Research article

Holistic exploration of *CHGA* and hsa-miR-137 in colorectal cancer via multi-omic data Integration

Hossein Safarpour [a,*], Javad Ranjbaran [b,1], Nafiseh Erfanian [c,1], Samira Nomiri [b], Afshin Derakhshani [d], Casimiro Gerarduzzi [e,f], Adib Miraki Feriz [c], Edris HosseiniGol [g], Samira Saghafi [a,h], Nicola Silvestris [i,**]

[a] *Cellular and Molecular Research Center, Birjand University of Medical Sciences, Birjand, Iran*
[b] *Department of Clinical Biochemistry, School of Medicine, Birjand University of Medical Sciences, Birjand, Iran*
[c] *Student Research Committee, Birjand University of Medical Sciences, Birjand, Iran*
[d] *Laboratory of Experimental Pharmacology, IRCCS Istituto Tumori Giovanni Paolo II, 70124, Bari, Italy*
[e] *Centre de Recherche de l'Hôpital Maisonneuve-Rosemont, Montréal, Québec, Canada*
[f] *Département de Médecine, Université de Montréal, Montréal, Québec, Canada*
[g] *Department of Computer Engineering, University of Birjand, Birjand, Iran*
[h] *Department of Internal Medicine, School of Medicine, Birjand University of Medical Sciences, Birjand, Iran*
[i] *Medical Oncology Unit, Department of Human Pathology "G. Barresi", University of Messina, Messina, Italy*

ARTICLE INFO

ABSTRACT

Colorectal cancer (CRC) ranks among the most widespread malignancies globally, with early detection significantly influencing prognosis. Employing a systems biology approach, we aimed to unravel the intricate mRNA-miRNA network linked to CRC pathogenesis, potentially yielding diagnostic biomarkers. Through an integrative analysis of microarray, Bulk RNA-seq, and single-cell RNA-seq data, we explored CRC-related transcriptomes comprehensively. Differential gene expression analysis uncovered crucial genes, while Weighted Gene Co-expression Network Analysis (WGCNA) identified key modules closely linked to CRC. Remarkably, CRC manifested its strongest correlation with the turquoise module, signifying its pivotal role. From the cohort of genes showing high Gene Significance (GS) and Module Membership (MM), and Differential Expression Genes (DEGs), we highlighted the downregulated Chromogranin A (*CHGA*) as a notable hub gene in CRC. This finding was corroborated by the Human Protein Atlas database, which illustrated decreased *CHGA* expression in CRC tissues. Additionally, *CHGA* displayed elevated expression in primary versus metastatic cell lines, as evidenced by the CCLE database. Subsequent RT-qPCR validation substantiated the marked downregulation of *CHGA* in CRC tissues, reinforcing the significance of our differential expression analysis. Analyzing the Space-Time Gut Cell Atlas dataset underscored specific *CHGA* expression in epithelial cell subclusters, a trend persisting across developmental stages. Furthermore, our scrutiny of colon and small intestine Enteroendocrine cells uncovered distinct *CHGA* expression patterns, accentuating its role in CRC pathogenesis. Utilizing the WGCNA algorithm and TargetScan database, we validated the downregulation of hsa-miR-137 in CRC, and integrated assessment highlighted its interplay with *CHGA*. Our findings advocate hsa-miR-137 and *CHGA* as promising CRC biomarkers, offering

---

valuable insights into diagnosis and prognosis. Despite proteomic analysis yielding no direct correlation, our multifaceted approach contributes comprehensive understanding of CRC's intricate regulatory mechanisms. In conclusion, this study advances hsa-miR-137 and *CHGA* as promising CRC biomarkers through an integrated analysis of diverse datasets and network interactions.

1. Introduction

Colorectal cancer (CRC) is one of the most common cancers of the digestive system, as well as the world's fourth most common malignancy, after breast, lung, and prostate cancer [1]. Over 1.9 million new CRC cases and 930,000 deaths were estimated in 2020. The burden of CRC is projected to increase to 3.2 million new cases and 1.6 million deaths by 2040 [2]. In 2020, men had a 25% higher incidence and mortality rate than women, and the colon is responsible for the vast majority of reported CRC cases, with the rectum accounting for fewer cases [3]. Despite the fact that the majority of CRC cases are sporadic (70%), a significant proportion of cases (25%) occur in patients with a family history of CRC or hereditary colorectal cancer syndromes (10 %) [4]. CRC prognosis is heavily dependent on early detection; however, better methods are needed for monitoring and screening to further reduce the rate of cancer mortality [5,6].

Screening methods of CRC such as fecal occult blood testing (FOBT), adaptable sigmoidoscopy, and colonoscopy are commonly regarded as aggressive, each with its own set of consequences. As a result, there is an urgent need to develop simple, less invasive procedures with high sensitivity and specificity in order to identify these patients with the fewest side effects and in the shortest amount of time possible. Currently, stool proteins, plasma-based DNA, and microRNAs (miRNAs) are employed as innovative noninvasive screening methods for colorectal neoplasia [7–9]. On the other hand, miRNAs are thought to be involved in the genesis of cancer since their aberrant expression has been associated with a number of disorders, including cancer. Recently, miRNAs have been receiving clinical significance to cancer development, making them potential candidates of non-invasive molecular biomarkers for cancer diagnosis.

MicroRNAs (miRNAs) are a type of noncoding regulatory RNA that is small in size (19–25 nt). MiRNAs suppress the expression of target mRNAs by complementary binding. More importantly, one miRNA may target hundreds of mRNAs; thus, miRNAs–mRNAs form networks that participate in a variety of cellular pathways, including proliferation, apoptosis, and differentiation. Some studies have discovered deregulated miRNA–mRNA networks, in CRC [10,11]. However, the use of some important clinical parameters, such as Tumor, Node and Metastasis (TNM) stage and survival, is limited, which may affect the ability to discover molecular mechanisms and biomarkers linked to cancer progression and prognosis. Furthermore, miRNA-target relationships have attracted extensive attention. In this regard, due to the fact that most genes and miRNAs are influenced by each other, the relationships between genes and miRNAs are of special interest [12].

Current approaches in transcriptomic studies, such as microarray-based monitoring, offer novel insights into the identification of new markers as diagnostic or therapeutic targets. A high throughput sequencing technique may reveal new candidate genes, provide a comprehensive molecular landscape of CRC, and provide novel molecular classifications that describe molecular tumor heterogeneity and allow for new targeted therapies. However, due to the wide variety of data generated by high throughput techniques, new approaches of obtaining significant correlations from highly multivariate datasets are required [13]. Weighted Gene Co-Expression Network Analysis (WGCNA) is a biology-based evolutionary method for investigating intrinsic transcriptome organization. WGCNA has been shown to be an effective method for detecting expression modules, hub genes, and miRNAs in a variety of ways [14–16]. WGCNA uses correlation networks to identify modules of highly correlated genes and miRNAs, as well as intramodular hub-genes and hub-miRNAs that have been identified as candidate biomarkers or therapeutic targets [17,18].

Consequently, the goal of this study was to identify the mRNA-miRNA network involved in CRC pathogenesis using WGCNA analysis and a systems biology approach, as well as to introduce important genes and miRNAs as diagnostic biomarkers and new gene targets in CRC treatment.

2. Materials and methods

The plan of this study is shown in (Chart. 1).

2.1. Data acquisition and differential gene expression analysis

The mRNA expression profile of GSE81558 and clinical information of CRC patients were obtained from the gene expression omnibus (GEO) database for this study (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE81558(. This dataset contains 23 CRC samples and 9 normal mucosa samples. Raw data underwent quantile normalization, followed by additional data processing, utilizing the R software packages (version 3.2.0). The Affymetrix-provided probe-set annotation file was employed to extract official gene symbols and probe IDs. Genes displaying coefficients of variation <0.1 were excluded. Subsequently, an average expression information assessment was conducted for each individual sample.

The process of identifying differential gene expression was conducted using the GEO2R tool. Genes displaying an adjusted p-value (Padj) of less than 0.05 were deemed as statistically significant.

The ComBat method [19] from the sva package in R was used for intra-platform batch effects removal. ComBat is widely recognized for its efficacy in mitigating batch-related variations, preserving biological signals. This approach assures the reliability and accuracy of our analysis results, as supported by previous studies [20].

2.2. Construction of co-expression network and identification of key modules and genes

We employed the WGCNA method to delve into gene expression data alterations in a network context specific to CRC. The analytical process began with the conversion of GSE81558 gene expression profiles into co-expression networks. This entailed initial computations of pairwise Pearson's correlation matrices, followed by the application of a soft thresholding power (β) to adjust the network's adherence to scale-free topology. The chosen value of β (β = 8) ensured optimal fitting within the scale-free framework, indicated by an R^2 value of 0.95. Further steps included the calculation of the topological overlap matrix (TOM) to gauge gene interconnectedness [21]. The dynamic tree cut algorithm was utilized to cluster genes into co-expression modules, emphasizing similarity in expression patterns.

To identify modules significantly related to evaluated clinical features, the expression profiles of each module were summarized using the Module Eigengene (ME), which represents the first principal component of a module. This ME provides insight into the expression pattern of the module for each instance and serves as a representative of gene expression levels within the module. Additionally, the gene significance (GS) values were employed to assess the association of individual genes with CRC. Moreover, the Module Membership (MM) was established through the correlation between the ME and the gene expression profile within each module. Strong connections between GS and MM signify that the pivotal elements within the modules are closely tied to the trait under scrutiny. This information was then utilized to construct a network and identify hub genes. To do this, we filtered out the high ranked genes from selected module based on the highest GS and MM values. The genes which passed the criteria were then compared to the differential expression gene (DEG) list and common genes were considered as hub-genes.

Subsequently, a comprehensive enrichment analysis was conducted to delve into the functional significance of the identified modules. This analytical exploration involved the utilization of the Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways, which were assessed through the sophisticated Enrichr database accessible at https://maayanlab.cloud/Enrichr. This database facilitated an in-depth investigation into the biological processes, molecular functions, cellular components, and intricate pathways that the modules potentially contribute to. Finally, STRING database (http://string-db.org/) was used for construction of PPI networks of selected hub gene.

2.3. Validation of chosen hub genes across external datasets and databases

To validate our hub gene selection process, we incorporated the GSE110224 microarray dataset (GPL570) with 34 samples (17 adjacent normal and 17 primary colorectal adenocarcinoma (COAD)) (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE110224). Employing GEO2R software, we conducted DEG analysis, highlighting gene expression disparities between these categories. This approach strengthens the reliability of our hub gene choices and underscores its relevance across diverse datasets, enhancing the credibility of our research outcomes.

To assess how the chosen hub gene impacts the overall survival (OS) of CRC patients, we turned to the GEPIA online database, accessible at http://gepia.cancer-pku.cn. This database allowed us to delve into the relationship between the hub gene and patients' survival outcomes. To further investigate the expression patterns of the filtered genes in both control and CRC groups, we employed two complementary resources: the GEPIA database and the Human Protein Atlas (HPA) database [22]. These platforms facilitated a comprehensive comparison of gene expression levels across different conditions. To validate our findings and gain insights into the hub gene's expression across various tumor states, including primary tumors, recurrent tumors, and metastatic tumors, we extracted relevant data from the TCGA-COAD database. This information was retrieved using the UCSC Cancer Browser, accessible at https://xenabrowser.net/(retrieved on May 18, 2023). By doing so, we could establish a broader picture of the hub gene's behavior within different cancer contexts. Furthermore, to extend our understanding of the hub gene's expression in both primary and metastatic cells, we tapped into the Cancer Cell Line Encyclopedia (CCLE) datasets. These datasets were accessed through the Xena browser, which provided us with valuable insights into the hub gene's expression patterns across a range of cancer cell lines.

2.4. Validating the expression of hub-genes by quantitative real-time PCR (qRT-PCR)

A total of six pairs of malignant and adjacent normal tissues from six CRC patients from Shahid Beheshti University of Medical Sciences Research Center for Gastroenterology and Liver Diseases Biobank were obtained to validate the expression of hub-gene. Ethical issues and patient satisfaction were considered based on the Helsinki Declaration, which was recognized by the ethics committee of Birjand University of Medical Sciences (IR.BUMS.REC.1398.139). These samples' total mRNA was extracted using TRIzol (Invitrogen; Thermo Fisher Research, Inc.) per the company's procedure. After quantification and qualification of RNA samples using NanoDrop spectrophotometer (Epoch spectrophotometer- BioTek) and standard agarose gel electrophoresis, RNA was reverse transcribed into cDNA using the RevertAid First Strand cDNA Synthesis Kit (Thermo Fisher Scientific, USA) according to the manufacturer's instructions. The RealQ Plus Master Mix Green (Amplicon) and Applied Biosystems 7500 Real-time PCR machine were used for qRT-PCR (Thermo Fisher Scientific, Inc., Waltham, MA, USA). To standardize data analysis, GAPDH was employed as an endogenous control to estimate expression levels. $2^{-\Delta\Delta CT}$ analysis followed. Each sample was triple-tested.

2.5. Expression characterization of selected hub-gene on single-cell RNA-seq data

In the subsequent phase, our objective was to comprehensively delineate the expression patterns of the chosen hub gene using Single-cell RNA sequencing (scRNA-seq) data. To accomplish this, we harnessed the capabilities of scRNA-seq data sourced from the Space-Time Gut Cell Atlas (obtained from the Cellxgene collections; https://cellxgene.cziscience.com). This extensive dataset offers a panoramic view of the complete spectrum of cellular diversity within the intestinal milieu, spanning a spectrum of developmental stages — from fetal to pediatric and adult donors. With an impressive repository of 428,000 intestinal cells at our disposal, this dataset encompasses a multitude of unique cellular contexts within the intestinal domain, totaling up to 11 distinct subsets.

2.6. Uncovering key MicroRNAs in CRC associated with the selected hub gene

To identify microRNAs with potential correlations to our chosen hub gene in CRC, we conducted a dual analysis involving both differential expression and co-expression assessments using the GSE108153 dataset (https://www.ncbi.nlm.nih.gov/geo/query/acc.
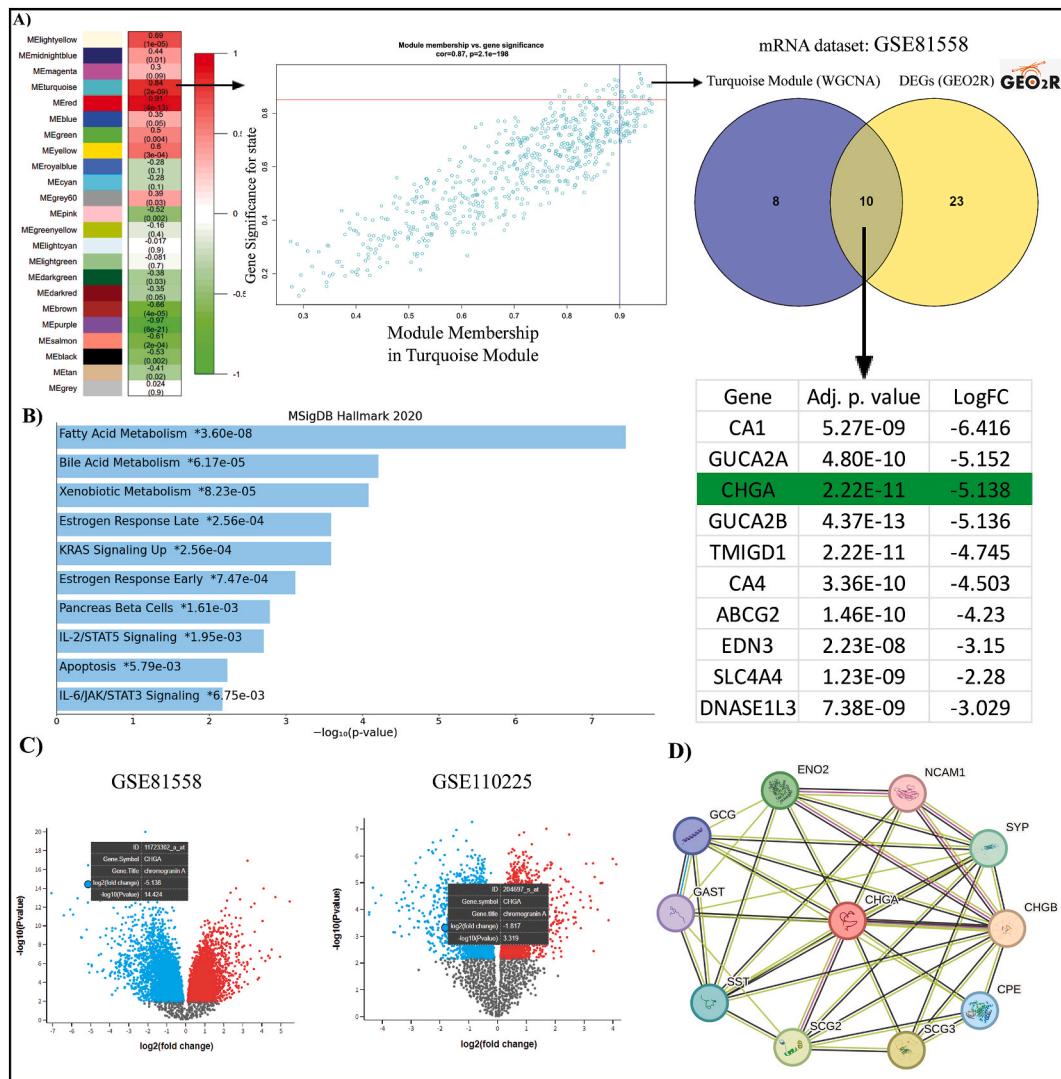

Fig. 1. WGCNA analysis and hub-gene selection for mRNA microarray dataset. A) Module-trait relationship. Each row corresponds to a module eigengene, and the column corresponds to CRC status. The numbers in each cell represent the corresponding correlation and p-value; Module features of GS and MM. Each point represents an individual within each module, which are plotted by GS on the y-axis and gene MM on the x-axis; **B)** Molecular signature hallmarks identified within the turquoise module through EnrichR database; **C)** Visualization of GSE81558 and GSE110225 DEG analysis. Volcanic diagram shows the significance of $-\log_{10}$(p.val) on the y-axis and the increase and decrease threshold of gene expression based on LogFC on the x-axis. Each gene is also marked as a dot (blue dots for upregulation and red dots for downregulation) on the graph; **D)** PPI network of *CHGA* gene using STRING database.

cgi?acc=GSE108153). This dataset, constructed on the GPL19730 platform, encompasses 21 CRC samples and 21 adjacent normal colorectal samples. This approach allowed us to not only unearth differentially expressed microRNAs but also explore their possible associations and interactions with the hub gene.

In the initial step, we performed a differential expression analysis utilizing GEO2R. This produced a list of Differentially Expressed miRNAs (DEMs) marked by LogFC$\geq$|0.5| and P-value <0.01. Simultaneously, after batch effect removal, our investigation expanded into co-expression analysis via WGCNA. Here, we meticulously evaluated all miRNAs based on their GS and MM values. Enhancing our analysis further, we turned to the TargetScan database (https://www.targetscan.org/vert_80/), from which we extracted a set of evolutionarily conserved miRNAs associated with our selected hub gene. This curated list added another layer of understanding, considering their evolutionary importance. By intersecting these three distinct lists – the DEMs, the WGCNA-selected module miRNAs, and the conserved miRNAs from TargetScan – we reached a pivotal juncture. This meticulous fusion of data sources unveiled a shared set of miRNAs that emerged as the central hub miRNA.

Furthermore, we leveraged the DepMap Portal (https://depmap.org/portal/) to delve into the expression patterns of hsa-miR-137 within 84 cell lines relevant to the bowel. Moreover, we assessed the correlation between selected hub gene and this microRNA by analyzing the copy number data available in the publicly accessible 23Q2 dataset of the DepMap Portal. To unveil potential links between hsa-miR-137 and selected hub gene at the protein level, we tapped into the proteomic dataset provided by the DepMap platform.

3. Results

3.1. Identification of DEGs between CRC and adjacent normal tissues

In parallel a set of 33 genes emerged as DEGs with a high degree of significance, reflected by a p-value threshold of less than 0.01 and a substantial fold change exceeding |3|.

3.2. WGCNA reveals gene modules associated with CRC

The co-expression analysis of GSE81558 incorporated expression profiles from 32 samples originating from two distinct colorectal tissue sources. For the WGCNA analysis, a subset of 4000 genes was scrutinized. Cluster analysis of the samples didn't reveal distinct clusters, ensuring robustness (Fig. S1A). To ensure a scale-free network, we conducted empirical analysis to determine the optimal β parameter. As shown in Fig. S1B, the scale-free topological model fit index (R2) and average connectivity stabilized at a β value of 7. Upon establishing weighting coefficients, we obtained the disTOM for the 4000 genes and identified 23 modules through mean linkage hierarchical clustering, each represented by a unique color (Fig. S1C).

3.3. Identification of key modules of CRC

A heat map, depicted in Fig. 1A, was generated to explore the connection between module eigenvalues and CRC. Columns present correlation coefficients and corresponding p-values, with red indicating positive correlations and green signifying negative ones. Darker colors denote larger correlation coefficients. Notably, CRC exhibited its strongest correlation with the turquoise and red modules. Relying on enrichment analysis outcomes, the designation of the turquoise module as the pivotal focal point for subsequent analysis is firmly substantiated. Pathway analysis on this module in MSigDB revealed KRAS, IL-2/STAT5, and IL-6/JAK/STAT3 related signatures were enriched (Fig. 1B).

3.4. CHGA was screened as a key gene in CRC pathogenicity

The examination of the characteristics (MM and GS) associated with the turquoise module (Fig. 1A) has yielded insightful findings, uncovering hub genes that exhibit a significant correlation with CRC pathogenesis. By identifying genes within the turquoise module that possess the highest MM and GS scores, a subsequent comparison was conducted against the DEG list. Genes that demonstrated similarity were ultimately designated as the definitive hub genes, as depicted in Fig. 1A. Through meticulous application of filters, a subset of 10 genes fulfilling the criteria was derived. Among these, *CHGA* emerged as the standout candidate, aligning with the applied parameters. Consequently, *CHGA* was chosen for further downstream analysis, representing a promising avenue to delve deeper into our investigative pursuits.

3.5. Characterization of CHGA expression by external databases

3.5.1. Validation of CHGA expression on another CRC microarray dataset

In the GSE81558 dataset, the expression of *CHGA* was notably reduced in CRC samples, exhibiting a LogFC of −5.1 (adjusted p-value: 2.22E-11). A similar pattern of decrease was also evident in the GSE110224 dataset, where the LogFC was recorded as −1.82 (adjusted p-value: 0.009). These consistent observations across different datasets underscore the significant and consistent down-regulation of *CHGA* expression in CRC samples (Fig. 1C).

3.5.2. Construction of protein-protein interaction (PPI) network of CHGA

To gain deeper insights into *CHGA* and its associated protein levels, we meticulously assembled a PPI network employing the STRING database. Illustrated in Fig. 1D, the core PPI network encompasses 11 nodes interconnected by 40 edges, revealing interactions involving genes such as *CHGB*, *CPE*, *ENO2*, *GAST*, *GCG*, *NCAM1*, *SCG2*, *SCG3*, *SST*, and *SYP*. Impressively, the PPI enrichment analysis yielded a remarkably low p-value of 6.47e-11, underscoring the significance and robustness of the observed interactions within this network.
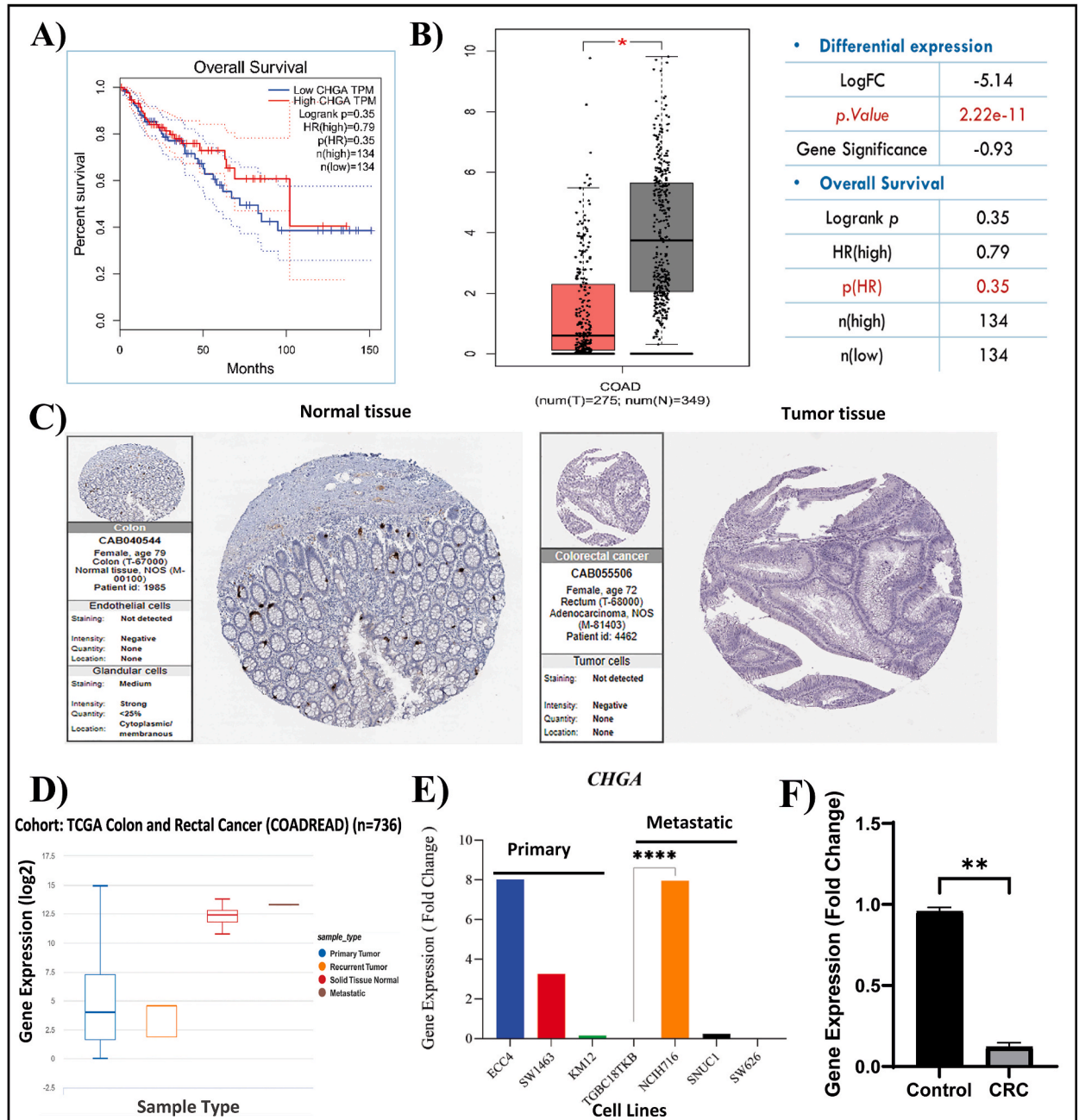

Fig. 2. *CHGA* expression characterization using external databases. A) The effect of changes in *CHGA* gene expression on survival and prognosis of the disease by GEPIA; **B)** The rate of change in gene expression. Red: tumor sample and gray: normal sample; **C)** Expression of hub genes in the Human Protein Atlas database, *CHGA* expression was downregulated in CRC tissues. (Left: normal tissue, right: cancerous tissue); **D)** *CHGA* expression profiles and clinicopathological data of patients with CRC from UCSC database; **E)** The analysis of the CCLE database revealed a high level of *CHGA* expression in primary (ECC4, SW1463, KM12, and TGBC18TKB) compared to metastatic (NCIH716, SNUC1, and SW626) cell lines; **F)** The expression pattern of *CHGA* in control and CRC tissues by qRT-PCR.

3.5.3. CHGA expression on Bulk RNA-seq data and histopathological images of CRC

Through an in-depth exploration of *CHGA* within the GEPIA database, we unearthed its paramount significance in both prognostic value and disease pathogenesis, evidenced by its distinct differential expression (Fig. 2A and B).

Furthermore, insights from the Human Protein Atlas Database provided a visual representation of *CHGA*'s expression dynamics within CRC tissues. The histopathological image revealed a downregulation of *CHGA* in CRC tissues (Fig. 2C).

Delving into the clinicopathological aspects, we sourced patient data with primary CRC from the UCSC Xena browser (https://xenabrowser.net/) (accessed on May 18, 2022) (Fig. 2D). Intriguingly, *CHGA* exhibited significantly varied expression across solid tissue normal (N = 51), primary (N = 380), recurrent tumors (N = 1), and metastatic tumors (N = 1). Indeed, *CHGA* exhibited a pronounced downregulation in primary and recurrent tumors compared to solid tissue normal (p < 0.0001). However, a contrasting
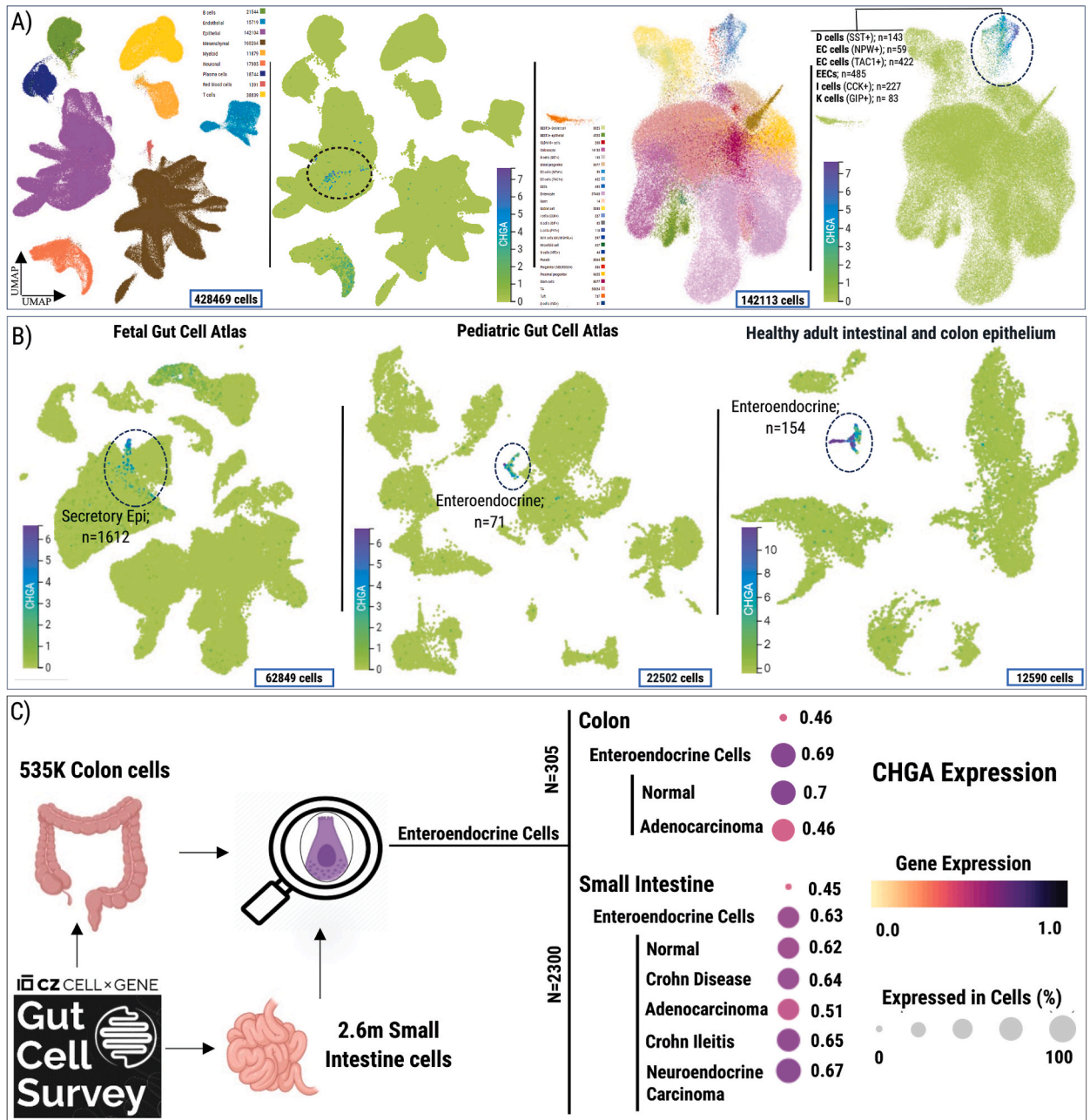

Fig. 3. The expression pattern of *CHGA* on single-cell RNA-seq data. A) specific expression of CHGA within epithelial cell sub-clusters in Space-Time Gut Cell Atlas dataset; **B)** *CHGA* showed a consistent trend across three developmental stages (Fetal, Pediatric, and Adult); **C)** *CHGA* expression was notably reduced in adenocarcinoma samples compared to normal cells.
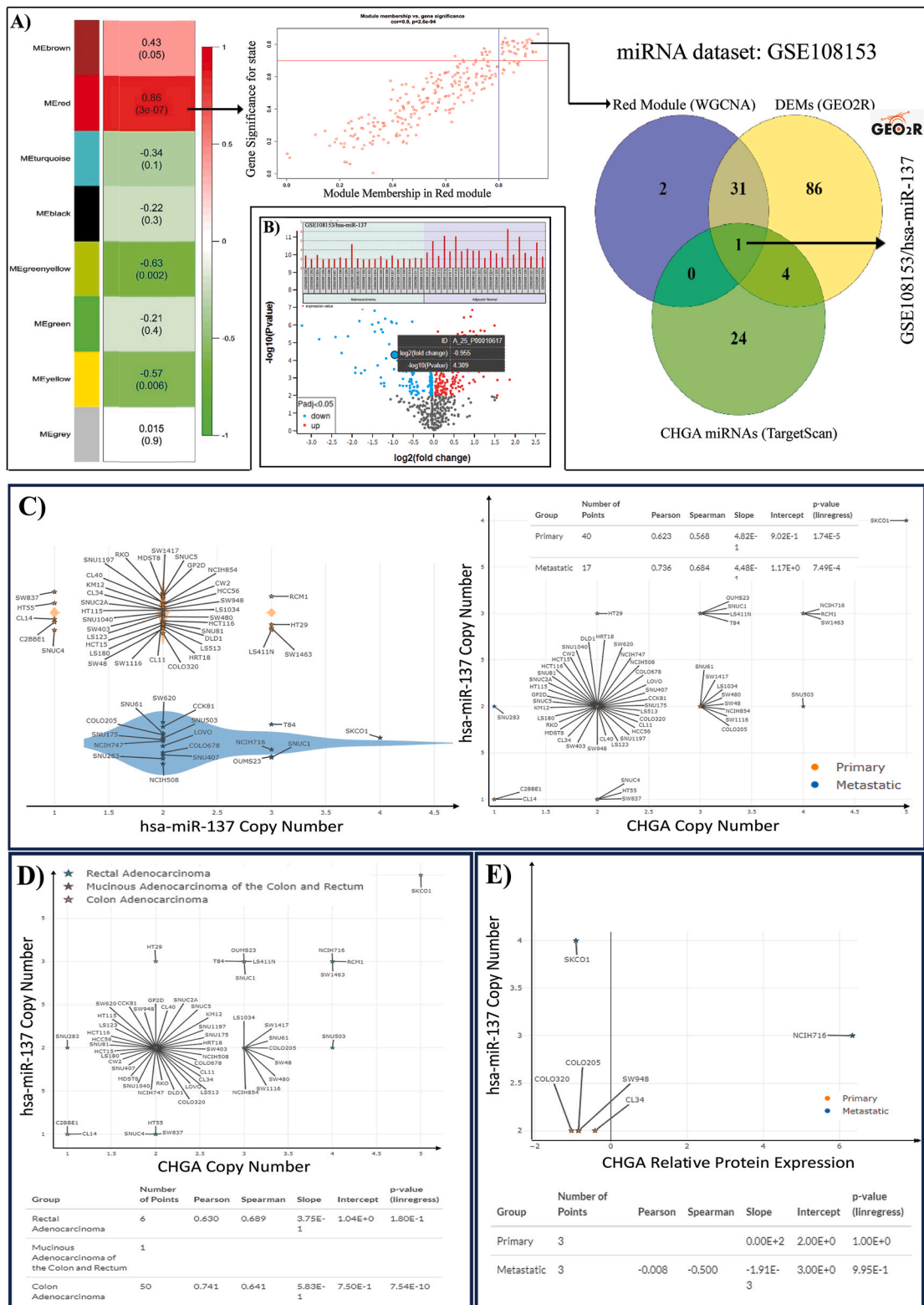
Fig. 4. WGCNA analysis and hub-miRNA selection for non-coding dataset. A) Module-trait relationship. Each row corresponds to a module eigengene, and the column corresponds to CRC status. The numbers in each cell represent the corresponding correlation and p-value; Module features of GS and MM. Each point represents an individual within each module, which are plotted by GS on the y-axis and gene MM on the x-axis; **B)** Visualization of GSE108153 DEG analysis. Volcanic diagram shows the significance of –log$_{10}$(p.val) on the y-axis and the increase and decrease

threshold of gene expression based on LogFC on the x-axis. Each gene is also marked as a dot (blue dots for upregulation and red dots for downregulation) on the graph; **C)** Distribution of hsa-miR-137 copy numbers across a diverse array of 57 bowel-related cell lines; **D)** uncovering potential transcriptomic correlations between *CHGA* and hsa-miR-137 within distinct CRC subtypes; **E)** Evaluation of significant correlation between *CHGA* and hsa-miR-137 at protein level.

pattern emerged in metastatic tissues, where *CHGA* expression was upregulated compared to normal tissues.

Consistent with these findings, data sourced from the CCLE database (https://portals.broadinstitute.org/ccle) portrayed a substantial upregulation of *CHGA* expression in primary cell lines (ECC4, SW1463, KM12, and TGBC18TKB) in contrast to metastatic cell lines (NCIH716, SNUC1, and SW626) (Fig. 2E). This cumulative evidence reinforces *CHGA*'s pivotal role in the context of CRC, as corroborated across multiple data sources and platforms.

3.6. Validation of CHGA expression changes by qRT-PCR on CRC tissues

To validate the precision and consistency of the transcriptome analysis, we opted to assess the transcript abundance of *CHGA* using qRT-PCR. Our investigation revealed a noteworthy reduction in *CHGA* expression within CRC samples compared to adjacent normal tissues (6.91-fold decrease, p-value = 0.002) (Fig. 2F). This alignment between the qRT-PCR outcomes and RNA-seq analyses substantiates the reliability of the RNA-seq data, reinforcing the accuracy and reproducibility of our findings.

3.7. Exploring CHGA expression patterns in single-cell RNA-seq data

The analysis of the Space-Time Gut Cell Atlas dataset, comprising approximately 428,000 cells, yielded a noteworthy finding. Specifically, our investigation highlighted the specific expression of *CHGA* within epithelial cells (Fig. 3A). Focusing on the epithelial cluster, which consisted of 142,113 cells, our examination revealed the presence of *CHGA* expression in distinct subclusters. Notably, these subclusters encompassed D cells (SST+) (N = 143), EC cells (NPW+) (N = 59), EC cells (TAC1+) (N = 422), EECs (N = 485), I cells (CCK+) (N = 227), and K cells (GIP+) (N = 83) (Fig. 3A).

Seeking to uncover the dynamics of *CHGA* expression throughout development, we turned to the Space-Time Gut Cell Atlas data. Fig. 3B unveiled a consistent trend across three developmental stages (Fetal, Pediatric, and Adult), with *CHGA* specifically expressed in epithelial cells (Enteroendocrine cells).

Transitioning to a more focused inquiry, we delved into the behavior of *CHGA* within Enteroendocrine cells across two distinct contexts: the colon (535,000 cells) and the small intestine (2.6 million cells). In the colon related Enteroendocrine cells (n = 305), which encompassed both normal and adenocarcinoma samples, a significant difference in *CHGA* expression emerged between the two groups. Specifically, *CHGA* expression was notably reduced in adenocarcinoma samples compared to normal cells, as depicted in Fig. 3C.

Remarkably, the story differed within the context of the small intestine related Enteroendocrine cells (n = 2300). Despite including various conditions such as normal, Crohn's disease, Crohn ileitis, and neuroendocrine carcinoma, *CHGA* expression exhibited no significant differences between these conditions and normal cells. Intriguingly, a significant decrease in *CHGA* expression was observed solely within adenocarcinoma samples in the intestine.

3.8. Identification of hsa-miR-137 as a potential non-coding candidate gene for CHGA

A set of 122 non-coding genes emerged as DEMs based on LogFC≥|0.5| and P-value <0.01.

The GSE108153 co-expression analysis involved 42 samples from distinct colorectal tissues. A subset of 4000 genes was analyzed for WGCNA, resulting in 20 sample outliers being excluded (Fig. S2A). To ensure a scale-free network, optimal β was determined via empirical analysis, stabilizing at 8 (Fig. S2B). Weighting coefficients were established, producing the disTOM for the 4000 genes. Through hierarchical clustering, 8 modules, each represented by a distinct color, were identified (Fig. S2C). CRC displayed a notable and compelling correlation, with the red module emerging as the primary and pivotal focus for our subsequent analytical endeavors (Fig. 4A).

Upon conducting a comprehensive search within the TargetScan database for the *CHGA* gene, a compilation of 29 specific miRNAs was unveiled.

Subsequently, a meticulous comparison was carried out among three distinct lists: the WGCNA red module output (34 miRNAs), the DEM list (122 miRNAs), and the TargetScan output (29 miRNAs) (Fig. 4A). Remarkably, the intersecting point among these lists converged on hsa-miR-137, designating it as the pivotal hub miRNA. This judicious selection was guided by its consistent presence across the three analyses, elevating hsa-miR-137 to a central role in the intricate regulatory network associated with our study.

The analysis conducted using GEO2R revealed a significant decrease in the expression of hsa-miR-137 within adenocarcinoma samples, as evidenced by the findings from GSE108153 (Fig. 4B).

Fig. 4C illustrates the distribution of hsa-miR-137 copy numbers across a diverse array of 57 bowel-related cell lines, encompassing both primary (40 cell lines) and metastatic COAD (17 cell lines) statuses. This valuable dataset was derived from the depmap portal's Copy Number Absolute resource. Notably, the presented graphical representation underscores a notable correlation between hsa-miR-137 copy numbers and *CHGA* copy numbers. In primary cell lines, this correlation is quantified by a Pearson correlation coefficient of 0.62, yielding a significant p-value of 1.74E-5. Meanwhile, among metastatic cell lines, the correlation becomes even more

pronounced, reaching a Pearson correlation coefficient of 0.73, with a p-value of 7.49E-4. These findings accentuate a robust and statistically significant relationship between hsa-miR-137 and *CHGA* copy numbers, particularly evident in the context of metastatic cell lines, thereby highlighting a potentially intricate interplay between these genetic factors in CRC progression.

Moreover, our exploration extended to leveraging the copy number absolute dataset, aiming to uncover potential transcriptomic correlations between *CHGA* and hsa-miR-137 within distinct CRC subtypes. Our investigation incorporated data from three distinct sources: rectal adenocarcinoma (n = 6 cell lines), mucinous adenocarcinoma of the colon and rectum (n = 1 cell line), and colon adenocarcinoma (n = 50 cell lines). In rectal adenocarcinoma cell lines, the correlation between *CHGA* and hsa-miR-137 is quantified by a Pearson correlation coefficient of 0.63. However, the accompanying p-value of 1.80E-1 renders this correlation statistically non-significant. In contrast, among colon adenocarcinoma cell lines, the correlation strengthens further, evidenced by a Pearson correlation coefficient of 0.74. This heightened correlation is underscored by a markedly significant p-value of 7.54E-10 (as depicted in Fig. 4D). These outcomes spotlight the intricate interplay between *CHGA* and hsa-miR-137 in CRC subtypes, with particularly noteworthy significance observed in the context of colon adenocarcinoma cell lines.

Delving deeper, we focused on the subset of cell lines (6 out of the initial 57) that had proteomic data available. Within this subset, which included COLO205, SW948, COLO320, CL34, SKCO1, and NCIH716, the examination extended to the protein level. Regrettably, our endeavors once again revealed the absence of a significant correlation between *CHGA* and hsa-miR-137 (Fig. 4E).

4. Discussion

CRC is one of the most common cancers worldwide, and advances in diagnostic approaches can be very effective in the treatment of patients. The accumulation of genetic and epigenetic abnormalities predisposes the colon epithelium to gradual changes with loss of cellular structure and the emergence of a benign adenoma, which can progress towards a malignant adenocarcinoma. The five-year survival rate for early-stage patients is over 90%, while it is only about 10% for later-stage patients. Approximately half of CRC patients are diagnosed in the later stages of the disease. As a result, improving the ability to detect CRC early is critical [2].

In this study we used WGCNA to identify most important hub-genes and their related hub-miRNAs in CRC. Among the hub genes, chromogranin A (*CHGA*; CgA) was selected based on the criteria that this gene was more important in terms of survival index [23]. On the other hand, miR-137 is one of the miR there that target this gene (based on the TargetScan database). Also, in terms of expression patterns, they had a logical match with each other.

CHGA is a 49 kDa glycoprotein gene located on chromosome 14 that is primarily expressed by endocrine and neuroendocrine cells [24]. *CHGA* can be secreted into the blood in its entirety or in small fragments after fragmentation. Although the precise function of *CHGA* peptides is not fully understood, they are believed to play a key role in cardiac function, catecholamines, parathyroid hormone secretion, carbohydrates, fat metabolism, immune properties, and reproduction. Chromogranin A and B have been proposed as promising biomarkers for CRC detection in the early stages [23].

In a meta-analysis study conducted by Rossi RE et al. (2018), *CHGA* as a follow-up marker demonstrated a sensitivity of 46%–100% and a specificity of 68%–90% in gastrointestinal malignancies, concluding that *CHGA* is more reliable in monitoring disease progression and response to treatment, as well as for early detection of relapse after treatment [25]. Based on a study on patients with pancreatic cancer conducted by Lee SH et al. (2018), they discovered that people with high levels of *CHGA* had higher rates of metastasis and that these intermediate patients had shorter overall survival. According to the findings of this study, high *CHGA* levels predict a poor prognosis in patients with pancreatic cancer, particularly in the metastatic stage [26].

In contrast to previous research, Zhang X et al. (2019) discovered that *CHGA* expression is significantly lower in the early stages of CRC when compared to healthy controls. This decrease in expression in the early stages of CRC resulted in the development of a new CRC diagnostic biomarker with high predictability and validated function. According to an examination of the protein-protein interaction network (PPI), *CHGA* is also involved in some of the pathways associated with *KRAS* and *TP53*. *CHGA* may also be regarded as a novel, promising, and potent biomarker for the early detection of CRC [23].

Herold et al. (2020) discovered that right-sided colon tumors were more common in the *CHGA*[+] group than in the *CHGA*[−] group after dividing CRC patients into two groups based on whether *CHGA* was positive or negative. Furthermore, stage I cancer was not found in this group, and its patients had poorer overall health. The prognosis for survival level was also significantly weak in this group. Furthermore, based on these findings, a new subtype of CRC was identified by differentiating *CHGA*-positive neuroendocrine cells [27].

Changing the expression of *CHGA* in different diseases yields different results. The mean plasma *CHGA* level in SCLC (small cell lung cancer), for example, is higher than in healthy specimens as well as in patients with chronic obstructive pulmonary disease, pulmonary adenocarcinoma, and large cell lung carcinoma. This implies that *CHGA* levels are related to disease levels. Studies of *CHGA* levels in breast cancer have shown that *CHGA* is present in both normal granular epithelial cells and breast cancer tissues. According to these studies, serum *CHGA* is not sensitive enough to identify subtypes of breast cancer that are rarely diagnosed with neuroendocrine therapy. Because of its early expression throughout the neuroendocrine system, *CHGA* has also been widely accepted as a biomarker for the evaluation of neuroendocrine tumors. It is also commonly used in the management of patients with gastro-intestinal or pancreatic tumors [28].

This study elucidated the intricate interplay between *CHGA* and hsa-miR-137 in CRC pathogenesis. Our exploration of hsa-miR-137 copy numbers across bowel-related cell lines highlighted a correlation with *CHGA* copy numbers. Transcriptomic correlations between *CHGA* and hsa-miR-137 in CRC subtypes unveiled context-dependent relationships, particularly in colon adenocarcinoma cell lines. Despite efforts at the protein level using a subset of cell lines with proteomic data, no significant correlation was found. These findings collectively underscore the complex interplay of *CHGA* and hsa-miR-137 in CRC, implying their potential roles as significant

biomarkers in CRC progression.

Mahmoudi and Cairns (2017) describe miR-137 as one of several miRNAs that play important roles in cellular biology [29]. According to research, miR-137 may play a dual role during tumorigenesis, the nature of which may vary depending on the type of tumor and the identity of the target messenger RNA (mRNA) [30]. Kashani et al. (2019) discovered that miR-137 acts as a tumor suppressor miRNA through hypermethylation events, and it was discovered that the methylation status of this miRNA has diagnostic and prognostic value in CRC [31].

Zhao et al. (2018) also demonstrated that decreasing and sponging miR-137 through *SNHG1* increases *RICTOR* expression and promotes tumorigenesis in CRC [30]. The results of RT-qPCR in Fasihi et al. (2018) also revealed a decrease in the expression of miR-137 in CRC tissues. Hsa-miR-137 was also identified as a tumor suppressor that works in CRC by regulating the Wnt signaling pathway [32]. Another study discovered that miR-137 regulates CRC cell proliferation, colony formation, migration, invasion, and metastasis [33]. Huang et al. (2016) discovered that miR-137 was silenced in human CRC tissues and polyps. Furthermore, decreased expression of miR-137 in various types of polyps leads to the development of CRC, and miR-137 expression is gradually reduced during the process of colon carcinogenesis [34]. According to ROC curve analysis, the loss of miR-137 expression in colon polyps can also be used as a biomarker to predict CRC potential. These studies' results confirm our findings, demonstrating that the increase in miR-137 expression and the decrease in *CHGA* gene expression are consistent, and that these two factors reinforce each other's effects in preventing CRC. Based on the studies, both *CHGA* gene and miR-137 have been investigated and evaluated separately in all types of cancers, especially CRC. In this study, we tried to have a novel approach of using the relationship between this gene and miR-137 and investigate its effect on CRC. More studies and tests, however, are required to confirm these findings in order to obtain more accurate information.

5. Conclusion

In this study, we employed the WGCNA method to identify gene modules and hub-genes that are specifically associated with the pathological mechanisms of CRC. Using this approach, we identified the *CHGA* gene as a key player in the development of CRC. We validated the expression level of *CHGA* in CRC patients using RT-qPCR analysis, which confirmed that *CHGA* was significantly downregulated in CRC tissues compared to non-cancerous tissues. Furthermore, we found that hsa-miR-137 is a potential regulatory element of *CHGA*, suggesting that the downregulation of *CHGA* may be mediated by this microRNA in CRC patients. Our findings highlight the importance of investigating the role of *CHGA*-specific non-coding RNAs in the pathogenesis of CRC. Overall, our results suggest that *CHGA* downregulation may be a crucial event in the development of CRC, and could serve as a potential diagnostic biomarker for CRC patients. We recommend that future studies should explore the protein level of *CHGA* and the role of *CHGA*-specific non-coding RNAs through more extensive investigations. Additionally, targeting the *CHGA* gene may represent a promising therapeutic strategy for the treatment of CRC.

Ethical considerations

ethical considerations and patients' personal satisfaction were observed based on Helsinki Declaration which were approved by the ethics committee of Birjand University of Medical Sciences (IR.BUMS.REC.1398.139). (https://ethics.research.ac.ir/EthicsProposalView.php?id=75410).

Data Availability

The datasets analyzed during the current study are public and available as below:
https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE81558;
https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE110224;
https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE108153;
Cellxgene collections; https://cellxgene.cziscience.com.

CRediT authorship contribution statement

Hossein Safarpour: Performed the experiments; Wrote the paper; Analyzed and interpreted the data. **Javad Ranjbaran:** Performed the experiments; Wrote the paper; Analyzed and interpreted the data. **Nafiseh Erfanian:** Performed the experiments; Wrote the paper; Analyzed and interpreted the data. **Samira Nomiri:** Performed the experiments; Wrote the paper; Analyzed and interpreted the data. **Afshin Derakhshani:** Contributed reagents, materials, analysis tools or data; Wrote the paper. **Casimiro Gerarduzzi:** Contributed reagents, materials, analysis tools or data; Wrote the paper. **Adib Miraki Feriz:** Contributed reagents, materials, analysis tools or data; Wrote the paper. **Edris HosseiniGol:** Contributed reagents, materials, analysis tools or data; Wrote the paper. **Samira Saghafi:** Contributed reagents, materials, analysis tools or data; Wrote the paper. **Nicola Silvestris:** Conceived and designed the experiments; Wrote the paper.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e27046.

References

[1] R.L. Siegel, K.D. Miller, A. Jemal, Cancer statistics, 2020, CA Cancer J Clin 70 (1) (2020) 7–30.
[2] M. Eileen, et al., Global burden of colorectal cancer in 2020 and 2040: incidence and mortality estimates from GLOBOCAN, Gut 72 (2) (2023) 338.
[3] M. Ahmed, Colon cancer: a clinician's perspective in 2019, Gastroenterol. Res. 13 (1) (2020) 1.
[4] A. Hadjipetrou, et al., Colorectal cancer, screening and primary care: a mini literature review, World J. Gastroenterol. 23 (33) (2017) 6049–6058.
[5] E. Danese, M. Montagnana, G. Lippi, Circulating molecular biomarkers for screening or early diagnosis of colorectal cancer: which is ready for prime time? Ann. Transl. Med. 7 (21) (2019) 610.
[6] A.C. Society, American Cancer Society (2021).
[7] G. Lech, et al., Colorectal cancer tumour markers and biomarkers: recent therapeutic advances, World J. Gastroenterol. 22 (5) (2016) 1745–1755.
[8] J.R. Bailey, A. Aggarwal, T.F. Imperiale, Colorectal cancer screening: stool DNA and other noninvasive Modalities, Gut Liver 10 (2) (2016) 204–211.
[9] A.L. Zygulska, P. Pierzchalski, Novel diagnostic biomarkers in colorectal cancer, Int. J. Mol. Sci. 23 (2) (2022).
[10] J.Y. Wang, et al., Comprehensive analysis of microRNA/mRNA signature in colon adenocarcinoma, Eur. Rev. Med. Pharmacol. Sci. 21 (9) (2017) 2114–2129.
[11] F. Wu, et al., Network analysis based on TCGA reveals hub genes in colon cancer, Contemp. Oncol. 21 (2) (2017) 136–144.
[12] X.G. Zhou, et al., Identifying miRNA and gene modules of colon cancer associated with pathological stage by weighted gene co-expression network analysis, OncoTargets Ther. 11 (2018) 2815–2830.
[13] G. Businello, F. Galuppini, M. Fassan, The impact of recent next generation sequencing and the need for a new classification in gastric cancer, Best Pract. Res. Clin. Gastroenterol. 50–51 (2021) 101730.
[14] L. Bo, et al., Screening of critical genes and MicroRNAs in blood samples of patients with Ruptured Intracranial Aneurysms by Bioinformatic analysis of gene expression data, Med Sci Monit 23 (2017) 4518–4525.
[15] X. Liu, et al., Identification of key gene modules in human Osteosarcoma by Co-expression analysis weighted gene Co-expression network analysis (WGCNA), J. Cell. Biochem. 118 (11) (2017) 3953–3959.
[16] Y. Tang, et al., Co-expression analysis reveals key gene modules and pathway of human coronary heart disease, J. Cell. Biochem. 119 (2) (2018) 2102–2109.
[17] Z. Rezaei, et al., Identification of early diagnostic biomarkers via WGCNA in gastric cancer, Biomed. Pharmacother. 145 (2022) 112477.
[18] S. Nomiri, et al., Prediction and validation of GUCA2B as the hub-gene in colorectal cancer based on co-expression network analysis: in-silico and in-vivo study, Biomed. Pharmacother. 147 (2022) 112691.
[19] W.E. Johnson, C. Li, A. Rabinovic, Adjusting batch effects in microarray expression data using empirical Bayes methods, Biostatistics 8 (1) (2007) 118–127.
[20] C. Chen, et al., Removing batch effects in analysis of expression microarray data: an evaluation of six batch adjustment methods, PLoS One 6 (2) (2011) e17238.
[21] P. Langfelder, S. Horvath, WGCNA: an R package for weighted correlation network analysis, BMC Bioinf. 9 (1) (2008) 1–13.
[22] T.H.P. Atlas, THE HUMAN PROTEIN ATLAS, 2022.
[23] X. Zhang, et al., Chromogranin-A expression as a novel biomarker for early diagnosis of colon cancer patients, Int. J. Mol. Sci. 20 (12) (2019).
[24] A. Ebert, et al., Chromogranin serves as novel biomarker of endocrine and gastric Autoimmunity, J. Clin. Endocrinol. Metab. 105 (8) (2020).
[25] R.E. Rossi, et al., Chromogranin A in the follow-up of Gastroenteropancreatic neuroendocrine Neoplasms: is it Really Game over? A Systematic review and meta-analysis, Pancreas 47 (10) (2018) 1249–1255.
[26] S.H. Lee, et al., Plasma chromogranin A as a prognostic marker in pancreatic Ductal adenocarcinoma, Pancreas 48 (5) (2019) 662–669.
[27] Z. Herold, et al., Histopathological chromogranin A-Positivity is associated with right-sided colorectal cancers and Worse prognosis, Cancers 13 (1) (2020).
[28] S. Gkolfinopoulos, et al., Chromogranin A as a valid marker in oncology: clinical application or false hopes? World J. Methodol. 7 (1) (2017) 9–15.
[29] E. Mahmoudi, M.J. Cairns, MiR-137: an important player in neural development and neoplastic transformation, Mol Psychiatry 22 (1) (2017) 44–55.
[30] S. Zhao, et al., Long noncoding RNA small Nucleolar RNA Host gene 1 (SNHG1) promotes renal cell carcinoma progression and metastasis by negatively regulating miR-137, Med Sci Monit 24 (2018) 3824–3831.
[31] E. Kashani, et al., The differential DNA hypermethylation patterns of microRNA-137 and microRNA-342 Locus in early colorectal Lesions and Tumours, Biomolecules 9 (10) (2019).
[32] A. Fasihi, et al., Introduction of hsa-miR-103a and hsa-miR-1827 and hsa-miR-137 as new regulators of Wnt signaling pathway and their relation to colorectal carcinoma, J. Cell. Biochem. 119 (7) (2018) 5104–5117.
[33] T. Chen, et al., Mecp2-mediated epigenetic silencing of miR-137 contributes to colorectal adenoma-carcinoma sequence and tumor progression via Relieving the suppression of c-Met, Sci. Rep. 7 (2017) 44543.
[34] Y.C. Huang, et al., Epigenetic silencing of miR-137 contributes to early colorectal carcinogenesis by impaired Aurora-A inhibition, Oncotarget 7 (47) (2016) 76852–76866.