ARTICLE

A network pharmacology approach to identify potential prognostic colorectal cancer biomarkers associated with gut microbiome metabolites

Trigger Jon B. Balitang[1], Carl James L. Padua[1], Mer-Hamid Ali D. Abuat[1], Jodelyn B. Melegrito[1], and Ma. Easter Joy V. Sajo*[2]

[1]Department of Biology, College of Science, University of the Philippines Baguio, Baguio City, Benguet, 2600 Philippines
[2]Philippine Genome Center, University of the Philippines, Diliman, Quezon City, 1101 Philippines

ABSTRACT

Colorectal cancer (CRC) is more challenging to treat as it progresses. Its progression may be influenced by various gut microbiome metabolites (GMMs). To expand treatment options for CRC, GMM-associated genes may serve as biomarkers for CRC prognosis. Using network pharmacology, this study aimed to identify GMM- and CRC-associated genes with prognostic value. Genes and GMMs were curated from online databases. KEGG pathway enrichment and protein-protein interaction (PPI) network analysis revealed that these genes are associated with CRC-related cellular signaling mechanisms, namely, the PI3K-Akt signaling pathway, among others. Survival analysis of TCGA-COAD and TCGA-READ patient cohorts showed five genes, including *CXCL8*, *HDAC2*, *PTGS2*, *MET*, and *PGR* as genes that may predict prognosis. Phenylpyruvic acid and ursodeoxycholic acid 3-sulfate showed the most favorable binding affinities for their respective proteins, CXCL8 and HDAC2. Overall, this study suggests that the GMM-associated genes CXCL8 and HDAC2 may play roles in CRC development and progression.

INTRODUCTION

Colorectal cancer (CRC) was the second leading cause of cancer-related death worldwide in 2020 (Xi and Xu 2021). Advanced-stage CRC is typically more difficult to treat, resulting in a poor prognosis. To mitigate poor survival outcomes, the gut microbiome and its metabolites (GMMs) are being explored as biomarkers for CRC prognosis and early detection (Zhang et al. 2021a, Avuthu and Guda 2022). The gut microbiome constitutes an ecosystem of diverse microorganisms that produce metabolites such as amino acids, amino acid by-products, lipids, modified bile acids, carnitine, and choline metabolites (Yang et al. 2019, Mardinoglu et al. 2015, Hashim et al. 2019). These metabolites are known to influence proteins and biological pathways critical to CRC progression.

The rising incidence of CRC, particularly among young adults, is often linked to Western dietary patterns (e.g., high fat and sugar content, processed foods), exposure to industrial pollution, or sedentary lifestyle (Stoffel and Murphy 2020). These environmental factors can lead to gut microbiome dysbiosis and altered metabolite levels, which can in turn influence the risk and development of CRC (Zhang et al. 2021a). Dysbiosis can promote inflammation in the colorectum and alter metabolic processes (Candela et al. 2014). For example, diets that enrich sulfur-metabolizing bacterial communities, and consequently, hydrogen sulfide levels, can increase the risk of developing CRC (Nguyen et al. 2020).

While traditional *in vitro* and *in vivo* methods for studying the gut microbiome and identifying CRC biomarkers are time-consuming, costly, and constrained by experimental limitations, computational techniques, such as network pharmacology, can address these issues. Network pharmacology is an *in silico* method used in drug discovery pipelines to screen chemical compounds for bioactivity through their interactions with target proteins and biological pathways (Moshkov et al. 2023). Large gene expression datasets generated from numerous wet-lab experiments can be analyzed through bioinformatics methodologies.

In this study, we employed a purely computational approach, using network pharmacology to screen potential prognostic biomarkers. This study identifies the biological pathways and mechanisms of CRC-associated protein-coding genes linked to these GMMs, the

Corresponding author
Email Address: **mvsajo@up.edu.ph**
Date received: 17 November 2025
Date revised: 19 February 2026
Date accepted: 01 March 2026
DOI: https://doi.org/10.54645/2026191GIZ-73

KEYWORDS

colorectal cancer, gut microbiome metabolites, network pharmacology, survival analysis

prognostic genes that significantly affect the survival rate of patients with CRC, and GMMs with strong binding affinities toward the proteins encoded by these prognostic genes. By investigating GMM-associated genes as potential prognostic biomarkers for CRC, this study aims to expand targeted treatment options. Ultimately, developing therapeutics tailored to individual patient gene expression profiles and their gut microbiome characteristics holds promise for improving patient survival outcomes.

MATERIALS AND METHODS

Collection of CRC-associated Genes

Using the keyword 'colorectal cancer', CRC-associated genes were obtained from the following databases: GeneCards (https://www.genecards.org/, accessed on 27th June 2024), Ensembl (https://asia.ensembl.org/index.html, accessed on 27th June 2024), OMIM (https://www.omim.org/, accessed on 27th June 2024), and PharmGKB (https://www.pharmgkb.org/, accessed on 27th June 2024). Differentially expressed genes (DEGs) from 30 paired colorectal tumor and normal tissue samples were obtained from microarray data of GSE74602 from the GEO database (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE74602, accessed on 27th June 2024). In this dataset, colon tumor tissue was sourced from the Singapore Tissue Bank. Patient demographic information was not reported. DEGs were analyzed using GEO2R, a built-in GEO database tool that utilizes R packages like limma for microarray data analysis. To account for the false discovery rate (FDR), the Benjamini & Hochberg method was utilized for *p*-value adjustment. Other default parameters included adjusted $p < 0.05$, $|log2\ fold\ change| > 0$, no limma precision weights, and no forced normalization.

Collection of GMM Target Genes

GMMs were downloaded from the Human Metabolome Database (HMDB) (https://hmdb.ca/, accessed on 1st July 2024). The following filters were applied: all metabolite statuses (detected, expected, or predicted), "feces" as biospecimen, "microbial" as origin, and all cellular locations. The Simplified Molecular Input Line Entry System (SMILES) notation, a string-based representation of chemical structures, of each metabolite (excluding those with fewer than 5 non-hydrogen atoms) was entered into the SwissTargetPrediction platform (https://www.swisstargetprediction.ch/, accessed on 1st July 2024) to predict target genes, that encode proteins likely to interact with the metabolites. Only genes with a predicted interaction probability greater than 0 were retained.

Identification of Common Genes among Datasets

Venny 2.1.0 (https://bioinfogp.cnb.csic.es/tools/venny/, accessed on 3rd July 2024) was used to generate a Venn diagram of the collected CRC-associated genes and the metabolite target genes. The overlapping genes were designated as CRC-associated genes linked to GMMs and were carried forward for subsequent analyses.

Identification of Hub Genes via Protein–Protein Interaction (PPI) Network

The common genes' gene symbols were entered into the STRING database (https://string-db.org/, accessed on 23rd October 2024) to predict interactions between their proteins. The parameters for PPI network construction included "Homo sapiens" as the species, a medium confidence score of ≥ 0.4, and an FDR stringency of < 0.05. Cytoscape v3.10.2 software was used to visualize the topology of the network. The topology parameters chosen were degree value, betweenness centrality (BC), and closeness centrality (CC). To identify top hub genes (genes with the most interactions with others) in the PPI network, a two-step screening process was applied based on a two-step screening process based on the methodology of Hu et al. (2023). Median values of degree, BC, and CC were used as thresholds. At each step, genes with values above the median for each parameter were retained. Genes that passed the second filter of median values were designated as the top hub genes.

KEGG Pathway Enrichment Analysis

Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis was performed using ShinyGo v0.741 (http://bioinformatics.sdstate.edu/, accessed on 2nd September 2024). KEGG pathway enrichment analysis was used to identify specific pathways or networks of genes involved in a particular biological function. KEGG terms were visualized in a bubble plot. The top hub genes were also subjected to KEGG pathway enrichment analysis, which was performed separately from the enrichment analysis of the initial common-gene dataset.

Identification of Prognostic Genes via Survival Analysis

The top hub genes were subjected to survival analysis using the GEPIA2 website (http://gepia2.cancer-pku.cn/, accessed on 3rd November 2024). Kaplan-Meier curves were plotted to show the relationship between the level of expression of each hub gene and the combined overall survival time of 461 colorectal adenocarcinoma and 172 rectum adenocarcinoma patients, whose tissue samples were obtained from the TCGA-COAD (https://portal.gdc.cancer.gov/projects/TCGA-COAD, accessed on 3rd November 2024) and TCGA-READ (https://portal.gdc.cancer.gov/projects/TCGA-READ, accessed on 3rd November 2024) datasets, respectively. Both datasets contain RNA-sequencing data of tumor and normal tissue samples. The racial distribution of patients showed White patients as the predominant group, followed by African-American, Asian, and patients with unreported races. All cancer stages of each cancer type are included in both datasets, but stage IIA is predominant. Patients' ages range from 31 to 90 years.

For the survival analysis parameters, the "group cutoff" was set to "median" to separate the patients into two groups (high and low expression) using the median gene expression value as the threshold. Genes associated with survival curves showing a statistically significant difference between groups ($p < 0.05$) were considered to have prognostic value and were considered prognostic genes.

Verification of Metabolite-Protein Binding via Molecular Docking

To assess potential interactions between the identified prognostic genes' proteins and GMMs used in SwissTargetPrediction, molecular docking simulations were performed in SwissDock (https://www.swissdock.ch/, accessed on 27th November 2024). Proteins encoded by the prognostic genes were searched in the UniProt database (https://www.uniprot.org/, accessed on 27th November 2024). The gene symbol and "human" were used as keywords. The parameters for the chosen proteins were as follows: "Homo sapiens" as the source organism, "PDB" as the database, "X-ray" as the method of protein structure identification, a resolution value of ≤ 2.00 Å (ideal for molecular docking), and no mutations. The protein structures were exported as PDB files. From the curated dataset of GMM target genes, the metabolite that had the highest probability of interaction with each prognostic gene was selected as the ligand for its corresponding protein and designated as a key GMM. The SMILES of each metabolite and the PDB file of its corresponding protein were entered and uploaded, respectively, into SwissDock, using the "Docking with AutoDock Vina" setting. Sampling exhaustivity was set to the maximum available value.

RESULTS AND DISCUSSION

Common Genes among Datasets

A total of 17,182 top DEGs (p-adj < 0.05) were identified from the GSE74602 dataset. Additionally, 11,301 CRC-associated genes were collated from the gene databases GeneCards, Ensembl, OMIM, and PharmGKB. SwissTargetPrediction generated 573 GMM target genes. After combining all gene datasets in a Venn diagram, 350 genes were found to be commonly shared between the CRC-associated genes and GMM target genes. Because two genes had redundant symbols, Figure 1 shows 350 genes instead of 352.

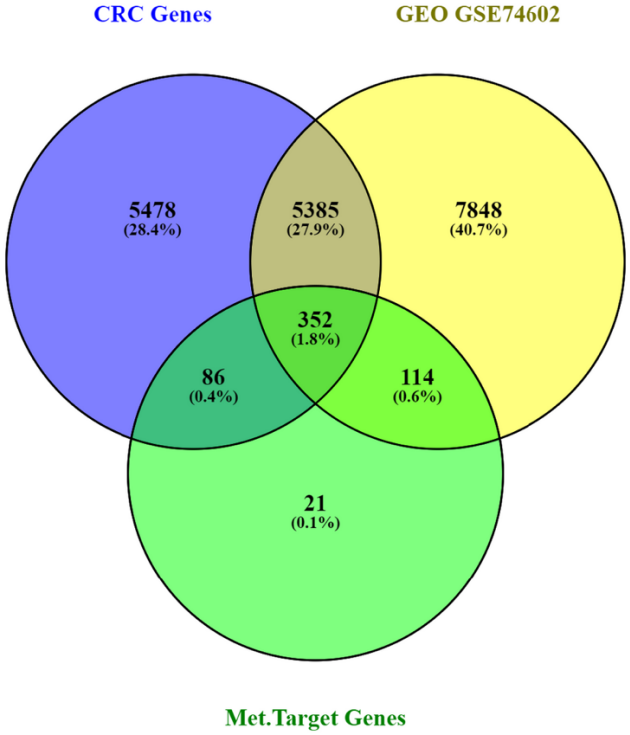

Figure 1: Venn diagram of CRC-associated genes and GMM target genes.

KEGG Pathway Enrichment Analysis

Figure 2A shows the top 10 enriched KEGG pathways associated with the 350 common genes. The fold enrichment value of the dot represents the extent of overrepresentation of the KEGG term in the dataset. The dot color and size represent the *p*-value and number of genes associated with the KEGG term, respectively. A redder dot indicates a lower *p*-value, which suggests a more statistically significant enrichment. A larger dot size and higher fold enrichment value suggest a stronger association of the genes in the dataset with the biological mechanism.

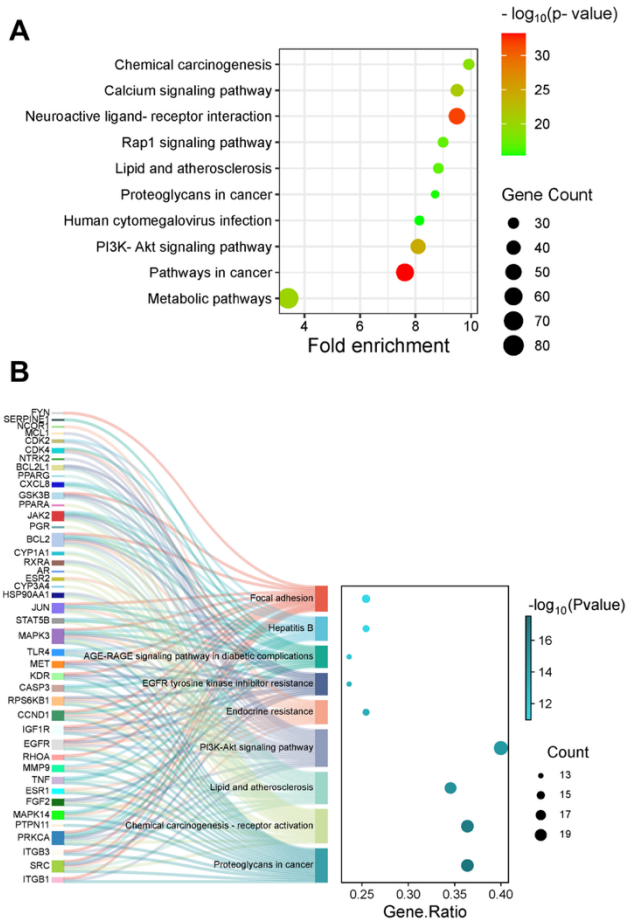

Figure 2: Top enriched KEGG pathways. (**A**) Pathways associated with the 350 common genes, and (**B**) Pathways associated with the top 55 hub genes.

Figure 2B presents the results of the KEGG pathway enrichment analysis of the 55 top hub genes. The terms "Chemical carcinogenesis", "Lipid and atherosclerosis", "Proteoglycans in cancer", and "PI3K-Akt signaling pathway" were among the top enriched pathways in both Figures 2A and 2B. In Figure 2B, the dot size represents the number of genes linked to a pathway. A larger dot and a higher gene ratio indicate higher pathway enrichment. A darker dot color indicates a smaller *p*-value, and thus higher significance of enrichment. The PI3K-Akt signaling pathway shows the highest enrichment values. The enrichment of the 'Pathways in cancer' term supports the association of the screened genes with CRC. The PI3K-Akt signaling pathway, which is a highly relevant biological pathway in both the initial 350 gene and 55 top hub gene collection, drives CRC progression by stimulating cancer cell growth, angiogenesis, and metastasis (Zhong et al., 2023).

PPI Network Analysis

Of the 350 common genes, 55 were identified as top hub genes from the PPI network topology screening process (Figure 3). In the network, nodes represent proteins and connecting edges represent protein–protein interactions. Proteins of hub genes are highlighted in red. Hub genes are nodes with values that were consistently in the upper median half of the selected network topology metrics (degree, BC, and CC) in each filtering step.

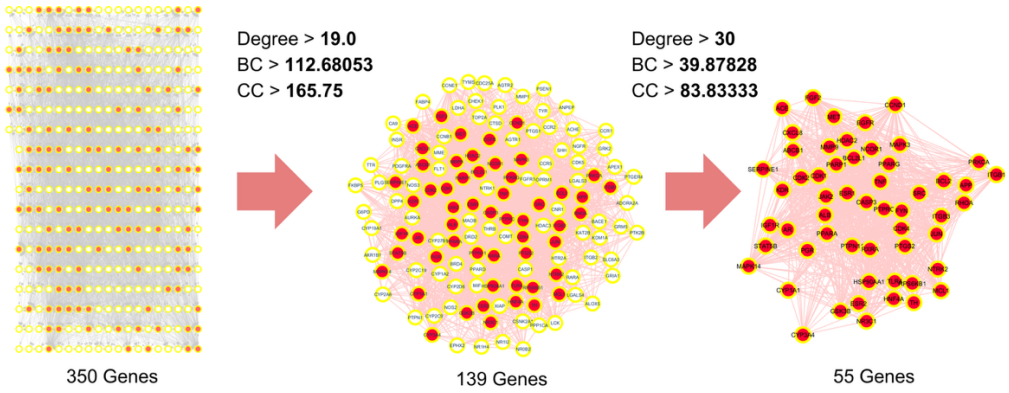

Figure 3: PPI network topology visualization of hub genes (represented by nodes in red) resulting from the two-step screening process.

Survival Analysis

Out of the 55 top hub genes, five genes, including CXCL8, HDAC2, PTGS2, MET, and PGR, showed statistically significant differences in overall survival between high and low gene expression groups (Figure 4). The survival curves of CXCL8, HDAC2, PTGS2, and MET show that a higher expression of these genes is associated with a better prognosis in patients with colorectal cancer. Conversely, a higher expression of PGR is associated with poorer prognosis.

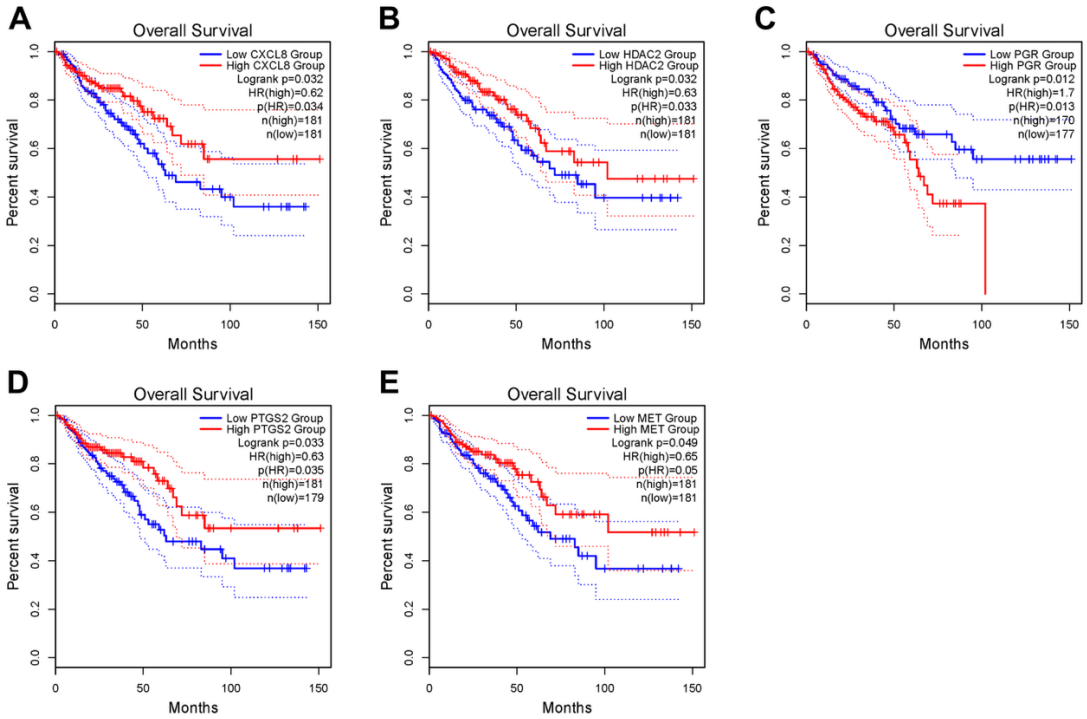

Figure 4: Survival analysis results based on expression levels of top hub genes. Overall Survival (OS) curves for high (red) vs. low (blue) expression of (A) CXCL8 (p = 0.032), (B) HDAC2 (p = 0.032), (C) PGR (p = 0.012), (D) PTGS2 (p = 0.033), and (E) MET (p = 0.049).

Contrary to the survival analysis results, *in vitro* and *in vivo* studies show that lower progesterone receptor (PGR) gene expression is linked to worse CRC prognosis, as progesterone acting through PGR can inhibit the growth of CRC cells by inducing cell cycle arrest and apoptosis (Zhang et al. 2021b). However, in a similar *in silico* study, survival analysis performed on gastric cancer patients showed that higher PGR expression is correlated with worse prognosis (Li and Zhou 2021). In contrast, higher expression of the prostaglandin-endoperoxide synthase 2 (PTGS2) gene has been linked to poorer CRC outcomes. PTGS2 is an enzyme that synthesizes prostaglandin and when glycosylated, it exists in a more stable form (gPTGS2), which has been associated with CRC (Venè et al. 2020). Patients whose tumors expressed PTGS2 were more likely to experience tumor recurrence and had worse CRC-specific survival (Kunzmann et al. 2013). However, the authors noted that confounding factors may have influenced CRC prognosis, particularly the tumor stage, which could have influenced both PTGS2 expression and patient outcomes.

Also in contrast, higher expression of the mesenchymal–epithelial transition factor (MET) gene is consistently associated with adverse clinical features in CRC. MET encodes the hepatocyte growth factor receptor (HGFR). Higher expression of MET was shown to be associated with increased tumor aggressiveness, metastasis, and poorer overall survival in patients with CRC (De Oliveira et al. 2009, Liu et al. 2015).

Some studies support the present finding that high expression of the histone deacetylase 2 (HDAC2) gene and interleukin-8 (CXCL8) genes lead to better overall survival. HDAC2 suppresses CRC metastasis by inhibiting epithelial-mesenchymal transition (EMT) and H19 and MMP14 expression (Hu et al. 2020). Low HDAC2 expression, observed in metastatic CRC, was associated with a worse prognosis in CRC patients (Hu et al. 2020). Deletion or knockdown of HDAC2 induces metastasis and EMT in CRC cells (Hu et al. 2020).

TCGA data and independent GEO cohorts showed that high CXCL8 expression was linked to better survival among CRC

patients (Li et al. 2021). An integrated analysis of tumor-infiltrating immune cell profiles showed a positive correlation between CXCL8 expression and anti-tumor immune response via the expression of dendritic cell activation markers CD80, CD83, and CD86 (Li et al. 2021). In *in vivo* animal models, the inhibition of the CXCL8 receptor (CXCR2) accelerated tumor progression (Li et al. 2021).

The discrepancies in survival outcomes observed in this study and in established studies may stem from dataset-specific factors in the TCGA cohorts used. The TCGA-COAD and TCGA-READ datasets are predominantly composed of White patients and Stage IIA colorectal adenocarcinoma samples that may have contributed to a demographic and clinical skew in the results. The prognostic behavior of these genes may differ between ethnicities, cancer stages, and variation in tumor molecular subtypes. Furthermore, the transcriptomic TCGA datasets used measure only mRNA levels, which do not always predict functional protein activity and downstream biological mechanisms that contribute to better or worse CRC prognosis.

Molecular Docking

A lower binding energy indicates a greater likelihood of stable binding between molecules. Binding energies below -5 kJ/mol are generally indicative of stable interactions. Molecular docking results show that each key GMM has a strong affinity for its corresponding protein (Figure 5). Ursodeoxycholic acid 3-sulfate and the HDAC2 protein (7KBG) showed the most favorable predicted binding at -29.217 kJ/mol, followed by lithocholic acid glycine conjugate and the PGR protein (1SQN) at -27.096 kJ/mol. The binding of phenylpyruvic acid and the CXCL8 protein (4XDX), and ferulic acid and the MET protein (4R1V) also showed favorable binding energies. The molecular docking of the PTGS2 protein was not performed because no PDB file of this protein had a resolution of less than 2.00 Å.

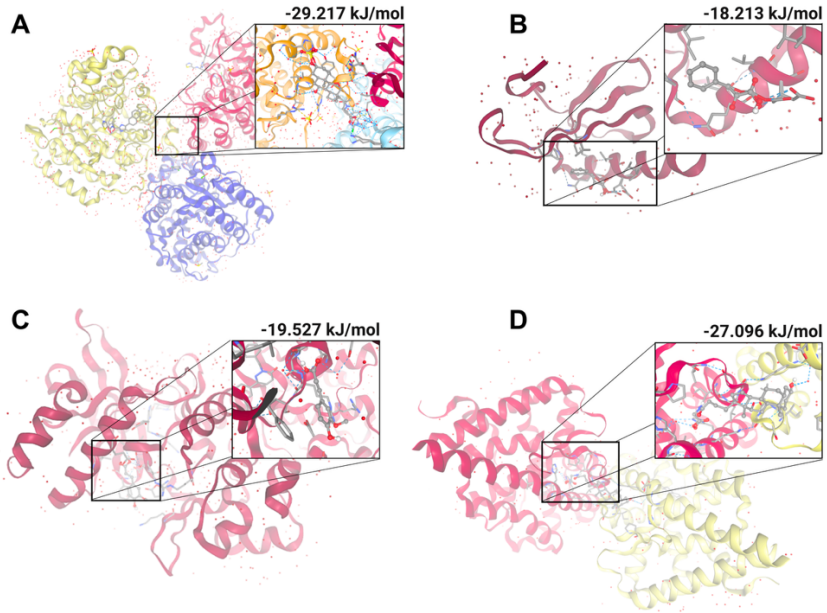

Figure 5: Molecular docking results of key GMMs and proteins encoded by potential prognostic genes. (**A**) Ursodeoxycholic acid 3-sulfate and HDAC2 protein (7KBG) binding; (**B**) Phenylpyruvic acid and CXCL8 protein (4XDX) binding; (**C**) Ferulic acid and MET protein (4R1V) binding; and (**D**) Lithocholic acid glycine conjugate and PGR protein (1SQN) binding.

Ursodeoxycholic acid (UDCA) is a secondary bile acid. Studies on the role of UDCA 3-sulfate, in particular, in CRC are lacking. Nonetheless, UDCA is known to have antiproliferative effects on colon cancer cells. UDCA can inhibit G1/S and G2/M phases, thereby inhibiting colon cancer cell-cycle progression (Kim et al., 2017). UDCA can increase the expression of cell cycle inhibitory proteins p27 and p21 and inhibit the expression of cell cycle regulators such as CDK2, CDK4, and CDK6. UDCA-based conjugates, but not UDCA itself, have shown inhibitory effects on HDAC2 (Caballero-Camino et al., 2021).

Studies directly linking between the other key metabolites and the proteins in this study are lacking. However, there is evidence of the key GMMs' involvement in CRC. Phenylpyruvic acid is a derivative of the amino acid phenylalanine. Phenylpyruvic acid may exert anti-tumor effects by competing against pyruvic acid in the lactate dehydrogenase metabolic pathway (Doğan & Doğan, 2022). This disrupts the normal lactic acid production in cancer cells and can thus inhibit the synthesis of amino acids needed by the cancer cells.

CONCLUSION

This study shows that the CRC-associated protein-coding genes linked to GMMs are most enriched in KEGG pathways involved in mechanisms of CRC progression. Specifically, the PI3K-Akt signaling pathway, whose dysregulation contributes to CRC metastasis, may serve as a GMM-associated pathway of interest. Higher expressions levels of PTGS2, MET, and PGR showed survival trends that differ from established findings possibly due to confounding factors such as tumor stage and patient ethnicity. While there are other studies that support the finding that higher CXCL8 or HDAC2 expression results in better CRC prognosis, this study did not stratify patients by demographic or clinical characteristics that may have variable influences on survival outcomes. In addition, transcriptomic data (mRNA levels) cannot fully elucidate downstream biological mechanisms and pathways that influence CRC prognosis. The discrepancies in the results warrant further analysis of larger, more diverse, and stratified cohorts, as well as further *in vitro* and *in vivo* experiments to validate computational predictions in this study.

The molecular docking simulations suggest binding interactions between GMMs with the highest binding probabilities and their protein targets. These results present evidence that GMMs, namely UDCA 3-sulfate, phenylpyruvic acid, ferulic acid, and

lithocholic acid glycine conjugate, may affect CRC-associated proteins involved in CRC progression. Hence, the metabolites, genes, and their respective proteins identified in this study represent potential prognostic CRC biomarkers and candidate molecular targets that may inform future mechanistic studies. Translation into clinical applications, including personalized therapeutic strategies integrating host genetics and gut microbiome profiles, will require rigorous experimental and clinical validation.

ACKNOWLEDGMENTS

The authors would like to express their gratitude to Dr. Arun Decano for her valuable guidance and insights during the conceptualization of this study. The authors also extend their appreciation to the Cordillera Region Health Research and Development Consortium (CRHRDC) for supporting the presentation and dissemination of this research in regional and national conferences, where valuable feedback and comments were received. This research did not receive any specific grant from funding agencies.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

CONTRIBUTIONS OF INDIVIDUAL AUTHORS

TJB Balitang is the primary author who designed and conceptualized the study in collaboration with the co-authors, performed data curation and analysis, and drafted the manuscript. CJL Padua and MAD Abuat conducted database searches and data extraction, assisted in data analysis and figure preparation, and participated in manuscript drafting. JB Melegrito provided guidance in data analysis, interpretation, and reviewing the manuscript. MEJV Sajo served as the senior and corresponding author and supervised the conceptualization and design of the study, critically reviewed, revised, and formatted the manuscript.

REFERENCES

Avuthu N, Guda C. Meta-analysis of altered gut microbiota reveals microbial and metabolic biomarkers for colorectal cancer. *Microbiol Spectr.* 2022; 10(4): e00013-22. https://doi.org/10.1128/spectrum.00013-22.

Caballero-Camino FJ, Rivilla I, Herraez E, Briz O, Santos-Laso A, Izquierdo-Sanchez L, Lee-Law PY, Rodrigues PM, Munoz-Garrido P, Jin S, Peixoto E, Richard S, Gradilone SA, Perugorria MJ, Esteller M, Bujanda L, Marin JJG, Banales JM, Cossío FP. Synthetic conjugates of ursodeoxycholic acid inhibit cystogenesis in experimental models of polycystic liver disease. *Hepatology.* 2021; 73(1): 186–203. https://doi.org/10.1002/hep.31216.

Candela M, Turroni S, Biagi E, Carbonero F, Rampelli S, Fiorentini C, Brigidi P. Inflammation and colorectal cancer: when microbiota-host mutualism breaks. *World J Gastroenterol.* 2014; 20(4): 908–922. https://doi.org/10.3748/wjg.v20.i4.908.

De Oliveira AT, Matos D, Logullo AF, da Silva SR, Neto RA, Longatto Filho A, Saad SS. MET is highly expressed in advanced stages of colorectal cancer and indicates worse prognosis and mortality. *Anticancer Res.* 2009; 29(11): 4807–4811.

Doğan A, Doğan AN. Investigation of the effects of sodium phenylpyruvate on pulmonary adenocarcinoma (A549) and mammary adenocarcinoma (MDA-MB-231) cell lines. *Caucasian Journal of Science.* 2022; 9(1): 20–34.

Hashim NAA, Ab-Rahim S, Suddin LS, Saman MSA, Mazlan M. Global serum metabolomics profiling of colorectal cancer. *Mol Clin Oncol.* 2019; 11(1): 3–14. https://doi.org/10.3892/mco.2019.1853.

Hu XT, Xing W, Zhao RS, Tan Y, Wu XF, Ao LQ, Li Z, Yao MW, Yuan M, Guo W, Li SZ, Yu J, Ao X, Xu X. HDAC2 inhibits EMT-mediated cancer metastasis by downregulating the long noncoding RNA H19 in colorectal cancer. *J Exp Clin Cancer Res.* 2020; 39(1): 270. https://doi.org/10.1186/s13046-020-01783-9.

Hu M, Yan H, Li H, Feng Y, Sun W, Ren Y, Ma L, Zeng W, Huang F, Jiang Z, Dong H. Use of network pharmacology and molecular docking to explore the mechanism of action of curcuma in the treatment of osteosarcoma. *Sci Rep.* 2023; 13(1): 9569. https://doi.org/10.1038/s41598-023-36687-z.

Kim EK, Cho JH, Kim E, Kim YJ. Ursodeoxycholic acid inhibits the proliferation of colon cancer cells by regulating oxidative stress and cancer stem-like cell growth. *PLoS One.* 2017; 12(7): e0181183. https://doi.org/10.1371/journal.pone.0181183.

Kunzmann AT, Murray LJ, Cardwell CR, McShane CM, McMenamin UC, Cantwell MM. PTGS2 (cyclooxygenase-2) expression and survival among colorectal cancer patients: a systematic review. *Cancer Epidemiol Biomarkers Prev.* 2013; 22(9): 1490–1497. https://doi.org/10.1158/1055-9965.EPI-13-0263.

Li E, Yang X, Du Y, Wang G, Chan DW, Wu D, Xu P, Ni P, Xu D, Hu Y. CXCL8-associated dendritic cell activation marker expression and recruitment as indicators of favorable outcomes in colorectal cancer. *Front Immunol.* 2021; 12: 667177. https://doi.org/10.3389/fimmu.2021.667177.

Li M, Zhou C. Progesterone receptor gene serves as a prognostic biomarker associated with immune infiltration in gastric cancer: a bioinformatics analysis. *Transl Cancer Res.* 2021; 10(6): 2663–2677. https://doi.org/10.21037/tcr-21-218.

Liu Y, Yu XF, Zou J, Luo ZH. Prognostic value of c-Met in colorectal cancer: a meta-analysis. *World J Gastroenterol.* 2015; 21(12): 3706–3710. https://doi.org/10.3748/wjg.v21.i12.3706.

Mardinoglu A, Shoaie S, Bergentall M, Ghaffari P, Zhang C, Larsson E, Bäckhed F, Nielsen J. The gut microbiota modulates host amino acid and glutathione metabolism in mice. *Mol Syst Biol.* 2015; 11(10): 834. https://doi.org/10.15252/msb.20156487.

Moshkov N, Becker T, Yang K, Horvath P, Dancik V, Wagner BK, Clemons PA, Singh S, Carpenter AE, Caicedo JC. Predicting compound activity from phenotypic profiles and chemical structures. *Nat Commun.* 2023; 14(1): 1967. https://doi.org/10.1038/s41467-023-37570-1.

Nguyen LH, Ma W, Wang DD, Cao Y, Mallick H, Gerbaba TK, Lloyd-Price J, Abu-Ali G, Hall AB, Sikavi D, Drew DA,

Mehta RS, Arze C, Joshi AD, Yan Y, Branck T, DuLong C, Ivey KL, Ogino S, Rimm EB, Song M, Garrett WS, Izard J, Huttenhower C, Chan AT. Association between sulfur-metabolizing bacterial communities in stool and risk of distal colorectal cancer in men. *Gastroenterology.* 2020; 158(5): 1313–1325. https://doi.org/10.1053/j.gastro.2019.12.029.

Stoffel EM, Murphy CC. Epidemiology and mechanisms of the increasing incidence of colon and rectal cancers in young adults. *Gastroenterology.* 2020; 158(2): 341–353. https://doi.org/10.1053/j.gastro.2019.07.055.

Venè R, Costa D, Augugliaro R, Carlone S, Scabini S, Casoni Pattacini G, Boggio M, Zupo S, Grillo F, Mastracci L, Pitto F, Minghelli S, Ferrari N, Tosetti F, Romairone E, Mingari MC, Poggi A, Benelli R. Evaluation of glycosylated PTGS2 in colorectal cancer for NSAIDs-based adjuvant therapy. *Cells.* 2020; 9(3): 683. https://doi.org/10.3390/cells9030683.

Xi Y, Xu P. Global colorectal cancer burden in 2020 and projections to 2040. *Transl Oncol.* 2021; 14(10): 101174. https://doi.org/10.1016/j.tranon.2021.101174.

Yang Y, Misra BB, Liang L, Bi D, Weng W, Wu W, Cai S, Qin H, Goel A, Li X, Ma Y. Integrated microbiome and metabolome analysis reveals a novel interplay between commensal bacteria and metabolites in colorectal cancer. *Theranostics.* 2019; 9(14): 4101–4114. https://doi.org/10.7150/thno.35186.

Zhang W, An Y, Qin X, Wu X, Wang X, Hou H, Song X, Liu T, Wang B, Huang X, Cao H. Gut microbiota-derived metabolites in colorectal cancer: the bad and the challenges. *Front Oncol.* 2021a; 11: 739648. https://doi.org/10.3389/fonc.2021.739648.

Zhang YL, Wen XD, Guo X, Huang SQ, Wang TT, Zhou PT, Li W, Zhou LF, Hu YH. Progesterone suppresses the progression of colonic carcinoma by increasing the activity of the GADD45α/JNK/c-Jun signalling pathway. *Oncol Rep.* 2021b; 45(6): 95. https://doi.org/10.3892/or.2021.8046.