Journal of Genetic Engineering and Biotechnology 24 (2026) 100713

Contents lists available at ScienceDirect

Journal of Genetic Engineering and Biotechnology

journal homepage: www.elsevier.com/locate/jgeb

The bioinformatics approach to identifying pathogenic variants for colorectal cancer (CRC)

Muhammad Ma'ruf [a], Lalu Muhammad Irham [b], Wirawan Adikusuma [c,d], Maulida Mazaya [c], Alfian Mubaraq [e], Didi Nurhadi Illian [f], Arini Sabilah Al Mustaqimah [e,g], Linda Chiuman [h], Hayssam M. Ali [i], Shofiyah Sabilah Al Mustaniroh [e], Mohammad Basyuni [e,j,*]

[a] Sekolah Tinggi Ilmu Kesehatan ISFI Banjarmasin, Banjarmasin 70123, Indonesia
[b] Faculty of Pharmacy, Universitas Ahmad Dahlan, Yogyakarta 55164, Indonesia
[c] Research Center for Computing, Research Organization for Electronics and Informatics, National Research and Innovation Agency (BRIN), Cibinong 16911, Indonesia
[d] Department of Pharmacy, Universitas Muhammadiyah Mataram, Mataram 83217, Indonesia
[e] Center of Excellence for Mangrove, Universitas Sumatera Utara, Medan 20155, Indonesia
[f] Department of Pharmacy, Faculty of Mathematics and Natural Sciences, Universitas Syiah Kuala, Banda Aceh 23111, Indonesia
[g] Faculty of Public Health, Universitas Sumatera Utara, Medan 20155, Indonesia
[h] Department of Physiology, Faculty of Medicine, Dentistry, and Health Sciences, Universitas Prima Indonesia, Medan 20118, Indonesia
[i] Department of Botany and Microbiology, College of Science, King Saud University, Riyadh 11451, Saudi Arabia
[j] Department of Forestry, Faculty of Forestry, Universitas Sumatera Utara, Medan 20155, Indonesia

ARTICLE INFO

Keywords:
Bioinformatics
Genetic variation
Genomic
Pathogenic variants
Colorectal cancer

ABSTRACT

Colorectal cancer (CRC) is the third most prevalent cancer globally, accounting for 9.6% of newly diagnosed cases and 9.3% of cancer-related deaths. It develops from the uncontrolled proliferation of glandular cells in the colon and rectum and is categorized into three primary types: sporadic, hereditary, and colitis-associated. While genetic susceptibility is a key factor in CRC pathogenesis, identifying high-impact pathogenic variants remains a significant challenge. This study integrates bioinformatics and population genetics approaches to identify CRC-associated single-nucleotide polymorphisms (SNPs) with potential clinical significance. CRC-associated SNPs were extracted from the Genome-Wide Association Studies (GWAS) Catalog, functionally annotated via HaploReg, and validated via Ensembl. In addition, expression quantitative trait locus (eQTL) data from the GTEx database were used to assess the effects of these variants on gene expression across human tissues. Our analysis identified three high-priority SNPs (rs9379084, rs3184504, and rs11557154) associated with the *RREB1, ATXN2, SH2B3,* and *DCAF12* genes, which exhibited marked allele frequency differences among populations. These findings suggest potential biomarkers for CRC risk assessment and highlight the importance of genetic screening across diverse populations.

1. Introduction

Colorectal cancer (CRC) is the third most common cancer worldwide, representing 9.6% of all newly diagnosed cases, and the second leading cause of cancer-related deaths, accounting for 9.3% of total cancer fatalities. It ranks as the third leading cause of cancer mortality in both men and women. Furthermore, the average incidence rate of CRC is significant in most developing countries, including Indonesia, with it being the third-highest cause of mortality among cancer patients.[1,2] The global incidence of CRC is expected to double by 2035, with the sharpest increase anticipated in less developed countries, largely driven by a significant rise in cases among older individuals.[3]

CRC is a disorder of the colon and rectum caused by the abnormal proliferation of glandular cells. It is classified into three main types: sporadic, hereditary, and colitis-associated. The global incidence of CRC is increasing, driven by both environmental and genetic factors. The risk increases with age, particularly in those with long-standing ulcerative colitis and Crohn's disease.[4] Colorectal cancer ranks as the second leading cause of cancer-related deaths, with nearly 935,000 fatalities estimated.[5] Globally, the incidence of cancer is increasing, accounting for 11% of all cancer diagnoses. The most significant increase in CRC cases and deaths is observed in medium to high Human Development

* Corresponding author at: Center of Excellence for Mangrove, Universitas Sumatera Utara, Medan 20155, Indonesia.
E-mail address: m.basyuni@usu.ac.id (M. Basyuni).

https://doi.org/10.1016/j.jgeb.2026.100713
Received 28 April 2025; Received in revised form 7 May 2026; Accepted 11 May 2026

Index (HDI) countries, where the adoption of Western lifestyles contributes to the increasing incidence.[6] Various factors, including health-related behaviors (such as smoking, obesity, and physical activity) and social determinants (such as education, income, and government healthcare investment), significantly influence cancer development. Hence, life expectancy should be a key consideration in developing effective cancer prevention and treatment strategies.[7] Current clinical management of CRC relies on colonoscopy and histopathological examination for diagnosis, followed by staging with imaging modalities and treatment with surgery, chemotherapy, radiotherapy, and, in selected cases, targeted agents (e.g., anti-EGFR, anti-VEGF) and immunotherapy. Despite these advances, many patients are still diagnosed at advanced stages, responses to systemic therapy are heterogeneous, and reliable prognostic or predictive biomarkers for guiding individualized treatment are limited. This highlights the need for molecular markers that could improve early detection, refine risk stratification, and inform therapeutic decision-making.

Detecting CRC symptoms may require DNA analysis, as genetic variations can impact disease progression. The Genome-Wide Association Studies (GWAS) catalog is a bioinformatics resource that compiles genetic variations, including single-nucleotide polymorphisms (SNPs), which are linked to liver fat content, liver enzymes, the development of nonalcoholic fatty liver disease, and predictive disease markers.[8,9] Genetic identification studies in humans aim to identify inherited genetic risk factors for diseases such as CRC. Despite the importance of gene mapping in CRC pathogenesis, information on genes linked to colorectal cancer risk has been limited. Therefore, this study utilized the GWAS catalog database to map genes associated with genetic variations across populations, highlighting their crucial role in CRC pathogenesis. Key gene variations affecting protein function were subsequently validated.

2. Materials and methods

2.1. Study design and workflow

This study employed an integrative bioinformatics and population genetics approach to identify colorectal cancer (CRC)-associated single-nucleotide polymorphisms (SNPs) with potential functional and clinical relevance. The analytical workflow consisted of three main stages: (i) identification and prioritization of CRC-associated SNPs, (ii) analysis of population-based allele frequency distributions, and (iii) evaluation of tissue-specific gene expression associated with prioritized SNPs.

A schematic overview of the study workflow is presented in Fig. 1. The methodological framework was adapted from previously published studies.[9–13]

2.2. Identification of CRC-associated SNPs

CRC-associated SNPs were retrieved from the NHGRI–EBI GWAS Catalog (http://www.ebi.ac.uk/gwas; accessed on 10 March 2024). Only variants reported to be significantly associated with CRC were included. To ensure robust association signals and minimize false positives due to multiple testing, a stringent genome-wide significance threshold of $p < 1 \times 10^{-8}$ was applied. Duplicate entries were removed, and the remaining SNPs were curated for downstream analyses.

2.3. Functional annotation and SNP prioritization

Functional annotation of the identified SNPs was performed using HaploReg version 4.2 (https://pubs.broadinstitute.org/mammals/haploreg; accessed on 15 March 2024). This tool was used to assess chromatin states, regulatory motifs, and potential effects on gene regulation.

Missense variants with predicted functional relevance and strong statistical significance were prioritized for further analysis. This approach enabled the identification of SNPs with potential biological and regulatory importance.

2.4. Validation of genetic variants

The genomic positions, gene annotations, and variant characteristics of prioritized SNPs were validated using the Ensembl Genome Browser (release 2024)[11] (https://www.ensembl.org; accessed on 22 March 2024). This step ensured consistency in variant annotation and gene mapping across reference databases.

2.5. Population allele frequency analysis

Allele frequency data for prioritized SNPs were obtained from Ensembl population genetics resources. Frequencies were analyzed across major continental populations, including African, American, East Asian, European, and South Asian groups, to evaluate population-

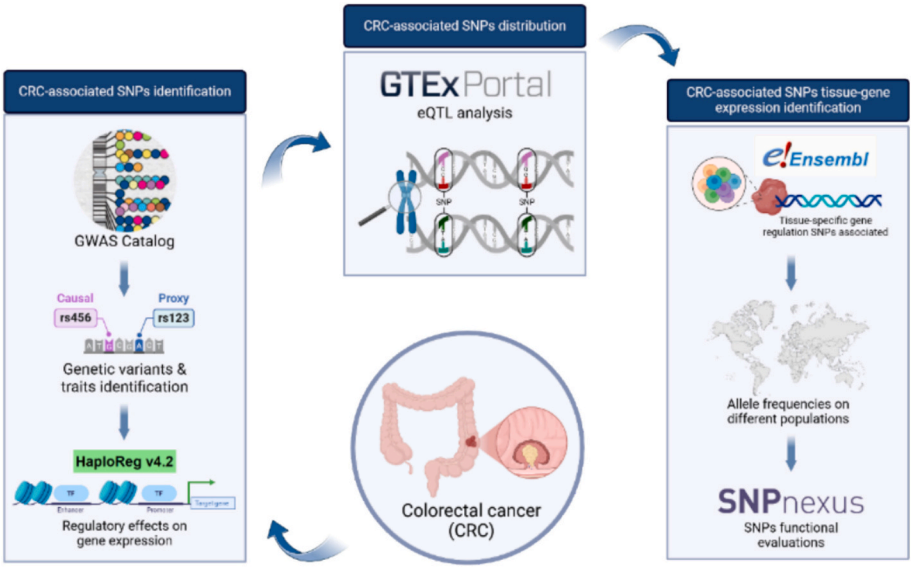

Fig. 1. illustrates the methodological workflow used in this study. The figure was created in BioRender. Mazaya, M. (2024), refer to https://BioRender.com/n09w794.

specific genetic variation and potential differences in CRC susceptibility.

2.6. Expression quantitative trait loci (eQTLs) and tissue-specific gene expression analysis

To investigate the regulatory effects of prioritized SNPs on gene expression, expression quantitative trait locus (eQTL) analysis was conducted using data from the GTEx Portal (https://www.gtexportal.org; accessed on 16 March 2024).[12]

Gene expression patterns were examined across multiple human tissues, with particular emphasis on tissues relevant to colorectal cancer, to identify potential functional consequences of the identified variants.

2.7. Additional functional evaluation of SNPs

Further functional characterization of the prioritized SNPs was conducted using the SNPnexus web server (https://www.snp-nexus.org; accessed on 25 March 2024).[13] This analysis evaluated the potential impact of variants on gene function, protein structure, and disease association. This multi-database validation strategy strengthened the biological relevance and interpretability of the identified CRC-associated variants.

2.8. Data presentation and visualization

Figures and tables were generated based on outputs obtained from the respective bioinformatics databases and analytical tools. The overall methodological workflow is illustrated in Fig. 1, while subsequent figures present tissue-specific gene expression profiles derived from the GTEx portal.

3. Results

3.1. Identification of CRC-associated SNPs

A total of 1551 CRC-associated SNPs were initially retrieved from the GWAS Catalog. After removing duplicate entries and applying a genome-wide significance threshold ($p < 1 \times 10^{-8}$), 17 unique SNPs were retained for further analysis (Table 1).

Functional annotation using HaploReg identified three high-priority missense variants (rs9379084, rs3184504, and rs11557154) located in the *RREB1*, *ATXN2/SH2B3*, and *DCAF12* genes, respectively (Table 2). These variants were selected based on their predicted functional impact and strong statistical significance.

Table 1
CRC-associated SNPs with genome-wide significance ($p < 1 \times 10^{-8}$).

SNP	*p*-value
rs1078643	2×10^{-27}
rs10936599	1×10^{-13}
rs1078643	2×10^{-13}
rs1078643	8×10^{-13}
rs9379084	2×10^{-12}
rs1078643	7×10^{-12}
rs11692435	7×10^{-12}
rs1078643	4×10^{-11}
rs3184504	2×10^{-10}
rs11557154	6×10^{-10}
rs1078643	1×10^{-9}
rs3184504	8×10^{-9}
rs11692435	1×10^{-8}
rs3184504	2×10^{-8}
rs10936599	3×10^{-8}
rs1446585	3×10^{-8}
rs7542665	4×10^{-8}

Table 2
Prioritized missense SNPs and associated genes.

Variation and risk alleles	*p* value	Gencode	Type of allele
rs9379084	2×10^{-12}	*RREB1*	Missense
rs3184504	2×10^{-10}	*ATXN2*	Missense
		SH2B3	Missense
rs11557154	6×10^{-10}	*DCAF12*	Missense

3.2. Distribution of CRC-associated SNPs

The allele frequency distribution of prioritized SNPs varied across continental populations, indicating potential population-specific genetic susceptibility to CRC (Table 3). Notably, rs3184504 exhibited high allele frequency in East Asian and African populations, whereas rs11557154 showed elevated frequency in East Asia compared to other populations.

3.3. Identification of the tissue gene expression of CRC-associated SNPs

To evaluate the tissue-specific gene expression of CRC-associated SNPs in human tissues, we utilized the GTEx portal database (http://www.gtexportal.org/), which provides gene expression data across various tissues. eQTL annotation highlights the most significant functional consequences of genetic variation.

Tissue gene expression for repressor of E1A-activated gene 1 (*RREB1*) can be accessed through the GTEx portal, a comprehensive database that provides gene expression data across multiple human tissues (Fig. 2). The GTEx portal is beneficial for researchers who want to study the regulation of gene expression in different tissues and how it is related to various diseases, including CRC. In terms of the correlation of *RREB1* expression in CRC, several studies have investigated the role of this gene in disease development and progression. For example, a study published in the journal "Gene Expression" reported that *RREB1* was significantly overexpressed in CRC tissues compared with normal tissues, suggesting that it may play a role in the tumorigenesis of this disease.[14]

ATAXIN 2 (*ATXN2*) is a gene that plays a crucial role in regulating cellular processes, including cell growth, differentiation, and apoptosis (Fig. 3). It is also linked to the development of various diseases, including cancer, neurodegenerative disorders, and metabolic conditions.[15]

The GTEx portal is a comprehensive database that provides access to gene expression data across multiple human tissues. This database is particularly useful for researchers who want to study the regulation of gene expression in different tissues and how it is related to various diseases. This explanation focuses on the SH2B3 gene, its expression across different tissues in the GTEx portal, and its correlation with CRC. *SH2B3* (SH2B Adaptor Protein 3) is a gene that plays crucial roles in various cellular processes, including cell growth, differentiation, and apoptosis (Fig. 4). It is also linked to the development of several diseases, such as cancer, cardiovascular disease, and metabolic disorders.[16]

The tissue gene expression data for *DCAF12*, also known as *DCAF12* (*CT102, DKFZP434O125, KIAA1892, MGC1058, TCC52, and WDR40A*), can be accessed through the GTEx portal, a comprehensive database that provides access to gene expression data across multiple human tissues (Fig. 5). The GTEx portal is particularly useful for researchers who want

Table 3
Distribution of allele frequencies of four SNPs across multiple continents.

SNP	Allele		Allele frequency				
	REF	ALT	Africa	America	East Asia	Europe	South Asia
rs9379084	G	A	0.006	0.111	0.156	0.132	0.106
rs3184504	T	C	0.981	0.746	0.997	0.536	0.931
rs11557154	C	T	0.011	0.215	0.601	0.137	0.235

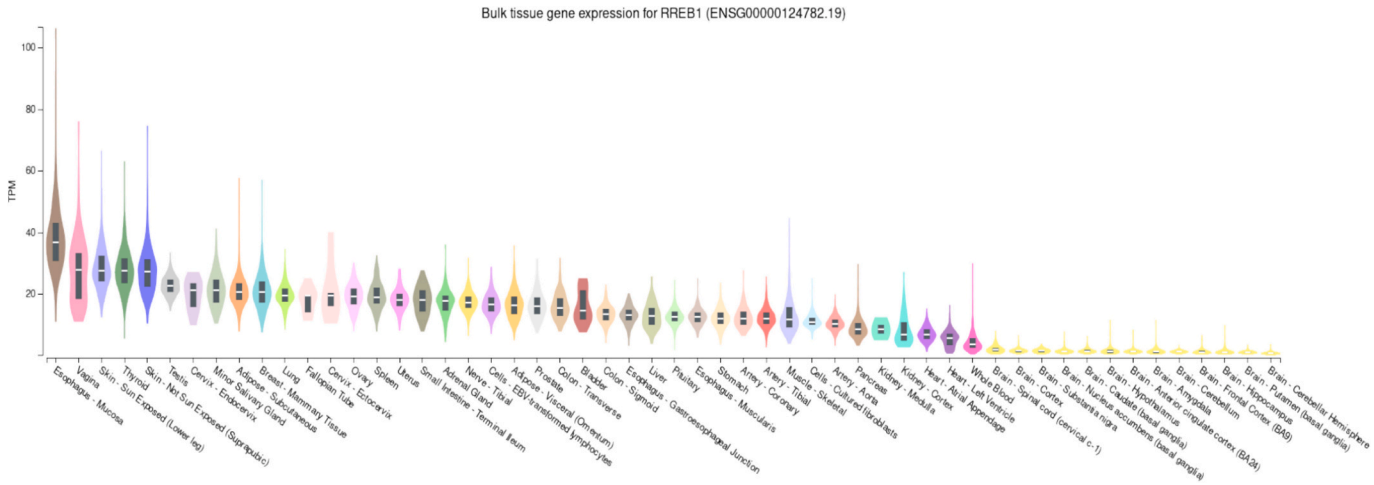

Fig. 2. Tissue gene expression of *RREB1* according to the database GTEx portal.

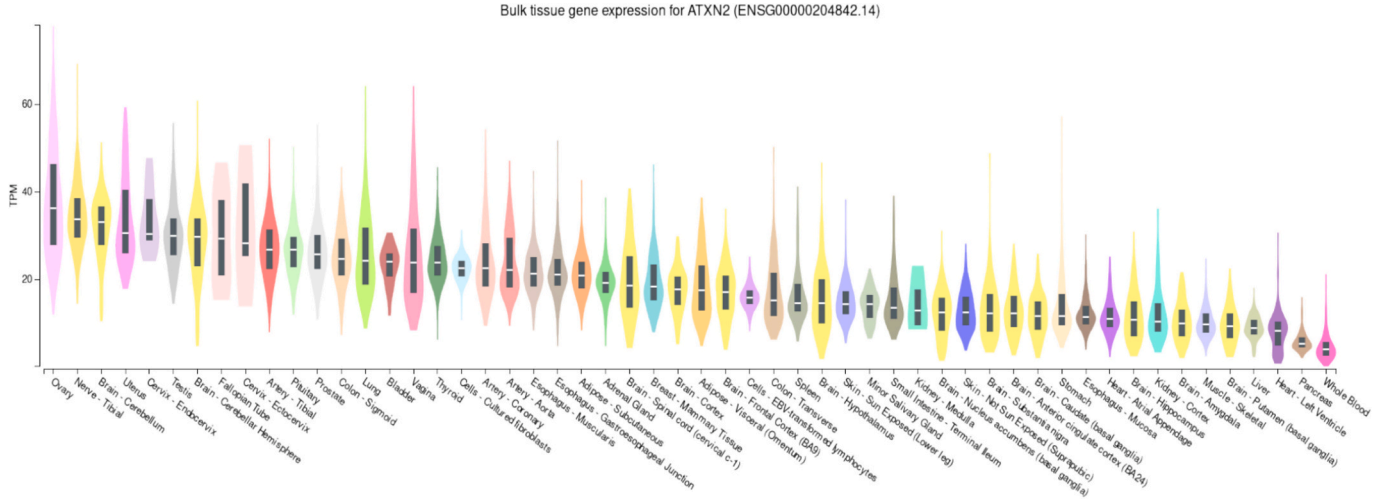

Fig. 3. Tissue gene expression of *ATXN2* according to the database GTEx portal.

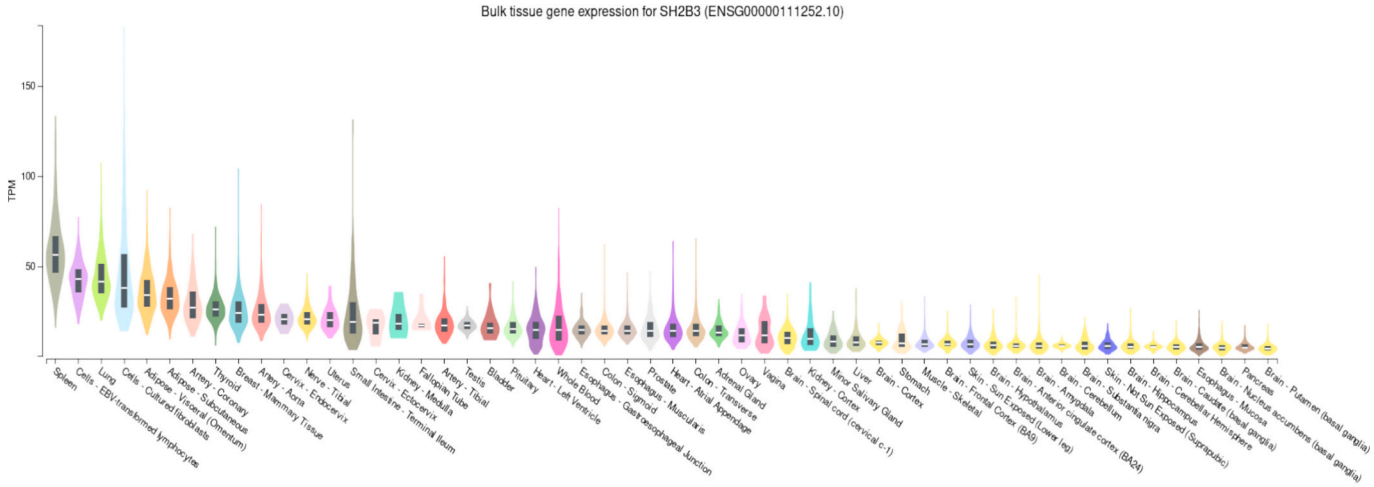

Fig. 4. Tissue gene expression of *SH2B3* according to the database GTEx portal.

to study the regulation of gene expression in different tissues and how it is related to various diseases. In this context, we focused on the expression of *DCAF12* across different tissues in the GTEx portal and its correlation with CRC.[17]

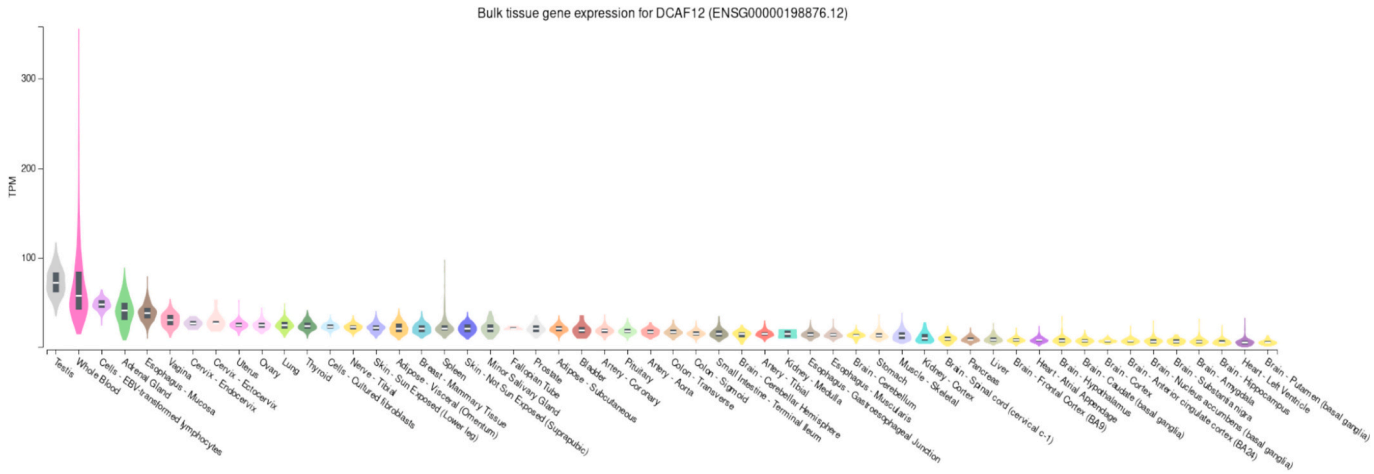

Fig. 5. Tissue gene expression of *DCAF12* according to the database GTEx portal.

4. Discussion

This study identified key missense variants associated with colorectal cancer (CRC) through an integrative bioinformatics approach combining GWAS data, functional annotation, and population genetics analysis. Among the prioritized variants, rs9379084 (*RREB1*), rs3184504 (*ATXN2/SH2B3*), and rs11557154 (*DCAF12*) emerged as potential contributors to CRC susceptibility, supported by their functional relevance and population-specific allele frequency patterns. The bioinformatics analysis integrates functional prediction models, allele frequency data, and gene expression databases to support the prioritization of these SNPs as key modulators of CRC risk. Moreover, population-specific allele frequency divergence highlights the potential genetic heterogeneity underlying CRC susceptibility across different ethnic groups, underscoring the importance of considering genetic diversity in disease risk assessment.

RREB1 functions as a transcription factor involved in cell proliferation, differentiation, and metabolic regulation. Its dysregulation has been linked to tumorigenesis, partly through modulation of the AKT signaling pathway and cancer cell metabolism. The identification of rs9379084 as a missense variant further supports its potential role in CRC pathogenesis. Additionally, it influences metabolic pathways such as fat development, fasting glucose homeostasis, and zinc transport, which are crucial for maintaining cellular integrity and preventing oncogenic transformation. Dysregulation of *RREB1* has been observed in various cancers, including CRC, where an imbalance in RREB1 activity disrupts normal cellular processes, potentially leading to uncontrolled cell proliferation.[18] Notably, posttranslational modifications, such as the acetylation and deacetylation of *RREB1, modulate its transcriptional activity*. This regulation is particularly significant in CRC and affects the AKT1 signaling pathway, a key mediator of cell survival and metabolism. By influencing AKT1 activity, *RREB1* plays a pivotal role in the glycolysis of CRC cells, a hallmark of cancer metabolism and tumorigenesis.[19] Thus, the involvement of RREB1 in transcriptional regulation and metabolic reprogramming underscores its importance as a critical factor in CRC pathogenesis.

The rs3184504 variant, associated with *ATXN2* and *SH2B3*, highlights the interplay between genetic susceptibility and immune regulation in CRC. *SH2B3* is involved in cytokine signaling and immune homeostasis, suggesting that immune-mediated mechanisms may contribute to CRC development. Meanwhile, *ATXN2* has been implicated in cell growth regulation and inflammatory pathways, reinforcing its relevance in cancer biology. This dual association implies that *ATXN2* plays a role in the shared genetic susceptibility pathways that underlie both CRC and IBD. Although the exact mechanisms remain unclear, *ATXN2* is believed to impact cell growth, differentiation, and immune regulation, all of which play crucial roles in the development and progression of CRC.[20] Furthermore, a study analyzing the involvement of m7G-lncRNAs in tumorigenesis via data from The Cancer Genome Atlas (TCGA) identified *ATXN2* as a key target regulated by m7G-lncRNAs, with elevated expression detected in CRC tissues.[21]

SH2B3, another susceptibility marker identified through GWAS, has been associated with CRC risk via the missense variant rs3184504 located at chromosome 12q24.12.[22] *SH2B3* encodes an adaptor protein involved in cytokine signaling and immune cell regulation, particularly in lymphocytes. Given the increasing recognition of immune dysregulation as a factor in tumor development, the role of this gene in immune modulation may provide a mechanistic link to CRC.[23] Although the direct associations of DCAF12 with CRC are limited in the literature, its expression has been noted across several cancer types, including CRC. Although less extensively characterized in CRC. *DCAF12* has been associated with cancer-related pathways and may serve as a prognostic biomarker. The identification of rs11557154 suggests a potential functional role that warrants further investigation. This finding underscores the potential of *DCAF12* and other differentially expressed genes (DEGs) as prognostic markers for CRC, suggesting the possibility of improved predictive accuracy over traditional staging systems such as TNM.[24,25]

Our analysis of allele frequency data from the Ensembl Genome Browser across different geographical regions, including Africa, America, East Asia, Europe, and Southeast Asia, revealed significant population-specific differences. For example, the rs3184504 (C) allele in SH2B3 is notably more prevalent in East Asian and African populations than in American, European, and South Asian populations. These variations in allele frequency emphasize the importance of incorporating population genetics into CRC risk assessment and developing region-specific screening strategies.[26]

Further expression analysis via data from the GTEx portal revealed significant overexpression of *RREB1* in CRC tissues, and its expression level correlated with disease progression. The GTEx portal provides valuable insights into tissue-specific gene expression, which can help elucidate the role of genes such as *RREB1* in CRC and identify potential therapeutic targets. The overexpression of *ATXN2* in CRC tissues, as revealed by GTEx data, is associated with poor prognosis, indicating its potential role in disease progression. Although the exact mechanisms by which *ATXN2* influences CRC remain fully explored, its regulatory effects on cell growth and immune modulation are likely contributing factors.[27] Finally, *SH2B3* expression analysis via GTEx highlights its potential involvement in CRC via its immune-regulatory functions. We can better understand how genetic variants influence disease pathogenesis by accessing large-scale gene expression data from resources such as GTEx and identifying new molecular targets for CRC therapy.[28]

While our study offers valuable insights into the potential genetic

factors driving CRC, several limitations exist. First, the functional effects of the identified SNPs have yet to be validated in experimental or clinical settings. Future studies integrating experimental validation and more diverse genomic datasets are needed to confirm these findings.[29] Additionally, the analysis relied heavily on publicly available datasets, which may contain biases, especially concerning population diversity. Although the allele frequency data from Ensembl provide a general overview, they may not fully reflect the genetic diversity in underrepresented populations. Additionally, this study focused primarily on missense variants, whereas other genetic factors, such as regulatory variants or structural alterations, may also play a role in CRC susceptibility but have not been extensively examined. Future research addressing these gaps could enhance our understanding of the genetic basis of CRC and refine risk prediction models across diverse populations. This study also highlights the importance of bioinformatics for identifying genetic risk factors in CRC by integrating large-scale genomic data, functional annotations, and population genetics. It enables efficient identification of disease-associated variants, streamlining candidate gene selection for further investigation.[30] Overall, this study demonstrates the utility of integrative bioinformatics approaches in identifying candidate genetic variants associated with CRC and provides a foundation for future functional and clinical investigations.

5. Conclusion

This study employed an integrative bioinformatics approach to identify candidate genetic variants associated with colorectal cancer (CRC). Three prioritized missense SNPs—rs9379084 (*RREB1*), rs3184504 (*ATXN2/SH2B3*), and rs11557154 (*DCAF12*)—were identified as potential contributors to CRC susceptibility based on their functional relevance and population-specific allele frequency patterns.

The observed variation in allele distribution across populations highlights the importance of incorporating population genetics into CRC risk assessment and supports the development of more targeted screening strategies.

Although further experimental validation is required, these findings provide a valuable foundation for future studies aimed at elucidating the molecular mechanisms of CRC and advancing precision medicine approaches.

CRediT authorship contribution statement

Muhammad Ma'ruf: Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Conceptualization. **Lalu Muhammad Irham:** Writing – review & editing, Methodology, Formal analysis, Data curation. **Wirawan Adikusuma:** Writing – review & editing, Investigation, Data curation. **Maulida Mazaya:** Writing – review & editing, Formal analysis. **Alfian Mubaraq:** Writing – review & editing, Data curation. **Didi Nurhadi Illian:** Writing – review & editing. **Arini Sabilah Al Mustaqimah:** Writing – review & editing, Investigation. **Linda Chiuman:** Writing – review & editing. **Hayssam M. Ali:** Writing – review & editing, Resources, Project administration. **Shofiyah Sabilah Al Mustaniroh:** Writing – review & editing, Data curation. **Mohammad Basyuni:** Writing – review & editing, Writing – original draft, Validation, Supervision, Resources, Project administration, Investigation, Funding acquisition, Data curation, Conceptualization.

Declaration of competing interest

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

Acknowledgment

This work was supported by The Center of Excellence for Mangrove, Universitas Sumatera Utara, through Penelitian Kolaborasi Unggulan Universitas Sumatera Utara (PKUU) EQUITY 2023-2024 [91/UN5.2.3.1/PPM/KPEP/2023] from the Indonesia Endowment Fund for Education and Ministry of Education, Culture, Research, and Technology, Republic Indonesia.

Data availability

The datasets generated and/or analyzed during the current study are publicly available in the Figshare repository at https://doi.org/10.6084/m9.figshare.31568566. These include the processed datasets underlying all figures and tables presented in this article.

In addition, the detailed analytical workflow and protocols used in this study have been deposited in protocols.io and are accessible at https://10.17504/protocols.io.kxygx641kg8j/v1.

All analyses were conducted using publicly available bioinformatics tools and databases as described in the Methods section; no custom code or proprietary software was developed.

The authors support open and transparent research practices and encourage reuse of these data and workflows to facilitate validation, reproducibility, and further research.

References

1. Gondhowiardjo S, et al. Five-year cancer epidemiology at the National Referral Hospital: hospital-based cancer registry data in Indonesia. *JCO Glob Oncol.* 2021;7:190–203. https://doi.org/10.1200/GO.20.00155.
2. Illian DN, et al. Current status, distribution, and future directions of natural products against colorectal cancer in Indonesia: a systematic review. *Molecules*. 2021. https://doi.org/10.3390/molecules26164984.
3. Papamichael D, et al. Treatment of colorectal cancer in older patients: International Society of Geriatric Oncology (SIOG) consensus recommendations 2013. *Ann Oncol.* 2015. https://doi.org/10.1093/annonc/mdu253.
4. Triantafillidis JK, Nasioulas G, Kosmidis PA. Colorectal cancer and inflammatory bowel disease: epidemiology, risk factors, mechanisms of carcinogenesis and prevention strategies. *Anticancer Res.* 2009;29:2727–2738.
5. Arnold M, Sierra MS, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global patterns and trends in colorectal cancer incidence and mortality. *Gut.* 2017;66(4):683–691. https://doi.org/10.1136/gutjnl-2015-310912.
6. Wong MCS, et al. Differences in incidence and mortality trends of colorectal cancer worldwide based on sex, age, and anatomic location. *Clin Gastroenterol Hepatol.* 2021;19:955–966.
7. Fidler MM, Soerjomataram I, Bray F. A global view on cancer incidence and national levels of the human development index. *Int J Cancer.* 2016;139(11):1–30.
8. Wang J, et al. Association of genetic risk score with NAFLD in an ethnically diverse cohort. *Hepatol Commun.* 2021;5(10):1689–1703. https://doi.org/10.1002/hep4.1751/suppinfo.
9. Sollis E, et al. The NHGRI-EBI GWAS catalog: knowledgebase and deposition resource. *Nucleic Acids Res.* 2023;51(1 D):D977–D985. https://doi.org/10.1093/nar/gkac1010.
10. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res.* 2012;40(D1). https://doi.org/10.1093/nar/gkr917.
11. Harrison PW, et al. Ensembl 2024. *Nucleic Acids Res.* 2024;52(D1):D891–D899. https://doi.org/10.1093/nar/gkad1049.
12. Aguet F, Ardlie KG, Lappalainen T. The GTEx Consortium Atlas of Genetic Regulatory Effects Across Human Tissues The GTEx Consortium* [Online]. Available: www.gtexportal.org; 2020.
13. Oscanoa J, Sivapalan L, Gadaleta E, Dayem Ullah AZ, Lemoine NR, Chelala C. SNPnexus: a web server for functional annotation of human genome sequence variation (2020 update). *Nucleic Acids Res.* 2020;48(W1):W185–W192. https://doi.org/10.1093/NAR/GKAA420.
14. Ding X, Duan H, Luo H. Identification of core gene expression signature and key pathways in colorectal cancer. *Front Genet.* 2020;11. https://doi.org/10.3389/fgene.2020.00045.
15. Birkenkamp-Demtroder K, et al. Gene expression in colorectal cancer. *Cancer Res.* 2002;62(15):4352–4363.
16. Cai S, Lu JX, Wang YP, Shi CJ, Yuan T, Wang XP. SH2B3, transcribed by STAT1, promotes glioblastoma progression through transducing IL-6/gp130 signaling to activate STAT3 signaling. *Front Cell Dev Biol.* 2021;9. https://doi.org/10.3389/fcell.2021.606527.
17. Yang Q, Feng M, Ma X, Li H, Xie W. Gene expression profile comparison between colorectal cancer and adjacent normal tissues. *Oncol Lett.* 2017;14(5):6071–6078. https://doi.org/10.3892/ol.2017.6915.
18. Deng YN, Xia Z, Zhang P, Ejaz S, Liang S. Transcription factor RREB1: from target genes toward biological functions. *Int J Biol Sci.* 2020;16(8):1463–1473. https://doi.org/10.7150/ijbs.40834.

19. Chen B, et al. miR-26a enhances colorectal cancer cell growth by targeting RREB1 deacetylation to activate AKT-mediated glycolysis. *Cancer Lett*. 2021;521:1–13. https://doi.org/10.1016/j.canlet.2021.08.017.
20. Orlando G, et al. Variation at 2q35 (PNKD and TMBIM1) influences colorectal cancer risk and identifies a pleiotropic effect with inflammatory bowel disease. *Hum Mol Genet*. 2016;25(11):2349–2359. https://doi.org/10.1093/hmg/ddw087.
21. Chen S, Dong R, Li Y, Wu H, Liu M. m7G-lncRNAs are potential biomarkers for prognosis and tumor microenvironment in patients with colon cancer. *Nan Fang Yi Ke Da Xue Xue Bao*. 2022;42(5):681–689. https://doi.org/10.12122/j.issn.1673-4254.2022.05.08.
22. Pande M, et al. Genetic susceptibility markers for a breast-colorectal cancer phenotype: exploratory results from genome-wide association studies. *PloS One*. 2018;13(4). https://doi.org/10.1371/journal.pone.0196245.
23. Garcia-Etxebarria K, et al. Genetic variants as predictors of the success of colorectal cancer treatments. *Cancers (Basel)*. 2023;15(19). https://doi.org/10.3390/cancers15194688.
24. Fernandez-Rozadilla C, et al. Deciphering colorectal cancer genetics through multiomic analysis of 100,204 cases and 154,587 controls of European and East Asian ancestries. *Nat Genet*. 2023;55(1):89–99. https://doi.org/10.1038/s41588-022-01222-9.
25. Pan F, et al. Prognosis prediction of colorectal cancer using gene expression profiles. *Front Oncol*. 2019;9(APR). https://doi.org/10.3389/fonc.2019.00252.
26. Deutelmoser H, et al. Genotype-based gene expression in colon tissue—prediction accuracy and relationship with the prognosis of colorectal cancer patients. *Int J Mol Sci*. 2020;21(21):1–13. https://doi.org/10.3390/ijms21218150.
27. Asting AG, Carén H, Andersson M, Lönnroth C, Lagerstedt K, Lundholm K. COX-2 gene expression in colon cancer tissue related to regulating factors and promoter methylation status. *BMC Cancer*. 2011;11. https://doi.org/10.1186/1471-2407-11-238.
28. Spitzner M, et al. A gene expression signature for chemoradiosensitivity of colorectal cancer cells. *Int J Radiat Oncol Biol Phys*. 2010;78(4):1184–1192. https://doi.org/10.1016/j.ijrobp.2010.06.023.
29. Sari DP, et al. Cytotoxic and antiproliferative activity of polyisoprenoids in seventeen mangroves species against WiDr colon cancer cells. *Asian Pac J Cancer Prev*. 2018;19(12):3393–3400. https://doi.org/10.31557/APJCP.2018.19.12.3.
30. Basyuni M, Wati R. Bioinformatics analysis of the oxidosqualene cyclase gene and the amino acid sequence in mangrove plants. *J Phys Conf Ser*. 2017;801(1). https://doi.org/10.1088/1742-6596/801/1/012011.