Am J Cancer Res 2024;14(3):1402-1418
www.ajcr.us /ISSN:2156-6976/ajcr0154417

Original Article

Multi-omics data-based analysis characterizes molecular alterations of the vesicle genes in human colorectal cancer

Xi Wang[1*], Min-Min Wu[2,3*], Wei Zhang[1,2*], Zhen-Qiong Liu[3], Meng-Qi Xu[3], Feng-Mei Zhang[3], Zhi-Qiang He[3], Dong-E Tang[1], Min Tang[3], Yong Dai[1,4,5]

[1]Clinical Medical Research Center, The First Affiliated Hospital (Shenzhen People's Hospital), Southern University of Science and Technology, Shenzhen 518055, Guangdong, China; [2]Health Science Center, South China Hospital, Shenzhen University, Shenzhen 518020, Guangdong, China; [3]Key Laboratory of Diagnostic Medicine Designated by The Chinese Ministry of Education, Chongqing Medical University, Chongqing 400016, China; [4]The First Affiliated Hospital, School of Medicine, Anhui University of Science and Technology, Huainan 232001, Anhui, China; [5]Department of Rheumatism and Immunology, The Key Laboratory of Inflammatory and Immunology Diseases, Peking University Shenzhen Hospital, Shenzhen 518036, Guangdong, China. [*]Equal contributors and co-first authors.

Received November 14, 2023; Accepted February 5, 2024; Epub March 15, 2024; Published March 30, 2024

Abstract: The role of vesicular genes in the development of colorectal cancer (CRC) is crucial. Analyzing alterations in these genes at multi-omics can aid in understanding the molecular pathways behind colorectal carcinogenesis and identifying potential treatment targets. However, studies on the overall alteration of vesicular genes in CRC are still lacking. In this study, we aimed to investigate the relationship between vesicle genetic alterations and CRC progression. To achieve this, we analyzed molecular alterations in CRC vesicle genes at eight levels, including mRNA, protein, and epigenetic levels. Additionally, we examined CRC overall survival-related genes that were obtained from a public database. Our analysis of chromatin structural variants, DNA methylation, chromatin accessibility, and proteins (including phosphorylation, ubiquitination, and malonylation), along with RNA-seq data from the TCGA database, revealed multiple levels of alterations in CRC vesicle genes in the collected tissue samples. We progressively examined the alterations of vesicle genes in mRNA and protein levels in CRC and discovered the hub genes. Further investigation identified the probable essential transcription factors. This study contributes to a thorough knowledge of the connection between vesicle gene alterations at multiple levels and the development of CRC and offers a theoretical framework for the identification of novel treatment targets.

Keywords: Human colorectal cancer, vesicle genes, multi-omics analysis, mechanisms of disease, therapeutic targets

Introduction

Colorectal cancer (CRC) ranked third in morbidity and second in mortality [1]. The main therapies for CRC are surgery and chemotherapy. However, its pathophysiology is complex and predisposes it to metastasis [2]. Unfortunately, around 25% of CRC patients experience distant metastases within the first year of therapy [3, 4]. Early detection is crucial to prompt treatment and to reduce the risk of metastases. Therefore, identifying potential targets for early detection of CRC requires a comprehensive understanding of the molecular processes involved in its development.

Vesicles, which are small organelles surrounded by a lipid bilayer, are critical to the growth and spread of CRC [5, 6]. Research has indicated that secretory vesicles, including exosomes, play a significant role in facilitating communication between tumor cells and the tumor microenvironment [7]. Studies have demonstrated that secretory vesicles can transport nucleic acids, proteins, and various other substances to distant target organs, creating pre-metastatic niches (PMN) that provide a survival advantage to tumor cells [8]. Additionally, endosomes and lysosomes are utilized for the uptake and degradation of cargo from both inside and outside the cell through endocytosis, phagocyto-

https://doi.org/10.62347/QFKD1805

sis, and autophagocytosis, all of which are linked to the development of colorectal cancer [2, 9]. In our previous research, we discovered that phosphorylation differential protein enrichment in CRC tumor cells is primarily responsible for the endocytosis process, which is critical in cell vesicle transport [10]. Therefore, it is essential to monitor the molecular alterations in CRC vesicle proteins for early diagnosis and treatment. However, there is still a lack of comprehensive studies on the alterations in vesicle proteins in CRC.

In this study, we analyzed multiple omics data from colorectal cancer tissue and normal paracancerous tissue cells (n=6-8). Our findings indicate a significant correlation between vesicular-related molecules in CRC tissue cells and mRNA, proteomics, DNA methylation, chromatin structural variation, chromatin openness, protein ubiquitination, as well as changes in protein phosphorylation and malonylation. In this study, we have combined our original data with the RNA-seq dataset obtained from The Cancer Genome Atlas (TCGA) of colorectal cancer patients (n=538). We investigated the changes in vesicle-related molecules across various omics levels in CRC patients and analyzed the protein interaction network and hub genes. Furthermore, we identified potential upstream transcription factors that regulate genes related to vesicles.

Material and methods

Patients

Samples of colorectal cancer were collected from Shenzhen People's Hospital with the approval of the ethics committee (No. LL-KY-2019213). The study participants were all volunteers who signed an informed consent form. The research included patients who had surgical resection without the use of chemotherapy or radiation, and those who were excluded due to hereditary factors. Each instance received a pathological analysis grade. Colorectal cancer tumor tissue sampling procedures were all completed within one hour of the patient's surgery. Both tumor and normal adjacent tissues were obtained from the colonic segment, and normal adjacent colorectal mucosa was collected at a distance of 5 cm from the tumor. The clinical information of the colorectal cancer (CRC) patients of this study was shown in Supplementary Table 1.

Whole genome bisulfite sequencing (WGBS)

The SMRTbell Express Template Prep Kit 2.0 (Pacific Bioscience, 100-938-900) was used for extracting genomic DNA and assessing its quality and concentration. The DNA samples were sonicated into smaller fragments and then converted with bisulfite. After that, the single-stranded DNA fragments were ligated using the EZ DNA Methylation-GoldTM Kit (ZYMO research, D5005) before undergoing PCR amplification. The HiSeq X10 sequencing platform (Illumina) was used for double-end sequencing of the sequencing libraries. The amplified products were then purified and checked for integrity using an Agilent 2100 Bioanalyzer. Finally, the outcomes were compared using differentially methylated areas.

Long-read whole genome sequencing

Genomic DNA of high quality was obtained and fragmented into 20 kb pieces (g-TUBETM, Covaris, 520079), which were enriched using magnetic beads (Pacific Bioscience, 100-317-100). Stem loops were added to the DNA fragments (Pacific Bioscience, 101-731-100) to facilitate sequencing by DNA polymerase. The sequencing process was carried out using the HiSeq X10 sequencing technology (Illumina), which generates paired-end 150 BP reads, after assessing the quality of the DNA.

Assay to detect transposase-accessible chromatin based on high throughput sequencing (ATAC-Seq)

The ATAC-seq procedure was executed following standard operating procedures. Initially, 50,000 single cell suspensions underwent purification and transposition. The remaining cell precipitate was combined with 50 L of Tn5 transposase and 1× TD buffer after cell lysis. The mixture was then incubated at 37°C for 30 minutes. Subsequently, the sample DNA was processed using the MinElute reaction clearing kit (QIAGEN, 51306). The TruePrep DNA Library Preparation Kit V2 by Vazyme Biotecho (TD501/TD502/TD503) was utilized to generate DNA libraries for Illumina sequencing. The quality of the libraries was evaluated in real-time using the StepOnePlus Real-Time Polymerase Chain

Reaction System. The length of the insert fragments was determined using the HS 2100 Bioanalyzer by Agilent. High-quality libraries generated from 150 BP paired-end readings were sequenced using the HiSeq X Ten sequencing technology (Illumina).

RNA-Seq dataset from the TCGA database

In this study, RNA-Seq datasets of 647 CRC patients and 51 healthy individuals were obtained from the TCGA database. Samples with uncertain survival periods were excluded, resulting in a total of 538 sample data files. The individual files were merged into a matrix file for analysis. Ensembl IDs were converted to gene symbols using the Ensembl database and separate files were merged into matrix files. The symbol matrix was utilized to extract datasets for both normal and CRC individuals. The R package 'Survival V' was employed to perform overall survival analysis.

Protein extraction and digestion

Tissue samples were processed by combining them with lysis buffer containing 1% cocktail and 8 M urea (Merck Millipore, 156535140). The resulting mixture was sonicated and then centrifuged at 12,000 g at 4°C for 10 minutes to remove debris. The protein-containing sample was subjected to treatment with 5 mM dithiothreitol at 56°C for 30 minutes. Following this, it was combined with 11 mM iodoacetamide and incubated in the dark at room temperature for an additional 15 minutes. The combination was then digested with trypsin at a 1:50 ratio for 12 hours, followed by a 1:100 ratio of trypsin for 4 hours. The resulting sample was then analyzed using LC-MS/MS.

Lysine ubiquitylated peptide enrichment

To dissolve the tryptic peptides, a NETN buffer was used, which contained 100 mM NaCl, 1 mM Ethylene diamine tetraacetic acid (EDTA), 50 mM Tris-HCl, and 0.5% Nonidet P (NP-40) at pH 8.0. The solution was then incubated with prewashed K-ε-GG antibody beads (PTM Biolabs) and gently shaken at 4°C overnight. The beads were washed four times with NETN buffer and twice with ddH_2O. The peptides were extracted from the beads using 0.1% trifluoroacetic acid (TFA). The resulting fractions were combined and dried using a vacuum. The peptides were then purified with C18 STAGE tips (Thermo Fisher). The peptides were analyzed using an EASY-nLC 1000 UPLC system coupled to a Q-Exactive plus hybrid quadrupole-Orbitrap mass spectrometer through online LC-MS/MS analysis.

Trypsin digestion and malonylation modification enrichment

Trichloroacetic acid was gradually added to each sample until a final concentration of 20% was achieved. The samples were then vortexed, and equal amounts of protein were precipitated at 4°C for a duration of 2 hours. The protein pellet was separated from the supernatant by centrifugation and washed with cold acetone for 1-2 times. After air drying the protein pellets, 200 mM triethylammonium bicarbonate (TEAB) was added to each sample, and the protein pellet was resuspended using ultrasonic sonication. Trypsin was added to the sample at a ratio of 1:50 (protease:protein, M/M) and left overnight. To prepare the protein for analysis, it was first reduced using DTT and then alkylated with iodoacetamide (IAA) at a final concentration of 11 mM. The resulting peptides were dissolved in an IP buffer solution and added to a cleaned pan anti-malonylysine antibody resin (PTM Bio, No. PTM-904). The peptide solution and antibody bead combination were then shaken overnight at 4°C. Following incubation, the resin was washed twice with deionized water and four times with IP buffer solution. The peptides bound to the resin were subsequently eluted three times using 0.1% trifluoroacetic acid. The resulting eluent was collected and dried under a vacuum. The peptides were further desalted using C18 tips and prepared for LC-MS/MS analysis.

LC-MS/MS analyses

LC-MS/MS analysis was performed using a standardized procedure. Peptide samples were digested with trypsin and dissolved in liquid chromatography mobile phase A (0.1% formic acid). An ultra-high phase apparatus was utilized for separation, and the resulting mixture was ionized using an NSI ion source before being fed into a timsTOF Pro mass spectrometer for examination. The mass spectrometry (MS) scan range used in this study was 100 to 1700 m/z. Fragmentation of precursors with charge states ranging from 0 to 5 was per-

formed, and the secondary mass spectra's dynamic exclusion was set to 30 seconds. To obtain secondary mass spectrometry data, Maxquant (v1.6.6.0) was used with the Homo sapiens 9606 SP 20191115 database containing 20380 sequences.

In this study, cysteine alkylation was considered a fixed modification, while methionine oxidation was considered a variable modification. For the analysis of ubiquitination, cysteine carbamidomethylation was designated as a fixed modification, while methionine oxidation, lysine ubiquitination, and N-terminal methionine ubiquitination of the protein were designated as variable modifications. In the malonylation analysis, Carbamidomethyl on Cys was designated as a fixed modification, while Malonyllysine on Lys and oxidation on Met were designated as variable modifications. For the phosphorylated proteomics analysis, Ethyl carbamate on Cys was designated as a fixed modification, while oxidation of Met, acetylation of protein N terminus, and phosphorylation of Ser, Thr, and Tyr were designated as variable modifications.

Data availability

The ATAC-Seq, WGBS, and long-read whole genome sequencing result data obtained in this study have been deposited in the Sequence Read Archive (SRA) under the number PRJNA693028. The mass spectrometry proteomics data of ubiquitination and malonylation have been unloaded and deposited onto the ProteomeXchange Consortium through the PRIDE partner repository: PXD028504 and PXD036533, respectively. The raw data of proteomic and phosphorylated proteomic also have been uploaded to the ProteomeXchange Consortium under accession numbers PXD021314 and PXD021318, respectively.

Statistical analyses

We used the *t*-test to compare gene expression, proteome composition, and phosphorylated proteome composition between the initial data collation done in Excel 2016 and the subsequent analysis performed using SPSS 25.0. To determine the significance level of the enrichment, we utilized the Fisher Exact Test and considered it significant at $P < 0.05$.

Results

Multi-omics analysis revealed a tremendous molecular alteration of vesicle genes in CRC patients

To comprehensively investigate the molecular alterations of vesicle genes (VGs) in colorectal cancer (CRC) tumor cells, we conducted a multi-omics analysis on tumor tissue samples collected from patients with CRC and adjacent paracancerous tissue. The analysis involved proteomics (n=8), DNA methylation (n=6), chromatin structural variation (n=6), chromatin openness (n=6), and changes in protein modification levels, including protein ubiquitination (n=8), protein phosphorylation (n=8), and protein malonylation (n=8), as shown in **Figure 1A**. This study collected surgical samples of CRC tumor tissue and paracancerous tissue from twelve individuals, eight of whom had early stage (I-II) cancer and four had advanced stage (III-IV) cancer. To ensure adequate sample size, two patients' samples from each stage of the disease were combined into one sample for testing.

To compare mRNA expression levels in tumor tissues of CRC patients with those in normal tissues, we downloaded RNA-seq datasets and survival data of 538 CRC patients and 51 healthy individuals from the TCGA database. Our findings indicate that 691 VGs' mRNA expression levels were up-regulated while 513 VGs' mRNA expression levels were down-regulated (**Figure 1B**). Additionally, 66 VGs had increased protein expression levels while 73 VGs had decreased protein expression levels (**Figure 1C**). The study also found that out of the total VGs analyzed, 256 showed hypermethylation while 432 showed hypomethylation. Additionally, chromatin openness was up-regulated in 565 VGs and down-regulated in 67 VGs. The levels of ubiquitin were found to be up-regulated in 32 VGs and down-regulated in 57 VGs. In terms of structural changes in CRC chromatin, 32 changes were attributed to exons while 210 were attributed to non-exons. These findings are depicted in **Figure 1D-G**. In this study, we observed that the phosphorylation levels of 21 and 29 VGs were up- and down-regulated, respectively (**Figure 1H**), and the malonylation levels of 4 and 10 VGs were up- and down-regulated, respectively (**Figure 1I**). To investigate the potential correlation

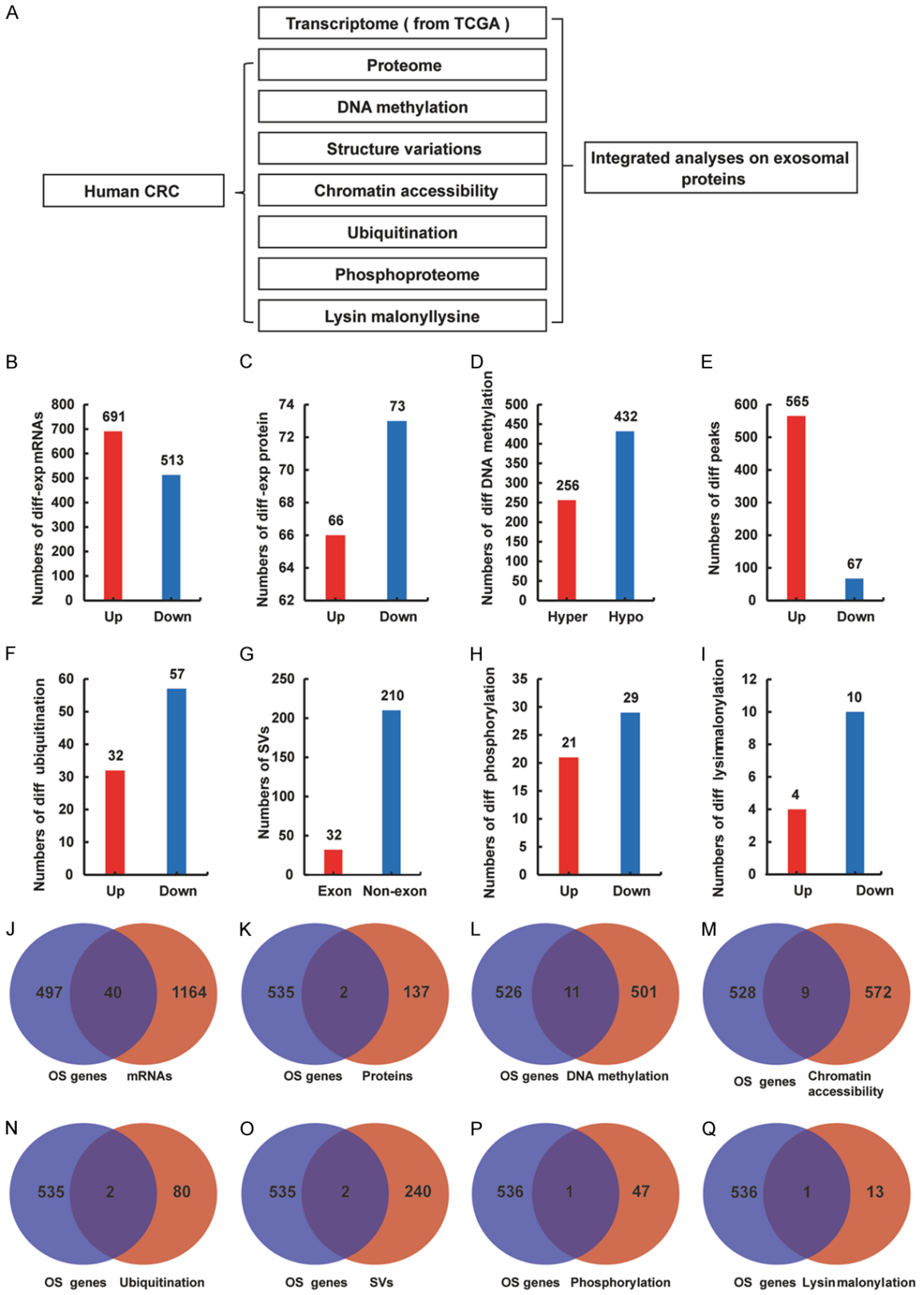


Figure 1. A multi-omics overview of alterations in vehicle genes (VGs) in CRC. (A) General overview of the analytical procedure. Histogram showed the changes of VGs in mRNA (B), protein (C), DNA methylation (D), chromatin accessibility (E), ubiquitination (F), chromatin structural variation (G), phosphorylation (H), and malonylation (I) in CRC. Venn

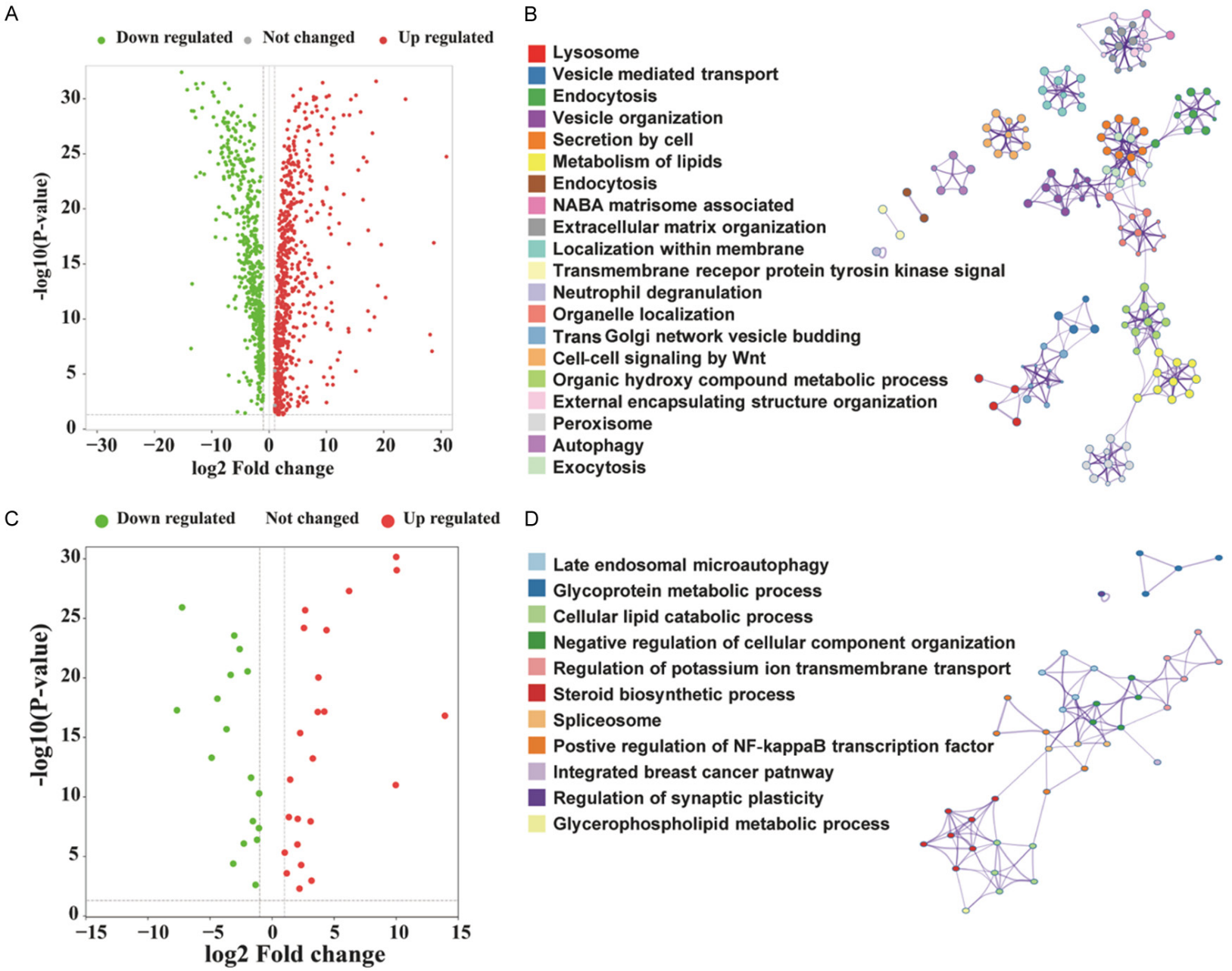
plots of the changes of VGs in mRNA (J), protein (K), DNA methylation (L), chromatin accessibility (M), ubiquitination (N), chromatin structural variation (O), phosphorylation (P), and malonylation (Q) in CRC and CRC survival-related genes.

between these molecular changes in VGs and the prognosis of CRC patients, we analyzed RNA-seq datasets obtained from TCGA. Our findings revealed that the overall survival rate of CRC patients was significantly associated with 537 overall survival rates-related genes (OS genes) (Supplementary Table 2). The Venn diagrams in **Figure 1J-Q** illustrate the overlap between the 537 genes and the VGs that experienced changes at the omics level. The study revealed that changes in mRNA levels, protein expression, DNA methylation, chromatin openness, protein ubiquitination, chromatin structure, protein phosphorylation, and protein malonylation in VGs are all linked to the overall survival rate of CRC. The corresponding numbers for each association are as follows: 40 for mRNA level alteration, 2 for protein expression change, 11 for DNA methylation difference, 9 for chromatin openness to change, 2 for protein ubiquitin alternation, 2 for chromatin structure variation, and 1 each for protein phosphorylation change and protein malonylation (Supplementary Tables 3, 4, 5, 6, 7, 8, 9 and 10). During the progression of CRC, there were significant changes observed in the expression levels of survival-related genes P4HA1 and IDUA. The degree of phosphorylation of CLINT1 at 299 Ser increased, while the level of ubiquitination decreased. Additionally, there was a reduction in ubiquitination levels at several loci of HSPA8. Furthermore, malonylation levels were observed to increase at the 493 Lys on GSPT1. Overall, these findings suggest that there were notable alterations in VGs during the progression of CRC.

Transcriptome analysis revealed that the mRNA levels of VGs in CRC patients' tumor tissues were significantly altered, and HSPA8 and HSPA1A might be potential hub genes

Based on the results presented in **Figure 1**, it was observed that the VGs in CRC displayed significant changes in mRNA levels and had the largest number of differential genes associated with survival. Subsequently, we conducted a further analysis of the vesicles with mRNA level changes in CRC, which revealed the presence of 691 mRNA up-regulated VGs and 513 down-regulated genes, as depicted in **Figure 2A**. Notably, the number of up-regulated genes was found to be much greater than the number of down-regulated genes. We utilized Metascape to examine the protein interactions between CRC and paracancerous tissues, focusing on differentially expressed VGs mRNA levels as shown in **Figure 2B**. Our findings indicate that the main biological processes impacted by these differentially expressed VGs in CRC include lysosome, vesicle-mediated transport, and endocytosis (**Figure 2B**). Additionally, we conducted a thorough analysis of the 40 intersected genes from **Figure 1J** in order to identify key genes associated with CRC survival. The results of the volcanic map analysis demonstrated that 23 VGs were up-regulated and 17 VGs were down-regulated in relation to the survival of CRC cancer patients (**Figure 2C**). Subsequently, we conducted a protein interaction network analysis for these 40 genes (**Figure 2D**) and utilized Cytohubba to identify key interaction genes (**Figure 2E**, **2F**). The top 10 key genes identified were HSPA8, GSPT1, RNASE1, NOXA1, EFTUD2, HSPA1A, VPS37D, AHD6, GDE1, and DPP7 (**Figure 2E**, **2F**).

In patients with CRC, immune cell infiltration is associated with a poorer prognosis [11]. Therefore, we obtained 1793 immune-related genes from the ImmPort Portal database and used a Venn diagram to identify two of the ten hub genes in **Figure 2F** that were associated with the immune system: HSPA8 and HSPA1A (**Figure 2G**). We found that HSPA8 mRNA levels were up-regulated in CRC tissues, while HSPA1A mRNA levels were down-regulated (Supplementary Table 11). Finally, we used the Timer database to investigate the immune infiltration of these two hub genes (**Figure 2H**). The findings demonstrated a close relationship between the HSPA8 and HSPA1A genes and a variety of immune cell infiltration processes, including B cells, CD8+ T cells, macrophages, neutrophils, and dendritic cells. Furthermore, immunohistochemical results retrieved from The Human Protein Atlas (HPA) database indicated that HSPA8 protein expression was significantly higher in CRC tumor tissues than in normal colon tissues (Supplementary Figure 1). Patients with elevated levels of HSPA8 in CRC exhibit a higher hazard ratio (HR=1.672) than


A
Down regulated
Not changed
Up regulated
-log10(P-value)
log2 Fold change
B
Lysosome
Vesicle mediated transport
Endocytosis
Vesicle organization
Secretion by cell
Metabolism of lipids
Endocytosis
NABA matrisome associated
Extracellular matrix organization
Localization within membrane
Transmembrane recepor protein tyrosin kinase signal
Neutrophil degranulation
Organelle localization
Trans Golgi network vesicle budding
Cell-cell signaling by Wnt
Organic hydroxy compound metabolic process
External encapsulating structure organization
Peroxisome
Autophagy
Exocytosis
C
Down regulated
Not changed
Up regulated
-log10(P-value)
log2 Fold change
D
Late endosomal microautophagy
Glycoprotein metabolic process
Cellular lipid catabolic process
Negative regulation of cellular component organization
Regulation of potassium ion transmembrane transport
Steroid biosynthetic process
Spliceosome
Postive regulation of NF-kappaB transcription factor
Integrated breast cancer patnway
Regulation of synaptic plasticity
Glycerophospholipid metabolic process

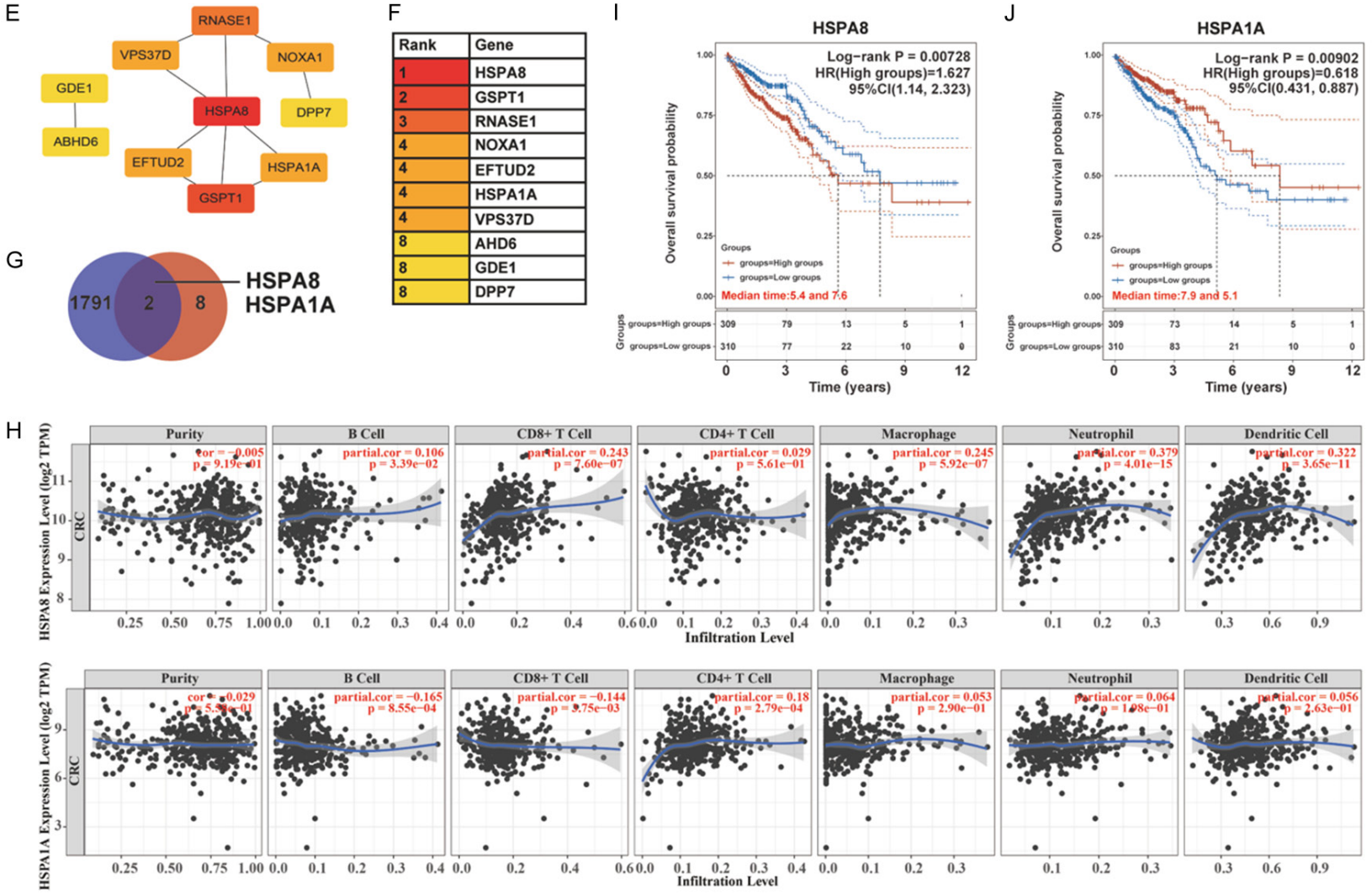


Figure 2. The overall picture of mRNA level varies. Volcanic map (A) and PPI (B) showed changes in the mRNA level of VGs. Volcano map (C) and PPI (D) show the survival-related vehicle molecules with altered mRNA levels. (E, F) The hub genes of survival-related VGs. (G) Venn diagram was performed for hub gene and immune-related molecules screened from (F). (H) To demonstrate the immune infiltration of HSPA8 and HSPA1A in CRC. (I, J) Kaplan-Meier survival analysis of the gene signature from TCGA dataset, comparison among different groups was made by log-rank test. HR (High exp) represents the hazard ratio of the low-expression sample relatives to the high-expression sample. HR > 1 indicates the gene is a risk factor, and HR < 1 indicates the gene is a protective factor. HR (95% CI), the median survival time (LT50) for different groups.

those with low HSPA8 expression. Conversely, patients with elevated levels of HSPA1A in CRC demonstrate a lower hazard ratio (HR=0.618) compared to those with low HSPA8 expression (**Figure 2I**, **2J**). These findings suggest that the vesicle gene HSPA8 is implicated as a risk factor in CRC, with patients exhibiting high HSPA8 expression associated with a poorer prognosis. Conversely, the vesicle gene HSPA1A is identified as a protective factor in CRC, with patients displaying high HSPA1A expression linked to a more favorable prognosis. These results underscore the potential hub gene role of VGs in CRC.

The overall changes of VGs at the protein level in CRC patients

Protein is a fundamental molecular unit that plays a crucial role in conducting biological functions and a wide range of living activities [12]. However, there is currently no comprehensive description of the alterations in vesicle transport-related components at the protein level in CRC. We conducted GO (**Figure 3A-C**) and KEGG (**Figure 3D**) analyses on the differentially expressed vesicle molecules in CRC at the protein level. Protein interaction analyses were conducted on the differential proteins using Metascape. The results showed significant relationships between these proteins, with enriched functions including neutrophil degranulation, lysosome, and extracellular matrix organization, as demonstrated in **Figure 3E**. The heatmap in **Figure 4A** clearly displays the 139 differential vesicle molecules. Subsequently, four critical subnetworks were identified through MCODE, as shown in **Figure 4B** and **4C**. The biological functions associated with MCODE1 involve extracellular matrix organization, collagen formation, collagen biosynthesis, and modifying enzymes. MCODE2 is primarily associated with neutrophil degranulation, leukocyte-mediated cytotoxicity, and cell killing. MCODE3 participates in G Alpha (Q) signaling events, GPCR downstream signaling, and signaling by GPCR. Finally, Cytohubba was used to screen out the top 10 key proteins (**Figure 4D**, **4E**), which were COL5A1, COL4A2, COL12A1, PLOD1, P4HA1, P4HA2, LEPRE1, COLGALT1, MMP2, and PCOLCE in sequence. The aforementioned results validate that vesicle genes manifest distinctive expression profiles at the protein level, engaging in diverse cellular functions and pathways such as enzyme modulation and other fundamental cellular processes.

Changes in post-translational modifications of vesicle proteins in CRC

Post-translational modification of proteins plays a crucial role in regulating intracellular tumor-related physiological and pathological processes, and modulating their biological activity [13-15]. Therefore, we identified genes altered at the phosphorylation, ubiquitination, and malonylation omics levels in CRC by analyzing the corresponding omics data. Specifically, we focused on genes that were altered at all three protein modification levels and performed KEGG pathway analysis on these genes. It's worth noting that the endocytosis pathway was significantly enriched (Supplementary Figure 3). In CRC, there were 80 proteins with up-regulated phosphorylation levels and 161 proteins with down-regulated phosphorylation levels. Additionally, 144 proteins had decreased levels of ubiquitination while 42 proteins had increased levels. The malonylation levels of up-regulated and down-regulated proteins in CRC were 433 and 471, respectively (**Figure 5A**). Changes in the modification level of CRC vesicle proteins may be closely related to tumor progression. Metascape then looked at the interactions of vesicle proteins that changed at these three modification levels in CRC (**Figure 5B**). Subsequently, four key subnetworks in the protein interaction network were further screened by the MCODE (**Figure 5C**, **5D**). It turns out that MCODE1's main functions are clathrin-mediated endocytosis, cargo recognition for clathrin-mediated endocytosis, and membrane trafficking. Copii-mediated vesicle transport, ER to Golgi anterograde transport, transport to the Golgi, and subsequent modification are MCODE2's main functions. The main function of MCODE3 is neutrophil degranulation. MCODE4 performs crucial functions in eukaryotic translation termination, nonsense mediated decay (NMD) independent of the exon junction complex (EJC), and nonsense-mediated decay (NMD). Overall, vesicle genes have undergone notable alterations at multiple levels, encompassing transcriptional regulation, protein expression, and post-translational modifications.

Comprehensive analysis disclosed potential TFs for the VGs in CRC patients

It is a widely accepted fact that transcription factors play a crucial role in regulating gene expression in eukaryotes. The results of the aforementioned research indicate that there

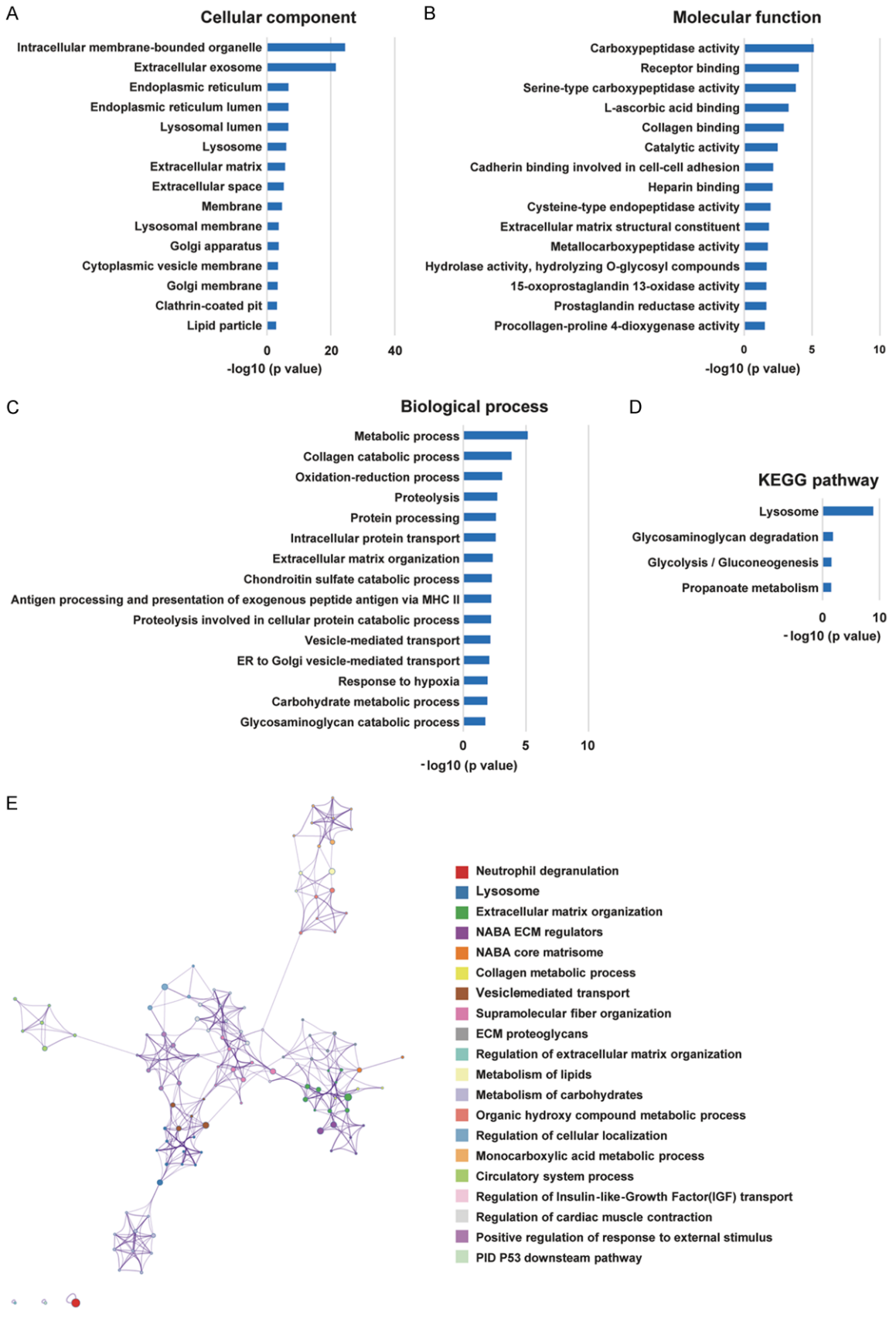
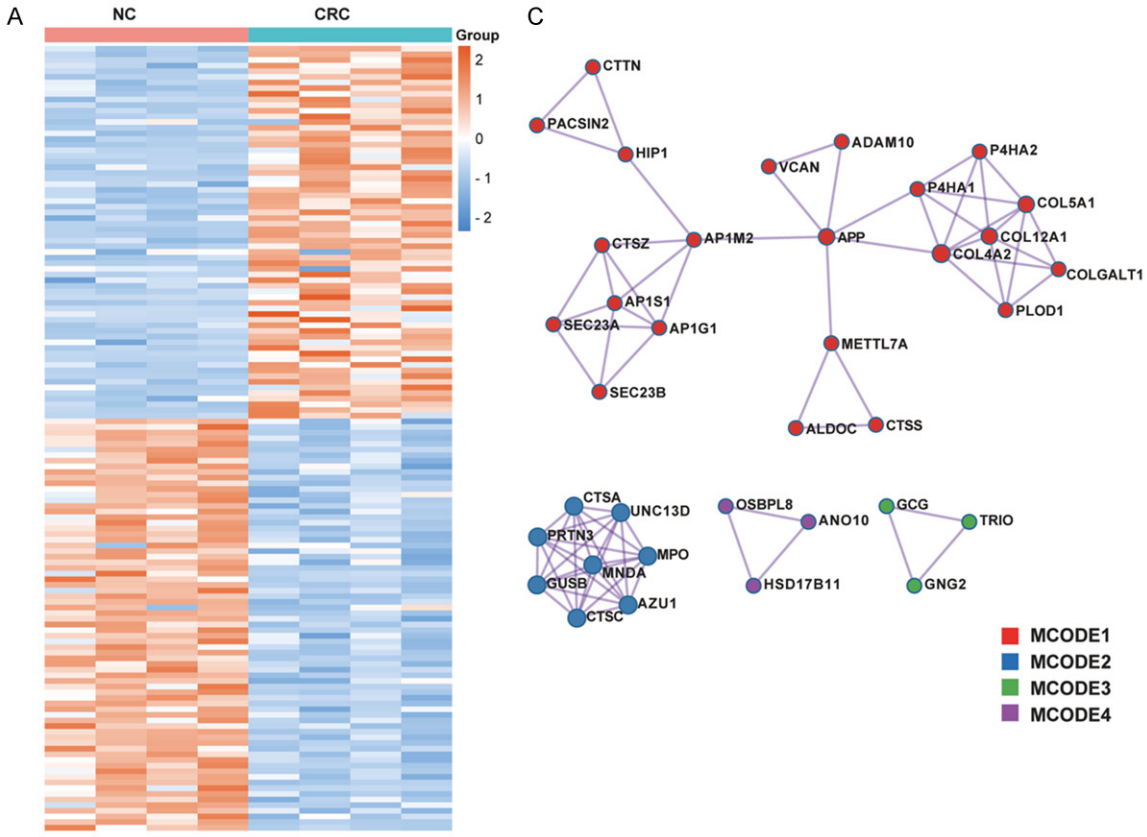
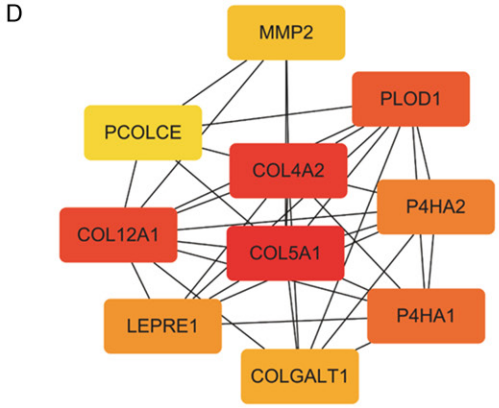


Figure 3. The general picture of changes in protein level. A-D. GO and KEGG analysis VGs with altered protein levels. E. PPI of altered VGs in protein level.



B

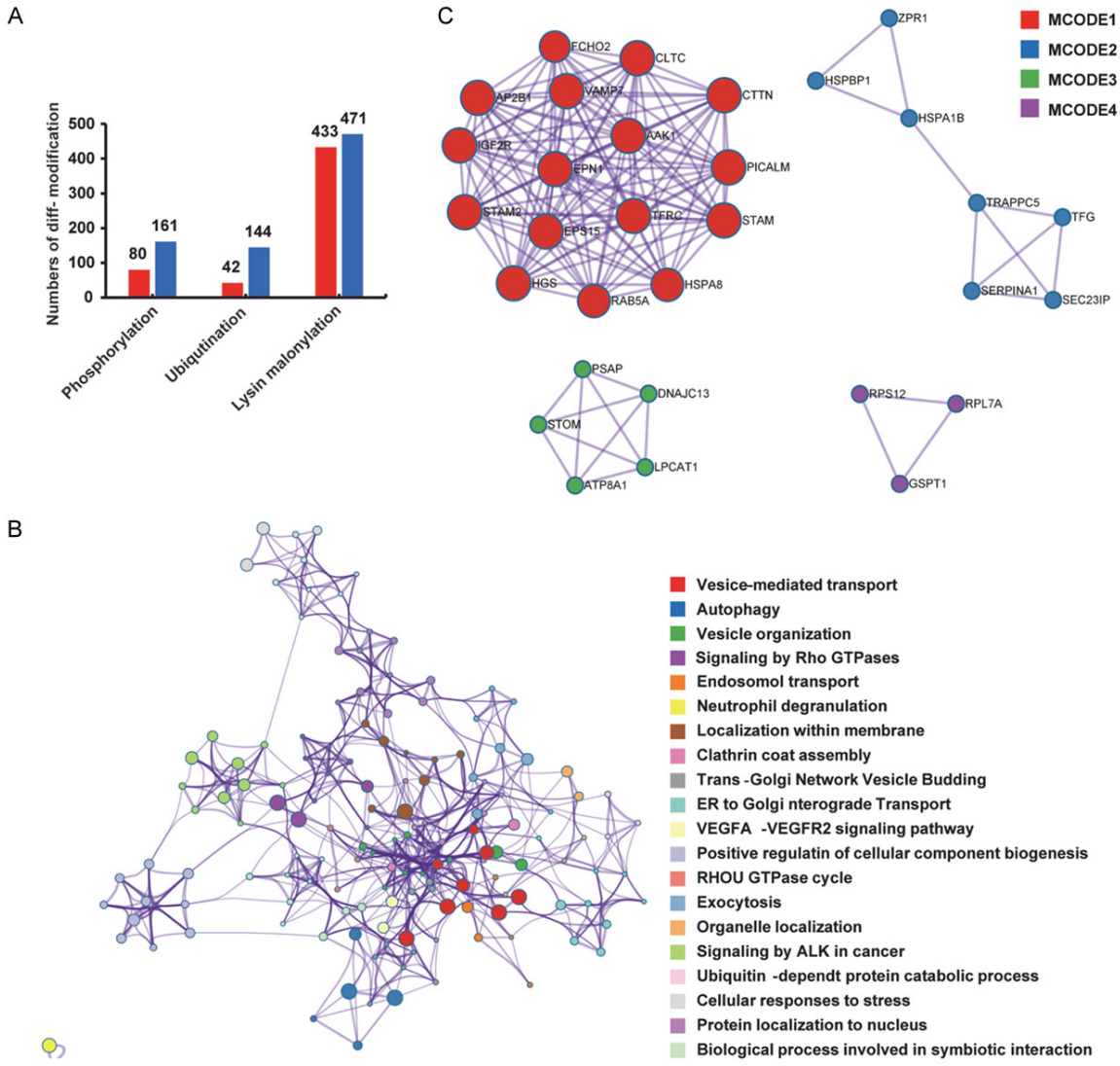
MCODE	GO	Description	Log10(P)
MCODE_1	R-HSA-1474244	Extracellular matrix organization	-16.3
MCODE_1	R-HSA-1474290	Collagen formation	-14.9
MCODE_1	R-HSA-1650814	Collagen biosynthesis and modifying enzymes	-13.5
MCODE_2	R-HSA-6798695	Neutrophil degranulation	-14.4
MCODE_2	GO:0001909	leukocyte mediated cytotoxicity	-6.6
MCODE_2	GO:0001906	Cell killing	-5.8
MCODE_3	R-HSA-416476	G alpha (q) signalling events	-6.4
MCODE_3	R-HSA-388396	GPCR downstream signalling	-5
MCODE_3	R-HSA-372790	Signaling by GPCR	-4.9



E

Rank	Protein
1	COL5A1
2	COL4A2
3	COL12A1
4	PLOD1
5	P4HA1
6	P4HA2
7	LEPRE1
8	COLGALT1
9	MMP2
10	PCOLCE

Figure 4. The overall picture of protein level variations and the hub genes. A. Heatmap of VGs. B, C. MCODE analysis of VGs with protein level changes. D, E. The analysis to search hub genes.



D

MCODE	GO	Description	Log10(P)
MCODE_1	R-HSA-8856828	Clathrin-mediated endocytosis	-37.4
MCODE_1	R-HSA-8856825	Cargo recognition for clathrin-mediated endocytosis	-29.5
MCODE_1	R-HSA-199991	Membrane Trafficking	-26.9
MCODE_2	R-HSA-204005	COPII -mediated vesicle transport	-9.1
MCODE_2	R-HSA-199977	ER to Golgi Anterograde Transport	-7.7
MCODE_2	R-HSA-948021	Transport to the Golgi and subsequent modification	-7.3
MCODE_3	R-HSA-6798695	Neutrophil degranulation	-9
MCODE_4	R-HSA-72764	Eukaryotic Translation Termination	-7.6
MCODE_4	R-HSA-975956	Nonsense Mediated Decay (NMD) independent of the Exon Junction Complex (EJC)	-7.5
MCODE_4	R-HSA-927802	Nonsense Mediated Decay (NMD)	-7.3

Figure 5. The whole representation of VGs changes at three modification levels (ubiquitin, phosphorylation, malonylation). A. Histogram display of gene changes at each modification level in CRC. B. PPI of VGs with three modification changes. C, D. Four main MCODE of PPI.

were significant changes in the mRNA expression levels of several VGs in CRC patients as compared to normal paracancerous tissues. This led us to investigate if there were any

abnormal molecular alterations in the upstream transcription factors that were responsible for these changes. Firstly, we screened out the VGs that showed alterations in both mRNA and protein levels in CRC, yielding a total of 85 genes (Supplementary Table 12). Subsequently, the hTFtarget database was searched for upstream transcription factors regulating these 85 genes, resulting in a total of 144 transcription factors (Supplementary Table 13). Eukaryotic DNA is further folded and compacted to form chromosomes with the help of additional structural proteins after joining with histones. Hence, chromosome opening is necessary for binding transcription factors with DNA [16]. We examined TFs with highly accessible DNA-binding domains in CRC tumors and paracancerous tissues using chromatin accessibility. Our findings revealed that, unlike CRC tissues, the DNA binding domain of transcription factors SP1, SP2, KLF5, ZBTB7B, GLIS1, ESRRA, NKX2-5, and USF2 were exclusively open in paracancerous tissues (**Figure 6A**). And they respectively regulated 45.9%, 2.4%, 28.2%, 3.5%, 5.9%, 1.2%, 1.2% of all down-regulated differentially expressed vesicle genes (**Figure 6C**). The DNA-binding sites of JUN, SOX4, STAT2, JUND, ASCL2, REST, and ELF1 were, however, exclusively open in CRC tissues (**Figure 6B**), and they can control 9.4%, 2.4%, 1.2%, 21.2%, 3.5%, 4.7%, and 1.2% of CRC up-regulated VGs, respectively (**Figure 6D**). We then checked for the levels of protein and phosphorylation of 144 potential transcription factors. The results showed that transcription factors CDX2, YY1, STAT2, and HSF1 were up-regulated in CRC (**Figure 6E** and Supplementary Figure 4). CBX3 phosphorylation was up-regulated at Ser 196 (**Figure 6F**). Additionally, the transcription factors YY1, CDX2, STAT2, HSF1, and phosphorylated CBX3 control 52.9%, 25.9%, 3.5%, 1.2%, and 12.9%, respectively, of the differentially expressed vesicle genes in CRC (**Figure 6G**, **6H**). In our study, we systematically investigated the changes in vesicle genes at the mRNA and protein levels in CRC, pinpointing key hub genes. Subsequent analysis unveiled potential critical transcription factors involved in this process.

Discussion

For a comprehensive search for precise diagnostic and therapeutic targets, it is crucial to fully uncover the molecular changes involved in the onset and progression of colorectal cancer (CRC). Our prior research has indicated that proteins exhibiting reduced levels of phosphorylation in CRC tumors are highly concentrated in endocytic vesicles. Vesicles are small organelles enclosed by a lipid bilayer, which include transport vesicles, secretory vesicles, endosomes, and lysosomes [17]. These vesicles play various biological roles in cells, such as intracellular transport of proteins and lipids, maintaining stability of intracellular material metabolism through uptake and absorption, and communication between cells through signal transmission [18]. Despite the increasing interest in vesicles in CRC, there is still a lack of multi-omics studies examining their molecular changes. To address this gap, we conducted a comprehensive analysis of proteomics (n=8), DNA methylation (n=6), chromatin structural variation (n=6), chromatin openness (n=6), protein ubiquitination (n=8), protein phosphorylation (n=8), and protein malonylation (n=8) in tumor and para-cancer tissues of patients with clinical colorectal cancer. Additionally, we incorporated RNA-seq data from public databases to further enhance our findings. The panoramic view of CRC reveals changes in the vesicle at each omics level, identifying critical molecules and transcription factors that influence the progression of the disease.

Our study utilized multi-omics analysis to identify alterations in various VGs in CRC at different levels, such as DNA, mRNA, protein, and protein post-translational modification. We also discovered two immune genes, HSPA8 and HSPA1A, that are associated with CRC survival. Notably, we found that HSPA8 expression was significantly higher in CRC tumor tissues compared to paracancerous tissues, and HSPA8 ubiquitination at several locations was significantly reduced. Ubiquitination plays a crucial role in protein degradation [19]. A decrease in the ubiquitination level of HSPA8 protein indicates a slower rate of protein disintegration. HSPA8, a stress protein, is involved in regulating hepatoma carcinoma cell growth and apoptosis. Its expression increases with disease progression and has been used as an early biomarker for hepatic carcinoma [20]. Our study found that there was an increase in mRNA and protein expression levels of HSPA8 in CRC when compared to paracancerous tissues.

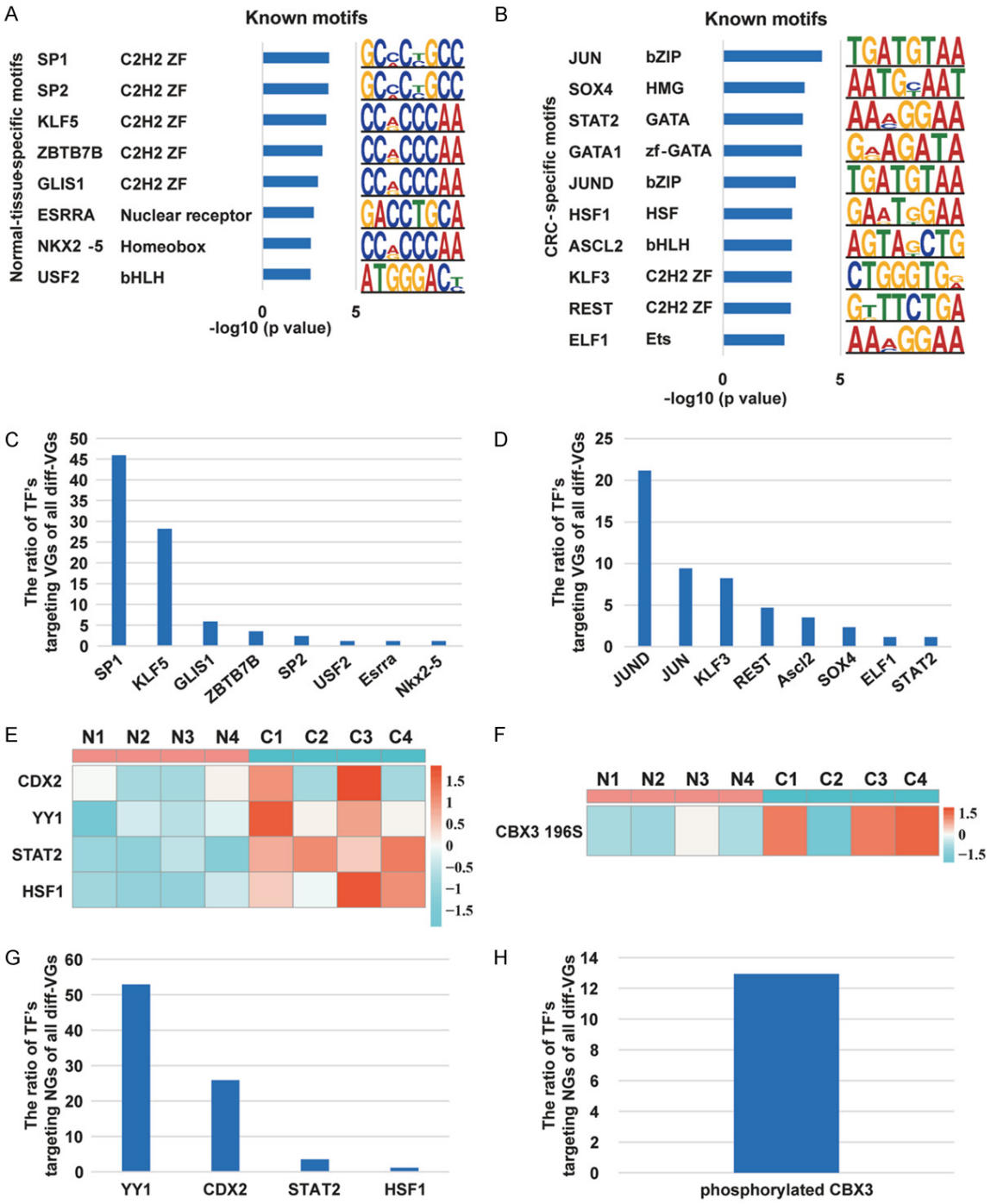


Figure 6. Proteogenomic alterations in TFs of the differentially-expressed VGs of CRC patients. (A) TFs unique in normal tissues. (B) Tissue-specific TFs of CRC. (C) The fraction of down-regulated VGs targeted by TFs in (A). (D) The propotion of up-regulated VGs targeted by TFs in (B). The changes of TF in protein level (E) and phosphorylation level (F). Altered TFs in (G) protein and (H) phosphorylation levels.

These results were consistent with a previous study [21], which suggests that HSPA8 could be a potential prognostic marker gene for CRC. Further research could build upon these findings and explore the potential clinical implications of HSPA8 in CRC management.

This study identified the top 10 key genes that were altered at the protein level in CRC using the Cytohubba tool. These genes include COL5A1, COL4A2, COL12A1, PLOD1, P4HA1, P4HA2, LEPRE1, COLGALT1, MMP2, and PCOLCE (**Figure 4D**, **4E** and Supplementary Figure 2).

Among these genes, the top three proteins that were significantly up-regulated in CRC tumor tissues were COL5A1, COL4A2, and COL12A1. Studies have shown that overexpression of COL5A1 in various types of cancer, including ovarian cancer [22], glioma [23], and lung adenocarcinoma [24], is linked to tumor growth, paclitaxel resistance, and tumor-infiltrating immune cells. In addition, research has found that the effectiveness of 5-fluorouracil (5-FU), the standard first-line treatment for CRC patients, is associated with COL4A2 sensitivity [25]. Bioinformatics research has revealed that the abnormal expression of COL5A1 and COL12A1 is linked to the development and prognosis of colorectal cancer (CRC) [26]. This discovery indicates that extracellular matrix-associated collagen could potentially serve as a valuable prognostic and therapeutic biomarker for CRC in the future.

Eukaryotic gene transcription is regulated through multiple levels, including chromatin openness, protein levels, and phosphorylation levels of relevant transcription factors. Our research focused on identifying upstream transcription factors associated with tumorigenesis by analyzing differential vesicle genes. Our findings revealed that SP1, KLF5, JUND, YY1, CDX2, and CBX3 were associated with tumorigenesis. Among these, SP1 [27], YY1 [28], and CDX2 [29] have been extensively studied as cancer-promoting proteins in colorectal cancer. SP1 is known to regulate the proliferation, apoptosis, and invasion of CRC tumor cells [30-32]. Similarly, YY1 has been found to promote the proliferation of colorectal cancer cells through the Mir-526B-3p/E2F1 axis, and higher expression of YY1 is associated with a poor prognosis for CRC patients [28]. CDX2 has been shown to modulate 5-fluorouracil metabolism, which is believed to have an impact on colorectal cancer (CRC) prognosis and chemotherapy effectiveness according to studies [33, 34]. However, the roles of KLF5, JUND, and CBX3 in CRC have not been fully elucidated yet.

In conclusion, the study highlights significant changes in the molecular makeup of vesicle-associated genes as CRC tumors progress. It is the first comprehensive analysis of the follicular molecules at multiple omics levels in CRC and identifies potential prognostic molecules and target genes with clinical value. This could aid in improving the survival and prognosis of patients with colorectal cancer.

Acknowledgements

This study was approved by the National Natural Science Foundation of China (No. 82003172), Shenzhen Fund for Guangdong Provincial High level Clinical Key Specialties (No. SZGSP001), the Postdoctoral Science Foundation of China (No. 2020M673065), Natural Science Foundation of Guangdong Province (No. 2019A1515-111138), and the science and technology plan of Shenzhen (No. JCYJ20180306140810282).

The study participants were all volunteers who signed an informed consent form.

Disclosure of conflict of interest

The authors declare that the research was conducted without any potential competing interest.

Abbreviations

CRC, colorectal cancer; PMN, pre-metastatic niches; TCGA, The Cancer Genome Atlas; VGs, vesicle genes; HPA, Human Protein Atlas; NMD, nonsense mediated decay; EJC, exon junction complex; WGBS, Whole genome bisulfite sequencing; ATAC-Seq, Assay to detect transposase-accessible chromatin based on high throughput sequencing.

Address correspondence to: Dong-E Tang and Yong Dai, Clinical Medical Research Center, The First Affiliated Hospital (Shenzhen People's Hospital), Southern University of Science and Technology, Shenzhen 518055, Guangdong, China. E-mail: donge66@126.com (DET); Tel: +86-0755-2294-2780; Fax: +86-0755-22942780; E-mail: daiyong22@aliyun.com; daiyong22@aust.edu.cn (YD); Min Tang, Key Laboratory of Diagnostic Medicine Designated by The Chinese Ministry of Education, Chongqing Medical University, Chongqing 400016, China. E-mail: catom@126.com

References

[1] Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A and Bray F. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2021; 71: 209-249.

[2] Qureshi-Baig K, Kuhn D, Viry E, Pozdeev VI, Schmitz M, Rodriguez F, Ullmann P, Koncina E, Nurmik M, Frasquilho S, Nazarov PV, Zuegel N, Boulmont M, Karapetyan Y, Antunes L, Val D, Mittelbronn M, Janji B, Haan S and Letellier E. Hypoxia-induced autophagy drives colorectal cancer initiation and progression by activating the PRKC/PKC-EZR (ezrin) pathway. Autophagy 2020; 16: 1436-1452.
[3] Mohammadi D. The twists and turns of colorectal cancer screening. Lancet Gastroenterol Hepatol 2017; 2: 10-11.
[4] Buccafusca G, Proserpio I, Tralongo AC, Rametta Giuliano S and Tralongo P. Early colorectal cancer: diagnosis, treatment and survivorship care. Crit Rev Oncol Hematol 2019; 136: 20-30.
[5] Ji Q, Zhou L, Sui H, Yang L, Wu X, Song Q, Jia R, Li R, Sun J, Wang Z, Liu N, Feng Y, Sun X, Cai G, Feng Y, Cai J, Cao Y, Cai G, Wang Y and Li Q. Primary tumors release ITGBL1-rich extracellular vesicles to promote distal metastatic tumor growth through fibroblast-niche formation. Nat Commun 2020; 11: 1211.
[6] Zhang Q, Liu RX, Chan KW, Hu J, Zhang J, Wei L, Tan H, Yang X and Liu H. Exosomal transfer of p-STAT3 promotes acquired 5-FU resistance in colorectal cancer cells. J Exp Clin Cancer Res 2019; 38: 320.
[7] Gao T, Liu X, He B, Nie Z, Zhu C, Zhang P and Wang S. Exosomal lncRNA 91H is associated with poor development in colorectal cancer by modifying HNRNPK expression. Cancer Cell Int 2018; 18: 11.
[8] Peinado H, Zhang H, Matei IR, Costa-Silva B, Hoshino A, Rodrigues G, Psaila B, Kaplan RN, Bromberg JF, Kang Y, Bissell MJ, Cox TR, Giaccia AJ, Erler JT, Hiratsuka S, Ghajar CM and Lyden D. Pre-metastatic niches: organ-specific homes for metastases. Nat Rev Cancer 2017; 17: 302-317.
[9] Auger C, Christou N, Brunel A, Perraud A and Verdier M. Autophagy and extracellular vesicles in colorectal cancer: interactions and common actors? Cancers (Basel) 2021; 13: 1039.
[10] Zhang W, Wu M, Gao X, Ma C, Xu H, Lin L, He J, Cai W, Zhong Y, Tang D, Tang M and Dai Y. Multi-platform-based analysis characterizes molecular alterations of the nucleus in human colorectal cancer. Front Cell Dev Biol 2022; 10: 796703.
[11] Picard E, Verschoor CP, Ma GW and Pawelec G. Relationships between immune landscapes, genetic subtypes and responses to immunotherapy in colorectal cancer. Front Immunol 2020; 11: 369.
[12] Berezovsky IN, Guarnera E, Zheng Z, Eisenhaber B and Eisenhaber F. Protein function machinery: from basic structural units to modulation of activity. Curr Opin Struct Biol 2017; 42: 67-74.
[13] Noujaim J, Payne LS, Judson I, Jones RL and Huang PH. Phosphoproteomics in translational research: a sarcoma perspective. Ann Oncol 2016; 27: 787-794.
[14] Sun T, Liu Z and Yang Q. The role of ubiquitination and deubiquitination in cancer metabolism. Mol Cancer 2020; 19: 146.
[15] Mokarram P, Albokashy M, Zarghooni M, Moosavi MA, Sepehri Z, Chen QM, Hudecki A, Sargazi A, Alizadeh J, Moghadam AR, Hashemi M, Movassagh H, Klonisch T, Owji AA, Los MJ and Ghavami S. New frontiers in the treatment of colorectal cancer: autophagy and the unfolded protein response as promising targets. Autophagy 2017; 13: 781-819.
[16] Hirano T. Chromosome dynamics during mitosis. Cold Spring Harb Perspect Biol 2015; 7: a015792.
[17] Cocucci E and Meldolesi J. Ectosomes and exosomes: shedding the confusion between extracellular vesicles. Trends Cell Biol 2015; 25: 364-372.
[18] Meldolesi J. Exosomes and ectosomes in intercellular communication. Curr Biol 2018; 28: R435-R444.
[19] van Wijk SJ, Fulda S, Dikic I and Heilemann M. Visualizing ubiquitination in mammalian cells. EMBO Rep 2019; 20: e46520.
[20] Xiang X, You XM and Li LQ. Expression of HSP90AA1/HSPA8 in hepatocellular carcinoma patients with depression. Onco Targets Ther 2018; 11: 3013-3023.
[21] Lim Y, Gang DY, Lee WY, Yun SH, Cho YB, Huh JW, Park YA and Kim HC. Proteomic identification of arginine-methylated proteins in colon cancer cells and comparison of messenger RNA expression between colorectal cancer and adjacent normal tissues. Ann Coloproctol 2022; 38: 60-68.
[22] Zhang J, Zhang J, Wang F, Xu X, Li X, Guan W, Men T and Xu G. Overexpressed COL5A1 is correlated with tumor progression, paclitaxel resistance, and tumor-infiltrating immune cells in ovarian cancer. J Cell Physiol 2021; 236: 6907-6919.
[23] Gu S, Peng Z, Wu Y, Wang Y, Lei D, Jiang X, Zhao H and Fu P. COL5A1 serves as a biomarker of tumor progression and poor prognosis and may be a potential therapeutic target in gliomas. Front Oncol 2021; 11: 752694.
[24] Liu W, Wei H, Gao Z, Chen G, Liu Y, Gao X, Bai G, He S, Liu T, Xu W, Yang X, Jiao J and Xiao J. COL5A1 may contribute the metastasis of lung adenocarcinoma. Gene 2018; 665: 57-66.
[25] Wang Y, Wei Q, Chen Y, Long S, Yao Y and Fu K. Identification of hub genes associated with

sensitivity of 5-fluorouracil based chemotherapy for colorectal cancer by integrated bioinformatics analysis. Front Oncol 2021; 11: 604315.
[26] Wu Y and Xu Y. Integrated bioinformatics analysis of expression and gene regulation network of COL12A1 in colorectal cancer. Cancer Med 2020; 9: 4743-4755.
[27] Bajpai R and Nagaraju GP. Specificity protein 1: its role in colorectal cancer progression and metastasis. Crit Rev Oncol Hematol 2017; 113: 1-7.
[28] Fang Z, Yang H, Chen D, Shi X, Wang Q, Gong C, Xu X, Liu H, Lin M, Lin J, Xu C and Shao J. YY1 promotes colorectal cancer proliferation through the miR-526b-3p/E2F1 axis. Am J Cancer Res 2019; 9: 2679-2692.
[29] Zarour LR, Anand S, Billingsley KG, Bisson WH, Cercek A, Clarke MF, Coussens LM, Gast CE, Geltzeiler CB, Hansen L, Kelley KA, Lopez CD, Rana SR, Ruhl R, Tsikitis VL, Vaccaro GM, Wong MH and Mayo SC. Colorectal cancer liver metastasis: evolving paradigms and future directions. Cell Mol Gastroenterol Hepatol 2017; 3: 163-173.
[30] Zhao Y, Zhang W, Guo Z, Ma F, Wu Y, Bai Y, Gong W, Chen Y, Cheng T, Zhi F, Zhang Y, Wang J and Jiang B. Inhibition of the transcription factor Sp1 suppresses colon cancer stem cell growth and induces apoptosis in vitro and in nude mouse xenografts. Oncol Rep 2013; 30: 1782-1792.
[31] Yonesaka K, Satoh T, Ueda S, Yoshida T, Takeda M, Shimizu T, Okamoto I, Nishio K, Tamura T and Nakagawa K. Circulating hepatocyte growth factor is correlated with resistance to cetuximab in metastatic colorectal cancer. Anticancer Res 2015; 35: 1683-1689.
[32] Pathi S, Li X and Safe S. Tolfenamic acid inhibits colon cancer cell and tumor growth and induces degradation of specificity protein (Sp) transcription factors. Mol Carcinog 2014; 53 Suppl 1: E53-61.
[33] Delhorme JB, Bersuder E, Terciolo C, Vlami O, Chenard MP, Martin E, Rohr S, Brigand C, Duluc I, Freund JN and Gross I. CDX2 controls genes involved in the metabolism of 5-fluorouracil and is associated with reduced efficacy of chemotherapy in colorectal cancer. Biomed Pharmacother 2022; 147: 112630.
[34] Ribeirinho-Soares S, Padua D, Amaral AL, Valentini E, Azevedo D, Marques C, Barros R, Macedo F, Mesquita P and Almeida R. Prognostic significance of MUC2, CDX2 and SOX2 in stage II colorectal cancer patients. BMC Cancer 2021; 21: 359.

Supplementary Table 1. The clinical information of the CRC patients of this study

		I-II stage		III-IV stage	
Age	≤65	0	0.00%	1	12.50%
	>65	4	50.00%	3	37.50%
Type	Adenocarcinoma	4	50.00%	4	50.00%
Gender	Male	1	12.50%	3	37.50%
	Female	3	37.50%	1	12.50%
Tumor stage	T1	1	12.50%	0	0.00%
	T2	0	0.00%	0	0.00%
	T3	2	0.00%	1	12.50%
	T4	1	12.50%	3	37.50%
Node stage	N0	4	50.00%	0	0.00%
	N1	0	0.00%	4	50.00%
	N2	0	0.00%	0	0.00%
Metastasis stage	M0	4	50.00%	1	12.50%
	M1	0	0.00%	3	37.50%
Location	Ascending colon	2	25.00%	1	12.50%
	Sigmoid colon	1	12.50%	3	37.50%
	Appendix colon	1	12.50%	0	0.00%
Therapy	Yes	0	0.00%	0	0.00%
	No	4	50.00%	4	50.00%

Supplementary Table 3. The intersections of OS-related genes and VGs differentially expressed in mRNA level in CRC

Gene	logFC	*p*Value	log2FC
ARV1	-0.362240898	4.00832E-07	-1.203338215
ERI1	0.616554469	9.88029E-07	2.048149613
KCNIP3	-1.464595353	5.03247E-14	-4.865280452
CA11	-0.684826021	8.25031E-07	-2.274942799
CHEK1	1.321749732	9.70847E-25	4.390757569
ACOX1	-1.324776833	5.57976E-19	-4.400813382
GDI1	0.407721004	4.94518E-09	1.35441986
EFTUD2	0.803409332	2.05225E-26	2.668868033
FUT4	0.622769035	6.98426E-09	2.068793954
HSPA8	0.682281907	4.2762E-16	2.266491436
TM7SF2	1.108795321	7.25163E-18	3.683338328
UBTD1	1.267169193	6.76645E-18	4.209444942
NOXA1	0.706845086	5.24755E-05	2.348088549
RFX1	0.356736609	0.000258816	1.185053365
DPP7	1.124443242	9.17692E-21	3.735319598
NR1H2	-0.312076595	5.04562E-11	-1.036696009
SENP8	-0.508070541	2.37814E-12	-1.687773804
VPS37D	0.957610789	0.001082331	3.181114184
HSPA1A	-0.464077662	1.08876E-08	-1.541632625
PRNP	-0.941085626	4.03846E-05	-3.12621878
ABCD3	-0.916327276	2.77157E-24	-3.043973323
RFT1	0.774623555	6.33459E-25	2.573243751
SCG2	-2.182076633	1.18325E-26	-7.248701674
ZNF132	-1.002783939	5.57193E-21	-3.331176141
CLINT1	-0.318124673	4.17444E-08	-1.056787289

AMH	4.193159177	1.49938E-17	13.92937328
GABRD	3.024168416	8.92901E-30	10.04607003
GSPT1	0.440501273	3.51369E-12	1.463313555
P4HA1	0.988602914	5.84096E-14	3.284067795
ADAMTS13	0.665992783	0.005040921	2.212380138
RNASE1	-1.103020797	2.02429E-16	-3.664155773
GDE1	-0.592755682	2.8199E-21	-1.969091754
IDUA	0.936014341	1.13592E-08	3.109372336
VGF	3.000721864	1.00232E-11	9.968182266
ARHGEF25	-2.30646821	5.19135E-18	-7.661921546
KCNIP2	-0.399677156	0.002432484	-1.327698772
FJX1	3.012394739	6.83315E-31	10.00695872
CHPF	1.86895484	4.98547E-28	6.208533591
ABHD6	-0.784933087	3.79698E-23	-2.607491274
TRIM8	0.30833668	4.74142E-06	1.024272281

Supplementary Table 4. The intersections of OS-related genes and VGs differentially expressed in protein level in CRC

Protein accession	P13674	P35475
Protein description	Prolyl 4-hydroxylase subunit alpha-1 OS=Homo sapiens OX=9606 GN=P4HA1 PE=1 SV=2	Alpha-L-iduronidase OS=Homo sapiens OX=9606 GN=IDUA PE=1 SV=2
Ratio	4.038	0.486
Regulated Type	Up	Down
P value	0.0000218	0.0071351
Gene name	*P4HA1*	*IDUA*
Mol. weight [kDa]	61.049	72.669
Sequence coverage [%]	32.6	13.3
MS/MS Counts	55	19
Peptides	15	6
Unique peptides	15	6
MS/MS Count I_II_N	4	4
MS/MS Count III_IV_N	3	3
MS/MS Count M_N	4	3
MS/MS Count V_N	3	5
MS/MS Count I_II_C	11	0
MS/MS Count III_IV_C	8	0
MS/MS Count M_C	11	3
MS/MS Count V_C	11	1
Peptides I_II_N	11	5
Peptides III_IV_N	9	3
Peptides M_N	9	5
Peptides V_N	6	6
Peptides I_II_C	13	4
Peptides III_IV_C	9	2
Peptides M_C	14	5
Peptides V_C	12	3
Unique peptides I_II_N	11	5
Unique peptides III_IV_N	9	3
Unique peptides M_N	9	5

Unique peptides V_N	6	6
Unique peptides I_II_C	13	4
Unique peptides III_IV_C	9	2
Unique peptides M_C	14	5
Unique peptides V_C	12	3
LFQ intensity I_II_N	260470	65001
LFQ intensity III_IV_N	217970	65870
LFQ intensity M_N	231590	58699
LFQ intensity V_N	195860	81176
LFQ intensity I_II_C	905530	19941
LFQ intensity III_IV_C	991700	34132
LFQ intensity M_C	676650	47177
LFQ intensity V_C	1084700	30234
I_II_N	0.457	1.293
III_IV_N	0.382	1.31
M_N	0.406	1.167
V_N	0.343	1.615
I_II_C	1.587	0.397
III_IV_C	1.738	0.679
M_C	1.186	0.938
V_C	1.901	0.601

Supplementary Table 5. The intersections of OS-related genes and VGs differentially expressed in DNA methylation in CRC

Genes	Direction
ACOX1	Hypo
FUT4	Hypo
DPP7	Hyper
SENP8	Hypo
PRNP	Hyper
AMH	Hyper
GABRD	Hyper
ADAMTS13	Hypo
IDUA	Hyper
FJX1	Hyper
TRIM8	Hyper

Supplementary Table 6. The intersections of OS-related genes and VGs changed in chromatin openness in CRC

	Genes
1	*CHEK1*
2	*ACOX1*
3	*RFX1*
4	*NR1H2*
5	*SENP8*
6	*PRNP*
7	*CLINT1*
8	*ABHD6*
9	*TRIM8*

Supplementary Table 7. The intersections of OS-related genes and VGs changed in protein ubiquitination in CRC

Subcellular localization	nucleus	cytoplasm	cytoplasm	cytoplasm	cytoplasm	cytoplasm	cytoplasm	cytoplasm	cytoplasm	cytoplasm
Protein accession	Q14677	P11142	P11142	P11142	P11142	P11142	P11142	P11142	P11142	P11142
Position	192	451	583	257	601	500	507	348	512	524
Amino acid	K	K	K	K	K	K	K	K	K	K
C_C/C_N Ratio	0.331	0.559	0.577	0.509	0.622	0.568	0.532	0.611	0.641	0.569
Regulated Type	Down	Down	Down	Down	Down	Down	Down	Down	Down	Down
Protein description	Clathrin interactor 1 OS=Homo sapiens OX=9606 GN=CLINT1 PE=1 SV=1	Heat shock cognate 71 kDa protein OS=Homo sapiens OX=9606 GN=HSPA8 PE=1 SV=1	Heat shock cognate 71 kDa protein OS=Homo sapiens OX=9606 GN=HSPA8 PE=1 SV=1	Heat shock cognate 71 kDa protein OS=Homo sapiens OX=9606 GN=HSPA8 PE=1 SV=1	Heat shock cognate 71 kDa protein OS=Homo sapiens OX=9606 GN=HSPA8 PE=1 SV=1	Heat shock cognate 71 kDa protein OS=Homo sapiens OX=9606 GN=HSPA8 PE=1 SV=1	Heat shock cognate 71 kDa protein OS=Homo sapiens OX=9606 GN=HSPA8 PE=1 SV=1	Heat shock cognate 71 kDa protein OS=Homo sapiens OX=9606 GN=HSPA8 PE=1 SV=1	Heat shock cognate 71 kDa protein OS=Homo sapiens OX=9606 GN=HSPA8 PE=1 SV=1	Heat shock cognate 71 kDa protein OS=Homo sapiens OX=9606 GN=HSPA8 PE=1 SV=1
Gene name	*CLINT1*	*HSPA8*	*HSPA8*	*HSPA8*	*HSPA8*	*HSPA8*	*HSPA8*	*HSPA8*	*HSPA8*	*HSPA8*
Localization probability	0.999968	1	0.987894	1	1	1	1	1	1	0.997637
PEP	0.021337	1.73084E-08	7.33322E-09	0.00381197	6.58467E-17	2.72656E-07	0.00067166	3.09046E-13	0.000882346	8.90469E-10
Score	48.785	148.9	78.888	143.56	135.5	120.65	121.67	110.88	105.36	166.61
Charge	2	2	3	2	2	2	3	2	3	2
Modified sequence	NK (1) SAFPFSDK	AMTK (1) DNNLLGK	CNEIINWLDK (0.988) NQTAEK (0.012)	DISENK (1) R	ELEK (1) VCNPIITK	ENK (1) ITITNDK	ITITNDK (1) GR	IQK (1) LLQDFF-NGK	LSK (1) EDIER	MVQEAEK (0.998) YK (0.002)
Mass error [ppm]	0.40583	-0.10491	0.45216	1.3439	1.4139	-1.012	2.1424	1.8059	2.5731	3.3444
C_C	0.501	0.75	0.763	0.711	0.794	0.757	0.729	0.787	0.807	0.758
C_N	1.514	1.342	1.322	1.398	1.277	1.332	1.37	1.289	1.259	1.331

Supplementary Table 8. The intersections of OS-related genes and VGs with chromatin structure variation in CRC

	Genes
1	CLINT1
2	FJX1

Supplementary Table 9. The intersections of OS-related genes and VGs differentially expressed in protein phosphorylation level in CRC

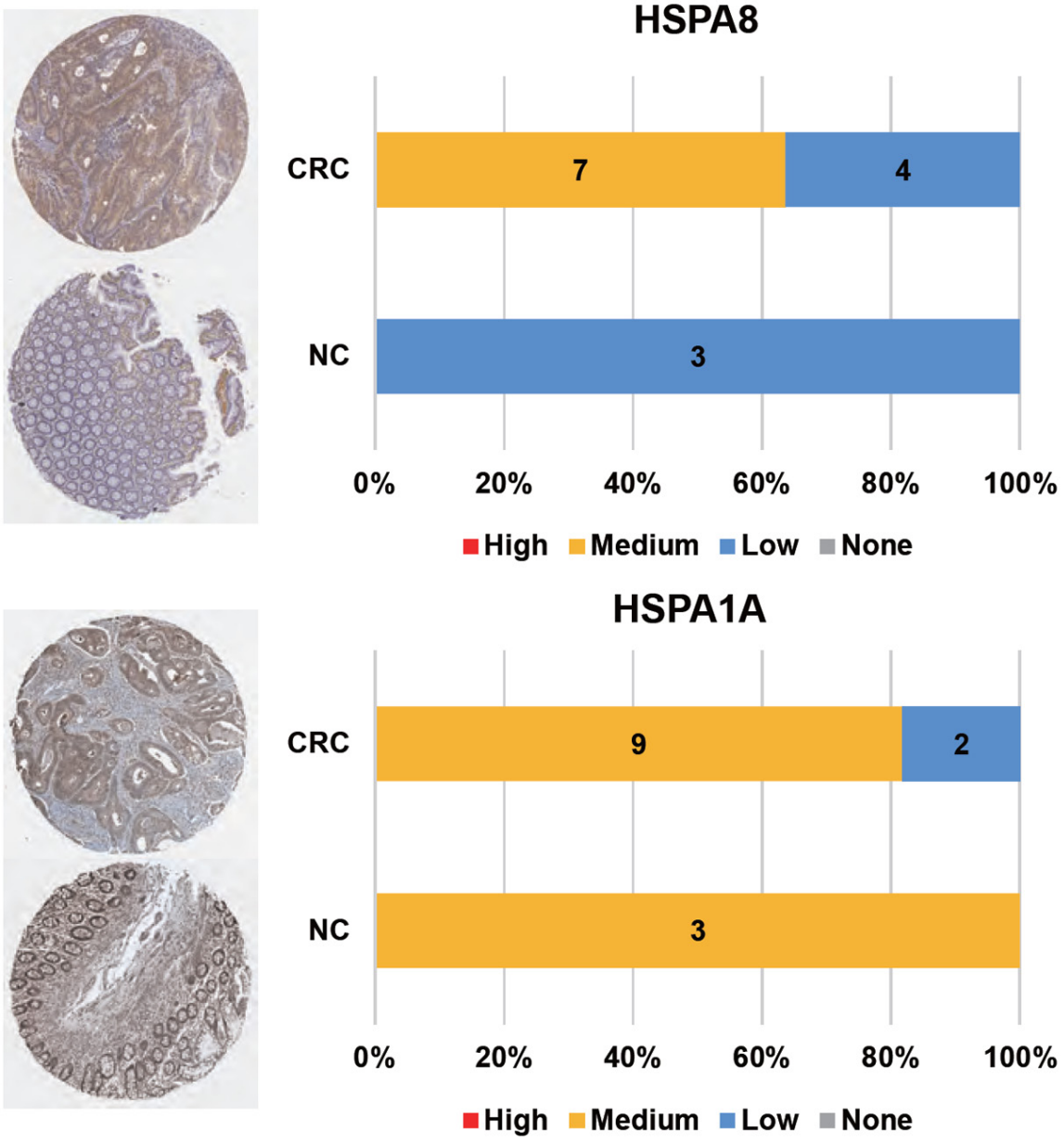
Protein accession	Q14677
Position	299
Ratio	1.671
Regulated Type	Up
P value	0.00026415
Amino acid	S
Protein description	Clathrin interactor 1 OS=Homo sapiens OX=9606 GN=CLINT1 PE=1 SV=1
Gene name	*CLINT1*
Localization probability	0.999946
PEP	0
Score	272.63
Modified sequence	TIDLGAAAHYTGDKAS (1) PDQNASTHTPQSSVK
Charge	4
Mass error [ppm]	2.5436
MS/MS Count	17
Gene name	*CLINT1*
III_IV_C	1.158730159
III_IV_N	0.747805267
I_II_C	1.244174265
I_II_N	0.812757202
M_C	1.538116592
M_N	0.748444444
V_C	1.197419355
V_N	0.765179225

Supplementary Table 10. The intersections of OS-related genes and VGs differentially expressed in protein propylene acylation level in CRC

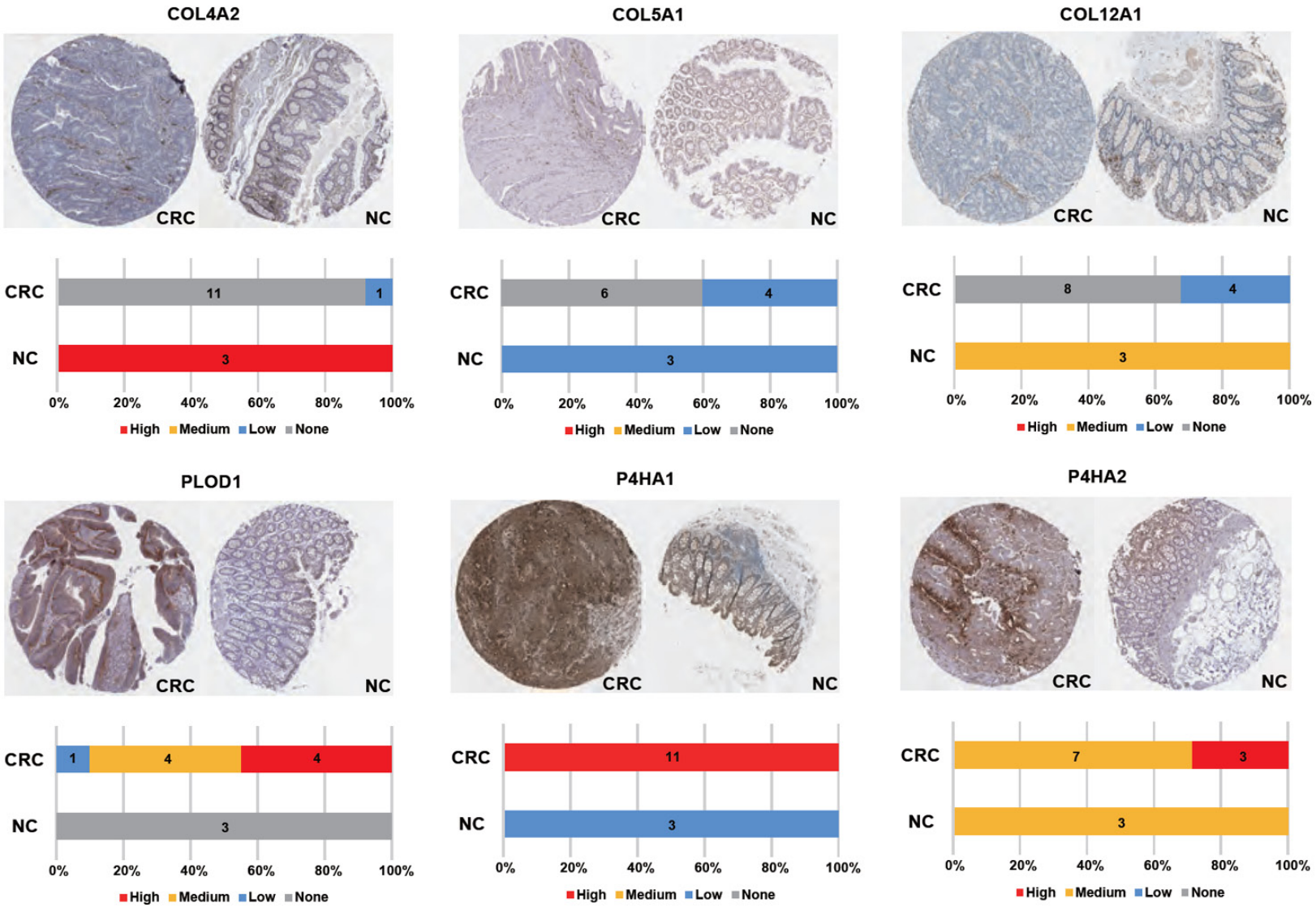
Subcellular localization	cytoplasm, nucleus
Protein accession	P15170
Position	493
Amino acid	K
C_C/C_N Ratio	1.727
Regulated Type	Up
Protein description	Eukaryotic peptide chain release factor GTP-binding subunit ERF3A OS=Homo sapiens OX=9606 GN=GSPT1 PE=1 SV=1
Gene name	*GSPT1*
Localization probability	0.999954
PEP	0.00816678
Score	46.068
Charge	2
Modified sequence	VLK (1) LVPEKD
Mass error [ppm]	0.4968
C_C	1.228
C_N	0.711

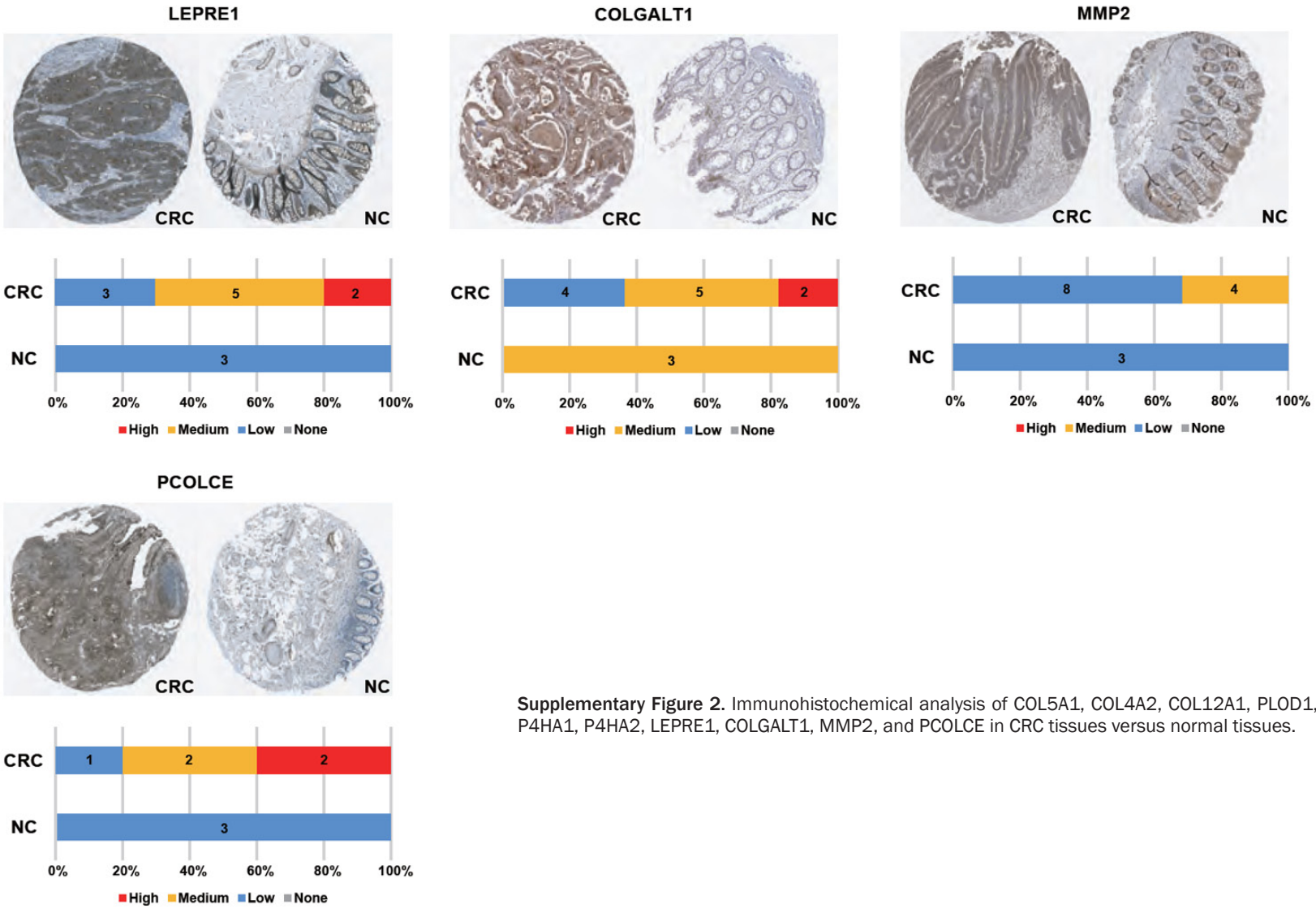
Supplementary Table 11. The mRNA expression of HSPA8 and HSPA1A in CRC cells versus normal cells

Gene	conMean	treatMean	logFC	*p*Value
HSPA8	205.5412804	329.827131	0.682281907	4.28E-16
HSPA1A	21.55032529	15.62257195	-0.464077662	1.09E-08

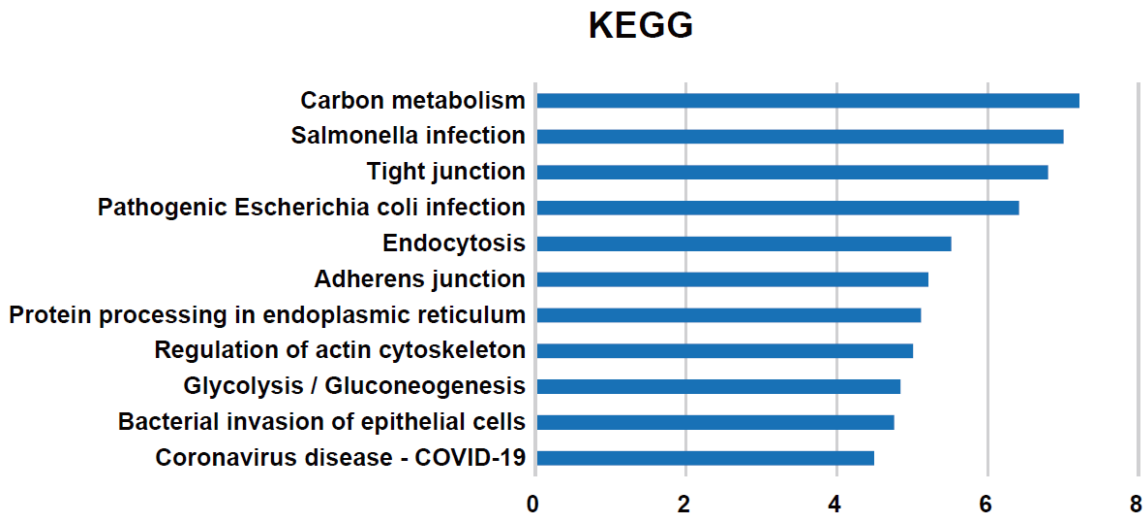


Supplementary Figure 1. Immunohistochemical analysis of HSPA8 and HSPA1A in CRC tissues versus normal tissues.


COL4A2
CRC
NC
CRC
NC
11
1
3
0%
20%
40%
60%
80%
100%
High
Medium
Low
None
COL5A1
CRC
NC
CRC
NC
6
4
3
0%
20%
40%
60%
80%
100%
High
Medium
Low
None
COL12A1
CRC
NC
CRC
NC
8
4
3
0%
20%
40%
60%
80%
100%
High
Medium
Low
None
PLOD1
CRC
NC
CRC
NC
1
4
4
3
0%
20%
40%
60%
80%
100%
High
Medium
Low
None
P4HA1
CRC
NC
CRC
NC
11
3
0%
20%
40%
60%
80%
100%
High
Medium
Low
None
P4HA2
CRC
NC
CRC
NC
7
3
3
0%
20%
40%
60%
80%
100%
High
Medium
Low
None



Supplementary Figure 2. Immunohistochemical analysis of COL5A1, COL4A2, COL12A1, PLOD1, P4HA1, P4HA2, LEPRE1, COLGALT1, MMP2, and PCOLCE in CRC tissues versus normal tissues.



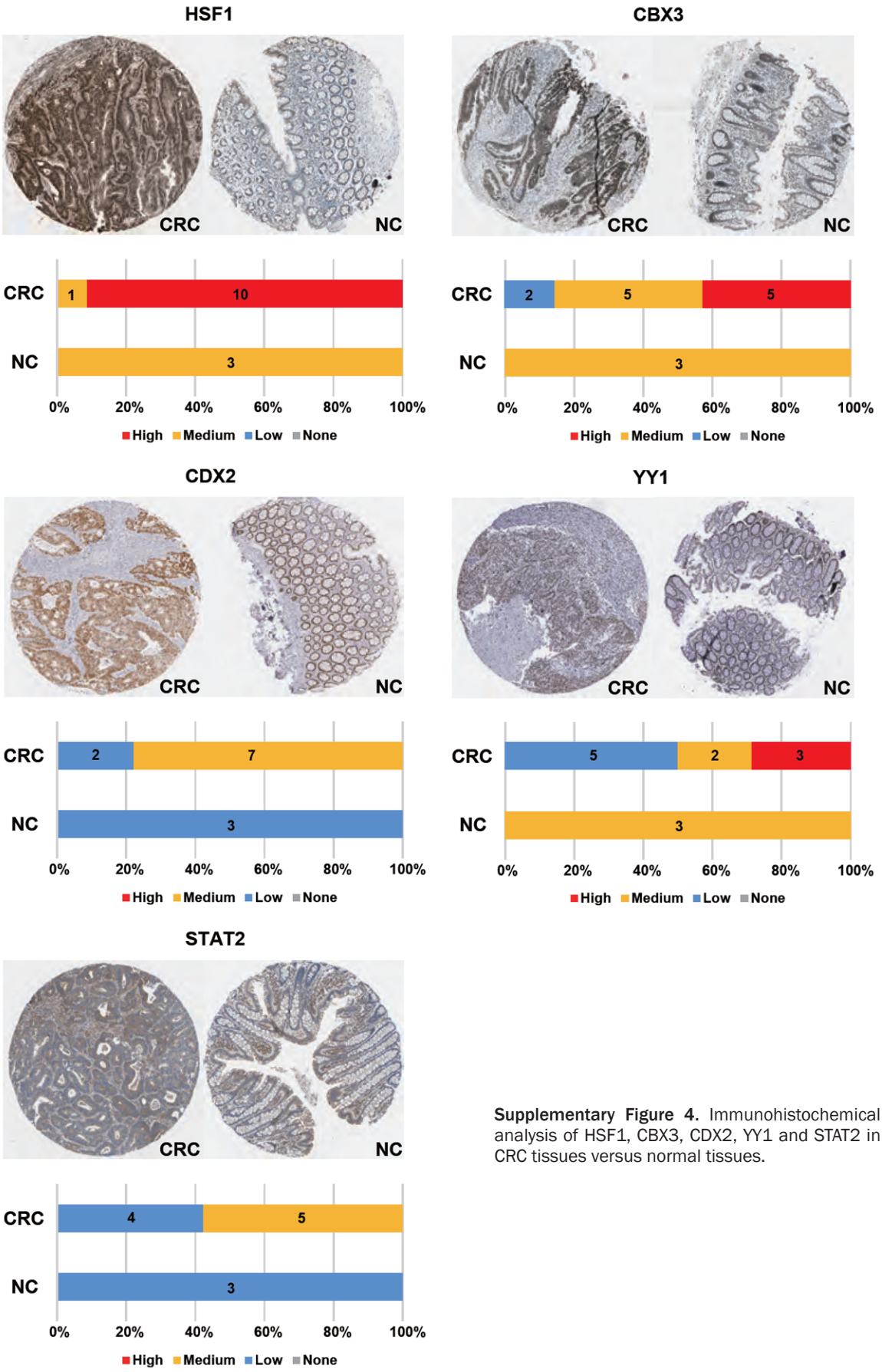
Supplementary Figure 3. KEGG analysis of differentially expressed genes in protein ubiquitination, phosphorylation and propylene acylation levels of CRC cells versus normal cells.

Supplementary Table 12. 85 genes changed both in mRNA and protein levels in CRC cells versus normal cells

List of 85 genes	RCN3 COL4A2 SERPINB5 OSBPL8 CD276 ZPR1 PLEK2 SQSTM1 ALDOC SLC7A5 P3H1 PLIN2 TOMM34 PRAF2 AGTRAP P4HA1 VCAN PDGFRB PDCD11 EEPD1 LDHA SBSPON SNX8 GPR-C5A COLGALT1 POP7 SPARC COL5A1 CEBPZ FKBP9 HIP1 TFRC SULT2B1 LPCAT1 DPY19L1 COL12A1 NSDHL PLOD1 RBP5 ITIH5 PACSIN2 ECHDC1 NOSTRIN AGR3 PLCG2 HSD11B2 TEP1 SLC44A2 CASP10 TSPAN8 SCP2 HSD17B11 HK2 PAG1 AKAP9 ATP8A1 BCAS1 CHGA CAPN9 SLC26A2 PBLD ST14 METTL7A CTSS TRAF3IP3 FAM177A1 GCG GCNT3 GLIPR2 CPM LIPA ANO10 TCEA3 CPE CTSA CTSZ GNG2 GSTM2 CPQ ECI2 TP53I3 HAPLN1 ZADH2 ACSS2 IQGAP2

Supplementary Table 13. 144 TFs of the nuclear genes which were differentially expressed in CRC cells versus normal cells at both mRNA and protein levels

List of 144 TFs	CDK9 CTCF SP1 CDK8 CDX2 CEBPB CLOCK CTNNB1 EHF ETS2 FOXD2 FOXG1 HNF4A HOXA1 HOXB13 JUND KLF5 MECOM MED1 MED12 MYC NIPBL RAD21 REPIN1 RXRA SMC1A SMC3 SOX2 TAL1 TCF4 ZNF250 E2F7 EP300 ETV5 FOXP1 GLI2 HINFP JUN KLF3 LYL1 MAFK MNT POLR2A TAF3 USF1 ATF3 E2F3 EGR1 FOSL1 GMEB2 NFYC RBCK1 TCF7L2 YY1 ZBTB33 ZNF12 ZNF143 ZNF263 ZC3H8 ZNF83 ELF1 ETV7 KMT2B ASCL2 CAMTA2 E2F2 HOXA4 MAX NFAT5 RELA RFX2 SIN3A TFDP1 VEZF1 REST TEAD4 ZFHX3 CBX3 GLIS1 USF2 ZBTB10 ZBTB7B NCOR1 BCL6 E2F8 ESRRA HOXA6 HOXC8 NFYA SP2 TERF1 VDR ZNF395 SRF MYBL2 ARNTL BHLHE40 EMX1 HBP1 NKX2-5 SMAD3 SOX4 STAT2 ZNF280D ZNF84 GMEB1 ELL2 FOXA2 GATA6 HOXC6 LHX2 RUNX1 TFAP2A BARX2 CTCFL TBX3 ELF2 CASP8AP2 NPAT HSF1 ESR1 TP73 DBP CREM DLX2 ELF3 EOMES FOXO3 GLI1 HOXA13 ISL1 PDX1 RORA ZFAT ZBTB17 CREB3 ATF5 MEF2C ZNF236 GATA1 JAZF1 NCOR2 MAFG ARID3A



Supplementary Figure 4. Immunohistochemical analysis of HSF1, CBX3, CDX2, YY1 and STAT2 in CRC tissues versus normal tissues.