Molecular Biology Reports (2024) 51:576
https://doi.org/10.1007/s11033-024-09522-5

ORIGINAL ARTICLE

Identification of genes related to ribosomal proteins in colorectal cancer: exploring their potential as biomarkers, prognostic indicators, and therapeutic targets

Negin Salehinia[1,4] · Aseel Kamil Mohammad Al-Mosawi[2,4] · Duaa Kamel Al-Moussawi[3,4] · Ensieh Sagheb Sadeghi[4] · Atefeh Zamani[4,5] · Mohammad Mahdevar[4,6]

Received: 10 February 2024 / Accepted: 3 April 2024
© The Author(s), under exclusive licence to Springer Nature B.V. 2024

Abstract

Background Colorectal cancer (CRC) ranks as the third most commonly diagnosed cancer in both females and males, underscoring the need for the identification of effective biomarkers.

Methods and results We assessed the expression levels of ribosomal proteins (RPs) at both mRNA and protein levels. Subsequently, leveraging the STRING database, we constructed a protein-protein interaction network and identified hub genes. The co-expression network of differentially expressed genes associated with CRC and their target hub RPs was constructed using the weighted gene co-expression network analysis algorithm. Gene ontology and molecular signatures database were conducted to gain insights into the biological roles of genes associated with the identified module. To confirm the results, the expression level of the candidate genes in the CRC samples compared to the adjacent healthy was evaluated by the RT-qPCR method. Our findings indicated that the genes related to RPs were predominantly enriched in biological processes associated with Myc Targets, Oxidative Phosphorylation, and cell proliferation. Also, results demonstrated that elevated levels of *GRWD1, MCM5, IMP4*, and *RABEPK* that related to RPs were associated with poor prognostic outcomes for CRC patients. Notably, *IMP4* and *RABEPK* exhibited higher diagnostic value. Moreover, the expression of *IMP4* and *RABEPK* showed a significant association with drug resistance using cancer cell line encyclopedia and genomics of drug sensitivity in cancer databases. Also, the results showed that the expression level of *IMP4* and *RABEPK* in cancerous samples was significantly higher compared to the adjacent healthy ones.

Conclusion The general results of this study have shown that many genes related to RPs are increased in cancer and could be associated with the death rate of patients. We also highlighted the therapeutic and prognostic potentials of RPs genes in CRC.

Keywords RT-qPCR · Co-expression · WGCNA · Prognosis · Targeted therapeutic

✉ Atefeh Zamani
atefeh.zamani20@gmail.com

✉ Mohammad Mahdevar
mahdevar416@gmail.com; mahdevar@resident.mui.ac.ir

[1] Department of Biology, Islamic Azad University, Qaemshahr, Iran

[2] Department of biology, College of Sciences, University of Thi Qar, Nasiriyah, Iraq

[3] General Directorate of Education in Thi-Qar, Ministry of Education, Al-Nasiriya City, Iraq

[4] Genius Gene, Genetics and Biotechnology Company, Isfahan, Iran

[5] Department of Biology, Faculty of Basic Sciences, Shahrekord Branch, Islamic Azad University, Shahrekord, Iran

[6] Department of Genetics and Molecular Biology, School of Medicine, Isfahan University of Medical Sciences, Isfahan, Iran

Published online: 25 April 2024

Introduction

Colorectal cancer (CRC) holds the position of the second leading cause of mortality on a global scale, primarily attributed to its pervasive nature of metastasis and inherent heterogeneity [1]. Even with progress in early detection and therapeutic approaches, the persistence of lymphatic and distant metastases remains a significant factor contributing to decreased survival rates [2]. It continues to be the primary cause of mortality for individuals diagnosed with CRC. So, identification of genes that may dominate a wide range of mechanisms could highlight the sensitive biomarkers for determining prognosis, monitoring recurrence, and enhancing the management of personalized therapy in CRC.

The process of ribosome biogenesis is a remarkably dynamic and well-coordinated sequence, involving the synthesis, modification, and assembly of ribosomal RNA (rRNA) with ribosomal proteins (RPs) to ultimately create fully mature ribosomes. This intricate process occurs within specialized subnuclear compartments recognized as nucleoli [3]. Increasing evidence indicates that post-translational modifications of ribosomal proteins, such as phosphorylation and ubiquitylation, may actively impact translational control and are associated with various human diseases [4]. The disruption of RPs can have consequences for ribosomal biogenesis, potentially compromising cell survival, growth, and proliferation. Research indicates that RPs extend beyond their primary ribosomal functions, playing roles in extra ribosomal activities such as DNA repair, replication, proliferation, apoptosis, and resistance to chemotherapy [5]. Numerous RPs exhibit overexpression in diverse human tumors, encompassing prostate cancer (PCa), colon adenocarcinomas, liver, pancreatic, gastric, lung, and breast cancer, among various others [6, 7]. For instance, Luo et al. have documented an elevation in the expression levels of RPS2 and RPS12 within tissues afflicted by colon cancer [8]. In addition, recent findings regarding somatic mutations in RP genes within hematological cancers and solid tumors underscore the connection between abnormalities in ribosome biogenesis and the process of oncogenic transformation [9]. Therefore, considering the expression pattern of RPs and identifying the genes associated with RPs in cancers, particularly in CRC, could play a crucial role in understanding the molecular mechanisms underlying the disease and developing targeted therapeutic approaches.

In this study, we employed RNA-Seq data from the TCGA dataset to elucidate the prognostic and biomarker significance of genes associated with RPs through bioinformatics and statistical analysis. Subsequently, we identified drugs capable of targeting our candidate genes and assessed the correlation between the expression levels of these genes and drug resistance and sensitivity. Finally, the validation of two genes related to ribosomal proteins in colorectal cancer tissues was conducted using qRT-PCR.

Martial and methods

Differential expression analysis of RPs in colorectal cancer vs. normal tissues using TCGA data and limma package

RNA expression data for 41 normal tissue samples and 483 CRC samples, along with their respective clinical information, were obtained from the TCGA database (https://portal.gdc.cancer.gov/). Differential gene analysis was conducted on colon cancer and normal tissue samples sourced from the TCGA dataset. Utilizing the "limma" package in the R language, the analysis employed normal tissue samples as controls, and the FDR method was employed to rectify the differential *p*-values. A selection criterion of $-0.5 \leq |\log FC| \geq 0.5$ and FDR < 0.05 was applied for differential gene selection, leading to the identification of significantly differentially expressed genes (DEGs). The list of RPs was achieved from a published article about RPs [10].

To confirm our results via using second cohort, we downloaded the raw data from both the GSE39582 and GSE22598 studies, which consisted of 583 primary gastric adenocarcinoma samples and 36 surrounding normal fresh frozen tissues. Following data acquisition, initial preprocessing steps were implemented. These steps included background light removal, data normalization using the RMA method, and transformation of the data into logarithmic mode with a base of 2. The resulting expression matrix obtained from these preprocessing procedures formed the foundation for all subsequent analyses conducted in the study. Additionally, we obtained clinical data from the GSE39582 dataset, which included parameters such as age at diagnosis, TNM stage, overall survival event (OS), and overall survival delay (in months). All the codes used in this study are given in supplementary materials.

Protein-protein interaction network

STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) is an online tool (https://string-db.org/) designed for protein interaction analysis. It leverages known protein-protein correlation data to establish upstream and downstream relationships among proteins [11]. The shared differentially expressed RPs were input into the STRING software to construct and visualize the protein-protein interaction (PPI) network. Additionally, the cytoHubba plugin within Cytoscape software (version Cytoscape_v3.6.1) was employed for the identification of hub genes. The top RPs

with the highest connection and maximal clique centrality (MCC) were selected as hub genes.

Exploring colorectal cancer co-expression networks: WGCNA-based module identification and trait correlation analysis

The CRC sample co-expression network was established using the WGCNA algorithm. The construction of the gene co-expression network utilized the "WGCNA" package within the R software. Hierarchical clustering analysis was executed through the Hclust function. The suitable soft threshold value (β) was determined using the "pickSoftThreshold" function, and the adjacency matrix underwent transformation into a topological overlap matrix (TOM). Subsequently, a hierarchical clustering tree was created to categorize similar gene expressions into distinct modules. The minimum gene count within each module was established at 30. To consolidate potentially similar modules, a cutting height threshold of 0.2 was established. Ultimately, the module eigengenes (ME) were computed to encapsulate the expression profiles of each gene, and the correlation between ME and traits was evaluated to pinpoint the most pertinent module for subsequent analysis [12, 13].

Table 1 Clinical data pertaining to CRC patients

Cancer name	characteristic	Number ($N=30$)
CRC	**Age**	
	< 50	9
	> 50	21
	Gender	
	Male	18
	Female	12
	TNM stage	
	I	5
	II	11
	III	9
	IV	5
	Tumor size	
	< 5cm	12
	> 5cm	18
	Tumor site	
	Right colon	6
	Left colon	7
	Cecum	2
	Sigmoid	10
	Rectum	5
	TNM.N	
	N0	17
	N1	13

Conducting intersection gene screening

Identification of candidate genes by overlapping critical and differentially expressed genes in the weighted gene Co-expression Network using the 'venn' package in R.

Gene ontology and MSigDB pathway analysis

The GO analysis of the identified differentially expressed RPs (DERPs) was conducted using a web-based tool database. Enrichment analysis for the obtained DERPs in the MSigDB pathway was performed using Enrichr. A significance threshold of $P < 0.05$ was applied for statistical significance.

Survival analysis

For survival analysis, gene expression values were stratified into low and high expression groups using the R language. The hazard ratio (HR) was calculated through a Cox regression model, and Kaplan-Meier estimations were utilized to generate survival curves. A significance threshold of $P < 0.05$ was applied to identify statistically significant differences.

Diagnostic significance of top hub genes: ROC curve analysis

The pROC package was employed for Receiver Operating Characteristic (ROC) curve analysis on hub genes. A cutoff value of AUC > 85% was established to assess the diagnostic significance of the identified hub genes.

Drug discovery, drug sensitivity and resistance

To identify drugs influencing the expression of candidate genes, this study utilized the DRUG bank, DSigDB, and other drug discovery datasets. Relevant drugs were selected based on keywords such as cancer, treatment, and the names of candidate genes. Moreover, the cancer cell line encyclopedia (CCLE) data and genomics of drug sensitivity in cancer (GDSC) data were analyzed using the R "PharmacoGx" package. This examination aimed to investigate the correlation between the expression levels of candidate genes and both drug resistance and sensitivity. Ultimately, this information was utilized to assess the correlation between the expression levels of candidate genes and the IC50 values of various drugs.

Sample collection

In this investigation, a total of 60 colon tissues, comprising 30 tumor samples and 30 adjacent healthy tissues, were sourced from the Tumor Bank of Iran (Table 1). The study adhered to bioethical standards outlined by the Ministry of Health, Treatment, and Medical Education of Iran, and it received approval from the research ethics committees of the College of Science, University of Tehran, under the accession code IR.UT.SCIENCE.REC.1401.007. Additionally, all cancer samples were verified by a pathologist to confirm their disease status.

RNA extraction, quantitative real-time PCR

Total RNA was extracted from tissues using TRIzol reagent (Sigma-Aldrich, Germany). Subsequently, DNase treatment (Sina Clone) was applied to eliminate DNA contamination. For all samples, cDNA synthesis was carried out using a cDNA synthesis kit (Yekta Tajhiz, Iran) following the provided procedure. The *IMP4* and *RABEPK* specific primers, designed by NCBI (https://www.ncbi.nlm.nih.gov/tools/primer-blast), had the following sequences: *IMP4*-F: 5'- CA TCACACACGGCTTCTCCT-3', R: 5'-CCAGACGGATC ATGTACAGCTT-3'; *RABEPK*-F: 5'-CCCACCATCCCCA AGAACAT-3', R: 5'-CAGGTCAGAGTGTTTGCGTC-3'. RT-qPCR was performed using SYBR Green PCR master mix (Yekta Tajhiz, Iran) and gene-specific primers. *GAPDH* was used as an internal control, and the expression levels of each gene in each sample were calculated using the $2^{-\Delta Ct}$ method.

Statistics

All data processing and statistical analyses were conducted using R software (version 4.1.0). The data are presented as mean ± SD. Statistical comparisons were carried out using either a two-tailed Student's t-test or a two-tailed non-parametric Mann–Whitney U-test, with statistical significance set at $P < 0.05$ ($*P < 0.5$; $**P < 0.1$; $***P < 0.01$; $****P < 0.001$). Correlations between two continuous variables were evaluated using Pearson's correlation coefficients. The Survminer package was utilized to determine the optimal cut-off value. Cox regression and Kaplan–Meier analyses were conducted through the survival package.

Results

Identification of DEGs

To explore the number of dysregulated genes in CRC, we conducted a differential gene expression analysis comparing tumor samples to normal samples. This comparison revealed a total of 13,173 differentially expressed genes, with 6,594 upregulated genes and 6,580 downregulated genes. Our outcomes were represented by volcano plot (Fig. 1A).

Analysis of PPI networks and identification of hub genes

RPs, which make up the structural components of the ribosome, play a crucial role in both the assembly and functioning of the ribosome. Among 77 achieved RPs from database, our investigation unveiled dysregulation in 25 RPs within CRC samples in comparison to normal samples (Fig. 1B). Utilizing transcriptome data from TCGA database, the expression levels of these genes were detailed in supplementary file (Table S1). To further explore the interactions between these RPs at the protein level, we constructed a PPI network using the STRING online database. Subsequently, employing the Cytoscape plug-in CytoHubba, we identified the top 5 genes in the network through two key topological analysis algorithms (MCC and Degree). The resulting hub genes are as follows: *RPL35, RPL36A, RPS17, RPS4X*, and *RPS19* (Fig. 1C and E). As reported in Table S1, the expression levels of *RPL35, RPL36A, RPS17, RPS4X*, and *RPS19* were up-regulated, and among all of these genes, *RPL36A* and *RPS17* had significantly higher expression pattern in CRC compared to normal samples (Fig. 2A-E, $P < 0.001$). To validate the expression levels of our hub genes at protein level, we assessed their protein levels using a proteomics database. Interestingly, elevated expression of hub genes was validated in CRC samples vs. normal samples (Fig. 2F-J, $P < 0.01$). These results suggest that the mentioned RPs could play an important role in the pathogenesis of CRC.

Building co-expression modules for CRC using WGCNA

The 483 samples of CRC were clustered utilizing the average linkage method, and the levels of 16,352 genes were used to create the co-expression gene networks, and a sample dendrogram was constructed. The soft-thresholding power was selected 12 which ensure a scale-free network R^2 (scale-free $R^2 = 0.9$). By using the dynamic tree cut method, a total of 21 modules were detected, and the clustering dendrograms of genes and a heat map for the module-trait association are illustrated in Fig. 3A-C. As shown in heatmap

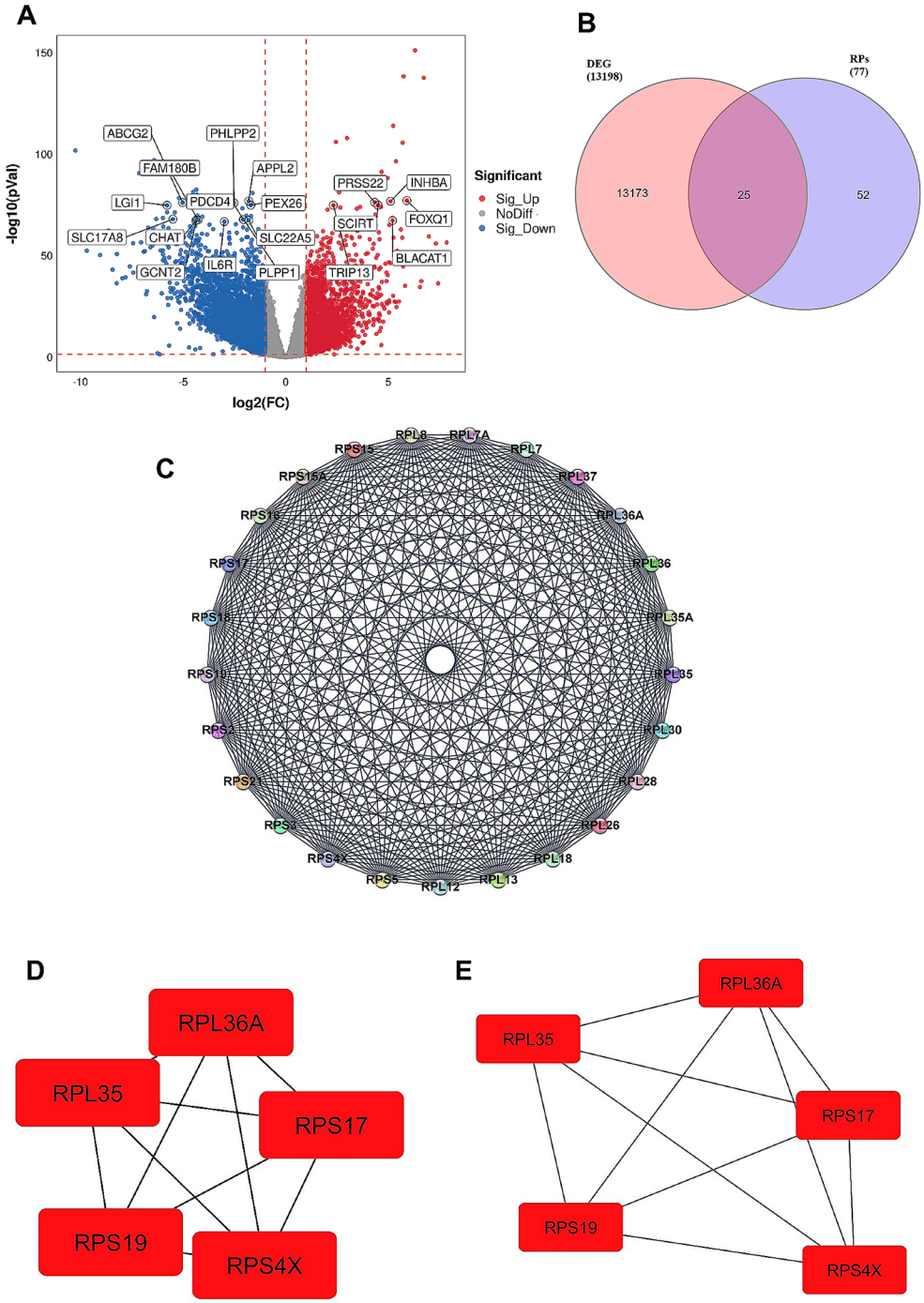

Fig. 1 Gene expression changes in CRC. (**A**) The volcano plot illustrates the Differentially Expressed Genes (DEGs) in cancer samples compared to the healthy group, utilizing TCGA data. Genes meeting the criteria of $-1 \leq |\log FC| > 1$ and Adj.*P*. value < 0.05 are highlighted with red and blue dots in the figure. (**B**) A Venn diagram displays the overlap of genes obtained from DEGs and ribosomal proteins (RPs). (**C**) The Protein-Protein Interaction (PPI) network is depicted for RPs based on the STRING database. (**D** and **E**) Genes with the highest degree and MCC scores were identified using the cytoHubba plugin

plot, 21 co-expression modules were discovered, and except for gray module, which is non-clustering genes, the green module was the most related module to five top RPs genes (Fig. 3B). So, the green module was considered for further analysis. Only genes exhibiting a correlation coefficient greater than 0.2 and a *p*-value less than 0.05 with all five top RP genes were extracted as candidate genes from the green module. 624 genes in green module were identified as candidates, of which 419 genes were significantly differentially expressed in CRC compared to normal samples (Fig. 4A). Therefore, we used these 419 RPs-related genes for continuing our study.

Functional annotation and pathway enrichment analysis for co-expressed genes

To gain a more comprehensive understanding of the functions of these 419 RPs-related genes, we conducted gene

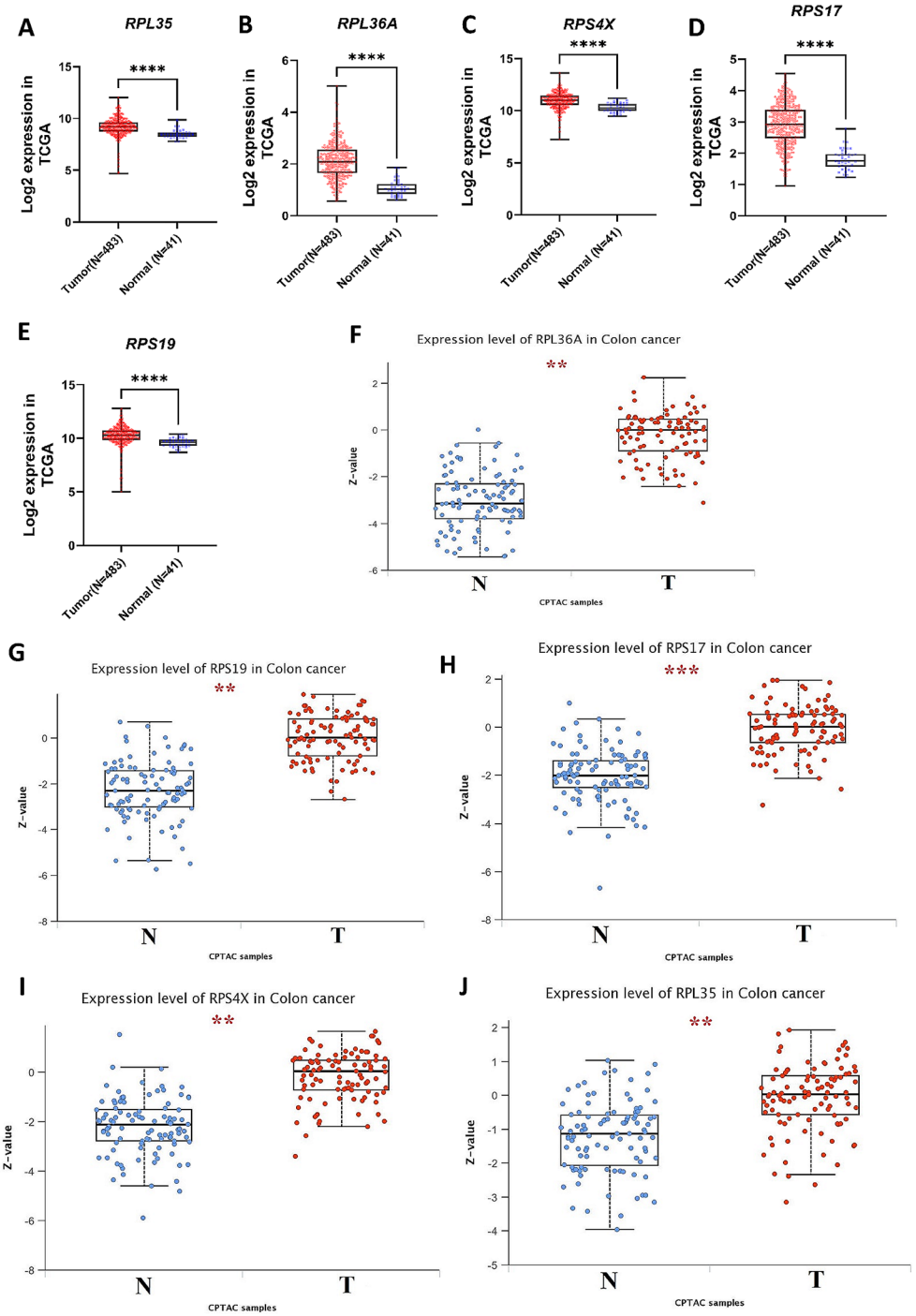

Fig. 2 Integrated Analysis of Hub Ribosomal Proteins: mRNA and Protein Expression. (**A-J**) The mRNA and protein expression levels of hub Ribosomal Proteins (RPs) were assessed using both TCGA and proteomics datasets. A significance cutoff of *P*-value less than 0.05 was applied

ontology (GO) annotation and the molecular signatures database (MSigDB) pathway enrichment analyses. Our enrichment analysis revealed that our RPs-related genes were enriched in 14 biological processes (BP), 19 cellular components (CC), and 5 molecular functions (MF) terms with FDR less than 0.01 (Fig. 4B). Our MSigDB pathway enrichment results also showed that RPs-related genes could act in Myc Targets V2, Oxidative Phosphorylation, Myc Targets V1, Adipogenesis, G2-M Checkpoint, mTORC1 Signaling, E2F Targets, and DNA Repair pathways with FDR less than 0.01 (Fig. 4C). Therefore, the first pathway (Myc Targets V2) including: *PUS1; PLK1; IMP4; NOC4L; RABEPK; RRP9; SRM; TBRG4; AIMP2; CDK4; GRWD1* and *MCM5* genes was selected as the top and most relevant pathway for our ongoing study on RPs genes (Fig. 4B and C). These results show that RP-related genes are involved in the main pathways related to cancer such as cell proliferation.

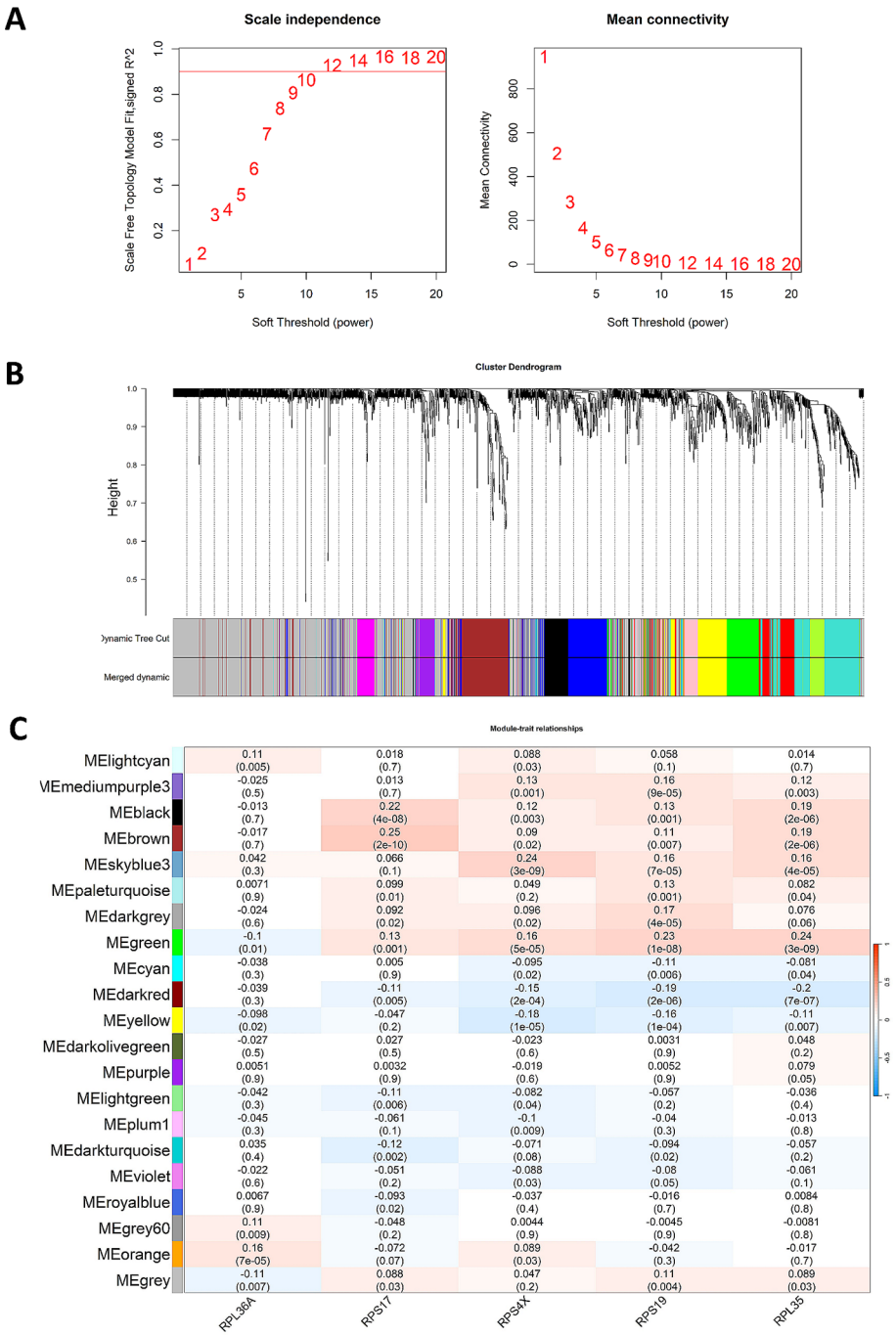

Fig. 3 WGCNA was employed to identify key module genes significantly associated with hub RPs scores. (**A**) The scale-free fit index and mean connectivity were analyzed across various soft-thresholding powers to determine the optimal soft-thresholding power. (**B**) The cluster dendrogram, based on gene expression, illustrates co-expression modules identified by different colors. (**C**) Heatmap of module-trait relationships was generated to identify key modules

Hub genes' survival curve analysis

Firstly, the expression levels of our candidate genes were separately represented by violine plot in CRC samples and normal samples. As shown in Fig. 5A and L, the levels of all candidate genes were significantly up-regulated in CRC samples compared to normal samples. To recognize the OS-related genes, univariate cox regressions analysis was performed. Patients were divided into the low- and high-expression categories utilizing the median gene level as cut-off. Our univariate cox results indicated that among all candidate genes, the high levels of *GRWD1, MCM5, IMP4*, and *RABEPK* was related to poor prognostic of patients in CRC (Supplementry file: Table S2, HR > 1 and logRank < 0.05). Moreover, the K-M curve was plotted to show the difference between the rate of survival between high and low expression using the survival package of R language (Fig. 6A and D). These results suggest that RP-related genes could be associated with the survival rate of CRC patients and play a role in malignancy.

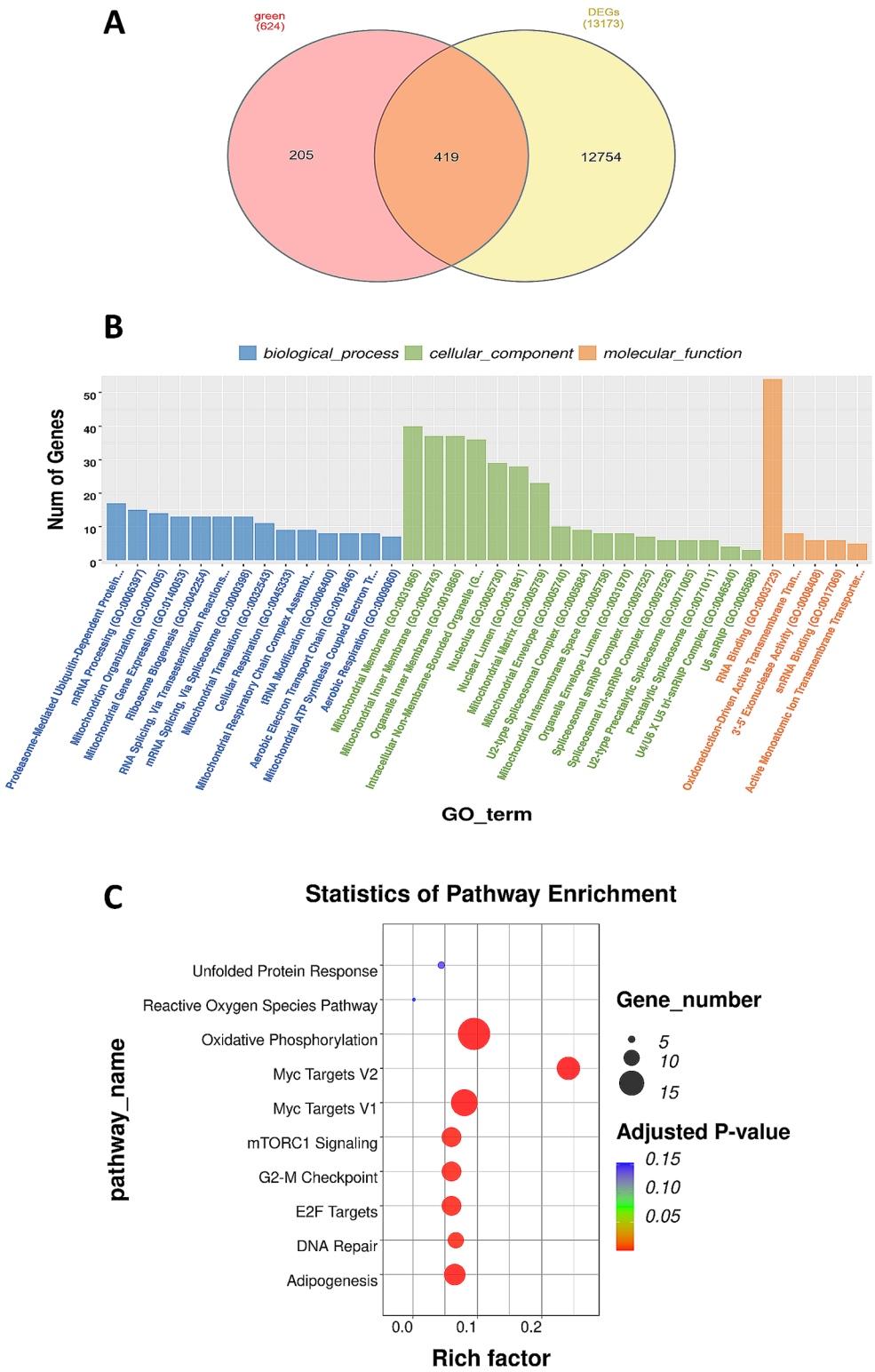

Fig. 4 Enrichment analysis diagrams. (**A**) A Venn diagram was constructed to illustrate the overlap of 419 candidate hub genes by intersecting the key green module genes with DEGs from CRC and control samples. (**B**) The candidate genes were subjected to pathway analysis using the gene ontology (GO) and molecular signatures database (MsigDB) databases to identify relevant biological pathways

Evaluation the diagnostic value of hub genes in CRC

To evaluate the potential of our candidate genes as biomarkers in CRC, we estimated their expression levels using the ROC metric with the TCGA-COAD normalized gene expression matrix. Our results showed that *PUS1*(AUC = 0.85, $P < 0.0001$), *PLK1* (AUC = 0.9, $P < 0.0001$), *IMP4* (AUC = 0.94, $P < 0.0001$), *NOC4L* (AUC = 0.87, $P < 0.0001$), *RABEPK* (AUC = 0.91, $P < 0.0001$), *RRP9* (AUC = 0.9, $P < 0.0001$), *SRM*

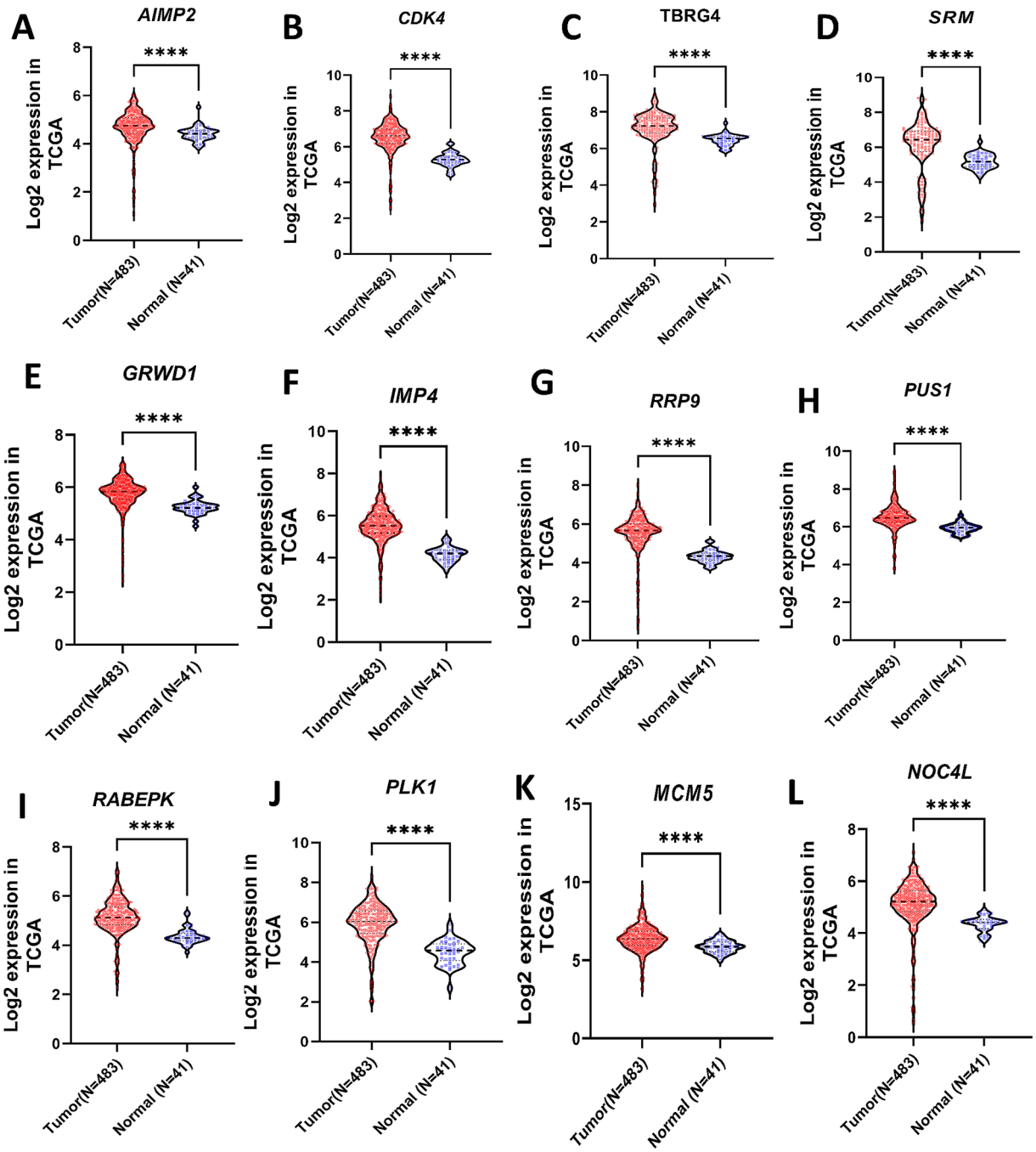

Fig. 5 Expression levels of 22 Hub Gene. (**A-L**) The expression levels of candidate genes were represented by box plot. A significance cutoff of *P*-value less than 0.05 was performed

(AUC = 0.81, $P = 0.002$), *TBRG4* (AUC = 0.86, $P < 0.0001$), *AIMP2* (AUC = 0.73, $P = 0.009$), *CDK4* (AUC = 0.91, $P = 0.001$), *GRWD1* (AUC = 0.87, $P = 0.002$), *MCM5* (AUC = 0.73, $P = 0.02$) genes could act as diagnostic markers in CRC. With the exception of *PUS1*, *NOC4L*, *SRM*, *TBRG4*, *AIMP2*, *GRWD1*, and *MCM5*, the remaining genes demonstrated a high efficiency in distinguishing tumor samples from normal ones (Fig. 7, A-L).

Identification of desirable drugs

To investigate the drugs with capability to have an influence on our candidate genes, some drug database including drug bank, DGIdb, and DSigDB were used. Our results indicate that alprostadil can target *CDK4, IMP4, RABEPK*, and *RRP9*, leading to a decrease in the expression levels of these genes (Table 2, $P < 0.05$). Additionally, ambroxol

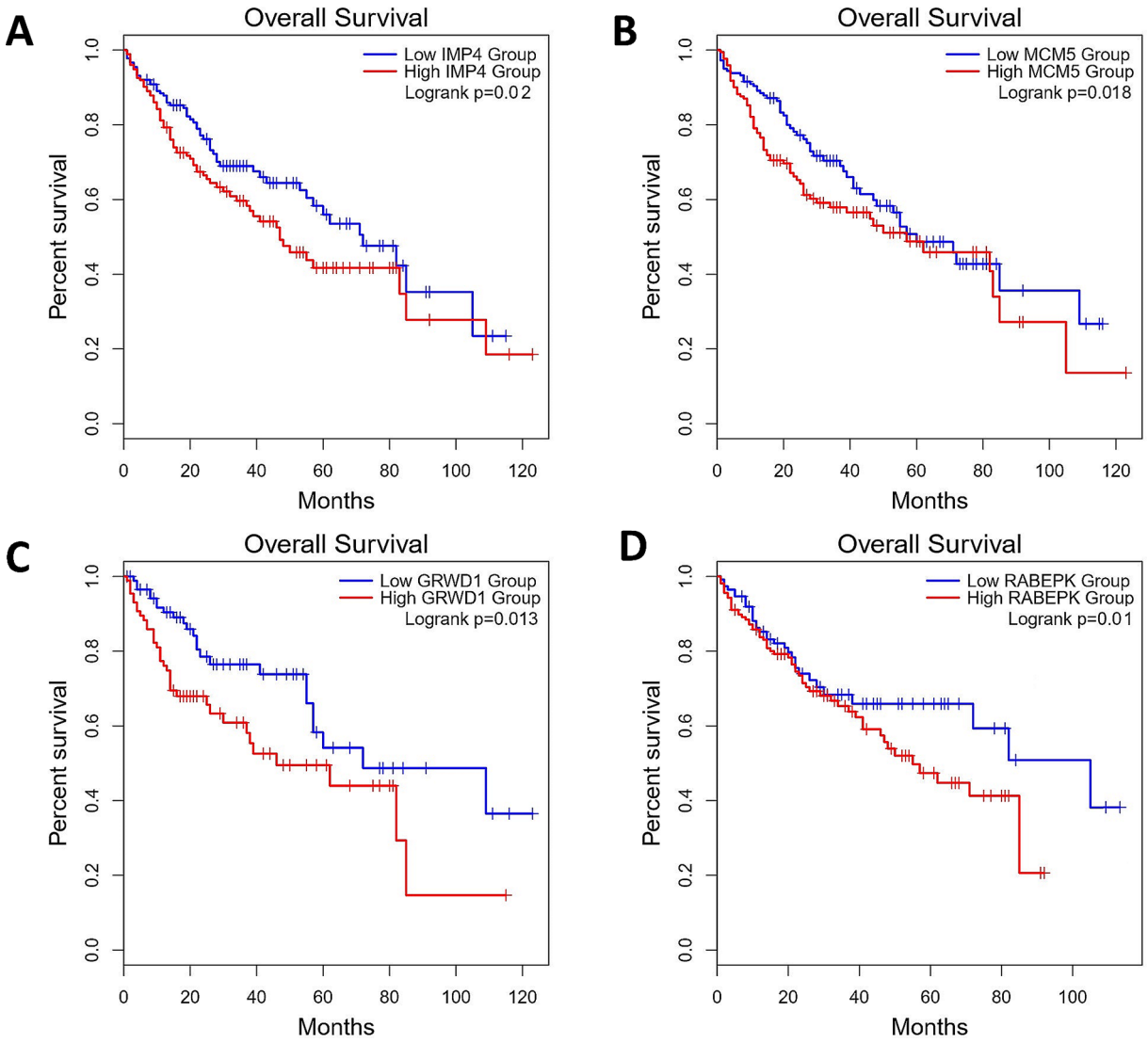

Fig. 6 Interaction of *GRWD1, MCM5, IMP4*, and *RABEPK* with genes associated with a poor prognosis. (**A-D**) All the genes under consideration displayed a noteworthy increase in expression in CRC, and this upregulation was identified to be correlated with an unfavorable prognosis for patients that were shown by Kaplan–Meier plots. The red represents of high expression and the blue represent the low expression

has been investigated for its potential to target *CDK4*, *IMP4, RABEPK*, and *RRP9*, thereby reducing their levels in pancreatic cancer (Table 2, $P < 0.05$). Palbociclib, an FDA-approved drug, has demonstrated its ability to target *CDK4* and *PLK1* in cancers such as breast cancer and CRC (Table 2, $P < 0.05$). Consequently, our findings suggest that the identified drugs may serve as effective agents in targeting genes related to RPs and associated pathways in CRC, both directly and indirectly.

The association between the levels of our candidate genes and drug resistance and sensitivity

We utilized data from the CCLE and GDSC databases to establish a link between the expression levels of candidate genes and drug resistance or sensitivity. To validate our findings, we employed the Pearson correlation test, as described in the materials and methods section. Our results revealed a significant correlation between the expression level of *IMP4* and resistance to Sorafenib (Fig. 8A, $P < 0.05$). Furthermore, the expression level of *RABEPK* was found to be associated with sensitivity to Irinotecan (Fig. 8B, $P < 0.05$). These findings underscore the potential correlation between the expression of specific candidate genes and drug resistance or sensitivity. However, for a comprehensive understanding of the role of these candidate genes, further investigations through in vitro and in vivo studies are imperative.

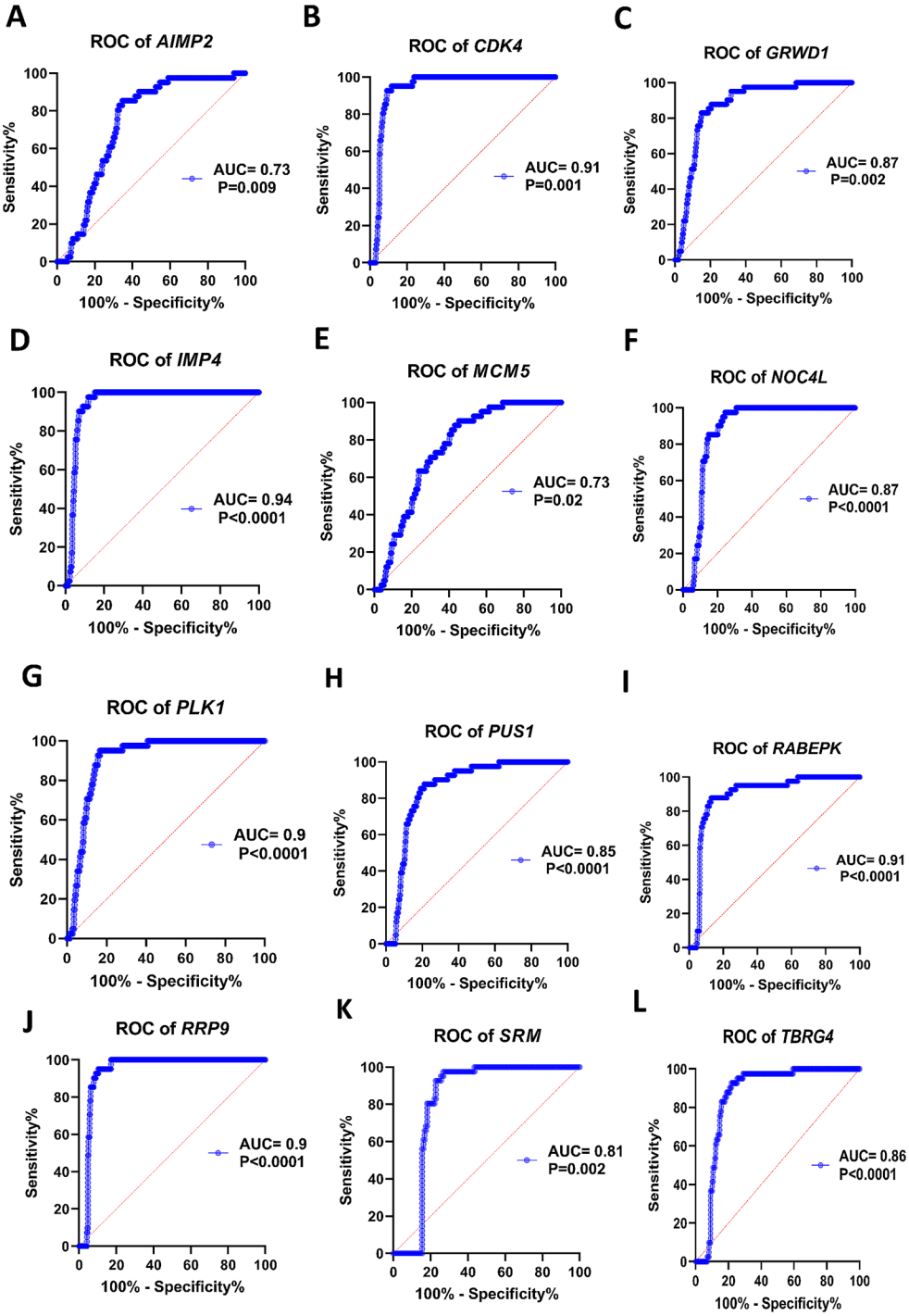

Fig. 7 ROC curve of Hub Gene. (**A-L**) ROC curves representing the identified hub genes in CRC samples compared to the healthy group are depicted utilizing TCGA data

Table 2 The potential of drugs to target candidate genes associated with RPs

Drug Names	*P*-value	Adjusted *P*-value	Genes
alprostadil	1.89E-06	7.64E-04	CDK4;IMP4;RABEPK; RRP9
ambroxol	4.52E-06	9.13E-04	CDK4;PLK1;IMP4;RRP9
Palbociclib	1.69E-04	0.017047	CDK4;PLK1

GEO cohort validation

To confirm the up-regulation of candidate genes in tumor samples compared to normal samples, we utilized GSE22598 and GSE39582 as validation cohorts, as described in the materials and methods section. Our findings indicate significant up-regulation of the expression levels of *PUS1, PLK1, IMP4, NOC4L, RRP9, SRM, CDK4*, and *MCM5* in both studies. Additionally, *RABEPK, TBRG4,*

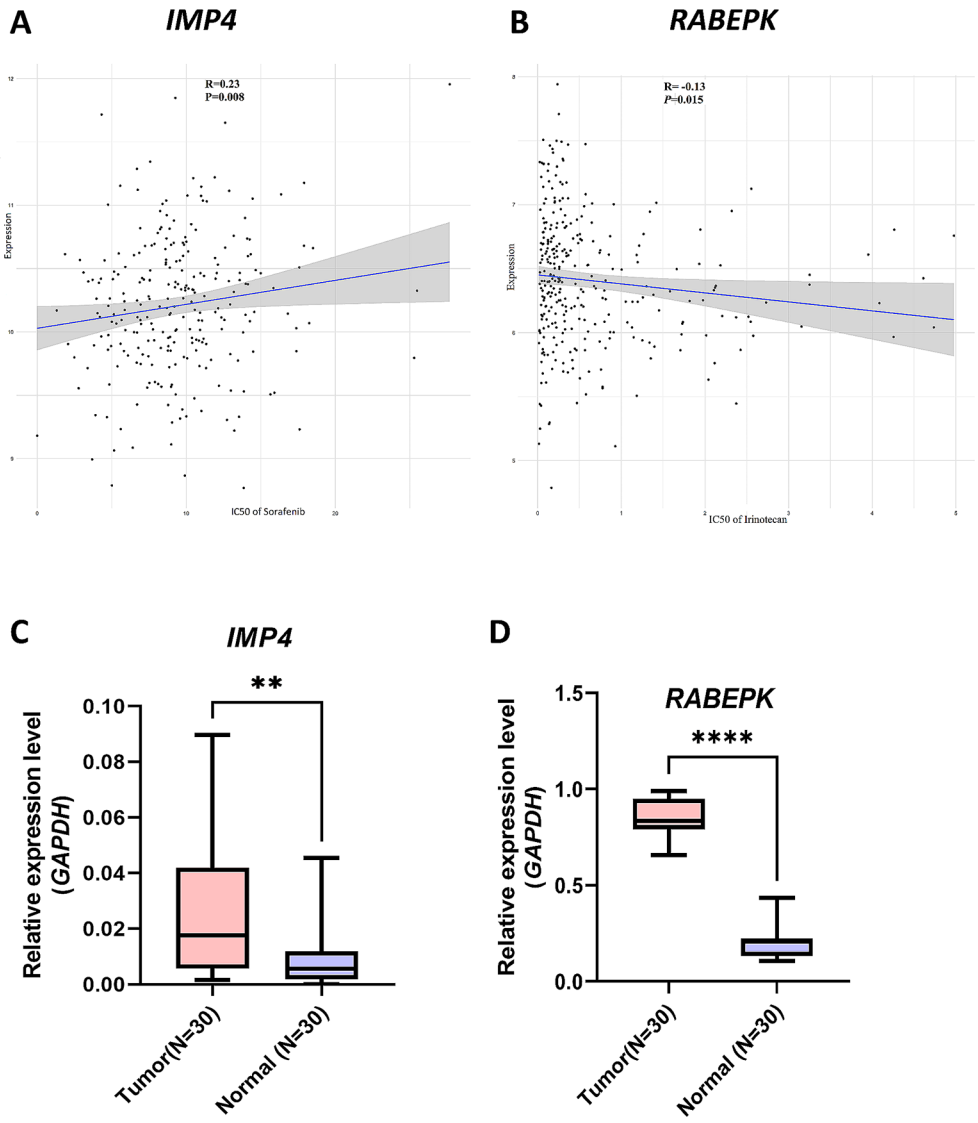

Fig. 8 The association between the expression levels of *IMP4* and *RABEPK* and their impact on drug resistance and sensitivities. qRT-PCR validation in colorectal cancer samples as compared to adjacent healthy tissue. (**A-B**) The correlation between candidate gene expression and drug resistance/sensitivity is demonstrated utilizing CCLE and GDSC data. A correlation test was conducted to evaluate the association between the expression levels of candidate genes and the respective IC50 values for each drug. (**C-D**) The assessment of *IMP4* and *RABEPK* levels in cancer samples relative to adjacent healthy tissue is depicted using the qRT-PCR method. The expression level of each gene in every sample was computed utilizing the 2-ΔCt method

Table 3 The change expression levels of selected genes in the second cohort

Study	GSE39582			GSE22598		
	logFC	*P*.Value	adj.*P*.Val	logFC	*P*.Value	adj.*P*.Val
PUS1	1.229835	2.99E-25	4.56E-27	1.07	5.34E-05	3.48E-06
PLK1	1.264995	1.42E-14	7.18E-16	8.36E-01	7.46E-03	1.55E-03
IMP4	0.834864	2.81E-15	1.29E-16	6.80E-01	4.20E-03	7.61E-04
NOC4L	0.616485	1.44E-08	1.69E-09	6.15E-01	7.90E-03	1.67E-02
RABEPK	0. 5418	4.79E-01	3.79E-01	7.20E-01	1.13E-04	8.86E-06
RRP9	1.133993	4.78E-22	1.02E-23	8.83E-01	8.46E-04	1.07E-04
SRM	0.505448	5.91E-04	1.68E-04	5.35E-01	2.35E-02	6.40E-03
TBRG4	0.086785	5.19E-01	4.19E-01	5.86E-01	1.02E-02	2.29E-03
AIMP2	0.02054	8.51E-01	7.98E-01	2.13E-01	4.45E-02	3.26E-02
CDK4	1.709589	4.86E-38	2.14E-40	1.62	1.11E-07	1.21E-09
GRWD1	0.49566	7.59E-10	6.82E-09	2.92E-02	8.06E-01	7.09E-01
MCM5	1.056463	8.70E-12	1.02E-10	7.44E-01	1.79E-02	4.58E-03

AIMP2, and *GRWD1* exhibited significant overexpression in one of the studies. These results are comprehensively presented in Table 3. Moreover, we employed the clinical data from the GSE39582 study for survival analysis of the candidate genes, as outlined in the materials and methods. Our analysis revealed a significant association between elevated levels of *MCM5*, *IMP4*, and *RABEPK* and poor prognosis in patients (Supplementary file: Figure S1, P-value < 0.05).

Elevated the levels of *IMP4* and *RABEPK* in CRC

Among OS-related genes, the AUC of *IMP4* and *RABEPK* were higher than other candidates, so these genes were chosen for experimental validation. The expression levels of *IMP4* and *RABEPK* mRNA in CRC were assessed using qRT-PCR in a cohort comprising 30 healthy tissues and 30 matched CRC samples. Our findings indicated a notable increase in the expression of both *IMP4* and *RABEPK* in CRC samples when compared to the corresponding adjacent healthy tissue (Fig. 8C and D, $P < 0.01$). In general, our results showed that the genes related to RPs could play an important role in CRC, especially *IMP4* and *RABEPK*, and could be suitable diagnostic, therapeutic and prognostic targets in CRC.

Discussion

The CRC poses a global concern, given its elevated mortality and morbidity rates. Across the world, medical systems have made concerted efforts to diminish the incidence of CRC through the implementation of innovative diagnostic and prognostic methods [14]. Nevertheless, it remains a prominent medical burden globally [15]. Identifying genes that correlate with key genes and pathways in CRC could pave the way for further insights and advancements in understanding the intricacies of the disease [16].

The pathogenesis of ribosomopathies, a heterogeneous group of diseases, stems from abnormalities in both the biogenesis and functioning of ribosomes [17, 18]. The identification of somatic mutations in RP genes in hematological cancers including T-cell acute lymphoblastic leukemia, chronic lymphocytic leukemia, and multiple myeloma, as well as in solid tumors such as breast cancer and melanoma, underscores the significance of ribosome biogenesis defects in potentially fostering oncogenic transformation [17, 19]. Unlike ribosomal defects associated with RP mutations in cancers, oncogene-driven cancers exhibit an increased activation of ribosome biogenesis, fostering uncontrolled growth, proliferation, and altered metabolism. Accumulating studies have focused on mechanisms of oncogenic potential RPs in tumor progression and cancer developing [20, 21]. Reports indicate that RPs are closely related to well-known genes and their associated pathways. Nucleolar stress can induce p53 stabilization through various mechanisms, encompassing post-translational modifications, protein-protein interactions, and an elevation in the translation rate of p53 mRNA. The process of p53 stabilization in response to nucleolar stress relies on the involvement of RPs [22–24]. In addition, there is demonstrated overexpression of *RPL31* in prostate cancer (PCa) tissues, impacting the levels of the tumor suppressor p53 and its associated targets, specifically the cell-cycle negative regulator p21 and the E3 ubiquitin ligase MDM2, which targets p53 [25]. Moreover, it has been reported that rise in expressions of some RPs are relevant to the activation of NFκB, Akt-mTOR-p70S6K signaling pathway that promotes neo angiogenesis in breast cancer and hepatocellular carcinoma [26, 27]. Therefore, these results support the idea that RPs could play a vital role in the progression of cancer. The identification of genes associated with RPs could open up new avenues in controlling the development of cancer cells.

The idea of recognizing the genes that are related to RPs could enable us to investigate their specific functions and interactions, providing valuable insights into the intricate mechanisms underlying cancer progression. So, through the analysis of TCGA-COAD transcriptome and the implementation of specific correlation clustering, genes related to RPs were discovered. The association of these genes with prognostic outcomes and drug resistance/sensitivity was subsequently evaluated. Our results indicated that the expression levels of mRNA and protein of *RPL35, RPL36A, RPS17, RPS4X*, and *RPS19* were up-regulated in tumor samples compared to the normal samples. Intriguingly, recent results also confirm that RPL35 and RPS19 were up-regulated in colon, breast cancer [28–30]. As we have mentioned above, RPs could act as key genes in some pathways. Remarkably, certain genes related to RPs were found to be enriched in pathways shared with the enriched RPs pathways. Additionally, other genes were enriched in critical cancer progression pathways, such as Myc Targets V2, which play a role in both the initiation and maintenance of tumorigenesis. Our findings indicate that the expression levels of all 12 RP-related genes, including *PUS1, PLK1, IMP4, NOC4L, RABEPK*, *RRP9, SRM*, *TBRG4, AIMP2, CDK4, GRWD1*, and *MCM5*, were up-regulated in CRC tumor samples compared to normal samples. Similar patterns of up-regulation have been observed in other cancers, as demonstrated in prior studies. For instance, Chenlu Lan et al. revealed elevated levels of PUS1 in Hepatocellular Carcinoma, and intriguingly, this increase was associated with a poor prognosis [31]. Moreover, *GRWD1* has been indicated that to be overexpressed in colon samples compare to normal samples [32]. Within the selected genes, elevated levels of MCM5, IMP4, and

RABEPK were notably linked to a poor prognosis in CRC patients. M. Burger et al. have reported an association between elevated levels of MCM5 and an unfavorable prognosis in CRC patients [33]. Notably, *IMP4* and *RABEPK* demonstrated the ability to distinguish CRC samples from normal samples with AUC exceeding 90%. These findings suggest that *IMP4* and *RABEPK* could serve as promising biomarkers for colorectal cancer.

IMP U3 small nucleolar ribonucleoprotein 4 (IMP4) functions as a constituent of U3 small nucleolar ribonucleoproteins and participates in the maturation process of 18 S ribosomal RNA. Furthermore, there are reports indicating that IMP4 functions as a novel telomeric DNA-binding protein, suggesting its potential significant roles in telomere-related processes [33]. Our findings revealed overexpression of *IMP4* in tumor samples, indicating its potential involvement in cancer-related pathways. Furthermore, it has been demonstrated that *IMP4* is up-regulated in lung cancer, and this overexpression is associated with promoting the progression of Lung adenocarcinoma by activating the ERK pathway [34]. Our results displayed that the levels of *IMP4* was up-regulated in CRC samples compared to normal samples. interestingly, our outcomes showed that the high level of *IMP4* was associated with poor prognostic of CRC patients. Moreover, our experimental results confirm the elevated levels of IMP4 in CRC compared to normal samples significantly. So, these results suggest that IMP4 could have a significant influence on the progression of CRC either directly or in conjunction with RPs.

RAB9 belongs to the Rab GTPase family, a group of proteins primarily situated on various organelle membranes. Working in conjunction with their specific effectors, members of this family play a central role in regulating vesicle-mediated transport. Furthermore, RAB9 holds significance in lysosomal biogenesis and the morphology of late endosomes. Additionally, it functions as a cellular target for specific pathogens, contributing to its pivotal role in pathogenic infections [35–37]. The protein known as p40, alternatively referred to as Rab9 effector protein with Kelch motifs, is encoded by the *RABEPK* gene in humans [38]. Our findings indicated an up-regulation of its levels in tumor samples compared to normal samples. Furthermore, our research unveiled an association between elevated levels of these genes and a poor prognosis in CRC samples.

Our results showed that the Alprostadil could target *CDK4, IMP4, RABEPK*, and *RRP9* in ovarian cell line and effect on some genes. Alprostadil is a pharmaceutical employed in the treatment and control of erectile dysfunction in males. It is also utilized for the transient maintenance of ductus arteriosus in newborns diagnosed with congenital heart diseases prior to undergoing surgical intervention [39, 40]. In rabbit corneas, PGE1(Alprostadil) demonstrated angiogenic properties according to Ziche et al. (1982), and it also promoted the formation of new capillaries in rat femoral veins [41, 42]. There is indirect evidence implicating prostaglandins 1 and 2 in angiogenesis, as inhibitors of cyclooxygenase (COX) have been demonstrated to hinder tumor growth, potentially through an anti-angiogenic mechanism [43]. Maria Grazia Cattaneo findings suggest that the impact of PGE1 is anti-angiogenic. This is supported by the inhibition of endothelial cell proliferation, migration, and cord formation in vitro, as well as the suppression of Matrigel vascularization in vivo [44]. Hence, our findings propose that Alprostadil could be deemed as a potentially effective treatment for CRC, notwithstanding the need for diverse experimental confirmations.

The combination of chemotherapy drugs with autophagy inhibitors represents an innovative and potent strategy in the field of cancer treatment. ambroxol has been utilized for the treatment of both acute and chronic bronchitis, as well as bronchial asthma, demonstrating a broad therapeutic window and minimal side effects even at high doses. In addition, ambroxol treatment has been observed to activate the coordinated lysosomal expression and regulation network through transcription factors such as EB. It serves as a molecular chaperone, modulating lysosomal biochemistry and thereby enhancing the well-being of patients with Gaucher disease [45, 46]. Administering elevated doses of ambroxol appears to yield superior clinical outcomes compared to standard doses in preventing postoperative complications among lung cancer patients [47]. ambroxol treatment resulted in a disruption of autophagic flux and hindered the degradation of intracellular cargo. This indicates that the accumulation of autophagosomes stemmed from a reduction in autophagosome turnover rather than an increase in autophagosome formation [48]. So, these outcomes showed that ambroxol could have deep influence on the lung cancer. Our results showed that ambroxol could target our selected genes and may decrease the levels of them in CRC. Ambroxol serves as a powerful inhibitor of neuronal sodium channels and, concurrently, functions as an anti-inflammatory medication [46]. A recent investigation indicated that Ambroxol may exert a beneficial effect on the inflamed intestinal epithelium in CRC [49]. Hence, our findings propose that Ambroxol might influence the progression of CRC by inhibiting unidentified pathways, and it could be employed as an effective treatment for CRC, potentially in combination with other drugs.

Conclusion

In summary, the bioinformatics approach employed in the present study unveiled the significant prognostic roles of *IMP4* and *RABEPK* as genes related to ribosomal proteins in the process of CRC development. These genes play a crucial role and emerge as potential biomarkers for CRC detection and promising therapeutic targets for the treatment of this cancer. Finally, the high levels of these genes were evaluated in 30 CRC samples compared to the 30 normal samples via using qRT-PCR.

Abbreviations

CRC	Colorectal cancer
TCGA	The cancer genome atlas
PPI	Protein-protein interaction
DEG	The differential expression gene
RPs	Ribosomal proteins
ROC	Receiver operating characteristic
AUC	The area under the curve
HR	Hazard ratio
CCLE	Cancer cell line encyclopedia
GDSC	Genomics of drug sensitivity in cancer
MCC	Maximal
MSigDB	Molecular Signatures Database, clique centrality
TMM	Trimmed mean of M values
WGCNA	Weighted gene co-expression network analysis algorithm
GO	Gene Ontology

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11033-024-09522-5.

Acknowledgements Thanks to Genius Gene center for assisting in laboratory work and data analysis.

Author contributions The study design was performed by N.S., A.Z. and M.M.Data analysis and experiments were done by N.S., A.K.M.A and D.K.AInterpretations of data were performed by N.S., A.Z., E.S.S and M.M.Bioinformatics analysis was performed by N.S., A.K.M.A., D.K.A and M.M. Manuscript writing was performed by N.S.The manuscript was finally approved by M.M and A.Z.

Funding Not available.

Data availability No datasets were generated or analysed during the current study.
Supporting and raw data are available upon a reasonable request to the corresponding author.

Declarations

Ethical approval and consent to participate All bioethical issues were reviewed and confirmed by the review board of Emam Khomeini hospital according to the criteria of the ministry of health and medical education of Iran. Written consent forms were acquired from all individuals participating in the study. Also, ethical considerations were approved by the committee related to access number IR.UT.SCIENCE.REC.1401.007.

Consent for publication All authors support the submission of the present manuscript to this journal.

Competing interests The authors declare no competing interests.

References

1. Siegel RL, Miller KD, Wagle NS, Jemal A (2023) Cancer statistics, 2023. Ca Cancer J Clin 73(1):17–48
2. Dekker E, Tanis PJ, Vleugels J, Kasi PM, Wallace M (2019) Pure-AMC. Lancet 394:1467–1480
3. Cmarko D, Smigova J, Minichova L, Popov A, Nucleolus (2008) The ribosome factory. Histol Histopathol
4. Simsek D, Barna M (2017) An emerging role for the ribosome as a nexus for post-translational modifications. Curr Opin Cell Biol 45:92–101
5. Zhou X, Liao W-J, Liao J-M, Liao P, Lu H (2015) Ribosomal proteins: functions beyond the ribosome. J Mol Cell Biol 7(2):92–104
6. Penzo M, Montanaro L, Treré D, Derenzini M (2019) The ribosome biogenesis—cancer connection. Cells 8(1):55
7. Goudarzi KM, LINDSTRöM MS (2016) Role of ribosomal protein mutations in tumor development. Int J Oncol 48(4):1313–1324
8. Luo M-J, Lai M-D (2001) Identification of differentially expressed genes in normal mucosa, adenoma and adenocarcinoma of colon by SSH. World J Gastroenterol 7(5):726
9. Farley-Barnes KI, Ogawa LM, Baserga SJ (2019) Ribosomopathies: old concepts, new controversies. Trends Genet 35(10):754–767
10. Yoshihama M, Uechi T, Asakawa S, Kawasaki K, Kato S, Higa S et al (2002) The human ribosomal protein genes: sequencing and comparative analysis of 73 genes. Genome Res 12(3):379–390
11. Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M et al (2016) The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible. Nucleic Acids Res. :gkw937
12. Yao M, Zhang C, Gao C, Wang Q, Dai M, Yue R et al (2021) Exploration of the shared gene signatures and molecular mechanisms between systemic lupus erythematosus and pulmonary arterial hypertension: evidence from transcriptome data. Front Immunol 12:658341
13. Deng M, Yin Y, Zhang Q, Zhou X, Hou G (2021) Identification of inflammation-related biomarker Lp-PLA2 for patients with COPD by comprehensive analysis. Front Immunol 12:670971
14. Siegel RL, Miller KD, Jemal A (2015) Cancer statistics, 2015. Cancer J Clin 65(1):5–29
15. Schmoll H, Van Cutsem E, Stein A, Valentini V, Glimelius B, Haustermans K et al (2012) ESMO Consensus Guidelines for management of patients with colon and rectal cancer. A personalized approach to clinical decision making. Ann Oncol 23(10):2479–2516
16. Chodary Khameneh S, Razi S, Shamdani S, Uzan G, Naserian S (2022) Weighted correlation network analysis revealed novel long non-coding RNAs for colorectal cancer. Sci Rep 12(1):2990
17. Kampen KR, Sulima SO, Vereecke S, De Keersmaecker K (2020) Hallmarks of ribosomopathies. Nucleic Acids Res 48(3):1013–1028

18. Sulima SO, Kampen KR, De Keersmaecker K (2019) Cancer biogenesis in ribosomopathies. Cells 8(3):229
19. Aspesi A, Ellis SR (2019) Rare ribosomopathies: insights into mechanisms of cancer. Nat Rev Cancer 19(4):228–238
20. Bywater MJ, Pearson RB, McArthur GA, Hannan RD (2013) Dysregulation of the basal RNA polymerase transcription apparatus in cancer. Nat Rev Cancer 13(5):299–314
21. Montanaro L, Treré D, Derenzini M (2013) The emerging role of RNA polymerase I transcription machinery in human malignancy: a clinical perspective. OncoTargets Therapy. :909–916
22. Pan W, Issaq S, Zhang Y (2011) The in vivo role of the RP-Mdm2-p53 pathway in signaling oncogenic stress induced by pRb inactivation and ras overexpression. PLoS ONE 6(6):e21625
23. Macias E, Jin A, Deisenroth C, Bhat K, Mao H, Lindström MS et al (2010) An ARF-independent c-MYC-activated tumor suppression pathway mediated by ribosomal protein-Mdm2 Interaction. Cancer Cell 18(3):231–243
24. Motlagh AV, Mahdevar M, Mirzaei S, Entezari M, Hashemi M, Hushmandi K et al (2022) Introduction of mutant TP53 related genes in metabolic pathways and evaluation their correlation with immune cells, drug resistance and sensitivity. Life Sci 303:120650
25. Maruyama Y, Miyazaki T, Ikeda K, Okumura T, Sato W, Horie-Inoue K et al (2014) Short hairpin RNA library-based functional screening identified ribosomal protein L31 that modulates prostate cancer cell growth via p53 pathway. PLoS ONE 9(10):e108743
26. Li W, Tan D, Zhang Z, Liang JJ, Brown RE (2008) Activation of Akt-mTOR-p70S6K pathway in angiogenesis in hepatocellular carcinoma. Oncol Rep 20(4):713–719
27. Stehlik C, De Martin R, Kumabashiri I, Schmid JA, Binder BR, Lipp J (1998) Nuclear factor (NF)-κB–regulated X-chromosome–linked iap gene expression protects endothelial cells from tumor necrosis factor α–induced apoptosis. J Exp Med 188(1):211–216
28. Lau TP, Roslani AC, Lian LH, Chai HC, Lee PC, Hilmi I et al (2014) Pair-wise comparison analysis of differential expression of mRNAs in early and advanced stage primary colorectal adenocarcinomas. BMJ open 4(8):e004930
29. Kondoh N, Schweinfest CW, Henderson KW, Papas TS (1992) Differential expression of S19 ribosomal protein, laminin-binding protein, and human lymphocyte antigen class I messenger RNAs associated with colon carcinoma progression and differentiation. Cancer Res 52(4):791–796
30. Markiewski MM, Vadrevu SK, Sharma SK, Chintala NK, Ghouse S, Cho J-H et al (2017) The ribosomal protein S19 suppresses antitumor immune responses via the complement C5a receptor 1. J Immunol 198(7):2989–2999
31. Lan C, Huang X, Liao X, Zhou X, Peng K, Wei Y et al (2023) PUS1 May Be a Potential Prognostic Biomarker and Therapeutic Target for Hepatocellular Carcinoma. Pharmacogenomics and Personalized Medicine. :337 – 55
32. Zhou X, Shang J, Liu X, Zhuang J-F, Yang Y-F, Zhang Y-Y et al (2021) Clinical significance and oncogenic activity of GRWD1 overexpression in the development of colon carcinoma. OncoTargets Therapy. :1565–1580
33. Burger M (2009) MCM2 and MCM5 as prognostic markers in colon cancer: a worthwhile approach. Springer, pp 197–198
34. Li R, Han Z, Ma W, Zhang L, Zhang X, Jiang Y et al (2022) IMP4 Silencing Inhibits the Malignancy of Lung Adenocarcinoma via ERK Pathway. Journal of Oncology. ;2022
35. Barr F, Lambright DG (2010) Rab gefs and gaps. Curr Opin Cell Biol 22(4):461–470
36. Stenmark H (2009) Rab GTPases as coordinators of vesicle traffic. Nat Rev Mol Cell Biol 10(8):513–525
37. McGourty K, Thurston TL, Matthews SA, Pinaud L, Mota LJ, Holden DW (2012) Salmonella inhibits retrograde trafficking of mannose-6-phosphate receptors and lysosome function. Science 338(6109):963–967
38. Díaz E, Schimmöller F, Pfeffer SR (1997) A novel Rab9 effector required for endosome-to-TGN transport. J Cell Biol 138(2):283–290
39. Vieillard V, Eychenne N, Astier A, Yiou R, Deffaux C, Paul M (eds) (2013) Physicochemical stability study of a new Trimix formulation for treatment of erectile dysfunction. Annales pharmaceutiques françaises. Elsevier
40. Coceani F, Olley P, Bishai I, Bodach E, Heaton J, Nashat M et al (1977) Prostaglandins and the control of muscle tone in the ductus arteriosus. Tissue Hypoxia Ischemia. :135–142
41. Diaz-Flores L, Gutierrez R, Valladares F, Varela H, Perez M (1994) Intense vascular sprouting from rat femoral vein induced by prostaglandins E1 and E2. Anat Rec 238(1):68–76
42. Ziche M, Jones J, Gullino PM (1982) Role of prostaglandin E1 and copper in angiogenesis. J Natl Cancer Inst 69(2):475–482
43. Masferrer JL, Leahy KM, Koki AT, Zweifel BS, Settle SL, Woerner BM et al (2000) Antiangiogenic and antitumor activities of cyclooxygenase-2 inhibitors. Cancer Res 60(5):1306–1311
44. Cattaneo MG, Pola S, Dehò V, Sanguini AM, Vicentini LM (2003) Alprostadil suppresses angiogenesis in vitro and in vivo in the murine Matrigel plug assay. Br J Pharmacol 138(2):377
45. Baranwal AK, Murthy AS, Singhi SC (2015) High-dose oral ambroxol for early treatment of pulmonary acute respiratory distress syndrome: an exploratory, randomized, controlled pilot trial. J Trop Pediatr 61(5):339–350
46. McNeill A, Magalhaes J, Shen C, Chau K-Y, Hughes D, Mehta A et al (2014) Ambroxol improves lysosomal biochemistry in glucocerebrosidase mutation-linked Parkinson disease cells. Brain 137(5):1481–1495
47. Wang X, Wang L, Wang H, Zhang H (2015) Perioperative lung protection provided by high-dose ambroxol in patients with lung cancer. Cell Biochem Biophys 73:281–284
48. Zhang X, Chen Q, Chen M, Ren X, Wang X, Qian J et al (2017) Ambroxol enhances anti-cancer effect of microtubule-stabilizing drug to lung carcinoma through blocking autophagic flux in lysosome-dependent way. Am J Cancer Res 7(12):2406
49. Mastrogamvraki N, Zaravinos A (2020) Signatures of co-deregulated genes and their transcriptional regulators in colorectal cancer. NPJ Syst Biology Appl 6(1):23

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.