data were generated within a single study under a consistent experimental framework, batch effects were not considered. Initial quality assessment was performed using FastQC (v0.12.1) [27] to detect low-quality reads, adapter contamination, and overrepresented sequences. Subsequent processing was carried out using QIIME 2 (amplicon-2024.10) [28]. Primer and adapter sequences, along with low-quality bases, were removed using the Cutadapt plugin [29] integrated within the QIIME 2 framework, a commonly used framework for microbial community studies. This initial processing was essential for maintaining the precision and reliability of microbiome assessment in both CRC and HC. Afterward, the paired end reads underwent quality filtering, trimming and denoising through the DADA2 [30] plugin within the QIIME 2 environment, allowing only high-quality sequences to be retained for downstream analyses. The denoised paired-end reads were then merged through the DADA2 workflow, preserving the default overlap and mismatch allowance to ensure accurate sequence merging. After merging, chimera removal and dereplication were performed using the DADA2 plugin within QIIME 2 to infer exact amplicon sequence variants (ASVs). The resulting ASV table contains the abundance of each unique sequence feature across all samples. This table formed the basis for subsequent downstream analyses.

2.3. Microbial diversity analysis

The assessment of bacterial diversity was conducted utilizing R software (v4.4.2). To evaluate alpha diversity, the analysis incorporated Chao1, Observed species, ACE, and Shannon metrics [31–33] calculated through the “phyloseq (v1.50.0)” R

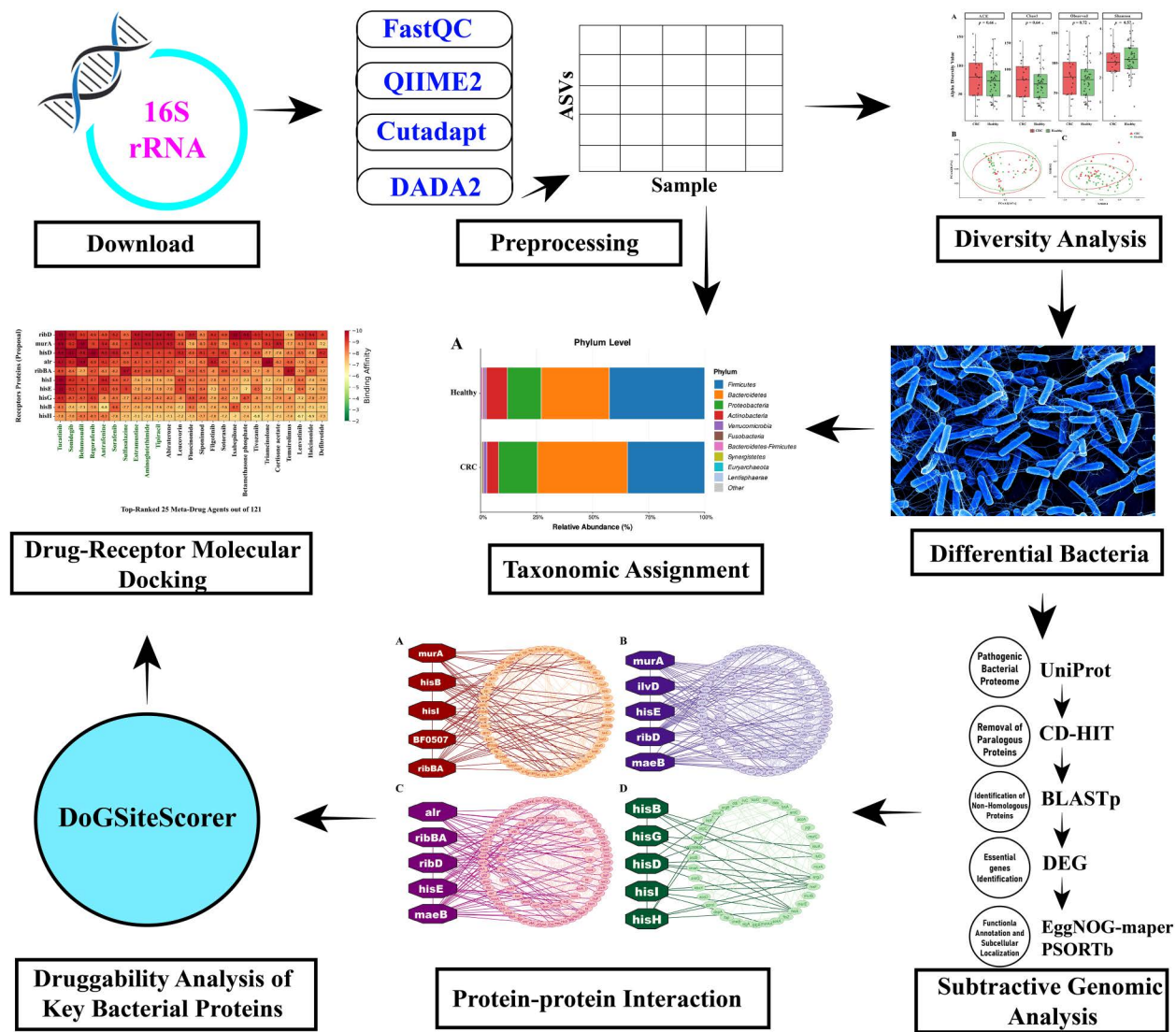


Fig 1. Workflow of the study.

<https://doi.org/10.1371/journal.pone.0343565.g001>

package [34,35]. The Chao1, Observed, and ACE indices quantify bacterial richness [36,37], whereas the Shannon index captures both species richness and community evenness [36]. Diversity assessments were conducted on rarefied count data standardized to the sequencing depth equals the minimum read count, using the “phyloseq” R package. Statistical comparisons of alpha diversity across groups were performed using the non-parametric Wilcoxon rank-sum test [38]. Conversely, beta diversity, which reflects differences in microbial composition between samples, was evaluated using the Bray-Curtis dissimilarity metric [39] based on the rarefied ASV abundance table, computed through the “phyloseq” package in R. Principal Coordinate Analysis (PCoA) was employed to graphically represent variations in microbial community structure among samples [40]. Beta diversity was also assessed using Jaccard dissimilarity and visualized through Non-metric Multidimensional Scaling (NMDS) [41]. Differences in bacterial community composition between the CRC and HCs were assessed using PERMANOVA based on Bray-Curtis and Jaccard distance metrics, implemented in the vegan (v2.6-10) package in R with 999 permutations [42].

2.4. Taxonomic profiling

Probabilistic taxonomic classification of the ASVs were carried out using a Naive Bayes classifier trained in QIIME 2, based on the Greengenes-SILVA-RDP (GSR) database, a manually curated and optimized 16S rRNA gene reference database [43]. Reference sequences and corresponding taxonomy for the V4 region (primers 515F/806R) were utilized to train the model through the feature-classifier plugin in QIIME 2 [44]. Taxonomic classification of ASVs was then inferred using the classify-sklearn method, though V4 amplicon-based classification provides probabilistic taxonomic assignments rather than definitive species identification and resolution at the species level may be limited particularly for complex genera. Relative abundance profiling was performed across hierarchical taxonomic levels (phylum, class, order, family, genus, and species) using QIIME 2 [45].

2.5. Differential abundance testing

Differentially abundant bacterial taxa between CRC and HC group were identified using a Zero-Inflated Gaussian Mixture Model (ZIGMM) [46] through the “metagenomeSeq”(version: 1.48.1) R package [47]. Taxa with adjusted p -value < 0.05 and $|\log_2 FC| > 1$ were defined as significantly differentially abundant. The false discovery rate (FDR) was controlled using the Benjamini-Hochberg correction to account for multiple hypothesis testing [48]. Taxa having greater effect size, species-level taxonomic annotations, and support from published articles about their associations with CRC, were considered to prioritize for further investigation.

2.6. Subtractive genomic approach for bKGs identification in CRC-associated bacteria

The full proteome of the selected bacterial taxa, along with the human reference proteome, were collected from UniProt database [49]. Reference proteomes were used as functional approximations, as the actual bacterial strains present in the patient samples remain uncharacterized and may differ from the reference strains analyzed. To remove paralogous sequences, the protein sequences were clustered using CD-HIT at a 60% sequence identity cutoff (word size = 4), ensuring that only non-paralogous proteins were retained for further analysis [50]. To exclude proteins with host similarity, all non-paralogous proteins were screened against the human proteome using BLASTp. Proteins sharing sequence identity greater than 30% and query coverage of at least 70% with any human proteins at an E-value threshold of 1×10^{-5} or lower were considered human homologs and excluded from further analysis [51]. Essential bacterial proteins were predicted by BLASTp searches against the Database of Essential Genes (DEG 10) [52], applying stringent thresholds of sequence identity $\geq 40\%$, query coverage $\geq 80\%$, bit score ≥ 100 , and E-value $\leq 1 \times 10^{-10}$ to select crucial proteins for bacterial survival [53–55]. Proteins failing to satisfy any of these conditions were excluded from downstream analysis. Functional characterization of the sequences covering enzyme commission (EC) numbers, KEGG pathway assignments was carried out with eggNOG-mapper [56]. PSORTb (v3.0) was employed to predict the subcellular locations of the candidate proteins, enabling distinction between cytoplasmic and membrane-associated proteins for effective drug-target prioritization [57–59]. Protein-protein interaction (PPI) networks of the cytoplasmic proteins of each taxa were generated using the STRING database with medium interaction confidence (0.4) and subsequently visualized in Cytoscape. [60,61]. Proteins exhibiting high interaction connectivity within the network were identified as hub proteins based on the Degree topological measure using the cytoHubba plugin in Cytoscape and considered potential therapeutic targets [62].

2.7. Druggability analysis of bKGs

Druggability of the target bKGs was analyzed using DoGSiteScorer [63], a pocket-detection and scoring module integrated into the ProteinPlus platform. The tool identifies potential ligand-binding cavities through a Difference-of-Gaussian-based segmentation procedure, which isolates regions of the protein surface likely to support small-molecule

interaction. For each protein, DoGSiteScorer computes a set of geometric and physicochemical features, including pocket volume and surface area, drug score (0–1), and simple score which provide an overall estimate of the pocket's suitability for binding drug-like compounds. Pockets with a drug score ≥ 0.5 were considered to have greater likelihood of being modulated by drug-like molecule [64]. When multiple pockets within a protein achieved comparable drug scores, the pocket with the greatest volume, surface area, and simple score was reported.

2.8. Molecular docking

To explore possible therapeutic agents for CRC, 121 drug compounds associated with CRC were intentionally extracted from the DrugBank database for drug repurposing perspective [65], a well-established and cost-effective strategy in computational drug discovery [66]. This is further supported by the evidence that non-antibiotic drugs can exhibit off-target antimicrobial activity [67], providing rationale for investigating CRC associated compounds as potential inhibitors of targets enriched in CRC-associated microbiota. The 3D structures of chosen drug molecules were obtained from the PubChem database [68], while structural models of the bacterial target proteins were sourced from AlphaFold [69], and SWISS-MODEL [70] databases. All molecular structures were systematically prepared prior to the molecular docking. The selected ligands were subjected to energy minimization, torsion tree generation, and charge assignment using Avogadro [71] and AutoDockTools4 [72]. Protein structures were refined by eliminating crystallographic water molecules, non-essential heteroatoms, and extraneous protein chains using Discovery Studio [73,74], followed by hydrogen addition, Kollman charge assignment, and grid box preparation in AutoDockTools4 [72]. Molecular docking of the preprocessed drug candidates with the target protein structures was performed using AutoDock Vina [75]. Blind docking was performed by defining a grid box encompassing the entire protein surface, allowing AutoDock Vina to globally explore all possible binding cavities without prior assumptions regarding active site location. The Binding Affinity Scores (BAS) from each docking result were assessed to identify the compounds showing the strongest predicted interactions with the target proteins.

2.9. Drug-likeness and ADME/T analysis

To evaluate the pharmacokinetic and toxicity profiles of the top ten candidate compounds, drug-likeness and ADME/T analyses were performed using the Deep-PK [76] and pkCSM [77]. Compound structures were submitted in Simplified Molecular Input Line Entry Specification (SMILES) format as input. Drug-likeness was assessed based on Lipinski's Rule of Five [78], which defines acceptable physicochemical boundaries as molecular weight ≤ 500 Da, hydrogen bond donors ≤ 5 , hydrogen bond acceptors ≤ 10 , and lipophilicity ($\log P$) ≤ 5 . ADME/T properties including absorption, distribution, metabolism, excretion, and toxicity parameters were predicted and compounds satisfying both drug-likeness and ADME/T criteria were shortlisted as potential candidates for further consideration. The parameters of gut-targeted antibacterial activity including intestinal luminal stability, bacterial membrane uptake, luminal availability, and activity under anaerobic conditions were not assessed in this study. Therefore, ADME/T findings are preliminary filters for drug-likeness and safety assessment, rather than definitive indicators of suitability for gut-targeted antibacterial activity.

2.10. Validation of docking protocol using decoy molecules

To assess docking specificity, 40 decoy molecules were generated for each proposed drug compound as negative controls using the Directory of Useful Decoys Enhanced (DUDE-Z) [79], which provides high-quality property-matched decoys using the ZINC database (S15 Table). Molecular docking was subsequently performed between each target protein and its corresponding decoy ligands. Docking reliability was evaluated by comparing binding affinity of the proposed drug compounds against their corresponding decoy sets, confirming that active compounds consistently achieved superior binding affinities relative to decoy molecules across all target proteins [80].

2.11. Binding site assessment of final selected complexes

To assess binding site consistency of the final shortlisted compounds, a post-docking evaluation was performed for the ADME/T-screened ligands in complex with the two principal target proteins. Potential druggable regions (drug score >0.5) within the protein structures were identified using DoGSiteScorer. Subsequently, the spatial agreement between docked poses and predicted cavities was analyzed to evaluate binding site correspondence.

2.12. Molecular dynamics (MD) simulation

To examine the dynamic behavior of the most promising protein-ligand complexes, molecular dynamics (MD) simulations were performed using YASARA software [81]. The three ligands Sulfasalazine, Tipiracil, and Aminoglutethimide which demonstrated superior docking scores, satisfied Lipinski Rule of Five criteria, and exhibited favorable ADME/T profiles, were selected for 100 ns simulations of their respective protein-ligand complexes under physiological environments. Trajectory data were recorded at 100 ps intervals and subsequently analyzed using YASARA macros [82], alongside the SciDAVis software package (<http://scidavis.sourceforge.net/>). To further evaluate the stability of the interactions throughout the simulation, binding free energies (ΔG_{bind}) were determined at 100 ps intervals employing the MM-PBSA approach implemented within YASARA [83].

3. Results

3.1. Preprocessing of bacterial 16S rRNA-sequence data

A total of 74 stool samples (24 CRC patients and 50 healthy controls) were included in the analysis. Following sequence processing, a high proportion of paired end reads (81%) were successfully merged, resulting in a feature table comprising 1,291 amplicon sequence variants (ASVs) across all samples. The dataset contained a total of 1,308,541 reads, with a median feature frequency of 17,981 counts per sample. Read counts ranged from 7,101 to 25,675 per sample, indicating adequate sequencing depth across samples. Prior to downstream analyses, unclassified and unassigned ASVs at the phylum level including 'Unclassified Bacteria' and 'Unclassified Archaea' were removed from the feature table using the "dplyr" (v1.1.4) package in R. After removing these, the feature table comprised 1,276 ASVs across 74 samples with median feature frequency of 17,980 counts per sample. The highest read count in a sample was 25,527 while the lowest was 7,101. Rarefaction analysis showed a progressive increase in ASV richness with greater sequencing depth, and the eventual plateau of the curves indicated that sampling effort was adequate and captured the majority of microbial diversity. Accordingly for diversity analysis, all samples were rarefied to the minimum sample read count. This ensured an even sampling effort for alpha and beta diversity comparisons. For differential abundance analysis, ASVs with prevalence below 5% of samples were excluded to minimize noise and retain biologically meaningful features. The final filtered feature table was normalized to provide a robust foundation for identifying taxa significantly associated with CRC.

3.2. Bacterial diversity analysis

Alpha and beta diversity analyses were performed to compare bacterial composition between HC and CRC samples. Alpha diversity was assessed using the Chao1, Observed, ACE, and Shannon indices (Fig 2A), and group differences were evaluated using the Wilcoxon rank-sum test. No statistically significant differences in alpha diversity were observed between CRC and HC across any of the indices assessed (Observed: p -value=0.72; Shannon: p -value=0.57; Chao1: p -value=0.64; ACE: p -value=0.66), suggesting that overall bacterial richness and evenness were comparable between groups.

Consequently, beta diversity was assessed using Bray-Curtis distance and Jaccard dissimilarity; the results were visualized using PCoA plots and NMDS plot respectively (Fig 2B-2C). Both metrics revealed different community composition patterns between groups. PERMANOVA test based on both metrics indicated significant differences in bacterial

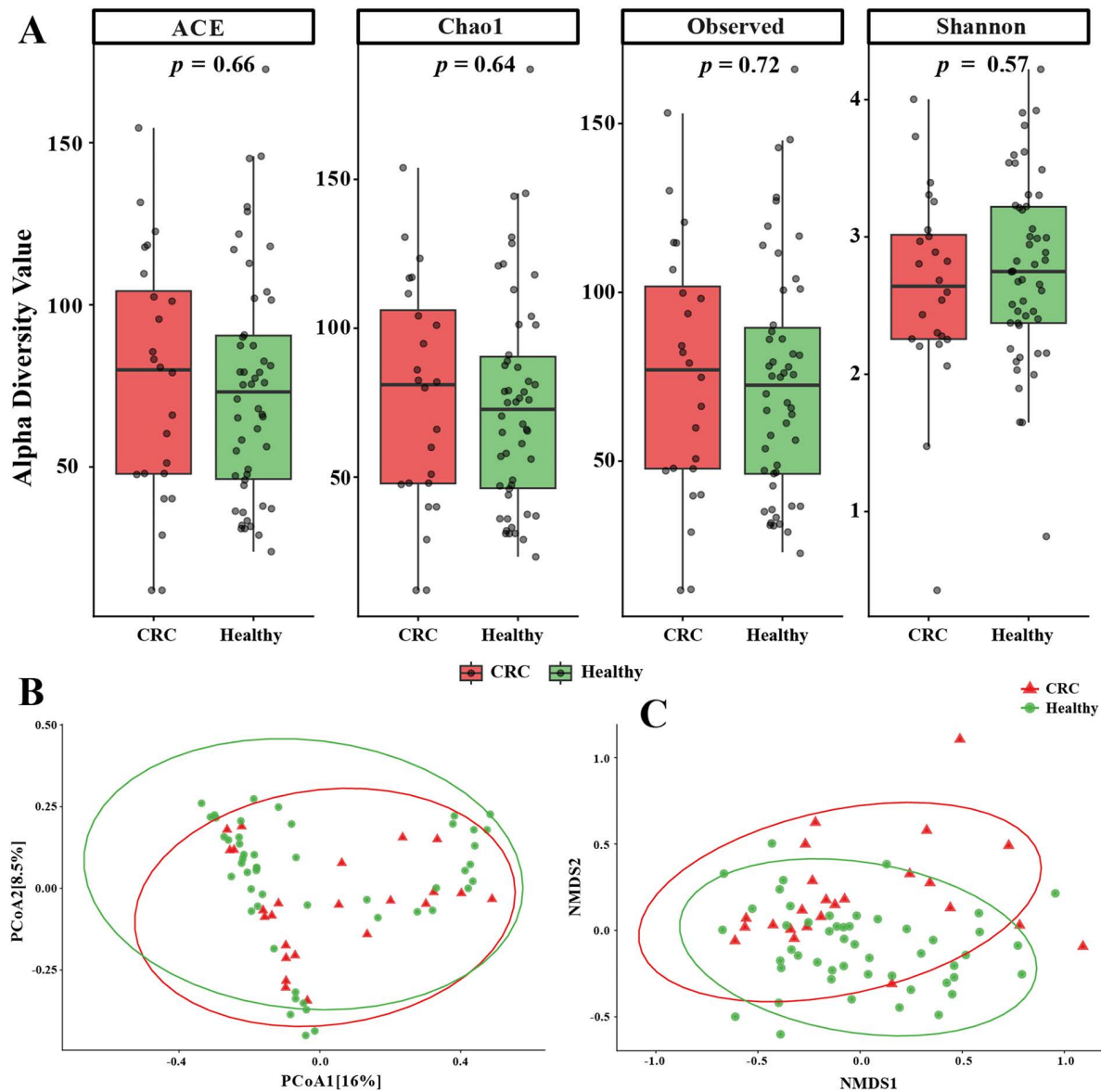


Fig 2. Alpha and beta diversity of the microbiota structure in CRC patients and HC. (A) Boxplots indicate no significant difference in alpha diversity indices (ACE, Chao1, Observed species, Shannon) between CRC and HC. (B) and (C) highlight the Beta diversity using Bray-Curtis and Jaccard dissimilarity exhibiting significantly distinct microbial community composition.

<https://doi.org/10.1371/journal.pone.0343565.g002>

community composition between CRC and HC groups (Bray-curtis: $F = 4.076$, p -value = 0.0472, Jaccard-dissimilarity: $F = 1.622$, p -value = 0.007). Though the PERMANOVA result based on Bray-Curtis dissimilarity showed borderline significance (p -value = 0.0472), community-level differences were additionally supported by Jaccard dissimilarity analysis (p -value = 0.007), providing complementary evidence for microbiome compositional separation between groups. A test for homogeneity of group dispersions based on Bray-Curtis distance (Bray-curtis: $F = 0.00998$, p -value = 0.92) and Jaccard dissimilarity (Jaccard: $F = 0.207$, p -value = 0.65) indicated no significant difference in variance between groups, suggesting

that the observed differences were not attributable to within-group variability, supporting the reliability of the PERMANOVA results.

3.3. Taxonomic profiling

Taxonomic inference of the gut microbiome identified 1,276 distinct ASVs, with samples containing an average of 17,568 reads per sample. These ASVs were distributed across 15 phyla, 26 classes, 40 orders, 97 families, and 236 genera, reflecting broad microbial diversity. Across all samples, the microbial community was predominantly composed of Firmicutes (39.88%), Bacteroidetes (33.55%), Proteobacteria (15.8%), and Actinobacteria (8.06%), with smaller contributions from Verrucomicrobia (1.05%) and Fusobacteria (0.61%). Compared to HC, CRC patients showed a higher relative abundance of Bacteroidetes (CRC: 40.22% vs. HC: 30.36%), Proteobacteria (CRC: 17.26% vs. HC: 15.11%), and Verrucomicrobia (CRC: 1.5% vs. HC: 0.84%) along with a decreased abundance of Firmicutes (CRC: 34.38% vs. HC: 42.52%) and Actinobacteria (CRC: 5.2% vs. HC: 9.44%) ([Fig 3A](#)) ([S1 Table](#)).

At the family level, Prevotellaceae (17.67%), Bacteroidaceae (11.74%), Enterobacteriaceae (11.69%), and Lachnospiraceae (8.10%) were identified as the most prevalent families ([S2 Table](#)). In the CRC group, elevated relative abundances were observed in Bacteroidaceae, Veillonellaceae, and Succinivibrionaceae, whereas Bifidobacteriaceae, Lactobacillaceae, Streptococcaceae, and Lachnospiraceae, exhibited a reduction in relative abundance ([Fig 3B](#)). Consistently, genus-level analysis revealed increased abundances of *Bacteroides*, *Phocaeicola*, *Succinivibrio*, and *Alistipes* (phylum Bacteroidetes and Proteobacteria) in CRC, alongside reduced levels of *Bifidobacterium* and *Streptococcus* (phylum Actinobacteria and Firmicutes) ([Fig 3C](#)) ([S3 Table](#)).

3.4. Identification of differentially abundant bacterial taxa

To identify significantly differentially abundant bacterial taxa, ZIGMM was applied using thresholds of adjusted p -value < 0.05 and $|\log_2FC| > 1$. Phylum level analysis revealed significant enrichment of Actinobacteria ($\log_2FC = -1.744$, $adj.p = 0.0074$) in HC group, while Fusobacteria ($\log_2FC = 1.13$, $adj.p = 0.0411$) was significantly enriched in CRC group ([S1 Fig](#)). At the family level, 15 taxa showed clear significant group-specific differences including Peptostreptococcaceae, Lactobacillaceae, Streptococcaceae, and Clostridiaceae which were enriched in HC group, whereas Bacteroidaceae, Rikenellaceae, Acidaminococcaceae, and Odoribacteraceae were more abundant in CRC group ([S2 Fig](#)), indicating a marked shift in gut microbial composition associated with CRC ([S4 Table](#)). At the genus level, 28 genera emerged as differentially abundant with 18 enriched and 10 depleted ([Fig 4A](#)) in CRC including *Romboutsia* (Peptostreptococcaceae; $\log_2FC = -4.09$, $adj.p = 4.68E-07$), *Bacteroides* (Bacteroidaceae; $\log_2FC = 3.49$, $adj.p = 0.00048$), *Clostridium* (Clostridiaceae; $\log_2FC = -3.38$, $adj.p = 6.66E-05$), *Alistipes* (Rikenellaceae; $\log_2FC = 2.54$, $adj.p = 0.00011$), *Streptococcus* (Streptococcaceae; $\log_2FC = -2.30$, $adj.p = 0.0309$), *Flavonifractor* (Firmicutes; $\log_2FC = 2.25$, $adj.p = 1.02E-06$). Complete genus-level results are compiled in [S5 Table](#). For identifying differentially abundant bacterial species, ZIGMM is used at the ASV level, revealing 42 differentially abundant ASVs between CRC and HC groups ([Fig 4B](#)), which were subsequently mapped to corresponding bacterial species.

Among these 42 species, 21 were enriched in CRC patients, while the remaining 21 were depleted compared to HCs including *Bacteroides fragilis* (*Bacteroides*; $\log_2FC = 2.23$, $adj.p = 0.00011$), *Bacteroides uniformis* (*Bacteroides*; $\log_2FC = 2.20$, $adj.p = 0.0014$), *Bacteroides ovatus* (*Bacteroides*; $\log_2FC = 3.14$, $adj.p = 1.12E-05$), *Flavonifractor plautii* (*Flavonifractor*; $\log_2FC = 1.90$, $adj.p = 3.65E-07$) and *Clostridium celatum* (*Firmicutes*; $\log_2FC = -2.02$, $adj.p = 1.14E-05$). All 42 differential bacterial species information are compiled in [S6 Table](#). The top 10 bacterial taxa were initially considered among the 42 significantly differentially abundant species. Among these top 10, taxa having greater effect size, species-level taxonomic annotations, and support from published articles about their associations with CRC, were prioritized for further investigation. Consequently, four bacterial taxa (*Bacteroides fragilis*, *Bacteroides ovatus*, *Bacteroides uniformis*, and *Flavonifractor plautii*) were selected. Although *Bacteroides uniformis* and *Bacteroides ovatus* are regarded

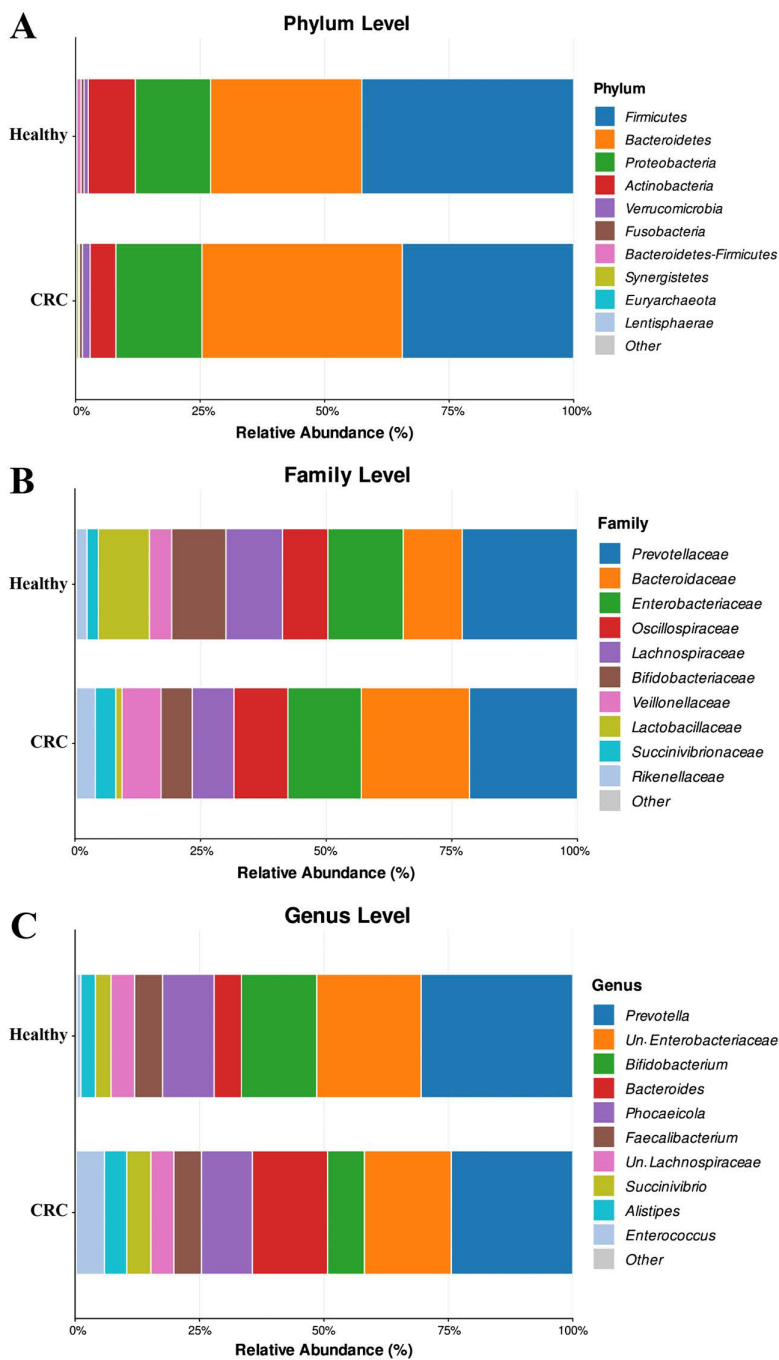


Fig 3. The stacked bar plots illustrate the relative composition of taxa in the CRC and HC groups at (A) phylum, (B) family, and (C) genus levels. The proportion of top 10 phylum, family and genera are displayed.

<https://doi.org/10.1371/journal.pone.0343565.g003>

as commensals, their association with CRC has been reported in several previous studies [19,84–103], and their inclusion is further supported by evidence that commensal bacteria may exhibit opportunistic pathogenic behavior depending on host immune status and environmental conditions, and may enhance the virulence of co-occurring pathogens in mixed

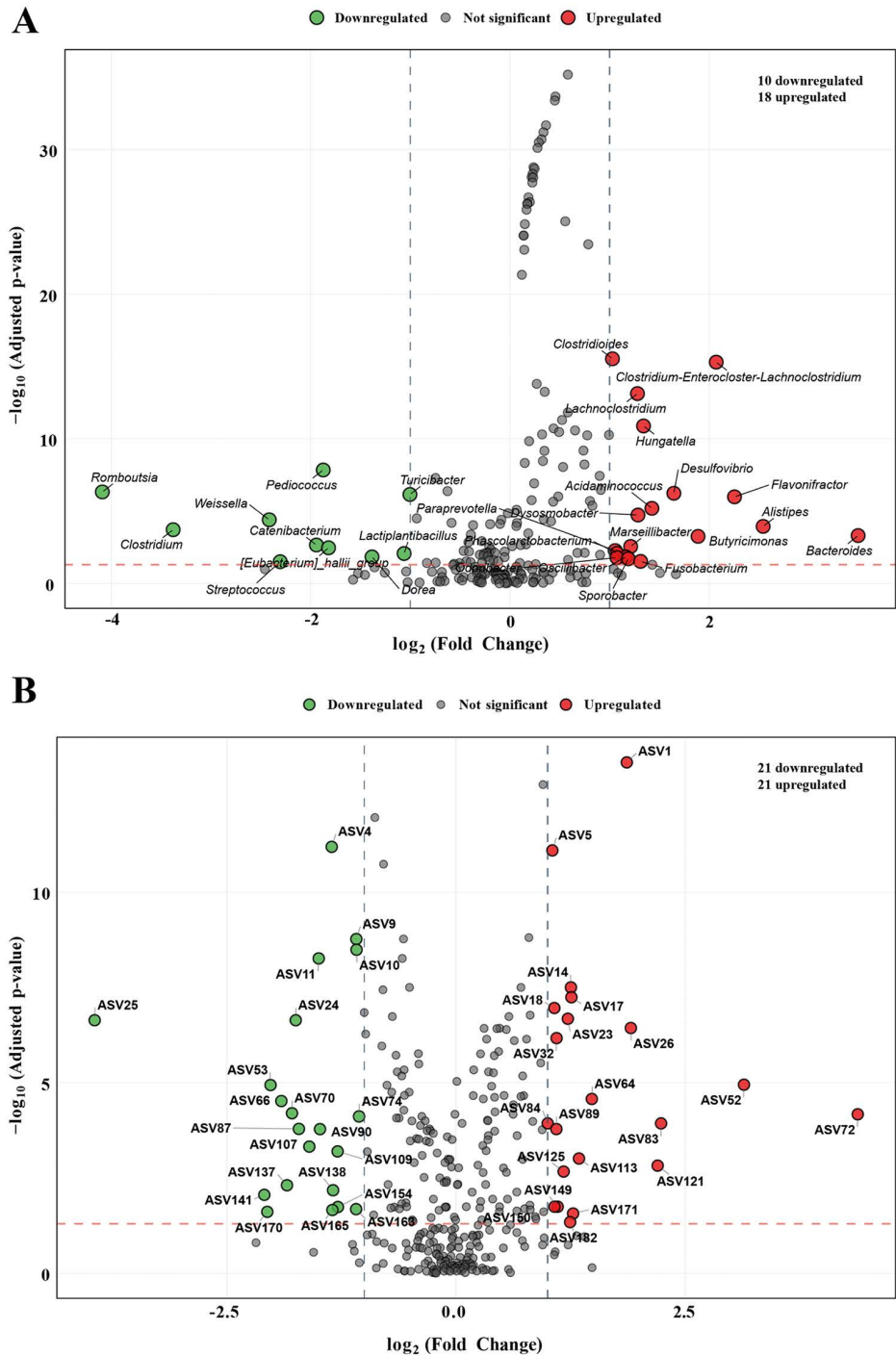


Fig 4. Differentially abundant taxa (A) genus and (B) species detected by ZIGMM ($|\log_2FC| > 1.0$ and $adj.p < 0.05$) between CRC and HC individuals.

<https://doi.org/10.1371/journal.pone.0343565.g004>

infections [104,105]. The top 10 most differentially abundant bacterial taxa, ranked by adjusted p-value and log₂FC, are presented in [Table 1](#).

3.5. Subtractive genomic approach for bKGs identification in CRC-associated bacteria

Subtractive genomic analysis was conducted on the four CRC-associated bacterial species identified through differential abundance analysis—*Bacteroides fragilis*, *Bacteroides uniformis*, *Bacteroides ovatus*, and *Flavonifractor plautii*. The complete reference proteomes for these organisms were retrieved from the UniProt database using accession IDs UP000006731, UP000004110, UP000473905, and UP000029585, respectively. These reference proteomes served as functional approximations as the actual bacterial strains present in the CRC samples may differ from the reference strains. A summary of the proteomic characteristics for each species is provided in [Table 2](#).

After collecting the bacterial proteomes, paralogous entries were removed using the CD-HIT with a 60% identity cutoff, identifying 149 paralogs in *BaF*, 199 in *BaU*, 451 in *BaO*, and 285 in *FlaP*. The remaining non-paralogous proteins were screened against the human proteome using BLASTp, applying a sequence identity threshold of ≤30%, an E-value cutoff of ≤1 × 10⁻⁵, and a minimum query coverage of ≥70%, yielding 3,737 non-homologous proteins in *BaF*, 4,084 in *BaU*, 5,240 in *BaO*, and 4,047 in *FlaP*. These non-homologous proteins were then considered as candidates for downstream drug-target prioritization.

Table 1. Top 10 significantly altered gut microbiota between CRC and HCs, based on log₂FC and adjusted p-value.

Phylum	Family	Genus	Species	log ₂ FC	Adjusted P-value
Bacteroidetes	Bacteroidaceae	<i>Bacteroides</i>	<i>spp.</i>	4.38	6.70E-05
Firmicutes	Peptostreptococcaceae	<i>Romboutsia</i>	<i>unclassified</i>	-3.94	2.28E-07
Bacteroidetes	Bacteroidaceae	<i>Bacteroides</i>	<i>ovatus</i>	3.14	1.12E-05
Bacteroidetes	Bacteroidaceae	<i>Bacteroides</i>	<i>fragilis</i>	2.23	0.0001
Bacteroidetes	Bacteroidaceae	<i>Bacteroides</i>	<i>uniformis</i>	2.20	0.001
Actinobacteria	Bifidobacteriaceae	<i>Bifidobacterium</i>	<i>unclassified</i>	-2.09	0.0087
Firmicutes	Lachnospiraceae	<i>Blautia</i>	<i>unclassified</i>	-2.05	0.024
Firmicutes	Clostridiaceae	<i>Clostridium</i>	<i>celatum</i>	-2.02	1.14E-05
Firmicutes	Oscillospiraceae	<i>Flavonifractor</i>	<i>plautii</i>	1.90	3.65E-07
Firmicutes	Clostridiaceae	<i>Clostridium</i>	<i>unclassified</i>	-1.90	3.01E-05

<https://doi.org/10.1371/journal.pone.0343565.t001>

Table 2. Steps involved in selecting key proteins of identified CRC-associated bacteria through subtractive genomics analysis.

S.N.	Subtractive approaches	<i>BaF</i> (strain: ATCC 25285)	<i>BaU</i> (strain: ATCC 8492)	<i>BaO</i> (strain: BIOML-A134)	<i>FlaP</i> (strain: 1_3_50AFAA)
1.	Proteome (Total sequence)	4234	4618	6058	4332
2.	Non-paralogous sequences	4085	4419	5607	4047
3.	Non-homologous proteins	3737	4084	5240	3614
4.	Essential proteins	665	607	658	322
5.	Essential metabolic proteins	152	153	155	115
6.	Subcellular localization (Cytoplasmic)	111	118	119	90
7.	PPI networking analysis	5	5	5	5
8.	Drug target analysis	5	5	5	5

**BaF*: *Bacteroides fragilis*, *BaU*: *Bacteroides uniformis*, *BaO*: *Bacteroides ovatus*, *FlaP*: *Flavonifractor plautii*.

<https://doi.org/10.1371/journal.pone.0343565.t002>

To determine the essentiality, the non-homologous protein sets were screened against the DEG10 database, applying thresholds of sequence identity $\geq 40\%$, query coverage $\geq 80\%$, bit score ≥ 100 , and E-value $\leq 1 \times 10^{-10}$, identifying 665 essential proteins in *BaF*, 607 in *BaU*, 658 in *BaO*, and 322 in *FlaP*. Metabolic pathway analysis of these essential non-homologous proteins was performed using the eggNOG-mapper server. Pathway analysis results showed the number of pathways that were shared with human pathways, and the number of proteins that are unassigned to KO identifiers. For *BaF*, out of 665 essential non-homologous proteins 230 proteins lacked KO assignments, 129 KO-annotated proteins were not linked to any pathway, and 166 pathways overlapped with those in humans. In *BaU*, 183 proteins lacked KO identifiers, 129 KO-assigned proteins did not map to any pathways, and 166 pathways were shared with human pathways. For *BaO*, 213 proteins lacked KO annotation, 137 KO-assigned proteins had no associated pathways, and 152 pathways were common to human pathways. In the case of *FlaP*, 32 proteins lacked KO identifiers, 87 KO-assigned proteins showed no pathway mapping, and 140 pathways were shared with human pathways. After eliminating the shared pathways, the analysis identified 72 unique bacterial pathways in *BaF*, 66 in *BaU*, 62 in *BaO*, and 68 in *FlaP*, and their corresponding unique bacterial pathway proteins are 152 in *BaF*, 153 in *BaU*, 155 in *BaO*, and 115 in *FlaP*. These pathway proteins were subsequently subjected to subcellular localization analysis (S7–S10 Tables). Prediction of subcellular localization offered valuable insights into the functional characteristics and therapeutic potential of these proteins.

Subcellular localization analysis of the pathway proteins revealed that in *BaF*, 111 were cytoplasmic, 21 cytoplasmic membrane-associated, 2 periplasmic, and 18 proteins with unknown localization. In *BaU*, 118 were cytoplasmic, 20 cytoplasmic membrane-associated, 1 periplasmic, 1 extracellular, and 13 proteins of unknown localization. For *BaO*, 119 were cytoplasmic, 22 cytoplasmic membrane-associated, 3 periplasmic, and 11 of unknown localization. Finally, *FlaP* exhibited 90 cytoplasmic proteins, 18 cytoplasmic membrane-associated, and 7 proteins of unknown localization (S11 Table).

As cytoplasmic proteins are frequently used as therapeutic targets [106], only proteins located in the cytoplasm were selected for subsequent analyses. Protein-protein interaction networks were generated for these cytoplasmic proteins in each bacterium using the STRING platform with medium interaction confidence (0.4). The resulting networks were visualized in Cytoscape, and the top five hub proteins for each bacterium were identified based on Degree topological measure using the cytoHubba plugin in Cytoscape (Fig 5). The top 5 hub proteins from each of the four bacteria yielded 20 proteins in total, from which duplicate entries appearing across multiple species were removed, resulting in a set of unique proteins. After removing overlapping proteins and proteins with unavailable structure information, a final set of 10 proteins (*ribD*, *ribBA*, *murA*, *alr*, *hisI*, *hisE*, *hisD*, *hisG*, *hisH*, and *hisB*) were retained as bKGs for subsequent analyses (S12 Table). These bKGs are conserved essential bacterial genes and are identified here as putative antibacterial targets, since they are essential for survival, have low similarity to human homologs [107,108].

3.6. Druggability analysis of bKGs

The druggability of each target protein was evaluated by identifying potential ligand-binding pockets and ranking them according to their predicted suitability for small-molecule binding. Multiple binding pockets were detected for all selected proteins. For each pocket, structural and physicochemical properties, including pocket volume (\AA^3), surface area (\AA^2), drug score, and simple score, were calculated. Pockets with a drug score ≥ 0.5 were considered to have a greater likelihood of being modulated by drug-like molecules. All ten bKGs had at least one pocket with drug scores ≥ 0.5 , confirming their predicted druggability and suitability as targets for small-molecule binding. The druggability parameters for the selected pockets are (see S3–S12 Figs for pocket visualization) compiled in Table 3.

3.7. Molecular docking

Molecular docking analyses were performed to assess the interaction strength between the prioritized target proteins and the curated ligand set. 3D-structures of the 10 selected bKGs (*ribD*, *ribBA*, *murA*, *alr*, *hisI*, *hisE*, *hisD*, *hisG*, *hisH*, and *hisB*) were obtained from the AlphaFold Protein Structure Database and SWISS-MODEL, and these structural models

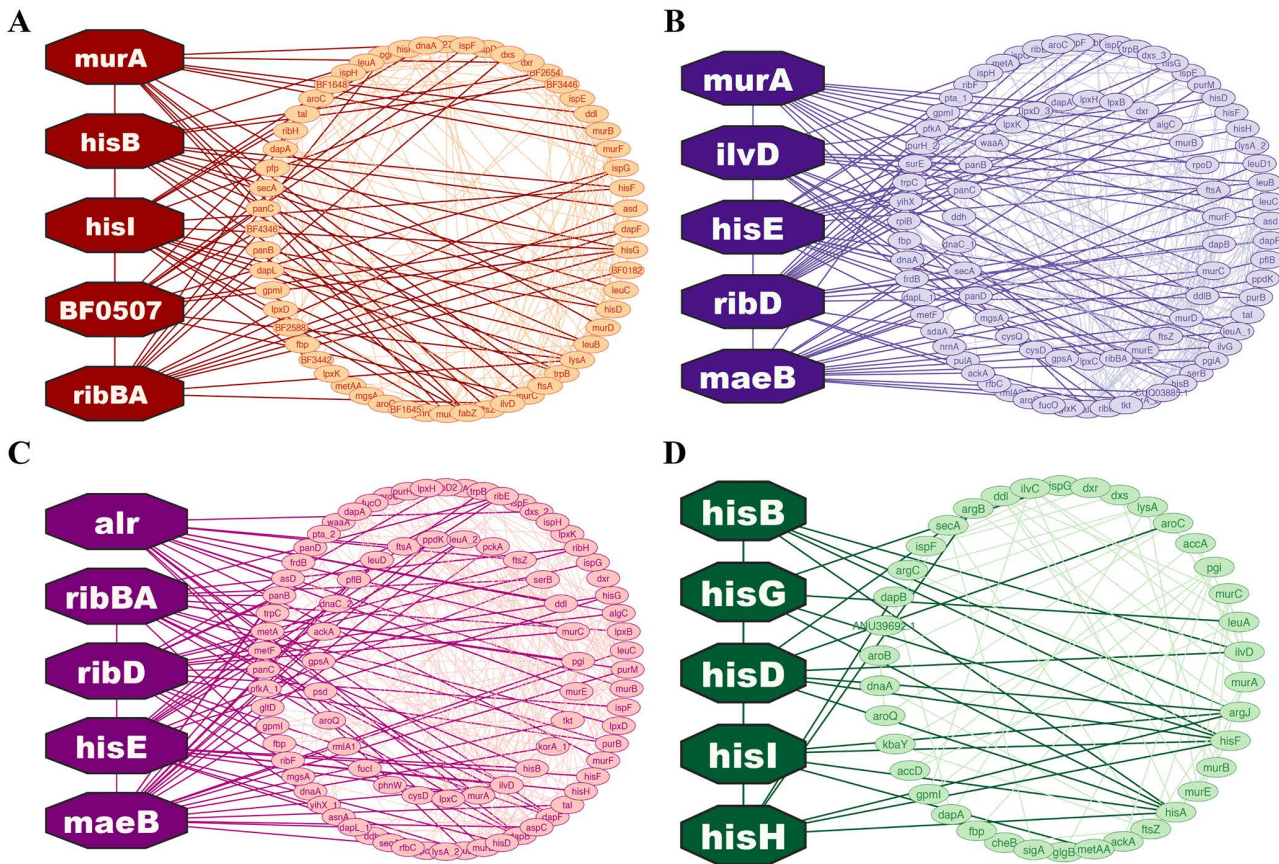


Fig 5. Protein-protein interaction networks depicting the top five hub proteins for (A) Bacteroides fragilis, (B) Bacteroides uniformis, (C) Bacteroides ovatus, and (D) Flavonifractor plautii.

<https://doi.org/10.1371/journal.pone.0343565.g005>

Table 3. Druggability information of bKGs (No. of binding pocket, highest binding pocket, pocket volume, surface area, drug score and simple scores).

Gene	Uniprot_id	No. of binding pockets	Highest binding pocket	Volume (Å ³)	Surface (Å ²)	Drug score	Simple score
<i>ribD</i>	A0A5M5DBX0	7	P_0	1232.19	1694.05	0.81	0.62
<i>ribBA</i>	Q5LHT1	11	P_0	2460.77	3128.46	0.8	0.63
<i>murA</i>	A0ABC9NCX7	12	P_0	1458.94	1408.86	0.81	0.61
<i>alr</i>	A0A5M5DT22	35	P_0	2701.34	2771.96	0.81	0.62
<i>hisI</i>	Q5LBD8	5	P_1	378.82	739.81	0.74	0.21
<i>hisE</i>	A0ABC9N566	6	P_0	670.08	1029.33	0.81	0.49
<i>hisD</i>	A0A096CN98	10	P_0	1978.71	2316.44	0.81	0.63
<i>hisG</i>	A0A096BB53	10	P_0	719.23	930.74	0.81	0.52
<i>hisH</i>	A0A096BA31	6	P_0	253.63	474.70	0.65	0.13
<i>hisB</i>	A0A096DFI2	7	P_1	203.97	305.98	0.54	0

<https://doi.org/10.1371/journal.pone.0343565.t003>

were subsequently used for molecular docking. A total of 121 candidate compounds associated with CRC were compiled from the DrugBank ([S13 Table](#)). Both protein and ligand structures were prepared according to the procedure outlined in the Methods section. Molecular docking was then executed for all protein-ligand combinations, generating BASs (kcal/mol). The 25 highest-scoring interactions were visualized in a matrix heatmap. Based on these results, the top 10 ligands Tucatinib, Sonidegib, Belumosudil, Regorafenib, Antrafenine, Sorafenib, Sulfasalazine, Estramustine, Aminoglutethimide, and Tipiracil-highlighted in green in [Fig 6](#), were selected as the most promising preliminary drug candidates for in-depth investigation against the proposed bKGs. The identified candidates represent computational predictions derived from a multi-step inference pipeline; experimental validation is therefore essential before any therapeutic relevance can be established.

3.8. Drug likeness and ADME/T analysis

Drug-likeness of the 10 lead drug candidates showing the strongest predicted average binding affinity toward the selected targets was evaluated using Lipinski's Rule of Five (ROF). This guideline considers molecular properties such as molecular weight (<500 Da), lipophilicity ($\log P \leq 5$), and limits on hydrogen-bond donors (≤ 5) and acceptors (≤ 10). Among 10 lead candidates, Belumosudil, Sulfasalazine, Aminoglutethimide, and Tipiracil are the promising preliminary candidates that follow all the drug-like characteristics without permitting any single violation. A comprehensive summary of the drug likeness properties for each compound is provided in [Table 4](#).

Evaluation of the ADME and toxicity profiles highlighted Sulfasalazine, Aminoglutethimide, and Tipiracil as the most promising preliminary therapeutic candidates based on computational predictions. All three compounds exhibited high human intestinal absorption (>90%), indicating strong predicted oral bioavailability and efficient absorption. Their Blood-Brain Barrier permeability was limited ($BBB < 0.3$) and not penetrable into the CNS, which is desirable for minimizing

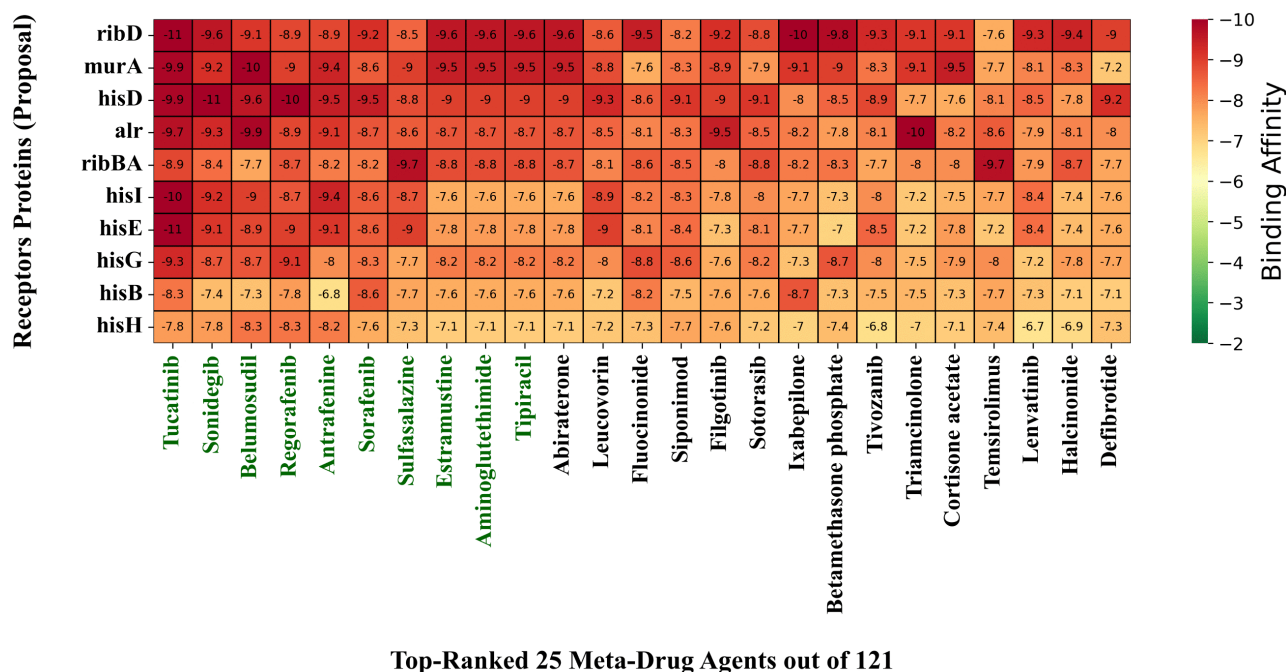


Fig 6. Heatmap representation of binding affinity scores between bacterial key gene (Y-axis) and candidate therapeutic compounds (X-axis, top 25 ranked). The matrix illustrates interaction strengths for 10 prioritized protein targets against 25 screened molecules. Green shading identifies the 10 compounds with superior binding affinity scores.

<https://doi.org/10.1371/journal.pone.0343565.g006>

Table 4. Drug likeness properties of top 10 drug candidates with highest binding affinities.

Compounds	Molecular Weight (<500)	LogP (≤ 5)	H-bond Donors (≤ 5)	H-bond Acceptors (≤ 10)	Polar Surface area (\AA^2)	Lipinski's Rule	
						Follow	Violation
TUCATINIB	480.532	5.094	2	10	110.85	3	1
SONIDEGIB	485.506	5.822	1	6	63.69	3	1
BELUMOSUDIL	452.518	4.82	3	8	104.82	4	0
REGORAFENIB	482.821	5.689	3	7	92.35	3	1
ANTRAFENINE	588.552	6.995	1	6	57.7	3	1
SORAFENIB	464.831	5.55	3	7	92.35	3	1
SULFASALAZINE	398.4	3.702	3	9	141.31	4	0
ESTRAMUSTINE	440.411	5.182	1	4	49.77	3	1
AMINOGLUTETHIMIDE	232.283	1.353	3	4	72.19	4	0
TIPIRACIL	242.666	0.29	3	6	92.81	4	0

<https://doi.org/10.1371/journal.pone.0343565.t004>

potential neurotoxic effects. Toxicity assessment further supported the overall predicted safety profiles of these compounds. Overall, the combination of computationally predicted absorption characteristics, constrained BBB entry, and acceptable toxicity profiles underscores their potential as preliminary drug candidates, requiring experimental validation for establishing therapeutic relevance. A comprehensive summary of the pharmacokinetic and toxicity profiles for each compound is provided in [S14 Table](#).

3.9. Validation of docking protocol using decoy molecules

To validate the molecular docking protocol, the BASs of the selected drug compounds were compared against the average BASs of corresponding decoy molecules for each target protein ([S15 Table](#)). In all cases, the selected compounds demonstrated superior binding affinity scores compared to the decoy molecules ([S16 Table](#)), confirming that the observed protein-ligand interactions were not random. These results support the reliability of the docking protocol and confirm that the selected compounds were distinguishable from inactive decoy molecules based on their binding affinity scores, validating the specificity of the docking results.

3.10. Binding site assessment of final selected complexes

The docking poses of the final selected complexes were cross-referenced with the predicted pocket residues of the proteins. For *ribBA*, the docked ligand occupied the top predicted pocket, with all interacting residues confirmed as members of the predicted pocket. For *ribD*, DoGSiteScorer predicted 3 pockets with comparable drug scores (0.81, 0.8, and 0.82), indicating multiple potential binding pockets across the surface. The docked ligand occupied second ranked pocket with all interacting residues fully contained within the predicted cavity ([S17 Table](#)). This reflects blind docking independently binding within the computationally predicted druggable pockets of the target proteins, providing a computational justification of the identified docking sites.

3.11. Molecular dynamics simulation

MD simulation analysis of the three selected protein-ligand complexes (*ribBA_sulfasalazine*, *ribD_aminoglutethimide*, and *ribD_tipiracil*) with superior docking scores, satisfaction of Lipinski Rule of Five criteria, and favorable ADME/T profiles, was performed to assess conformational stability and binding consistency of each compound throughout the 100 ns trajectory. The evaluation focused on key metrics including RMSD, RMSF, and MM-PBSA binding free energy. RMSD indicates the average atomic displacement over time, reflecting complex stability; as depicted in [Fig 7](#). With

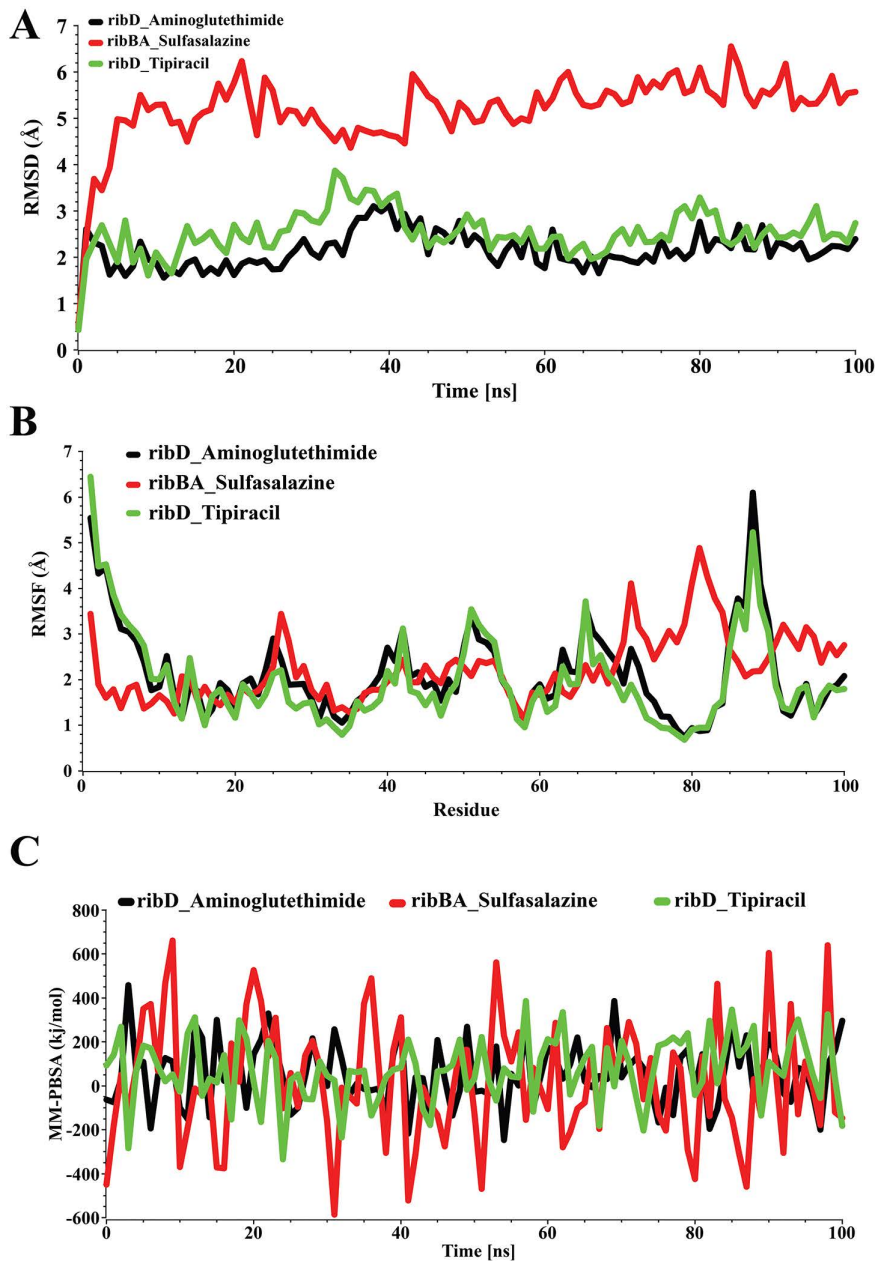


Fig 7. MD simulation results: (A) the root means square deviation (RMSD), (B) the root mean square fluctuation (RMSF), and (C) the binding free energy (MM-PBSA) of the top-ranked drug-target complexes with 3 candidate drugs for a duration of 100-ns simulation.

<https://doi.org/10.1371/journal.pone.0343565.g007>

average RMSD values of 2.145 Å, and 2.518 Å *ribD_aminoglutethimide* and *ribD_tipiracil* exhibited better stability than *ribBA_sulfasalazine* (average RMSD 5.189 Å). Here, *ribD_aminoglutethimide* exhibited the least fluctuation, suggesting greater stability (Fig 7A). RMSF analysis revealed that *ribD*-Tipiracil exhibited the lowest residue flexibility with an average value of 1.980 Å, followed by *ribD*-Aminoglutethimide (2.140 Å) and *ribBA*-Sulfasalazine (2.234 Å), indicating more consistent and stable interactions (Fig 7B). MM-PBSA binding free energy, revealed good stability with average scores of 56.24 kJ/mol (*ribD_aminoglutethimide*), 18.67 kJ/mol (*ribBA_sulfasalazine*), and 74.21 kJ/mol (*ribD_tipiracil*).

*ribD*_aminoglutethimide demonstrated good binding, followed by *ribD*_tipiracil, with *ribBA*_sulfasalazine showing comparatively weaker interaction (Fig 7C).

4. Discussion

The gut microbiome has been increasingly associated with colorectal carcinogenesis, with studies suggesting its potential involvement in inflammatory processes and cellular signaling pathways [109]. Throughout cancer development, a complex interplay has been observed among the gut microbial community, the tumor associated microbiota, and the host immune system [110,111]. CRC pathogenesis represents a multifactorial process associated with inherited genetic alterations, lifestyle-related risk factors, and compositional shifts in gut microbiota. Accumulating evidence suggests that gut microbial alterations may be associated with treatment resistance and tumor evolution, emphasizing the growing need to develop microbiome-guided strategies for cancer management [20,112]. Therapeutic agents that act on human proteins and those that target microbial proteins provide two complementary avenues of treatment. Modulating host proteins may influence internal biological pathways, whereas targeting bacterial proteins may help in reshaping gut microbial activity, which may affect metabolism and disease outcomes [113–115]. This study utilized gut microbiome derived 16S rRNA sequencing data to identify bKGs as putative antibacterial targets within CRC-associated bacterial taxa and to screen candidate drug compounds through computational drug repurposing. Our analysis revealed significant differences in the overall microbial profile between CRC patients and the HC group. The observed phylum-level alterations in CRC specifically, increased Fusobacteria (previously associated with colorectal carcinogenesis) abundance contrasted with reduced Actinobacteria (known to produce SCFA and possesses anti-inflammatory properties) represent compositional shifts reported in CRC-associated microbiome studies [116]. At the family level, high enrichment of the families Bacteroidaceae, Rikenellaceae, and Acidaminococcaceae in CRC [91,117,118] and high enrichment of Bifidobacteriaceae, Lactobacillaceae, and Lachnospiraceae [119–121] in HC group corroborates findings with other studies. Differential abundance analysis uncovered 42 bacterial species that were significantly altered in CRC compared to HC individuals. Bacterial species including *Bacteroides fragilis*, *Flavonifractor plautii*, *Bacteroides uniformis*, and *Bacteroides ovatus* were significantly enriched in CRC group, while taxa including *Romboutsia*, *Bifidobacterium*, and *Blautia* remained significantly depleted. The depletion of *Romboutsia*, *Bifidobacterium*, and *Blautia* consistent with emerging literature on gut microbiota compositional alterations associated with CRC [119,122,123]. The elevated abundance of *Bacteroides* species observed in CRC patients compared to HC subjects is well documented in several microbiome studies [124,125]. Enterotoxigenic *Bacteroides fragilis* has emerged as a potential diagnostic biomarker for CRC and has been associated with unfavorable prognosis result [126]. *Bacteroides fragilis* has been reported to promote tumor development through various mechanisms, including modulation of NF- κ B signaling, induction of DNA damage, enhancement of polyamine metabolism, stimulation of TH17 immune responses, and activation of stem cell functions [127]. *Flavonifractor plautii*, a major flavonoid-degrading bacterium, is known for its elevated flavonoid degradation, potentially reducing the beneficial effects and bioavailability of flavonoids in CRC. Moreover, its association with the catechol cleavage pathway which processes catechols produced during flavonoid breakdown further highlights its role in gut flavonoid metabolism [88]. *Bacteroides uniformis* and *Bacteroides ovatus* were enriched in CRC in some studies [84,101] and have been associated with bacteremia which has been linked to increased CRC risk [85], though they have a report of being present as gut commensal or as member of healthy gut [128,129]. Among the 42 altered bacteria, four bacteria were selected (*Bacteroides fragilis*, *Bacteroides ovatus*, *Bacteroides uniformis*, and *Flavonifractor plautii*) for further investigation. Although *Bacteroides uniformis* and *Bacteroides ovatus* are regarded as commensals, their association with CRC has been reported in several previous studies [84–86,90–93,97,101–103], and their selection is further supported by evidence that commensal bacteria may exhibit opportunistic pathogenic behavior depending on host immune status and environmental conditions, and may enhance the virulence of co-occurring pathogens in mixed infections [104,105]. Analysis of the top four CRC-associated bacteria identified 10 conserved essential bKGs (*ribD*, *ribBA*, *murA*, *alr*, *hisI*, *hisE*, *hisD*, *hisG*, *hisH*, and *hisB*) within these taxa.

These bKGs are fundamental for bacterial survival, mediating the biosynthesis of riboflavin, histidine, and cell wall component, and may serve as putative antibacterial targets, given their essentiality, low similarity to human homologs [130]. Bacterial essential genes involved in peptidoglycan biosynthesis and riboflavin metabolism have been reported to indirectly influence host immune signaling through the release of bacterial cell wall fragments and metabolic intermediates that interact with host [131,132]. Essential genes involved in histidine biosynthesis have also been reported to influence host-microbiome crosstalk [133,134], which may indicate their indirect role in disease development. The genes *ribD* and *ribBA* play critical roles in bacterial riboflavin biosynthesis, providing the precursor required for FMN and FAD cofactors that support redox processes and core metabolic functions. Their inhibition may offer a potential strategy to compromise the metabolic fitness and growth of CRC-enriched bacterial taxa [135–137]. The *his* genes collectively encode enzymes responsible for histidine biosynthesis, a highly regulated metabolic pathway in bacterial systems, disruption of these genes has been associated with histidine auxotrophy and loss of bacterial virulence supporting their candidacy as putative antibacterial targets within the CRC-associated bacterial taxa identified in this study [138,139]. The genes *murA* and *alr* encode key enzymes involved in peptidoglycan biosynthesis, rendering them attractive targets for antibacterial intervention [140–142]. Collectively, these 10 genes represent conserved essential metabolic targets without human homologs which may have potential influence on host-microbes interaction, making them potentially suitable candidates for selective antibacterial intervention against CRC-associated bacterial taxa. To identify therapeutic compounds capable of targeting the prioritized bKGs, a comprehensive in-silico drug screening analysis prioritized three compounds (Sulfasalazine, Aminoglutethimide, and Tipiracil) as antibacterial agents based on their strong predicted binding affinity against the identified bKGs, compliance with Lipinski Rule of Five drug-likeness criteria and favorable pharmacokinetic (limited BBB penetration, limited CNS penetration, high intestinal absorption) and toxicity profiles (non-mutagenic) (see S14 Table). MD simulation analysis further supported the structural stability of the selected complexes. Sulfasalazine is an anti-bacterial agent that acts on colonic bacterial azoreductase enzyme and is metabolized to the anti-inflammatory metabolite and anti-bacterial metabolites sulfapyridine [143]. Sulfasalazine is also a well-established anti-inflammatory agent, exerts anti-tumor effects by promoting apoptosis and tumor shrinkage through blockade of the plasma membrane-associated system xc⁻ cystine transporter [144]. Sulfasalazine demonstrated strong predicted binding affinity with *murA*, *ribBA*, and *hisE* may be suggesting potential interference with peptidoglycan and riboflavin biosynthesis in CRC-associated bacterial taxa. Aminoglutethimide and Tipiracil demonstrated favorable predicted binding affinity with *ribD*, *murA*, and *hisD*, suggesting potential influence on riboflavin and peptidoglycan biosynthesis pathways and bacterial growth. Notably, Tipiracil along with Trifluridine has been approved for use in metastatic CRC patients that have failed or are unsuitable for standard chemotherapeutic and biologic treatments as well as for patients with unresectable advanced or recurrent CRC [145], further supporting the translational relevance of these computationally identified candidates. Collectively, these multi-layered analyses provide preliminary support for exploring microbiome-guided therapeutic strategies relevant to CRC and identifying a set of putative microbial gene-drug interactions as candidates for further investigation. These findings represent computational predictions that warrant experimental validation to confirm their antibacterial efficacy, functional relevance, and therapeutic potential in the context of CRC-associated gut microbiota.

5. Strengths and limitations

One of the advantages of this study is the use of a comprehensive computational pipeline that integrates 16S rRNA-based microbial profiling, statistical analysis, subtractive genomics, protein-protein interaction network analysis, molecular docking and ADME/T analysis to explore bKGs as putative antibacterial targets in CRC-associated bacterial taxa and prioritize candidate therapeutics. Together, these approaches offer a systematic and integrative framework for identifying potential therapeutic targets. Despite these strengths, several limitations should be acknowledged. All conclusions are drawn from in-silico analyses and therefore require confirmation through laboratory experiments including in vitro and in vivo validation. This study focused exclusively on the bacterial targets and did not incorporate host molecular pathways. The

workflow involves multiple sequential computational inferences which may introduce cumulative uncertainty. Taxonomic assignments from V4 16S amplicon sequencing are probabilistic rather than definitive, and the reference proteomes used represent functional approximations rather than actual patient-derived strains. The drug compounds were sourced from DrugBank based on CRC association for repurposing investigation and may not represent the most appropriate library for bacterial enzyme inhibition. Molecular docking results reflect predicted binding affinities based on computational models, and evidence of drug efficacy needs experimental confirmation. Additionally, blind docking approach was employed with no benchmarking against known inhibitors due to the absence of experimentally resolved structures of the target proteins, and docking was performed using a single engine (AutoDock Vina) without replication across alternative platforms. Furthermore, gut-targeted antibacterial parameters including intestinal luminal stability, bacterial membrane uptake, and activity under anaerobic conditions could not be included in the ADME/T analysis, due to computational limitations. Finally, the dataset comprises a relatively modest single cohort without available clinical metadata, and the broader biological influence of the identified bKGs on host metabolic processes remains to be established through functional and longitudinal studies.

6. Conclusions

This study employed an integrative in silico framework comprising 16S rRNA-based microbial profiling, subtractive genomics, PPI network analysis, molecular docking and ADME/T analysis to investigate gut microbiota compositional alteration associated with CRC and to identify bKGs in CRC-associated gut microbiota as potential targets for antibacterial therapy. Analysis of 16S rRNA sequencing data revealed significant alterations in bacterial community composition between CRC and HC groups. Differential abundance analysis identified several CRC-associated bacterial taxa, including *Bacteroides fragilis*, *Bacteroides uniformis*, *Bacteroides ovatus*, and *Flavonifractor plautii* which have been previously associated with CRC in observational studies. Using a subtractive genomics approach and PPI network analysis, ten essential bKGs (*ribD*, *ribBA*, *murA*, *alr*, *hisI*, *hisE*, *hisD*, *hisG*, *hisH* and *hisB*) were identified from these CRC-associated bacterial taxa as putative antibacterial targets as these proteins are of conserved essential bacterial metabolic proteins. Subsequent computational drug screening identified ten candidate compounds, of which Sulfasalazine, Aminoglutethimide, and Tipiracil exhibited favorable in silico pharmacokinetic and toxicity profiles that support their potential suitability for repurposing as antibacterial agents in CRC. Finally, this study proposes a microbiome-guided strategy that integrates bKGs identification with computational drug repurposing to explore antibacterial therapeutic options relevant to CRC-associated microbiome. Targeting CRC-associated altered bacteria and their essential bKGs may offer a promising complementary treatment avenue. However, experimental validation and clinical studies are necessary to confirm the efficacy, safety and translational potential of the identified drug candidates while future functional and patient-specific microbiome analyses may further support personalized CRC interventions.

Supporting information

S1 Fig. Phylum level differential analysis between HC and CRC group.

(TIF)

S2 Fig. Differentially Abundant Family ($|\log_2FC| > 1$ and $\text{adj.p-value} < 0.05$).

(TIF)

S3 Fig. *ribD* protein (pocket with highest drug score).

(TIF)

S4 Fig. *ribBA* protein (pocket with highest drug score).

(TIF)

S5 Fig. *murA* protein (pocket with highest drug score).
(TIF)

S6 Fig. *alr* protein (pocket with highest drug score).
(TIF)

S7 Fig. *hisI* protein (pocket with highest drug score).
(TIF)

S8 Fig. *hisE* protein (pocket with highest drug score).
(TIF)

S9 Fig. *hisH* protein (pocket with highest drug score).
(TIF)

S10 Fig. *hisG* protein (pocket with highest drug score).
(TIF)

S11 Fig. *hisD* protein (pocket with highest drug score).
(TIF)

S12 Fig. *hisB* protein (pocket with highest drug score).
(TIF)

S1 Table. Percentage of Each phylum.
(XLSX)

S2 Table. Percentage of each family's contribution to the overall microbiome.
(XLSX)

S3 Table. Percentage of each genera's contribution to the overall microbiome.
(XLSX)

S4 Table. Differentially abundant microbial family ($|\log_2FC| > 1$ and $\text{adj.p-value} < 0.05$).
(XLSX)

S5 Table. Differentially abundant microbial genera ($|\log_2FC| > 1$ and $\text{adj.p-value} < 0.05$).
(XLSX)

S6 Table. 42 differentially abundant bacteria ($|\log_2FC| > 1$ and $\text{adj.p-value} < 0.05$).
(XLSX)

S7 Table. Unique pathogenic pathway proteins of *BaF*.
(XLSX)

S8 Table. Unique pathogenic pathway proteins of *BaU*.
(XLSX)

S9 Table. Unique pathogenic pathway proteins of *BaO*.
(XLSX)

S10 Table. Unique pathogenic pathway proteins of *FlaP*.
(XLSX)

S11 Table. Proteins of 4 bacteria with different locations.

(XLSX)

S12 Table. Key-genes selection using degree method.

(XLSX)

S13 Table. Drug collection details.

(XLSX)

S14 Table. Pharmacokinetic and toxicity profiles of the top 10 drugs with highest binding affinity.

(DOCX)

S15 Table. The identifiers of the Decoy molecules generated for SULFASALZINE, AMINOGLUTETHIMIDE, and TIPIRACIL.

(DOCX)

S16 Table. The average binding affinity scores (BASs) of the decoy molecules and the BASs of suggested drug candidates in kcal/mol with the receptors.

(DOCX)

S17 Table. The list of ligand binding residues and pockets residues.

(DOCX)

Author contributions

Conceptualization: Nibas Kumar Pal, Md. Kaderi Kibria, Md. Nurul Haque Mollah.

Data curation: Nibas Kumar Pal, Md. Kaderi Kibria, Dipto Kundu.

Formal analysis: Nibas Kumar Pal, Md. Feroj Ahmed.

Investigation: Nibas Kumar Pal, Md. Foyzal Ahmed, Md. Nurul Haque Mollah.

Methodology: Nibas Kumar Pal, Md. Kaderi Kibria, Md. Shariful Islam, Md. Nurul Haque Mollah.

Project administration: Md. Nurul Haque Mollah.

Resources: Nibas Kumar Pal, Md. Kaderi Kibria, Tasfia Noor, Md. Nurul Haque Mollah.

Software: Nibas Kumar Pal, Mohammad Ali, Md. Feroj Ahmed, Md. Shariful Islam, Ovi Sharma, Md. Al Noman, Md. Foyzal Ahmed.

Supervision: Md. Nurul Haque Mollah.

Validation: Nibas Kumar Pal, Md. Shariful Islam, Md. Nurul Haque Mollah.

Visualization: Nibas Kumar Pal, Md. Abdul Latif, Tasfia Noor.

Writing – original draft: Nibas Kumar Pal.

Writing – review & editing: Md. Nurul Haque Mollah.

References

1. Cancer That Develops in the Colon (the Longest Part of the Large Intestine) and/or the Rectum (the Last Several Inches of the Large Intestine before the Anus). 2011. <https://www.cancer.gov/publications/dictionaries/cancer-terms/def/colorectal-cancer>
2. Colon cancer - symptoms and causes. Mayo Clin. 2025.
3. Xi Y, Xu P. Global colorectal cancer burden in 2020 and projections to 2040. Transl Oncol. 2021;14. <https://doi.org/10.1016/j.tranon.2021.101174>

4. Siegel RL, Wagle NS, Cercek A, Smith RA, Jemal A. Colorectal cancer statistics, 2023. *CA Cancer J Clin.* 2023;73:233–54. <https://doi.org/10.3322/caac.21772>
5. Senthakumaran T, Tannæs TM, Moen AEF, Brackmann SA, Jahanlu D, Rounge TB, et al. Detection of colorectal-cancer-associated bacterial taxa in fecal samples using next-generation sequencing and 19 newly established qPCR assays. *Mol Oncol.* 2025;19(2):412–29. <https://doi.org/10.1002/1878-0261.13700> PMID: 38970464
6. Fadlallah H, El Masri J, Fakhereddine H, Youssef J, Chemaly C, Doughan S, et al. Colorectal cancer: recent advances in management and treatment. *World J Clin Oncol.* 2024;15(9):1136–56. <https://doi.org/10.5306/wjco.v15.i9.1136> PMID: 39351451
7. Choi S, Chung J, Cho M-L, Park D, Choi SS. Analysis of changes in microbiome compositions related to the prognosis of colorectal cancer patients based on tissue-derived 16S rRNA sequences. *J Transl Med.* 2021;19(1):485. <https://doi.org/10.1186/s12967-021-03154-0> PMID: 34844611
8. Wong CC, Yu J. Gut microbiota in colorectal cancer development and therapy. *Nat Rev Clin Oncol.* 2023;20(7):429–52. <https://doi.org/10.1038/s41571-023-00766-x> PMID: 37169888
9. Sánchez-Alcoholado L, Ramos-Molina B, Otero A, Laborda-Illanes A, Ordóñez R, Medina JA, et al. The role of the gut microbiome in colorectal cancer development and therapy response. *Cancers (Basel).* 2020;12(6):1406. <https://doi.org/10.3390/cancers12061406> PMID: 32486066
10. Yurtseven B, Aydemir E, Ayaz F. The role of intestinal microbiota and immune system interactions in autoimmune diseases. *Immunotargets Ther.* 2025;14:1347–72. <https://doi.org/10.2147/ITT.S569016> PMID: 41334562
11. Koliarakis I, Lagkouvardos I, Vogiatzoglou K, Tsamandouras I, Intze E, Messaritakis I, et al. Circulating bacterial DNA in colorectal cancer patients: the potential role of *Fusobacterium nucleatum*. *Int J Mol Sci.* 2024;25(16):9025. <https://doi.org/10.3390/ijms25169025> PMID: 39201711
12. Abreu MT, Peek RM Jr. Gastrointestinal malignancy and the microbiome. *Gastroenterology.* 2014;146(6):1534–1546.e3. <https://doi.org/10.1053/j.gastro.2014.01.001> PMID: 24406471
13. Osman M-A, Neoh H-M, Ab Mutalib N-S, Chin S-F, Jamal R. 16S rRNA gene sequencing for deciphering the colorectal cancer gut microbiome: current protocols and workflows. *Front Microbiol.* 2018;9:767. <https://doi.org/10.3389/fmicb.2018.00767> PMID: 29755427
14. Xiao L, Liu S, Wu Y, Huang Y, Tao S, Liu Y, et al. The interactions between host genome and gut microbiome increase the risk of psychiatric disorders: mendelian randomization and biological annotation. *Brain Behav Immun.* 2023;113:389–400. <https://doi.org/10.1016/j.bbi.2023.08.003> PMID: 37557965
15. The interaction between gut microbiota and host DNA methylation in the pathogenesis and therapy of inflammatory bowel disease.
16. Gao G, Zhao X, Li Q, He C, Zhao W, Liu S, et al. Genome and metagenome analyses reveal adaptive evolution of the host and interaction with the gut microbiota in the goose. *Sci Rep.* 2016;6:32961. <https://doi.org/10.1038/srep32961> PMID: 27608918
17. Levy M, Thaiss CA, Elinav E. Metagenomic cross-talk: the regulatory interplay between immunogenomics and the microbiome. *Genome Med.* 2015;7:120. <https://doi.org/10.1186/s13073-015-0249-9>
18. Sun Y, Gan Z, Liu S, Zhang S, Zhong W, Liu J, et al. Metagenomic and transcriptomic analysis reveals crosstalk between intratumor mycobiome and hosts in early-stage nonsmoking lung adenocarcinoma patients. *Thorac Cancer.* 2025;16(2):e15527. <https://doi.org/10.1111/1759-7714.15527> PMID: 39853685
19. Feng Q, Liang S, Jia H, Stadlmayr A, Tang L, Lan Z, et al. Gut microbiome development along the colorectal adenoma-carcinoma sequence. *Nat Commun.* 2015;6:6528. <https://doi.org/10.1038/ncomms7528> PMID: 25758642
20. Yu J, Feng Q, Wong SH, Zhang D, Liang QY, Qin Y, et al. Metagenomic analysis of faecal microbiome as a tool towards targeted non-invasive biomarkers for colorectal cancer. *Gut.* 2017;66(1):70–8. <https://doi.org/10.1136/gutjnl-2015-309800> PMID: 26408641
21. Ahn J, Sinha R, Pei Z, Dominianni C, Wu J, Shi J, et al. Human gut microbiome and risk for colorectal cancer. *J Natl Cancer Inst.* 2013;105(24):1907–11. <https://doi.org/10.1093/jnci/djt300> PMID: 24316595
22. Zeller G, Tap J, Voigt AY, Sunagawa S, Kultima JR, Costea PI, et al. Potential of fecal microbiota for early-stage detection of colorectal cancer. *Mol Syst Biol.* 2014;10(11):766. <https://doi.org/10.15252/msb.20145645> PMID: 25432777
23. Thomas AM, Manghi P, Asnicar F, Pasolli E, Armanini F, Zolfo M, et al. Metagenomic analysis of colorectal cancer datasets identifies cross-cohort microbial diagnostic signatures and a link with choline degradation. *Nat Med.* 2019;25(4):667–78. <https://doi.org/10.1038/s41591-019-0405-7> PMID: 30936548
24. Fujita K, Kubota Y, Ishida H, Sasaki Y. Irinotecan, a key chemotherapeutic drug for metastatic colorectal cancer. *World J Gastroenterol.* 2015;21(43):12234–48. <https://doi.org/10.3748/wjg.v21.i43.12234> PMID: 26604633
25. Mahmud S, Ajadee A, Sarker A, Ahmmmed R, Noor T, Pappu MAA, et al. Exploring common genomic biomarkers to disclose common drugs for the treatment of colorectal cancer and hepatocellular carcinoma with type-2 diabetes through transcriptomics analysis. *PLoS One.* 2025;20(3):e0319028. <https://doi.org/10.1371/journal.pone.0319028> PMID: 40127075
26. Gamage BD, Ranasinghe D, Sahankumari A, Malavige GN. Metagenomic analysis of colonic tissue and stool microbiome in patients with colorectal cancer in a South Asian population. *BMC Cancer.* 2024;24(1):1124. <https://doi.org/10.1186/s12885-024-12885-4> PMID: 39256724
27. Babraham Bioinformatics - FastQC A Quality Control Tool for High Throughput Sequence Data. 2025.
28. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet C, Al-Ghali GA, et al. QIIME 2: reproducible, interactive, scalable, and extensible microbiome data science. *PeerJ Preprints.* 2018.

29. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* 2011;17:10. <https://doi.org/10.14806/ej.17.1.200>
30. 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(7):581–3. <https://doi.org/10.1038/nmeth.3869> PMID: 27214047
31. Ning Y, Yang G, Chen Y, Zhao X, Qian H, Liu Y, et al. Characteristics of the urinary microbiome from patients with gout: a prospective study. *Front Endocrinol (Lausanne).* 2020;11:272. <https://doi.org/10.3389/fendo.2020.00272> PMID: 32508748
32. Feng Z-H, Li Q, Liu S-R, Du X-N, Wang C, Nie X-H, et al. Comparison of composition and diversity of bacterial microbiome in human upper and lower respiratory tract. *Chin Med J (Engl).* 2017;130(9):1122–4. <https://doi.org/10.4103/0366-6999.204934> PMID: 28469109
33. Leung PHM, Subramanya R, Mou Q, Lee KT-W, Islam F, Gopalan V, et al. Characterization of mucosa-associated microbiota in matched cancer and non-neoplastic mucosa from patients with colorectal cancer. *Front Microbiol.* 2019;10:1317. <https://doi.org/10.3389/fmicb.2019.01317> PMID: 31244818
34. Kibria MK, Ali MA, Yaseen M, Khan IA, Bhat MA, Islam MA, et al. Discovery of bacterial key genes from 16S rRNA-Seq profiles that are associated with the complications of SARS-CoV-2 infections and provide therapeutic indications. *Pharmaceuticals (Basel).* 2024;17(4):432. <https://doi.org/10.3390/ph17040432> PMID: 38675393
35. McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One.* 2013;8(4):e61217. <https://doi.org/10.1371/journal.pone.0061217> PMID: 23630581
36. Roswell M, Dushoff J, Winfree R. A conceptual guide to measuring species diversity. *Oikos.* 2021;130(3):321–38. <https://doi.org/10.1111/oik.07202>
37. Fan L, Wu W, Qiu L, Song C, Meng S, Zheng Y, et al. Methanogenic community compositions in surface sediment of freshwater aquaculture ponds and the influencing factors. *Antonie Van Leeuwenhoek.* 2018;111(1):115–24. <https://doi.org/10.1007/s10482-017-0932-5> PMID: 28840355
38. Kim B, Lee S, Lee Y-J, Kang Y-M, Rhee MH, Kwak D, et al. Preliminary insights into the gut microbiota of Captive Tigers in Republic of Korea: influence of geographic and individual variation. *Microorganisms.* 2025;13(6):1427. <https://doi.org/10.3390/microorganisms13061427> PMID: 40572315
39. Legendre P, Borcard D, Peres-Neto PR. Analyzing beta diversity: partitioning the spatial variation of community composition data. *Ecol Monogr.* 2005;75:435–50. <https://doi.org/10.1890/05-0549>
40. McCafferty J, Mühlbauer M, Gharabeh RZ, Arthur JC, Perez-Chanona E, Sha W, et al. Stochastic changes over time and not founder effects drive cage effects in microbial community assembly in a mouse model. *ISME J.* 2013;7(11):2116–25. <https://doi.org/10.1038/ismej.2013.106> PMID: 23823492
41. Jannel L, Guilhaumon F, Valade P, Chabanet P, Borie G, Grondin H, et al. eDNA metabarcoding, a promising tool for monitoring aquatic biodiversity in the estuaries of Reunion Island (South-West Indian Ocean). *Environmental DNA.* 2024;6(6). <https://doi.org/10.1002/edn3.70044>
42. Chen J, Bittinger K, Charlson ES, Hoffmann C, Lewis J, Wu GD, et al. Associating microbiome composition with environmental covariates using generalized UniFrac distances. *Bioinformatics.* 2012;28(16):2106–13. <https://doi.org/10.1093/bioinformatics/bts342> PMID: 22711789
43. Molano L-AG, Vega-Abellana S, Manichanh C. GSR-DB: a manually curated and optimized taxonomical database for 16S rRNA amplicon analysis. *mSystems.* 2024;9(2):e0095023. <https://doi.org/10.1128/msystems.00950-23> PMID: 38189256
44. Bokulich NA, Kaehler BD, Rideout JR, Dillon M, Bolyen E, Knight R, et al. Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. *Microbiome.* 2018;6(1):90. <https://doi.org/10.1186/s40168-018-0470-z> PMID: 29773078
45. Ahmad A, Yang W, Chen G, Shafiq M, Javed S, Ali Zaidi SS, et al. Analysis of gut microbiota of obese individuals with type 2 diabetes and healthy individuals. *PLoS One.* 2019;14(12):e0226372. <https://doi.org/10.1371/journal.pone.0226372> PMID: 31891582
46. Paulson JN, Stine OC, Bravo HC, Pop M. Differential abundance analysis for microbial marker-gene surveys. *Nat Methods.* 2013;10(12):1200–2. <https://doi.org/10.1038/nmeth.2658> PMID: 24076764
47. Paulson J. MetagenomeSeq: statistical analysis for sparse high-throughput sequencing. *BioconductorJp.* 2014:1–20.
48. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B.* 1995;57:289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
49. Boutet E, Lieberherr D, Tognolli M, Schneider M, Bansal P, Bridge AJ, et al. UniProtKB/swiss-prot, the manually annotated section of the UniProt knowledgebase: how to use the entry view. *Methods Mol Biol.* 2016;1374:23–54. https://doi.org/10.1007/978-1-4939-3167-5_2 PMID: 26519399
50. Huang Y, Niu B, Gao Y, Fu L, Li W. CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics.* 2010;26(5):680–2. <https://doi.org/10.1093/bioinformatics/btq003> PMID: 20053844
51. P D. Important databases and tools to identify promising drug targets by subtractive genomics approach – a review. *Int J Res Eng Technol.* 2015;4(6):453–5. <https://doi.org/10.15623/ijret.2015.0406077>
52. Luo H, Lin Y, Gao F, Zhang C-T, Zhang R. DEG 10, an update of the database of essential genes that includes both protein-coding genes and noncoding genomic elements. *Nucleic Acids Res.* 2014;42:D574–80. <https://doi.org/10.1093/nar/gkt1131> PMID: 24243843
53. Hossain T, Kamruzzaman M, Choudhury TZ, Mahmood HN, Nabi AHMN, Hosen MI. Application of the Subtractive genomics and molecular docking analysis for the identification of novel putative drug targets against Salmonella Enterica Subsp. Enterica Serovar Poona. *Biomed Res Int.* 2017;2017:3783714. <https://doi.org/10.1155/2017/3783714>

54. Watson AK, Lannes R, Pathmanathan JS, Méheust R, Karkar S, Colson P, et al. The Methodology Behind Network Thinking: Graphs to Analyze Microbial Complexity and Evolution. In: Anisimova M, editor. *Evolutionary Genomics: Statistical and Computational Methods*. New York, NY: Springer New York; 2019. pp. 271–308.
55. Pearson WR. An introduction to sequence similarity (“homology”) searching. *Curr Protoc Bioinformatics*. 2013;Chapter 3:3.1.1–3.1.8. <https://doi.org/10.1002/0471250953.bi0301s42> PMID: [23749753](https://pubmed.ncbi.nlm.nih.gov/23749753/)
56. Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, et al. EggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Research*. 2019;47(D309–D314). <https://doi.org/10.1093/nar/gky1085>
57. Nakai K, Horton P. PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localization. *Trends Biochem Sci*. 1999;24(1):34–6. [https://doi.org/10.1016/s0968-0004\(98\)01336-x](https://doi.org/10.1016/s0968-0004(98)01336-x) PMID: [10087920](https://pubmed.ncbi.nlm.nih.gov/10087920/)
58. Gardy JL, Spencer C, Wang K, Ester M, Tusnády GE, Simon I, et al. PSORT-B: Improving protein subcellular localization prediction for Gram-negative bacteria. *Nucleic Acids Res*. 2003;31(13):3613–7. <https://doi.org/10.1093/nar/gkg602> PMID: [12824378](https://pubmed.ncbi.nlm.nih.gov/12824378/)
59. Yu NY, Wagner JR, Laird MR, Melli G, Rey S, Lo R, et al. PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics*. 2010;26(12):1608–15. <https://doi.org/10.1093/bioinformatics/btq249>
60. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res*. 2019;47(D1):D607–13. <https://doi.org/10.1093/nar/gky1131> PMID: [30476243](https://pubmed.ncbi.nlm.nih.gov/30476243/)
61. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003;13(11):2498–504. <https://doi.org/10.1101/gr.1239303> PMID: [14597658](https://pubmed.ncbi.nlm.nih.gov/14597658/)
62. Zhang S, Xiang X, Liu L, Yang H, Cen D, Tang G. Bioinformatics analysis of hub genes and potential therapeutic agents associated with gastric cancer. *Cancer Manag Res*. 2021;13:8929–51. <https://doi.org/10.2147/CMAR.S341485> PMID: [34876855](https://pubmed.ncbi.nlm.nih.gov/34876855/)
63. Volkamer A, Kuhn D, Rippmann F, Rarey M. DoGSiteScorer: a web server for automatic binding site prediction, analysis and druggability assessment. *Bioinformatics*. 2012;28(10):2074–5. <https://doi.org/10.1093/bioinformatics/bts310>
64. Alzyoud L, Bryce RA, Al Sorkhy M, Atatreh N, Ghattas MA. Structure-based assessment and druggability classification of protein-protein interaction sites. *Sci Rep*. 2022;12(1):7975. <https://doi.org/10.1038/s41598-022-12105-8> PMID: [35562538](https://pubmed.ncbi.nlm.nih.gov/35562538/)
65. Anonymous. DrugBank - Open data drug and drug target database. 2013.
66. Pushpakom S, Iorio F, Eyers PA, Escott KJ, Hopper S, Wells A, et al. Drug repurposing: progress, challenges and recommendations. *Nat Rev Drug Discov*. 2019;18:41–58. <https://doi.org/10.1038/nrd.2018.168>
67. Maier L, Pruteanu M, Kuhn M, Zeller G, Telzerow A, Anderson EE, et al. Extensive impact of non-antibiotic drugs on human gut bacteria. *Nature*. 2018;555(7698):623–8. <https://doi.org/10.1038/nature25979> PMID: [29555994](https://pubmed.ncbi.nlm.nih.gov/29555994/)
68. Kim S, Chen J, Cheng T, Gindulyte A, He J, He S, et al. PubChem in 2021: new data content and improved web interfaces. *Nucleic Acids Res*. 2021;49:D1388–95. <https://doi.org/10.1093/nar/gkaa971>
69. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. *Nature*. 2021;596(7873):583–9. <https://doi.org/10.1038/s41586-021-03819-2> PMID: [34265844](https://pubmed.ncbi.nlm.nih.gov/34265844/)
70. Schwede T, Kopp J, Guex N, Peitsch MC. SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Res*. 2003;31:3381–5. <https://doi.org/10.1093/nar/gkg520>
71. Hanwell MD, Curtis DE, Lonie DC, Vandermeersch T, Zurek E, Hutchison GR. Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *J Cheminform*. 2012;4:17. <https://doi.org/10.1186/1758-2946-4-17>
72. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, et al. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J Comput Chem*. 2009;30(16):2785–91. <https://doi.org/10.1002/jcc.21256> PMID: [19399780](https://pubmed.ncbi.nlm.nih.gov/19399780/)
73. Ahmed MF, Al Noman M, Ahmed MF, Latif MA, Pappu MAA, Islam MS, et al. In-Silico discovery of Pediatric Acute-Myeloid-Leukemia (pAML) causing druggable molecular signatures highlighting their pathogenetic processes and therapeutic agents through single-cell RNA-Seq profile analysis. *PLoS One*. 2025;20(10):e0335410. <https://doi.org/10.1371/journal.pone.0335410> PMID: [41171828](https://pubmed.ncbi.nlm.nih.gov/41171828/)
74. Islam MS, Mollah MMH, Ahsan MA, Ali M, Al Noman M, Ahmed MF, et al. In-silico identification of genetic variants associated with chronic lymphocytic leukemia for diagnostic and therapeutic applications. *Sci Rep*. 2026;16(1):14545. <https://doi.org/10.1038/s41598-026-45456-7> PMID: [41866418](https://pubmed.ncbi.nlm.nih.gov/41866418/)
75. Eberhardt J, Santos-Martins D, Tillack AF, Forli S. AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings. *J Chem Inf Model*. 2021;61:3891–8. <https://doi.org/10.1021/acs.jcim.1c00203>
76. Myung Y, de Sá AGC, Ascher DB. Deep-PK: deep learning for small molecule pharmacokinetic and toxicity prediction. *Nucleic Acids Res*. 2024;52(W1):W469–75. <https://doi.org/10.1093/nar/gkae254> PMID: [38634808](https://pubmed.ncbi.nlm.nih.gov/38634808/)
77. Pires DEV, Blundell TL, Ascher DB. pkCSM: predicting small-molecule pharmacokinetic and toxicity properties using graph-based signatures. *J Med Chem*. 2015;58(9):4066–72. <https://doi.org/10.1021/acs.jmedchem.5b00104> PMID: [25860834](https://pubmed.ncbi.nlm.nih.gov/25860834/)
78. Lipinski CA. Lead- and drug-like compounds: The rule-of-five revolution. *Drug Discov Today Technol*. 2004;1:337–41. <https://doi.org/10.1016/j.ddtec.2004.11.007>

79. Stein RM, Yang Y, Balius TE, O'Meara MJ, Lyu J, Young J, et al. Property-unmatched decoys in docking benchmarks. *J Chem Inf Model*. 2021;61(2):699–714. <https://doi.org/10.1021/acs.jcim.0c00598> PMID: [33494610](https://pubmed.ncbi.nlm.nih.gov/33494610/)
80. Huang S-Y, Zou X. An iterative knowledge-based scoring function to predict protein-ligand interactions: I. Derivation of interaction potentials. *J Comput Chem*. 2006;27(15):1866–75. <https://doi.org/10.1002/jcc.20504> PMID: [16983673](https://pubmed.ncbi.nlm.nih.gov/16983673/)
81. Ozvoldik K, Stockner T, Krieger E. YASARA model-interactive molecular modeling from two dimensions to virtual realities. *J Chem Inf Model*. 2023;63(20):6177–82. <https://doi.org/10.1021/acs.jcim.3c01136> PMID: [37782001](https://pubmed.ncbi.nlm.nih.gov/37782001/)
82. Krieger E, Koraimann G, Vriend G. Increasing the precision of comparative models with YASARA NOVA—a self-parameterizing force field. *Proteins*. 2002;47(3):393–402. <https://doi.org/10.1002/prot.10104> PMID: [11948792](https://pubmed.ncbi.nlm.nih.gov/11948792/)
83. Mitra S, Dash R. Structural dynamics and quantum mechanical aspects of shikonin derivatives as CREBBP bromodomain inhibitors. *J Mol Graph Model*. 2018;83:42–52. <https://doi.org/10.1016/j.jmgm.2018.04.014> PMID: [29758466](https://pubmed.ncbi.nlm.nih.gov/29758466/)
84. Liang L, Kong C, Li J, Liu G, Wei J, Wang G, et al. Distinct microbes, metabolites, and the host genome define the multi-omics profiles in right-sided and left-sided colon cancer. *Microbiome*. 2024;12(1):274. <https://doi.org/10.1186/s40168-024-01987-7> PMID: [39731152](https://pubmed.ncbi.nlm.nih.gov/39731152/)
85. Justesen US, Nielsen SL, Jensen TG, Dessau RB, Møller JK, Coia JE, et al. Bacteremia with anaerobic bacteria and association with colorectal cancer: a population-based cohort study. *Clin Infect Dis*. 2022;75(10):1747–53. <https://doi.org/10.1093/cid/ciac259> PMID: [35380653](https://pubmed.ncbi.nlm.nih.gov/35380653/)
86. Chen H, Jiao J, Wei M, Jiang X, Yang R, Yu X, et al. Metagenomic analysis of the interaction between the gut microbiota and colorectal cancer: a paired-sample study based on the GMrepo database. *Gut Pathog*. 2022;14(1):48. <https://doi.org/10.1186/s13099-022-00527-8> PMID: [36564826](https://pubmed.ncbi.nlm.nih.gov/36564826/)
87. Iadsee N, Chuaypen N, Techawiwattanaboon T, Jinato T, Patcharatrakul T, Malakorn S, et al. Identification of a novel gut microbiota signature associated with colorectal cancer in Thai population. *Sci Rep*. 2023;13(1):6702. <https://doi.org/10.1038/s41598-023-33794-9> PMID: [37095272](https://pubmed.ncbi.nlm.nih.gov/37095272/)
88. Gupta A, Dhakan DB, Maji A, Saxena R, P K VP, Mahajan S, et al. Association of flavonifractor plautii, a flavonoid-degrading bacterium, with the gut microbiome of colorectal cancer patients in India. *mSystems*. 2019;4(6):e00438–19. <https://doi.org/10.1128/mSystems.00438-19> PMID: [31719139](https://pubmed.ncbi.nlm.nih.gov/31719139/)
89. Yang Y, Du L, Shi D, Kong C, Liu J, Liu G, et al. Dysbiosis of human gut microbiome in young-onset colorectal cancer. *Nat Commun*. 2021;12(1):6757. <https://doi.org/10.1038/s41467-021-27112-y> PMID: [34799562](https://pubmed.ncbi.nlm.nih.gov/34799562/)
90. Ma Y, Qu R, Zhang Y, Jiang C, Zhang Z, Fu W. Progress in the study of colorectal cancer caused by altered gut microbiota after cholecystectomy. *Front Endocrinol (Lausanne)*. 2022;13:815999. <https://doi.org/10.3389/fendo.2022.815999> PMID: [35282463](https://pubmed.ncbi.nlm.nih.gov/35282463/)
91. He T, Cheng X, Xing C. The gut microbial diversity of colon cancer patients and the clinical significance. *Bioengineered*. 2021;12(1):7046–60. <https://doi.org/10.1080/21655979.2021.1972077> PMID: [34551683](https://pubmed.ncbi.nlm.nih.gov/34551683/)
92. Li H, Sheng D, Jin C, Zhao G, Zhang L. Identifying and ranking causal microbial biomarkers for colorectal cancer at different cancer subsites and stages: a Mendelian randomization study. *Front Oncol*. 2023;13:1224705. <https://doi.org/10.3389/fonc.2023.1224705> PMID: [37538123](https://pubmed.ncbi.nlm.nih.gov/37538123/)
93. Yuan Y, Chen Y, Yao F, Zeng M, Xie Q, Shafiq M, et al. Microbiomes and resistomes in biopsy tissue and intestinal lavage fluid of colorectal cancer. *Front Cell Dev Biol*. 2021;9:736994. <https://doi.org/10.3389/fcell.2021.736994> PMID: [34604238](https://pubmed.ncbi.nlm.nih.gov/34604238/)
94. Ding X, Ting NL-N, Wong CC, Huang P, Jiang L, Liu C, et al. Bacteroides fragilis promotes chemoresistance in colorectal cancer, and its elimination by phage VA7 restores chemosensitivity. *Cell Host Microbe*. 2025;33(6):941–956.e10. <https://doi.org/10.1016/j.chom.2025.05.004> PMID: [40446807](https://pubmed.ncbi.nlm.nih.gov/40446807/)
95. Wang T, Cai G, Qiu Y, Fei N, Zhang M, Pang X, et al. Structural segregation of gut microbiota between colorectal cancer patients and healthy volunteers. *ISME J*. 2012;6(2):320–9. <https://doi.org/10.1038/ismej.2011.109> PMID: [21850056](https://pubmed.ncbi.nlm.nih.gov/21850056/)
96. Png CW, Chua YK, Law JH, Zhang Y, Tan KK. Alterations in co-abundant bacteriome in colorectal cancer and its persistence after surgery: a pilot study. *Scientific Reports*. 2022;12:9829. <https://doi.org/10.1038/s41598-022-14203-z>
97. Ruiz-Malagón AJ, Rodríguez-Sojo MJ, Redondo E, Rodríguez-Cabezas ME, Gálvez J, Rodríguez-Nogales A. Systematic review: the gut microbiota as a link between colorectal cancer and obesity. *Obes Rev*. 2025;26(4):e13872. <https://doi.org/10.1111/obr.13872> PMID: [39614602](https://pubmed.ncbi.nlm.nih.gov/39614602/)
98. Pateriya D, Malwe AS, Sharma VK. CRCpred: an AI-ML tool for colorectal cancer prediction using gut microbiome. *Comput Biol Med*. 2025;195:110592. <https://doi.org/10.1016/j.compbiomed.2025.110592> PMID: [40570762](https://pubmed.ncbi.nlm.nih.gov/40570762/)
99. Cheng Y, Ling Z, Li L. The intestinal microbiota and colorectal cancer. *Front Immunol*. 2020;11:1–13. <https://doi.org/10.3389/fimmu.2020.615056>
100. Murovec B, Deutsch L, Stres B. Predictive modeling of colorectal cancer using exhaustive analysis of microbiome information layers available from public metagenomic data. *Front Microbiol*. 2024;15:1426407. <https://doi.org/10.3389/fmicb.2024.1426407> PMID: [39252839](https://pubmed.ncbi.nlm.nih.gov/39252839/)
101. Zhou P, Dai Z, Xie Y, Li T, Xu Z, Huang Y, et al. Differences in tissue-associated bacteria between metastatic and non-metastatic colorectal cancer. *Front Microbiol*. 2023;14:1133607. <https://doi.org/10.3389/fmicb.2023.1133607> PMID: [37362927](https://pubmed.ncbi.nlm.nih.gov/37362927/)
102. Chénard T, Malick M, Dubé J, Massé E. The influence of blood on the human gut microbiome. *BMC Microbiol*. 2020;20(1):44. <https://doi.org/10.1186/s12866-020-01724-8> PMID: [32126968](https://pubmed.ncbi.nlm.nih.gov/32126968/)
103. Liang X, Li H, Tian G, Li S. Dynamic microbe and molecule networks in a mouse model of colitis-associated colorectal cancer. *Sci Rep*. 2014;4:4985. <https://doi.org/10.1038/srep04985> PMID: [24828543](https://pubmed.ncbi.nlm.nih.gov/24828543/)
104. Stacy A, Fleming D, Lamont RJ, Rumbaugh KP, Whiteley M. A Commensal bacterium promotes virulence of an opportunistic pathogen via cross-respiration. *mBio*. 2016;7(3):e00782–16. <https://doi.org/10.1128/mBio.00782-16> PMID: [27353758](https://pubmed.ncbi.nlm.nih.gov/27353758/)

105. Tewari N, Dey P. Navigating commensal dysbiosis: Gastrointestinal host-pathogen interplay orchestrating opportunistic infections. *Microbiol Res.* 2024;286:127832. <https://doi.org/10.1016/j.micres.2024.127832> PMID: [39013300](https://pubmed.ncbi.nlm.nih.gov/39013300/)
106. Ahammad I, Jamal TB, Lamisa AB, Bhattacharjee A, Zinan N, Hasan Chowdhury MZ, et al. Subtractive genomics study of *Xanthomonas oryzae* pv. *Oryzae* reveals repurposable drug candidate for the treatment of bacterial leaf blight in rice. *J Genet Eng Biotechnol.* 2024;22(1):100353. <https://doi.org/10.1016/j.jgeb.2024.100353> PMID: [38494267](https://pubmed.ncbi.nlm.nih.gov/38494267/)
107. Nazarshodeh E, Marashi S-A, Gharaghani S. Structural systems pharmacology: A framework for integrating metabolic network and structure-based virtual screening for drug discovery against bacteria. *PLoS One.* 2021;16(12):e0261267. <https://doi.org/10.1371/journal.pone.0261267> PMID: [34905555](https://pubmed.ncbi.nlm.nih.gov/34905555/)
108. Sakharkar KR, Sakharkar MK, Chow VTK. Biocomputational strategies for microbial drug target identification. In: Champney WS, editor. *New antibiotic targets.* Totowa, NJ: Humana Press; 2008. pp. 1–9.
109. Rebersek M. Gut microbiome and its role in colorectal cancer. *BMC Cancer.* 2021;21:1325. <https://doi.org/10.1186/s12885-021-09054-2>
110. Inamura K. Colorectal cancers: an update on their molecular pathology. *Cancers.* 2018;10(26).
111. Scott AJ, Alexander JL, Merrifield CA, Cunningham D, Jobin C, Brown R, et al. International Cancer Microbiome Consortium consensus statement on the role of the human microbiome in carcinogenesis. *Gut.* 2019;68(9):1624–32. <https://doi.org/10.1136/gutjnl-2019-318556> PMID: [31092590](https://pubmed.ncbi.nlm.nih.gov/31092590/)
112. Gao R, Gao Z, Huang L, Qin H. Gut microbiota and colorectal cancer. *Eur J Clin Microbiol.* 2017;36(5):757–69. <https://doi.org/10.1007/s10096-016-2881-8>
113. Wu G, Zhao N, Zhao L. Microbial-host isozyme: A novel target in “drug the bug” strategies for diabetes. *Cell Metab.* 2023;35:1677–9. <https://doi.org/10.1016/j.cmet.2023.09.008>
114. Kanwal A, Kanwar N, Bharati S, Srivastava P, Singh SP, Amar S. Exploring new drug targets for type 2 diabetes: success, challenges and opportunities. *Biomedicines.* 2022;10(2):331. <https://doi.org/10.3390/biomedicines10020331> PMID: [35203540](https://pubmed.ncbi.nlm.nih.gov/35203540/)
115. Vatsha P, Goswami A, Mondal A, Mondal A, Chakra B, Ratha SK, et al. Evolving novel drug targets for type 2 diabetes: a mechanistic approach towards-success, challenges and opportunities. *Cancer Clin J.* 2023;3:1017.
116. Illescas O, Rodríguez-Sosa M, Gariboldi M. Mediterranean diet to prevent the development of Colon diseases: a meta-analysis of gut microbiota studies. *Nutrients.* 2021;13(7):2234. <https://doi.org/10.3390/nu13072234> PMID: [34209683](https://pubmed.ncbi.nlm.nih.gov/34209683/)
117. Park J, Kim N-E, Yoon H, Shin CM, Kim N, Lee DH, et al. Fecal microbiota and gut microbe-derived extracellular vesicles in colorectal cancer. *Front Oncol.* 2021;11:650026. <https://doi.org/10.3389/fonc.2021.650026> PMID: [34595105](https://pubmed.ncbi.nlm.nih.gov/34595105/)
118. Hoang T, Kim MJ, Park JW, Jeong S-Y, Lee J, Shin A. Nutrition-wide association study of microbiome diversity and composition in colorectal cancer patients. *BMC Cancer.* 2022;22(1):656. <https://doi.org/10.1186/s12885-022-09735-6> PMID: [35701733](https://pubmed.ncbi.nlm.nih.gov/35701733/)
119. Chen W, Liu F, Ling Z, Tong X, Xiang C. Human intestinal lumen and mucosa-associated microbiota in patients with colorectal cancer. *PLoS One.* 2012;7(6):e39743. <https://doi.org/10.1371/journal.pone.0039743> PMID: [22761885](https://pubmed.ncbi.nlm.nih.gov/22761885/)
120. Mutignani M, Penagini R, Gargari G, Guglielmetti S, Cintollo M, Airolidi A, et al. Blood bacterial DNA, intestinal adenoma and colorectal cancer. *medRxiv.* 2021.
121. Youssef O, Lahti L, Kokkola A, Karla T, Tikkanen M, Ehsan H, et al. Stool microbiota composition differs in patients with stomach, colon, and rectal neoplasms. *Dig Dis Sci.* 2018;63(11):2950–8. <https://doi.org/10.1007/s10620-018-5190-5> PMID: [29995183](https://pubmed.ncbi.nlm.nih.gov/29995183/)
122. Mangifesta M, Mancabelli L, Milani C, Gaiani F, de'Angelis N, de'Angelis GL, et al. Mucosal microbiota of intestinal polyps reveals putative biomarkers of colorectal cancer. *Sci Rep.* 2018;8(1):13974. <https://doi.org/10.1038/s41598-018-32413-2> PMID: [30228361](https://pubmed.ncbi.nlm.nih.gov/30228361/)
123. Kim MJ, Song M-H, Ji Y-S, Park JW, Shin Y-K, Kim S-C, et al. Cell free supernatants of *Bifidobacterium adolescentis* and *Bifidobacterium longum* suppress the tumor growth in colorectal cancer organoid model. *Sci Rep.* 2025;15(1):935. <https://doi.org/10.1038/s41598-024-83048-5> PMID: [39762302](https://pubmed.ncbi.nlm.nih.gov/39762302/)
124. Bonnet M, Buc E, Sauvanet P, Darcha C, Dubois D, Pereira B, et al. Colonization of the human gut by *E. coli* and colorectal cancer risk. *Clin Cancer Res.* 2014;20(4):859–67. <https://doi.org/10.1158/1078-0432.CCR-13-1343> PMID: [24334760](https://pubmed.ncbi.nlm.nih.gov/24334760/)
125. Wong SH, Yu J. Gut microbiota in colorectal cancer: mechanisms of action and clinical applications. *Nat Rev Gastroenterol Hepatol.* 2019;16(11):690–704. <https://doi.org/10.1038/s41575-019-0209-8> PMID: [31554963](https://pubmed.ncbi.nlm.nih.gov/31554963/)
126. Haghi F, Goli E, Mirzaei B, Zeighami H. The association between fecal enterotoxigenic *B. Fragilis* with Colorectal Cancer. *BMC Cancer.* 2019;19:879. <https://doi.org/10.1186/s12885-019-6115-1>
127. Liu QQ, Li CM, Fu LN, Wang HL, Tan J, Wang YQ, et al. Enterotoxigenic *Bacteroides fragilis* induces the stemness in colorectal cancer via upregulating histone demethylase JMJD2B. *Gut Microbes.* 2020;12:1788900. <https://doi.org/10.1080/19490976.2020.1788900>
128. Wexler AG, Goodman AL. An insider's perspective: *Bacteroides* as a window into the microbiome. *Nat Microbiol.* 2017;2:17026. <https://doi.org/10.1038/nmicrobiol.2017.26> PMID: [28440278](https://pubmed.ncbi.nlm.nih.gov/28440278/)
129. Fultz R, Ticer T, Ihekweazu FD, Horvath TD, Haidacher SJ, Hoch KM, et al. Unraveling the metabolic requirements of the gut commensal *Bacteroides ovatus*. *Front Microbiol.* 2021;12:745469. <https://doi.org/10.3389/fmicb.2021.745469> PMID: [34899632](https://pubmed.ncbi.nlm.nih.gov/34899632/)
130. Haag NL, Velk KK, Wu C. Potential antibacterial targets in bacterial central metabolism. *Int J Adv Life Sci.* 2012;4(1–2):21–32. PMID: [24151543](https://pubmed.ncbi.nlm.nih.gov/24151543/)

131. Solopova A, Bottacini F, Venturi Degli Esposti E, Amaretti A, Raimondi S, Rossi M, et al. Riboflavin biosynthesis and overproduction by a derivative of the human gut commensal *Bifidobacterium longum* subsp. *infantis* ATCC 15697. *Front Microbiol.* 2020;11:573335. <https://doi.org/10.3389/fmicb.2020.573335> PMID: 33042083
132. Orsini Delgado ML, Gamas Magalhaes J, Morra R, Cultrone A. Muropeptides and muropeptide transporters impact on host immune response. *Gut Microbes.* 2024;16(1):2418412. <https://doi.org/10.1080/19490976.2024.2418412> PMID: 39439228
133. Xu H, Pan LB, Yu H, Han P, Fu J, Zhang ZW, et al. Gut microbiota-derived metabolites in inflammatory diseases based on targeted metabolomics. *Front Pharmacol.* 2022;13.
134. Dwivedy A, Ashraf A, Jha B, Kumar D, Agarwal N, Biswal BK. De novo histidine biosynthesis protects *Mycobacterium tuberculosis* from host IFN- γ mediated histidine starvation. *Commun Biol.* 2021;4:410. <https://doi.org/10.1038/s42003-021-01926-4>
135. Fuentes Flores A, Sepúlveda Cisternas I, Vásquez Solís de Ovando JI, Torres A, García-Angulo VA. Contribution of riboflavin supply pathways to *Vibrio cholerae* in different environments. *Gut Pathog.* 2017;9:64. <https://doi.org/10.1186/s13099-017-0214-9> PMID: 29163672
136. Chengalroyen MD, Mehaffy C, Lucas M, Bauer N, Raphela ML, Oketade N, et al. Modulation of riboflavin biosynthesis and utilization in mycobacteria. *Microbiol Spectr.* 2024;12(8):e0320723. <https://doi.org/10.1128/spectrum.03207-23> PMID: 38916330
137. Gnanagobal H, Cao T, Hossain A, Vasquez I, Chakraborty S, Chukwu-Osazuwa J, et al. Role of riboflavin biosynthesis gene duplication and transporter in *Aeromonas salmonicida* virulence in marine teleost fish. *Virulence.* 2023;14(1):2187025. <https://doi.org/10.1080/21505594.2023.2187025> PMID: 36895132
138. Dwivedy A, Ashraf A, Jha B, Kumar D, Agarwal N, Biswal BK. De novo histidine biosynthesis protects *Mycobacterium tuberculosis* from host IFN- γ mediated histidine starvation. *Commun Biol.* 2021;4. <https://doi.org/10.1038/s42003-021-01926-4>
139. Del Duca S, Semenzato G, Esposito A, Liò P, Fani R. The operon as a conundrum of gene dynamics and biochemical constraints: what we have learned from histidine biosynthesis. *Genes (Basel).* 2023;14(4):949. <https://doi.org/10.3390/genes14040949> PMID: 37107707
140. Garbinski LD, Rosen BP, Yoshinaga M. Organoarsenicals inhibit bacterial peptidoglycan biosynthesis by targeting the essential enzyme MurA. *Chemosphere.* 2020;254:126911. <https://doi.org/10.1016/j.chemosphere.2020.126911> PMID: 32957300
141. Zhou J, Cai Y, Liu Y, An H, Deng K, Ashraf MA, et al. Breaking down the cell wall: still an attractive antibacterial strategy. *Front Microbiol.* 2022;13:952633. <https://doi.org/10.3389/fmicb.2022.952633> PMID: 36212892
142. Liu Y, Breukink E. The membrane steps of bacterial cell wall synthesis as antibiotic targets. *Antibiotics (Basel).* 2016;5(3):28. <https://doi.org/10.3390/antibiotics5030028> PMID: 27571111
143. Ben Mrid R, Bouchmaa N, Ainani H, El Fatimy R, Malka G, Mazini L. Anti-rheumatoid drugs advancements: new insights into the molecular treatment of rheumatoid arthritis. *Biomed Pharmacother.* 2022;151:113126. <https://doi.org/10.1016/j.biopha.2022.113126>
144. Ma M-Z, Chen G, Wang P, Lu W-H, Zhu C-F, Song M, et al. Xc⁻ inhibitor sulfasalazine sensitizes colorectal cancer to cisplatin by a GSH-dependent mechanism. *Cancer Lett.* 2015;368(1):88–96. <https://doi.org/10.1016/j.canlet.2015.07.031> PMID: 26254540
145. Burness CB, Duggan ST. Trifluridine/tipiracil: a review in metastatic colorectal cancer. *Drugs.* 2016;76(14):1393–402. <https://doi.org/10.1007/s40265-016-0633-9> PMID: 27568360