参赛队员姓名：吴若予 Ruoyu Wu

中学：南京市金陵中学

省份： 江苏省

国家/地区： 中国

指导教师姓名：杜前明 Qianming Du

指导教师单位： 南京市第一医院

论文题目：A Bioinformatics-Based Screening and Identification of Biomarkers and Therapeutic Target for Colorectal Cancer

A Bioinformatics-Based Screening and Identification of Biomarkers and Therapeutic Target for Colorectal Cancer

Ruoyu Wu and Qianming Du

Abstract

Colorectal cancer (CRC) is one of the most frequently diagnosed cancers worldwide with high mortality rate. Most CRC patients are already at terminal stages when been diagnosed, The first-line therapy for CRC remains as FOLFOX/FOLFIRI chemotherapy, which comes with limited therapeutic effects and serious side effects. Therefore, identification of novel biomarkers and therapeutical targets for CRC is needed. This study was carried out to identify novel key genes for CRC prognosis and therapy. We identified 323 differentially expressed genes (DEGs) between tumor and adjacent normal tissues from three independent CRC datasets from Gene Expression Omnibus (GEO). GO and KEGG pathway enrichment analysis was performed for these DEGs and it showed that the functions of these DEGs involved mostly in the regulation of cell proliferation and metabolism. Among these DEGs, CTHRC1, FXYD5, and SULT2B1 were demonstrated to have prognostic significance for CRC, which could potentially be used as biomarkers for unfavorable prognosis. We also found that these three genes were overexpressed in CRC tumor tissues compared with normal counterparts. Knocking down SULT2B1 in HCT116 and HT29 cells by shRNA suppressed cell proliferation, whereas knocking down CTHRC1 or FXYD5 had no such effects. We next designed Locked Nucleic Acids (LNAs) to knockdown SULT2B1 in HCT116 and HT29 cells, which significantly inhibited cell proliferation. Collectively, we have identified CTHRC1, FXYD5, and SULT2B1 as potential prognostic biomarkers and SULT2B1 as a promising therapeutic target for CRC, and LNAs that we designed could be used for the targeted treatment for CRC.

Key Words: Colorectal cancer; differentially expressed genes (DEGs); sulfotransferase family 2B member 1 (SULT2B1), Locked Nucleic Acid

Table of Contents

Introduction

Colorectal cancer (CRC) is the third most frequently diagnosed cancer and accounts for approximately 10% of all cancer-related death worldwide [1, 2]. Although new treatments for CRC have emerged, including laparoscopic surgery for primary tumors, more aggressive resection for metastatic diseases, radiotherapy for rectal tumor, neoadjuvant chemotherapy and targeted therapies, the outcomes remain dismal in CRC patients [3-5]. Long-term survival for CRC is hardly improved over the past few decades.

In recent years, differentially expressed genes (DEGs) when comparing tumor samples with normal counterparts could be identified through gene expression profiling [6-8]. Still, the number of identified functional genes is limited, and studies are needed to identify more genes to better understand the mechanism underlying the pathogenesis of CRC. Thus, we used bioinformatics tools in this study to identify novel key genes in CRC from three gene expression profiles from GEO database and accessed the clinical relevance of these identified DEGs in the prognosis of the disease. The expression and importance of the DEGs were then tested using clinical samples. And we also explored the possibility of using these DEGs as potential therapeutic targets.

Among the approaches to develop targeted drugs, antisense oligonucleotides have attracted great attention. This is partly due to the fact that many causative proteins for diseases have be facing difficulties to target by conventional methods including low-molecular-weight chemicals as well as antibodies [9]. The antisense nucleic acid strategy is based on the hybridization-base synthesis of oligonucleotide to bind to its complementary mRNA. Locked Nucleic Acid (LNA) is a new nucleic acid analog that has unprecedented properties as a potential oligonucleotide drug, including good affinity, specificity, stability and easy delivery [10, 11].

Due to its RNA-like conformation, LNA is not subject to RNase H cleavage when bound to its target mRNA. Therefore, LNA oligonucleotides are designed as gapmers which typically comprise a central DNA phosphorothioate segment, which recruits RNase H when the gapmer hybridizes with mRNA, flanked at both ends by 2 to 4 LNA monomers. LNA gapmers have shown extraordinary activities in several in vivo animal models for diseases [12, 13]. In this study, we used 17-18 LNA gapmer as the antisense oligonucleotides (as shown in **Figure 1**), which structured as 3 LNAs distributed in each side to enhance the affinity to target sequences, a gap of DNA monomers in the center, which are linked by phosphorothioate (PTO) backbones [14]. The design of LNA gapmers were strictly based on the criterion previously reported by Jaschinski, F. et al [15].

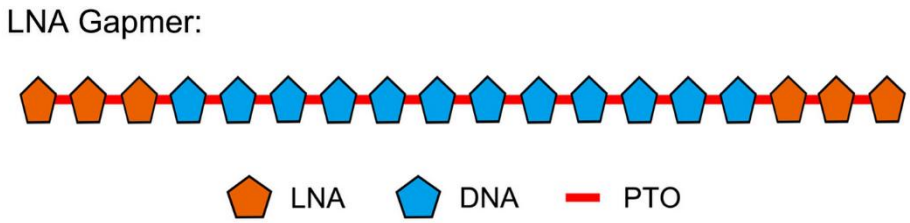


Figure 1. Schematic diagram of locked nucleic acid (LNA) structure, PTO: phosphorothioate

Materials and Methods

Human colorectal cancer tissue microarray and clinical samples

The human colorectal cancer tissue microarrays were obtained from Shanghai Outdo Biotech. Tumor and adjacent normal tissues were obtained from 21 CRC patients from Nanjing First Hospital. Written informed consent files were obtained from all patients recruited in this study.

Cell lines and culture

HCT116 and HT29 cell lines were obtained from the Cell Bank of Shanghai Institute of Biochemistry and Cell Biology. Cells were cultured in DMEM medium (Gibco) with 10% FBS; HT29 cells were cultured in McCoy's 5A medium (Gibco) with 10% FBS.

Cell transfection

Lentivirus plasmid pLVX-shRNA2 encoding shRNA sequences against human CTHRC1, FXYD5, and SULT2B1 were transfected into HEK293T cell along with packaging plasmids PsPAX2 and pMD2.G by using transfection reagent (Beyotime) to generate indicated shRNA lentivirus particles (targeting sequences are listed in **Table S1**). HCT116 and HT29 cells were transfected with lentivirus, and efficiency was verified by western blotting assay.

Supplementary Table S1. shRNA sequences used in this study

Target Gene	shRNA ID	Sequence (5' to 3')
FXYD5	1#	GATCCCGTTGAAAGATACCACGTCCATTCAAGAGATGGACGTGGTATCTTTCAACGTTTTTTG
	2#	GATCCGAGACACACAAGAGCACCAAATTCAAGAGATTTGGTGCTCTTGTGTGTCTCTTTTTTG
SULT2B1	1#	GATCCCCAGAAGTTGCCAGGTGAATATTCAAGAGATATTCACCTGGCAACTTCTGGTTTTTTG
	2#	GATCCGATCGAGATCATCTGCTTAATTTCAAGAGAATTAAGCAGATGATCTCGATCTTTTTTG
CTHRC1	1#	GATCCCGCATCATTATTGAAGAACTATTCAAGAGATAGTTCTTCAATAATGATGCGTTTTTTG
	2#	GATCCGCGTTGGTATTTCACATTCAATTCAAGAGATTGAATGTGAAATACCAACGCTTTTTTG

For LNA knockdown of SULT2B1, 2 μmol/L LNAs were transfected into HCT116 or HT29 for 72 h without assistance of transfection reagent, and knockdown efficiency was verified by western blotting assay (LNA sequences are listed in **Supplementary Table S2**).

Supplementary Table. S2 LNA sequence used in this study

LNA Name	Length	Sequence (5' to 3')
LNA 1#	17	+T*+T*+C*T*T*C*A*G*C*T*C*C*A*A*+G*+G*+C

LNA 2#	17	+A*+C*+A*C*C*C*G*A*C*C*A*G*T*T*+C*+C*+T
LNA 3#	18	+G*+T*+G*C*G*G*G*A*C*G*A*C*G*A*C*+A*+T*+C
LNA 4#	18	+G*+G*+C*G*A*C*T*G*G*A*A*G*A*A*C*+C*+A*+C
Ctrl Oligo	18	+C*+G*+T*T*T*A*G*G*C*T*A*T*G*T*A*+C*+T*+T

+ represent LNA-modified nucleotides and * represent phosphorothioated backbones.

Bioinformatics analysis of online dataset

mRNA expression matrix and clinical features of three colorectal cancer GEO datasets (GSE81582, GSE89076, and GSE106582) were downloaded from https://www.ncbi.nlm.nih.gov/geo/, differential expression genes (DEGs) were identified by limma package of R (parameter setting: $\log_2$FC >1, *P* value <0.05). Integration analysis of DEGs identified from the three GEO datasets was carried out by RobustRankAggreg package of R [16], and visualized by pheatmap package of R. GO and KEGG enrichment analysis and visualization of DEGs were carried out by clusterProfiler package of R. mRNA expression matrix and clinical features of TCGA-COAD dataset was downloaded from https://portal.gdc.cancer.gov/, and survival analysis was carried out by survival package of R, visualized by survminer package of R.

Western blotting

For tissue samples from CRC patients, 100 mg tissue was cut into 1-mm pieces and homogenized in RIPA lysis buffer (Beyotime). Total protein lysate was obtained by centrifugation (12000 x *g*, 5 minutes) and qualified by BCA method with a commercial kit (Beyotime). Lysates were adjusted to 30 μg protein for each well were separated by SDS-PAGE gel, then transferred to PVDF membrane for immunoblotting analysis using the following primary antibodies obtained from Proteintech: anti-CTHRC1, anti-FXYD5, and anti-SULT2B1 and anti-β-actin (Bioworld). Then, PVDF membranes were incubated with goat anti-mouse IgG or goat anti-rabbit IgG as secondary antibodies. Protein bands were analyzed using chemiluminescence system (Tanon). The intensity of bands was analyzed by Image J software.

Real-time PCR

For tissue samples from CRC patients, 100 mg tissue was cut into 1-mm pieces and homogenize by a glass homogenizer in 1 mL Trizol reagent (Life Technologies) under ice bath condition. Total RNA was qualified by a micro-spectrophometer. cDNA were synthesized from total RNA templates via reverse transcription reaction by EasyScript® First-Strand cDNA Synthesis Super Mix (Transgen) according to the manufacturer's instruction. mRNA expression detection was carried out by LightCycler 96 real-time PCR system (Roche) using AceQ qPCR SYBR Master

Mix (Vazyme) in a 20 μL reaction system. The primer sequence of targeted genes used in this study were listed in **Table S3**.

Supplementary Table S3. Primers sequences used in this study

Gene symbol	Forward (5' to 3')	Reverse (5' to 3')
FXYD5	CTCACCATCGTTGGCCTGATT	TCCATGATAGTTGAGTCTGCTGA
SULT2B1	GTTGCCAGGTGAATACTTCCG	CCCGCACATCTTGGGTGTT
CTHRC1	CAATGGCATTCCGGGTACAC	GTACACTCCGCAATTTTCCCAA

Colony formation assay

For shRNA related experiments, HCT116 and HT29 cells with anti-CTHRC1 1-2#, anti-FXYD5 1-2#, and SULT2B1 1-2# transfected were seeded in 6-well dishes. After 24 h culture, cells were replaced with fresh medium and continually cultured for another 14 days. For LNA related experiments, HCT116 and HT29 cells were seeded in 6-well plates. After 24 h culture, fresh medium containing 1 μmol/L LNAs (1-4#) was added and cultured for another 14 days.

At the end of culture, cells were washed three times with PBS, and fixed with 4% paraformaldehyde for 15 min, then washed three times with PBS. Cells were stained with 0.1% crystal violet staining solution for 3 min and then washed three times with PBS, dried in the air at room temperature. Counting for colony was carried out under an inverted microscope. Images for colony were captured by a digital camera.

MTT assay

HCT116 and HT29 cells were seeded in 96-well plates with a density of 5 x 10^3 cells/well. After 24 h culture, cells were replaced with fresh medium contained 5, 2.5, 1.25, 0.625, 0.3125, and 0.15625 μmol/L LNAs and incubated for another 72 h.

At the end of culture, 20 μL methylthiazolyldiphenyl-tetrazolium bromide (MTT, 5 mg/mL) were added to each well, and incubated at 37℃ for 4 h. Supernatant was removed and 150 μL DMSO was added to each well, shocking for 10 min to dissolve MTT. The optical density (OD) for each well was detected at 570 nm by a microplate reader. Cell suppression ratio was calculated as: $(1\text{-}OD_{\text{treatment well}}/OD_{\text{control well}}) \times 100\%$.

Statistical analysis

Quantitative data are presented as mean ± SD, and categorical data are presented as frequency. Statistical analysis was performed by using SPSS software. One way ANOVA was used for multiple group comparison. Paired samples t-test was used for paired quantitative variable

comparison between two groups; independent samples t-test was used for unpaired quantitative variable comparison between two groups. McNemar's test was used for paired categorical variable comparison between two groups; Chi-square test was used for unpaired categorical variable comparison between two groups. *P* value < 0.05 was considered as statistically significant.

Results

1. Identification of differentially expressed genes (DEGs)

1.1 DEGs have been identified from three GEO datasets of colorectal cancer

We analyzed three independent clinical datasets of colorectal cancer (CRC) from GEO, and identified 701 upregulated and 835 downregulated genes in CRC tumor tissues of GSE81582, 1457 upregulated and 2572 downregulated genes in GSE89076, and 212 upregulated and 363 downregulated genes in GSE106582, when compared with adjacent normal tissues (**Figure 2**).

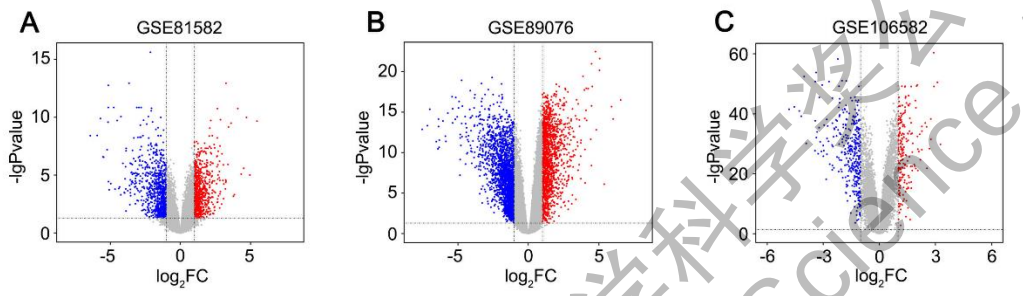


Figure 2. A-C, gene expression profiles of colorectal tumor tissues which is normalized to adjacent normal tissues; data are downloaded from GEO database, and the GEO accession IDs are: GSE81582, GSE89076, and GSE106582; FC, fold change; red dots represents up-regulated genes ($log_2FC > 1$, P value < 0.05), and blue dots stands for down-regulated ones ($log_2FC > -1$, P value < 0.05).

1.2 Integration analysis of DEGs identified from GEO datasets

To integratedly analyze the DEGs identified from these three cohorts, Robust Rank Aggregation algorithm was used and 94 upregulated and 229 downregulated genes were identified with significance scores < 0.05 (**Figure 3**).

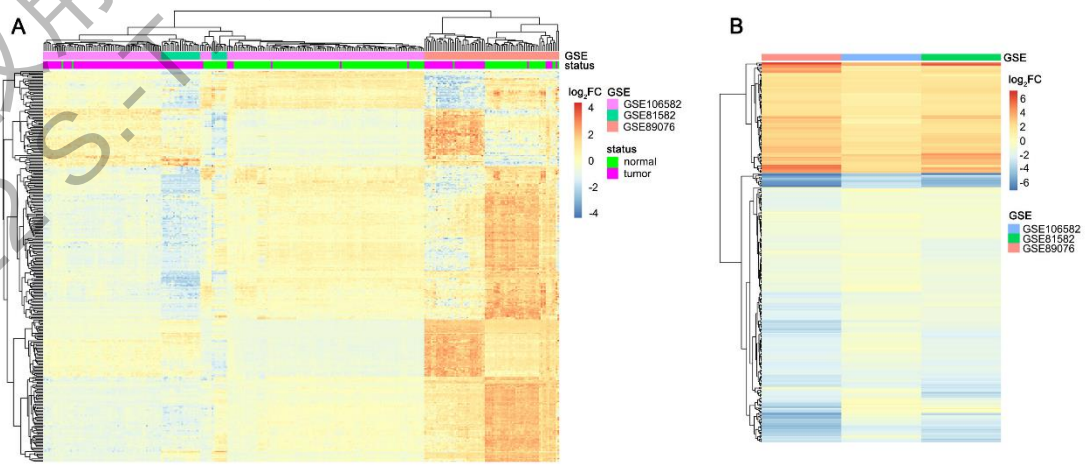


Figure 3. A, heat map for DEGs identified by Robust Rank Aggregation analysis of the three datasets (significant score < 0.05, and 323 DEGs have been identified). **B**, Fold change heat map for DEGs identified by Robust Rank Aggregation analysis.

1.3 Functional enrichment analysis of identified DEGs

We next run Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genome (KEGG) enrichment analysis for these 323 identified DEGs, and it was found the biological functions of the DEGs mainly involved in the regulation of cell proliferation and metabolism, which are critical features for tumor cell growth and survival during cancer progression (**Figure 4**).

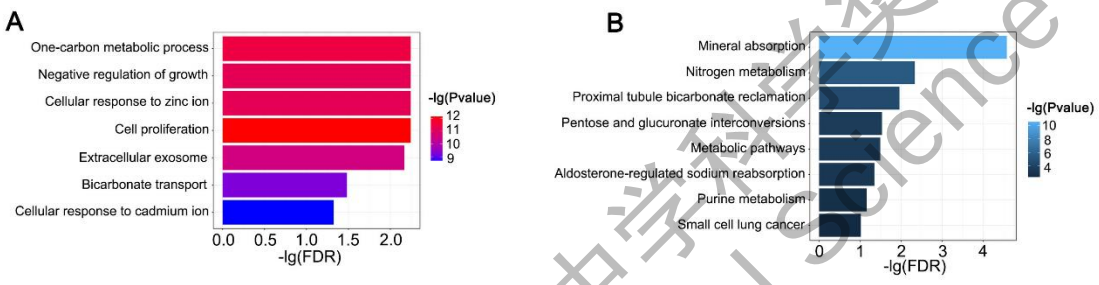


Figure 4. A, GO enrichment analysis. **B**, KEGG enrichment analysis. FDR, false discovery rate.

2. Three DEGs have prognostic significance in colorectal cancer

2.1 Survival analysis of colorectal cancer patients from TCGA database according to DEGs mRNA expression

To discover the prognostic significance of the DEGs we identified, the correlation between DEG expression and survival of CRC patients from TCGA-COAD dataset was accessed. We found that the mRNA levels of three genes, namely collagen triple helix repeat containing 1 (CTHRC1), FXYD domain containing ion transport regulator 5 (FXYD5), and sulfotransferase family 2B member 1 (SULT2B1), are significantly correlated with the outcomes of CRC patients' survival (patients with high mRNA expression have poorer overall survival, while those with low expression of these three genes have better OS; **Figure 5**).

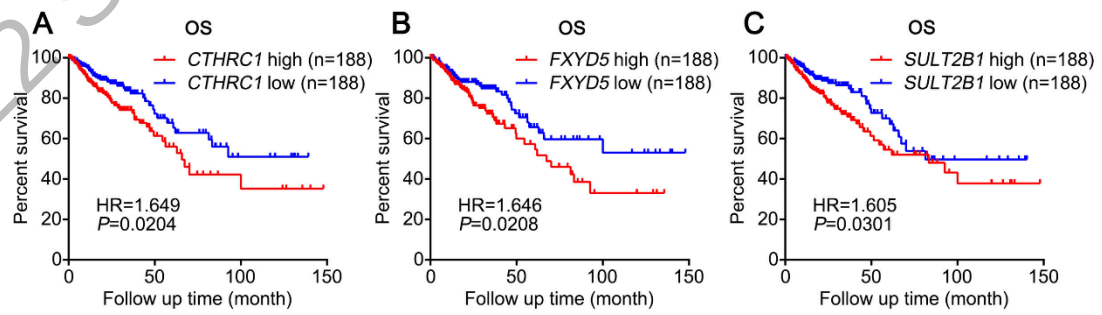


Figure 5. A-C, overall survival (OS) analysis of colorectal cancer patients from TCGA database according to *CTHRC1*, *FXYD5*, and *SULT2B1* mRNA expressions.

2.2 CTHRC1, FXYD5, and SULT2B1 expression validation and prognostic value analysis in CRC

CRC tissue microarray (containing tissue sections of pairs of tumor and adjacent tissues from 83 CRC patients) was next used to look at the protein expressions of CTHRC1, FXYD5, and SULT2B1 in human CRC tissue samples, and we found that the expression of these genes are reversed correlated with survival outcomes of patients, with higher gene expression correlates with worse overall survival and disease free survival **(Figure 6**), indicating that the expressions of CTHRC1, FXYD5, and SULT2B1 may have prognostic significance for CRC.

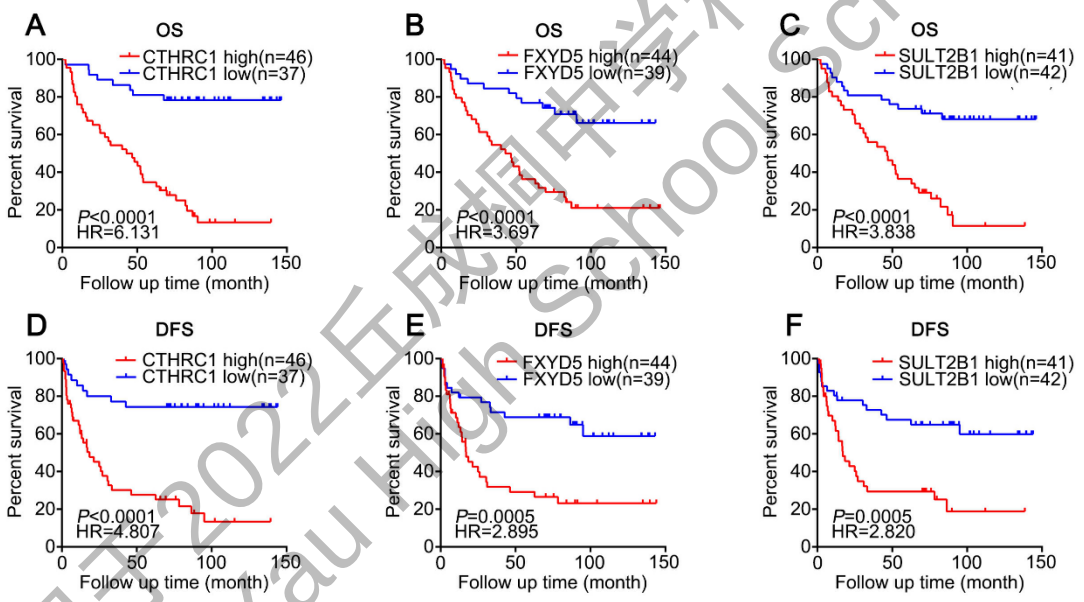


Figure 6. A-C, Overall survival (OS) analysis of colorectal cancer patients from clinical tissue microarray corresponds to CTHRC1, FXYD5, and SULT2B1 expressions. **D-F**, disease-free survival (DFS) analysis of colorectal cancer patients from clinical tissue microarray corresponds to CTHRC1, FXYD5, and SULT2B1 expressions. HR, Hazard ratio.

3. CTHRC1, FXYD5, and SULT2B1 are identified as prognostic biomarkers for colorectal cancer

3.1 Cox regression analysis of CTHRC1, FXYD5, and SULT2B1

Given that the expressions of CTHRC1, FXYD5, and SULT2B1 in tumor tissues are closely

correlated with the outcomes of CRC patients, we next analyzed whether they could be used as prognostic biomarkers for CRC. Cox regression analysis showed that the expression of CTHRC1, FXYD5, and SULT2B1 in tumor tissues are independent risk factors of CRC patients' outcomes (**Table 1** and **Figure 7**).

Table 1. Univariate and multivariate Cox regression analysis of colorectal cancer tissue microarray

Variables	Univariate analysis			Multivariate analysis		
	HR	95% CI	p value	HR	95% CI	p value
CTHRC1	6.47	2.99-13.98	0.000**	2.96	1.22-7.17	0.016*
FXYD5	3.80	1.95-7.37	0.000**	2.71	1.22-6.00	0.014*
SULT2B1	4.07	2.11-7.84	0.000**	2.27	1.13-4.55	0.022*
T stage	1.71	1.08-2.72	0.023*	2.44	1.38-4.30	0.002**
N stage	2.00	1.12-3.57	0.020*	1.27	0.67-2.41	0.463
M stage	7.64	3.93-14.83	0.000**	3.10	1.35-7.09	0.007*
Gender	0.97	0.54-1.74	0.910			
Age	1.00	0.98-1.02	0.809			

HR, Hazard ratio; CI, confidence interval; *, $P < 0.01$; **, $P < 0.01$.

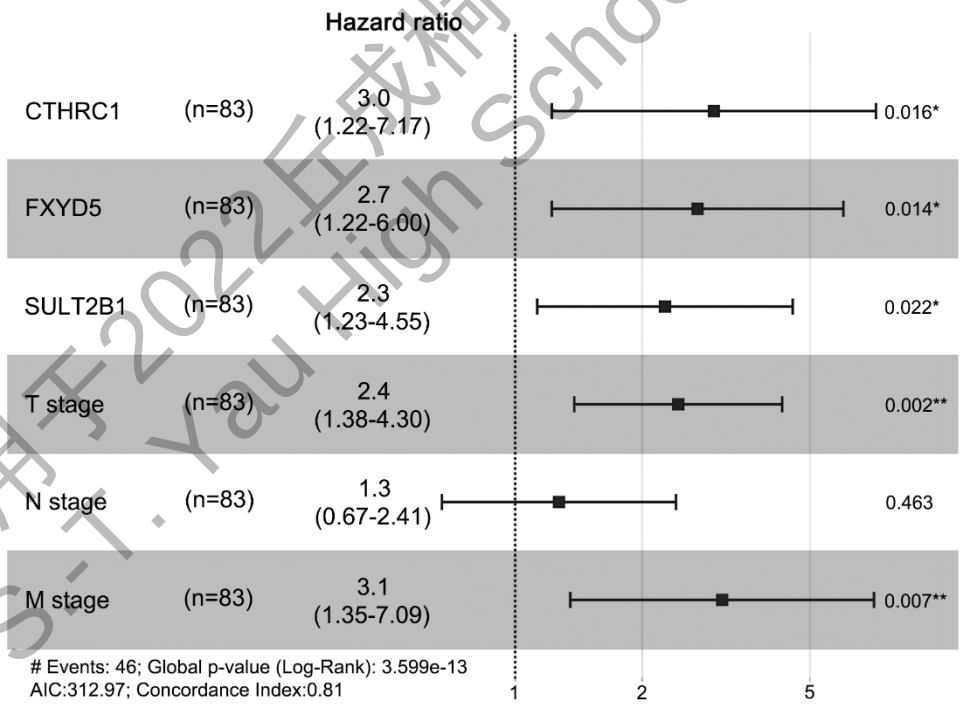


Figure 7. Forest plot for visualization of multivariate Cox regression analysis. *, $P < 0.01$; **, $P < 0.01$.

3.2 Logistic regression analysis of CTHRC1, FXYD5, and SULT2B1

Logistic regression analysis also showed that the expressions of CTHRC1, FXYD5, and

SULT2B1 in tumor tissues are significant risk factors for CRC patients' prognosis, which can be used to predict outcomes of CRC patients (**Table 2** and **Figure 9**). These results above concluded that CTHRC1, FXYD5, and SULT2B1 can be used as prognostic biomarkers for colorectal cancer.

Table 2. Logistic regression analysis of colorectal cancer tissue microarray

Variables	Crude OR(95%CI)	Adj. OR(95%CI)	Wald's test	LR-test
CTHRC1	17.22 (5.77~51.35)	186.4 (11.84~2934.12)	< 0.001**	< 0.001**
FXYD5	7.65 (2.87~20.37)	47.96 (4.82~476.88)	< 0.001**	< 0.001**
SULT2B1	9.2 (3.34~25.32)	164.47 (9.28~2916.56)	< 0.001**	< 0.001**

OR, odds ratio; LR-test, likelihood ratio test; *, $P < 0.01$; **, $P < 0.01$.

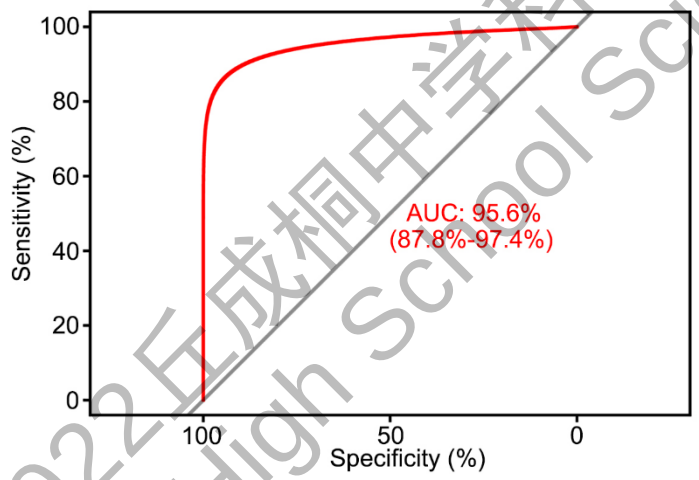


Figure 8. ROC analysis of logistic regression model using CTHRC1, FXYD5, and SULT2B1 expressions for colorectal cancer survival prediction. AUC, area under the curve.

4. CTHRC1, FXYD5, and SULT2B1 are up-regulated in tumor tissues of colorectal cancer

4.1 Comparison of protein levels of CTHRC1, FXYD5, and SULT2B1 in tumor tissues and adjacent normal tissues from clinical samples

We detected protein levels of these three genes in CRC tissue samples. The results showed that protein levels of CTHRC1, FXYD5, and SULT2B1, accessed by Western blot, were up-regulated in tumor tissues when compared with adjacent normal tissues from 21 clinical samples obtained from Nanjing First Hospital (**Figure 9**), which was in line with the results from the CRC tissue microarray.

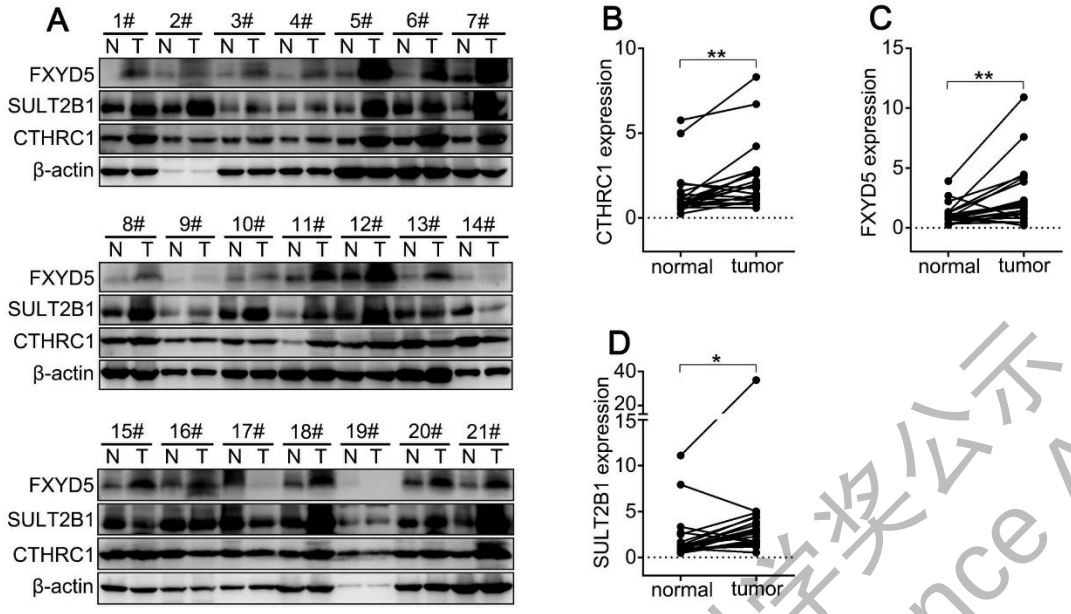


Figure 9. A-D, protein bands and quantification of CTHRC1, FXYD5, and SULT2B1 in adjacent normal and tumor tissues from colorectal cancer patients; N: adjacent normal tissue, T: tumor tissue. *, $P < 0.05$; **, $P < 0.01$.

4.2. Comparison of mRNA levels of CTHRC1, FXYD5, and SULT2B1 in tumor tissues and adjacent normal tissues from clinical samples

We next detected mRNA levels of these three genes in CRC tissue samples. The results showed that mRNA levels of them, accessed by real-time PCR, were up-regulated in tumor tissues when compared with adjacent normal tissues from 21 clinical samples obtained from Nanjing First Hospital (**Figure 10**).

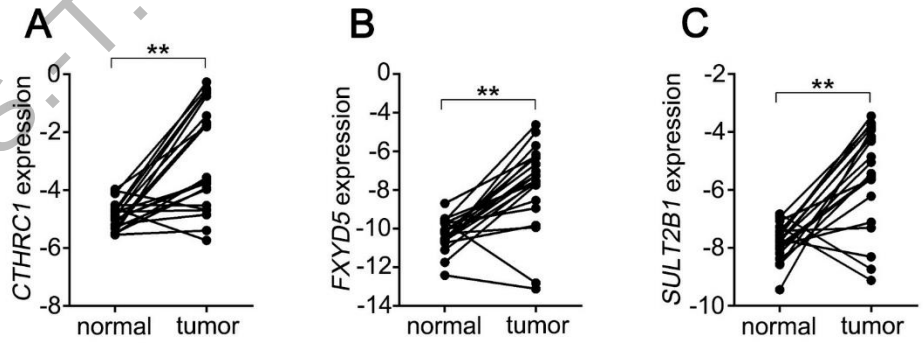


Figure 10. Relative mRNA levels of *CTHRC1*, *FXYD5*, and *SULT2B1* in adjacent normal and

tumor tissues from colorectal cancer patients. *, $P < 0.05$; **, $P < 0.01$.

5. SULT2B1 knockdown impairs colorectal cancer cell proliferation

5.1 SULT2B1 knockdown by shRNA inhibits colorectal cancer cell proliferation

Since we have demonstrated that CTHRC1, FXYD5, and SULT2B1 can be used as biomarkers for CRC prognosis and are up-regulated in CRC, we next checked whether CTHRC1, FXYD5, or SULT2B1 have influences on CRC cell proliferation. We designed shRNAs to knockdown CTHRC1, FXYD5, and SULT2B1 in HCT116 and HT29 cells, and as illustrated in **Figure 11**, the expressions of these three genes were greatly diminished in both cell lines transfected with shRNAs, demonstrating that these shRNAs were transfected with good knocking down efficiency.

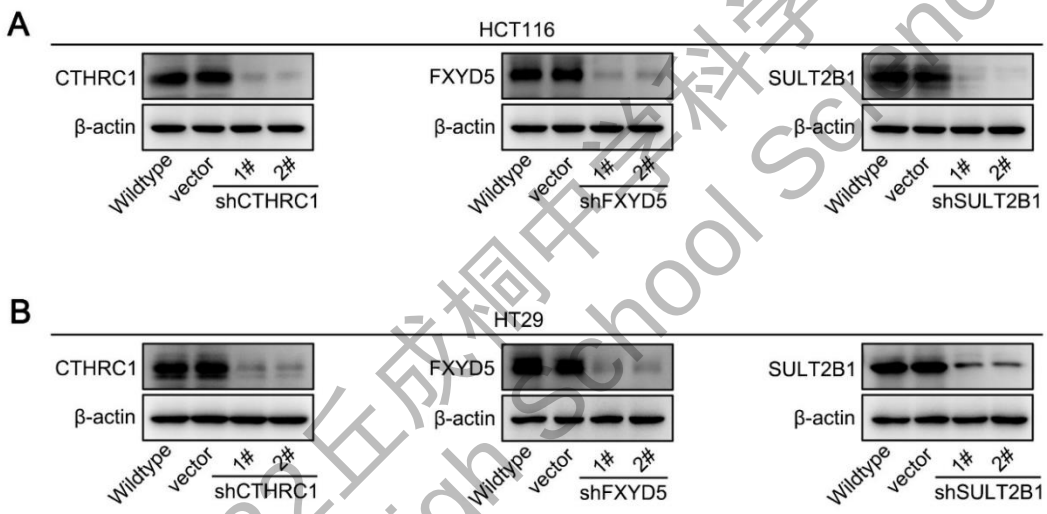


Figure 11. Verification of shRNA induced CTHRC1, FXYD5, and SULT2B1 knockdown in HCT116 and HT29 cells by Western blotting assay.

Colony formation assay was next carried out to determine the influences of CTHRC1, FXYD5, and SULT2B1 knockdown on cell proliferation. It was found that SULT2B1 knockdown by shRNA significantly suppressed colony formation in HCT116 and HT29 cells, whereas CTHRC1 or FXYD5 knockdown has no significant effects (**Figure 12**).

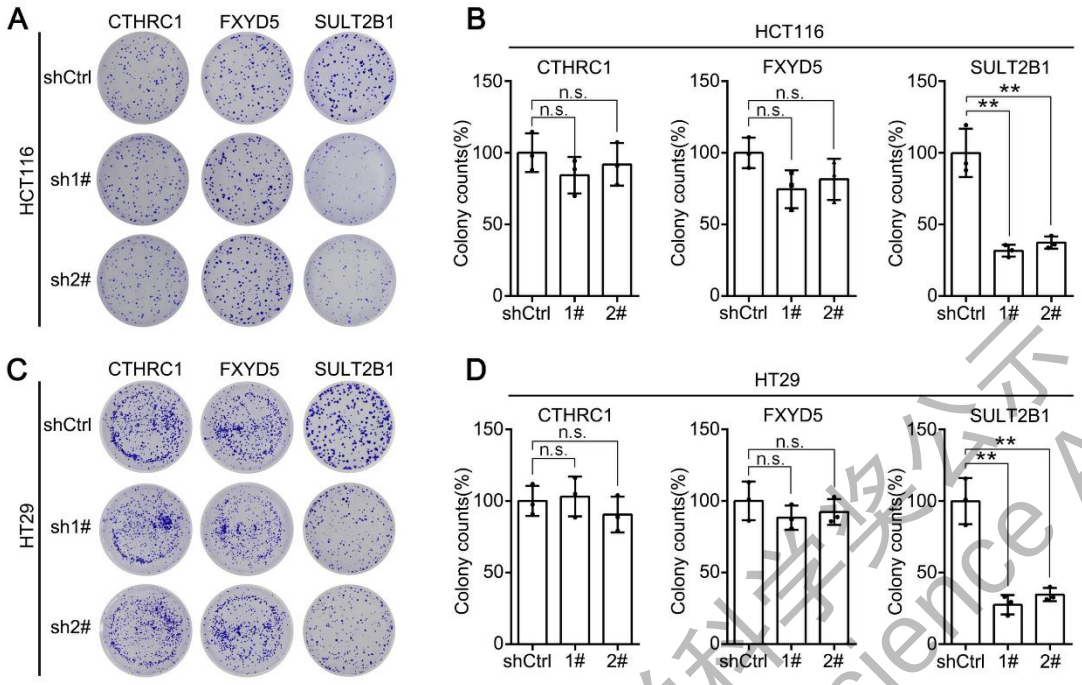


Figure 12. A-B, representative cell colony photos and quantification of colony formation assay in HCT116 cells after transfected with shRNA against CTHRC1, FXYD5 or SULT2B1. **C-D**, representative cell colony photos and quantification of colony formation assay in HT29 cells after transfected with shRNA against CTHRC1, FXYD5 or SULT2B1. **, $P < 0.01$; n.s., no significant differences.

5.2 Anti-SULT2B1 locked nucleic acid inhibits colorectal cancer cell proliferation

We next designed anti-SULT2B1 locked nucleic acids (LNAs) and detected proliferation suppression activity in HCT116 and HT29 cells, and it was found that LNAs 1#, 2#, 3#, and 4# successfully knocked down the expression of SULT2B1 (**Figure 13)**, and they significantly suppressed cell proliferation activity with IC50 ranged 1~2 μM (**Figure 14**). LNA 1#, 2#, 3#, and 4# also markedly suppressed colony formation in both cells (**Figure 15**). These results indicated that SULT2B1 may play an important role in CRC cell proliferation and warrant future investigation towards therapeutic target for CRC. In addition, LNAs targeting SULT2B1 could be used for the treatment of CRC.

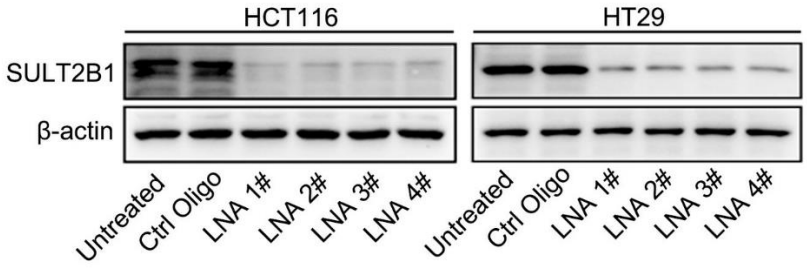
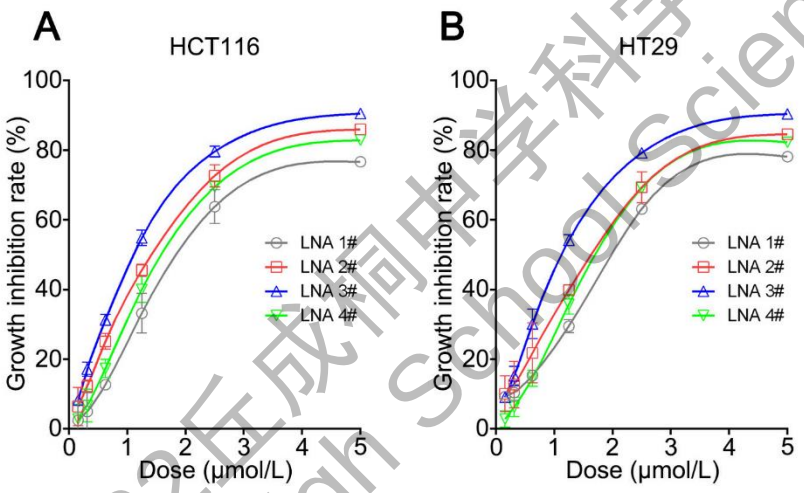


Figure 13. Assessment of LNA-induced SULT2B1 knockdown in HCT116 and HT29 cells by Western blotting assay.



C

	LNA	1#	2#	3#	4#
HCT116	IC50	1.996	1.328	1.019	1.628
HT29	IC50	1.930	1.457	1.046	1.729

Figure 14. Growth inhibition curves (**A-B**) and calculated IC50s (**C**) of LNAs after 72 h treatment in HCT116 and HT29 cells, respectively.

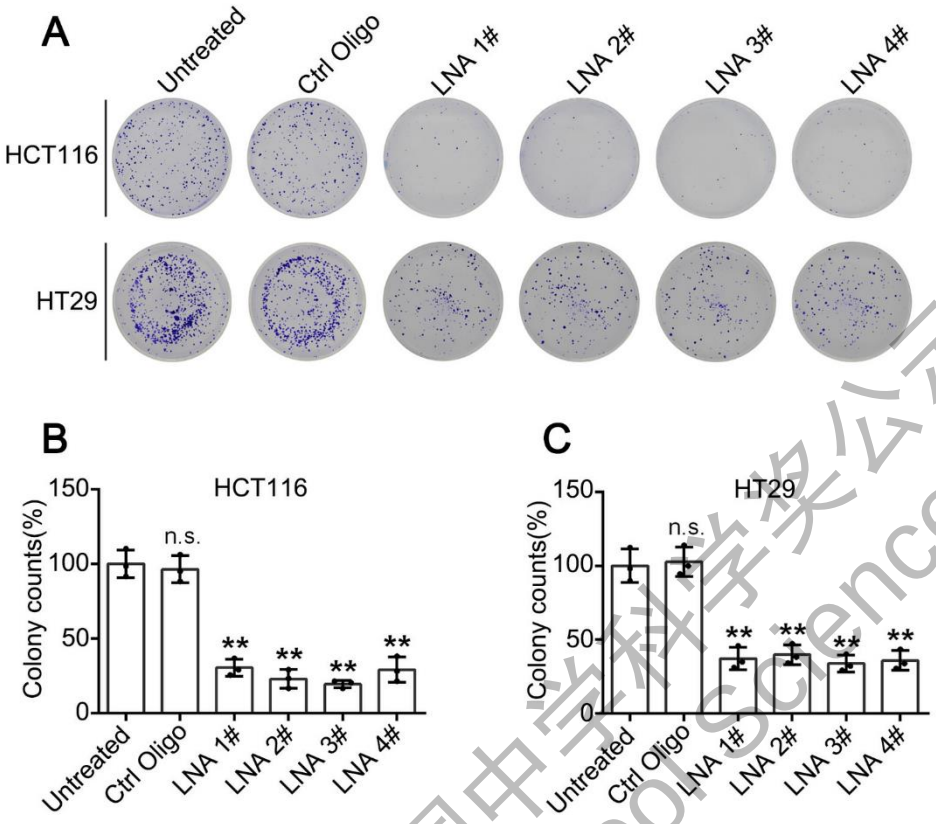


Figure 15. Representative cell colony photos (**A**) and quantification of colony formation assays (**B**-**C**) in HCT116 and HT29 cells after treated with 1 μM LNAs. For **B** and **C**, **, $P < 0.01$; n.s., no significant differences for compared with untreated group.

Discussion

Over the past few decades, CRC has become one of the most common cancers, and its incidence is expected to increase. Despite major advances in treatment, especially in chemotherapy and targeted therapy, mortality rate remains high. Accurate diagnosis and prognosis are challenging to improve CRC outcomes. Various therapeutic methods are been used to meet the clinical needs to treat CRC, and targeted therapy, focusing on genes that play essential roles during tumor pathogenesis, has attracted great attention and made a lot of progress in recent years [17]. In this study, we tried to identify novel biomarkers and therapeutics targets for CRC.

SULT2B1 (sulfotransferase family 2B member 1) is a member of sulfotransferases super family, which selectively catalyzes cholesterol into cholesterol sulfate. Cholesterol sulfate is involved in a variety of important biological progressions, like regulates keratinocytes differentiation, skin barrier formation as well as T cell signaling transduction [18-20]. The pathological role of SULT2B1 in cancer seems contradictory and the molecular mechanism remains largely unexplored. Evidences have indicated that SULT2B1 improves cancer cell growth by blocking tumor necrosis factor signaling in prostate cancer, also enhances epithelial mesenchymal transition via interacting with beta-catenin/ matrix metallopeptidase 7 (MMP7) signaling in liver cancer [21, 22]; while evidence also indicates that SULT2B1 is a suppressor for esophageal squamous cell carcinoma by upregulation period circadian regulator 1 (PER1) mRNA expression [23]. Further investigations shall be done to detail the complex role of SULT2B1 in cancer progression as well as decipher the underlying mechanisms.

All 4 LNAs that have been designed and used in our study have shown significant suppression activity for cell proliferation. We are currently carrying out an in vivo experiment to see if LNA #3, the most potent LNA among the four LNAs we designed, would have therapeutics effects in CRC xenograft mouse model (as illustrated in **Figure 16**). Antisense oligonucleotides including LNAs has developed very rapidly over the past few decades, and several LNA oligonucleotides are currently under investigation for the treatment of cancer, metabolic diseases and infectious diseases [24, 25]. Given the difficulties that traditional drug designs have encountered with numerous protein targets, the development of nucleic acid drug is therefore promising.

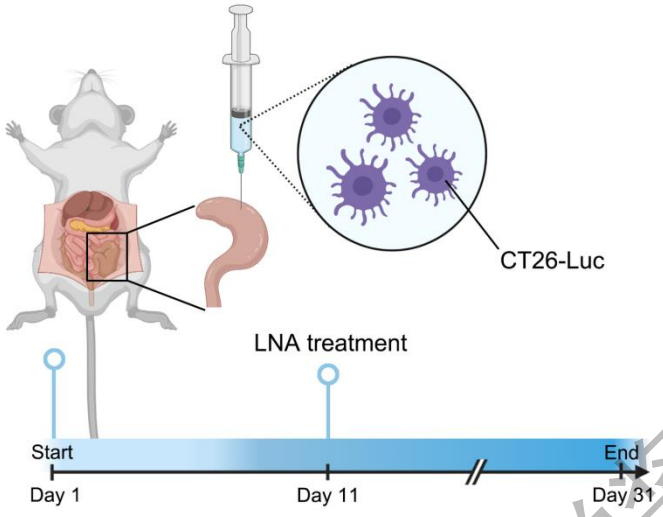


Figure 16. Illustration of future in vivo experiment plan.

In conclusion, we identified 3 differentially expressed genes, namely CTHRC1, FXYD5, and SULT2B1, from three independent CRC datasets from GEO, and validated them as biomarkers for unfavorable prognosis. We also found that they were all up-regulated in CRC tumor tissues as compared to normal counterparts, while only knocking down SULT2B1 suppressed cancer cell proliferation. LNAs were designed to knockdown SULT2B1 in HCT116 and HT29 cells, which significantly inhibited cell proliferation, indicating the therapeutic potential of these LNAs for the treatment of CRC, as well as the use of SULT2B1 as a therapeutic target. Collectively, we have identified that CTHRC1, FXYD5, and SULT2B1 may serve as potential prognostic biomarkers and SULT2B1 as a promising therapeutic target for CRC, and LNAs targeting SULT2B1 have therapeutic potential for CRC.

References

1. Keum, N. and E. Giovannucci, *Global burden of colorectal cancer: emerging trends, risk factors and prevention strategies.* Nat Rev Gastroenterol Hepatol, 2019. **16**(12): p. 713-732.
2. Siegel, R.L., et al., *Colorectal cancer statistics, 2020.* CA Cancer J Clin, 2020. **70**(3): p. 145-164.
3. Kuipers, E.J., et al., *Colorectal cancer.* Nat Rev Dis Primers, 2015. **1**: p. 15065.
4. Biller, L.H. and D. Schrag, *Diagnosis and Treatment of Metastatic Colorectal Cancer: A Review.* JAMA, 2021. **325**(7): p. 669-685.
5. Dekker, E., et al., *Colorectal cancer.* Lancet, 2019. **394**(10207): p. 1467-1480.
6. Sui, W., et al., *Circular RNA and gene expression profiles in gastric cancer based on microarray chip technology.* Oncol Rep, 2017. **37**(3): p. 1804-1814.
7. Tantai, J.C., X.F. Pan, and H. Zhao, *Network analysis of differentially expressed genes reveals key genes in small cell lung cancer.* Eur Rev Med Pharmacol Sci, 2015. **19**(8): p. 1364-72.
8. Wang, M., et al., *Identification of Key Genes and Prognostic Value Analysis in Hepatocellular Carcinoma by Integrated Bioinformatics Analysis.* Int J Genomics, 2019. **2019**: p. 3518378.
9. Frieden, M. and H. Orum, *Locked nucleic acid holds promise in the treatment of cancer.* Curr Pharm Des, 2008. **14**(11): p. 1138-42.
10. Hagedorn, P.H., et al., *Locked nucleic acid: modality, diversity, and drug discovery.* Drug Discov Today, 2018. **23**(1): p. 101-114.
11. Smith, C.I.E. and R. Zain, *Therapeutic Oligonucleotides: State of the Art.* Annu Rev Pharmacol Toxicol, 2019. **59**: p. 605-630.
12. Fluiter, K., et al., *In vivo tumor growth inhibition and biodistribution studies of locked nucleic acid (LNA) antisense oligonucleotides.* Nucleic Acids Res, 2003. **31**(3): p. 953-62.
13. Fluiter, K., et al., *On the in vitro and in vivo properties of four locked nucleic acid nucleotides incorporated into an anti-H-Ras antisense oligonucleotide.* Chembiochem, 2005. **6**(6): p. 1104-9.
14. Frieden, M. and H. Ørum, *Locked nucleic acid holds promise in the treatment of cancer.* Curr Pharm Des, 2008. **14**(11): p. 1138-42.
15. Jaschinski, F., H. Korhonen, and M. Janicot, *Design and Selection of Antisense Oligonucleotides Targeting Transforming Growth Factor Beta (TGF-β) Isoform mRNAs for the Treatment of Solid Tumors*, in *Gene Therapy of Solid Cancers: Methods and Protocols*, W. Walther and U. Stein, Editors. 2015, Springer New York: New York, NY. p. 137-151.
16. Kolde, R., et al., *Robust rank aggregation for gene list integration and meta-analysis.* Bioinformatics (Oxford, England), 2012. **28**(4): p. 573-580.
17. Fonseca, A.L., et al., *Bioinformatics Analysis of the Human Surfaceome Reveals New Targets for a Variety of Tumor Types.* Int J Genomics, 2016. **2016**: p. 8346198.
18. Hanley, K., et al., *Cholesterol sulfate stimulates involucrin transcription in keratinocytes by increasing Fra-1, Fra-2, and Jun D.* J Lipid Res, 2001. **42**(3): p. 390-8.
19. Presland, R., *Function of Filaggrin and Caspase-14 in Formation and Maintenance of the Epithelial Barrier.* Dermatol Sinica, 2009. **27**: p. 1-14.
20. Wang, F., et al., *Inhibition of T cell receptor signaling by cholesterol sulfate, a naturally occurring derivative of membrane cholesterol.* Nature Immunology, 2016. **17**(7): p. 844-850.
21. Vickman, R.E., et al., *Cholesterol Sulfotransferase SULT2B1b Modulates Sensitivity to Death Receptor Ligand TNFα in Castration-Resistant Prostate Cancer.* Molecular Cancer Research : MCR, 2019. **17**(6): p. 1253-1263.

22. Yang, X., et al., *SULT2B1b promotes epithelial-mesenchymal transition through activation of the β-catenin/MMP7 pathway in hepatocytes.* Biochemical and Biophysical Research Communications, 2019. **510**(4): p. 495-500.
23. Li, Z., et al., *The promoter hypermethylation of SULT2B1 accelerates esophagus tumorigenesis via downregulated PER1.* Thoracic Cancer, 2021. **12**(24): p. 3370-3379.
24. Morris, K.V. and J.S. Mattick, *The rise of regulatory RNA.* Nat Rev Genet, 2014. **15**(6): p. 423-37.
25. Frieden, M. and H. Orum, *The application of locked nucleic acids in the treatment of cancer.* IDrugs, 2006. **9**(10): p. 706-11.

Acknowledgement

I thank my mentor Dr. Qianming Du from Nanjing First Hospital, for bringing me into the field of cancer biology and guiding me throughout this project. I would also like to thank Tao Li for his guidance on dataset analysis and experimental design, especially the design of the LNAs, and Chenchen Jin for teaching me and helping me with all the wet lab techniques.

致谢

感谢南京市第一医院的杜前明老师的理论以及实验指导，南京市第一医院中心实验室提供实验条件，以及李涛和金晨晨两位学长学姐在数据分析，实验设计和实验操作方面的指导和帮助。此论文选题是基于对结肠癌的在线数据库分析。吴若予同学完成了数据分析，实验以及论文撰写和图片制作等过程。在此，对杜前明老师及其实验室所有同学的悉心指导以及无私帮助表示感谢。