

The possibility of prognostic and functional values of the 8q24 and 20q13 chromosomal bands in colorectal cancer

Seyed Ahmadreza Siadat*

Department of Biology, School of Science, Shiraz University, Shiraz, Iran

ABSTRACT

Colorectal cancer (CRC) remains a major global health concern, especially given its increasing incidence among younger individuals. While genome-wide association studies (GWAS) have identified numerous CRC-associated polymorphisms, their spatial distribution and functional implications are not fully understood. This study examined the locations of 1,346 CRC-linked polymorphisms across chromosomal bands. The results revealed significant nonrandom clustering across thirteen chromosomal bands: 1q41, 6p21, 8q24, 9q34, 10p14, 10q25, 11q12, 12p13, 15q13, 18q21, 19q13, 20p12, and 20q13. Functional enrichment analysis of genes within these bands revealed several Gene Ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways. Reciprocal chromosomal enrichment confirmed that many of these terms and pathways were not randomly localized within the same bands, highlighting their potential biological significance. Survival analysis using TCGA data identified three KEGG pathways and 33 GO terms mapped to nine of the thirteen bands that were significantly associated with poor prognosis. Notably, the 8q24 and 20q13 regions were enriched for differentially expressed genes and survival-associated terms yet showed no significant enrichment for genes with high somatic mutation rates. These results imply that 8q24 and 20q13 act as regulatory hotspots rather than mutation-driven regions. Overall, this integrative approach identified functionally and clinically relevant genomic regions that may contribute to inherited CRC risk and progression, providing valuable targets for the development of diagnostic and prognostic biomarkers.

Keywords: Chromosomal Bands; Colorectal Cancer; Enrichment Analysis; Genome-Wide Association Studies; Polymorphism

INTRODUCTION

Colorectal cancer (CRC) is a common malignancy and the second leading cause of cancer-related deaths worldwide. Despite advances in prevention, screening, and treatment, the incidence of CRC is increasing, particularly among younger individuals. This highlights the need for a better understanding of its genetic landscape [1]. Genome-wide association studies (GWAS) have identified numerous genetic variants associated with CRC risk, shedding light on inherited predispositions [2-4]. However, although these individual risk loci have been extensively cataloged, the functional significance, spatial distribution, and collective impact of these variants across chromosomal regions remain poorly understood. This knowledge gap

*Corresponding Author: Department of Biology, School of Science, Shiraz University, Shiraz 71467-13565, Iran;
Email: ahmadreza.siadat@yahoo.com

hinders their translation into clinical applications. Investigating the genomic organization and chromosomal localization of these polymorphisms could reveal nonrandom clustering patterns and highlight regulatory hotspots involved in tumorigenesis.

Previous studies have demonstrated that the nonrandom distribution of cancer-associated polymorphisms may indicate chromosomal regions containing key oncogenes, tumor suppressors, or elements involved in gene regulation [5-7]. Other studies have shown that alterations in specific chromosomal bands can affect gene expression, disrupt cellular pathways, and promote malignant transformation [8, 9]. Despite these observations, a comprehensive analysis of chromosomal bands enriched for CRC-associated polymorphisms and their functional implications is lacking. Therefore, the present study sought to address this gap.

MATERIALS AND METHODS

The current study consisted of five steps (Fig. 1). They are described below.

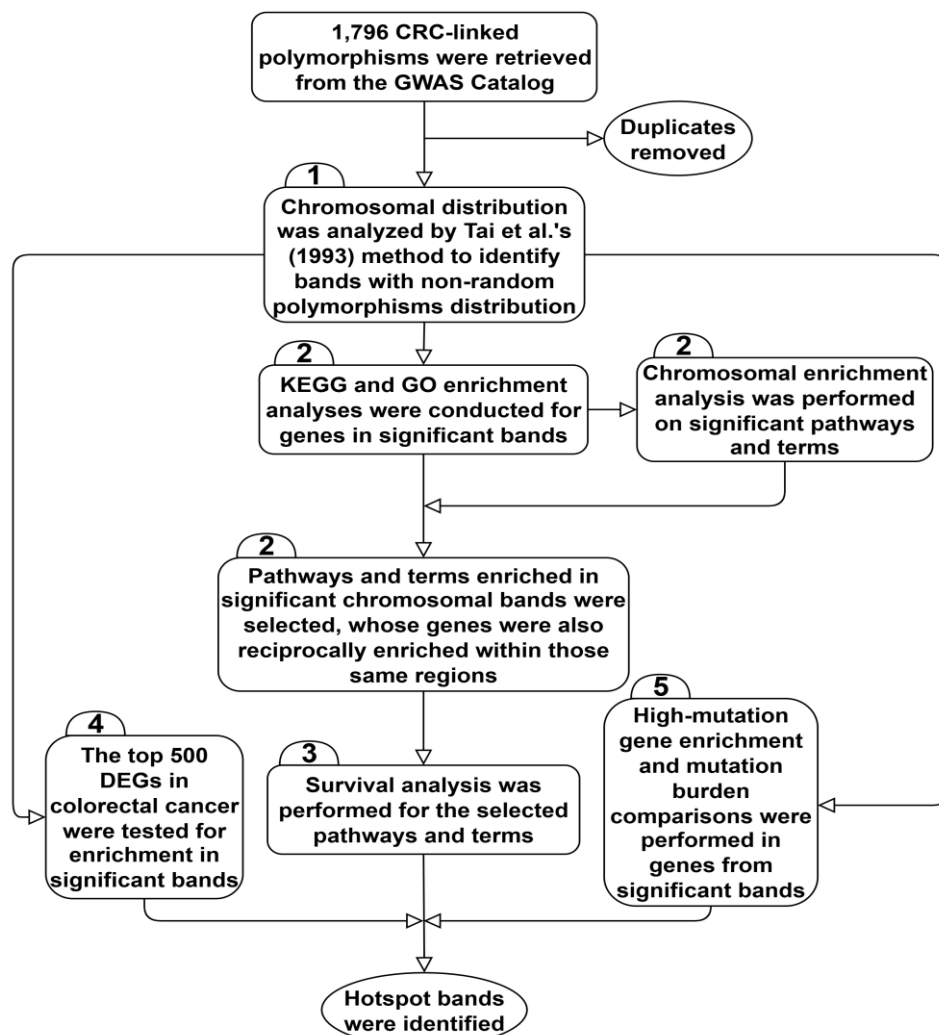


Figure 1: Diagram illustrating the overall workflow employed in the study.

Step 1: Chromosomal Distribution of CRC-Associated Polymorphisms: On July 12, 2024, we obtained the polymorphisms from the [MONDO_0005575](https://www.ebi.ac.uk/mondo/0005575) dataset, consisting of 75 GWAS on colorectal cancer (see Supplementary Data). We subsequently evaluated the nonrandom distribution of these polymorphic sites across human chromosomes [10].

Step 2: Enrichment Analyses: Two enrichment analyses were performed in this step. First, we used the biomaRt package in R to extract all genes located within chromosomal bands harboring nonrandom CRC-associated polymorphic sites from Ensembl. We then performed Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses on these gene sets using the clusterProfiler package in R. The GO terms were categorized into biological processes (BP), molecular functions (MF), and cellular components (CC). Second, we used the AnnotationDbi and KEGGREST packages to retrieve genes associated with each GO term and KEGG pathway, respectively. Finally, we conducted a chromosomal location enrichment analysis of these genes using the msigdbR package.

Step 3: Survival Analysis: Survival analysis was conducted using the cSurvival web tool (<https://tau.cmmt.ubc.ca/cSurvival>) [11]. based on the TCGA-COAD and TCGA-READ datasets. Kaplan-Meier estimates and log-rank tests were used to analyze the relationship between KEGG pathways and GO terms expression levels and overall survival.

Step 4: Differentially Expressed Genes (DEGs) Analysis: The top 500 differentially expressed genes from TCGA-COAD, including 250 upregulated and 250 downregulated genes, were obtained via UALCAN (<https://ualcan.path.uab.edu/index.html>) [12]. Enrichment of these DEGs within chromosomal bands was assessed using hypergeometric testing.

Step 5: Mutation Burden Analysis in Chromosomal Bands: Somatic mutation data for COAD and READ were obtained from The Cancer Genome Atlas (TCGA) via the TCGAbiolinks package. Mutations within coding sequences (CDS) were normalized using the following formula:

$$\text{Adjusted Mutation Count} = \frac{\text{Total Mutations}}{\text{CDS Length} + Q1} \times \log_2(\text{CDS Length})$$

where Q1 represents the first quartile of gene lengths, mitigating gene size bias. Gene length information was obtained from Ensembl using the biomaRt package. This adjustment accounts for gene size while emphasizing the relative mutation burden in longer genes. Genes with adjusted mutation counts above the median were classified as high-mutation genes. A hypergeometric test was used to assess their enrichment within chromosomal bands. Moreover, the mean adjusted mutation burden of genes in each band was compared with background genes using a *t*-test.

Statistical Analyses and Data Visualization: Statistical analyses were conducted in R, and significant bands, along with their polymorphism burden, were visualized using the chromPlot package. To account for multiple hypothesis testing and control the false discovery rate, the Benjamini-Hochberg method was used to adjust *p*-values. Statistical significance was defined as an adjusted *p*-value less than 0.05.

RESULTS

A total of 1,796 CRC-associated polymorphisms were compiled from 75 GWAS studies. After removing duplicates, 1,346 unique variants were retained for further analysis. Of these variants, 637 were located within protein-coding genes, with the majority (577) occurring in intronic regions. The remaining 709 variants were mapped to non-coding regions, including 503 in unannotated intergenic elements. To investigate their genomic distribution, we examined the location of these variants across chromosomal bands. The results revealed significant nonrandom clustering across thirteen chromosomal bands (1q41, 6p21, 8q24, 9q34, 10p14, 10q25, 11q12, 12p13, 15q13, 18q21, 19q13, 20p12, and 20q13) (Table 1, Figure S1).

Table 1: Chromosomal distribution of colorectal cancer-associated polymorphisms

Chromosomal band	Number of polymorphic sites	F	df	Adj p-Value
1q41	18	4.55	2658, 36	5.59×10^{-5}
6p21	27	4.15	2640, 54	4.66×10^{-7}
8q24	32	2.71	2630, 64	2.46×10^{-4}
9q34	18	3.96	2658, 36	3.83×10^{-4}
10p14	10	4.27	2674, 20	4.18×10^{-2}
10q25	18	3.25	2658, 36	5.46×10^{-3}
11q12	13	3.99	2668, 26	9.49×10^{-3}
12p13	21	3.46	2652, 42	4.70×10^{-4}
15q13	18	7.73	2658, 36	1.89×10^{-8}
18q21	20	2.66	2654, 40	2.90×10^{-2}
19q13	26	2.34	2642, 52	2.59×10^{-2}
20p12	27	5.09	2640, 54	5.92×10^{-9}
20q13	29	3.28	2636, 58	1.88×10^{-5}

Note: total of 1,346 CRC-associated polymorphisms were analyzed and bands that bearing non-randomly polymorphic sites were shown in the Table.

Next, we performed two enrichment analyses. First, we retrieved all genes located within the aforementioned bands and subjected them to enrichment analysis to elucidate their functional significance. This analysis revealed a total of 207 GO terms and 10 KEGG pathways, which are summarized in Tables S1 and S2, respectively. We know that genes involved in biological pathways are located on different chromosomes. Second, to investigate whether the genes associated with the aforementioned KEGG pathways and GO terms are nonrandomly distributed and are primarily located within the thirteen specified chromosomal bands, we conducted an additional analysis. We retrieved all genes associated with these pathways and terms and tested them for enrichment within the previously identified chromosomal bands (see Tables S3 and S4). This analysis revealed that all chromosomal bands, except 10p14, were significantly enriched for genes associated with at least one respective KEGG pathway and/or GO term. These results suggest the functional relevance of these chromosomal bands.

Following the previous enrichment analyses, in the third step we performed survival analysis on the KEGG pathways and GO terms that were mutually enriched within the aforementioned 12 chromosomal bands. This analysis evaluated the clinical relevance of our findings. The goal was to determine whether functionally enriched GO terms and KEGG pathways were associated with patient prognosis. The analysis identified three KEGG pathways and 33 GO terms that were significantly associated with survival in CRC (Table 2). Notably, these prognostic pathways and terms were associated with nine chromosomal bands: 8q24, 9q34, 10q25, 11q12, 12p13, 15q13, 18q21, 19q13, and 20q13.

In the fourth step of the present study, we examined the randomness distribution of the top 500 DEGs from the COAD project on the chromosomal bands associated with CRC that we identified in the first step. Our analysis revealed that only two chromosomal bands, 8q24 and 20q13, were significantly enriched with the DEGs (Table 3).

Lastly, we examined the somatic mutation burden of genes located on chromosomal bands associated with CRC. These bands were identified in the first step of the study. However, no statistically significant association was found between somatic mutation burden and chromosomal bands (Table S6).

DISCUSSION

The present study identified thirteen chromosomal bands that exhibiting a non-random distribution of CRC-associated GWAS polymorphisms. This finding suggests the functional relevance of these chromosomal bands in CRC. These regions have previously been associated

with cancer progression. For example, 8q24 contains regulatory elements that control MYC expression [13, 14].

Table 2: Survival analysis of KEGG pathways and Gene Ontology terms bidirectionally enriched in chromosomal bands

Category	Term ID	Description	Chromosomal band	Adj p-value
MF	GO:0001786	Phosphatidylserine binding	8q24	4.46×10^{-2}
MF	GO:0033130	Acetylcholine receptor binding	8q24	1.09×10^{-2}
CC	GO:0005930	Axoneme	9q34	4.46×10^{-2}
CC	GO:0036126	Sperm flagellum	9q34	2.45×10^{-2}
BP	GO:0070085	Glycosylation	9q34	4.46×10^{-2}
CC	GO:0097014	Ciliary plasm	9q34	1.09×10^{-2}
MF	GO:0004935	Adrenergic receptor activity	10q25	2.45×10^{-2}
KEGG	hsa04740	Olfactory transduction	11q12	1.63×10^{-2}
BP	GO:0001580	Detection of chemical stimulus involved in sensory perception of bitter taste	12p13	4.46×10^{-2}
BP	GO:0001906	Cell killing	12p13	2.70×10^{-2}
MF	GO:0008527	Taste receptor activity	12p13	6.55×10^{-3}
MF	GO:0033038	Bitter taste receptor activity	12p13	2.45×10^{-2}
BP	GO:0050909	Sensory perception of taste	12p13	1.09×10^{-2}
BP	GO:0050912	Detection of chemical stimulus involved in sensory perception of taste	12p13	4.46×10^{-2}
BP	GO:0050913	Sensory perception of bitter taste	12p13	4.46×10^{-2}
CC	GO:0070820	Tertiary granule	12p13	2.41×10^{-2}
CC	GO:0070821	Tertiary granule membrane	12p13	4.46×10^{-2}
MF	GO:0140375	Immune receptor activity	12p13	4.46×10^{-2}
KEGG	hsa04742	Taste transduction	12p13	1.63×10^{-2}
CC	GO:0000137	Golgi cis cisterna	15q13	1.63×10^{-2}
MF	GO:0005231	Excitatory extracellular ligand-gated monoatomic ion channel activity	15q13	1.09×10^{-2}
MF	GO:1904315	Transmitter-gated monoatomic ion channel activity involved in regulation of postsynaptic membrane potential	15q13	4.46×10^{-2}
KEGG	hsa05033	Nicotine addiction	15q13	3.76×10^{-2}
MF	GO:0004857	Enzyme inhibitor activity	18q21	4.46×10^{-2}
BP	GO:0043086	Negative regulation of catalytic activity	18q21	4.46×10^{-2}
MF	GO:0001228	DNA-binding transcription activator activity, RNA polymerase II-specific	19q13	1.72×10^{-2}
BP	GO:0002765	Immune response-inhibiting signal transduction	19q13	6.55×10^{-3}
MF	GO:0004875	Complement receptor activity	19q13	4.46×10^{-2}
MF	GO:0032396	Inhibitory MHC class I receptor activity	19q13	4.77×10^{-2}
MF	GO:0033691	Sialic acid binding	19q13	4.46×10^{-2}
CC	GO:0070821	Tertiary granule membrane	19q13	4.46×10^{-2}
MF	GO:0140375	Immune receptor activity	19q13	4.46×10^{-2}
MF	GO:0004857	Enzyme inhibitor activity	20q13	4.46×10^{-2}
BP	GO:0006959	Humoral immune response	20q13	4.46×10^{-2}
BP	GO:0019730	Antimicrobial humoral response	20q13	3.51×10^{-2}
BP	GO:0043086	Negative regulation of catalytic activity	20q13	4.46×10^{-2}

Note: The full table is provided in Table S5.

Table 3: Chromosomal enrichment analysis of differentially expressed genes (DEGs) in CRC-associated chromosomal bands

Chromosomal bands	Number of genes in each band	Number of DEGs in each band	Adj p-value
1q41	36	2	1.00
6p21	277	7	1.00
8q24	170	11	8.50×10^{-3}
9q34	211	4	1.00
10p14	16	0	1.00
10q25	56	1	1.00
11q12	204	4	1.00
12p13	187	3	1.00
15q13	33	0	1.00
18q21	81	3	1.00
19q13	780	15	1.00
20p12	41	0	1.00
20q13	217	14	1.69×10^{-3}

Note: The comparison was performed assuming a total gene count of 25,000 and a total of 500 DEGs. Significant enrichment of DEGs within chromosomal bands is indicated by adjusted p-values less than 0.05.

These regions may play critical roles in CRC pathogenesis by harboring key regulatory elements, tumor suppressor genes, or oncogenes. As previously suggested, genetic polymorphisms in certain chromosomal regions could be leveraged to develop a laboratory diagnostic test for mass screening programs, enabling the identification of individuals at high risk for various diseases [15, 16]. Similarly, genetic variations in the 13 chromosomal regions identified in this study could facilitate the development of diagnostic kits for colorectal cancer.

Survival analysis revealed that three KEGG pathways and 33 GO terms, associated with nine chromosomal bands, were significantly linked to poor CRC survival (Table 2). Many of these pathways and terms, such as those involved in olfactory and taste receptor signaling, immune regulation, microRNA (miRNA) transcription control, and enzyme inhibition, have been previously linked to CRC processes, including tumor growth, immune evasion, and microenvironmental remodeling [17-23]. These results suggest that these chromosomal bands and their associated pathways could serve as indicators of CRC prognosis.

The present study showed that the 8q24 and 20q13 chromosomal bands were enriched for DEGs and six GO terms associated with CRC patient survival. The 8q24 region, which harbors the MYC oncogene, plays a central role in cell proliferation, and its amplification has been linked to increased CRC susceptibility [13, 24]. Similarly, 20q13 is involved in the disruption of cell cycle regulation and apoptosis, with recurrent amplifications and deletions contributing to tumorigenesis [25-27]. Notably, amplification of the 20q13.33 sub-band has been reported as an early and frequent event in sporadic colorectal cancer and proposed as a potential biomarker for early tumor detection [26]. Importantly, neither of these regions showed enrichment for genes with high somatic mutation rates, and their average mutation burden was comparable to background levels. These findings support the idea that 8q24 and 20q13 primarily act as transcriptional regulatory hotspots rather than being involved in functional alterations due to somatic mutations.

Collectively, this study revealed a nonrandom distribution of CRC-associated GWAS polymorphisms across thirteen chromosomal bands. Enrichment analyses highlighted 8q24 and 20q13 as regulatory hotspots, enriched for differentially expressed genes and survival-associated pathways, but not for highly mutated genes. These findings underscore the functional and prognostic importance of these regions and suggest their potential utility as genetic biomarkers for CRC diagnosis, prognosis, and population-level screening.

Acknowledgements: I wish to express my sincere gratitude to Professor Mostafa Saadat for his invaluable guidance and support, which were instrumental in the advancement of this research.

Conflict of Interest: The author declares no conflicts of interest.

Ethics approval and consent to participate: Not required.

Authors' Contribution: SARS: Conceptualization, Methodology, Investigation, Formal Analysis, Data Curation, Visualization, Validation, and Writing.

REFERENCES

1. Burnett-Hartman AN, Lee JK, Demb J, Gupta S. An Update on the Epidemiology, Molecular Characterization, Diagnosis, and Screening Strategies for Early-Onset Colorectal Cancer. *Gastroenterology* 2021; 160:1041-1049.
2. Fernandez-Rozadilla C, Timofeeva M, Chen Z, Law P, Thomas M, Schmit S, *et al.* Deciphering colorectal cancer genetics through multi-omic analysis of 100,204 cases and 154,587 controls of European and east Asian ancestries. *Nat Genet* 2023; 55:89-99.
3. Huyghe JR, Bien SA, Harrison TA, Kang HM, Chen S, Schmit SL, *et al.* Discovery of common and rare genetic risk variants for colorectal cancer. *Nat Genet* 2019; 51:76-87.
4. Law PJ, Timofeeva M, Fernandez-Rozadilla C, Broderick P, Studd J, Fernandez-Tajes J, *et al.* Association analyses identify 31 new risk loci for colorectal cancer susceptibility. *Nat Commun* 2019; 10:2154.
5. Landi MT, Chatterjee N, Yu K, Goldin LR, Goldstein AM, Rotunno M, *et al.* A genome-wide association study of lung cancer identifies a region of chromosome 5p15 associated with risk for adenocarcinoma. *American journal of human genetics* 2009; 85 5:679-691.
6. Mahjoub G, Saadat M. Non-random distribution of gastric cancer susceptible loci on human chromosomes. *EXCLI J* 2018; 17:802-807.
7. Saify K, Saadat M. Non-random distribution of breast cancer susceptibility loci on human chromosomes. *Breast Cancer Res Treat* 2012; 136:315-318.
8. Popat S, Zhao D-b, Chen Z, Pan H, Shao Y-f, Chandler IP, Houlston RS. Relationship between chromosome 18q status and colorectal cancer prognosis: a prospective, blinded analysis of 280 patients. *Anticancer Res* 2007;27 1B:627-633.
9. van Dijk E, Biesma HD, Cordes M, Smeets D, Neerincx M, Das S, Eijk PP, Murphy V, Barat A, Bacon O, Prehn JHM, Betge J, Gaiser T, Fender B, Meijer GA, McNamara DA, Klinger R, Koopman M, Ebert MPA, Kay EW, Hennessey BT, Verheul HMW, Gallagher WM, O'Connor DP, Punt CJA, Loupakis F, Lambrechts D, Byrne AT, van Grieken NCT, Ylstra B. Loss of chromosome 18q11.2-q12.1 is predictive for survival in patients with metastatic colorectal cancer treated with bevacizumab. *J Clin Oncol* 2018; 36 20:2052-2060.
10. Tai JJ, Hou CD, Wang-Wuu S, Wang CH, Leu SY, Wu KD. A method for testing the nonrandomness of chromosomal breakpoints. *Cytogenet Cell Genet* 1993; 63:147-150.
11. Cheng X, Liu Y, Wang J, Chen Y, Robertson AG, Zhang X, Jones SJM, Taubert S. cSurvival: a web resource for biomarker interactions in cancer outcomes and in cell lines. *Brief Bioinform* 2022; 23.
12. Chandrashekar DS, Bashel B, Balasubramanya SAH, Creighton CJ, Ponce-Rodriguez I, Chakravarthi BV, Varambally S. UALCAN: a portal for facilitating tumor subgroup gene expression and survival analyses. *Neoplasia* 2017; 19:649-658.
13. Ahmadiyeh N, Pomerantz MM, Grisanzio C, Herman P, Jia L, Almendro V, He HH, Brown M, Liu XS, Davis M, Caswell JL, Beckwith CA, Hills A, Macconail L, Coetzee GA, Regan MM, Freedman ML. 8q24 prostate, breast, and colon cancer risk loci show tissue-specific long-range interaction with MYC. *Proc Natl Acad Sci USA* 2010; 107:9742-9746.

14. Ghoussaini M, Song H, Koessler T, Al Olama AA, Kote-Jarai Z, Driver KE, *et al.* Multiple loci with different cancer specificities within the 8q24 gene desert. *J Nat Cancer Inst* 2008; 100:962-966.
15. Saadat M. Distribution of preeclampsia-related genes on human chromosomes. *Taiwan J Obstet Gynecol* 2022;61:909-910.
16. Saadat M. Enrichment analysis and chromosomal distribution of gout susceptible loci identified by genome-wide association studies. *EXCLI J* 2023;22:1146-1154.
17. Bi S, Zhu J, Huang L, Feng W, Peng L, Leng L, Wang Y, Shan P, Kong W, Zhu S. Comprehensive analysis of the function and prognostic value of TAS2Rs family-related genes in colon cancer. *Int J Mol Sci* 2024;25.
18. Duan S, Guo W, Xu Z, He Y, Liang C, Mo Y, Wang Y, Xiong F, Guo C, Li Y, Li X, Li G, Zeng Z, Xiong W, Wang F. Natural killer group 2D receptor and its ligands in cancer immune escape. *Mol Cancer* 2019;18:29.
19. Fonseca A, Ramalhte SV, Mestre A, das Neves RP, Marreiros A, Castelo-Branco P, Roberto VP. Correction for: Identification of colorectal cancer associated biomarkers: an integrated analysis of miRNA expression. *Aging (Albany NY)* 2022; 14:2014.
20. Morita R, Hirohashi Y, Torigoe T, Ito-Inoda S, Takahashi A, Mariya T, Asanuma H, Tamura Y, Tsukahara T, Kanaseki T, Kubo T, Kutomi G, Mizuguchi T, Terui T, Ishitani K, Hashino S, Kondo T, Minagawa N, Takahashi N, Taketomi A, Todo S, Asaka M, Sato N. Olfactory receptor family 7 subfamily C member 1 is a novel marker of colon cancer-initiating cells and is a potent target of immunotherapy. *Clin Cancer Res* 2016;22:3298-3309.
21. Pan C, Li B, Simon MC. Moonlighting functions of metabolic enzymes and metabolites in cancer. *Molecular Cell* 2021; 81:3760-3774.
22. Rasool M, Karim S, Haque A, Alharthi M, Chaudhary AG, Natesan Pushparaj P. Deciphering gene expression signatures in liver metastasized colorectal cancer in stage IV colorectal cancer patients. *J King Saud University - Science* 2024;36:103415.
23. Najafi S, Salehi Z, Mashayekhi F, Saidi-Saedi H. Expression patterns of circRFX3 and miR-587 in colorectal cancer patients. *Mol Biol Res Commun* 2025;14:243-248.
24. Dang CV. MYC on the path to cancer. *Cell* 2012; 149:22-35.
25. Aust DE, Muders M, Köhler A, Schmidt M, Diebold J, Müller C, Löhrs U, Waldman FM, Baretton GB. Prognostic relevance of 20q13 gains in sporadic colorectal cancers: a FISH analysis. *Scand J Gastroenterol* 2004; 39:766-772.
26. Bui V-M-H, Mettling C, Jou J, Sun HS. Genomic amplification of chromosome 20q13.33 is the early biomarker for the development of sporadic colorectal carcinoma. *BMC Medical Genomics* 2020; 13:149.
27. Kobayashi Y, Esato K, Noshima S, Oga A, Sasaki K. Detection of 20q13 gain by dual-color FISH in breast cancers. *Anticancer Res* 2000; 20:531-535.