scientific reports

OPEN

Identification of expression profiles and prognostic value of RFCs in colorectal cancer

Md Misbah [1,4,5]✉, Manoj Kumar [2], Abul Kalam Najmi [3] & Mymoona Akhtar [4,6]✉

Colorectal cancer (CRC) ranks among the most prevalent cancers globally, with its incidence closely tied to DNA damage. The Replication Factor C (RFC) complexes comprises five protein subunits: RFC1, RFC2, RFC3, RFC4, and RFC5. These RFC complexes play crucial roles in DNA replication, repair pathways, activities post DNA damage, and ATP-dependent processes during DNA synthesis. However, the impact of RFC complexes proteins on CRC prognosis remains unclear. To explore this, we employed a computational analysis approach, utilizing platforms such as the DepMap portal, GEPIA, DAVID Bioinformatics for KEGG pathway analysis, Human Protein Atlas (HPA), STRING, and TIMER. Our results indicate that the mRNA levels of RFC1 and RFC5 were the least expressed among CRC cell lines compared to other RFC complex subunits. Notably, low RFC1 and RFC5 expression was correlated with poor prognosis in terms of CRC patients' overall survival (OS). Immunohistochemical results from the Human Protein Atlas demonstrated medium staining for RFC1, RFC2, and RFC5 in CRC tissues. Furthermore, the low expression of RFC1 and RFC5 showed a significant correlation with high expression levels of miR-26a-5p and miR-636, impacting cell proliferation through mismatch repair, DNA replication, and the nucleotide excision repair pathway. Although the precise functions of RFC1 in cancer are still unknown, our findings suggest that the small-molecule single target, CHEMBL430483, and multiple target molecules could be potential treatments for CRC. In conclusion, the elevated expression of miR-26a-5p and miR-636 targeting RFC1 and RFC5 expression holds promise as a potential biomarker for early-stage CRC detection. These insights provide novel directions and strategies for CRC therapies.

Abbreviations

RFCs	Replication factor C
CRC	Colorectal cancer
CCLE	Cancer cell line encyclopedia
COAD	Colon adenocarcinoma
READ	Rectal adenocarcinoma
GEPIA	Gene expression profiling interactive analysis
TIMER	Tumor immune estimation resources
OS	Overall survival
STRING	Search tool for the retrieval of interacting genes/proteins

Colorectal cancer (CRC) is the world's third most common epithelial malignancy. According to the GLOBOCON database, around 147,950 individuals were diagnosed with CRC, and 53,200 patients succumbed to CRC in the United States in 2020[1]. Surgery is the primary intervention for CRC patients in the early stage of diagnosis; however, it proves ineffective in metastatic cases. About 20% of patients exhibit micro-metastasis or metastatic (m)CRC post-surgery[2,3]. Biomarkers like mitogen-activated protein kinase 1 (MAPK1), phosphatidylinositol 3-kinase (PI3K), and others are presently employed for CRC treatment[4,5]. An increasing number of biomarkers

[1]International Ph.D. Program in Medicine, College of Medicine, Taipei Medical University, Taipei 110, Taiwan. [2]Centre for Translational and Clinical Research, School of Chemical and Life Sciences, Jamia Hamdard, New Delhi 110062, India. [3]Department of Pharmacology, School of Pharmaceutical Education and Research, Jamia Hamdard, New Delhi, India. [4]Bioinformatics Infrastructure Facility, Jamia Hamdard, New Delhi, India. [5]Kusumraj Institute of Pharmacy, Bikram, Patna, Bihar, India 801104. [6]Drug Design and Medicinal Chemistry Lab, Department of Pharmaceutical Chemistry, School of Pharmaceutical Education and Research, Jamia Hamdard, New Delhi, India. ✉email: d142109019@tmu.edu.tw; makhtar@jamiahamdard.ac.in

are being utilized for the diagnosis and selection of therapy for CRC patients[6]. Hence, identifying valuable biomarkers for patient identification is urgently needed in clinical practice.

The replication factor C (RFC) complexes function as clamp loaders responsible for loading and unloading proliferating cell nuclear antigen (PCNA) onto DNA, and play a role in various DNA replication and repair pathways, activities following DNA damage, and are also engaged in an ATP-dependent process during DNA synthesis[7–11]. The RFC family consists of five protein subunits, *RFC1, RFC2, RFC3, RFC4,* and *RFC5*[12]; which are strongly connected with tumor growth and metastasis[13].

The Gene Expression Profiling Interactive Analysis (GEPIA), Human Protein Atlas (HPA) database, The Cancer Genome Atlas (TCGA) database, and Tumor Immune Estimation Resource (TIMER) were used for determining the gene expression level of RFCs[14]. Furthermore, protein–protein-interaction (PPI) and Kyoto Encyclopedia Genes and Genomes (KEGG) analysis aided in identifying the pathway for prognostic markers in CRC patients[15,16].

MicroRNAs, short non-coding RNAs, can alter oncogene factors and mechanisms, by binding to the 3′-untranslated region (UTR) of their target messenger (m)RNAs leading to translational suppression. Onco-suppressor miRNAs can activate apoptosis, induce cell cycle arrest, impede cell viability, and tumor progression[17]. However, some previous studies identified, upregulated miR-26a-5p was shown to be a tumor suppressor oncomir in CRC[18] and miR-636 was also identified onco-suppressor in many solid tumors[19] The objective of this study was to identify treatment response biomarkers and comprehend the processes of the RFC family in colorectal cancer. In essence, the study aimed to discover microarray analysis predicting RFC's complex genes and underlying pathways in CRC patients.

Results

The expression level of RFCs in CRC using CCLE and GEPIA2

A schematic diagram showing a summary of all methodological techniques in a schematic diagram (Fig. 1a). The mRNA expression levels of RFC family members in CRC cell lines are presented through CCLE analysis (Fig. 1b). The heat map illustrates high mRNA expression of RFCs in certain cell lines, denoted by red and green colors. Additionally, we assessed RFC1, RFC2, RFC3, RFC4, and RFC5 expression in CRC using the GEPIA web tool analysis software. According to GEPIA analysis (Fig. S1 and Fig. 2), RFC2, RFC3, RFC4, and RFC5 displayed significant and distinct expressions in CRC. The boxplot comparing normal and colorectal adeno-carcinoma indicated significant differences in RFC expression in CRC (Figs. S1A,B, Fig. 2A), except for RFC1, which showed no significant difference.

We further investigated the association between the expression levels of selected RFC genes and clinico-pathological parameters. Figure 2B reveals that mRNA expression levels of RFC5 significantly differ across different tumor stages of CRC. However, the expression levels of RFC1, RFC2, RFC3, and RFC4 did not exhibit significant differences in various tumor stages. The RFC5 result aligns with a prior study indicating its relevance to cancer progression[20].

Genes associated with colorectal cancer patient's survival and KEGG pathway analysis

Genes associated with colorectal cancer patient survival were analyzed. The GEPIA online web tool was utilized for survival analysis, revealing that low RFC1 and RFC5 levels were associated with a poor prognosis for CRC patients compared to RFC2, RFC3, and RFC4, as depicted in Fig. 3A.

The five RFC genes were subjected to KEGG pathway analysis using the DAVID online tool. Three KEGG pathways, namely mismatch repair, DNA replication, and nucleotide excision repair, were identified, and all reached statistical significance (FDR value < 0.25 and P-value < 0.05) for RFC1, RFC2, RFC3, RFC4, and RFC5 (Table 1). The expression pattern of candidates for mismatch repair, DNA replication, and nucleotide excision repair signaling are shown in Fig. 3B, where it shows that the five genes are downregulated.

The protein expression level of RFCs in CRC

The protein expression levels of five RFCs genes in 12 tissue samples from colorectal cancer patients were validated using the database Human Protein Atlas (HPA). IHC (immunohistochemistry) images of immunoreactivity expression in cancer specimens were examined (Fig. 4), and the staining intensity was manually scored.

The IHC image scoring categorized staining intensity as negative, weak, moderate, or strong based on the proportion of stained cells (> 75 percent, 25–75 percent, or < 25 percent, respectively), determining the protein expression score.

In colorectal cancer tissues, the staining levels of *RFC1*, *RFC2*, *RFC4*, and *RFC5* protein expression were modest. Overall, *RFC3* exhibited weak intensity while *RFC1* and *RFC5* displayed moderate intensity with > 75% quantity. In contrast, RFC2 and RFC4 demonstrated moderate intensity with a quantity ranging from 25 to 75%. Given that RFC1 and RFC5 exhibited intensity levels greater than 75%, they emerged as more promising biomarkers for colorectal cancer development.

Identifying the potential miRNA target for the candidate's genes

To identify potential upstream regulators for those miRNAs, we employed MirWalk, webtools, and GSE29623 (CRC miRNA's) for miRNA prediction. The analysis revealed that 32 miRNAs were predicted to target RFC1, while 20 miRNAs were predicted to target RFC5 (Fig. 5A,B). Subsequently, we narrowed down the selection to 20 miRNAs for further analysis (Table S1).
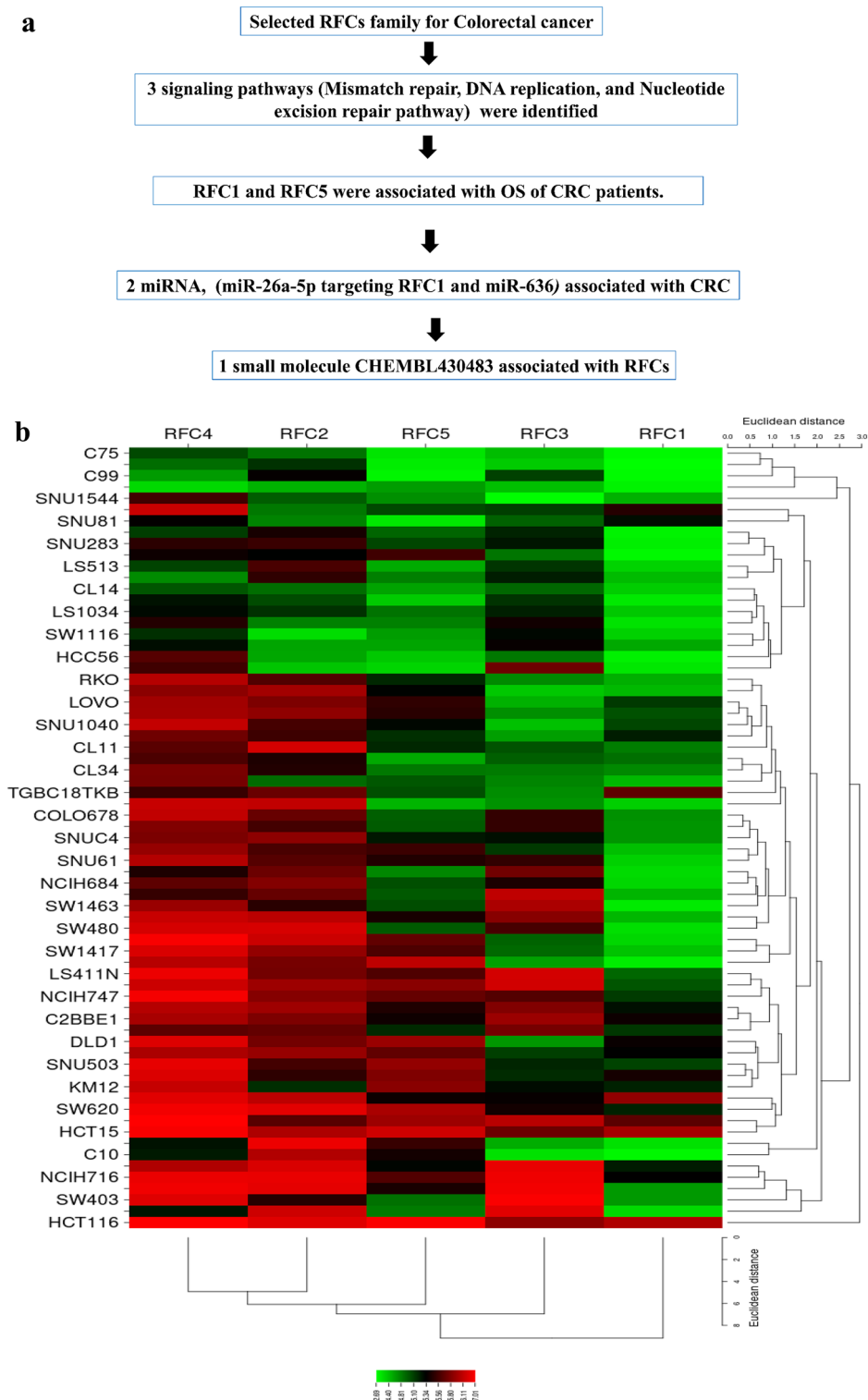
Figure 1. (**a**) Schematic diagram summarizing the study. (**b**) Heat map showing replication factor C subunits (RFC) gene mRNA expression in CRC cell lines (CCLEs) with red color signifying-overexpression, green color suggesting underexpression, and black colors show no expression.

miRNAs associated with survival of patients with colorectal cancer

The overall survival analysis of targeted miRNA expression levels of CRC tissue samples was analyzed and validated using the GSE29623 database. A Kaplan–Meir plot was generated using statistical software SPSS version
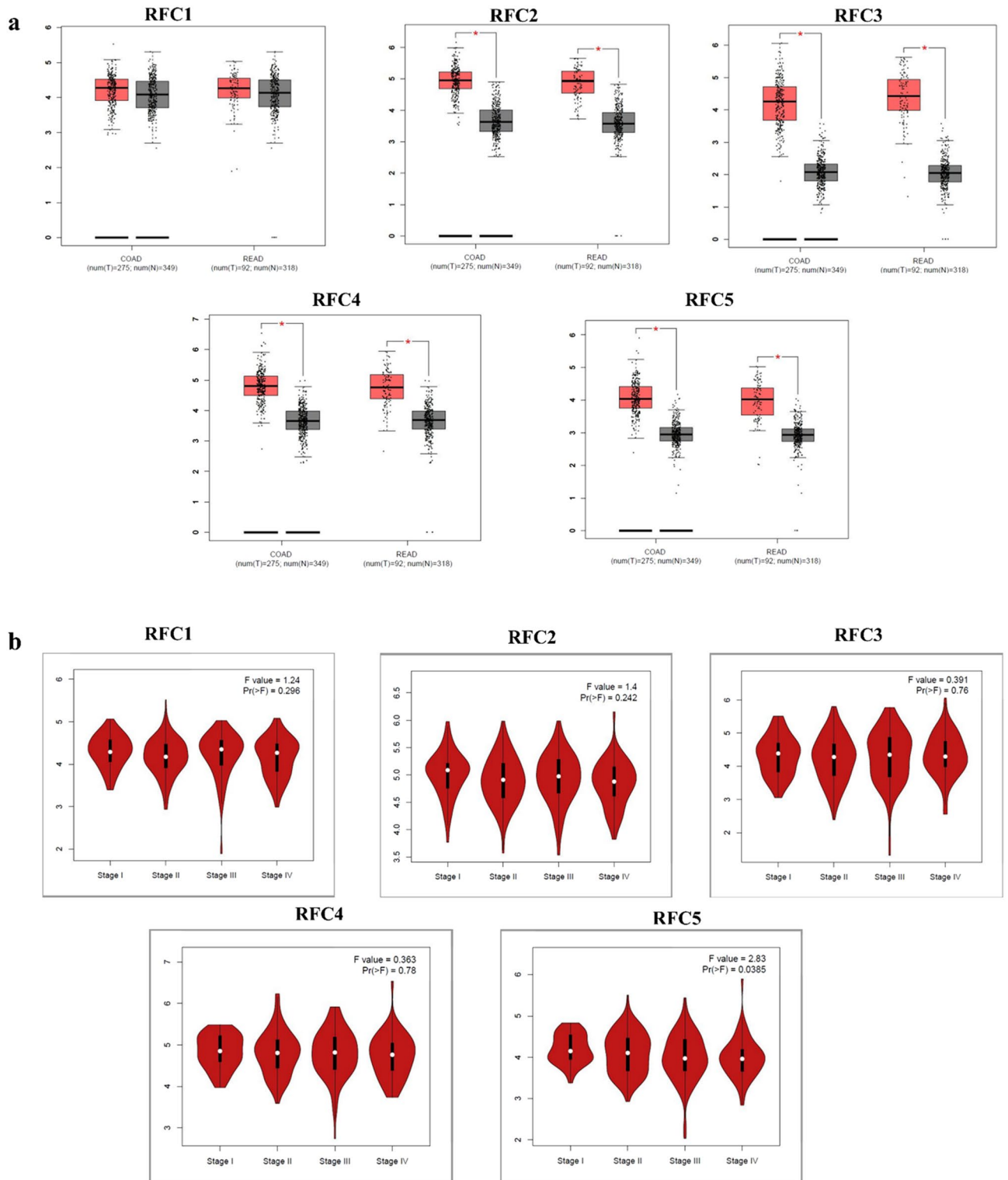
Figure 2. Gene expression profiles of (**a**) *RFC1*, *RFC2*, *RFC3*, *RFC4*, and *RFC5* in colorectal cancer (CRC). Boxplot showing transcriptional levels of Replication Factor C subunits gene (RFCs) in colon adenocarcinoma (COAD) ($n = 275$) vs. normal samples ($n = 349$) and rectal adenocarcinoma (READ) ($n = 92$) vs. normal tissues ($n = 318$) using the GEPIA web tool based on TCGA database. Black colors show transcriptional levels in normal tissues, while red colors show DEG transcriptional levels in COAD and READ tissues. A one-way ANOVA was used for the differential analysis with a statistically significant value of $P < 0.05$. (**b**) All stages of CRC are shown for cancer progression of the five RFCs. A violin plot shows different stages of cancer with log2 (transcripts per million (TPM) + 1) of genes in stages I to IV. A *t*-test was used with the statistically significant $p < 0.05$. The Pr($> F$) < 0.05, followed by Student's *t*-test.
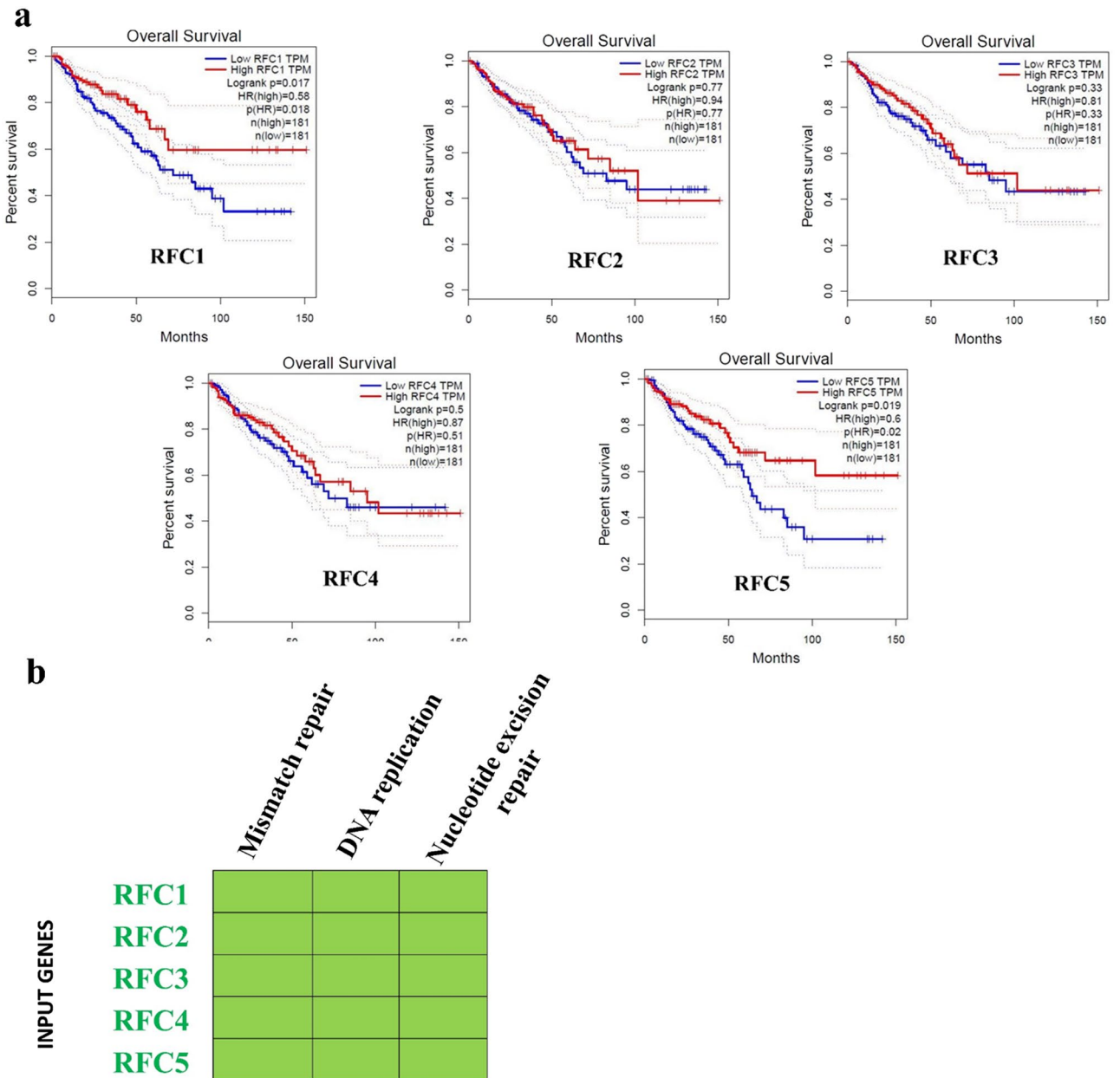
Figure. 3. (**a**) Kaplan–Meier survival curves presenting the prognostic relationship between high and low expressions of replication factor C subunits (RFC) genes to overall survival (OS) in CRC patients using the GEPIA (TCGA) database patient samples. Survival curves were plotted using GEPIA online. Specific replication factor C subunits gene (RFC) expression levels were selected by the median value. Results are visually presented by Kaplan–Meier survival plots, and p values were calculated using log-rank statistics. Patient number (n) = 362, p = log-rank p-value with $p < 0.05$ considered significant. (**b**) Significant KEGG pathways and genes are involved. Gene enrichment analysis shows that KEGG pathways were significantly enriched in David pathway online analysis and genes involved in the pathways (the pathways are in order of their enrichment from left to right), $FDR < 0.05$.

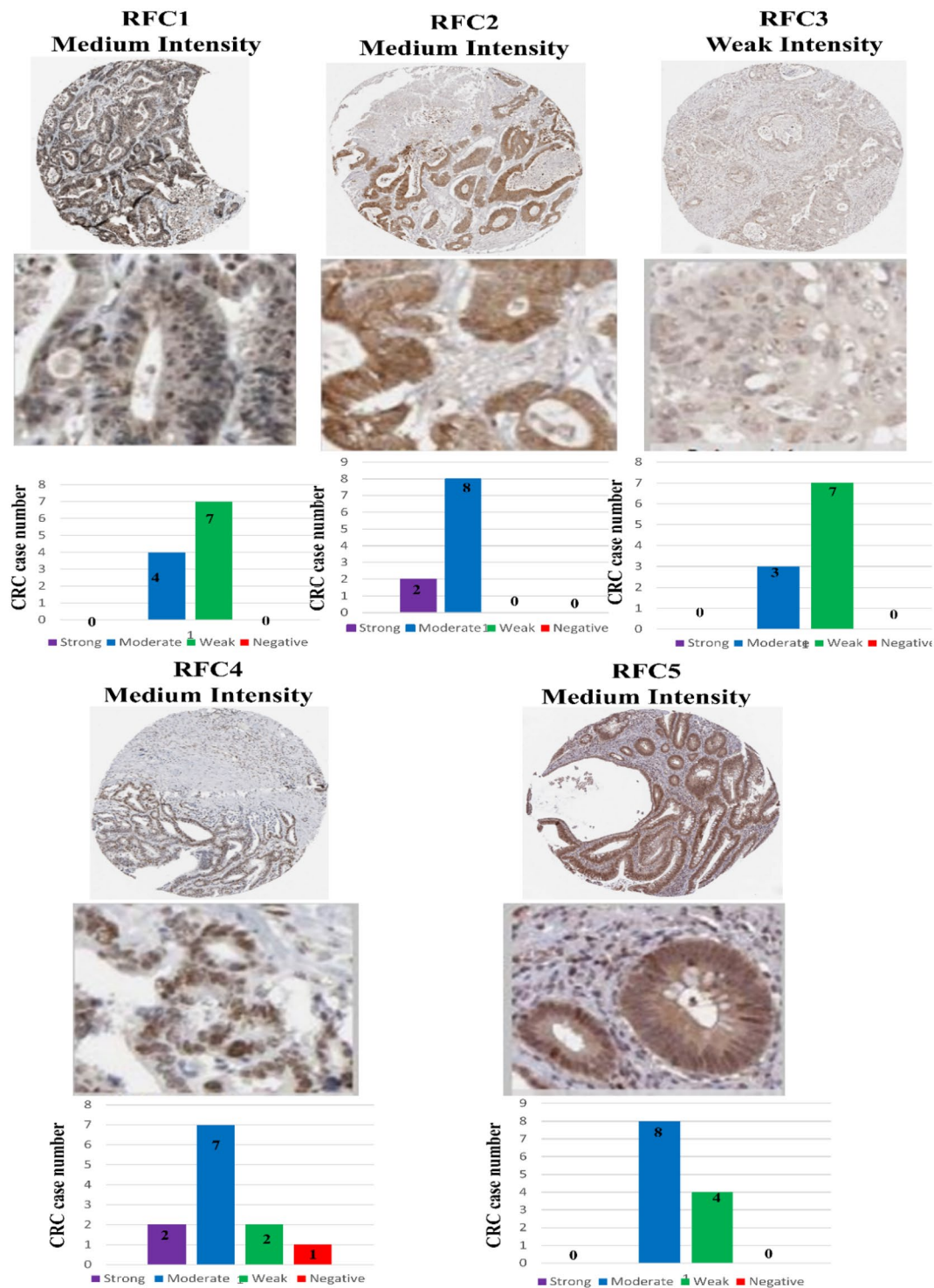| KEGG pathway	Count	P-value	Genes	FDR
hsa03430:Mismatch repair	5	9.50E−11	*RFC5, RFC3, RFC4, RFC1, RFC2*	1.90E−10
hsa03030:DNA replication	5	6.32E−10	*RFC5, RFC3, RFC4, RFC1, RFC2*	6.32E−10
hsa03420:Nucleotide excision repair	5	1.91E−09	*RFC5, RFC3, RFC4, RFC1, RFC2*	1.91E−09

Table 1. Enriched KEGG pathways.

Figure 4. Protein expressions of five genes. The protein expression analysis used the HPA database of colorectal cancer (CRC) tissue samples. IHC images show the intensity and staining of replication factor C subunit genes (RFCs). Manual scoring of IHC data for staining intensity (negative, weak, moderate, or strong) and proportion of stained cells (>75%, 25%~75%, or 25%) as determined by the protein expression score.

22.0. The correlation of four miRNAs with the clinical outcome of CRC was further validated as shown in Fig. 5C and D. High expression of miR-26a-5p and miR-636 showed significant association (P = 0.024 and P = 0.050) with poor prognosis of CRC patients. However, the other seventeen miRNAs did not exhibit a statistically significant correlation (P > 0.05) with the overall survival of CRC patients (Table S2).
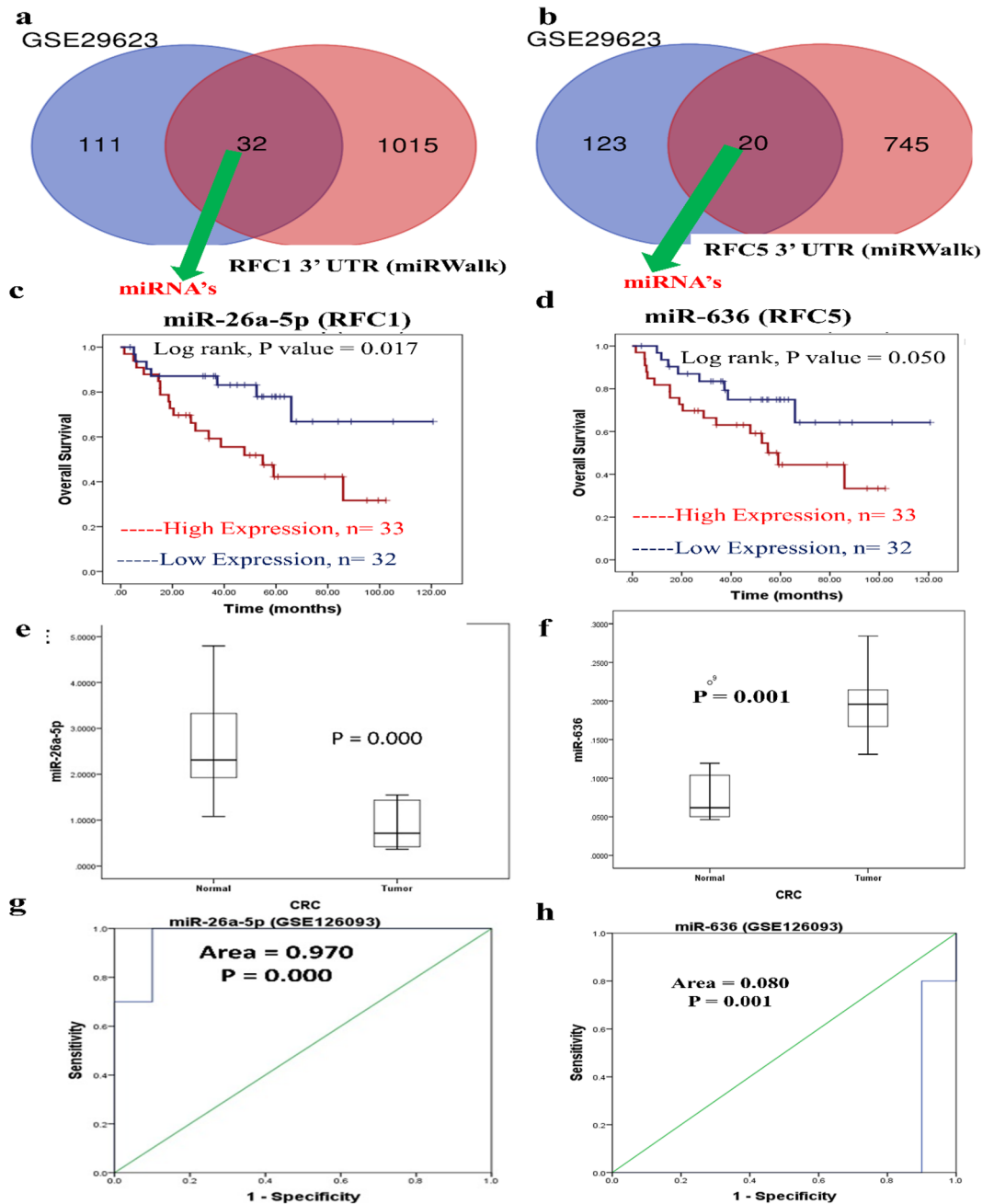
Figure 5. Venn-plot diagram showing miRNA targeting (**a**) *RFC1* 3′UTR and (**b**) *RFC5* 3′UTR in CRC identified from mirWalk vs GSE29623. Kaplan–Meier survival curves presenting prognostic relationships between high and low expressions of specific micro (mi)RNAs to overall survival (OS) using the GSE29623 database patient samples. (**c**) miR-26a-5p, and (**d**) miR-636 survival curves were plotted using SPSS 22.0. Specific miRNA expression levels were selected by the median value. Results are visually presented by Kaplan–Meier survival plots, and *P* values were calculated using log-rank statistics. Patient number (n) = 65, p = log-rank *P*-value, with $P \leq 0.05$ considered significant. (**e**, **f**) GEO database analysis of normal vs. CRC tissues using the GSE126093 dataset. Expression levels miR-26a-5p, and miR-636 in colorectal cancer patients. $P < 0.05$ was considered significant. (**g**, **h**) ROC curve for normal vs. colorectal cancer (CRC) patients. MiR-26a-5p and miR-636 were considered significant.

Clinical validation of miR-26a-5p and miR-636 on CRC patients

The GEO dataset GSE126093 includes miRNA profiles from tissues of 20 CRC patients, with 10 CRC tissues and their corresponding normal-appearing tissues (NATs). Patients with tumor size > 5 cm, lymph node metastases (Lx group), stage III–IV, or metastases were higher than those in patients with < 5 cm tumor size, without

lymph node metastases (L0), I–II stage or non-metastases, respectively[21]. The expression levels of miR-26a-5p were significantly lower in CRC patients (P = 0.000) whereas miR-636 were significantly higher in CRC patients (P = 0.001) (Fig. 5E,F). The ROC analysis for sensitivity was carried out using SPSS 22.0 (Fig. 5G,H). From the overall result, a double mRNA was selected as it was significantly associated with the OS.

Functional interaction protein–protein interaction network

We used the Gene MANIA online web tool to investigate the functional interactions of the miRNA targets, and the Reactome mismatch repair, DNA replication, and nucleotide excision repair pathway RFC's were used to analyze the functional roles of these molecules. The analysis revealed an interaction network involving 21 other related genes, the 5 RFCs targets (*RFC1, RFC2, RFC3, RFC4, RFC5*) were entered, and a total of 1856 links were observed. Two types of interaction, physical and co-expression interaction were involved, with co-expression being the most frequent type of interaction (8.01%) among them (Fig. 6A).

Furthermore, we loaded the RFCs into the STRING database (https://string-db.org/) to extract protein–protein interaction (PPI) pairs. Subsequently, these pairs were imported into Cytoscape software to identify hub genes, as illustrated in Fig. 6B and C.

A high level of miR-26a-5p and miR-636 can suppress only a low level of *RFC1 and RFC5*, it also regulates the mismatch repair, DNA replication, and nucleotide excision repair pathway.

The elevated expression of miR-26a-5p and miR-636, leading to the targeted suppression of RFC1 and RFC5 in CRC, appears to activate the mismatch repair, DNA replication, and nucleotide excision repair signaling pathway. The interaction between these miRNAs and their target genes, along with their regulatory mechanism, is summarized in Fig. 6D.

Association of RFC1 and RFC5 genes with immune cell infiltration

The relationship was examined between the *RFC1* gene with immune cell infiltration and an inflammatory response in CRC patients. The TIMER database, an online web tool, was used to predict the link between the *RFC1* and *RFC5* gene expression to immune infiltration in CRC patients (Fig. 7, Fig. S2). The results enunciated correlation between *RFC1* and *RFC5* cluster of differentiation CD4 + T cell, macrophages, and neutrophils were correlated in COAD (Colon Adenocarcinoma) and READ (Rectal Adenocarcinoma) patients.

ERBB2, KRAS, and PTEN correlations with RFC1

The *RFC1* and *RFC5* correlation with the mutation of BRAF, ERBB2, KRAS, and PTEN was studied and it was found that the correlation of *RFC1* with BRAF, ERBB2, KRAS and PTEN is medium to low with R value of 0.51, − 0.031, 0.21 and 0.35 respectively. Similarly, the correlation of *RFC5* with BRAF, ERBB2, KRAS and PTEN is low with R value of 0.11, − 0.08, 0.11, 0.15 respectively (Fig. 8A).

Small molecule predictions and docking analysis

Through the WEB-based GEne SeT AnaLysis Toolkit (WebGestalt, http://www.webgestalt.org), drug predictions were made using the single gene found by the survival analysis and HPA database. As presented by Table 2, RFC1 gene-targeted genes were finally considered to be druggable, with a p value of 1.210e-10. Compound (ID: PA16471233) (CHEMBL430483) was identified as potential small molecule (Fig. 8B).

To study the interaction of the identified compound (CHEMBL430483) with the target protein docking studies were performed using Schrodinger Maestro 12.9. Two-dimensional (2D) and three-dimensional (3D) binding of CHEMBL430483 with proteins have been provided (Fig. 8C,D) and Table 3. This CHEMBL430483 has the docking score of − 7.072 with protein RFC1, the analysis of the binding pocket in the 2D interaction showed that it major binds to the pocket containing amino acid Tyr590, Pro592, Lys591, Val507, Trp586, Glu159, Arg156. Two hydrogen bond bonds are formed with amino acids Tyr590 and Glu159 which are important for interaction with the target.

Discussion

Our study revealed that *RFC1* and *RFC5* are two significant antitumor mRNA and inhibitors of tumor progression. *RFC1* and *RFC5* were downregulated in colorectal adenocarcinoma and functionally suppressed the CRC. MiR-26a-5p and miR-636 overexpression by competitively binding *RFC1 and RFC5* mRNA 3'UTR leads to mismatch repair, DNA replication, and the nucleotide excision repair signaling pathway of colorectal adenocarcinoma.

This study, used previously, published mRNA expression of RFC complex in different cancers. A computational analysis was performed by defining the RFCs correlated with miRNAs. KEGG pathway enrichment analyses were done, and a protein–protein interaction (PPI) network was performed to identify network genes. Furthermore, overall survival (OS) was obtained to determine survival biomarkers for identified colorectal cancer (CRC) patients[22].

Replication factor C complexes play a crucial role in unloading and loading processivity clamps from DNA. They have been identified as involved in repair pathways and multiple DNA replication. The RFC (Ctf18) variant complex specifically is required to activate the intra-S-phase checkpoint at stalled replication forks and aids the establishment of sister chromatid cohesion. Unlike other RFC complexes, RFC (Ctf18) contains two non-RFC subunits, Dcc1 and Ctf8[11]. RFC1 is the largest subunit (140 kDa) of the RFC complex.

Previous studies have shown that *RFC2* the second largest subunit (40 kDa)[23] among the RFC complexes, was upregulated and associated with some tumor tissues such as choriocarcinoma tissue and nasopharyngeal carcinoma (NPC) tissue[24,25]. Other studies have shown that high expression of *RFC2* is associated with poor survival in CRC, glioblastoma, and hepatocellular carcinoma and aids in predicting breast cancer progression and
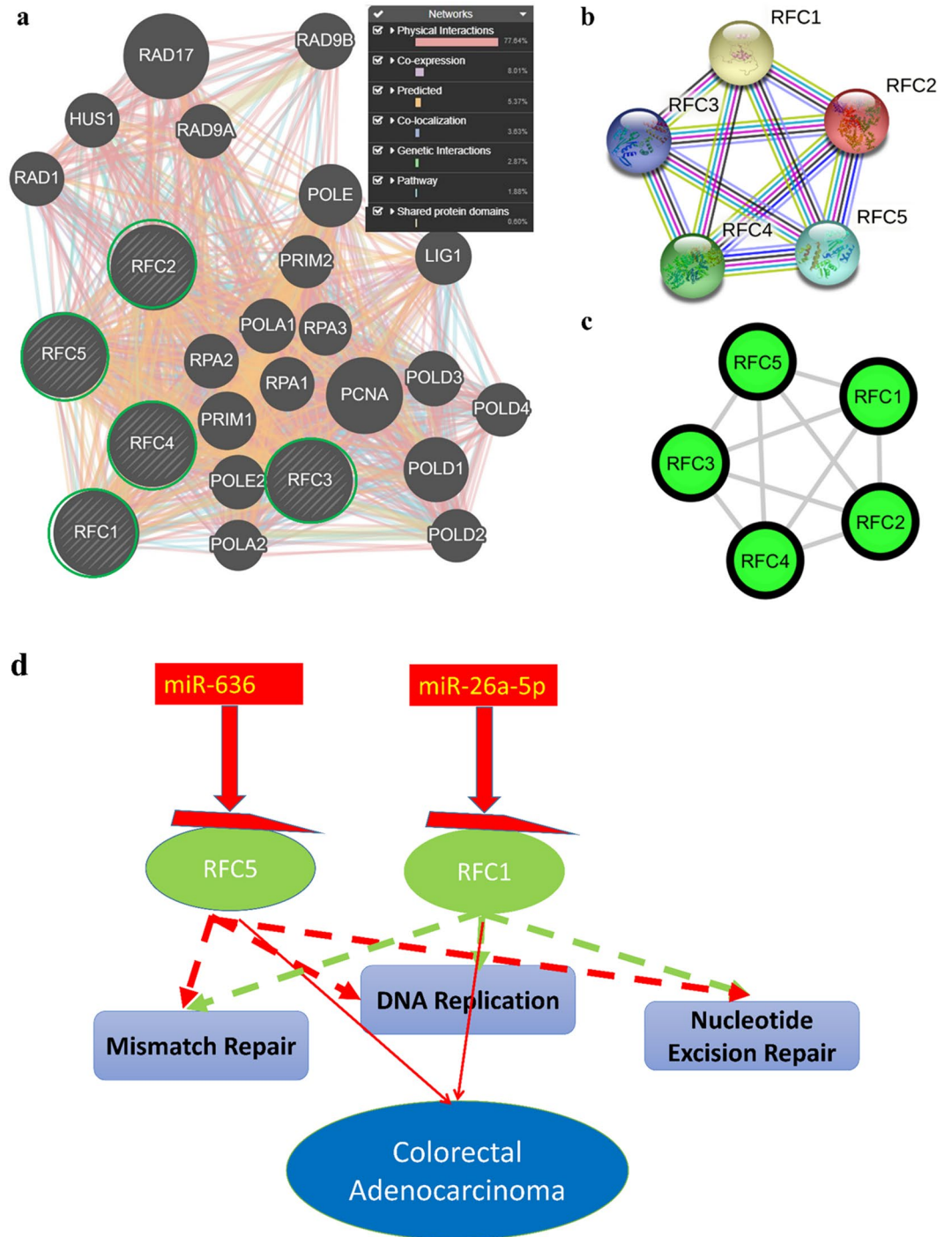
Figure 6. Gene interaction network (**a**) and (**b**, **c**) protein–protein interaction (PPI) network of replication factor C subunits genes (RFCs). In (**a**), input genes are indicated by stripes, with green circles representing downregulated genes in colorectal adenocarcinoma. In (**b**, **c**), PPI pairs were imported into Cytoscape software as described in "Methods and materials". Green nodes represent downregulated genes. The lines represent the interactive relationship between nodes. The highlighted DEGs represent a degree of ≥ 2. (**d**) Gene interaction network and pathway enrichment summary of common micro (mi)RNA targets. This schematic summary shows possible interactions of miRNAs and their colorectal adenocarcinoma targets. The red background represents upregulation and the green background indicates downregulation in colorectal adenocarcinoma, revealing significant expression in the respective validation dataset. Thick dashed red and green lines represent significant interactions with pathways. Colorectal adenocarcinoma is represented by a blue background.
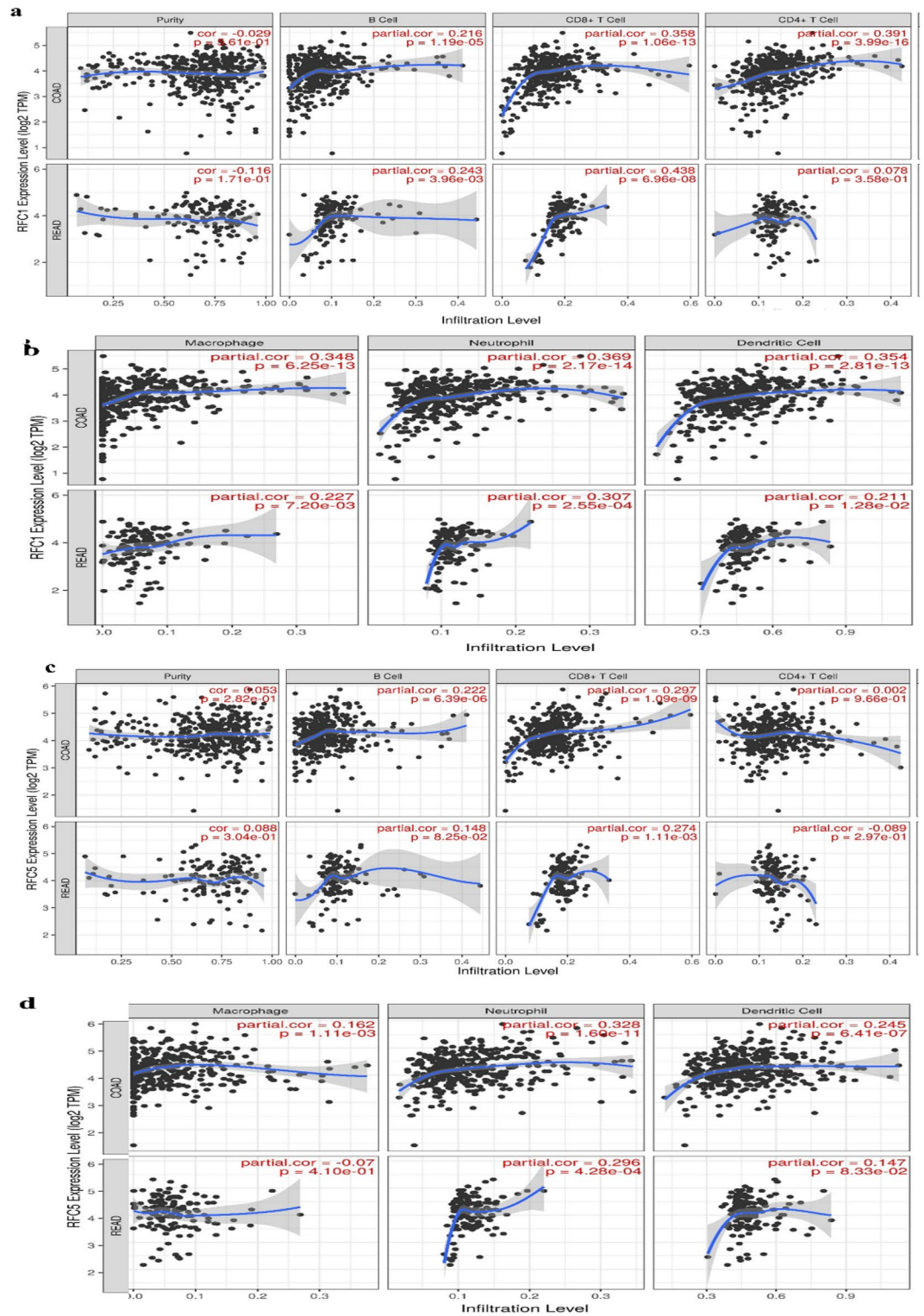
Figure 7. Immune filtration of the replication factor C subunit 1 (*RFC1* and *RFC5*) gene. Spearman correlations between the differentially expressed *RFC1* and *RFC5* gene and immune cell infiltration in (**a**) colon adenocarcinoma (COAD) and (**b**) rectal adenocarcinoma (READ) patients. The TIMER web tool was used for the analysis of correlations between immune infiltration of the immune cell markers of B cells, CD4+ cells, CD8+ cells, T cells, macrophages, neutrophils, and dendritic cells vs. the *RFC1* and *RFC5* gene. Statistically, significance was accepted at $P < 0.05$ for Spearman correlations.
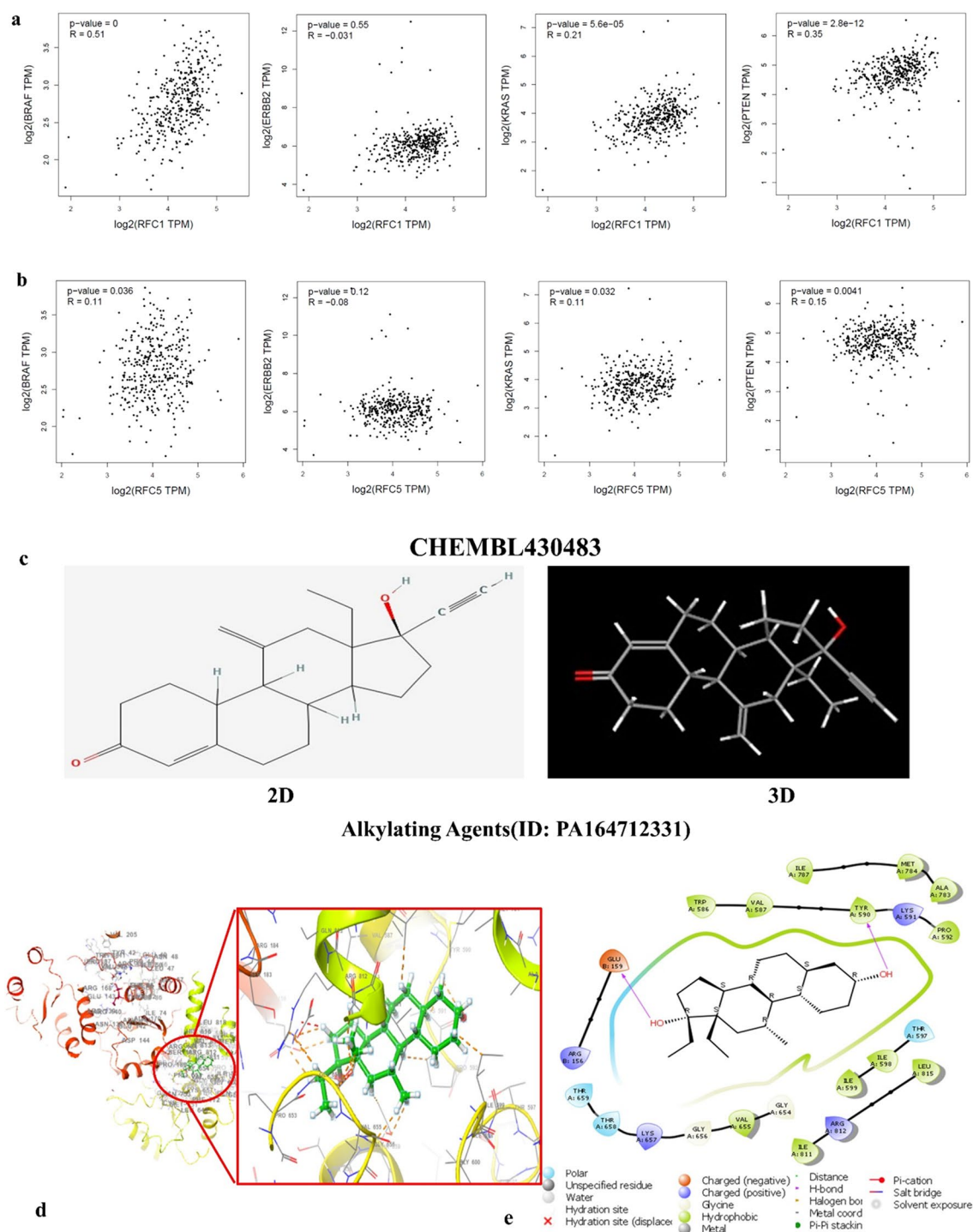
Figure 8. Mutation of BRAF, ERBB2, KRAS, and PTEN correlation with *RFC1* and *RFC5* (**a, b**) BRAF mutation, ERBB2 mutation, KRAS mutation and PTEN mutation. (**c**) chemical structure. (**d**) Binding of CHEMBL430483 against the RFC1 (6VVO) protein. (**e**) 2D interaction of CHEMBL430483against RFC1 (6VVO) protein.

metastasis[13,26,27]. Our results revealed that *RFC2* expression is insignificant in typical vs. tumor of CRC patients and high intensity in IHA protein analysis. However, the results are inconsistent with the overall survival data of CRC patients.

The *RFC3* gene is one of the small subunits (38 kDa) of the RFC complexes has been reported preferentially blind to proliferating cell nuclear antigen (PCNA) and formed a complex. Also reports related to attenuating the *RFC3* can inhibit tumor cell proliferation are present[28]. Although the *RFC3* is a tumor suppressor gene, it has been associated with poor survival in triple-negative breast cancer, ovarian tumor, lung adenocarcinoma,

Drug	Gene symbol	Gene name	P-value	FDR
Alkylating agents (ID: PA16471233)	RFC1 RFC2 RFC3 RFC4 RFC5	Replication factor C1-5	1.210e−10	2.224e−7

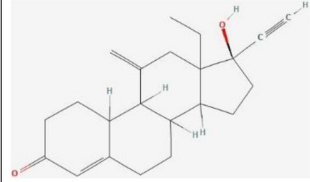**Table 2.** Small molecules target therapy.

S.no.	Structure	Target protein	Docking score (XP)
1	(CHEMBL430483)	RFC1	− 7.072

Table 3. Identification of drug molecule using GEne SeT AnaLysis Toolkit.

esophageal adenocarcinoma, hepatocellular carcinoma, suggesting that RFC3 may be a potential risk oncogenic gene involved in tumorigenesis[29–34]. In our results, RFC3 expression is insignificant in typical vs. tumor of CRC patients but the results are inconsistent with overall survival data of CRC patients.

The replication factor C subunit 4 (RFC4) has been reported to be involved in DNA replication as a clamp loader of PCNA[35]. The RFC4 has also been identified in previous studies as a tumor suppressor gene and has been associated with poor prognosis in CRC, HCC, cervical cancer, oral tongue squamous cell carcinoma, NSCLC, and esophageal squamous cell carcinoma[36–41]. In our results, RFC4 expression was significant in typical vs. tumor tissue of CRC patients but showed inconsistent results with overall survival data of CRC patients.

Overall, RFC5 (36 kDa) is a tumor suppressor and indicates poor survival in many cancers such as lung cancer and glioblastoma[42–44]. In our results, RFC5 expression was significant in normal vs. tumor tissue of CRC patients, and showed high intensity with IHA protein analysis, and consistent results with overall survival data of CRC patients. However, RFC5 is targeted with associated miR-636 which is why we selected RFC1 and RFC5 as a prognostic marker for colorectal adenocarcinoma.

We explored the potential mechanisms of RFC1 and RFC5 that mediated colorectal adenocarcinoma by focusing on potential microRNA. Here we identified that high expression of miR-26a-5p and miR-636 were associated with a poor prognosis of CRC patients. Recently, microRNAs have become famous for cancer treatment however, some previous studies identified, upregulated miR-26a-5p to be a tumor suppressor oncomir in CRC and sponge or a mediating oncomir that regulated autophagy, cell migration, cell proliferation, and invasion via the PI3K-AKT pathway in CRC[18,45–51]. miR-636 was identified as an onco-suppressor in lung cancer, nasopharyngeal carcinoma, cervical cancer, endometrial cancer, ovarian cancer, hepatocellular carcinoma[19,52–61]. High expression of miR-26a-5p and miR-636 has been correlated with poor OS of CRC patients, which was found consistent with our results.

The tumor microenvironment plays a critical role in the cancer progression of metastatic cancer, and tumor-associated macrophages (TAMs) form essential components of the tumor microenvironment. High TAMs is associated with invasion, migration, and IL6 for tumor progression of CRC metastasis[62]. Tumor infiltration is associated with six cells B cells, CD8+, CD4+, macrophages, neutrophils, and dendritic cells[63]. Our results showed that the RFC1 and RFC5 are associated with CD8+, CD4+, neutrophils, and macrophages. Which means the RFC1 and RFC5 can be used as a prognostic tumor marker for colorectal adenocarcinoma.

Our research is first to show a link between the RFC1 and RFC5 with tumor prognosis. This study showed that RFC1 and RFC5 is associated with overall survival and the prognosis of CRC in patients. Some previous studies also showed that the miR-26a-5p and miR-636 targeted RFC1 and RFC5 in the mismatch repair, DNA replication, nucleotide excision repair pathway, mediated colorectal adenocarcinoma, and our results are consistent with those of a previous study. Moreover, we also predicted a small molecule, CHEMBL430483 targeted RFC1 through webGestalt analysis. The interaction assessments of CHEMBL430483 docked positions, supports that it can be developed as potent inhibitor.

Nevertheless, some limitations exist in our study. It was difficult to collect sufficient CRC patient samples and to carry out the in vitro and in vivo studies or find a suitable public database to evaluate the clinical significance of miR-26a-5p or targeted RFC1 and miR-636 targeted RFC5 in terms of expression levels and CRC progression.

As differential expression miRNA data was limited, we only used twenty miRNAs in this investigation, leading to only two miR-26a-5p and miR-636 being identified as linked with CRC patients.

In the future, the tumor-suppressive role of miR-26a-5p targeted RFC1 and miR-636 targeted RFC5 in CRC progression needs to be further investigated in a larger cohort of patients.

We believe that miR-26a-5p targeted *RFC1* and miR-636 targeted *RFC5* in the mismatch repair, DNA replication, and nucleotide excision repair pathway, play an essential role by mediating colorectal adenocarcinoma progression.

Materials and methods

Data collection

The mRNA sequencing data, molecular categories, and clinical information of colorectal cancer patients, as well as other cancer types, were sourced from the TCGA, COAD, and READ databases (https://tcga-data.nci.nih.gov/, with links to COAD and READ accessed on 24 December 2021). Immunohistochemistry (IHC) staining data were retrieved from the Human Protein Atlas, available at https://www.proteinatlas.org/ (accessed on 24 December 2021), and GEPIA, available at http://gepia.cancer-pku.cn/index.html (accessed on 24 December 2021), respectively. The expression patterns of normal colon and tumor tissues were obtained from the TCGA database. The STRING database, accessible at https://www.string-db.org/ (accessed on 24 December 2021), and Cytoscape software (version 3.4.0, http://www.cytoscape.org), were utilized to construct the protein–protein interaction (PPI) network and enrichment pathway. The immune cell content file of the TCGA samples was acquired from the TIMER database (https://cistrome.shinyapps.io/timer/, accessed on 24 December 2021).

Expression analysis of RFCs in CRC

The CCLE dataset (https://portals.broadinstitute.org/ccle, accessed on 28 December 2021) was employed to illustrate the mRNA expression levels of RFCs in cancer cell lines. The expression data are presented in a heatmap using the CIMminer web tools. Additionally, this study utilized the GEPIA databases, accessible at http://gepia.cancer-pku.cn/index.html, to examine the expression levels of RFCs in both normal and CRC tissues. The analysis applied a threshold of an absolute log base 2 of the fold change (Log2FC) set to 1, and the q value set to 0.05[64].

Validation of RFCs genes in CRC patients

To validate the role of RFC genes, we conducted a survival analysis using the GEPIA database, which is accessible at http://gepia.cancer-pku.cn/detail.php?gene=&clicktag=survival. The Kaplan–Meier survival analysis graph was generated for a selected cohort of 362 CRC patients with both mutation and RNA sequence data. Additionally, we predicted the survival of miRNA using a specific cohort from GSE29623. The overall analysis was performed using the statistical software SPSS version 22.0 (SPSS, Chicago, IL, USA, www-01.ibm.com).

Pathway and enrichment analysis

The Database for Annotations, Visualization, and Integrated Discovery (DAVID bioinformatics, available at https://david.ncifcrf.gov/), was utilized to differentiate the expression genes based on their cellular components, molecular functions, and biological processes, utilizing resources from the Gene Ontology knowledgebase (GO, available at http://www.geneontology.org/)[65]. DAVID was employed for the enrichment analysis of RFC genes, and the pathway analysis was conducted with reference to the Kyoto Encyclopedia of Genes and Genomes (KEGG, http://www.genome.jp/kegg/) database, using FDR < 0.25 as the cutoff point[66].

Protein expression analysis of RFCs in CRC

The intensity of RFC proteins in CRC tissues within the human body was examined using the Human Protein Atlas (HPA) database, accessible at https://www.proteinatlas.org/[67]. The HPA database offers over 700 antibodies to human proteins for matching with 400,000 high-resolution images. The following formula was employed to assess each intensity and fraction combination, automatically converting it into a protein expression level score. The scores were categorized as follows: negative—not detected; weak—not detected; weak combined with either 25–75 percent or 75 percent—low; moderate—low; moderate combined with either 25–75 percent or 75 percent—medium; strong—medium; strong combined with either 25–75 percent or 75 percent—high.

Construction of protein–protein interaction network

A protein–protein interaction network was constructed using the online web tool 'STRING' (http://www.string-db.org/)[68]. This tool provides information on known and predicted protein interactions, derived from four sources: genomic context, co-expression, high-throughput experiments, and previous knowledge. A score of 0.4 (medium confidence) was selected as the cutoff criterion. PPI pairs were then analyzed using Cytoscape software (version 3.4.0, http://www.cytoscape.org) with the CytoNCA app. Hub genes, representing highly connected genes, were identified by calculating the degree value (the number of lines connecting the genes) with a cutoff of ≥ 2.

Prediction of microRNA for RFCs

MiRNAs were predicted using the GSE29623 cohort study, encompassing 143 miRNA samples in colon adenocarcinomas, and miRWalk websites for 1047 miRNAs of RFC1 and 765 miRNAs of RFC5 3'UTR. A target score of < 0.70 was selected in the miRWalk webtools. Additionally, a Venn diagram was employed to illustrate common miRNAs and their shared gene targets[69].

Clinical examination of the microRNA

For the clinical examination of differentially expressed (DE) miRNAs, a survival analysis was conducted using CRC patients. The CRC metabase for tissue expression of miRNA, including GSE29623 with a total of 65 cases and GSE126093 with 20 cases, was selected. Expression profiles were compared based on low or high expression

using the Mann–Whitney U test. Statistical analysis was performed using SPSS version 22.0 (SPSS, Chicago, IL, USA, www-01.ibm.com) for plotting Kaplan–Meier curves, boxplots, and ROC curves. A P-value < 0.05 was considered statistically significant.

Clinical relevance of mutation of BRAF, ERBB2, KRAS, and PTEN
GEPIA, available at http://gepia.cancer-pku.cn/index.html (accessed on 19 September 2022), was utilized to examine the correlation of mutations with RFC1 and RFC5.

Immune infiltration analysis RFCs in CRC
The TIMER database (available at http://timer.cistrome.org/; accessed on 24 December 2021) was employed for the analysis of immune infiltration. As a comprehensive database offering analysis of immune infiltrates in various cancer types[70,71], it was used in this study to investigate the involvement of RFCs in immune infiltrates in CRC. Scatterplots were used to present the relationship between gene expression and estimated infiltrate values, with the level of significance set at P < 0.05.

Small-molecule drug-targeting therapy for DEGs
The drugs linked to DEGs were predicted using the web-based GEne SeT AnaLysis Toolkit (WebGestalt, http://www.webgestalt.org), an integrated system for gene analysis[72].

Docking study
Molecular docking studies were validated to confirm the selected drug to study the interaction of drug with the protein. For docking investigations, the 3D structures of the target proteins (PBD ID-6VVO) was acquired from the Protein Data Bank (PDB) at (https://www.rcsb.org/)[73]. Compound (ID: PA16471233) (CHEMBL ID 430483) identified from webGestalt was downloaded from the PubChem databases (https://pubchem.ncbi.nlm.nih.gov/)[74]. The molecular docking software, maestro v12.8 (Schrödinger), was used to perform molecular docking and visualize interactions[75].

Conclusions
Our study concluded that high expression of miR-26a-5p targeting *RFC1* and miR-636 targeting *RFC5* expression in the mismatch repair, DNA replication, nucleotide excision repair pathway is expected to be a potential biomarker for detecting CRC at an early stage and small molecule CHEMBL430483 can be developed for the treatment of the disease. Our findings may provide novel directions and strategies for CRC therapies.

Data availability
The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding authors.

References
1. Siegel, R. L. & Miller, K. D. Colorectal cancer statistics. *Cancer* **2020**(70), 145–164. https://doi.org/10.3322/caac.21601 (2020).
2. Ruers, T. *et al.* Local treatment of unresectable colorectal liver metastases: Results of a randomized phase II trial. *J. Natl. Cancer Inst.* https://doi.org/10.1093/jnci/djx015 (2017).
3. Scheele, J., Stangl, R. & Altendorf-Hofmann, A. Hepatic metastases from colorectal carcinoma: Impact of surgical resection on the natural history. *Br. J. Surg.* **77**, 1241–1246. https://doi.org/10.1002/bjs.1800771115 (1990).
4. Pancione, M. *et al.* Emerging insight into MAPK inhibitors and immunotherapy in colorectal cancer. *Curr. Med. Chem.* **24**, 1383–1402. https://doi.org/10.2174/0929867324666170227114356 (2017).
5. Papadatos-Pastos, D., Rabbie, R., Ross, P. & Sarker, D. The role of the PI3K pathway in colorectal cancer. *Crit. Rev. Oncol. Hematol.* **94**, 18–30. https://doi.org/10.1016/j.critrevonc.2014.12.006 (2015).
6. Binefa, G., Rodriguez-Moranta, F., Teule, A. & Medina-Hayas, M. Colorectal cancer: From prevention to personalized medicine. *World J. Gastroenterol.* **20**, 6786–6808. https://doi.org/10.3748/wjg.v20.i22.6786 (2014).
7. Johnson, A., Yao, N. Y., Bowman, G. D., Kuriyan, J. & O'Donnell, M. The replication factor C clamp loader requires arginine finger sensors to drive DNA binding and proliferating cell nuclear antigen loading. *J. Biol. Chem.* **281**, 35531–35543. https://doi.org/10.1074/jbc.M606090200 (2006).
8. Oku, T. *et al.* Functional sites of human PCNA which interact with p21 (Cip1/Waf1), DNA polymerase delta and replication factor C. *Genes Cells* **3**, 357–369. https://doi.org/10.1046/j.1365-2443.1998.00199.x (1998).
9. Tsurimoto, T. & Stillman, B. Purification of a cellular replication factor, RF-C, that is required for coordinated synthesis of leading and lagging strands during simian virus 40 DNA replication in vitro. *Mol. Cell. Biol.* **9**, 609–619. https://doi.org/10.1128/mcb.9.2.609-619.1989 (1989).
10. Tsurimoto, T. & Stillman, B. Functions of replication factor C and proliferating-cell nuclear antigen: Functional similarity of DNA polymerase accessory proteins from human cells and bacteriophage T4. *Proc. Natl. Acad. Sci. USA* **87**, 1023–1027. https://doi.org/10.1073/pnas.87.3.1023 (1990).
11. Wade, B. O., Liu, H. W., Samora, C. P., Uhlmann, F. & Singleton, M. R. Structural studies of RFC(C)(tf18) reveal a novel chromatin recruitment role for Dcc1. *EMBO Rep.* **18**, 558–568. https://doi.org/10.15252/embr.201642825 (2017).
12. Yao, N. Y. & O'Donnell, M. The RFC clamp loader: Structure and function. *Sub-cell. Biochem.* **62**, 259–279. https://doi.org/10.1007/978-94-007-4572-8_14 (2012).
13. Ji, Z., Li, J. & Wang, J. Up-regulated RFC2 predicts unfavorable progression in hepatocellular carcinoma. *Hereditas* **158**, 17. https://doi.org/10.1186/s41065-021-00179-9 (2021).
14. Prayugo, F. B. *et al.* Expression profiles and prognostic value of FABPs in colorectal adenocarcinomas. *Biomedicines* https://doi.org/10.3390/biomedicines9101460 (2021).

15. Makondi, P. T. *et al.* Prediction of novel target genes and pathways involved in bevacizumab-resistant colorectal cancer. *PLoS ONE* **13**, e0189582. https://doi.org/10.1371/journal.pone.0189582 (2018).
16. Makondi, P. T., Chu, C. M., Wei, P. L. & Chang, Y. J. Prediction of novel target genes and pathways involved in irinotecan-resistant colorectal cancer. *PLoS ONE* **12**, e0180616. https://doi.org/10.1371/journal.pone.0180616 (2017).
17. Chen, K. C. & Juo, S. H. MicroRNAs in atherosclerosis. *Kaohsiung J. Med. Sci.* **28**, 631–640. https://doi.org/10.1016/j.kjms.2012.04.001 (2012).
18. Xu, M. *et al.* lncRNA SNHG6 regulates EZH2 expression by sponging miR-26a/b and miR-214 in colorectal cancer. *J. Hematol. Oncol.* **12**, 3. https://doi.org/10.1186/s13045-018-0690-5 (2019).
19. Li, Y., Huo, J., He, J. & Ma, X. LncRNA MONC suppresses the malignant phenotype of endometrial cancer stem cells and endometrial carcinoma cells by regulating the MiR-636/GLCE axis. *Cancer Cell* **21**, 331. https://doi.org/10.1186/s12935-021-01911-1 (2021).
20. Li, Y. *et al.* Multifaceted regulation and functions of replication factor C family in human cancers. *Am. J. Cancer Res.* **8**, 1343–1355 (2018).
21. Chen, Z. *et al.* Hsa_circ_101555 functions as a competing endogenous RNA of miR-597-5p to promote colorectal cancer progression. *Oncogene* **38**, 6017–6034. https://doi.org/10.1038/s41388-019-0857-8 (2019).
22. Sun, F., Liang, W. & Qian, J. The identification of CRNDE, H19, UCA1 and HOTAIR as the key lncRNAs involved in oxaliplatin or irinotecan resistance in the chemotherapy of colorectal cancer based on integrative bioinformatics analysis. *Mol. Med. Rep.* **20**, 3583–3596. https://doi.org/10.3892/mmr.2019.10588 (2019).
23. Noskov, V. N., Araki, H. & Sugino, A. The RFC2 gene, encoding the third-largest subunit of the replication factor C complex, is required for an S-phase checkpoint in Saccharomyces cerevisiae. *Mol. Cell. Biol.* **18**, 4914–4923. https://doi.org/10.1128/mcb.18.8.4914 (1998).
24. Cui, J. Q., Shi, Y. F. & Zhou, H. J. Expression of RFC2 and PCNA in different gestational trophoblastic diseases. *Chin. J. Cancer* **23**, 196–200 (2004).
25. Xiong, S., Wang, Q., Zheng, L., Gao, F. & Li, J. Identification of candidate molecular markers of nasopharyngeal carcinoma by tissue microarray and in situ hybridization. *Med. Oncol.* **28**(Suppl 1), S341-348. https://doi.org/10.1007/s12032-010-9727-5 (2011).
26. Ho, K. H. *et al.* Xanthohumol regulates miR-4749-5p-inhibited RFC2 signaling in enhancing temozolomide cytotoxicity to glioblastoma. *Life Sci.* **254**, 117807. https://doi.org/10.1016/j.lfs.2020.117807 (2020).
27. Hu, T. *et al.* RFC2, a direct target of miR-744, modulates the cell cycle and promotes the proliferation of CRC cells. *J. Cell. Physiol.* **235**, 8319–8333. https://doi.org/10.1002/jcp.29676 (2020).
28. Yao, Z. *et al.* shRNA-mediated silencing of the RFC3 gene suppresses hepatocellular carcinoma cell proliferation. *Int. J. Mol. Med.* **36**, 1393–1399. https://doi.org/10.3892/ijmm.2015.2350 (2015).
29. Gong, S. *et al.* RFC3 induces epithelial-mesenchymal transition in lung adenocarcinoma cells through the Wnt/β-catenin pathway and possesses prognostic value in lung adenocarcinoma. *Int. J. Mol. Med.* **44**, 2276–2288. https://doi.org/10.3892/ijmm.2019.4386 (2019).
30. He, Z. Y. *et al.* Up-regulation of RFC3 promotes triple negative breast cancer metastasis and is associated with poor prognosis via EMT. *Transl. Oncol.* **10**, 1–9. https://doi.org/10.1016/j.tranon.2016.10.004 (2017).
31. Lockwood, W. W. *et al.* Integrative genomics identified RFC3 as an amplified candidate oncogene in esophageal adenocarcinoma. *Clin. Cancer Res.* **18**, 1936–1946. https://doi.org/10.1158/1078-0432.ccr-11-1431 (2012).
32. Shen, H. *et al.* Overexpression of RFC3 is correlated with ovarian tumor development and poor prognosis. *Tumour Biol.* **35**, 10259–10266. https://doi.org/10.1007/s13277-014-2216-2 (2014).
33. Shen, H. *et al.* ShRNA-mediated silencing of the RFC3 gene suppress ovarian tumor cells proliferation. *Int. J. Clin. Exp. Pathol.* **8**, 8968–8975 (2015).
34. Zhou, J. *et al.* Downregulation of hsa_circ_0011946 suppresses the migration and invasion of the breast cancer cell line MCF-7 by targeting RFC3. *Cancer Manag. Res.* **10**, 535–544. https://doi.org/10.2147/cmar.s155923 (2018).
35. Xiang, J. *et al.* Levels of human replication factor C4, a clamp loader, correlate with tumor progression and predict the prognosis for colorectal cancer. *J. Transl. Med.* **12**, 320. https://doi.org/10.1186/s12967-014-0320-0 (2014).
36. Liu, L. *et al.* An RFC4/Notch1 signaling feedback loop promotes NSCLC metastasis and stemness. *Nat. Commun.* **12**, 2693. https://doi.org/10.1038/s41467-021-22971-x (2021).
37. Tang, X. *et al.* Identification of key candidate genes and small molecule drugs in cervical cancer by bioinformatics strategy. *Cancer Manag. Res.* **10**, 3533–3549. https://doi.org/10.2147/cmar.s171661 (2018).
38. Wang, J. *et al.* The upregulated expression of RFC4 and GMPS mediated by DNA copy number alteration is associated with the early diagnosis and immune escape of ESCC based on a bioinformatic analysis. *Aging* **13**, 21758–21777. https://doi.org/10.18632/aging.203520 (2021).
39. Wang, X. C. *et al.* Genome-wide RNAi screening identifies RFC4 as a factor that mediates radioresistance in colorectal cancer by facilitating nonhomologous end joining repair. *Clin. Cancer Res.* **25**, 4567–4579. https://doi.org/10.1158/1078-0432.ccr-18-3735 (2019).
40. Yang, W. X., Pan, Y. Y. & You, C. G. CDK1, CCNB1, CDC20, BUB1, MAD2L1, MCM3, BUB1B, MCM2, and RFC4 may be potential therapeutic targets for hepatocellular carcinoma using integrated bioinformatic analysis. *BioMed. Res.* **2019**, 1245072. https://doi.org/10.1155/2019/1245072 (2019).
41. Zhang, J., Wang, L. & Xie, X. RFC4 promotes the progression and growth of oral tongue squamous cell carcinoma in vivo and vitro. *J. Clin. Lab.* **35**, e23761. https://doi.org/10.1002/jcla.23761 (2021).
42. Peng, W. X. *et al.* FoxM1-mediated RFC5 expression promotes temozolomide resistance. *Cell Biol. Toxicol.* **33**, 527–537. https://doi.org/10.1007/s10565-017-9381-1 (2017).
43. Wang, M. *et al.* Identification of RFC5 as a novel potential prognostic biomarker in lung cancer through bioinformatics analysis. *Oncol. Lett.* **16**, 4201–4210. https://doi.org/10.3892/ol.2018.9221 (2018).
44. Zhao, X. *et al.* AEG-1 knockdown sensitizes glioma cells to radiation through impairing homologous recombination via targeting RFC5. *DNA Cell Biol.* **40**, 895–905. https://doi.org/10.1089/dna.2020.6287 (2021).
45. Wang, X. *et al.* LncRNA SNHG6 promotes chemoresistance through ULK1-induced autophagy by sponging miR-26a-5p in colorectal cancer cells. *Cancer Cell Int.* **19**, 234. https://doi.org/10.1186/s12935-019-0951-6 (2019).
46. Lopez-Urrutia, E. *et al.* MiR-26a downregulates retinoblastoma in colorectal cancer. *Tumour Biol.* **39**, 1010428317695945. https://doi.org/10.1177/1010428317695945 (2017).
47. Coronel-Hernandez, J. *et al.* Cell migration and proliferation are regulated by miR-26a in colorectal cancer via the PTEN-AKT axis. *Cancer Cell Int.* **19**, 80. https://doi.org/10.1186/s12935-019-0802-5 (2019).
48. Liu, B. *et al.* LINC01296/miR-26a/GALNT3 axis contributes to colorectal cancer progression by regulating O-glycosylated MUC1 via PI3K/AKT pathway. *J. Exp. Clin. Cancer Res.* **37**, 316. https://doi.org/10.1186/s13046-018-0994-x (2018).
49. Liu, B. *et al.* Correction to: LINC01296/miR-26a/GALNT3 axis contributes to colorectal cancer progression by regulating O-glycosylated MUC1 via PI3K/AKT pathway. *J. Exp. Clin. Cancer Res.* **38**, 142. https://doi.org/10.1186/s13046-019-1140-0 (2019).
50. Zhang, M., Duan, W. & Sun, W. LncRNA SNHG6 promotes the migration, invasion, and epithelial-mesenchymal transition of colorectal cancer cells by miR-26a/EZH2 axis. *OncoTargets Ther.* **12**, 3349–3360. https://doi.org/10.2147/OTT.S197433 (2019).
51. Xu, J. *et al.* Exosomal MALAT1 sponges miR-26a/26b to promote the invasion and metastasis of colorectal cancer via FUT4 enhanced fucosylation and PI3K/Akt pathway. *J. Exp. Clin. Cancer Res.* **39**, 54. https://doi.org/10.1186/s13046-020-01562-6 (2020).

52. De Tomi, E., Campagnari, R., Orlandi, E., Cardile, A. & Zanrè, V. Upregulation of miR-34a-5p, miR-20a-3p and miR-29a-3p by onconase in A375 melanoma cells correlates with the downregulation of specific onco-proteins. *J. Mol. Sci.* https://doi.org/10.3390/ijms23031647 (2022).

53. Deng, L., Wang, C., He, C. & Chen, L. Bone mesenchymal stem cells derived extracellular vesicles promote TRAIL-related apoptosis of hepatocellular carcinoma cells via the delivery of microRNA-20a-3p. *Cancer Biomark. A* 30, 223–235. https://doi.org/10.3233/cbm-201633 (2021).

54. Guo, W. W., Feng, M. M., Li, S. F. & Wei, L. H. Circular RNA circ_0023404 serves as a miR-636 sponge to promote malignant behaviors in cervical cancer cells through upregulation of CYP2S1. *Kaohsiung J. Med. Sci.* 38, 218–229. https://doi.org/10.1002/kjm2.12478 (2022).

55. Hu, Q. L. & Xu, Z. P. miR-636 represses cell survival by targeting CDK6/Bcl-2 in cervical cancer. *Kaohsiung J. Med. Sci.* 36, 328–335. https://doi.org/10.1002/kjm2.12181 (2020).

56. Jang, J. Y. *et al.* ANT2 suppression by shRNA restores miR-636 expression, thereby downregulating Ras and inhibiting tumorigenesis of hepatocellular carcinoma. *Exp. Mol. Med.* 45, e3. https://doi.org/10.1038/emm.2013.1 (2013).

57. Ma, J., Zhou, C. & Chen, X. miR-636 inhibits EMT, cell proliferation and cell cycle of ovarian cancer by directly targeting transcription factor Gli2 involved in Hedgehog pathway. *Cancer Cell* 21, 64. https://doi.org/10.1186/s12935-020-01725-7 (2021).

58. Song, J., Shi, W., Gao, Z., Liu, X. & Wang, W. Downregulation of circRNA_100876 inhibited progression of NSCLC in vitro via targeting miR-636. *Technol. Cancer Res. Treat.* 19, 1533033. https://doi.org/10.1177/1533033820951817 (2020).

59. Wang, Y. *et al.* Hsa_circ_0007637 facilitates nasopharyngeal carcinoma progression by sponging miR-636/TPD52 axis. *Cancer Manag. Res.* 13, 9439–9452. https://doi.org/10.2147/cmar.s328207 (2021).

60. Zhao, J., Yan, W., Huang, W. & Li, Y. Circ_0010235 facilitates lung cancer development and immune escape by regulating miR-636/PDL1 axis. *Cancer* 13, 965–976. https://doi.org/10.1111/1759-7714.14338 (2022).

61. Zhou, B. *et al.* Association of CASC18/miR-20a-3p/TGFB2 ceRNA axis with occult lymph node metastasis in tongue squamous cell carcinoma. *Mol. Med.* 27, 85. https://doi.org/10.1186/s10020-021-00345-9 (2021).

62. Wei, C. *et al.* Crosstalk between cancer cells and tumor associated macrophages is required for mesenchymal circulating tumor cell-mediated colorectal cancer metastasis. *Mol. Cancer* 18, 64. https://doi.org/10.1186/s12943-019-0976-4 (2019).

63. Zhang, Y. *et al.* Single-cell transcriptome analysis reveals tumor immune microenvironment heterogenicity and granulocytes enrichment in colorectal cancer liver metastases. *Cancer Lett.* 470, 84–94. https://doi.org/10.1016/j.canlet.2019.10.016 (2020).

64. Li, C., Tang, Z., Zhang, W., Ye, Z. & Liu, F. GEPIA2021: Integrating multiple deconvolution-based analysis into GEPIA. *Nucleic Acids Res.* 49, W242–W246. https://doi.org/10.1093/nar/gkab418 (2021).

65. da Huang, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4, 44–57. https://doi.org/10.1038/nprot.2008.211 (2009).

66. Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M. & Tanabe, M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* 44, D457-462. https://doi.org/10.1093/nar/gkv1070 (2016).

67. Uhlén, M. *et al.* Proteomics: Tissue-based map of the human proteome. *Science* 347, 1260419. https://doi.org/10.1126/science.1260419 (2015).

68. Szklarczyk, D. *et al.* STRING v10: Protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* 43, D447-452. https://doi.org/10.1093/nar/gku1003 (2015).

69. Hur, B. *et al.* Venn-diaNet: Venn diagram based network propagation analysis framework for comparing multiple biological experiments. *BMC Bioinform.* 20, 667. https://doi.org/10.1186/s12859-019-3302-7 (2019).

70. Li, T. *et al.* TIMER: A web server for comprehensive analysis of tumor-infiltrating immune cells. *Cancer Res.* 77, e108–e110. https://doi.org/10.1158/0008-5472.can-17-0307 (2017).

71. Li, T. *et al.* TIMER2.0 for analysis of tumor-infiltrating immune cells. *Nucleic Acids Res.* 48, W509–W514. https://doi.org/10.1093/nar/gkaa407 (2020).

72. Zhang, B., Kirov, S. & Snoddy, J. WebGestalt: An integrated system for exploring gene sets in various biological contexts. *Nucleic Acids Res.* 33, W741-748. https://doi.org/10.1093/nar/gki475 (2005).

73. Berman, H. M. *et al.* The protein data bank. *Acta Crystallogr. D* 58, 899–907. https://doi.org/10.1107/s0907444902003451 (2002).

74. Kim, S. *et al.* PubChem 2019 update: improved access to chemical data. *Nucleic Acids Res.* 47, D1102-d1109. https://doi.org/10.1093/nar/gky1033 (2019).

75. Shrivastava, A., Kumar, J. & Akhter, M. In-silico assessment of various PDB entries of PfLDH enzyme for their use in SBDD. *Chem. Inf.* 2, 1–9 (2016).

Author contributions

Conceptualization, M.M. and M.A.; methodology, M.M.; software, M.M.; validation, M.M., M.A. and M.K.; formal analysis, M.M.; investigation, M.M and M.A.; resources, M.A.; data curation, M.M.; writing—original draft preparation, M.M.; writing—review and editing, M.K. and M.A; visualization, M.M. and M.A.; supervision, M.A and A.K.N.; project administration, M.M.; All authors have read and agreed to the published version of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-56361-2.

Correspondence and requests for materials should be addressed to M.M. or M.A.

Reprints and permissions information is available at www.nature.com/reprints.

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