Diagnostic and prognostic potential of the novel biomarker nuclear cap binding protein subunit 2 (*NCBP2*) in colon adenocarcinoma

He Bu[1], Ting Cao[2], Xuesong Li[1], Yuekun Guo[1], Jiajia Guo[1], Yuxin Wang[1], Yunlong Sun[3], Donghong Wang[1]

[1]Department of Gastroenterology, The Third Affiliated Hospital of Qiqihar Medical University, Qiqihar, China; [2]Institute of Pathology of Qiqihar Medical University, Qiqihar, China; [3]Staidson Biopharmaceuticals Company Limited, Beijing, China

Contributions: (I) Conception and design: H Bu, T Cao; (II) Administrative support: Y Sun; (III) Provision of study materials or patients: Y Wang; (IV) Collection and assembly of data: Y Guo, J Guo; (V) Data analysis and interpretation: X Li, D Wang; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Donghong Wang. Department of Gastroenterology, The Third Affiliated Hospital of Qiqihar Medical University, Taishun Street 27, Tiefeng District, Qiqihar 161000, China. Email: wangdh0804@163.com.

Background: Colon adenocarcinoma (COAD) is an incurable malignancy and the third most common tumor worldwide. Advances in biomarkers screening have greatly contributed to explore the new diagnostic and prognostic biomarkers for the early detection and prognostic of COAD. However, the heterogeneity-specific nature of COAD in patients of different cancer stages, different races, genders and age are still the major challenge to clinical treatment.

Methods: Gene expression, copy number (CN), and dependency score (DS) data were obtained from the Cancer Cell Line Encyclopedia (CCLE), and linear regression analyses were performed using R language. We conducted receiver operating characteristic (ROC) curve analysis and compared the area under the ROC curve area under the curve (AUC) values to evaluate the sensitivity and specificity of nuclear cap binding protein subunit 2 (*NCBP2*) for the diagnosis of COAD in The Cancer Genome Atlas (TCGA) database. Survival analysis was performed in the distinct *NCBP2* expression groups. In vitro experiments and bioinformatics analysis were used to investigate the molecular mechanisms of *NCBP2* in COAD and its biological roles. A Connectivity Map (Cmap) was used to identify potential small molecule targeted drugs for *NCBP2* in COAD.

Results: Through the linear regression analysis of DS, CN, and gene expression, we determined that *NCBP2* met our criteria. The mean AUC of the ROC curve of *NCBP2* was 0.940±0.050. Survival analysis showed that high *NCBP2* expression was associated with a worse prognosis [hazard ratio (HR) =1.98, P<0.007]. *NCBP2* knockdown inhibited COAD cell proliferation and caused G0/G1 phase arrest in COAD cells.

Conclusions: NCBP2 was the novel diagnostic and prognostic biomarker of in COAD. Our research had implications for the treatment of colon cancer.

Keywords: Colon adenocarcinoma (COAD); novel biomarker; nuclear cap binding protein subunit 2 (*NCBP2*)

Introduction

Colon adenocarcinoma (COAD) is a malignant tumor that occurs in the colon or rectum (1,2), and ranks third in morbidity and second in mortality among all cancers worldwide. In Europe, approximately 250,000 new patients are diagnosed with COAD each year, accounting for about 9% of all malignancies (3,4). The incidence of COAD has increased with industrialization and urbanization, and it is more common in more affluent countries (5,6). Unfortunately, the lack of targeted therapeutic agents for COAD has resulted in poor therapeutic outcomes with

the existing clinical treatment strategies (7,8). Therefore, exploring the pathogenesis of COAD and proposing new clinical treatment strategies is crucial.

Recently, microarray technology has attracted much attention in the scientific research community, because the use of microarray technology can simultaneously screen thousands of differentially expressed messenger RNAs (mRNAs), microRNAs (miRNAs) and long non-coding RNAs (lncRNAs) (9-11). Undoubtedly, it plays a key role in the development and progression of diseases. In addition, the technology also facilitates further analysis of key genes to explore potential molecular targets and diagnostic biomarkers (12).

Numerous studies have shown that the pathogenesis and the prognosis of colon cancer are influenced by a combination of factors, including environmental and genetic factors (13,14). Recent studies have increasingly begun to focus on the role of genetic background variability in the development of colon cancer (6,7). For instance, the colorectal cancer risk of patients carrying copy number variations (CNVs) of human leukocyte antigen (HLA)-DQA1 is significantly higher than that of patients without CNVs (7). CNVs are large structural mutations involving genomic duplication or deletion of more than 1 kb, and are a possible key explanation for missing heritability in colorectal cancer (15,16). Therefore, identifying the underlying CNVs in patients with COAD may potentially provide a clinical strategy for better treatment outcomes.

Our study aims to explore potential prognosis-related biomarkers and their functions in COAD. Using bioinformatics analysis, nuclear cap binding protein subunit 2 (*NCBP2*) was identified as a key gene associated with the development of COAD. According to our findings, increased *NCBP2* copy numbers (CNs) cause overexpression and further tissue and brain architecture changes by influencing the cell cycle and apoptosis, which accelerates the pathogenesis of COAD. Based on this, we used Connectivity Map (Cmap) analysis to screen the potential small molecular drugs and provide a possible strategy for COAD treatment. We present the following article in accordance with the STARD and the MDAR reporting checklists (available at https://jgo.amegroups.com/article/view/10.21037/jgo-22-665/rc).

Methods

Data collection

The research design was showed in Figure S1. Both COAD-related and clinical data were acquired from The Cancer Genome Atlas (TCGA) data portal (https://portal.gdc.cancer.gov/). We downloaded the data of 330 colorectal cancer patients and 41 normal subjects. We also collected data on cell lines including gene expression, CN, and dependency score (DS) from depmap (https://depmap.org/portal/download/). The DS is a measure used to assess gene dependency through essentiality screens by computationally adjusting for the copy-number impact, allowing for unbiased gene dependency interpretation at all CN levels. The Declaration of Helsinki (as revised in 2013) was followed when conducting the study.

Linear regression analysis of the DS, gene expression, and CN

We performed linear regression analysis on gene expression, DS, and CN data from depmap using the rsm package (https://CRAN.R-project.org/package=rsm) in R. In the first linear regression analysis, gene expression was the dependent variable used to evaluate the DS. The following formula was developed: $Y = b1 \times X1$, where b1 represented gene expression and X1 denoted DS, and the cut-off criteria were $R < -0.4$ and $P < 0.05$.

CN was the dependent variable in the second linear regression study, which was utilized to assess the independent variable (gene expression). The following formula was developed: $Y = b1 \times X1$, where b1 was CN and X1 was gene expression. The cut-off criteria were $R > 0.6$ and $P < 0.05$.

Differentially-expressed genes analysis, receiver operating characteristic (ROC) curve analysis, and survival analysis

The differentially-expressed genes were identified using the "limma" package in R/Bioconductor, and those displaying an overlap region in the Venn diagram were screened and used in matrices construction. The cutoff criteria were log2 fold change (FC) >0 and adjusted $P < 0.05$. Furthermore, survival analysis was applied and the cutoff criteria were hazard ratio (HR) >1 and adjusted $P < 0.05$.

After identifying the key genes, we used the survival package in the R platform to analyze patient survival in the first, third, and fifth years. Finally, the accuracy of *NCBP2* expression in predicting COAD prognosis was evaluated using time-dependent ROC curves.

Five-fold cross-validation and logistic regression analysis

We used a logistic regression model to perform five-

fold cross-validation in the TCGA database to further investigate the efficiency in identifying COAD (6,7). To prevent the size mismatch between the two groups, the 70 percent tumor and control samples were picked randomly to produce a new dataset. Furthermore, we created the logistic regression model using the "scikit-learn" module in Python software (17) (version 3.6, https://www.python.org), and a confusion matrix with two categories (positive and negative) was displayed.

To assess the effectiveness of this classification model, the precision, recall, accuracy, and F1 score were added. The accuracy of the classifier is calculated as the number of true positives divided by the number predicted to be positive. Recall is expressed as the ratio of true positives to actual positives. The fraction of accurately anticipated outcomes in all samples represents accuracy. Since precision and recall interact with each other and their values cannot be ideally large at the same time, we considered the calculation of the F1 score as a combined evaluation metric. The F1 score is the harmonic mean of precision and recall.

Biological function and pathway enrichment analysis

We performed a gene set enrichment analysis (GSEA) with the Kyoto Encyclopedia of Genes and Genomes (KEGG) gene sets as the reference gene set to investigate the potential mechanism between different *NCBP2* expression groups in cell lines.

Cell culture and transient transfection

Human COAD cell lines (SW620, HT-29, HCT116) and normal intestinal epithelial cell line HIEC-6 from American Technical Ceramics Corporation (ATCC, https://www.atcc.org/) were respectively cultured in F12 and Dulbecco's modified Eagle's medium (DMEM, Gibco, Carlsbad, CA, USA) containing 10% fetal bovine serum (FBS). All cell lines were maintained in a humidified incubator at 37 ℃ and 5% carbon dioxide (CO_2). We transfected the negative control (NC) and *NCBP2* small interfering RNAs (siRNAs, Sigma-Aldrich, USA) into COAD cells using Lipofectamine 2000 (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's recommendations. The target sequences of the siRNAs were UUCUCCGAACGUGUCACGUdTdTA CGUGACGUGGAGAAdTdT (NCsi), GGGUGACAAU GAAGAACAAdTdTUUGUUCUUCAUUGUCACCCdT dT (*NCBP2*-si1), and GCUGAAGAACAUCACCUAAdTd TUUAGGUGAUGUUCUUCAGCdTdT (*NCBP2* -si2).

Western blot

Total proteins were extracted using lysis buffer containing protease inhibitors, and the concentration was calculated. Polyvinylidene fluoride (PVDF) membranes (Millipore, Bedford, MA, USA) containing proteins separated on Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis (SDS-PAGE) gels were blocked with 5% non-fat dry milk and incubated with primary antibodies for overnight at 4 ℃. Enhanced chemiluminescence (ECL) detection reagent (Millipore) was used to identify immunoblots according to the manufacturer's instructions. The antibodies used were as follows: β-ACTIN (A1978, Sigma-Aldrich, RRID: AB_476692), glyceraldehyde-3-phosphate dehydrogenase (GAPDH, Santa Cruz Biotechnology, USA, sc-47724), and *NCBP2* (Abcam, UK, ab91556).

Cell proliferation assay and cell cycle analysis

The Cell Counting Kit-8 (CCK-8) assay was used to determine the cell viability, and then the cells were collected and counted. In 96-well plates, we plated the specific number of tumor cells and treated with different concentrations of temozolomide (TMZ) or dimethyl sulfoxide control. After a while, the wells were filled with the CCK-8 solution. COAD cells were detected at 450 nm of absorbance in a microplate reader (Shenzhen Rongjin Technology Co., Ltd., Shenzhen, China). The cells were fixed in 70% ethanol overnight at 4 ℃ and stained with RNase A containing Propidium Iodide for cell cycle analysis (Sigma-Aldrich). Flow cytometry and the ModFit software (Verity Software House, Inc., USA) were used to determine the cell cycle distribution. The si-NC, si-*NCBP2*, and blank control groups were created.

Statistical analysis

Kaplan-Meier analysis for overall survival was carried out based on the gene expression level using GraphPad Prism 8 software (GraphPad Corporation, USA) and the log-rank test, with the cut-off level set at the median value. The link between clinicopathological parameters of COAD patients was investigated using Chi-squared analysis in line with the varied expression levels of *NCBP2*. *NCBP2* expression was significantly linked with age (median value). The ggplot2 package on the R platform was used to produce volcano charts. Statistical significance was defined as $P < 0.05$.

Results

Identification of COAD-related genes after data integration

At present, many tumor treatment drugs kill tumor cells and at the same time greatly affect the survival of normal cells. Therefore, we sought to identify tumor cell-targeting drugs that are less lethal to normal cells. We also found that tumor patients have significantly increased gene expression due to increased CNs, which is not common in normal people. So, we aimed to identify genes that are CN driven in tumor patients and those in which tumor cells become more sensitive to knockdown when these genes are highly expressed.

We collected gene expression, CN, and DS data from the Cancer Cell Line Encyclopedia (CCLE), and linear regression analyses were performed. Through the linear regression analyses of DS and gene expression and CN and gene expression, we identified 340 and 855 related genes, respectively (*Figure 1A,1B*). We then conducted survival and differential expression analyses and found that *NCBP2* met our criteria (*Figure 1C-1E*).

Identification of key genes for diagnosis of COAD

We conducted ROC curve analysis and compared the area under the curve (AUC) values to evaluate the sensitivity and specificity of *NCBP2* for the diagnosis of COAD in TCGA database. From the ROC curve of *NCBP2*, a mean AUC value of (0.940±0.050) (*Figure 2A*) was obtained. The average values of accuracy, precision, recall, F measure, and AUC were 0.90, 0.92, 0.97, 0.95, and 0.89, respectively (*Figure 2B*). These findings demonstrated that *NCBP2* performed well in distinguishing COAD samples from normal controls.

Survival analysis of NCBP2 and ROC curve in COAD

Survival analysis was conducted in the distinct *NCBP2* expression groups. High *NCBP2* expression was shown to be associated with a worse prognosis and a higher chance of cancer-related mortality. Compared to individuals with low *NCBP2* expression, those with high *NCBP2* expression had a lower median survival time (*Figure 3A,3B*). According to a time-dependent ROC curve, the *NCBP2* gene expression level had predictive significance for the long-term survival of COAD patients. The AUC of the ROC curves of 1-, 3-, and 5-year survival were 0.532, 0.608, and 0.660, respectively (*Figure 3C*).

NCBP2 was highly expressed in COAD and increased COAD cell proliferation

We then detected the protein expression of *NCBP2* by western blot and the results were similar to the messenger RNA (mRNA) expression in COAD tissues. According to the different *NCBP2* expression levels, the correlation between several clinicopathological characteristics of oral squamous cell carcinoma (OSCC) patients was explored (*Table 1*). High expression of *NCBP2* (median value) was statistically associated with aneuploidy score (P=1.22E-06), fraction genome altered (P=1.23E-09), tumor stage (P=0.01375), and T stage (P=0.01378).

NCBP2 was highly expressed in the COAD cell lines (SW620, HT-29, and HCT116) compared to the normal intestinal epithelial cell line (HIEC-6) (*Figure 4A*). The CCK8 results in COAD cells were considerably decreased following *NCBP2* elimination (*Figure 4B,4C*).

Genes that are positively associated with NCBP2 are enriched in the cell cycle

To further investigate the potential biological role of *NCBP2* in COAD, we used R/Bioconductor tools to perform a KEGG pathway analysis on the identified genes. The KEGG pathway analysis revealed a link between *NCBP2* and the cell cycle (*Figure 5A*). *NCBP2* knockdown caused COAD cells to arrest in the G0/G1 phase, demonstrating the function of *NCBP2* in cell cycle progression regulation (*Figure 5B,5C*).

Tridihexethyl inhibited COAD cell proliferation via NCBP2

To identify the targeted small-molecule drug for *NCBP2*, we also used Cmap for drug screening (18,19). Based on the Cmap results (Table S1), Tridihexethyl was identified as the most potent medicine for the treatment of COAD. Therefore, we focused the rest of our research on it.

We analyzed the cell proliferation rate using the CCK8 assay and found that it was decreased in Tridihexethyl relative to the empty vector + sh-control SW620 cells, but this was abrogated by *NCBP2* overexpression (Tridihexethyl + pcDNA-*NCBP2*) (*Figure 6*), suggesting that Tridihexethyl inhibited COAD cell proliferation via *NCBP2*.

Discussion

Clinically, COAD is an incurable tumor and is the third
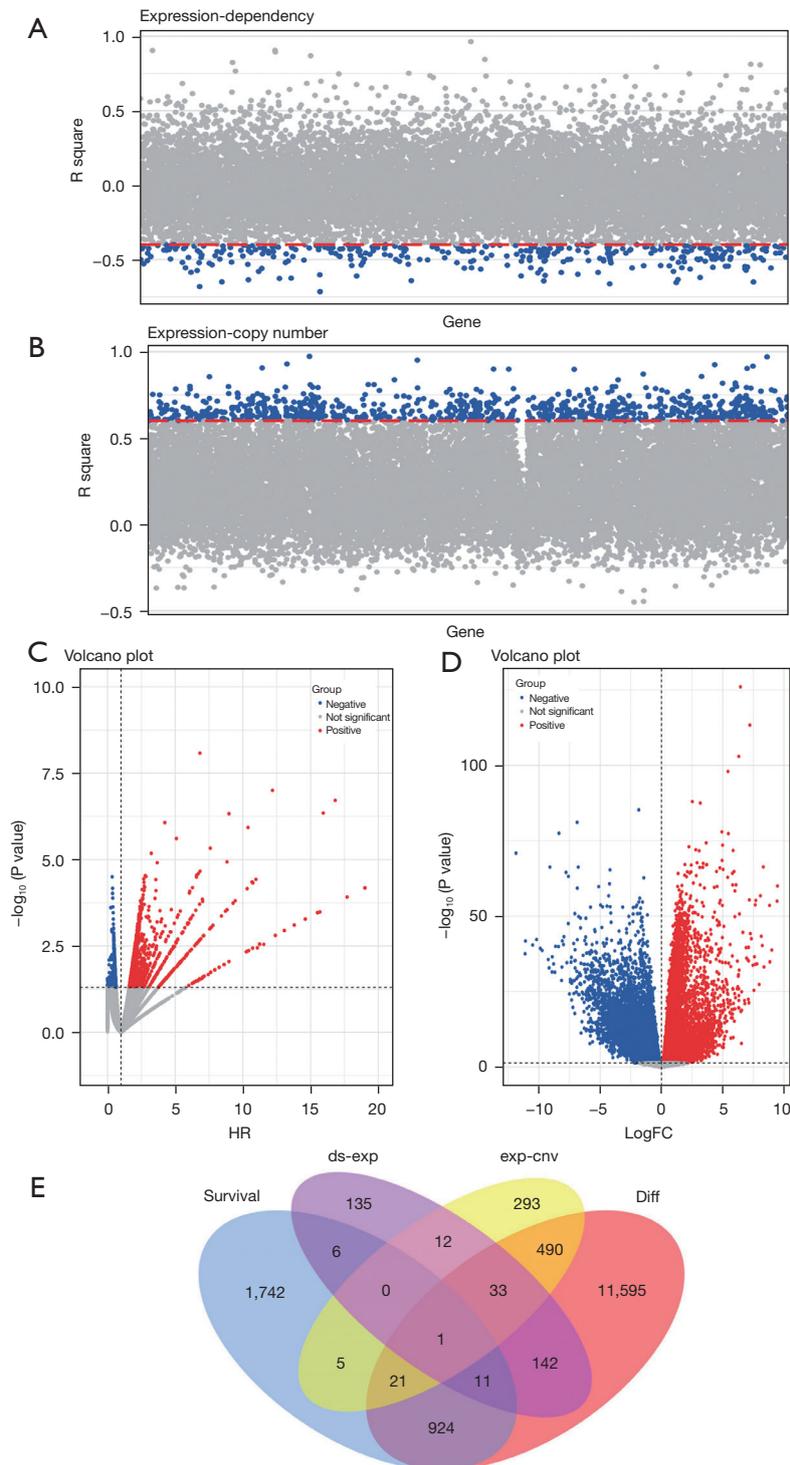
Figure 1 Identification of COAD-related genes after data integration. (A) Linear regression analysis of expression and dependency score; (B) linear regression analysis of expression and copy number; (C,D) volcano plots of gene expression. Red/blue symbols classify the upregulated/ downregulated genes according to the criteria above; (E) Venn diagram was used to screen *NCBP2*. COAD, colon adenocarcinoma; *NCBP2*, nuclear cap binding protein subunit 2.
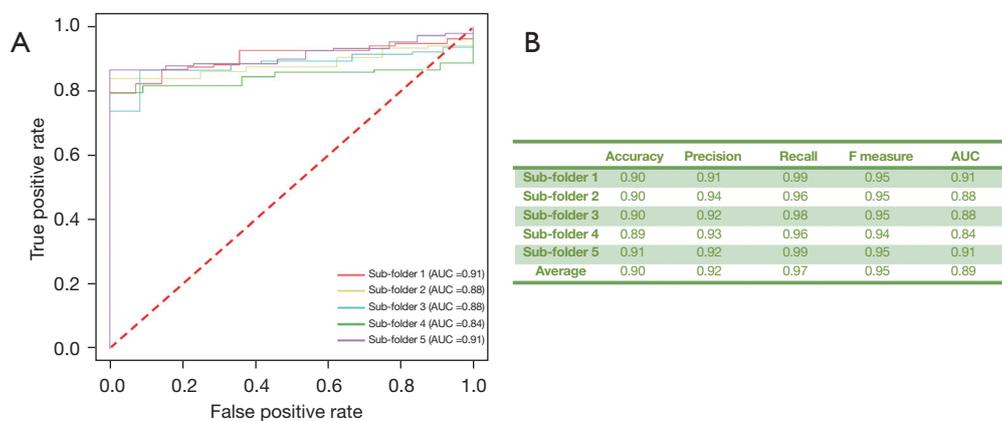
Figure 2 Identification of key genes for COAD diagnosis. (A) ROC curve of *NCBP2* in the five-fold cross-validation; (B) evaluation metrics of each fold. All data are represented by mean ± SD. AUC, area under the curve; COAD, colon adenocarcinoma; ROC, receiver operating characteristic; *NCBP2*, nuclear cap binding protein subunit 2; SD, standard deviation.
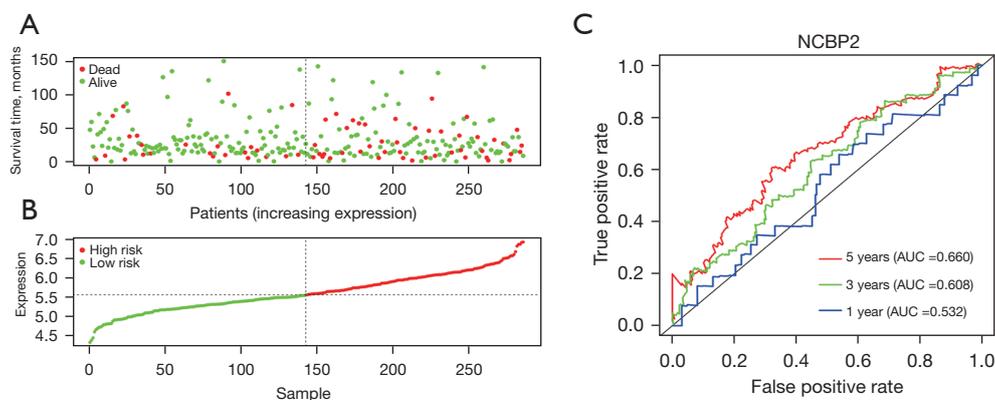
Figure 3 Survival analysis of *NCBP2* and ROC curve in COAD. (A,B) Expression of *NCBP2* and COAD patients' survival time scatter gram; (C) time-dependent ROC curves of *NCBP2* expression in predicting COAD OS. *NCBP2*, nuclear cap binding protein subunit 2; AUC, area under the curve; ROC, receiver operating characteristic; COAD, colon adenocarcinoma; OS, overall survival.

most common tumor in the world (20). The onset of COAD is insidious, and thus, patients will only have obvious clinical symptoms in the late stage of onset. Unfortunately, existing clinical measurement strategies are still unable to accurately detect the existence of or predict the development of COAD (21,22). At present, despite the noticeable development of gene targeting therapy, clinically effective strategies against COAD are still needed to improve the conditions of patients and delay tumor progression (23).

Gene-targeted therapy is a commonly used method for the treatment of COAD patients. In current clinical research, the transfer of tumor factors is directly related to genetic changes (24). With the development of bioinformatics methods, researchers have been able to discover interesting biomarkers for COAD, and several relevant papers have been published. In our research, *NCBP2* was found to affect the progress of COAD.

Our research aimed to identify potential oncogenes with diagnostic and prognostic significance. *NCBP2* expression was confirmed in TCGA database and clinical samples, demonstrating that *NCBP2* was overexpressed in tumorous tissues compared to normal tissues. Survival analysis was conducted to further investigate the role of *NCBP2* in COAD. The biological roles of *NCBP2* in COAD cells were investigated, and we found that higher *NCBP2* expression

Table 1 Correlations between NCBP2 and clinicopathological characteristics

| Characteristics	NCBP2 expression		P value
	High	Low	
Age, years			0.8604
>50	192	196	
≤50	31	29	
New neoplasm event post initial therapy indicator			0.2989
Yes	149	148	
No	46	34	
Aneuploidy score			1.22E-06
High	138	83	
Low	86	135	
Buffa hypoxia score			2.06E-02
High	68	74	
Low	89	54	
Fraction genome altered			1.23E-09
High	138	77	
Low	74	141	
MSI MANTIS score			0.002603
High	91	121	
Low	123	89	
MSI sensor score			0.05703
High	99	121	
Low	121	101	
Tumor stage			0.01375
I–II	111	141	
III–IV	106	82	
T stage			0.01378
T1–2	46	43	
T3–4	177	181	
M stage			0.01375
M0	149	177	
M1	39	22	
N stage			0.01378
N0	118	148	
Not N0	105	77	

NCBP2, nuclear cap binding protein subunit 2; MSI, microsatellite instability.
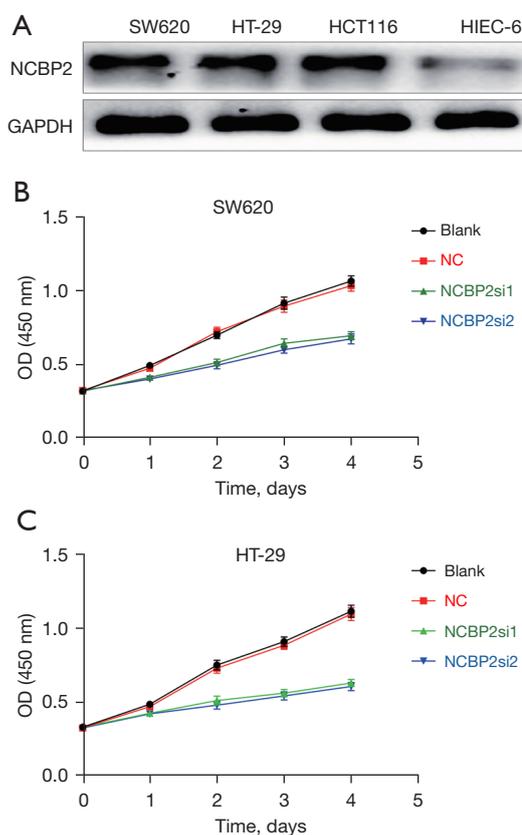
Figure 4 *NCBP2* was highly expressed in COAD and increased COAD cell proliferation. (A) *NCBP2* was highly expressed in COAD cell lines (SW620, HT-29, and HCT116) compared to the normal intestinal epithelial cell line (HIEC-6); (B) OD values of the SW620 cell line after *NCBP2* knockdown; (C) OD values of the HT-29 cell line after *NCBP2* knockdown. *NCBP2*, nuclear cap binding protein subunit 2; GAPDH, glyceraldehyde-3-phosphate dehydrogenase; OD, optical density; NC, negative control; COAD, colon adenocarcinoma.

was significantly associated with a poor prognosis in COAD patients in independent cohorts.

Based on previous research, uncontrolled cell proliferation is a hallmark of cancer (25), which is caused by abnormal cell cycle regulatory genes (26). A cancer study has shown that several cell cycle inhibitors have anticancer properties (27). The oncogenic function of *NCBP2* in COAD development was first discovered in our work. Depletion of *NCBP2* expression resulted in G0/G1 phase arrest and suppression of COAD cell growth, and cell cycle blockage was a result of the interaction between *NCBP2* and cell cycle-related molecules.
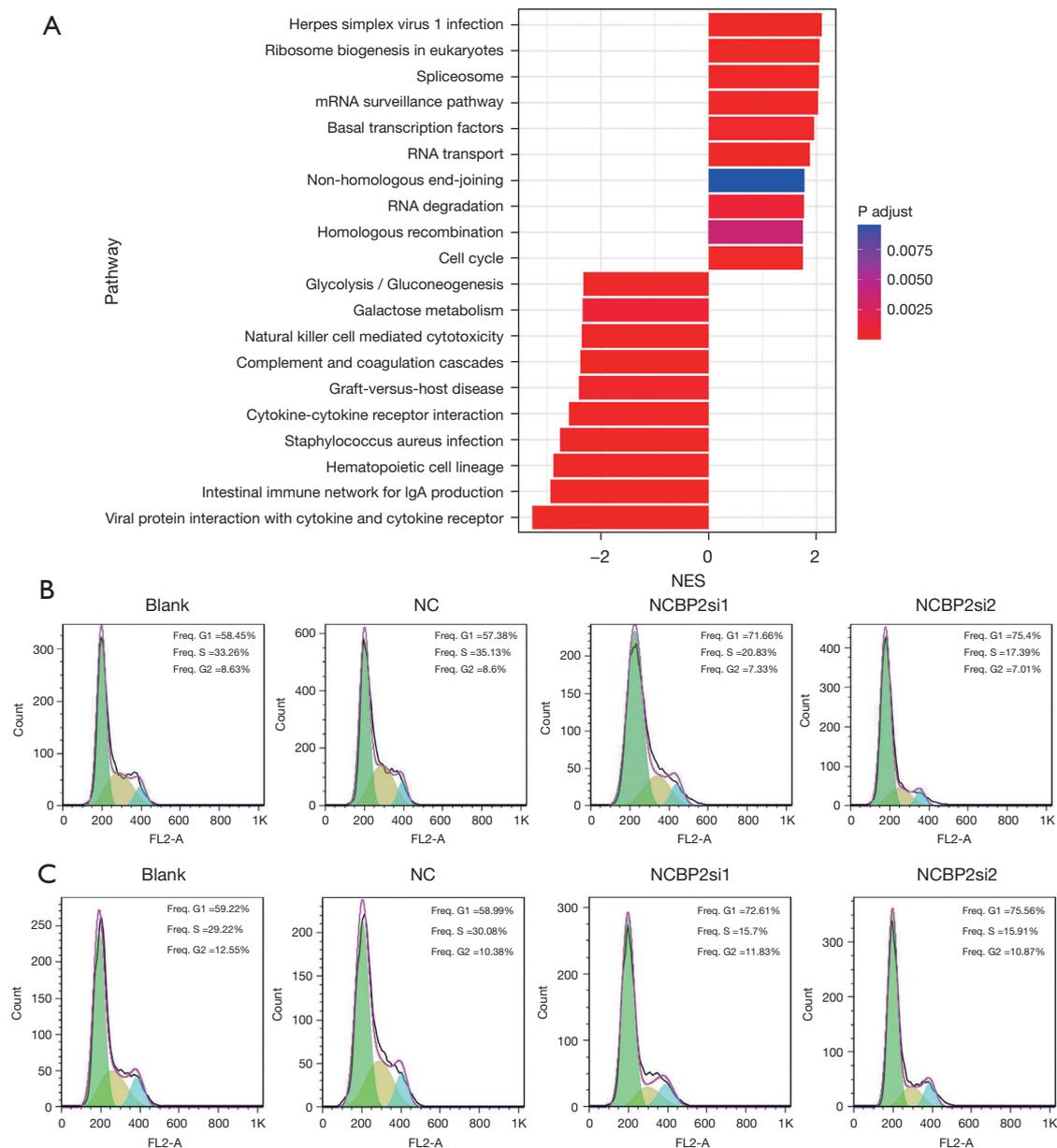
Figure 5 Genes that are positively associated with *NCBP2* are enriched in the cell cycle. (A) KEGG pathway enrichment analysis of genes that are positively associated with *NCBP2*; (B) representative images of Blank, NC siRNA-, and *NCBP2* siRNA-transfected HT-29 cells were analyzed in the cell cycle assay; (C) representative images of Blank, NC siRNA-, and *NCBP2* siRNA-transfected SW620 cells were analyzed in the cell cycle assay. NES, normalized enrichment score; NC, negative control; *NCBP2*, nuclear cap binding protein subunit 2; KEGG, Kyoto Encyclopedia of Genes and Genomes.

The tumor microenvironment (TME) plays a crucial role as both a positive and negative regulator of cancer hallmarks, and gaining a better understanding of the TME might lead to new insights into tumor development and prognosis. Immunohistochemical research, for example, demonstrates that tumors substantially infiltrated by tumor-infiltrating lymphocytes are linked to a better patient prognosis (28). Patients with a high level of *NCBP2* expression may have a poor response to immunotherapy, highlighting the potential prognostic utility of *NCBP2* for
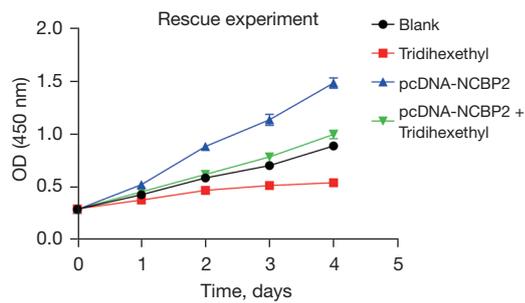
Figure 6 Tridihexethyl induces apoptosis by increasing DNA strand breaks. Human colon cell lines SW620 were cultured for 96 h with exemestane at the indicated concentrations, and cell growth was determined by the OD values in each group. OD, optical density; *NCBP2*, nuclear cap binding protein subunit 2.

COAD prognosis. Furthermore, the link between the TME and *NCBP2* may affect the curative efficacy of radiotherapy in COAD. Nevertheless, further research is required to fully understand the network of interactions between *NCBP2* and invading stromal and immune cells.

Non-homologous end joining and homologous recombination are the two most common strategies to repair DNA double-strand breaks (29,30), which are the most harmful type of DNA damage. A multitude of factors determine the course that should be taken, and the cell cycle is one of them. Effective DNA end processing is confined to the S and G2 phases and is regulated by cyclin-dependent kinases (CDK) activity (31,32). In addition, cells include checkpoints that stop cell cycle progression in the event of DNA damage, providing an opportunity for DNA repair (33). The activity of particular CDK complexes is influenced by several DNA damage checkpoints. Moreover, *NCBP2* is thought to be a DNA damage sensor (34), and the positive association between *NCBP2* and its function in DNA damage repair procedures might alter cell cycle progression in COAD. In this study, we identified G0/G1 phase arrest following the reduction of *NCBP2* expression. However, additional research is needed to validate *NCBP2* as a possible target for the treatment of COAD, since it may alter the impact of radiation and enhance cellular resistance to medicines via DNA damage repair pathways.

Our research combined COAD microarray data with relatively large samples from TCGA database and used GSEA to explore *NCBP2*'s potential biological role and clinicopathological importance in COAD. In conclusion, *NCBP2* displayed a definite increasing trend in COAD, and

its elevated expression caused aberrant COAD cell cycle transition. As a result, inhibiting *NCBP2* might be a viable method for COAD prevention and therapy.

However, there are still some limitations in our research that need to be considered carefully. Firstly, the mechanisms of action of *NCBP2* and its upstream and target genes were not thoroughly investigated. Secondly, we did not perform relevant mouse experiments to validate the results of our analysis. Therefore, more studies are needed to validate our conclusions.

Footnote

Reporting Checklist: The authors have completed the STARD and the MDAR reporting checklists. Available at https://jgo.amegroups.com/article/view/10.21037/jgo-22-665/rc

Data Sharing Statement: Available at https://jgo.amegroups.com/article/view/10.21037/jgo-22-665/dss

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://jgo.amegroups.com/article/view/10.21037/jgo-22-665/coif). All authors report that this study was funded by the Program of Qiqihar Medical Academy of Sciences (No. QMS12017M-13). YS is an employee of Staidson Biopharmaceuticals Company Limited. The authors have no other conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the

original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: https://creativecommons.org/licenses/by-nc-nd/4.0/.

References

1. Sterpetti AV, Sapienza P. Adenocarcinoma in the transposed colon: High grade active inflammation versus low grade chronic inflammation. Eur J Surg Oncol 2019;45:1536-41.

2. Boyce J, Tawagi K, Cole JT. Primary colon adenocarcinoma with choriocarcinoma differentiation: a case report and review of the literature. J Med Case Rep 2020;14:220.

3. Labianca R, Beretta GD, Kildani B, et al. Colon cancer. Crit Rev Oncol Hematol 2010;74:106-33.

4. Wu C. Systemic Therapy for Colon Cancer. Surg Oncol Clin N Am 2018;27:235-42.

5. Angell HK, Bruni D, Barrett JC, et al. The Immunoscore: Colon Cancer and Beyond. Clin Cancer Res 2020;26:332-9.

6. Lichtenstein P, Holm NV, Verkasalo PK, et al. Environmental and heritable factors in the causation of cancer--analyses of cohorts of twins from Sweden, Denmark, and Finland. N Engl J Med 2000;343:78-85.

7. Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. Nature 2012;487:330-7.

8. Bacolod MD, Barany F. Molecular profiling of colon tumors: the search for clinically relevant biomarkers of progression, prognosis, therapeutics, and predisposition. Ann Surg Oncol 2011;18:3694-700.

9. Zhang Y, Zheng Y, Fu Y, et al. Identification of biomarkers, pathways and potential therapeutic agents for white adipocyte insulin resistance using bioinformatics analysis. Adipocyte 2019;8:318-29.

10. Chen YJ, Guo YN, Shi K, et al. Down-regulation of microRNA-144-3p and its clinical value in non-small cell lung cancer: a comprehensive analysis based on microarray, miRNA-sequencing, and quantitative real-time PCR data. Respir Res 2019;20:48.

11. Chen DL, Lu YX, Zhang JX, et al. Long non-coding RNA UICLM promotes colorectal cancer liver metastasis by acting as a ceRNA for microRNA-215 to regulate ZEB2 expression. Theranostics 2017;7:4836-49.

12. Lemuth K, Rupp S. Microarrays as Research Tools and Diagnostic Devices. RNA and DNA Diagnostics 2015:259-80.

13. D'Erchia AM, Atlante A, Gadaleta G, et al. Tissue-specific mtDNA abundance from exome data and its correlation with mitochondrial transcription, mass and respiratory activity. Mitochondrion 2015;20:13-21.

14. Hosmer DW Jr, Wang CY, Lin IC, et al. A computer program for stepwise logistic regression using maximum likelihood estimation. Comput Programs Biomed 1978;8:121-34.

15. Lupski JR. Structural variation in the human genome. N Engl J Med 2007;356:1169-71.

16. Stranger BE, Forrest MS, Dunning M, et al. Relative impact of nucleotide and copy number variation on gene expression phenotypes. Science 2007;315:848-53.

17. Banday AH, Saeed BA, Al-Masoudi NA. Synthesis, Aromatase Inhibitory, Antiproliferative and Molecular Modeling Studies of Functionally Diverse D-Ring Pregnenolone Pyrazoles. Anticancer Agents Med Chem 2021;21:1671-9.

18. Lamb J. The Connectivity Map: a new tool for biomedical research. Nat Rev Cancer 2007;7:54-60.

19. Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin 2021;71:209-49.

20. Müller AC, Guido S. Introduction to machine learning with Python: a guide for data scientists. O'Reilly Media, Inc., 2016.

21. McCarthy B, Casey D, Devane D, et al. Pulmonary rehabilitation for chronic obstructive pulmonary disease. Cochrane Database Syst Rev 2015;(2):CD003793.

22. Gloeckl R, Marinov B, Pitta F. Practical recommendations for exercise training in patients with COPD. Eur Respir Rev 2013;22:178-86.

23. Miao Y, Wang J, Ma X, et al. Identification prognosis-associated immune genes in colon adenocarcinoma. Biosci Rep 2020;40:BSR20201734.

24. Chiang AC, Massagué J. Molecular basis of metastasis. N Engl J Med 2008;359:2814-23.

25. Senga SS, Grose RP. Hallmarks of cancer-the new testament. Open Biol 2021;11:200358.

26. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell 2011;144:646-74.

27. Suski JM, Braun M, Strmiska V, et al. Targeting cell-cycle machinery in cancer. Cancer Cell 2021;39:759-78.

28. Stanton SE, Disis ML. Clinical significance of tumor-infiltrating lymphocytes in breast cancer. J Immunother Cancer 2016;4:59.

29. Ranjha L, Howard SM, Cejka P. Main steps in DNA double-strand break repair: an introduction to homologous recombination and related processes. Chromosoma 2018;127:187-214.

30. Wright WD, Shah SS, Heyer WD. Homologous recombination and the repair of DNA double-strand breaks. J Biol Chem 2018;293:10524-35.

31. Zhang M, Zhang L, Hei R, et al. CDK inhibitors in cancer therapy, an overview of recent development. Am J Cancer Res 2021;11:1913-35.

32. MacKenzie AM, Lacefield S. CDK Regulation of Meiosis:

Lessons from S. cerevisiae and S. pombe. Genes (Basel) 2020;11:723.

33. Wei P, Li Y, Wu L, et al. Serum cortisol levels and adrenal gland size in patients with chronic obstructive pulmonary disease. Am J Transl Res 2021;13:8150-7.

34. Kugeratski FG, Atkinson SJ, Neilson LJ, et al. Hypoxic cancer-associated fibroblasts increase NCBP2-AS2/HIAR to promote endothelial sprouting through enhanced VEGF signaling. Sci Signal 2019;12:eaan8247.

Table S1 Cmap results

Cmap name	Mean	N	Enrichment	P	Specificity
Tridihexethyl	−0.567	4	−0.906	0.00014	0
Ascorbic acid	0.577	4	0.85	0.00072	0
Kaempferol	0.579	4	0.848	0.00076	0.0052
Zimeldine	0.367	5	0.797	0.00084	0
Trimethoprim	−0.302	5	−0.768	0.00124	0
Debrisoquine	0.488	4	0.815	0.00219	0
Bupivacaine	0.321	4	0.782	0.0042	0
Ouabain	0.274	4	0.778	0.00464	0.0909
Stock1n-35874	0.651	2	0.947	0.00499	0.0435
Conessine	−0.381	4	−0.77	0.00567	0
Fluvastatin	0.448	4	0.755	0.00688	0
Brinzolamide	0.403	4	0.754	0.00704	0.0394
Naltrexone	−0.247	5	−0.683	0.00733	0.0393
Terguride	0.402	8	0.549	0.00836	0
Hycanthone	0.459	4	0.742	0.00849	0.0933
Mg-262	0.55	3	0.836	0.00881	0.1946
Parthenolide	0.447	4	0.738	0.00907	0.1916
Imidurea	−0.448	3	−0.834	0.00915	0.0108
Amprolium	−0.449	5	−0.666	0.00969	0.0385
Scopolamine N-oxide	−0.441	5	−0.661	0.01061	0.0427
Dequalinium chloride	0.263	4	0.729	0.01076	0.0073
Tiapride	−0.472	5	−0.661	0.01077	0.0057
Lactobionic acid	−0.23	4	−0.723	0.01196	0.0056
Atracurium besilate	0.537	3	0.815	0.01244	0.0231
Ag-012559	−0.552	3	−0.812	0.01322	0.0163
Fludroxycortide	−0.504	5	−0.645	0.01388	0.0376
Molsidomine	−0.378	4	−0.703	0.01605	0.0129
Loxapine	0.479	4	0.7	0.01661	0.0106
Suxibuzone	−0.425	4	−0.698	0.01765	0.0057
Levodopa	0.472	5	0.635	0.0184	0.0175

Cmap, connectivity map.
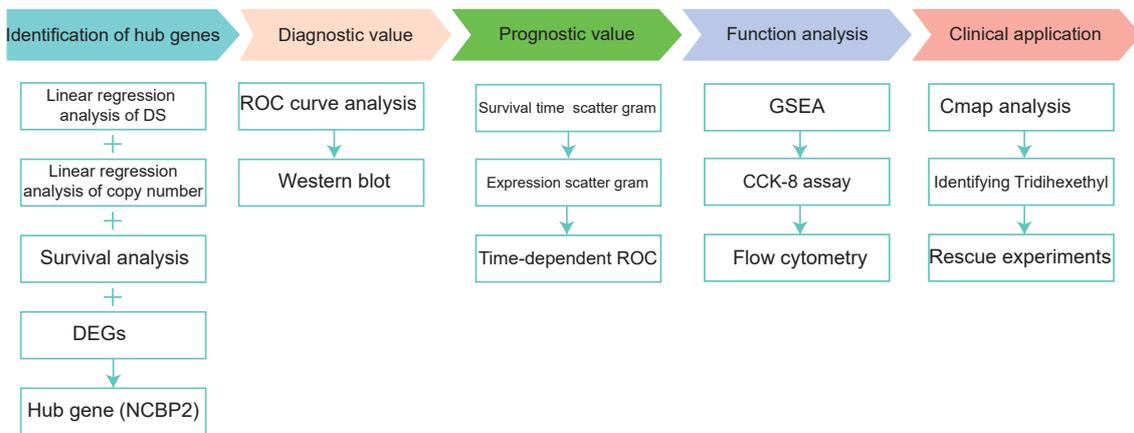
Figure S1 Flow chart. DS, dependency score; DEGs, differentially expressed genes; *NCBP2*, nuclear cap binding protein subunit 2; ROC, receiver operating characteristic; GSEA, gene set enrichment analysis; CCK-8, Cell Counting Kit-8.