Comprehensive analysis of ceRNA network of ERCC4 in colorectal cancer

Huixin Hu[1,2,3,4,*], Songyi Liu[1,2,3,4,*], Aining Chu[1,2,3], Jing Chen[1,2,3], Chengzhong Xing[1,2,3,4] and Jingjing Jing[1,2,3]

[1] Tumor Etiology and Screening Department of Cancer Institute and General Surgery, the First Hospital of China Medical University, Shenyang, China
[2] Key Laboratory of Cancer Etiology and Prevention in Liaoning Education Department, the First Hospital of China Medical University, Shenyang, China
[3] Key Laboratory of GI Cancer Etiology and Prevention in Liaoning Province, the First Hospital of China Medical University, Shenyang, China
[4] Department of Anorectal Surgery in Liaoning Province, the First Hospital of China Medical of China Medical University, Shenyang, China
* These authors contributed equally to this work.

Corresponding authors
Chengzhong Xing,
xcz1966@126.com
Jingjing Jing, hellojjjing@163.com

ABSTRACT

Objective: ERCC4 is one of the most significant molecules of Nucleotide Excision Repair (NER), which has been researched due to its high expression in colorectal cancer (CRC). This study aimed to find out the ceRNA (competitive endogenous RNA) network of *ERCC4* in CRC.

Methods and Materials: Pan cancer mRNA expression of *ERCC4* was evaluated using TCGA database. The protein expression of ERCC4 was evaluated based on the Human Protein Atlas (HPA). We screened DElncRNAs and DEmiRNAs in two groups of *ERCC4*^high and *ERCC4*^low expression in CRC. Then a lncRNA-miRNA-*ERCC4* regulatory network was constructed based on DElncRNAs and DEmiRNAs using Starbase database and visualized by Cytoscape software. Kaplan-Meier analysis was performed to evaluate the prognostic value of the ceRNA network. Further, RT-PCR was performed to validate the expression of the representative molecules in the ceRNA network in CRC and normal tissues. The relationship between drug sensitivity and these molecules were also evaluated using RNAactDrug database.

Results: *ERCC4* was overexpressed in a variety of tumors at mRNA levels, including CRC. High expression of ERCC4 was also observed on protein level in CRC. A total of 1,885 DElncRNAs and 68 DEmiRNAs were identified from CRC samples in *ERCC4*^high and *ERCC4*^low expression groups. Predicted by the Starbase database, we got interacting miRNAs and lncRNAs of *ERCC4* from the DEmiRNAs and DElncRNAs, and a lncRNA-miRNA-*ERCC4* regulatory network was constructed. Kaplan-Meier survival curves results showed that miR-200c-3p (hazard ratio [HR] = 0.62, P = 0.032), MALAT1 (HR = 1.54, P = 0.016), and AC005520.2 (hazard ratio [HR] = 1.75, P = 0.002) were significantly associated with the prognosis of CRC. After validation by RT-PCR, we found that *ERCC4* and MALAT1 were up-regulated in CRC compared with normal tissues, while miR-200c-3p was down-regulated. A strong negative correlation was observed between MALAT1 and miR-200c-3p. Drug sensitivity analysis showed that *ERCC4*, miR-200c and MALAT1 were all associated with Cisplatin.

Conclusion: We constructed a ceRNA network of *ERCC4* in CRC, of which the MALAT1-miR-200c-3p-*ERCC4* axis may be involved in the development, prognosis

and chemotherapy sensitivity of CRC. These findings might provide novel clues and insights on the molecular mechanisms of ERCC4 and NER pathway in CRC.

INTRODUCTION

DNA damage caused by endogenous or exogenous genetic toxicants which contribute to genomic instability may lead to a variety of cancers (*Iyama & Wilson, 2013*). DNA repair systems play a critical role in maintaining the integrity and stability of the genome, of which nucleotide excision repair (NER) can identify and amend multiple types of potential DNA damage. The NER process consists of four main steps namely damage identification, damage partitioning and unwinding, damage incision and new strand synthesis (*de Laat, Jaspers & Hoeijmakers, 1999*; *Liu et al., 2014*; *Slyskova et al., 2012*). Different key proteins are involved in different NER steps. ERCC4 (XPF) is one of the indispensable molecules of NER, which can determine the activity of NER and is responsible for the removal of UV-C photoproducts and large volume adducts from DNA (*Enzlin & Scharer, 2002*). Besides, the heterodimer of ERCC4-ERCC1 is involved with the 5′ incision step of the NER pathway (*Enzlin & Scharer, 2002*). ERCC4-ERCC1 can repair damaged DNA in both replicating and non-replicating cells as a structure-specific endonuclease (*Evans et al., 1997*; *Mu, Hsu & Sancar, 1996*; *Sijbers et al., 1996*). Therefore, the abnormal expression of *ERCC4* could seriously affect the NER pathway.

Many researches have mentioned the role of *ERCC4* in cancers, such as bladder cancer, gastric cancer and oral cancer (*Li & Ma, 2018*; *Qiu et al., 2014*; *Sa et al., 2020*; *Wang et al., 2010*; *Wei et al., 2014*). Colorectal cancer (CRC) is a malignant tumor which is the third cause of cancer death in China (*Zhang et al., 2021*). Previous studies also focused on the relationship between *ERCC4* polymorphisms and the risk of CRC (*He et al., 2014*; *Kabzinski et al., 2015*; *Yang, Li & Li, 2015*). In recent studies, the post-transcriptional regulation of non-coding RNA plays an important role in the development of cancer (*Anastasiadou, Jacob & Slack, 2018*; *Wei et al., 2017*). *Salmena et al. (2011)* put forward the hypothesis of ceRNA, which expands our insight into the pathogenesis of CRC (*Wei et al., 2017*; *Yuan et al., 2019*). For example, *Zhu et al. (2019)* suggested that LINC00365 is an oncogene in CRC, which can regulate the expression of several mRNAs by sponging miRNAs. H19 was found to regulate PI3K-Akt signaling pathway through ceRNA network, and predict the poor prognosis of CRC (*Zhong et al., 2019*). Since *ERCC4* plays a significant role in CRC tumorigenesis, it is necessary to explore the related ceRNA regulatory network to better understand its mechanism in the development and progression of CRC. Although there have been studies investigating the relationship between *ERCC4* expression and the risk of CRC based on small sample size (*Zhang et al., 2019*), the ceRNA network regulation of *ERCC4* in CRC has not been identified.

In this study, we screened DElncRNAs and DEmiRNAs in two groups of *ERCC4*^{high} and *ERCC4*^{low} expression in CRC. Then a lncRNA-miRNA-*ERCC4* regulatory network was constructed based on DElncRNAs and DEmiRNAs using bioinformatics methods. Further, RT-PCR was performed to validate the expression of the representative molecules in the ceRNA network in CRC and normal tissues. The relationship between drug sensitivity and these molecules were also evaluated. Our research might be useful for further studies of the molecular mechanism of *ERCC4* in CRC.

MATERIALS AND METHODS

Data preparation and processing

We firstly collected the gene expression information of 33 different kinds of tumors in TCGA database (http://cancergenome.nih.gov/). Further, the mRNA sequencing (mRNA seq) data of 698 CRC samples and miRNA sequencing (miRNA seq) data of 630 CRC samples were also gained from TCGA database. All raw RNA sequencing (RNA seq) data (lncRNAs, miRNAs, and mRNAs) were normalized as fragments per kilobase of exon model per million mapped fragment reads. The clinical information (survival status and survival time) were downloaded from UCSC XENA (https://xenabrowser.net/). Human Protein Atlas (HPA) database (http://www.proteinatlas.org/) was used to verify the protein expression level of ERCC4 in CRC.

Identification of DElncRNAs and DEmiRNAs

We further performed differential expression analysis to identify DElncRNAs and DEmiRNAs in *ERCC4*^{high} and *ERCC4*^{low} CRC samples, with the median of *ERCC4* expression level as the cutoff value for grouping. We determined the DElncRNA with a threshold of $|logFC| > 0.1$ and $P < 0.05$, and a threshold of $|logFC| > 0.5$ and $P < 0.05$ for DEmiRNA. Volcano plots of the DElncRNAs and DEmiRNAs were visualized using GraphPad Prism 8 software (version 8.3.0).

Construction of ceRNA network

Based on the hypothesis that lncRNAs compete with miRNAs as natural sponges in the cytoplasm and indirectly regulate mRNA expression, the ceRNA network was constructed through the following steps: (1) StarBase 3.0 (https://starbase.sysu.edu.cn/) was used to predict the potential miRNAs that interact with *ERCC4*. (2) The overlapping miRNAs intersected with DEmiRNA were selected and visualized using Venn diagram package in R software. (3) LncRNAs that potentially interact with overlapping miRNA were predicted using StarBase. (4) The intersection of the predicted lncRNAs and DElncRNAs was calculated and visualized by the Venn diagram. (5) Based on the identified lncRNAs and miRNAs, we constructed and visualized a lncRNA-miRNA-*ERCC4* triple regulation network using Cytoscape software (https://www.cytoscape.org/) (*Shannon et al., 2003*).

Survival analysis

We performed Kaplan-Meier analysis and log-rank test on DElncRNAs, DEmiRNAs, and *ERCC4* in the ceRNA network using survminer package in R software, thus to determine
their relationships with OS of CRC patients in TCGA database. The log-rank P value < 0.05 was considered statistically significant.

Real-time PCR validation

We enrolled a total of 67 cases of CRC and adjacent normal tissue for validation, which were collected under enterectomy at the First Hospital of China Medical University. The clinical parameters of them were listed in File S1.

Total RNA was extracted using RNAiso Plus (Takara, Kusatsu, Japan) and converted into cDNA using PrimeScript RT Kit (Takara, Kusatsu, Japan). Amplification was performed with SYBR Green (SYBR Premix Ex Taq II; Takara, San Jose, CA, USA). Each response of lncRNA was normalized to β-actin, and U6 for miRNA. The study was approved by the Human Ethics Committee of China Medical University. The IRB approval number is 2016-155-2.

Drug sensitivity analysis based on multiomics data

RNAactDrug (http://bio-bigdata.hrbmu.edu.cn/RNAactDrug) is a comprehensive database of relationship between RNAs and drug sensitivity, which integrated three large-scale pharmacogenomic databases respectively called Cancer Cell Line Encyclopedia (CCLE), Drug Sensitivity in Cancer (GDSC) and CellMiner. We explored the association between drug sensitivity and *ERCC4*, miR-200c and MALAT1 at expression, mutation, CNV, and methylation level. Pearson correlation analysis was used, and statistical significance was set at $P < 0.05$.

Statistical analysis

The statistical analysis was conducted through SPSS 20.0 software (IL, Chicago) and R software. The difference of RNA expression between CRC and non-tumor tissues was compared by t-tests, and the fold change value was used to evaluate the difference. We applied correlation analysis to find out the Pearson correlation coefficients (r) and corresponding P value. $P < 0.05$ was considered as statistically significant.

RESULTS

The expression and prognostic value of *ERCC4* in CRC

To investigate the possible role of *ERCC4* in CRC, based on TCGA data, we found that *ERCC4* was overexpressed in many kinds of tumors on mRNA level, including CRC (Fig. 1A). In addition, based on the HPA database, high expression of ERCC4 was also observed on protein level in CRC (Fig. 1B). Since *ERCC4* was up-regulated in tumor specimens, we then investigated the clinical significance of *ERCC4* expression in CRC patients. According to the Kaplan-Meier survival curves, as shown in Fig. 1C, our results showed that higher expression of *ERCC4* was correlated with poor overall survival (OS) in colon cancer patients.
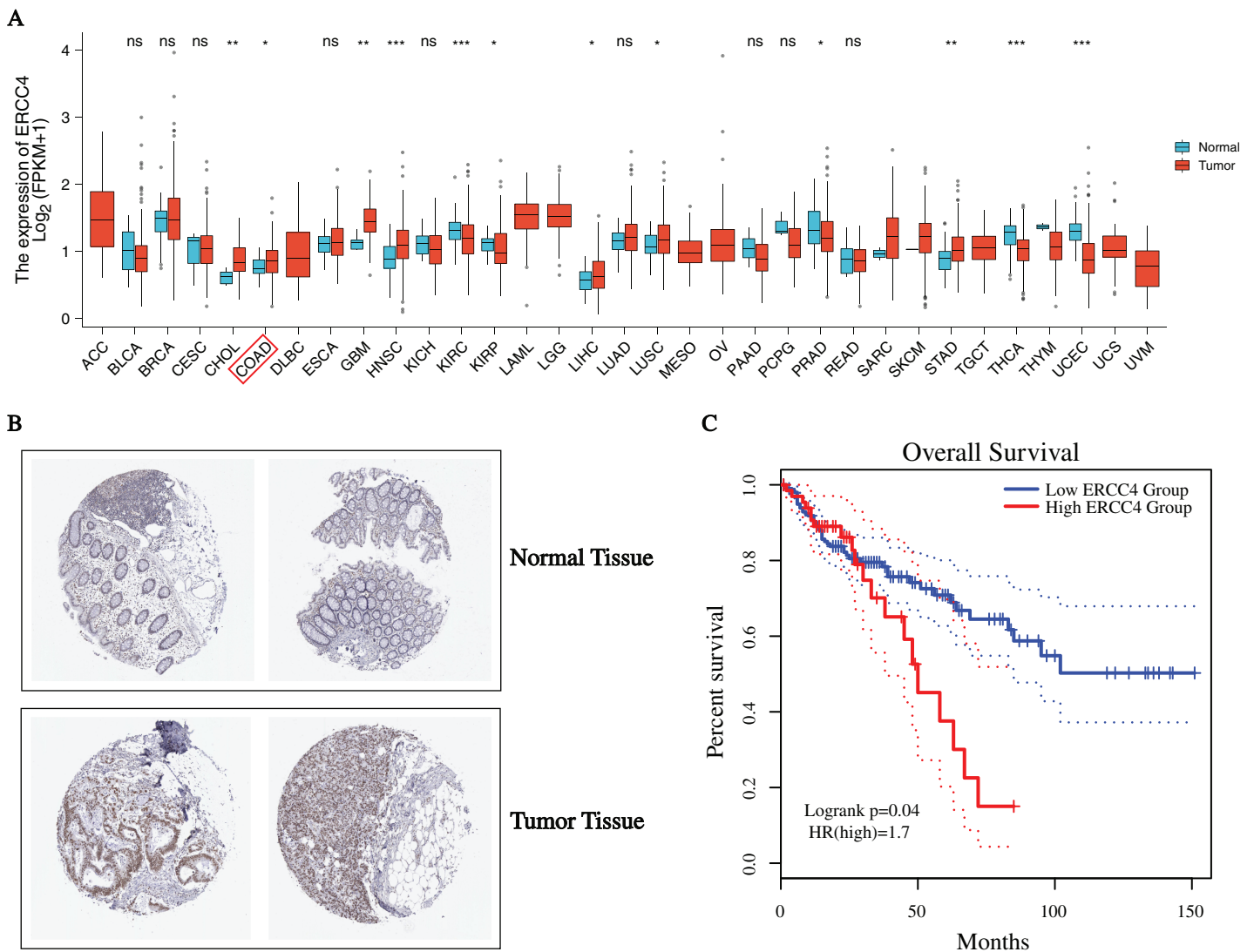
Figure 1 **The expression and prognostic value of ERCC4 in CRC.** (A) The mRNA expression level of ERCC4 in 33 different kinds of tumors based on TCGA database. (B) The protein expression level of ERCC4 in normal colorectal tissue and CRC tissue based on the Human Protein Atlas database. (C) The survival analysis of ERCC4 in colon cancer. $P = 0.04$. *: $P < 0.05$, **: $P < 0.01$, ***: $P < 0.001$.

Identification of DEmiRNAs and DElncRNAs

In total, 1,885 DElncRNAs (412 upregulated and 1,473 downregulated), 68 DEmiRNAs (7 upregulated and 61 downregulated) were identified from CRC samples in $ERCC4^{high}$ and $ERCC4^{low}$ expression groups (Figs. 2A, 2B, File S2).

Construction of ceRNA network of *ERCC4*

Predicted by the Starbase database, we got interacting miRNAs of *ERCC4* from the DEmiRNAs, and only one overlapping miRNA (miR-200c-3p) was identified (Fig. 2C). Further, 10 lncRNAs interacting with miR-200c-3p were identified from DElncRNAs
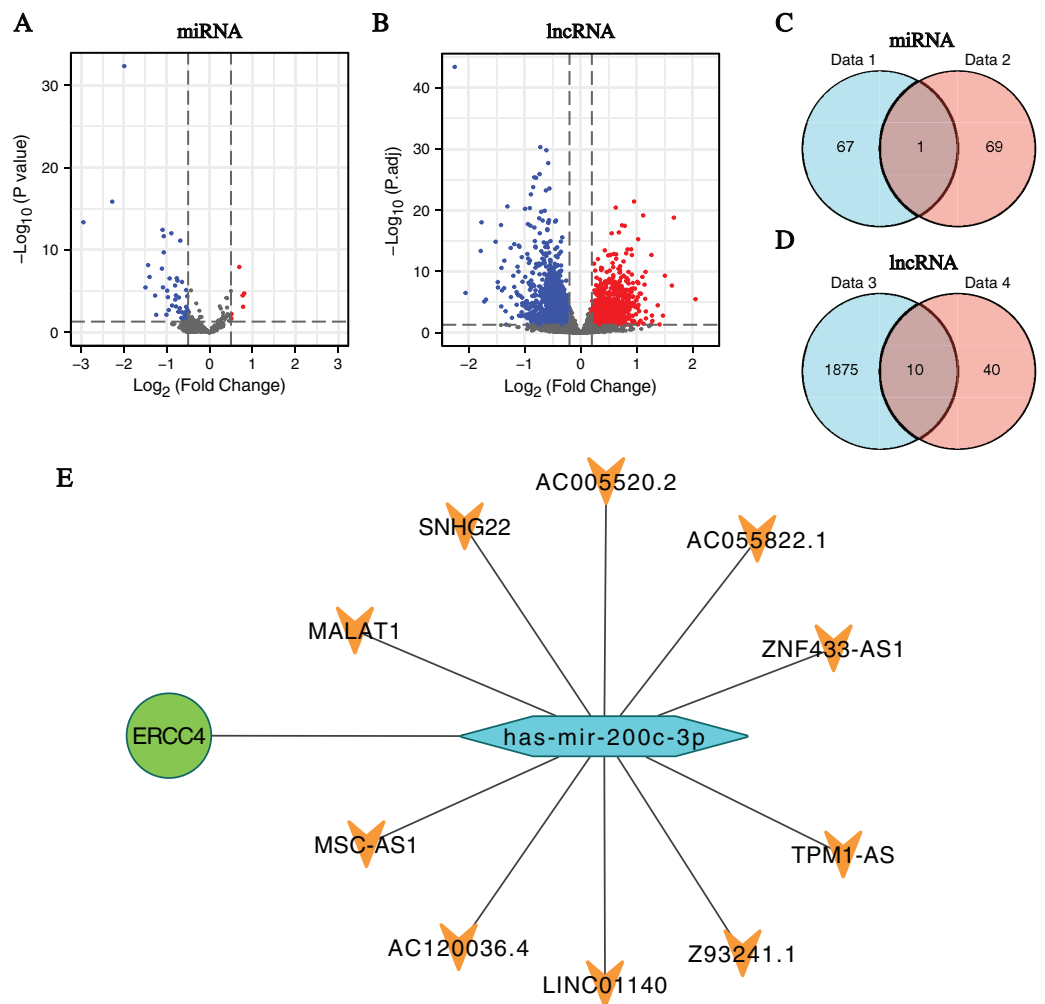
Figure 2 DElncRNAs and DEmiRNAs between the expression of ERCC4high and ERCC4low in CRC samples and the lnRNA-miRNA-ERCC4 triple regulatory network. (A) Visualized volcano plots of the DEmiRNAs in ERCC4high and ERCC4low CRC samples. (B) Visualized volcano plots of the DElncRNAs in ERCC4high and ERCC4low CRC samples. (C) The intersection of the predicted miRNA and DEmiRNA. (D) The intersection of the predicted lncRNA and DElncRNA. (E) The triple network of lncRNA-miRNA-ERCC4.

(Fig. 2D). Based on the theory of ceRNA regulatory network, and the obtained results of above DEmiRNAs and DElncRNAs, we constructed a lncRNA-miRNA-*ERCC4* regulatory network in Fig. 2E.

Survival analysis of the molecules in the ceRNA network

To investigate the clinical values of the molecules in the ceRNA network in CRC, Kaplan-Meier survival curves were used to evaluate the prognostic significance. The results showed that miR-200c-3p (hazard ratio [HR] = 0.62, $P = 0.032$), MALAT1 (hazard ratio [HR] = 1.54, $P = 0.016$), and AC005520.2 (hazard ratio [HR] = 1.75, $P = 0.002$) were significantly associated with the prognosis of CRC (Figs. 3A–3F).
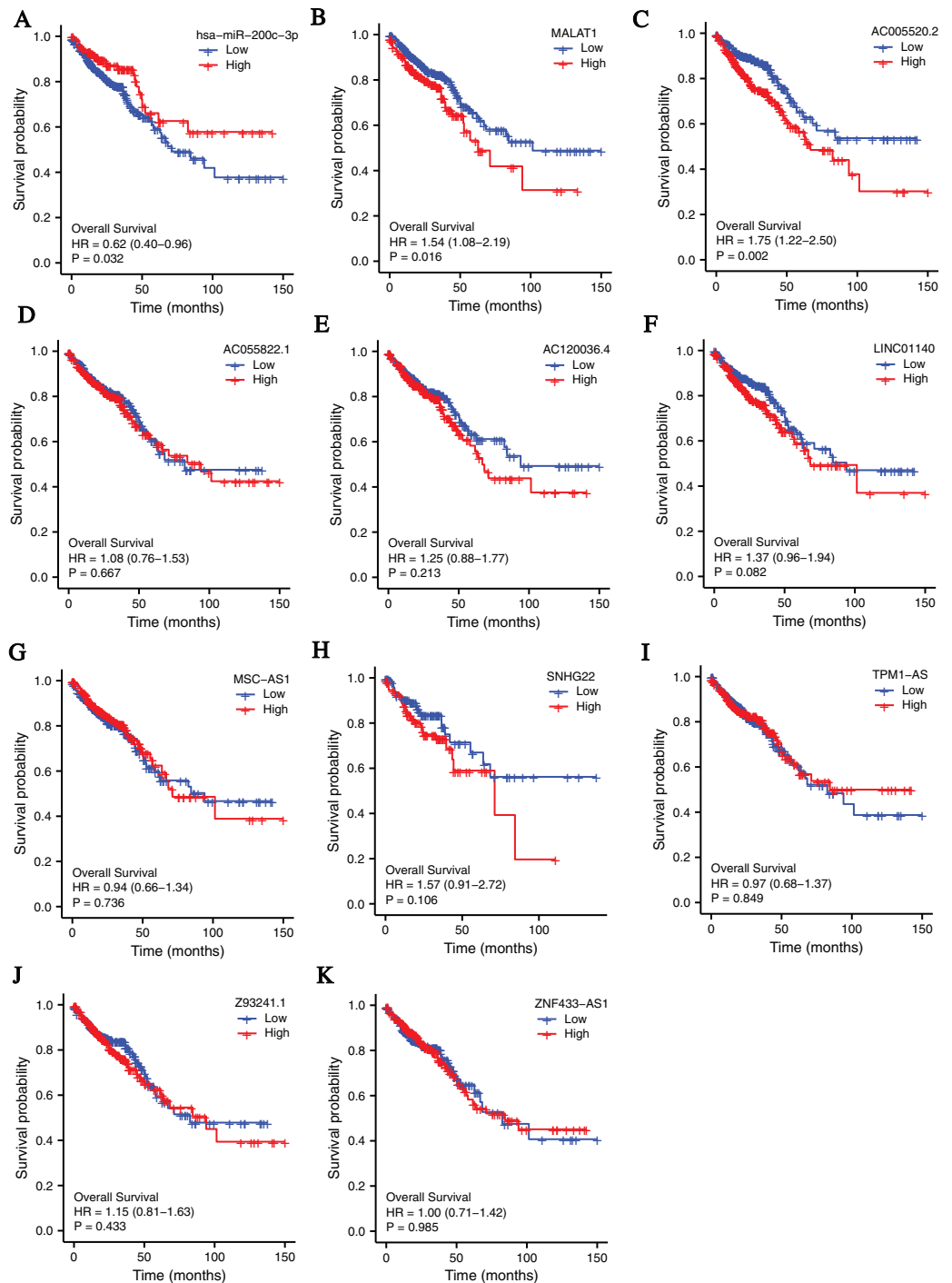
Figure 3 The lncRNAs and miRNAs in the ceRNA network and their relationships with CRC prognosis base on TCGA database. (A) Higher expression of has-miR-200c-3p had a longer OS in CRC. HR = 0.62, *P* = 0.032 . (B) Lower expression of MALAT1 had a longer OS in CRC. HR = 1.54, *P* = 0.016. (C) Lower expression of AC005520.2 had a longer OS in CRC. HR = 1.75, *P* = 0.002. (D–K) AC055822.1, AC120036.4, LINC01140, MSC-AS1, SNHG22, TPM1-AS, Z9324.1 and ZNF433-AS1 are not related to CRC prognosis. *P* < 0.05 was statistically significant.
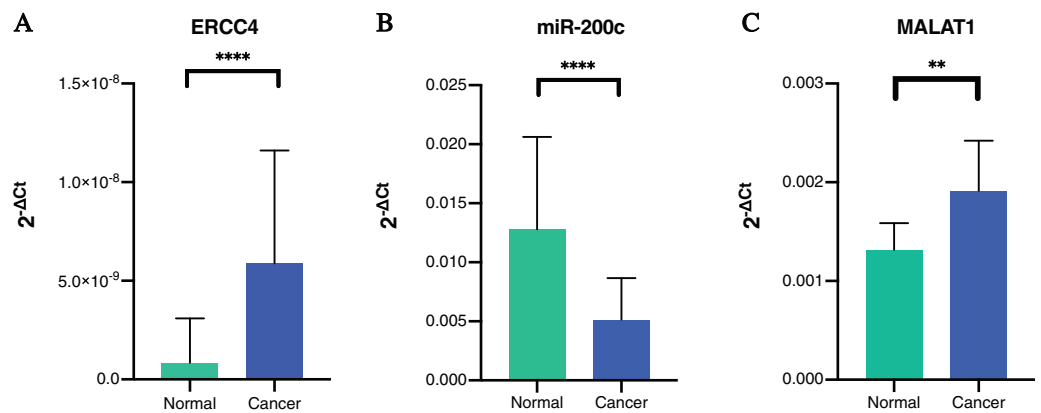
Figure 4 The expressions of ERCC4, miR-200c and MALAT1 in CRC and normal tissues based on RT-PCR. (A) ERCC4 was up-regulated in CRC compared with normal tissues. $P < 0.05$. (B) MiR-200c was down-regulated in CRC comparing with normal tissues. $P < 0.05$. (C) MALAT1 was up-regulated in CRC compared with normal tissues. $P < 0.05$.
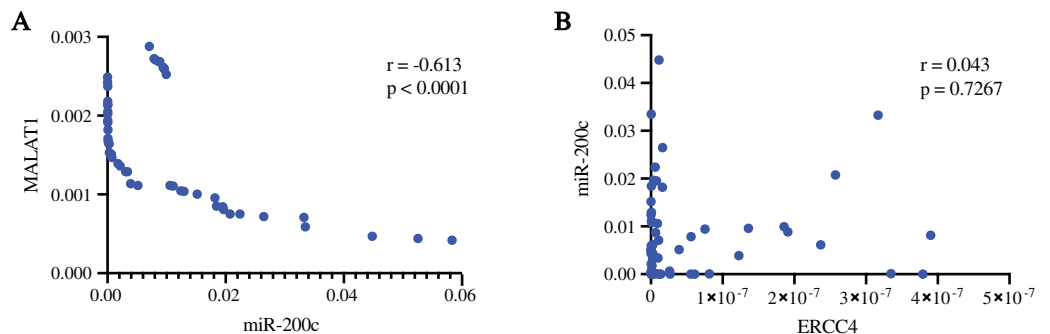
Figure 5 The associations of MALAT1, miR-200c-3p, and ERCC4 based on RT-PCR. (A) There is a strong negative correlation between MALAT1 and miR-200c-3p. r = −0.613, $P < 0.0001$. (B) The correlation between miR-200c-3p and ERCC4 was not statistically significant. $P = 0.723$.

RT-PCR verification of the molecules in the ceRNA network

To further verify our bioinformatic analysis results, we performed RT-PCR experiment to explore the expression of the molecules in the ceRNA network in CRC and normal tissues. Because MALAT1 showed a prognostic value in CRC, and it was a well-studied lncRNA in cancers, we selected *ERCC4*, miR-200c-3p and MALAT1 for verification. Through differential expression analysis, it could be found out that *ERCC4* and MALAT1 were up-regulated in CRC compared with normal tissues, while miR-200c-3p was down-regulated in CRC compared with normal tissues, which were all with statistical significance (Fig. 4).

In addition, we validated the association of MALAT1-miR-200c-3p-*ERCC4* by correlation analysis (Fig. 5). A strong negative correlation (r = −0.613) was found between MALAT1 and miR-200c-3p ($P < 0.0001$), while the correlation between miR-200c-3p and ERCC4 was not statistically significant ($P = 0.723$).
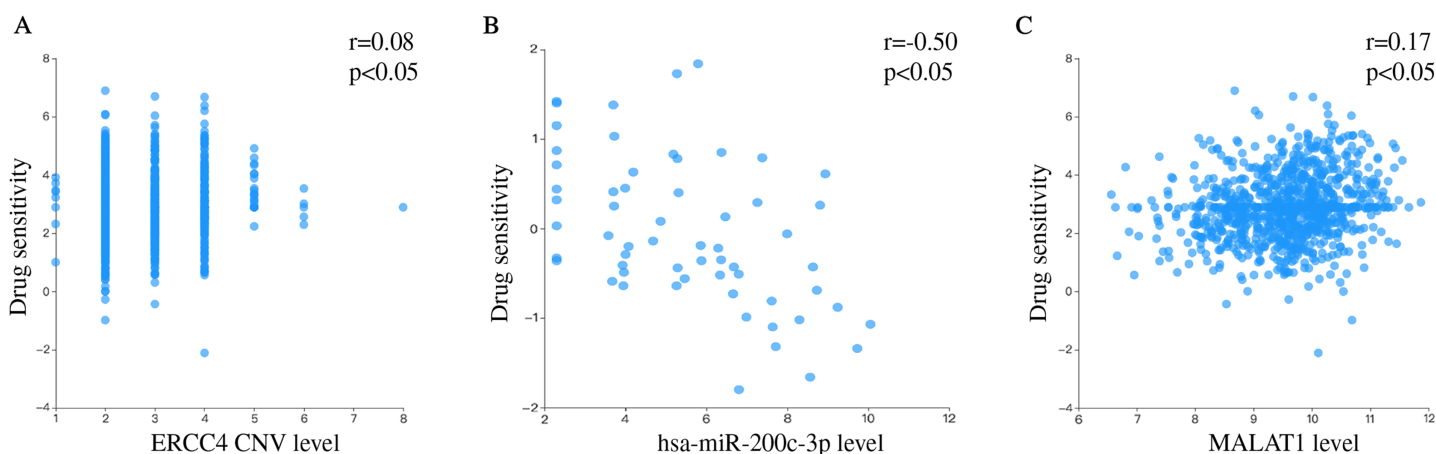
Figure 6 Drug sensitivity analysis of ERCC4, hsa-miR-200c-3p and MALAT1. (A) The correlation between drug sensitivity and ERCC4 (Copy number variations, CNV). Pearson $r = 0.08$. Pearson $P < 0.05$. (B) The correlation between drug sensitivity and hsa-miR-200c-3p (gene expression). Pearson $r = −0.50$. Pearson $P < 0.05$. (C) The correlation between drug sensitivity and MALAT1 (gene expression). Pearson $r = 0.17$. Pearson $P < 0.05$.

The correlation between *ERCC4*, miR-200c-3p, MALAT1 and drug sensitivity

We separately explored the relationship between drug sensitivity and *ERCC4*, miR-200c and MALAT1 using RNAactDrug database. The results were shown in File S3. In addition, we found that *ERCC4*, miR-200c and MALAT1 were all associated with Cisplatin (Fig. 6).

DISCUSSION

NER pathway plays an important role in identifying and repairing DNA damage, in which ERCC4 is one of the unnegligible molecules. Abnormal expression of *ERCC4* can lead the imbalance between DNA damage and repair (*Neven et al., 2018*). *ERCC4* has been found with high-expression in cancer tissues, including CRC (*McDaniel & Schultz, 2008*). Although there have been a few studies on *ERCC4* in several cancers, its epigenetic regulation is still not fully understood. Therefore, we attempted to establish a *ERCC4*-related ceRNA triple network in CRC to provide a better understanding of the biological mechanisms of *ERCC4* in CRC.

Firstly, we evaluated the expression and prognostic value of *ERCC4* in CRC using TCGA data. *ERCC4* was found to be upregulated in CRC tissues compared with normal tissues both on mRNA and protein levels. It was also associated with prognosis of colon cancer patients. Then 1885 DElncRNAs and 68 DEmiRNAs were sorted out from CRC samples in *ERCC4*[high] and *ERCC4*[low] expression groups. One miRNA and 10 lncRNAs were then identified from the DElncRNAs and DEmiRNAs through bioinformatics prediction using Starbase database. Based on these results, a lncRNA-miRNA-*ERCC4* regulatory network was constructed. We also investigated the clinical values of the molecules in the ceRNA network in CRC, and found that miR-200c-3p, MALAT1 and AC005520.2 were significantly associated with the prognosis of CRC.
Further, we performed RT-PCR experiment to verify the expression of the molecules in the ceRNA network in CRC and normal tissues. By comprehensively considering the prognostic value and research status, we selected *ERCC4*, miR-200c-3p and MALAT1 for verification. Our results showed that *ERCC4* and MALAT1 were up-regulated in CRC, while miR-200c-3p was down-regulated in CRC. In addition, we validated the association of MALAT1-miR-200c-3p-*ERCC4* by correlation analysis. A strong negative correlation was observed between MALAT1 and miR-200c-3p. The drug sensitivity analysis showed that *ERCC4*, miR-200c and MALAT1 were all associated with Cisplatin. Therefore, the MALAT1-miR-200c-3p-*ERCC4* axis may be closely associated with the development, prognosis and chemotherapy sensitivity of CRC and deserve further study.

ERCC4, as a key gene in NER, is mainly involved in the removal of damage fragments (*Friedberg, 2001*). ERCC4 domain is also necessary to form a tight complex with ERCC1, a structure specific DNA repair endonuclease responsible for 5'-primer cleavages during DNA excision and repair (*Tripsianes et al., 2005*; *Tsodikov et al., 2005*). MALAT1 is located on human chromosome 11q13 and can generate an 8 kb length transcript of long non-coding and nuclear-retained (*Ji et al., 2003*; *Lin et al., 2007*; *Zhang, Hamblin & Yin, 2017*). It can act at the transcriptional, post-transcriptional and translational levels, thus it may be involved in various biological processes such as DNA damage repair, metabolism, and cell signaling (*Castro-Oropeza et al., 2018*). MiR-200c is highly studied because it has an impact on development, stemness, proliferation, epithelial–mesenchymal transition (EMT), metastasis and therapy resistance (*Mutlu et al., 2016*). MiR-200c is predicted to target *SIRT1* which was found to be negatively regulated by Poly polymerase-2, an enzyme that plays significant roles in the DNA damage response (*Nabih, 2020*). Therefore, these non-coding RNAs in ceRNA network may competitively regulate *ERCC4* to affect the balance between DNA damage and repair, and participate in the development of CRC.

MALAT1 as a highly conserved lncRNA, has been found to be overexpressed in several human neoplasms and to promote tumor cell invasion and metastasis (*Castro-Oropeza et al., 2018*). It has been reported regulating mRNA expression and then contributing to CRC progression by sponging miRNA in a few studies, like miR-26a/26b (*Xu et al., 2020*), miR-106b-5p (*Zhuang et al., 2019*), and miR-203 (*Wu et al., 2018*). The miR-200 family played an important role in EMT. It was mentioned that miR-200c-3p was not only differentially expressed in CRC, but also may inhibit CRC migration and invasion, and promote apoptosis after LPS stimulation (*Huang et al., 2020*; *Jiang et al., 2020*). The miR-200c was directly target *ZEB1* and *ZEB2* which are the *CDH1* transcription inhibitors. Inhibition of miR-200c reduces the expression of *CDH1* and therefore induces EMT (*Gregory et al., 2008*; *Park et al., 2008*). MALAT1 has been shown to induce EMT during endometriosis (*Du et al., 2019*) and metastasis in clear cell kidney carcinoma mouse models (*Xiao et al., 2015*) *via* the miR200/*ZEB2* axis by acting as a sponge for the miR200 family. Sreekumar R found that *ZEB2*-dependent EMT transcriptional programme drives therapy resistance by activating nucleotide excision repair genes *ERCC1* and *ERCC4* in CRC (*Sreekumar et al., 2021*). These results demonstrated that MALAT1, miR-200 and *ERCC4* were all connected to the EMT. Therefore, the

MALAT1-miR-200c-3p-*ERCC4* axis may be involved in the malignant behaviour of CRC by affecting EMT.

DNA repair has an essential role in resistance to platinum agents. Platinum agents inhibit DNA replication and transcription by binding to DNA molecules in the form of platinum DNA adducts (*Altaha et al., 2004*). ERCC4 is an important protein in the NER pathway, which is a major part of the DNA repair pathway (*Fuss & Cooper, 2006*; *Gillet & Scharer, 2006*). CRC cells expressing *ZEB2* developed resistance to oxaliplatin while *ERCC4* is induced upon *ZEB2* expression (*Sreekumar et al., 2021*). Forced expression of miR-192-5p in SGC7901/DDP cells can significantly inhibit the expression of *ERCC4*, making the gastric cancer cells more sensitive to cisplatin *in vitro* and *in vivo* (*Xie et al., 2019*). MiR-138-5p regulated the cisplatin resistance by modulating the expression of the DNA repair proteins ERCC4 and ERCC1 in gastric cancer cells (*Ning et al., 2019*). During repeated treatments of oxaliplatin, the down-regulated expression of miR-200c could lead the resistance to oxaliplatin (*Tanaka et al., 2015*). It is worth noting that miR-200c and miR-200b can reverse the resistance of epithelial ovarian cancer to cisplatin by targeting DNA methyltransferase (DNMT) (*Liu et al., 2019a*). *Zhang, Li & Zhang (2020)* found that MALAT1 modulated *ZFP91* through sponging miR-22-3p to promote oxaliplatin resistance in gastric cancer cells. MALAT1 was significantly up-regulated in CRC cells and tissues which wre resistant to oxaliplatin. MALAT1 modulated *ADAM17 via* sponging miR-324-3p to enhance oxaliplatin resistance in CRC cells (*Fan et al., 2020*). *Liu et al. (2019b)* has fingered out that the down-regulated of MALAT1 improved the chemotherapeutic drug sensitivity and inhibited the cisplatin resistance of the bladder cancer. Interestingly, our results showed that *ERCC4*, miR-200c and MALAT1 were all associated with Cisplatin, which suggested that MALAT1-miR-200c-3p-*ERCC4* axis may play a role in the Platinum-based chemotherapy for CRC patients.

However, the limitation of this study should be acknowledged. The sample size is relatively small and a larger sample verification is still needed. The biological relationship of MALAT1-miR-200c-3p-*ERCC4* axis and the biological functions they participate in are still unclear, and further *in vivo* and *in vitro* studies are needed. It would make more sense if we could obtain the chemotherapy information of patients in TCGA. We will further explore the underlying association between platinum-based chemotherapy and the network in future research.

CONCLUSION

In summary, we constructed a ceRNA network of *ERCC4* in CRC, of which the MALAT1-miR-200c-3p-*ERCC4* axis may be involved in the development, prognosis and chemotherapy sensitivity of CRC. This study will help to further reveal the molecular mechanism of *ERCC4* in CRC and provide new clues and insights for the further study of NER pathway. However, our results are only based on bioinformatic analysis and tissue validation of small samples, further *in vivo* and *in vitro* studies are needed to investigate the biological relationship of MALAT1-miR-200c-3p-*ERCC4* axis and its biological function.

ABBREVIATIONS

ACC	Adrenocortical carcinoma
BLCA	Bladder Urothelial carcinoma
BRCA	Breast Invasive Carcinoma
ceRNA	competitive endogenous RNA
CCLE	Cancer Cell Line Encyclopedia
CESC	Cervical squamous cell carcinoma and endocervical adenocarcinoma
CHOL	Cholangiocarcinoma
COAD	Colon adenocarcinoma
CRC	Colorectal cancer
DLBC	Lymphoid Neoplasm Diffuse Large B-cell Lymphoma
DNMT	DNA methyltransferase
EMT	epithelial-mesenchymal transition
ESCA	Esophageal carcinoma
GBM	Glioblastoma multiforme
GDSC	Drug Sensitivity in Cancer and CellMiner
HNSC	Head and Neck squamous cell carcinoma
HPA	the Human Protein Atlas
KICH	Kidney Chromophobe
KIRC	Kidney renal clear cell carcinoma
KIRP	Kidney renal papillary cell carcinoma
LAML	Acute Myeloid Leukemia
LGG	Brain Lower Grade Glioma
LIHC	Liver hepatocellular carcinoma
LUAD	Lung adenocarcinoma
MESO	Mesothelioma
NER	Nucleotide Excision Repair
OS	Overall survival
OV	Ovarian serous cystadenocarcinoma
PAAD	Pancreatic adenocarcinoma
PCPG	Pheochromocytoma and Paraganglioma
PRAD	Prostate adenocarcinoma
READ	Rectum adenocarcinoma
SARC	Sarcoma
SKCM	Skin Cutaneous Melanoma
STAD	stomach and esophageal carcinoma
TGCT	Testicular Germ Cell Tumors
THCA	Thyroid carcinoma
THYM	Thymoma

UCEC	Uterine Corpus Endometrial Carcinoma
UCS	Uterine Carcinosarcoma
UVM	Uveal melanoma

Funding

This work was supported by the National Natural Science Foundation of China (Ref. No. 31800124) and the Key Project of the Natural Science Foundation of Liaoning Province (Ref. No. 20180540037). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:
National Natural Science Foundation of China: 31800124.
Natural Science Foundation of Liaoning Province: 20180540037.

Competing Interests

The authors declare that they have no competing interests.

Author Contributions

- Huixin Hu performed the experiments, authored or reviewed drafts of the paper, and approved the final draft.
- Songyi Liu analyzed the data, prepared figures and/or tables, and approved the final draft.
- Aining Chu analyzed the data, prepared figures and/or tables, and approved the final draft.
- Jing Chen analyzed the data, prepared figures and/or tables, and approved the final draft.
- Chengzhong Xing conceived and designed the experiments, authored or reviewed drafts of the paper, and approved the final draft.
- Jingjing Jing conceived and designed the experiments, authored or reviewed drafts of the paper, and approved the final draft.

Human Ethics

The following information was supplied relating to ethical approvals (*i.e.*, approving body and any reference numbers):

This study was approved by the Ethics Committee of the First Affiliated Hospital of China Medical University Shenyang, China.

Data Availability

The basic information of the cases and original data of qRT-PCR are available in the Supplemental Files.

Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj.12647#supplemental-information.

REFERENCES

Altaha R, Liang X, Yu JJ, Reed E. 2004. Excision repair cross complementing-group 1: gene expression and platinum resistance. *International Journal of Molecular Medicine* **14**:959–970.

Anastasiadou E, Jacob LS, Slack FJ. 2018. Non-coding RNA networks in cancer. *Nature Reviews Cancer* **18(1)**:5–18 DOI 10.1038/nrc.2017.99.

Castro-Oropeza R, Melendez-Zajgla J, Maldonado V, Vazquez-Santillan K. 2018. The emerging role of lncRNAs in the regulation of cancer stem cells. *Cellular Oncology* **41(6)**:585–603 DOI 10.1007/s13402-018-0406-4.

de Laat WL, Jaspers NG, Hoeijmakers JH. 1999. Molecular mechanism of nucleotide excision repair. *Genes & Development* **13(7)**:768–785 DOI 10.1101/gad.13.7.768.

Du Y, Zhang Z, Xiong W, Li N, Liu H, He H, Li Q, Liu Y, Zhang L. 2019. Estradiol promotes EMT in endometriosis via MALAT1/miR200s sponge function. *Reproduction* **157(2)**:179–188 DOI 10.1530/REP-18-0424.

Enzlin JH, Scharer OD. 2002. The active site of the DNA repair endonuclease XPF-ERCC1 forms a highly conserved nuclease motif. *The EMBO Journal* **21(8)**:2045–2053 DOI 10.1093/emboj/21.8.2045.

Evans E, Moggs JG, Hwang JR, Egly JM, Wood RD. 1997. Mechanism of open complex and dual incision formation by human nucleotide excision repair factors. *The EMBO Journal* **16(21)**:6559–6573 DOI 10.1093/emboj/16.21.6559.

Fan C, Yuan Q, Liu G, Zhang Y, Yan M, Sun Q, Zhu C. 2020. Long non-coding RNA MALAT1 regulates oxaliplatin-resistance via miR-324-3p/ADAM17 axis in colorectal cancer cells. *Cancer Cell International* **20(1)**:473 DOI 10.1186/s12935-020-01549-5.

Friedberg EC. 2001. How nucleotide excision repair protects against cancer. *Nature Reviews Cancer* **1(1)**:22–33 DOI 10.1038/35094000.

Fuss JO, Cooper PK. 2006. DNA repair: dynamic defenders against cancer and aging. *PLOS Biology* **4(6)**:e203 DOI 10.1371/journal.pbio.0040203.

Gillet LC, Scharer OD. 2006. Molecular mechanisms of mammalian global genome nucleotide excision repair. *Chemical Reviews* **106(2)**:253–276 DOI 10.1021/cr040483f.

Gregory PA, Bert AG, Paterson EL, Barry SC, Tsykin A, Farshid G, Vadas MA, Khew-Goodall Y, Goodall GJ. 2008. The miR-200 family and miR-205 regulate epithelial to mesenchymal transition by targeting ZEB1 and SIP1. *Nature Cell Biology* **10(5)**:593–601 DOI 10.1038/ncb1722.

He XF, Liu LR, Wei W, Liu Y, Su J, Wang SL, Shen XL, Yang XB. 2014. Association between the XPG Asp1104His and XPF Arg415Gln polymorphisms and risk of cancer: a meta-analysis. *PLOS ONE* **9(5)**:e88490 DOI 10.1371/journal.pone.0088490.

Huang G, Wei B, Chen Z, Wang J, Zhao L, Peng X, Liu K, Lai Y, Ni L. 2020. Identification of a four-microRNA panel in serum as promising biomarker for colorectal carcinoma detection. *Biomarkers in Medicine* **14(9)**:749–760 DOI 10.2217/bmm-2019-0605.

Iyama T, Wilson DM III. 2013. DNA repair mechanisms in dividing and non-dividing cells. *DNA Repair* **12(8)**:620–636 DOI 10.1016/j.dnarep.2013.04.015.

Ji P, Diederichs S, Wang W, Boing S, Metzger R, Schneider PM, Tidow N, Brandt B, Buerger H, Bulk E, Thomas M, Berdel WE, Serve H, Muller-Tidow C. 2003. MALAT-1, a novel

noncoding RNA, and thymosin beta4 predict metastasis and survival in early-stage non-small cell lung cancer. *Oncogene* **22(39)**:8031–8041 DOI 10.1038/sj.onc.1206928.

Jiang Y, Ji X, Liu K, Shi Y, Wang C, Li Y, Zhang T, He Y, Xiang M, Zhao R. 2020. Exosomal miR-200c-3p negatively regulates the migraion and invasion of lipopolysaccharide (LPS)-stimulated colorectal cancer (CRC). *BMC Molecular and Cell Biology* **21(1)**:48 DOI 10.1186/s12860-020-00291-0.

Kabzinski J, Majsterek I, Dziki A, Mik M. 2015. The role of the XPF gene polymorphism (Xrcc4) Ser835ser in the risk of malignant transformation of cells in the colorectal cancer. *Polish Journal of Surgery* **87(2)**:83–85 DOI 10.1515/pjs-2015-0023.

Li P, Ma Y. 2018. Correlation of xeroderma pigmentosum complementation group F expression with gastric cancer and prognosis. *Oncology Letters* **16**:6971–6976 DOI 10.3892/ol.2018.9529.

Lin R, Maeda S, Liu C, Karin M, Edgington TS. 2007. A large noncoding RNA is a marker for murine hepatocellular carcinomas and a spectrum of human carcinomas. *Oncogene* **26(6)**:851–858 DOI 10.1038/sj.onc.1209846.

Liu J, He C, Xing C, Yuan Y. 2014. Nucleotide excision repair related gene polymorphisms and genetic susceptibility, chemotherapeutic sensitivity and prognosis of gastric cancer. *Mutation Research* **765(1)**:11–21 DOI 10.1016/j.mrfmmm.2014.04.008.

Liu P, Li X, Cui Y, Chen J, Li C, Li Q, Li H, Zhang X, Zu X. 2019b. LncRNA-MALAT1 mediates cisplatin resistance via miR-101-3p/VEGF-C pathway in bladder cancer. *Acta Biochimica et Biophysica Sinica* **51**:1148–1157 DOI 10.1093/abbs/gmz112.

Liu J, Zhang X, Huang Y, Zhang Q, Zhou J, Zhang X, Wang X. 2019a. miR-200b and miR-200c co-contribute to the cisplatin sensitivity of ovarian cancer cells by targeting DNA methyltransferases. *Oncology Letters* **17**:1453–1460 DOI 10.3892/ol.2018.9745.

McDaniel LD, Schultz RA. 2008. XPF/ERCC4 and ERCC1: their products and biological roles. *Yeast Membrane Transport* **637(5)**:65–82 DOI 10.1007/978-0-387-09599-8_8.

Mu D, Hsu DS, Sancar A. 1996. Reaction mechanism of human DNA repair excision nuclease. *Journal of Biological Chemistry* **271(14)**:8285–8294 DOI 10.1074/jbc.271.14.8285.

Mutlu M, Raza U, Saatci O, Eyupoglu E, Yurdusev E, Sahin O. 2016. miR-200c: a versatile watchdog in cancer progression, EMT, and drug resistance. *Journal of Molecular Medicine* **94(6)**:629–644 DOI 10.1007/s00109-016-1420-5.

Nabih HK. 2020. Crosstalk between NRF2 and Dicer through metastasis regulating MicroRNAs; mir-34a, mir-200 family and mir-103/107 family. *Archives of Biochemistry and Biophysics* **686(10)**:108326 DOI 10.1016/j.abb.2020.108326.

Neven KY, Saenen ND, Tarantini L, Janssen BG, Lefebvre W, Vanpoucke C, Bollati V, Nawrot TS. 2018. Placental promoter methylation of DNA repair genes and prenatal exposure to particulate air pollution: an ENVIRONAGE cohort study. *The Lancet Planetary Health* **2(4)**:e174–e183 DOI 10.1016/S2542-5196(18)30049-4.

Ning J, Jiao Y, Xie X, Deng X, Zhang Y, Yang Y, Zhao C, Wang H, Gu K. 2019. miR1385p modulates the expression of excision repair crosscomplementing proteins ERCC1 and ERCC4, and regulates the sensitivity of gastric cancer cells to cisplatin. *Oncology Reports* **41**:1131–1139 DOI 10.3892/or.2018.6907.

Park SM, Gaur AB, Lengyel E, Peter ME. 2008. The miR-200 family determines the epithelial phenotype of cancer cells by targeting the E-cadherin repressors ZEB1 and ZEB2. *Genes & Development* **22(7)**:894–907 DOI 10.1101/gad.1640608.

Qiu J, Wang X, Meng X, Zheng Y, Li G, Ma J, Ye G, Li Y, Li J. 2014. Attenuated NER expressions of XPF and XPC associated with smoking are involved in the recurrence of bladder cancer. *PLOS ONE* **9(12)**:e115224 DOI 10.1371/journal.pone.0115224.

Sa MC, Conceicao TS, de Moura Santos E, de Morais EF, Galvao HC, de Almeida Freitas R. 2020. Immunohistochemical expression of TFIIH and XPF in oral tongue squamous cell carcinoma. *European Archives of Oto-Rhino-Laryngology* **277**(3):893–902 DOI 10.1007/s00405-019-05757-2.

Salmena L, Poliseno L, Tay Y, Kats L, Pandolfi PP. 2011. A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? *Cell* **146**(3):353–358 DOI 10.1016/j.cell.2011.07.014.

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research* **13**(11):2498–2504 DOI 10.1101/gr.1239303.

Sijbers AM, de Laat WL, Ariza RR, Biggerstaff M, Wei YF, Moggs JG, Carter KC, Shell BK, Evans E, de Jong MC, Rademakers S, de Rooij J, Jaspers NG, Hoeijmakers JH, Wood RD. 1996. Xeroderma pigmentosum group F caused by a defect in a structure-specific DNA repair endonuclease. *Cell* **86**(5):811–822 DOI 10.1016/s0092-8674(00)80155-5.

Slyskova J, Korenkova V, Collins AR, Prochazka P, Vodickova L, Svec J, Lipska L, Levy M, Schneiderova M, Liska V, Holubec L, Kumar R, Soucek P, Naccarati A, Vodicka P. 2012. Functional, genetic, and epigenetic aspects of base and nucleotide excision repair in colorectal carcinomas. *Clinical Cancer Research* **18**(21):5878–5887 DOI 10.1158/1078-0432.CCR-12-1380.

Sreekumar R, Al-Saihati H, Emaduddin M, Moutasim K, Mellone M, Patel A, Kilic S, Cetin M, Erdemir S, Navio MS, Lopez MA, Curtis N, Yagci T, Primrose JN, Price BD, Berx G, Thomas GJ, Tulchinsky E, Mirnezami A, Sayan AE. 2021. The ZEB2-dependent EMT transcriptional programme drives therapy resistance by activating nucleotide excision repair genes ERCC1 and ERCC4 in colorectal cancer. *Molecular Oncology* **15**(8):2065–2083 DOI 10.1002/1878-0261.12965.

Tanaka S, Hosokawa M, Yonezawa T, Hayashi W, Ueda K, Iwakawa S. 2015. Induction of epithelial-mesenchymal transition and down-regulation of miR-200c and miR-141 in oxaliplatin-resistant colorectal cancer cells. *Biological & Pharmaceutical Bulletin* **38**(3):435–440 DOI 10.1248/bpb.b14-00695.

Tripsianes K, Folkers G, Ab E, Das D, Odijk H, Jaspers NG, Hoeijmakers JH, Kaptein R, Boelens R. 2005. The structure of the human ERCC1/XPF interaction domains reveals a complementary role for the two proteins in nucleotide excision repair. *Structure* **13**(12):1849–1858 DOI 10.1016/j.str.2005.08.014.

Tsodikov OV, Enzlin JH, Scharer OD, Ellenberger T. 2005. Crystal structure and DNA binding functions of ERCC1, a subunit of the DNA structure-specific endonuclease XPF-ERCC1. *Proceedings of the National Academy of Sciences of the United States of America* **102**(32):11236–11241 DOI 10.1073/pnas.0504341102.

Wang M, Wang M, Yuan L, Wu D, Zhang Z, Yin C, Fu G, Wei Q, Zhang Z. 2010. A novel XPF-357A>C polymorphism predicts risk and recurrence of bladder cancer. *Oncogene* **29**(13):1920–1928 DOI 10.1038/onc.2009.484.

Wei ZH, Guo WH, Wu J, Suo WH, Fu GH. 2014. A nonsense mutation in the Xeroderma pigmentosum complementation group F (XPF) gene is associated with gastric carcinogenesis. *Gene* **537**(2):238–244 DOI 10.1016/j.gene.2013.12.061.

Wei JW, Huang K, Yang C, Kang CS. 2017. Non-coding RNAs as regulators in epigenetics (Review). *Oncology Reports* **37**(1):3–9 DOI 10.3892/or.2016.5236.

Wu C, Zhu X, Tao K, Liu W, Ruan T, Wan W, Zhang C, Zhang W. 2018. MALAT1 promotes the colorectal cancer malignancy by increasing DCP1A expression and miR203 downregulation. *Molecular Carcinogenesis* **57**(10):1421–1431 DOI 10.1002/mc.22868.

Xiao H, Tang K, Liu P, Chen K, Hu J, Zeng J, Xiao W, Yu G, Yao W, Zhou H, Li H, Pan Y, Li A, Ye Z, Wang J, Xu H, Huang Q. 2015. LncRNA MALAT1 functions as a competing endogenous RNA to regulate ZEB2 expression by sponging miR-200s in clear cell kidney carcinoma. *Oncotarget* **6(35)**:38005–38015 DOI 10.18632/oncotarget.5357.

Xie X, Huang N, Zhang Y, Wei X, Gao M, Li M, Ning J, Liu W, Zhao Q, Wang H, Gu K. 2019. MiR-192-5p reverses cisplatin resistance by targeting ERCC3 and ERCC4 in SGC7901/DDP cells. *Journal of Cancer* **10(4)**:1039–1051 DOI 10.7150/jca.25814.

Xu J, Xiao Y, Liu B, Pan S, Liu Q, Shan Y, Li S, Qi Y, Huang Y, Jia L. 2020. Exosomal MALAT1 sponges miR-26a/26b to promote the invasion and metastasis of colorectal cancer via FUT4 enhanced fucosylation and PI3K/Akt pathway. *Journal of Experimental & Clinical Cancer Research* **39(1)**:54 DOI 10.1186/s13046-020-01562-6.

Yang H, Li G, Li WF. 2015. Association between ERCC1 and XPF polymorphisms and risk of colorectal cancer. *Genetics and Molecular Research* **14(1)**:700–705 DOI 10.4238/2015.January.30.13.

Yuan W, Li X, Liu L, Wei C, Sun D, Peng S, Jiang L. 2019. Comprehensive analysis of lncRNA-associated ceRNA network in colorectal cancer. *Biochemical and Biophysical Research Communications* **508(2)**:374–379 DOI 10.1016/j.bbrc.2018.11.151.

Zhang X, Hamblin MH, Yin KJ. 2017. The long noncoding RNA Malat1: its physiological and pathophysiological functions. *RNA Biology* **14(12)**:1705–1714 DOI 10.1080/15476286.2017.1358347.

Zhang N, Hu X, Du Y, Du J. 2021. The role of miRNAs in colorectal cancer progression and chemoradiotherapy. *Biomedicine & Pharmacotherapy* **134(10207)**:111099 DOI 10.1016/j.biopha.2020.111099.

Zhang Z, Li M, Zhang Z. 2020. lncRNA MALAT1 modulates oxaliplatin resistance of gastric cancer via sponging miR-22-3p. *OncoTargets and Therapy* **13**:1343–1354 DOI 10.2147/OTT.

Zhang Y, Wu S, Zhou X, Huang F, Chen R, Wang Y, Wu J. 2019. Association between nucleotide excision repair gene polymorphism and colorectal cancer risk. *Journal of Clinical Laboratory Analysis* **33(8)**:e22956 DOI 10.1002/jcla.22956.

Zhong ME, Chen Y, Zhang G, Xu L, Ge W, Wu B. 2019. LncRNA H19 regulates PI3K-Akt signal pathway by functioning as a ceRNA and predicts poor prognosis in colorectal cancer: integrative analysis of dysregulated ncRNA-associated ceRNA network. *Cancer Cell International* **19(1)**:148 DOI 10.1186/s12935-019-0866-2.

Zhu Y, Bian Y, Zhang Q, Hu J, Li L, Yang M, Qian H, Yu L, Liu B, Qian X. 2019. Construction and analysis of dysregulated lncRNA-associated ceRNA network in colorectal cancer. *Journal of Cellular Biochemistry* **120(6)**:9250–9263 DOI 10.1002/jcb.28201.

Zhuang M, Zhao S, Jiang Z, Wang S, Sun P, Quan J, Yan D, Wang X. 2019. MALAT1 sponges miR-106b-5p to promote the invasion and metastasis of colorectal cancer via SLAIN2 enhanced microtubules mobility. *EBioMedicine* **41(1)**:286–298 DOI 10.1016/j.ebiom.2018.12.049.