

RESEARCH ARTICLE

Novel diagnostic and prognostic biomarkers of colorectal cancer: Capable to overcome the heterogeneity-specific barrier and valid for global applications

Yasir Hameed¹, Muhammad Usman¹, Shufang Liang², Samina Ejaz^{3*}

1 Department of Biotechnology, Institute of Biochemistry, Biotechnology and Bioinformatics, The Islamia University of Bahawalpur, Bahawalpur, Pakistan, **2** State Key Laboratory of Biotherapy and Cancer Center, West China Hospital, Sichuan University and Collaborative Innovation Center for Biotherapy, Chengdu, 610041, P.R. China, **3** Department of Biochemistry, Institute of Biochemistry, Biotechnology and Bioinformatics, The Islamia University of Bahawalpur, Bahawalpur, Pakistan

* saminaejazsyed@yahoo.com, samina.ejazsyed@iub.edu.pk



OPEN ACCESS

Citation: Hameed Y, Usman M, Liang S, Ejaz S (2021) Novel diagnostic and prognostic biomarkers of colorectal cancer: Capable to overcome the heterogeneity-specific barrier and valid for global applications. PLoS ONE 16(9): e0256020. <https://doi.org/10.1371/journal.pone.0256020>

Editor: Sripathi M. Sureban, University of Oklahoma Health Sciences Center, UNITED STATES

Received: December 27, 2020

Accepted: July 28, 2021

Published: September 2, 2021

Copyright: © 2021 Hameed 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 datasets underlying the results presented in the study are available in Gene Expression Omnibus (GEO) database (available at: <https://www.ncbi.nlm.nih.gov/geo/>) and can be retrieved through accession numbers, GSE17538, GSE29623, GSE81558, GSE21510, GSE44076, GSE24514, GSE32323, GSE110225, GSE20916, GSE73360, GSE44861, GSE23878, GSE41258, GSE110224, GSE35279, GSE87211, GSE117606, GSE28000, GSE21815,

Abstract

Introduction

The heterogeneity-specific nature of the available colorectal cancer (CRC) biomarkers is significantly contributing to the cancer-associated high mortality rate worldwide. Hence, this study was initiated to investigate a system of novel CRC biomarkers that could commonly be employed to the CRC patients and helpful to overcome the heterogenetic-specific barrier.

Methods

Initially, CRC-related hub genes were extracted through PubMed based literature mining. A protein-protein interaction (PPI) network of the extracted hub genes was constructed and analyzed to identify few more closely CRC-related hub genes (real hub genes). Later, a comprehensive bioinformatics approach was applied to uncover the diagnostic and prognostic role of the identified real hub genes in CRC patients of various clinicopathological features.

Results

Out of 210 collected hub genes, in total 6 genes (CXCL12, CXCL8, AGT, GNB1, GNG4, and CXCL1) were identified as the real hub genes. We further revealed that all the six real hub genes were significantly dysregulated in colon adenocarcinoma (COAD) patients of various clinicopathological features including different races, cancer stages, genders, age groups, and body weights. Additionally, the dysregulation of real hub genes has shown different abnormal correlations with many other parameters including promoter methylation, overall survival (OS), genetic alterations and copy number variations (CNVs), and CD8+T immune cells level. Finally, we identified a potential miRNA and various chemotherapeutic drugs via miRNA, and real hub genes drug interaction network that could be used in the treatment of CRC by regulating the expression of real hub genes.

GSE75970, GSE39582, GSE9348, GSE22598, GSE113513, GSE74602, GSE101502, GSE2509, GSE7621, GSE14333, GSE4183, and GSE4107.

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

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

Conclusion

In conclusion, we have identified six real hub genes as potential biomarkers of CRC patients that could help to overcome the heterogeneity-specific barrier across different clinicopathological features.

1. Introduction

Colorectal cancer (CRC) is the most common cancer and is one of the leading causes of cancer-related deaths worldwide [1]. Genetic mutations and altered expression levels of various tumor suppressor genes including Adenomatous Polyposis Coli (APC), Tumor Protein 53 (TP53) and Breast Cancer gene (BRCA1) have long been considered as an important cause of CRC, which affect the development, progression, and metastasis of CRC through a variety of regulatory pathways [2].

Advances in biomarkers screening technologies have greatly helped to discover the reliable novel diagnostic and prognostic biomarkers for the early detection and treatments of CRC. However, due to the heterogeneity-specific nature of the available diagnostic and prognostic biomarkers the recurrence and metastasis of CRC in patients of different races, different cancer stages, genders, age groups, and body weight are still, not fully addressed and remain the major challenge to clinical treatment [3,4].

In the recent times, microarray technology has become very popular among the scientist as it enables them to screen thousands of differentially expressed messenger RNAs (mRNAs), microRNAs (miRNAs), and long noncoding RNAs (lncRNAs) together [5–7], which play an important role in the development and progression of a disease. Besides, this technology has also helped to perform the in-depth analyses of the key genes to explore potential molecular targets and diagnostic biomarkers [8].

The Gene Expression Omnibus database (GEO, available at: <http://www.ncbi.nlm.nih.gov/geo/>) is an open-access gene expression database, which is created, and maintained by the National Center for Biotechnology Information (NCBI) [9,10]. This database is used to store and freely distribute the microarrays, next generation sequencing, and various other forms of high-throughput functional genomic datasets to the researchers worldwide [10].

GEO database is the most attractive platform for the researchers to re-evaluate and re-analyze the microarray datasets through different bioinformatics approaches for the identification of disease-specific novel potential biomarkers. Many researchers have previously utilized the CRC microarray expression datasets to identify the potential biomarkers as hub genes. However, by considering the fact that, biomarkers are highly race, cancer stage, genders, age, and body weight-specific biomolecules and knowing that CRC microarray expression datasets, available in the GEO database, contain information extracted from patients of different races, different cancer stages, genders, age groups, and body weights. Moreover, on daily basis new datasets are added to the GEO database and each dataset suggests different CRC related hub genes. Hence, it has become clinically impossible to globally employ the hub genes which have been reported by individual studies through extensive analysis of specific GEO's-CRC related microarray expression datasets [11] as potential diagnostic and prognostic biomarkers to the CRC patients of all the races, cancer stages, genders, age groups, and body weights. Therefore, in this study, we planned to re-analyze the already reported CRC-related hub genes through a multi-layered bioinformatics based strategy to detect few more closely CRC-linked hub genes

(real hub genes) that could commonly be used as potential diagnostic and prognostic biomarkers for CRC patients of different clinicopathological features.

For this purpose, the already reported hub genes will be extracted from those published studies that utilized the CRC-related GEO expression datasets. Then, all these hub genes will be added to a single pool to establish a consolidated set of most significantly dysregulated hub genes exhibiting the high degree of centrality among the analyzed genes. Followed by this step, the pool of hub genes will be subjected to pathway enrichment analysis, PPI network construction and identification of the most centralized genes (real hub genes) and their underlying pathways [12]. Next, the differential expression and validation analysis of the identified real hub genes in normal and CRC patients of different clinicopathological features will be carried out via multiple authentic platforms such as GEPIA database [13], GENT2 database, UALCAN database [14] and cBioPortal database [15] using numerous TCGA colon adenocarcinoma (COAD) datasets consisting of a large cohort of normal individuals and COAD patients. Following this, we will investigate the correlation of the real hub genes expressions with their promoter methylation level, genetic alterations, copy number variations (CNVs), overall survival (OS) and CD8+ T immune cells levels in CRC patients relative to normal controls. Additionally, in an attempt to understand the regulatory mechanisms, miRNAs interaction patterns will be studied to explore the role of miRNAs, if any, as mediators of the real hub genes' expression behavior. Similarly the effect of various chemotherapeutic drugs on expression profile of the identified real hub genes will be deduced through hub genes-drug interaction network analysis. The information obtained could help to regulate the real hub genes expression during the treatment of CRC.

Taken together, this detailed mega-scale study based upon information retrieved from the analysis of large number of datasets and reported hub genes is therefore expected to find some common CRC biomarkers which can be exploited for the diagnostic and prognostic purposes of worldwide CRC patients of different clinicopathological features and thus helpful to overcome the heterogenetic-specific barrier. The information retrieved can further help in predicting the treatment outcomes in CRC patients.

2. Material and methods

2.1 Literature search and hub genes extraction

A PubMed based search was performed to search all the studies which analyzed the CRC expression microarray datasets available on GEO (available at: <https://www.ncbi.nlm.nih.gov/geo/>) database and identified the hub genes. For this purpose, two keywords “Hub genes AND colon cancer” and “Hub genes AND colorectal cancer” were searched on PubMed separately with the “Original article” filter. In total 108 studies appeared which were further explored to filter out the studies having desired information. In total, 21 studies were selected which were found to collectively analyze more than 30 CRC expression based microarray datasets retrieved from the GEO database and identify numerous hub genes. We extracted all the hub genes from these studies and assembled them in the form of a single pool.

2.2 Pathway enrichment analysis

The Database for Annotation, Visualization and Integrated Discovery (DAVID, available at: <https://david.ncifcrf.gov>) integrates biological data as well as analytical tools to systematically and comprehensively annotate the biological functions [16]. We used the DAVID database for pathway enrichment analysis of the pooled hub genes. A p-value < 0.05 indicated the statistically significant differences.

2.3 Protein–protein interaction (PPI) network construction and mining the real hub genes

The protein–protein interaction (PPI) analysis is important to interpret the molecular mechanisms of the key pathways in carcinogenesis. In the present study, we utilized the Search Tool for the Retrieval of Interacting Genes (STRING) database (available at: <https://string-db.org/>) [17] to construct the PPI network of all the pooled hub genes. The six real hub genes present in the PPI network were then identified through Cytohubba application of the Cytoscape tool (version:3.7.1) [18], which can explore important nodes and fragile motifs in a network by several topological algorithms including degree-edge percolated component and degree of centrality.

2.4 GEPIA dataset analysis

GEPIA (available at: <http://gepia.cancer-pku.cn/>) is an online platform of retrieved data from the UCSC Xena database (available at: <https://xena.ucsc.edu/>), which in-houses the expression data of 9736 tumor samples and 8587 normal samples [13]. In this study, the transcriptional expression levels of the real hub genes were analyzed in COAD patients relative to control. For this purpose the Colon adenocarcinoma (COAD) dataset was utilized which includes 275 tumor and 349 normal samples. A t-test is used for the statistics purpose in GEPIA. The expression level of real hub genes in GEPIA was normalized as transcript per million (TPM) reads, and a p-value < 0.05 was considered to be statistically significant. We also utilized this database for the correlational analysis between real hub genes expression and OS duration of the COAD patients.

2.5 GENT2 dataset analysis

Gene Expression database of Normal and Tumor tissues 2 (GENT2, available at: <http://gent2.apex.kr>) is an online platform that provides a user-friendly overview of the gene expression patterns across different normal and tumor tissues compiled from publically available GEO datasets. GENT2 contains the expression data of more than 68,000 samples and has several useful functions. For example, GENT2 provides gene expression analysis option across 72 different cancerous tissues. GENT2 also provides an option to study the differential expression and its prognostic significance based on tumor subtypes. Additionally, GENT2 provides a meta-analysis of survival information to provide users more reliable prognostic value of a gene of interest. A t-test is used for the statistics purpose in GENT2. In the present study, this platform was used for further validation of the GEPIA based results of real hub genes expression patterns examined in COAD patients relative to controls [19]. The expression level of real hub genes in GENT2 was normalized as transcript per million (TPM) reads, and a p-value of < 0.05 was considered to be statistically significant.

2.6 UALCAN dataset analysis

The UALCAN (available at; <http://ualcan.path.uab.edu/>) is an online publicly available web-portal that offers in-depth analysis of data from TCGA. In the present study, this database was used for the genes promoters' methylation analysis and validation of variations detected in the real hub genes' mRNA and protein expression profiles of COAD patients of different clinicopathological features relative to normal controls. In UALCAN t-test was used for the statistics purpose. The mRNAs' expression levels of real hub genes were normalized as transcript per million (TPM) reads. While corresponding proteins expression levels were normalized as z-value, the promoters' methylation levels were normalized as beta (β) value, and a p-value of < 0.05 value was considered to be statistically significant.

2.7 cBioPortal analyses

An open-source tool, cBioPortal (Available at: <http://www.cbioportal.org>) developed by the Computation Biology Center located at Sloan Kettering, was utilized to summarize all the possible transcriptional changes, mutual expression tendencies, and overall survival through Kaplan-Meier analysis, by presenting the results as OncoPrint. In this study, the cBioPortal database was used to analyze genetic variations such as (amplifications, deep deletions, and mutations) in the real hub genes in COAD patients.

2.8 Real hub genes and infiltrating level of CD8+ T cells in COAD patients

TIMER (available at: <https://cistrome.shinyapps.io/timer/>) is a web resource for systematic evaluations of the clinical impact of different immune cells in diverse cancer types [20]. In the present study, this database was used to find the Spearman correlation between the levels of real hub genes' expression and CD8+ T immune cells. In TIMER a t-test was used for the statistics purpose and a p-value of < 0.05 was considered to be statistically significant.

2.9 The miRNA–real hub gene interaction network analysis

The miRNAs, targeting the real hub genes, were predicted through miRNA target prediction databases “Regulatory Network Repository of Transcription Factor and microRNA Mediated Gene Regulations (RegNetwork) database” (available at: <http://regnetworkweb.org/>). The RegNetwork database contains information of experimentally validated regulatory elements of gene expression including transcription factors (TFs) and miRNAs. This platform provides a user-friendly interface for the submission of query of interest and allows the finding of combinatorial and synergic regulatory relationships among TFs, miRNAs and genes [21]. A co-expression network based on the correlation analysis of real hub genes and miRNAs associated with the cancer was then developed by Cytoscape software. In the network, interaction between the miRNAs and real hub gene was represented by an arrow. The numbers of arrows in the networks indicated the contribution of one miRNA in the expression regulation of the surrounding genes.

2.10 Real hub gene-drug interaction network analysis

The Comparative Toxicogenomics Database (CTD, available at: <http://ctdbase.org/>) has been employed to obtain the information of chemotherapeutic drugs that could reduce or enhance the mRNAs or proteins expression levels of the genes of interest [22]. Briefly, all the real hub genes were searched in the CTD database, and hub gene-drug interaction networks were visualized using Cytoscape software.

3. Results

3.1 Literature search and hub genes extraction

In total 21 studies were selected, some of them have identified the hub genes in individual microarray dataset of CRC [23–25] while others have used the combination of multiple microarray datasets of CRC [26–28]. We extracted all hub genes reported in literature and pooled them after normalizing the duplicated genes, hence, a pool of 210 hub genes from 31 microarray datasets, containing 3128 CRC and 877 normal samples, were selected for further analysis (Table 1). Raw data without normalization can be seen in S1 Table (Supporting Information).

Table 1. Details of CRC microarray based expression datasets and the identified hub genes.

Dataset	No. samples C/N	Source of origin	Extracted hub genes	References
GSE17538 GSE29623	244/0	USA	HCLS1, EVI2B, CD48, DSN1, ALDH1A1, TUBAL3, RRM2, SHMT2, PGM1, CCT6A, IDH3A, HSPA2, UGP2, RUVBL1, CDK1, CCNB1, MAD2L1, AHCY, HSPH1, BUB1B, DARS, KIF18A, NUF2, CENPF, ERCC6L, SPC25, CXCL1, CXCL12, SST, NPY, PPY, LPAR1, CCL19, GNG4, CXCL5, CXCL2, CXCL3, NPY1R, CCL28, TIMP1, SPP1, MMP3, GCG, KIT, BMP2, COL1A1, CXCL8, CXCL11, NMU, PPBP, COL3A1, TGFR1, CD44, MMP1, SPARC, MYC, COL1A2, CA2, SERPINB5, CCND1, PTGS2, MET, GDF15, ECT2, CKAP2, CSE1L, ABCE1, RFC3, TPx2, SLC4A4, SLC26A3, CLCA4, SLC25A2, CEACAM7, PDGFRB, FZD2, PRKCB, FPR2, KIF2C, KIF20A, DLGAP5, NCAPG, UBE2C, EXO1, CDC45, CXCL10, DTL, PF4, CEP55, P2RY14, ADNP, CDK4, CEBPB, CENPA, CENPH, CENPN, RFC2, SSTR1, SSTR2, HCAR3, APLN, CXCR2, SAA1, PMCH, GAL, CXCR1, CCL23, CXCL6, HTR1D, GALR1, CNR1, AGT, FPR1, PTGDR2, CCR8, INSL5, F2RL2, GCC, GRP, OXTR, GPR4, NPSR1, UTS2B, PROK2, AGTR1, EDN3, CHRM1, DK1, CCNA2, PLK1, KIF11, MELK, NUSAP1, MCM4, RFC4, PTTG1, CHEK1, CENPE, ITGA2, BGN, SULF1, FA, THBS2, CTHRC1, COL5A2, AQP8, CLCA4, GUCA2B, MS4A12, GUCA2A, ABCG2, CLDN8, ZG16, PKIB, CA4, BEST4, CA1, MT1M, CD177, HSD17B2, ADH1C, CLCA1, FOXQ, KRT23, LY6G6D, MMP7, CDH3, CST1, CRNDE, DPEP1, EPHX4, CLDN1, CEL, CLDN2, SLC35D3, COL11A1, SLC01B3, CKMT2, NUDT21, GNB1, CLINT1, EGFR, HRas, Akt1, CDKN1a, PCNA, NAT1, NAT2, PLAGL2, POFUT1, TOP2A, ACLY, VEDFA, GMP5, ENO1, ACTA2, AURKA, CDC42, TEX11, QKI, CAV, FN1, ARHGEP6, JUP, WNT2, WNT5A, WNT11, PYY, CCL20	[11,23–25,27–43]
GSE81558 GSE21510	130/0	USA		
GSE44076	42/9	Spain		
GSE24514 GSE32323	148/0	Japan		
GSE110225	98/148	Spain		
GSE20916 GSE73360	37/12	Finland		
GSE44861	28/16	Japan		
GSE23878	30/30	Greece		
GSE41258	108/37	Poland		
GSE110224	62/30	Italy		
GSE35279	111/111	USA		
GSE87211	35/24	Saudi Arabia		
GSE117606	328/62	Israel		
GSE28000 GSE21815	17/17	Greece		
GSE75970	74/5	Japan		
GSE39582	230/133	USA		
GSE9348 GSE22598	102/102	Belgium		
GSE113513	86/26	USA		
GSE74602	123/9	Japan		
GSE101502	8/0	China		
GSE2509	585/0	France		
GSE7621	70/12	Singapore		
GSE14333	18/18	Japan		
GSE4183	14/14	China		
GSE4107	30/30	Singapore		
	3/3	China		
	6/0	Italy		
	16/9	USA		
	290/0	Australia		
	45/8	Hungary		
	10/12	Singapore		
Total = 31	Total = 3128/877		Total = 210	

C = Cancerous, N = Normal, USA = United States of America.

<https://doi.org/10.1371/journal.pone.0256020.t001>

3.2 KEGG pathway analysis of the pooled hub genes

KEGG pathway enrichment analysis suggested that all the 210 pooled hub genes were significantly enriched in pathways including ‘Chemokine signaling pathway’, ‘Pathways in cancer’, ‘Cell cycle’, ‘PI3K-Akt signaling pathway’, and ‘Cytokine-cytokine receptor interaction pathway’. The Top 10 KEGG terms related to the hub genes are enlisted in Fig 1 and Table 2).

3.3 PPI network construction, real hub genes selection and their pathway enrichment

A PPI network of all the 210 pooled hub genes was constructed through an online available STRING database containing 204 nodes and 2561 edges. A Cytohubba analysis was carried out through Cytoscape software to identify more closely CRC relevant few genes (real hub genes). Based on the degree of centrality, the group of the identified real six hub genes included CXCL12, CXCL8, AGT, GNB1, GNG4, and CXCL1 (Fig 2 and Table 3).

Furthermore, the pathway enrichment analysis of the identified real hub, using David tool, has shown the significant ($p < 0.05$) involvement of five genes in “Chemokine signaling pathway”, “Pathways in cancer”, and “Cytokine-cytokine receptor interaction” pathways (Table 4).

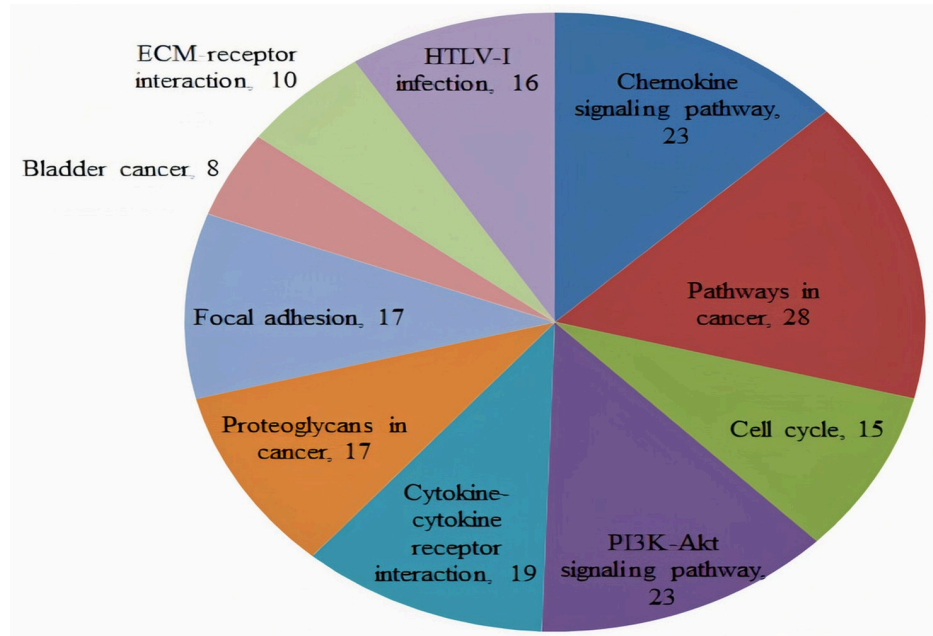


Fig 1. A heatmap of the KEGG pathways to portray role of all the CRC related hub genes (n = 210) identified during present study.

<https://doi.org/10.1371/journal.pone.0256020.g001>

3.4 Bioinformatics based expression analysis of the identified real hub genes in normal individuals and colon adenocarcinoma (COAD) patients

In order to analyze and validate the differential mRNA expression of the identified real hub genes in normal and COAD patients of different clinicopathological features (different race, cancer stages, genders, age groups, and body weights), a detailed bioinformatics analysis was

Table 2. KEGG pathway analysis details of the 210 pooled hub genes, extracted from the various GEO microarray CRC expression datasets.

Pathway ID	Pathway Name	Gene count	p-value	Gene name
04062	Chemokine signaling pathway	23	<0.05	CXCL1, HRAS, CXCL5, CXCL3, CXCL2, CXCR1, CXCL8, CCL19, CXCR2, PF4, CXCL6, CXCL11, CCL28, CXCL12, CXCL10, AKT1, CDC42, CCR8, CCL23, PPBP, CCL20, GNB1, GNG4
05200	Pathways in cancer	28	<0.05	WNT5A, HRAS, PTGS2, CXCL8, LPAR1, KIT, CXCL12, MMP1, WNT2, AKT1, AGTR1, CDC42, GNG4, MYC, FN1, EGFR, BMP2, MET, ITGA2, FZD2, CDK4, PRKCB, JUP, CDKN1A, CCND1, GNB1, PDGFRB, WNT11
04110	Cell cycle	15	<0.05	CDK1, CHEK1, PTTG1, CDK4, MCM4, CCNB1, CDC45, CDKN1A, CCND1, MAD2L1, PLK1, PCNA, BUB1B, MYC, CCNA2
04151	PI3K-Akt signaling pathway	23	<0.05	EGFR, HRAS, COL3A1, MET, ITGA2, LPAR1, KIT, CDK4, COL5A2, AKT1, CCND1, CDKN1A, GNB1, CHRM1, COL1A2, PDGFRB, COL1A1, GNG4, THBS2, MYC, COL11A1, SPP1, FN1
04060	Cytokine-cytokine receptor interaction	19	<0.05	CXCL1, BMP2, CXCL5, CXCL3, CXCL2, CXCR1, CXCL8, CCL19, CXCR2, PF4, CXCL6, CXCL11, CCL28, CXCL12, CXCL10, CCR8, CCL23, PPBP, CCL20
05205	Proteoglycans in cancer	17	<0.05	EGFR, WNT5A, HRAS, HCLS1, MET, ITGA2, FZD2, PRKCB, AKT1, WNT2, CDC42, CDKN1A, CCND1, CD44, WNT11, MYC, FN1
04510	Focal adhesion	17	<0.05	EGFR, HRAS, MET, COL3A1, ITGA2, COL5A2, PRKCB, AKT1, CDC42, CCND1, COL1A2, PDGFRB, COL1A1, THBS2, COL11A1, SPP1, FN1
05219	Bladder cancer	8	<0.05	EGFR, CDKN1A, HRAS, CCND1, CXCL8, CDK4, MYC, MMP1
04512	ECM-receptor interaction	10	<0.05	CD44, COL3A1, COL1A2, ITGA2, COL1A1, COL11A1, THBS2, COL5A2, SPP1, FN1
05166	HTLV-I infection	16	<0.05	WNT5A, HRAS, CHEK1, PTTG1, FZD2, CDK4, WNT2, AKT1, CDKN1A, CCND1, MAD2L1, PCNA, PDGFRB, BUB1B, WNT11, and MYC

<https://doi.org/10.1371/journal.pone.0256020.t002>

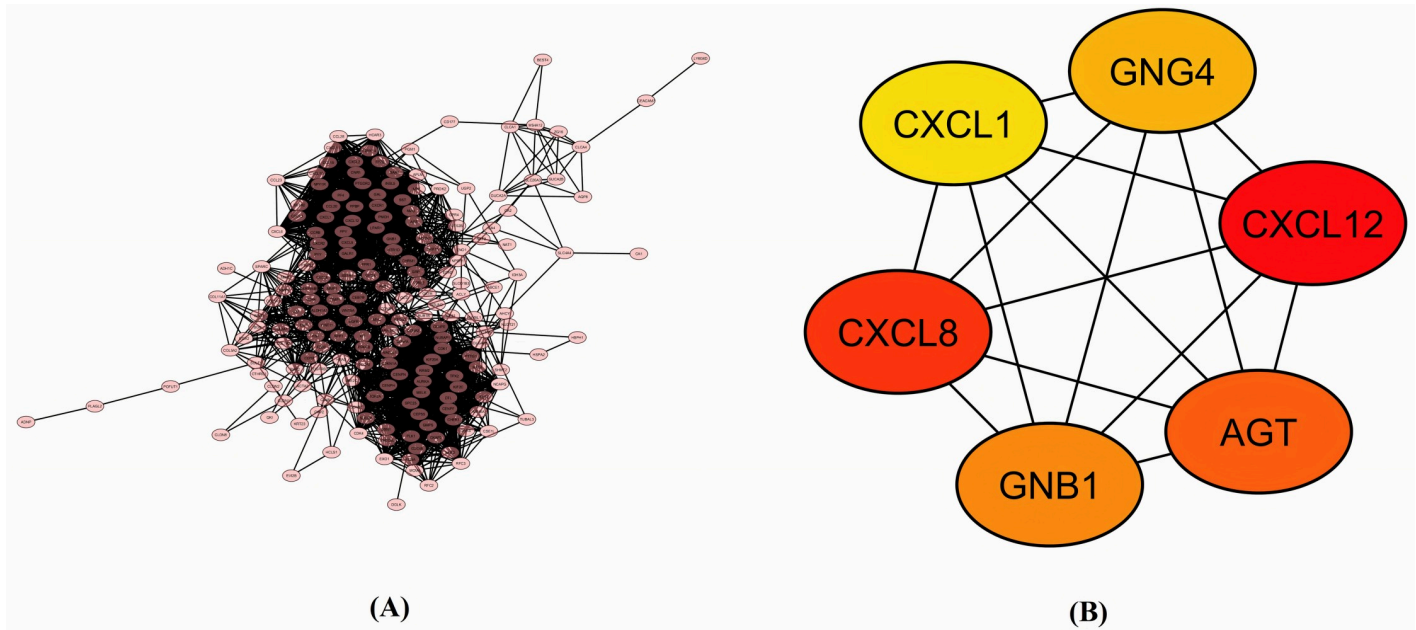


Fig 2. (A) A PPI network of all the 210 extracted hub genes. (B) A network of six real hub genes identified on the basis of degree of centrality.

<https://doi.org/10.1371/journal.pone.0256020.g002>

carried out. For the said purpose we considered three different online available platforms including, GEPIA; it retrieved mRNA data from UCSC Xena server which contained 275 COAD samples paired with 349 normal samples and used to analyze the mRNA expression of the real hub genes in the present study, GENT2; it retrieved mRNA data from GEO database which contained 477 COAD samples paired with 91 normal samples and utilized to validate the real hub genes expression in the present study, and finally the UALCAN; it retrieved mRNA data from TCGA database which contained 286 COAD samples paired with 41 normal samples and used to validate the real hub genes expression at protein level (in overall COAD cases relative to controls) and at mRNA level (in COAD patients of different clinicopathological features relative to controls).

Taken together the results of these three databases, we observed and validated the significant ($p < 0.05$) down-regulation of CXCL12 while significant ($p < 0.05$) overexpression of CXCL12, CXCL8, AGT, GNB1, GNG4, and CXCL1 real hub genes at mRNA level in COAD patients of different clinicopathological features (patients race, cancer stages, genders, age

Table 3. List of the real hub genes identified from the PPI network of the extracted 210 CRC related hub genes.

Sr. No	Name of the gene	Degree of centrality	No. Nodes	Closeness of centrality
1	CXCL12	68	68	0.50
2	CXCL8	66	66	0.53
3	AGT	61	61	0.47
4	GNB1	60	60	0.49
5	GNG4	59	59	0.47
6	CXCL1	56	56	0.47

Degree of centrality = It is the number of links incident upon a node.

Closeness of centrality = It is a measure of the average shortest distance from one node to other node.

<https://doi.org/10.1371/journal.pone.0256020.t003>

Table 4. Details of the KEGG pathway analysis of the identified real hub genes.

Pathway ID	Pathway Name	Gene count	p-value	Gene name
04062	Chemokine signaling pathway	05	<0.05	CXCL1, GNB1, CXCL8, GNG4, CXCL12
05200	Pathways in cancer	04	<0.05	GNB1, CXCL8, GNG4, CXCL12
04060	Cytokine-cytokine receptor interaction	03	<0.05	CXCL1, CXCL8, CXCL12

<https://doi.org/10.1371/journal.pone.0256020.t004>

groups, and body weight) relative to the normal controls. Moreover, at protein level, UALCAN based results confirmed the significant ($p < 0.05$) overexpression of CXCL8, GNG4, and CXCL1 while the insignificant ($p > 0.05$) down-regulation of CXCL12, AGT, and GNB1 (Figs 3–10). This scenario of inverse correlation between mRNA and protein expression levels of AGT, and GNB1 indicates the abnormal posttranscriptional regulation which probably decreases the half-life of these proteins and results in their down-regulation.

3.5 Promoters methylation levels of the real hub genes in normal individuals and COAD patients

The variations in the degree of promoter's methylation have regulatory impact on the expression behavior of the onco and proto-onco genes and thus have been closely linked with cancer development [44]. To document the variations, if any, examined in COAD patients as compared to the normal controls we studied the promoters' methylation levels of the real hub genes via UALCAN platform containing information of 13 normal and 313 COAD samples. A diverse pattern of real hub genes promoters' methylation levels were observed in COAD patients relative to normal controls. Approximately 30%, 20%, 60%, 57%, 56%, and 55% COAD patients have had the similar promoter methylation levels of CXCL12, CXCL8, AGT, GNB1, GNG4, and CXCL1, respectively, as examined in case of the normal controls. Moreover, 15%, 10%, 20%, 15%, 24%, and 45% COAD patients were found positive for the hypomethylation of CXCL12, CXCL8, AGT, GNB1, GNG4, and CXCL1 promoters, respectively, relative to normal controls. While the remaining 55%, 70%, 20%, 38%, 20%, and 0% COAD patients were found positive for the hypermethylation of CXCL12, CXCL8, AGT, GNB1, GNG4, and CXCL1 promoters, respectively, relative to normal controls. Ultimately, the overall comparison between COAD and normal groups by applying the t-test revealed that CXCL12, CXCL8, AGT, GNB1, and GNG4) were significantly ($p < 0.05$) hypermethylated while CXCL1

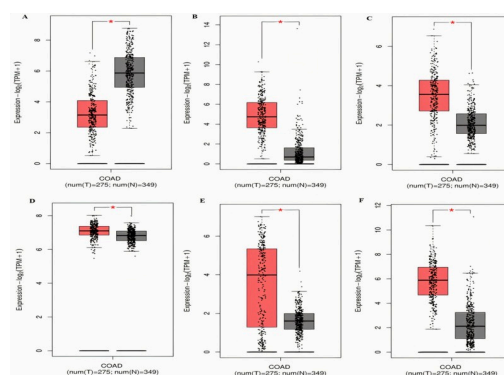


Fig 3. Box plots showing the relative mRNA expression levels of real hub genes in normal and COAD patients (Information retrieved via GEPIA database). The box plots shows the relative mRNA expression of: (A) CXCL12, (B) CXCL8, (C) AGT, (D) GNB1, (E) GNG4, and (F) CXCL1; in COAD patients and normal samples. A p-value of < 0.05 was considered significant.

<https://doi.org/10.1371/journal.pone.0256020.g003>

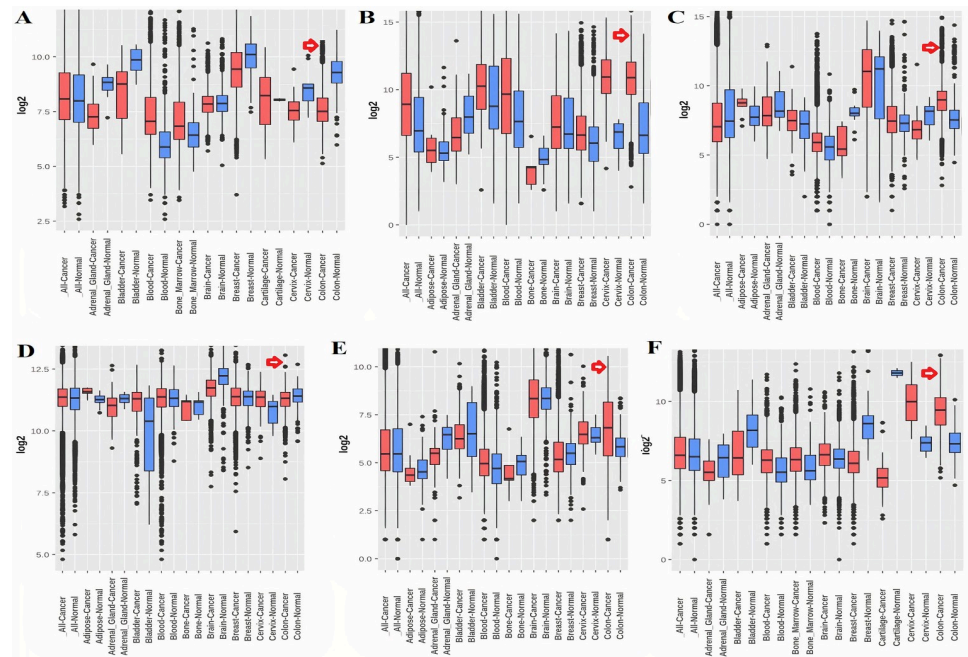


Fig 4. Box plots showing the relative expression levels of real hub genes in normal and COAD patients (Information retrieved from GENT2 database). The box plots shows the relative mRNA expression of: (A) CXCL12, (B) CXCL8, (C) AGT, (D) GNB1, (E) GNG4, and (F) CXCL1; in COAD patients and normal samples. A p-value of <0.05 was considered significant.

<https://doi.org/10.1371/journal.pone.0256020.g004>

was significantly ($p < 0.05$) hypomethylated in COAD samples relative to normal controls (Fig 11). Collectively, these results suggested that promoter methylation level positively correlates with the mRNA expression level of CXCL8, AGT, GNB1, and GNG4 while negatively correlates with the mRNA expression level of CXCL1 and CXCL12 in COAD samples.

3.6 Amplification, deletion, mutation, and fusion of real hub genes in COAD

Genetic alteration and CNVs are the common sources of gene expression dysregulation [45]. In this study, information related to the real hub genes-related genetic alterations and CNVs information were obtained from three different TCGA COAD datasets (TCGA firehose legacy, TCGA nature 2012 and TCGA PanCancer Atlas containing data of 619, 266 and 87 COAD samples, respectively), via cBioPortal platform. Results of the analysis revealed the varying degrees of genetic alterations and CNVs among all the real hub genes in 1482 analyzed COAD samples. Out of which AGT displayed the highest incidence rate of genetic alterations and CNVs (2.2%, 32/1482) followed by the GNB1 which showed the incidence rate of 1.4% (21/1482). Following GNB1, the other real hub genes including GNG4, CXCL8, CXCL1, and CXCL12 showed the 1% (15/1484), 0.6% (15/1482), 0.4% (6/1482) and, and 0.2% (3/1482) incidence rate of genetic alterations and CNVs in COAD samples (Fig 12). In CXCL8 and AGT, missense mutations accounted for most of the changes, while in CXCL12 and GNB1, deep deletions were identified as the most frequent changes. Moreover, in case of GNG4 and CXCL1 deep amplifications were highlighted as the most common changes in the queried samples.

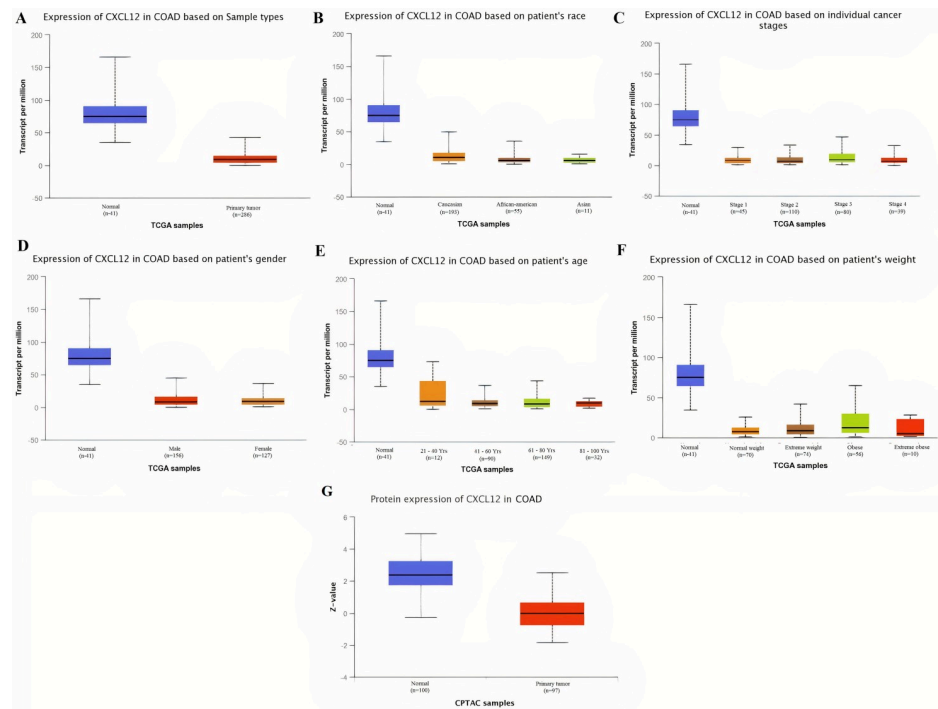


Fig 5. Box plots showing the relative expression levels of CXCL12 in normal and COAD samples of different clinicopathological features via UALCAN database. Relative mRNA expression of CXCL12: (A) in normal individuals and COAD patients; (B) in normal individuals and COAD patients of different races; (C) in normal individuals and COAD patients of different cancer stages; (D) in normal individuals and COAD patients of different genders; (E) in normal individuals and COAD patients of different age groups; and, (F) in normal individuals and COAD patients of different body weights; (G) Relative protein expression of CXCL12 in normal individuals and COAD patients. A p-value of <0.05 was considered significant. Normal weight = BMI greater than or equal to 18.5 and BMI less than 25, Extreme weight = BMI greater than or equal to 25 and BMI less than 30, Obese = BMI greater than or equal to 30 and BMI less than 40, and Extreme Obese = BMI greater than 40.

<https://doi.org/10.1371/journal.pone.0256020.g005>

3.7 Prognostic values of the real hub genes in the COAD patients

Correlation analysis among the mRNA's expression levels of the real hub genes and the overall survival (OS) of the COAD patients was performed via GEPIA database and using data of 275 normal and 349 COAD samples. Results revealed that the higher mRNAs' expression levels of GNG4 [HR: 1.7, $p>0.05$] are significantly correlated with the reduced OS duration of the COAD patients, therefore, suggested as good prognostic biomarker for predicting the OS duration of the COAD patients. Results also revealed that down-regulation of CXCL12 [HR: 1.1, $p>0.05$] and up-regulations of CXCL8 [HR: 0.61, $p>0.05$], AGT [HR: 0.64, $p>0.05$], GNB1 [HR: 0.57, $p<0.05$], and CXCL1 [HR: 0.69, $p<0.05$] are the bad prognostic biomarkers for predicting the OS duration of the COAD patients as they do not correlated with the reduced OS duration (Fig 13).

3.8 Real hub genes and infiltrating levels of CD8+ T cells in COAD patients

The functions of, and interactions between, the innate and adaptive immune systems are vital for the anticancer immunity. Cytotoxic T cells expressing cell-surface CD8 are the most powerful effectors in the anticancer immune response and form the backbone of current successful cancer immunotherapies [46]. In the current study, the Spearman correlation between the expression of real hub genes and CD8+ T cell infiltration was calculated using TIMER

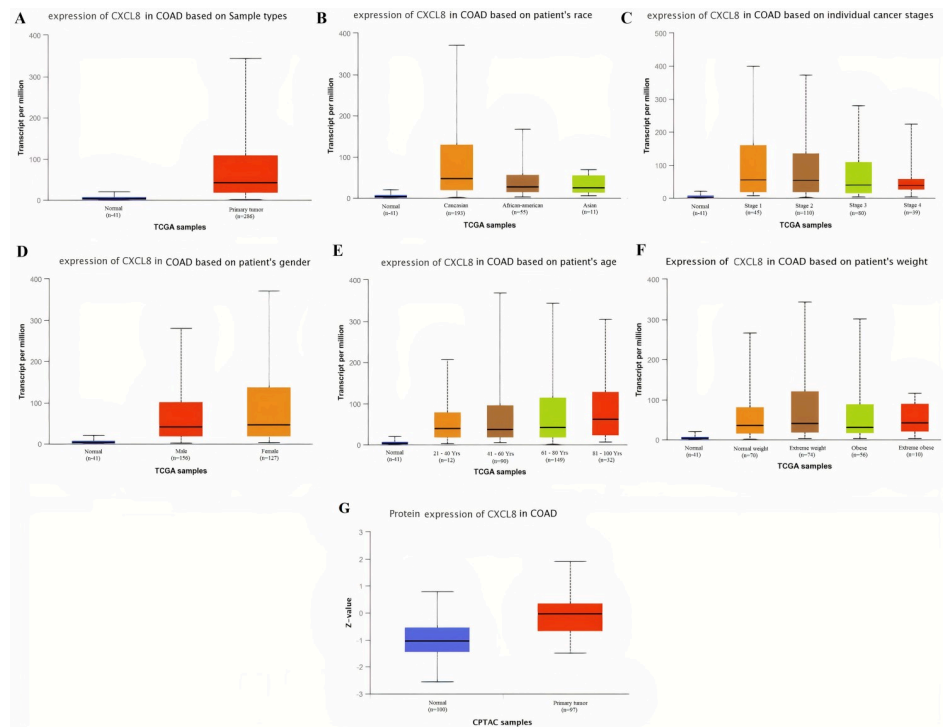


Fig 6. Box plots showing the relative expression levels of CXCL8 in normal and COAD samples of different clinicopathological features via UALCAN database. Relative mRNA expression of CXCL8: (A) in normal individuals and COAD patients; (B) in normal individuals and COAD patients of different races; (C) in normal individuals and COAD patients of different cancer stages; (D) in normal individuals and COAD patients of different genders; (E) in normal individuals and COAD patients of different age groups; and, (F) in normal individuals and COAD patients of different body weights; (G) Relative protein expression of CXCL8 in normal individuals and COAD patients. A p-value of <0.05 was considered significant. Normal weight = BMI greater than or equal to 18.5 and BMI less than 25, Extreme weight = BMI greater than or equal to 25 and BMI less than 30, Obese = BMI greater than or equal to 30 and BMI less than 40, and Extreme Obese = BMI greater than 40.

<https://doi.org/10.1371/journal.pone.0256020.g006>

database. Results revealed a significant ($p>0.05$) positive correlation between the mRNA expression of the CXCL12, GNB1 and CXCL1 and CD8+ T immune cells' infiltration while a significant ($p>0.05$) negative correlation between the mRNA expression of CXCL8, AGT and GNG4 and CD8+ T immune cells' infiltration (Fig 14).

3.9 Investigation of miRNAs-real hub genes interaction network

In this study, the RegNetwork database was used for predicting the miRNAs targeting real hub genes to investigate the regulatory relationships between real hub genes and miRNAs. The co-expression network was developed by Cytoscape software and was based on the correlation analysis between the hub genes and miRNAs (Fig 15). The numbers of miRNAs and mRNAs in the network were 200 and 6, respectively. In the miRNA-real hub gene network, the numbers of arrows indicates the contribution of each miRNA in the expression regulation of the surrounding real hub genes. The miRNA-real hub gene interaction analysis findings revealed that mir-1-3p targets the most (four) of the genes including CXCL8, CXCL12, CXCL1, and GNB1. The miR-1-3p has been widely proven as a tumor suppressor [47]. In CRC, contradicting reports were found reporting expression variations in miR-1-3p. Mainly, miR-1-3p was found down-regulated in different studies investigating its association with CRC [47–49]. However, one exceptional study has also reported its up-regulation [50]. In view of our

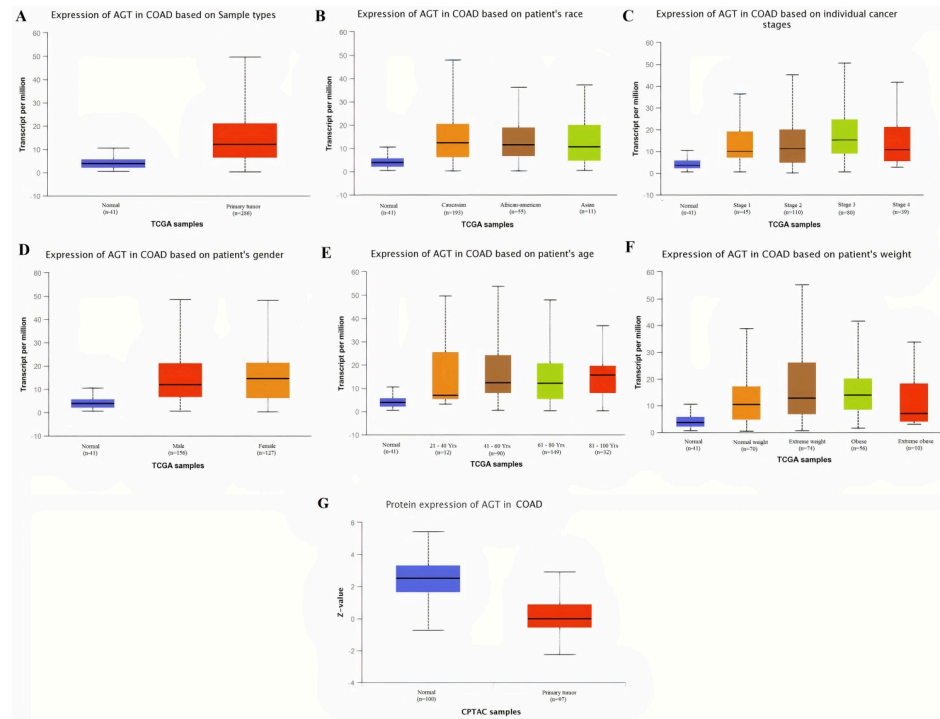


Fig 7. Box plots showing the relative expression levels of AGT in normal and COAD samples of different clinicopathological features via UALCAN database. Relative mRNA expression of AGT: (A) in normal individuals and COAD patients; (B) in normal individuals and COAD patients of different races; (C) in normal individuals and COAD patients of different cancer stages; (D) in normal individuals and COAD patients of different genders; (E) in normal individuals and COAD patients of different age groups; and, (F) in normal individuals and COAD patients of different body weights; (G) Relative protein expression of AGT in normal individuals and COAD patients. A p-value of <0.05 was considered significant. Normal weight = BMI greater than or equal to 18.5 and BMI less than 25, Extreme weight = BMI greater than or equal to 25 and BMI less than 30, Obese = BMI greater than or equal to 30 and BMI less than 40, and Extreme Obese = BMI greater than 40.

<https://doi.org/10.1371/journal.pone.0256020.g007>

findings, we speculate that four real hub genes (CXCL8, CXCL12, CXCL1, and GNB1) function as an oncogenes and down-regulation of mir-1-3p may lead to the up-regulation of these genes as a different axis (mir-1-3p/CXCL8 or CXCL12, CXCL1, and GNB1) in the pathogenesis of CRC. Therefore, it will be worth to therapeutically target the mir-1-3p/CXCL8, or CXCL12, CXCL1, and GNB1 axis in the treatment of CRC (Fig 15).

3.10 Real hub genes-drug interaction network analysis

In order to explore the relationship between real hub genes and available cancer therapeutic drugs, a gene-drug interaction network was developed using CTD database and analyzed via Cytoscape. In view of the gene-drug interaction network analysis results, it was observed that mRNA expression of the identified real hub genes including CXCL12, CXCL8, AGT, GNB1, GNG4 and CXCL1 could potentially be influenced by a variety of drugs. For example, valproic acid and furan could reduce the expression levels of CXCL12 while estradiol and coumestrol could elevate CXCL8 expression levels (Fig 16).

4. Discussion

CRC is the most common type of gastrointestinal tumor and one of the leading causes of cancer-related deaths worldwide [26]. Although recent advances in CRC screening and treatment

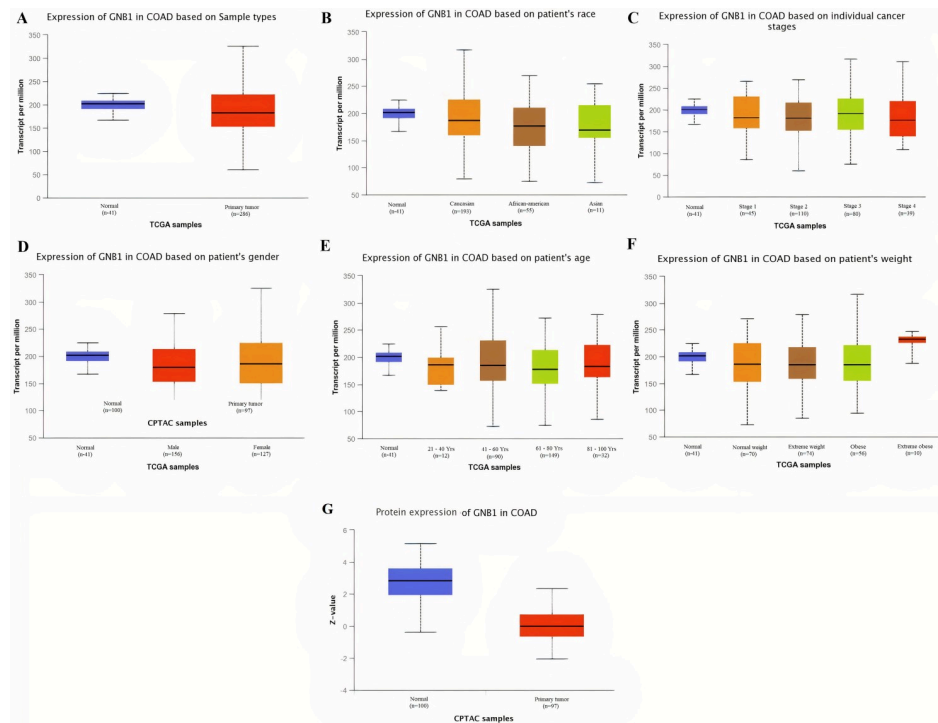


Fig 8. Box plots showing the relative expression levels of GNB1 in normal and COAD samples of different clinicopathological features via UALCAN database. Relative mRNA expression of GNB1: (A) in normal individuals and COAD patients; (B) in normal individuals and COAD patients of different races; (C) in normal individuals and COAD patients of different cancer stages; (D) in normal individuals and COAD patients of different genders; (E) in normal individuals and COAD patients of different age groups; and, (F) in normal individuals and COAD patients of different body weights; (G) Relative protein expression of GNB1 in normal individuals and COAD patients. A p-value of <0.05 was considered significant. Normal weight = BMI greater than or equal to 18.5 and BMI less than 25, Extereme weight = BMI greater than or equal to 25 and BMI less than 30, Obese = BMI greater than or equal to 30 and BMI less than 40, and Extreme Obese = BMI greater than 40.

<https://doi.org/10.1371/journal.pone.0256020.g008>

approaches have proven helpful in the management of disease, its worldwide prevalence is on a rise due to the heterogeneity-specific nature of the available biomarkers used for its detection and monitoring the prognosis [3,4]. It is, therefore, crucial to uncover the molecular mechanisms underlying CRC initiation, development, and progression for the identification of some novel diagnostic and prognostic biomarkers that could be used for detection and monitoring the prognosis and treatment of CRC over the heterogeneity-specific barrier.

In the present study, we conducted a PubMed based search to identify all the studies which utilized GEO-based CRC microarray expression datasets to explore the hub genes. In total 21 studies were identified which collectively utilized more than 30 CRC microarray expression datasets from the GEO. We extracted all the identified hub genes from each of the studies and pooled them to make a set of 210 hub genes representing all the CRC microarray expression datasets utilized by these studies. Then, we performed the pathway enrichment and PPI network construction and analysis of the extracted hub genes to identify the few more closely CRC-related genes (real hub genes) and their underlying pathways. We further performed the expression analysis and validation of the real hub genes, and correlation analysis of their expressions with their promoter methylation and OS survival duration of the COAD patients through a comprehensive multi-layered bioinformatics approach. We also aimed to identify the genetic alterations and to perform the miRNA-real hub genes interaction network analysis and real hub genes-drug interaction network analysis of the identified six real hub genes.

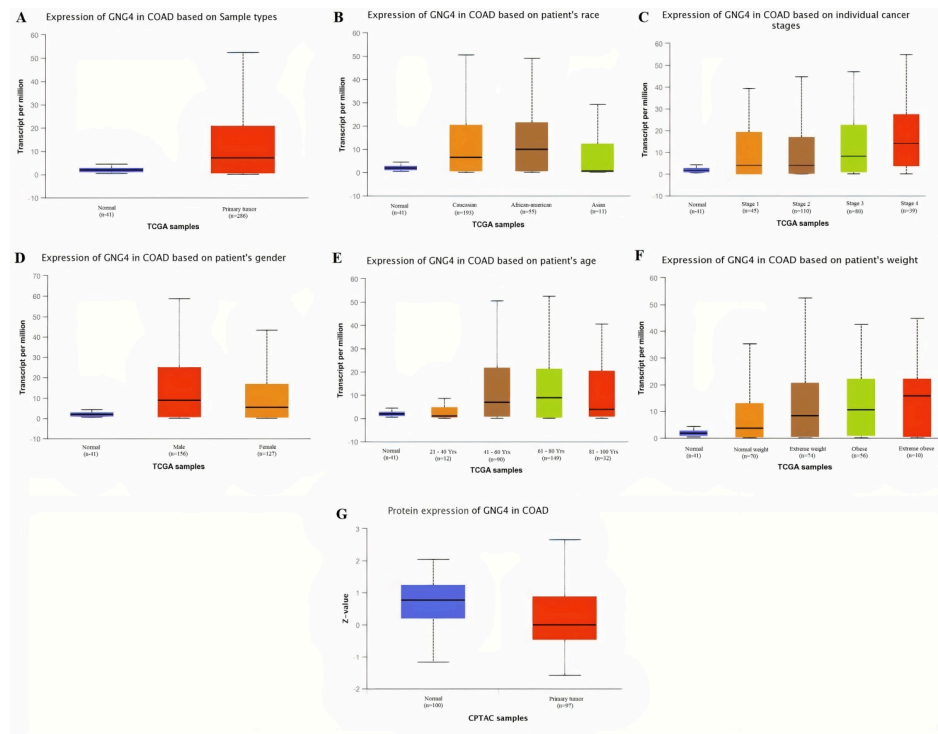


Fig 9. Box plots showing the relative expression levels of GNG4 in normal and COAD samples of different clinicopathological features via UALCAN database. Relative mRNA expression of GNG4: (A) in normal individuals and COAD patients; (B) in normal individuals and COAD patients of different races; (C) in normal individuals and COAD patients of different cancer stages; (D) in normal individuals and COAD patients of different genders; (E) in normal individuals and COAD patients of different age groups; and, (F) in normal individuals and COAD patients of different body weights; (G) Relative protein expression of GNG4 in normal individuals and COAD patients. A p-value of <0.05 was considered significant. Normal weight = BMI greater than or equal to 18.5 and BMI less than 25, Extreme weight = BMI greater than or equal to 25 and BMI less than 30, Obese = BMI greater than or equal to 30 and BMI less than 40, and Extreme Obese = BMI greater than 40.

<https://doi.org/10.1371/journal.pone.0256020.g009>

KEGG pathway enrichment analysis revealed that all the extracted 210 hub genes were significantly enriched in various pathways including ‘Chemokine signaling pathway’, ‘Pathways in cancer’, ‘Cell cycle’, ‘PI3K-Akt signaling pathway’, and Cytokine-cytokine receptor interaction pathway (Fig 1 and Table 2). Furthermore, PPI network of the extracted 210 hub genes illustrated the overview of their functional connections, of which top 6 real hub genes were selected as CXCL12, CXCL8, AGT, GNB1, GNG4 and CXCL1.

KEGG pathway analysis of the 6 real hub genes demonstrated that these genes were significantly ($p < 0.05$) enriched in pathways including “Chemokine signaling pathway”, “Pathways in cancer”, and “Cytokine-cytokine receptor interaction”. In a broader sense, all these pathways basically represent the “Cytokine-cytokine receptor interaction” pathway because cytokine is a general term used for all signaling molecules while chemokine are the specific cytokines that function by attracting cells to sites of infection/inflammation and this pathway also belongs to the category of the “Pathways in cancer”.

Inflammation is an essential component of the tumor microenvironment and one of the hallmarks of cancer [51]. Chemokine’s, are a family of small, secreted, and structurally related cytokines with a crucial role in inflammation and immunity [52]. They are also key mediators of cancer-related inflammation being present at the tumor site for pre-existing chronic inflammatory conditions but also being the target of oncogenic pathways [53]. Initially, chemokine’s

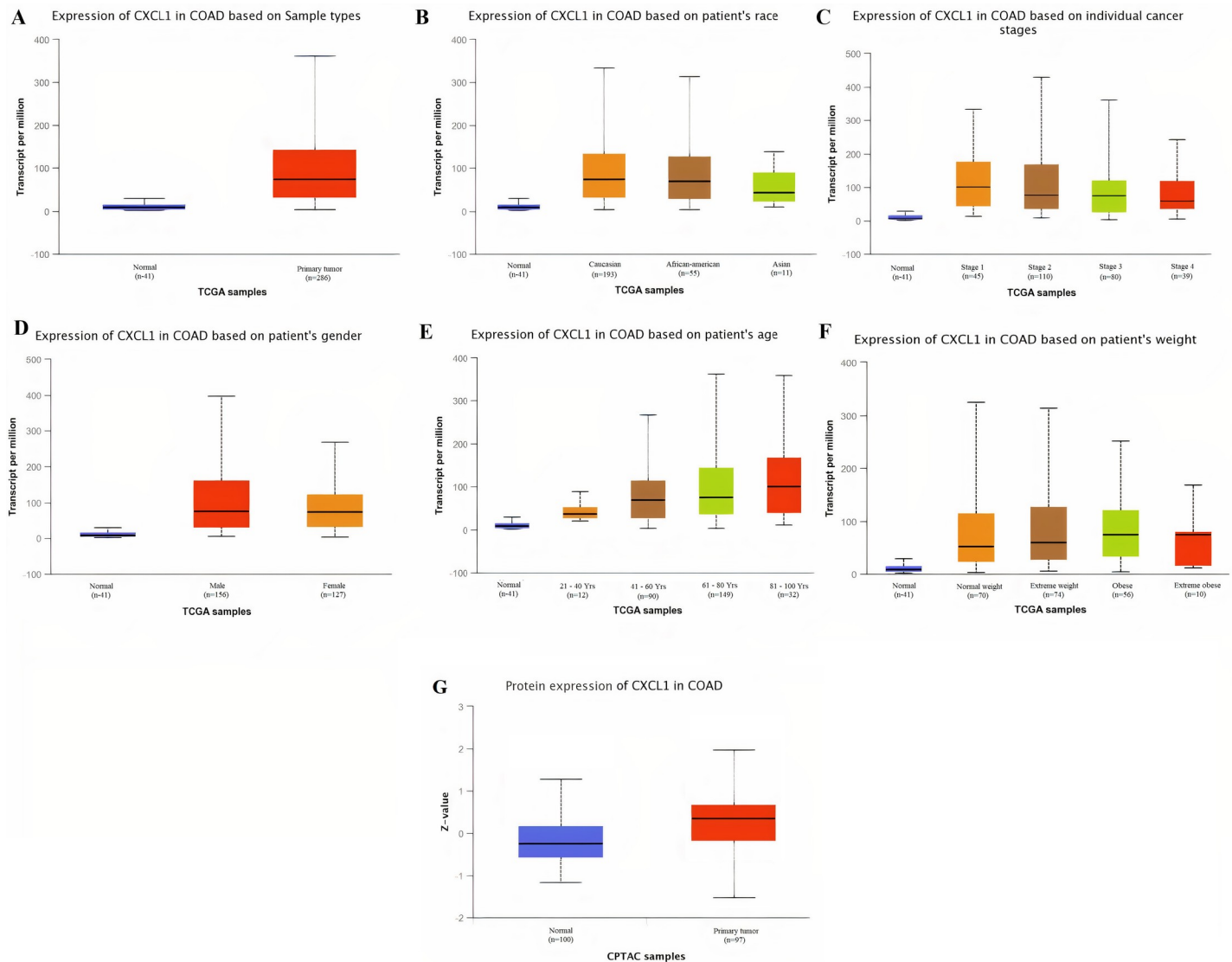


Fig 10. Box plots showing the relative expression levels of CXCL1 in normal and COAD samples of different clinicopathological features via UALCAN database. Relative mRNA expression of CXCL1: (A) in normal individuals and COAD patients; (B) in normal individuals and COAD patients of different races; (C) in normal individuals and COAD patients of different cancer stages; (D) in normal individuals and COAD patients of different genders; (E) in normal individuals and COAD patients of different age groups; and, (F) in normal individuals and COAD patients of different body weights; (G) Relative protein expression of CXCL1 in normal individuals and COAD patients. A p-value of <0.05 was considered significant. Normal weight = BMI greater than or equal to 18.5 and BMI less than 25, Extreme weight = BMI greater than or equal to 25 and BMI less than 30, Obese = BMI greater than or equal to 30 and BMI less than 40, and Extreme Obese = BMI greater than 40.

<https://doi.org/10.1371/journal.pone.0256020.g010>

were identified playing a prominent role in determining the composition of tumor stroma, where they were found to directly affect the cancer cell proliferation and metastasis [54,55].

The identified real hub gene CXCL12 (stromal cell-derived factor 1) is an extracellular chemokine, which binds to CXCR4, a G-protein coupled receptor (GPCR) and have been well-recognized as a factor involved in the cancer metastasis [56]. The CXCL12 is normally secreted by Kupffers and endothelial cells in the liver, which is the most common site for CRC metastases [57]. Previously, various studies have suggested that overexpression of CXCR4 is correlated with the poor survival and liver metastasis in CRC [58,59]. In addition, the elevated expression level of CXCR4 has also been observed in hypoxia due to the activity of hypoxia-inducible

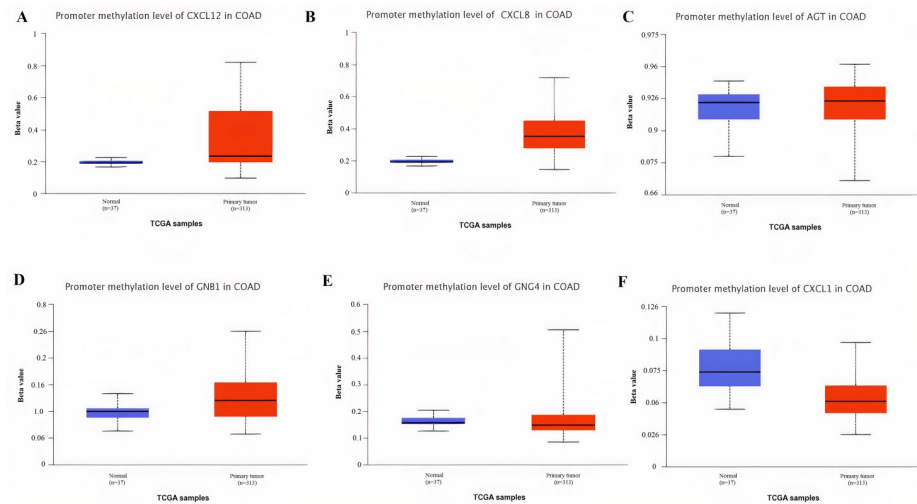


Fig 11. Box plot showing real hub genes promoter methylation levels in normal individuals and COAD patients via UALCAN database. The box plot shows the promoter methylation levels of: (A) CXCL12 (B) CXCL8 (C) AGT (D) GNB1 (E) GNG4 (F) CXCL1 in normal individuals and COAD patients. A p-value of <0.05 was considered significant.

<https://doi.org/10.1371/journal.pone.0256020.g011>

factor $1-\alpha$ [60]. CXCR7 is another receptor that interacts with CXCL12 in CRC cells. Earlier, Wang *et al.* [61] assessed the expression of CXCL12, CXCR4, and CXCR7 in CRC and found that the expression of both CXCL12 and CXCR7 were significantly up-regulated in CRC samples as compared to controls [62]. Contrary to this, however, in the present study, we demonstrated the unusual significant ($p < 0.05$) down-regulation of CXCL12 at mRNA levels in COAD patients of different clinicopathological features including different races, cancer stages, genders, age groups, and body weights as compared to the normal controls. Our results also showed the down-regulation of CXCL12 at translational level in COAD patients as compared to controls. We also investigated the correlation of CXCL12 down-regulation with its promoter methylation status and genetic alterations. In view of the results, it was noticed that significant ($p < 0.05$) hyper methylation is the possible cause of CXCL12's down-regulation in COAD rather than the genetic alterations which were observed in a least count (0.2%) of the analyzed samples. Taken together, our data suggested CXCL12's down-regulation as a novel potential diagnostic biomarker in CRC patients of different races, cancer stages, genders, age groups, and body weights.

The real hub gene CXCL8, also known as neutrophil-activating factor (NAF), and interleukin 8 (IL-8), was the first chemokine identified as a leukocyte chemo-attractant [63]. CXCL8 controls the leukocyte trafficking during homeostasis and inflammation by interacting with (GPCR) receptor CXCR1 [64]. Many previous studies [65–67] have already demonstrated its

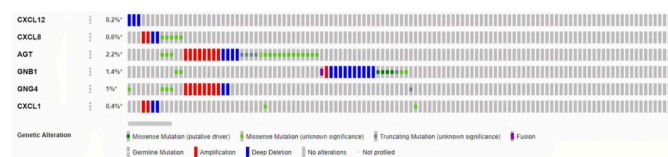


Fig 12. Frequency of the genetic alterations and CNVs of the real hub genes in COAD patients.

<https://doi.org/10.1371/journal.pone.0256020.g012>

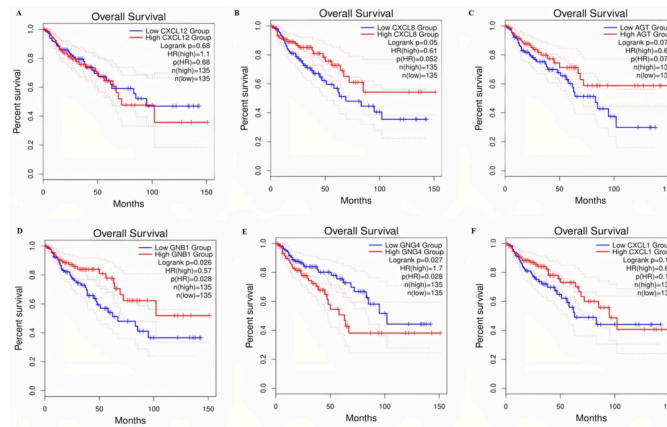


Fig 13. The COAD related prognostic information of the real hub genes obtained via GEPIA database. The calculated prognostic value of: (A) CXCL12, (B) CXCL8, (C) AGT, (D) GNB1, (E) GNG4, (F) CXCL1 gene. Blue color indicates this low expression while red color indicates the high expression of a gene.

<https://doi.org/10.1371/journal.pone.0256020.g013>

up-regulation in CRC patients relative to controls, however, none of these studies generalized CXCL8 in CRC patients of different clinicopathological features (different races, cancer stages, genders, age groups, and body weights). But in our study, we observed and generalized the significant higher mRNA expression of CXCL8 in COAD patients of different clinicopathological features including different races, cancer stages, genders, age groups, and body weights as compared to the normal controls. Our results also showed CXCL8’s up-regulation at translation level in COAD patients with respect to controls. We further revealed that CXCL8 was significantly hyper methylated in COAD patients group relative to controls and also genetically altered in a least proportion (0.6%) of the COAD samples. Regarding unexpected hyper methylation, our findings contradict the commonly accepted association of hyper methylation and down regulated gene expression; therefore, we recommend further in-depth research to explore the clear role of promoter methylation level in the expression regulation of CXCL8.

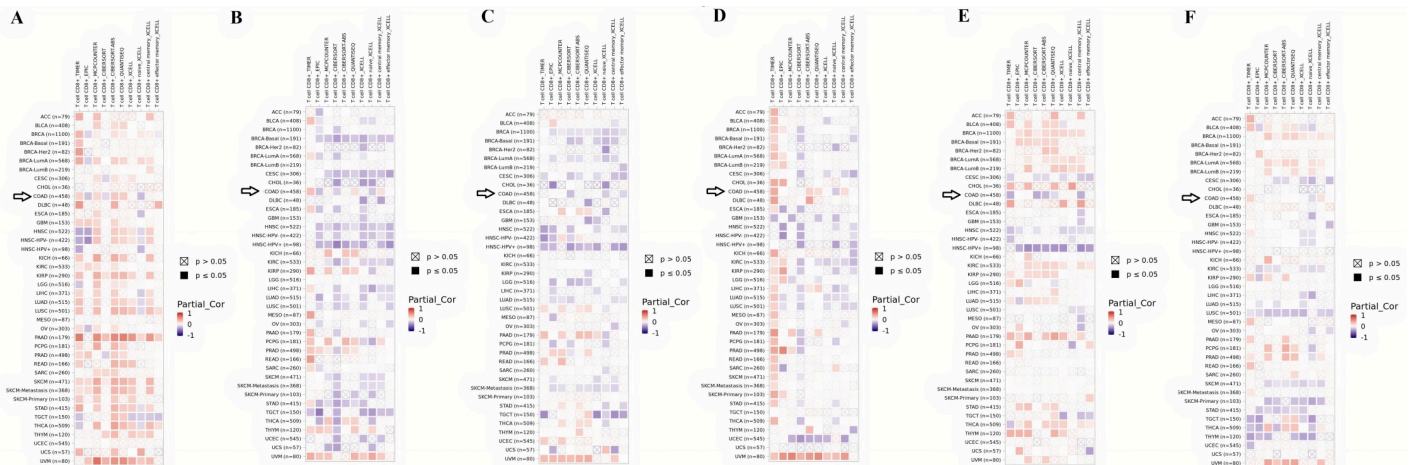


Fig 14. TIMER based Spearman correlational analysis between the expression of real hub genes and CD8+ T immune cell infiltration in COAD. TIMER based Spearman correlational analysis between: (A) CXCL12 expression and CD8+ T immune cell infiltration; (B) CXCL8 expression and CD8+ T immune cells’ infiltration; (C) AGT expression and CD8+ T immune cells’ infiltration; (D) GNB1 expression and CD8+ T immune cells’ infiltration; (E) GNG4 expression and CD8+ T immune cells’ infiltration; (F) CXCL1 expression and CD8+ T immune cells’ infiltration. Red color box represents the positive correlation; blue color represents the negative correlation while white color represents no correlation. A p-value of <0.05 was considered as significant.

<https://doi.org/10.1371/journal.pone.0256020.g014>

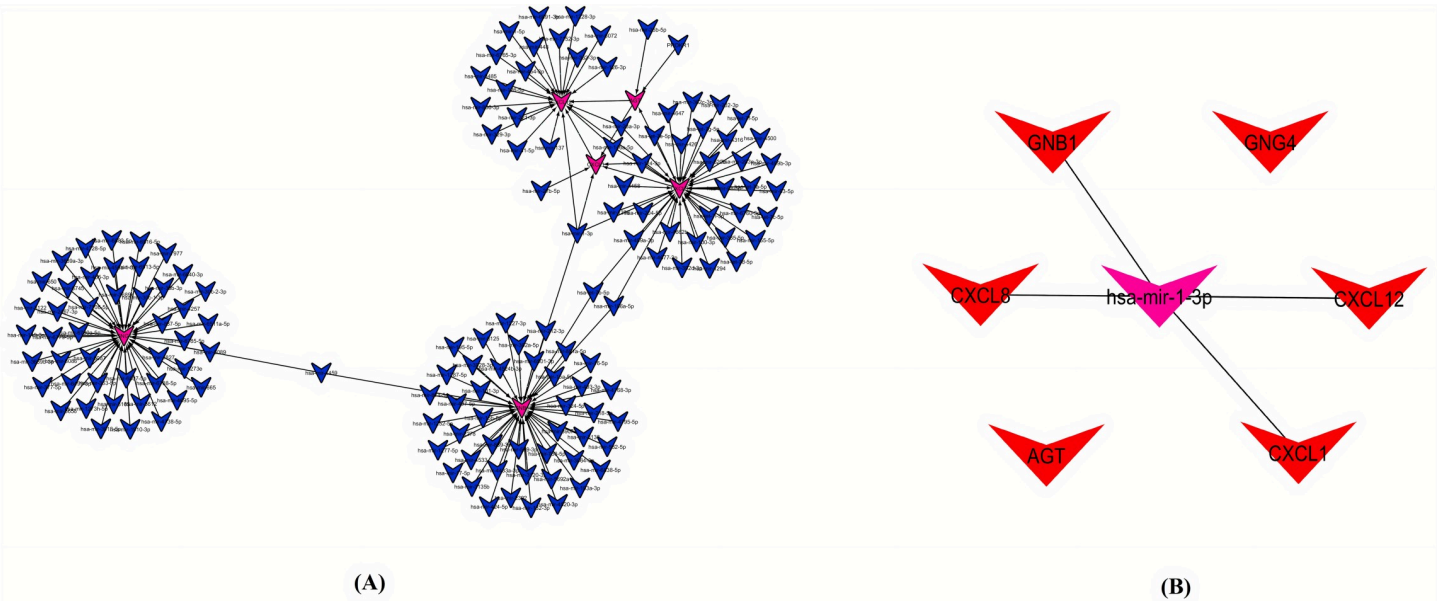


Fig 15. The miRNA–real hub gene interaction network. The pink circular node represents the miRNA. The blue v shape node represents the hub gene while pink and red v shape node represents the mir-1-3p/CXCL8 or CXCL12, CXCL1, and GNB1 axis. The arrow shape represents the interaction between the miRNAs and real hub genes.

<https://doi.org/10.1371/journal.pone.0256020.g015>

Collectively, our data suggested CXCL8 up-regulation as a novel potential diagnostic biomarker in CRC patients of different races, cancer stages, genders, age groups, and body weights.

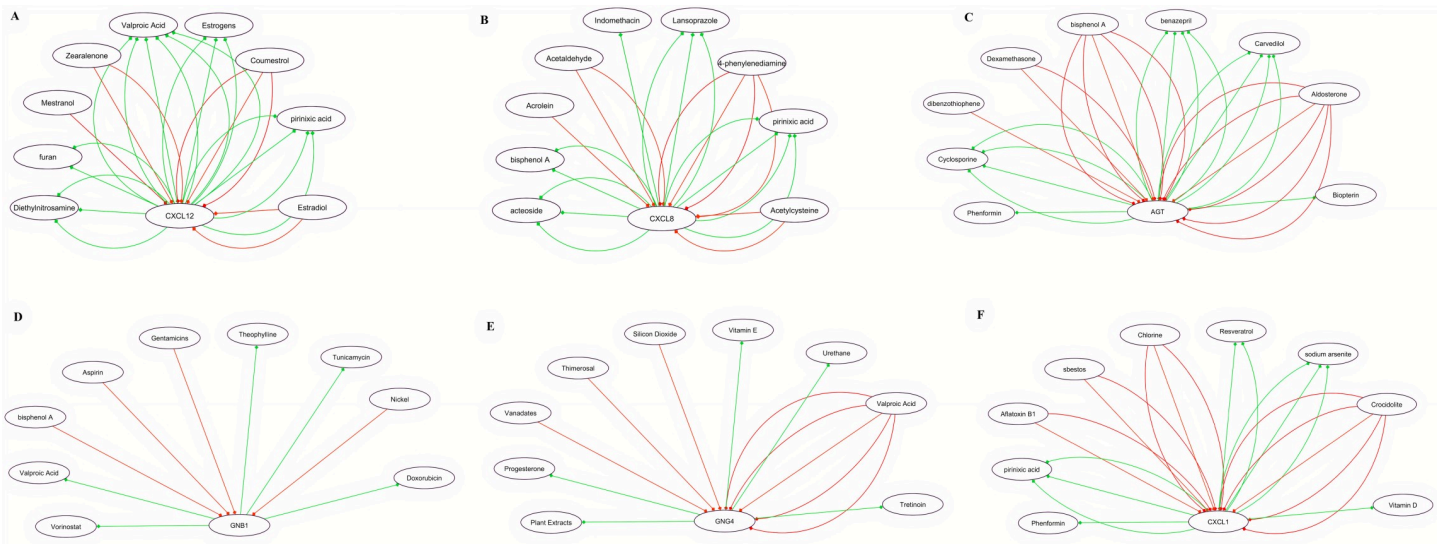


Fig 16. Real hub gene-drug interaction network of the identified real hub genes. Panels A–F indicates the available chemotherapeutic drugs in CTD that decrease or increase the expression levels of the real hub genes. (A) Real hub gene-drug network of CXCL12; (B) Real hub gene-drug network of CXCL8; (C) Real hub gene-drug network of AGT; (D) Real hub gene-drug network of GNB1; (E) Real hub gene-drug network of GNG4; (F) Real hub gene-drug network of CXCL1. Red arrows indicate the chemotherapeutic drugs that could increase the expression levels of the real hub genes, while green arrows indicate the chemotherapeutic drugs that could decrease the expression levels of the real hub genes. The numbers of arrows between chemotherapeutic drugs and real hub genes in the network represent the supported numbers of previous studies reported in literature.

<https://doi.org/10.1371/journal.pone.0256020.g016>

The real hub gene ATG (Angiotensinogen) is an essential component of the renin-angiotensin system (RAS) which is a potent regulator of blood pressure. ATG is a precursor of Angiotensin-II (A-II) and mainly produced in hepatocytes [68]. So far, a single study has been found in published scientific literature reporting overexpression of AGT in CRC [69], however, this study lacks the information regarding clinicopathological features-specific expression variations in AGT. In the current study, we observed the significant ($p < 0.05$) higher expression of ATG at mRNA levels in COAD patients of different clinicopathological features including different races, cancer stages, genders, age groups, and body weights as compared to the normal controls. In contrast to this, the results of present study also reported the lower expression of AGT at translation level in COAD patients relative to controls. This inverse correlation between AGT mRNA and translation levels indicates the defect in post-translation modifications which might reduce the half-life of AGT protein and results in its reduction. Our results further reported that AGT was significantly hyper methylated in COAD patients' group than the controls and also genetically altered in a small proportion (2.2%) of the COAD samples. Regarding a positive correlation between AGT mRNA expression and hyper methylation, we recommend further in-depth research to explore the clear role of promoter methylation levels in the expression regulation of AGT. Overall, the results of our study suggested ATG's up-regulation as a novel potential diagnostic biomarker in CRC patients of different races, cancer stages, genders, age groups, and body weights.

The real hub gene GNB1 encodes for $G\beta_1$ which is a beta (β) subunit of the guanine nucleotide-binding protein that forms heterotrimeric complexes with G protein subunits α and γ . The $G\beta$ subunit is joined to $G\gamma$ subunit to form a $G\beta\gamma$ complex which activates the RAS pathway, a signaling pathway responsible to maintain cell proliferation, cell adhesion, and cell migration [70]. Best to our knowledge, no study has yet reported GNB1 expression variation in CRC. However, one study has demonstrated its down-regulation in Clear-cell renal cell carcinoma (ccRCC) patients [71]. In the present study, we observed the significant ($p < 0.05$) up-regulation of GNB1 mRNA in COAD patients of different clinicopathological features including different races, cancer stages, genders, age groups, and body weights as compared to the normal controls. Our results also showed the down-regulation of GNB1 at translation level in COAD patients relative to controls. This inverse correlation between GNB1 mRNA and translation levels indicates the abnormalities in post-translation modification mechanisms which might reduce the half-life of GNB1 protein and results in its reduction. We further revealed that GNB1 was significantly hyper methylated in group of COAD patients relative to control group and also genetically altered in a least proportion (1.4%) of the COAD samples. Regarding unexpected hyper methylation, our findings challenges the classical concept of methylation, therefore, further work is required to be done to get a more detailed view of correlation between expression and methylation of GNB1 in COAD. Taken together, our results suggested that GNB1 up-regulation may be considered as a novel potential diagnostic biomarker in CRC patients of different races, cancer stages, genders, age groups, and body weights.

The real hub gene GNG4 encodes for the γ subunit of the G protein trimmer which potentially functions as a positive regulator of the RAS pathway that is responsible to maintain cell proliferation, cell adhesion, and cell migration [72]. Best to our knowledge, nothing has been reported in references regarding its expression variations in CRC, however, one earlier study [73] has evaluated its down-regulation in glioblastoma (GBM). Present study demonstrated that GNG4 was significantly ($p < 0.05$) overexpressed at mRNA level in CRC patients of different races, cancer stages, genders, age groups, and body weights. Results of this study also demonstrated the up-regulation of GNG4 at translation level in COAD patients relative to controls. We further reported that GNG4 was significantly hyper methylated in COAD patients group than the controls and also genetically altered in a small proportion (1%) of the

COAD samples. This scenario of GNG4 overexpression and hyper methylation challenges the classical view where hyper methylation is always related with the down-regulation. In total, the results of our study suggested GNG4's up-regulation as a novel potential diagnostic biomarker in CRC patients of different races, cancer stages, genders, age groups, and body weights.

The real hub gene Chemokine (C-X-C motif) ligand 1 (CXCL1), also known as GRO- α , belongs to the G protein-coupled receptor family that specifically binds to the CXC chemokine receptor 2 which activate the RAS (Rat sarcoma) pathway in cell proliferation [74]. Different previous studies have demonstrated the role of CXCL1's up-regulation in CRC metastasis and progression [75–77], however, none of these studies generalized CXCL1 in CRC patients of different clinicopathological features (different races, cancer stages, genders, age groups, and body weights). But this study identified the significant ($p < 0.05$) up-regulation of CXCL1 mRNA in COAD patients of different races, cancer stages, genders, age groups, and body weights. Moreover, results of this study also showed the up-regulation of CXCL1 at translation level in COAD patients relative to controls. The correlation analysis between the CXCL1 expression and methylation status revealed the expected significant ($p < 0.05$) negative correlation which strengthened the role hypomethylation in the up-regulation of CXCL1. Lastly, the results of CXCL1 genetic alteration analysis revealed that CXCL1 expression is unlikely to be the effect of genetic alterations as alterations were noticed in a very small proportion (0.4%) of the COAD patients. Collectively, our data suggested CXCL1 up-regulation as a novel potential diagnostic biomarker in CRC patients of different races, cancer stages, genders, age groups, and body weights. The details of crosstalk between the real hub genes involved pathways in context to CRC is given in Fig 17.

The OS survival analysis of the real hub genes demonstrated that high expression of GNG4 was the good prognostic biomarkers while CXCL12, CXCL8, AGT, GNB1, and CXCL1 were the bad prognostic biomarkers for predicting the OS duration of COAD patients.

To further clarify the underlying mechanisms of real hub genes in CRC tumorigenesis, we performed the correlation analysis between the real hub genes expression and CD8+ T immune cells' infiltration in COAD. The CD8+ T immune cells are known as the major drivers of the anticancer immunity [79] and earlier, CD8+ T immune cells infiltration was utilized

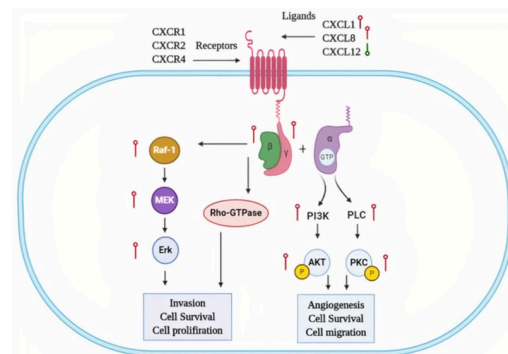


Fig 17. Crosstalk between the real hub genes involved pathways. At the cellular level, CXCL1, CXCL8 or CXCL12 binds to the GPCRs (CXCR1, CXCR2, or CXCR4) and activates the G protein. In this study, the up-regulation of GPCRs ligands (CXCL1 and CXCL8) and G protein subunits (β and γ) collectively are supposed to up-regulate the various downstream pathways. For example, the Heterotrimeric $G\alpha$ subunit up-regulation in this study further up-regulates its main effectors PLC and PI3K to induce hyper phosphorylation of PKC and Akt, respectively. The oncogenic roles of these two signaling pathways have already been reported to activate various respective transcription factors associated with angiogenesis, cell survival, and migration of tumor cells. On the other hand, β and γ G protein subunits up-regulation in this study further up-regulates the Rho-GTPase family and Raf-1/MAP/Erk signaling cascade which has been earlier reported to contribute in the cell invasion, cell survival, and cell proliferation [78].

<https://doi.org/10.1371/journal.pone.0256020.g017>

as a diagnostic marker for the early detection of laryngeal squamous cell carcinoma abbreviated as LSCC [80]. Furthermore, *Trojan et al.* have also successfully used CD8+ T immune cells' infiltration for the personalized immunotherapy trials in LSCC [81]. Our results revealed a significant ($p>0.05$) positive correlation between the mRNA expression levels of the CXCL12, GNB1 and CXCL1 and CD8+ T immune cells' infiltration while a significant ($p>0.05$) negative correlation between the mRNA expression of CXCL8, AGT and GNG4 and CD8+ T immune cells' infiltration. Taken together, these correlations have highlighted the new aspect of the real hub genes in CRC tumorigenesis via regulating CD8+ T immune cells' infiltration. Best to our knowledge, this study is the first study to investigate the spearman correlation coefficient between the expression of real hub genes (CXCL12, CXCL8, AGT, GNB1, GNG4, and CXCL1) and CD8+ T immune cells' infiltration in CRC. These correlations may bring new ideas for the treatment of CRC patients who do not benefit from the existing immune checkpoint inhibitors/regulators.

The miRNAs are the small non-coding RNA molecules (~22 nucleotides) responsible for the degradation and translation of the mRNAs in plants and animals [82]. In CRC, conflicting reports were found reporting expression variations in miR-1-3p. Mainly, miR-1-3p was found down-regulated in different studies investigating association of miR-1-3p with CRC [47–49]. However, one exceptional study has also reported its up-regulation [50]. In view to the results of present study, we speculate that four real hub genes (CXCL8, CXCL12, CXCL1, and GNB1) function as an oncogenes and down-regulation of mir-1-3p may lead to the up-regulation of these genes as a different axis (mir-1-3p/CXCL8 or CXCL12, CXCL1, and GNB1) in the pathogenesis of CRC. Best to our knowledge, this study is the first study to report the tumorigenic role of mir-1-3p together with CXCL8, CXCL12, CXCL1, and GNB1 in CRC. In addition, we have also identified various drugs (Fig 16) that could be used in the treatment of CRC by regulating the expression of real hub genes.

Some worth monitoring limitations of the present study are inevitable. Firstly, the lack of ample data regarding expression status of real hub genes in CRC patients of Pakistani population might be one limitation. Secondly, we only considered the top six real hub genes, however, the remaining genes needed to be considered in the future experiments. Finally, the information related to various clinical parameters of COAD patients like family history of cancer, tumor location, smoking, and alcohol drinking history is not available on the utilized databases and thus has not been analyzed in detail and discussed in the manuscript.

5. Conclusion

In summary, we have identified the a panel of six differentially expressed real hub genes (hub genes of the hub genes) including CXCL12, CXCL8, AGT, GNB1, GNG4, and CXCL1 and their underlying molecular pathways that could be employed as a possible diagnostic and prognostic biomarkers in CRC patients of different races, cancer stages, genders, age groups, and body weight and thus may help to overcome the heterogeneity-specific barrier. However, prior to clinical implementation there is need to launch extensive biological investigations especially for the under-represented populations in the expression datasets used during present study.

Supporting information

S1 Table. Row list of the CRC-associated hub genes extracted from the previous studies. (DOCX)

Author Contributions

Conceptualization: Yasir Hameed, Shufang Liang, Samina Ejaz.

Data curation: Yasir Hameed, Muhammad Usman.

Formal analysis: Yasir Hameed.

Investigation: Samina Ejaz.

Methodology: Yasir Hameed, Shufang Liang, Samina Ejaz.

Software: Yasir Hameed, Muhammad Usman, Samina Ejaz.

Supervision: Shufang Liang, Samina Ejaz.

Validation: Muhammad Usman, Samina Ejaz.

Visualization: Yasir Hameed, Muhammad Usman.

Writing – original draft: Yasir Hameed.

Writing – review & editing: Samina Ejaz.

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. *CA: Cancer J. Clin.* 2020; 70: 7–30.
2. Network TCGA. Comprehensive molecular characterization of human colon and rectal cancer. *Nature.* 2012; 487: 330–7. <https://doi.org/10.1038/nature11252> PMID: 22810696
3. Mármol I, Sánchez-de-Diego C, Pradilla Dieste A, Cerrada E, Rodriguez Yoldi MJ. Colorectal Carcinoma: A General Overview and Future Perspectives in Colorectal Cancer. *Int J Mol Sci.* 2017; 18: 197. <https://doi.org/10.3390/ijms18010197> PMID: 28106826
4. Molinari C, Marisi G, Passardi A, Matteucci L, De Maio G, Ulivi P. Heterogeneity in Colorectal Cancer: A Challenge for Personalized Medicine? *Int J Mol Sci.* 2018; 19: 3733. <https://doi.org/10.3390/ijms19123733> PMID: 30477151
5. Zhang Y, Zheng Y, Fu Y, Wang C. Identification of biomarkers, pathways and potential therapeutic agents for white adipocyte insulin resistance using bioinformatics analysis. *Adipocyte.* 2019; 8: 318–29. <https://doi.org/10.1080/21623945.2019.1649578> PMID: 31407623
6. Chen YJ, Guo YN, Shi K, Huang HM, Huang SP, Xu WQ, et al. Down-regulation of microRNA-144-3p and its clinical value in non-small cell lung cancer: a comprehensive analysis based on microarray, miRNA-sequencing, and quantitative real-time PCR data. *Respir Res.* 2019; 20: 019–0994.
7. Chen DL, Lu YX, Zhang JX, Wei XL, Wang F, Zeng ZL, et al. Long non-coding RNA UICLM promotes colorectal cancer liver metastasis by acting as a ceRNA for microRNA-215 to regulate ZEB2 expression. *Theranostics.* 2017; 7: 4836–49. <https://doi.org/10.7150/thno.20942> PMID: 29187907
8. Lemuth K, Rupp S. Microarrays as Research Tools and Diagnostic Devices. *RNA and DNA Diagnostics.* 2015: 259–80.
9. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, et al. NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res.* 2013; 41: 27.
10. Barrett T, Edgar R. Mining microarray data at NCBI's Gene Expression Omnibus (GEO)*. *Methods Mol Biol.* 2006; 338: 175–90. <https://doi.org/10.1385/1-59745-097-9:175> PMID: 16888359
11. Lv J, Li L. Hub Genes and Key Pathway Identification in Colorectal Cancer Based on Bioinformatic Analysis. *Biomed Res. Int.* 2019; 2019: 1545680-. <https://doi.org/10.1155/2019/1545680> PMID: 31781593
12. Creixell P, Reimand J, Haider S, Wu G, Shibata T, Vazquez M, et al. Pathway and network analysis of cancer genomes. *Nat Methods.* 2015; 12: 615–21. <https://doi.org/10.1038/nmeth.3440> PMID: 26125594
13. Tang Z, Li C, Kang B, Gao G, Zhang Z. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res.* 2017; 45: W98–W102. <https://doi.org/10.1093/nar/gkx247> PMID: 28407145
14. Chandrashekar DS, Bashel B, Balasubramanya SAH, Creighton CJ, Ponce-Rodriguez I, Chakravarthi B, et al. UALCAN: A Portal for Facilitating Tumor Subgroup Gene Expression and Survival Analyses. *Neoplasia (New York, NY).* 2017; 19: 649–58. <https://doi.org/10.1016/j.neo.2017.05.002> PMID: 28732212

15. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* 2012; 2: 401–4. <https://doi.org/10.1158/2159-8290.CD-12-0095> PMID: 22588877
16. Huang DW, Sherman BT, Tan Q, Collins JR, Alvord WG, Roayaei J, et al. The DAVID Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists. *Genome Biol.* 2007; 8: R183–R. <https://doi.org/10.1186/gb-2007-8-9-r183> PMID: 17784955
17. von Mering C, Huynen M, Jaeggi D, Schmidt S, Bork P, Snel B. STRING: a database of predicted functional associations between proteins. *Nucleic acids research.* 2003; 31: 258–61. <https://doi.org/10.1093/nar/gkg034> PMID: 12519996
18. 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: 2498–504. <https://doi.org/10.1101/gr.1239303> PMID: 14597658
19. Park SJ, Yoon BH, Kim SK, Kim SY. GENT2: an updated gene expression database for normal and tumor tissues. *BMC Med Genomics.* 2019; 12: 101. <https://doi.org/10.1186/s12920-019-0514-7> PMID: 31296229
20. Li T, Fu J, Zeng Z, Cohen D, Li J, Chen Q, et al. TIMER2.0 for analysis of tumor-infiltrating immune cells. *Nucleic acids research.* 2020; 48: W509–W14. <https://doi.org/10.1093/nar/gkaa407> PMID: 32442275
21. Liu ZP, Wu C, Miao H, Wu H. RegNetwork: an integrated database of transcriptional and post-transcriptional regulatory networks in human and mouse. *Database.* 2015;2015. <https://doi.org/10.1093/database/bav095> PMID: 26424082
22. Mattingly CJ, Colby GT, Forrest JN, Boyer JL. The Comparative Toxicogenomics Database (CTD). *Environ Health Perspect.* 2003; 111: 793–5. <https://doi.org/10.1289/ehp.6028> PMID: 12760826
23. Luo T, Wu S, Shen X, Li L. Network cluster analysis of protein-protein interaction network identified biomarker for early onset colorectal cancer. *Mol Biol Rep.* 2013; 40: 6561–8. <https://doi.org/10.1007/s11033-013-2694-0> PMID: 24197691
24. He J, Liu W. Identification of disrupted pathways associated with colon cancer based on combining protein-protein interactions and pathway data. *J Cancer Res Ther.* 2018; 14: S998–S1003. <https://doi.org/10.4103/0973-1482.191063> PMID: 30539836
25. Dong Z, Lin W, Kujawa SA, Wu S, Wang C. Predicting MicroRNA Target Genes and Identifying Hub Genes in IIA Stage Colon Cancer Patients Using Bioinformatics Analysis. *Biomed Res Int.* 2019;7. <https://doi.org/10.1155/2019/6341967> PMID: 30881993
26. Chen H, Li N, Ren J, Feng X, Lyu Z, Wei L, et al. Participation and yield of a population-based colorectal cancer screening programme in China. *Gut.* 2019; 68: 1450–7. <https://doi.org/10.1136/gutjnl-2018-317124> PMID: 30377193
27. Wei S, Chen J, Huang Y, Sun Q, Wang H, Liang X, et al. Identification of hub genes and construction of transcriptional regulatory network for the progression of colon adenocarcinoma hub genes and TF regulatory network of colon adenocarcinoma. *J. Cell. Physiol.* 2020; 235: 2037–48. <https://doi.org/10.1002/jcp.29067> PMID: 31612481
28. Gong B, Kao Y, Zhang C, Sun F, Gong Z, Chen J. Identification of Hub Genes Related to Carcinogenesis and Prognosis in Colorectal Cancer Based on Integrated Bioinformatics. *Mediators Inflamm.* 2020; 2020: 5934821. <https://doi.org/10.1155/2020/5934821> PMID: 32351322
29. Yuan Y, Chen J, Wang J, Xu M, Zhang Y, Sun P, et al. Identification Hub Genes in Colorectal Cancer by Integrating Weighted Gene Co-Expression Network Analysis and Clinical Validation in vivo and vitro. *Front Oncol.* 2020; 10: 638. <https://doi.org/10.3389/fonc.2020.00638> PMID: 32426282
30. Zhou H, Yang Z, Yue J, Chen Y, Chen T, Mu T, et al. Identification of potential hub genes via bioinformatics analysis combined with experimental verification in colorectal cancer. *Mol. Carcinog.* 2020; 59. <https://doi.org/10.1002/mc.23165> PMID: 32064687
31. Yang W, Ma J, Zhou W, Li Z, Zhou X, Cao B, et al. Identification of hub genes and outcome in colon cancer based on bioinformatics analysis. *Cancer Manag Res.* 2018; 11: 323–38. <https://doi.org/10.2147/CMAR.S173240> PMID: 30643458
32. Dai GP, Wang LP, Wen YQ, Ren XQ, Zuo SG. Identification of key genes for predicting colorectal cancer prognosis by integrated bioinformatics analysis. *Oncol Lett.* 2020; 19: 388–98. <https://doi.org/10.3892/ol.2019.11068> PMID: 31897151
33. Wang Y, Zheng T. Screening of hub genes and pathways in colorectal cancer with microarray technology. *Pathol Oncol Res.* 2014; 20: 611–8. <https://doi.org/10.1007/s12253-013-9739-5> PMID: 24504536
34. Wang YR, Meng LB, Su F, Qiu Y, Shi JH, Xu X, et al. Insights regarding novel biomarkers and the pathogenesis of primary colorectal carcinoma based on bioinformatic analysis. *Comput Biol Chem.* 2020; 85: 4. <https://doi.org/10.1016/j.compbiolchem.2020.107229> PMID: 32058945

35. Wang X, Hu S, Ji W, Tang Y, Zhang S. Identification of genes associated with clinicopathological features of colorectal cancer. *J Int Med Res.* 2020; 48: 300060520912139. <https://doi.org/10.1177/0300060520912139> PMID: 32281438
36. Rahman MR, Islam T, Gov E, Turanlı B, Gulfidan G, Shahjaman M, et al. Identification of Prognostic Biomarker Signatures and Candidate Drugs in Colorectal Cancer: Insights from Systems Biology Analysis. *Medicina.* 2019; 55. <https://doi.org/10.3390/medicina55010020> PMID: 30658502
37. 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> PMID: 32153633
38. Guo Y, Bao Y, Ma M, Yang W. Identification of Key Candidate Genes and Pathways in Colorectal Cancer by Integrated Bioinformatical Analysis. *Int J Mol Sci.* 2017; 18: 722. <https://doi.org/10.3390/ijms18040722> PMID: 28350360
39. Yu C, Chen F, Jiang J, Zhang H, Zhou M. Screening key genes and signaling pathways in colorectal cancer by integrated bioinformatics analysis. *Mol Med Rep.* 2019; 20: 1259–69. <https://doi.org/10.3892/mmr.2019.10336> PMID: 31173250
40. Chen L, Lu D, Sun K, Xu Y, Hu P, Li X, et al. Identification of biomarkers associated with diagnosis and prognosis of colorectal cancer patients based on integrated bioinformatics analysis. *Gene.* 2019; 692: 119–25. <https://doi.org/10.1016/j.gene.2019.01.001> PMID: 30654001
41. Qi C, Chen Y, Zhou Y, Huang X, Li G, Zeng J, et al. Delineating the underlying molecular mechanisms and key genes involved in metastasis of colorectal cancer via bioinformatics analysis. *Oncol Rep.* 2018; 39: 2297–305. <https://doi.org/10.3892/or.2018.6303> PMID: 29517105
42. Liu F, Ji F, Ji Y, Jiang Y, Sun X, Lu Y, et al. In-depth analysis of the critical genes and pathways in colorectal cancer. *Int J Mol Med.* 2015; 36: 923–30. <https://doi.org/10.3892/ijmm.2015.2298> PMID: 26239303
43. Lv Y, Xie B, Bai B, Shan L, Zheng W, Huang X, et al. Weighted gene coexpression analysis indicates that PLAGL2 and POFUT1 are related to the differential features of proximal and distal colorectal cancer. *Oncol Rep.* 2019; 42: 2473–85. <https://doi.org/10.3892/or.2019.7368> PMID: 31638246
44. Wajed SA, Laird PW, DeMeester TR. DNA methylation: an alternative pathway to cancer. *Ann Surg.* 2001; 234: 10–20. <https://doi.org/10.1097/0000658-200107000-00003> PMID: 11420478
45. Toft M, Ross OA. Copy number variation in Parkinson's disease. *Genome Med.* 2010; 2: 62. <https://doi.org/10.1186/gm183> PMID: 20828427
46. Ziai J, Gilbert HN, Foreman O, Eastham-Anderson J, Chu F, Huseni M, et al. CD8+ T cell infiltration in breast and colon cancer: A histologic and statistical analysis. *PloS one.* 2018; 13: e0190158. <https://doi.org/10.1371/journal.pone.0190158> PMID: 29320521
47. Wang JY, Huang JC, Chen G, Wei DM. Expression level and potential target pathways of miR-1-3p in colorectal carcinoma based on 645 cases from 9 microarray datasets. *Mol Med Rep.* 2018; 17: 5013–20. <https://doi.org/10.3892/mmr.2018.8532> PMID: 29393467
48. Wu Y, Pu N, Su W, Yang X, Xing C. Downregulation of miR-1 in colorectal cancer promotes radioresistance and aggressive phenotypes. *Journal of Cancer.* 2020; 11: 4832–40. <https://doi.org/10.7150/jca.44753> PMID: 32626530
49. Xu W, Zhang Z, Zou K, Cheng Y, Yang M, Chen H, et al. MiR-1 suppresses tumor cell proliferation in colorectal cancer by inhibition of Smad3-mediated tumor glycolysis. *Cell Death Dis.* 2017; 8: e2761–e. <https://doi.org/10.1038/cddis.2017.60> PMID: 28471448
50. Li Z, Gu X, Fang Y, Xiang J, Chen Z. microRNA expression profiles in human colorectal cancers with brain metastases. *Oncology letters.* 2012; 3: 346–50. <https://doi.org/10.3892/ol.2011.497> PMID: 22740910
51. Colotta F, Allavena P, Sica A, Garlanda C, Mantovani A. Cancer-related inflammation, the seventh hallmark of cancer: links to genetic instability. *Carcinogenesis.* 2009; 30: 1073–81. <https://doi.org/10.1093/carcin/bgp127> PMID: 19468060
52. Griffith JW, Sokol CL, Luster AD. Chemokines and chemokine receptors: positioning cells for host defense and immunity. *Annu. Rev. Immunol.* 2014; 32: 659–702. <https://doi.org/10.1146/annurev-immunol-032713-120145> PMID: 24655300
53. Mantovani A, Savino B, Locati M, Zammataro L, Allavena P, Bonecchi R. The chemokine system in cancer biology and therapy. *Cytokine Growth Factor Rev.* 2010; 21: 27–39. <https://doi.org/10.1016/j.cytogfr.2009.11.007> PMID: 20004131
54. Caronni N, Savino B, Recordati C, Villa A, Locati M, Bonecchi R. Cancer and chemokines. *Tumor Immunology.* 2016. p. 87–96. https://doi.org/10.1007/978-1-4939-3338-9_8 PMID: 27033218
55. Chow MT, Luster AD. Chemokines in cancer. *Cancer Immunol. Res.* 2014; 2: 1125–31. <https://doi.org/10.1158/2326-6066.CIR-14-0160> PMID: 25480554

56. Müller A, Homey B, Soto H, Ge N, Catron D, Buchanan ME, et al. Involvement of chemokine receptors in breast cancer metastasis. *nature*. 2001; 410: 50–6. <https://doi.org/10.1038/35065016> PMID: 11242036
57. Matsusue R, Kubo H, Hisamori S, Okoshi K, Takagi H, Hida K, et al. Hepatic stellate cells promote liver metastasis of colon cancer cells by the action of SDF-1/CXCR4 axis. *Ann. Surg. Oncol.* 2009; 16: 2645–53. <https://doi.org/10.1245/s10434-009-0599-x> PMID: 19588204
58. Kim J, Takeuchi H, Lam ST, Turner RR, Wang H-J, Kuo C, et al. Chemokine receptor CXCR4 expression in colorectal cancer patients increases the risk for recurrence and for poor survival. *J Clin Oncol.* 2005; 23: 2744–53. <https://doi.org/10.1200/JCO.2005.07.078> PMID: 15837989
59. Ottaiano A, Franco R, Talamanca AA, Liguori G, Tatangelo F, Delrio P, et al. Overexpression of both CXC chemokine receptor 4 and vascular endothelial growth factor proteins predicts early distant relapse in stage II-III colorectal cancer patients. *Clin Cancer Res.* 2006; 12: 2795–803. <https://doi.org/10.1158/1078-0432.CCR-05-2142> PMID: 16675573
60. Romain B, Hachet-Haas M, Rohr S, Brigand C, Galzi J-L, Gaub M-P, et al. Hypoxia differentially regulated CXCR4 and CXCR7 signaling in colon cancer. *Molecular cancer.* 2014; 13: 58. <https://doi.org/10.1186/1476-4598-13-58> PMID: 24629239
61. Wang M, Yang X, Wei M, Wang Z. The Role of CXCL12 Axis in Lung Metastasis of Colorectal Cancer. *Journal of Cancer.* 2018; 9: 3898. <https://doi.org/10.7150/jca.26383> PMID: 30410593
62. Kaplan RN, Riba RD, Zacharoulis S, Bramley AH, Vincent L, Costa C, et al. VEGFR1-positive haematopoietic bone marrow progenitors initiate the pre-metastatic niche. *Nature.* 2005; 438: 820–7. <https://doi.org/10.1038/nature04186> PMID: 16341007
63. Baggiolini M. CXCL8—the first chemokine. *Front Immunol.* 2015; 6: 285. <https://doi.org/10.3389/fimmu.2015.00285> PMID: 26106391
64. Ramjeesingh R, Leung R, Siu C-H. Interleukin-8 secreted by endothelial cells induces chemotaxis of melanoma cells through the chemokine receptor CXCR1. *The FASEB journal.* 2003; 17: 1292–4. <https://doi.org/10.1096/fj.02-0560fje> PMID: 12738812
65. Ning Y, Lenz HJ. Targeting IL-8 in colorectal cancer. *Expert Opin Ther Targets.* 2012; 16: 491–7. <https://doi.org/10.1517/14728222.2012.677440> PMID: 22494524
66. Ogawa R, Yamamoto T, Hirai H, Hanada K, Kiyasu Y, Nishikawa G, et al. Loss of SMAD4 Promotes Colorectal Cancer Progression by Recruiting Tumor-Associated Neutrophils via the CXCL1/8-CXCR2 Axis. *Clin Cancer Res.* 2019; 25: 2887–99. <https://doi.org/10.1158/1078-0432.CCR-18-3684> PMID: 30705034
67. Dabkeviciene D, Jonusiene V, Zitkute V, Zalyte E, Grigaitis P, Kirvelienu V, et al. The role of interleukin-8 (CXCL8) and CXCR2 in acquired chemoresistance of human colorectal carcinoma cells HCT116. *Medical oncology (Northwood, London, England).* 2015; 32: 258.
68. Dzau VJ. Implications of local angiotensin production in cardiovascular physiology and pharmacology. *Am J Cardiol.* 1987; 59: A59–A65. [https://doi.org/10.1016/0002-9149\(87\)90178-0](https://doi.org/10.1016/0002-9149(87)90178-0) PMID: 3028117
69. Shimomoto T, Ohmori H, Luo Y, Chihara Y, Denda A, Sasahira T, et al. Diabetes-associated angiotensin activation enhances liver metastasis of colon cancer. *Clin. Exp. Metastasis.* 2012; 29: 915–25. <https://doi.org/10.1007/s10585-012-9480-6> PMID: 22552372
70. Madukwe JC, Garland-Kuntz EE, Lyon AM, Smrcka AV. G protein β subunits directly interact with and activate phospholipase C ϵ . *J Biol Chem.* 2018; 293: 6387–97. <https://doi.org/10.1074/jbc.RA118.002354> PMID: 29535186
71. Chen C, Chi H, Min L, Junhua Z. Downregulation of guanine nucleotide-binding protein beta 1 (GNB1) is associated with worsened prognosis of clearcell renal cell carcinoma and is related to VEGF signaling pathway. *J BUON.* 2017; 22: 1441–6. PMID: 29332336
72. Kishibuchi R, Kondo K, Soejima S, Tsuboi M, Kajjira K, Kawakami Y, et al. DNA methylation of GHSR, GNG4, HOXD9 and SALL3 is a common epigenetic alteration in thymic carcinoma. *Int J Oncol.* 2020; 56: 315–26. <https://doi.org/10.3892/ijo.2019.4915> PMID: 31746370
73. Pal J, Patil V, Mondal B, Shukla S, Hegde AS, Arivazhagan A, et al. Epigenetically silenced GNG4 inhibits SDF1 α /CXCR4 signaling in mesenchymal glioblastoma. *Genes Cancer.* 2016; 7: 136–47. <https://doi.org/10.18632/genesandcancer.105> PMID: 27382437
74. Xiang Z, Jiang DP, Xia GG, Wei ZW, Chen W, He Y, et al. CXCL1 expression is correlated with Snail expression and affects the prognosis of patients with gastric cancer. *Oncol. Lett.* 2015; 10: 2458–64. <https://doi.org/10.3892/ol.2015.3614> PMID: 26622871
75. Divella R, Daniele A, De Luca R, Simone M, Naglieri E, Savino E, et al. Circulating levels of VEGF and CXCL1 are predictive of metastatic organotropism in patients with colorectal cancer. *Anticancer Res.* 2017; 37: 4867–71. <https://doi.org/10.21873/anticancer.11895> PMID: 28870907

76. Wang D, Sun H, Wei J, Cen B, DuBois RN. CXCL1 is critical for premetastatic niche formation and metastasis in colorectal cancer. *Cancer research*. 2017; 77: 3655–65. <https://doi.org/10.1158/0008-5472.CAN-16-3199> PMID: 28455419
77. Triner D, Xue X, Schwartz AJ, Jung I, Colacino JA, Shah YM. Epithelial hypoxia-inducible factor 2 α facilitates the progression of colon tumors through recruiting neutrophils. *Mol Cell Biol*. 2017; 37: e00481–16. <https://doi.org/10.1128/MCB.00481-16> PMID: 27956697
78. Cheng Y, Ma X-I, Wei Y-q, Wei X-W. Potential roles and targeted therapy of the CXCLs/CXCR2 axis in cancer and inflammatory diseases. *Biochimica et Biophysica Acta (BBA)—Reviews on Cancer*. 2019; 1871: 289–312. <https://doi.org/10.1016/j.bbcan.2019.01.005> PMID: 30703432
79. Xia A, Zhang Y, Xu J, Yin T, Lu X-J. T Cell Dysfunction in Cancer Immunity and Immunotherapy. *Front Immunol*. 2019; 10: 1719. <https://doi.org/10.3389/fimmu.2019.01719> PMID: 31379886
80. Jiang T, Shi J, Dong Z, Hou L, Zhao C, Li X, et al. Genomic landscape and its correlations with tumor mutational burden, PD-L1 expression, and immune cells infiltration in Chinese lung squamous cell carcinoma. *J. Hematol*. 2019; 12: 1–13.
81. Trojan A, Urosevic M, Dummer R, Giger R, Weder W, Stahel RA. Immune activation status of CD8+ T cells infiltrating non-small cell lung cancer. *Lung Cancer*. 2004; 44: 143–7. <https://doi.org/10.1016/j.lungcan.2003.11.004> PMID: 15084378
82. Ling H, Fabbri M, Calin GA. MicroRNAs and other non-coding RNAs as targets for anticancer drug development. *Nat Rev Drug Discov*. 2013; 12: 847–65. <https://doi.org/10.1038/nrd4140> PMID: 24172333