Comprehensive Analysis of Differentially Expressed Long Non-Coding RNA-mRNA in the Adenoma–Carcinoma Sequence of DNA Mismatch Repair Proficient Colon Cancer

Wenkun Li
Beijing Shijitan Hospital, Capital Medical University
Qian Li
Beijing Shijitan Hospital, Capital Medical University
Jiang Ge
Beijing Shijitan Hospital, Capital Medical University
Yun Wang
Peking University Ninth School of Clinical Medicine
Nanshan Li
Beijing Shijitan Hospital, Capital Medical University
Yueqiong Lao
Peking University Ninth School of Clinical Medicine
Yadan Wang
Beijing Shijitan Hospital, Capital Medical University
Kuiliang Liu
Beijing Shijitan Hospital, Capital Medical University
Chunmei Guo
Beijing Shijitan Hospital, Capital Medical University
Wu Lin
Beijing Shijitan Hospital, Capital Medical University
Guojun Jiang
Beijing Shijitan Hospital, Capital Medical University
Nan Wei
Beijing Shijitan Hospital, Capital Medical University
Canghai Wang
Beijing Shijitan Hospital, Capital Medical University
Hong Liu
Beijing Shijitan Hospital, Capital Medical University
Jing Wu (bjsjtyywj@ccmu.edu.cn)
Beijing Friendship Hospital, Capital Medical University https://orcid.org/0000-0001-8838-8692

Research article

Keywords: Colon cancer, long non-coding RNA, DNA mismatch repair proficient, adenoma–carcinoma sequence, microarray

Posted Date: July 27th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-47627/v1

Abstract

Background DNA mismatch repair proficient colon cancer (pMMR CC) is the most common subtype of sporadic CC. We aimed to investigate the role of long non-coding RNAs (lncRNAs) in pMMR CC carcinogenesis.

Methods In the present study, we conducted transcriptomic analysis of lncRNAs-mRNAs in five low-grade intraepithelial neoplasia (LGIN), five high-grade intraepithelial neoplasia (HGIN), four pMMR CC, and five normal control (NC) tissues. Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment pathway, and Co-Expression Network analyses were performed to elucidate the functions of lncRNAs and mRNAs as well as their interactions. Quantitative real-time polymerase chain reaction (qRT-PCR) was used to validate five dysregulated lncRNAs in a large set of colon tissues. Receiver-operating characteristic (ROC) curves were employed to evaluate the performance of the candidate lncRNAs.

Results A set of 5783 differentially expressed lncRNAs and 4483 differentially expressed mRNAs were detected among the LGIN, HGIN, pMMR CC, and NC samples. These differentially expressed lncRNAs and mRNAs were assigned to 275 significant GO terms and 179 significant KEGG enriched pathways. qRT-PCR confirmed that the expression of five selected lncRNAs (ENST00000521815, ENST00000603052, ENST00000609220, NR_026543, and ENST00000545920) were consistent with the microarray data. ROC analysis showed that four lncRNAs (ENST00000521815, ENST00000603052, ENST00000609220, and NR_026543) had larger area under the ROC curve (AUC) values than serum carcinoembryonic antigens, thereby distinguishing NC from pMMR CC.

Conclusions Several lncRNAs play various roles in the adenoma–carcinoma sequence and may serve as potential biomarkers for the early diagnosis of pMMR CC.

Background

Colorectal cancer (CRC) is the third most commonly diagnosed cancer worldwide, comprising over 1.8 million new cases in 2018. The disease is also the second leading cause of cancer deaths [1]. More than 70% of CRC cases are located in the colon, which is also called colon cancer (CC) [2]. With the development of the economy and the westernization of diets and lifestyles, the incidence of CRC is likely to increase in Asia. According to the latest data from the Chinese National Cancer Center, CRC ranks fifth in both incidence and mortality among cancers in China [3].

CRC is a heterogeneous disease with a progressive accumulation of genetic and epigenetic alterations [4–6]. About two-thirds of CRC cases arise with unknown contributions from germline factors or a significant family history of cancer or inflammatory bowel disease, defined as sporadic CRC. Three major genetic mechanisms underlie the pathogenesis of sporadic CRC [7], namely chromosomal instability (CIN), microsatellite instability (MSI), and the epigenomic CpG island methylator phenotype (CIMP). CIN, the first described and the most common pathway, accounts for 80%-85% of sporadic CRC [8]. CIN CRCs are non-hypermutated with DNA proficient mismatch repair (pMMR) functions [7]. The mechanism responsible for CIN is different from that responsible for MSI and CIMP. Understanding the molecular mechanisms underlying CRC is critically important for the clinical prognosis and therapeutic response. A meta-analysis of 63 eligible studies (10126 patients with CRC) showed a worse prognosis for CRC patients with CIN/pMMR. CIN and MSI status can be evaluated as a predictor of prognosis [9]. Additionally, stage II CRC patients with defective mismatch repair (dMMR) or high-level MSI (MSI-H) have better prognosis but no benefits from fluorouracil chemotherapy [10, 11].

More than 90% of colorectal tumorigenesis cases follow the adenoma–carcinoma sequence (ACS), with multiple gene mutations and abnormal activation of signaling pathways [12, 13]. Mutations in the adenomatous polyposis coli (APC) gene occur early during colorectal tumorigenesis, followed by the activation of the KRAS gene and the inactivation of the *TP*53 gene, which are often associated with CIN [14].

Long non-coding RNAs (lncRNAs) are a class of non-coding RNAs with more than 200 nucleotides that are not translated into proteins. lncRNAs can regulate gene expression at different levels, including epigenetically, transcriptionally, and post-

transcriptionally [15]. Existing studies have shown the aberrant expression of lncRNAs is associated with tumorigenesis, metastasis, and prognosis in various cancers. For instance, decreased expression of the lncRNA SATB2-AS1 promotes metastasis and influences the tumor microenvironment in CRC by regulating SATB2, thereby resulting in poor survival [16]. Overexpression of the lncRNA UCA1 contributes to the immune escape of cancer cells and protects PDL1 expression from miRNA repression in gastric carcinoma, serving as a potentially novel target for immunotherapy [17]. Aberrant expression of lncRNAs is involved in the evolution of several types of tumors; however, the differential expression and the role of lncRNAs in ACS for CRC remain to be elucidated.

Previous studies on the lncRNA expression profiles of CC were generally performed on all cancers in the colon together including rectal cancer, thereby neglecting the heterogeneity of molecular subtypes of CC. Although often linked together, CC and rectal cancer are different when it comes to treatment. Thus far, the differential expression profile for lncRNAs in the normal colonic mucosa-pMMR adenoma-pMMR sporadic CC sequence has not been reported. In the present study, we aimed to characterize the expression profile of lncRNAs in the malignant evolution process for pMMR sporadic CC using transcriptome microarray technology. Subsequently, differentially expressed lncRNAs were investigated using bioinformatics analysis and validated with quantitative real-time polymerase chain reaction (qRT-PCR). The findings provide useful candidates for the diagnosis and treatment of CC.

Methods

Clinical samples

A total of 244 sporadic colonic tissues, including 63 adenomas of low-grade intraepithelial neoplasia (LGIN), 32 adenomas of high-grade intraepithelial neoplasia (HGIN), 66 pMMR CC, eight dMMR CC, and 75 normal control (NC) samples from adjacent (≥ 10 cm) hyperplastic or inflammatory polyps were sampled at Beijing Shijitan Hospital (Beijing, China) from 2018 to 2019. The samples were immediately placed in RNALater solution (Cat. No. 76104, Qiagen Co, GmBH, Germany) for 24 hours at 4 °C. Freshly frozen samples were stored at -80℃ for RNA extraction. None of the patients received targeted therapy, chemotherapy, radiotherapy, or intervention therapy.

Screening of pMMR samples and total RNA isolation

As in our previous study [18], multiplex PCR and immunohistochemistry were used to determine the MMR status. pMMR colonic tissues were defined as tissue samples with a low frequency of MSI (MSI-L) or microsatellite stability (MSS, no marker tested). Total RNA was extracted from the tissue samples using RNAiso Plus (Code No. 9109, TAKARA, Japan). RNA integrity was confirmed with a NanoDrop ND-2000 spectrophotometer (Thermo Fisher Scientific, Rochester, NY, USA) and Agilent Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA, US), and further purified with an RNeasy mini kit (Cat. # 74106, Qiagen, GmBH, Germany) and RNase-Free DNase Set (Cat. #79254, Qiagen Co, GmBH, Germany). RNA samples with integrity ≥ 7.0 and a 28S:18S ratio ≥ 0.7 were used for further experiments.

Microarrays analysis

The lncRNA and mRNA expression profiles in multistage colonic mucosa tissues (five NC, five LGIN, five HGIN, and four pMMR CC samples) were determined with the Agilent custom SBC Human (4*180 K) lncRNA Microarray V6.0 (Product No. G4862A-074348; Agilent Technologies, Santa Clara, CA). Total RNA was amplified and labeled with the Low Input Quick Amp WT Labeling Kit (Cat. No. 5190 – 2943, Agilent Technologies, Santa Clara, CA, US) following the manufacturer's instructions. Labeled cRNAs were purified with the RNeasy mini kit (Cat. No. 74106, Qiagen, GmBH, Germany). Each slide was hybridized with 1.65 μg Cy3-labeled cRNA using a Gene Expression Hybridization Kit (Cat. No. #5188–5242, Agilent Technologies, Santa Clara, CA, US) in a hybridization oven (Cat. No. # G2545A, Agilent Technologies, Santa Clara, CA, US) for 17 h. The slides were washed in staining dishes (Cat. No. 121, Thermo Shandon, Waltham, MA, US) with the Gene Expression Wash Buffer Kit (Cat.

No. #5188–5327, Agilent Technologies, Santa Clara, CA, US) following the manufacturer's instructions and scanned with an Agilent Microarray Scanner (Cat. No. # G2565CA, Agilent Technologies, Santa Clara, CA, US) with default settings (Dye channel: Green, Scan resolution = 3 μm, PMT 100%, 20 bit). The data were extracted with Feature Extraction software 10.7 (Agilent Technologies, Santa Clara, CA, US).

Data processing and identification of DEGs

Series test of cluster (STC) analysis

Raw data were normalized with the Quantile algorithm in GeneSpring Software 11.0 (Agilent Technologies, Santa Clara, CA, US). Cluster analysis was performed to systematically and intuitively display the relatedness between samples. Differential genes expression (DEGs) among the four groups were filtered with the F test in the random variance model (RVM) using P values < 0.05 and Benjamini-Hochberg false discovery rate (FDR) < 0.05 as the significant cutoff criteria [19]. STC analysis was employed to identify the most probable set of clusters responsible for the observed ACS series by profiling the dynamic nature of the temporal gene expression. NC, LGIN, HGIN, and pMMR CC were set as different points to identify significant trending models related to multistage colonic mucosa tissue and their associated DEGs by STC analysis. Clusters with P values < 0.05/26 were considered statistically significant.

Gene Ontology and Kyoto Encyclopedia of Genes and Genomes pathway analysis Gene ontology (GO) analysis was performed to annotate DEGs in three categories: molecular functions, biological processes, and cellular components. Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis was performed to identify the functional pathways and pathways for molecular interactions, reactions, and relation networks. The statistically significant threshold was set as P value < 0.01 and FDR < 0.05.

Construction of co-expression and transcription factor (TF) regulatory network

To identify co-expressed lncRNA-mRNA pairs and hub DEGS, the lncRNA-mRNA co-expression network was constructed based on the gene expression value and the Pearson's correlation coefficient (PCC) between the differentially expressed lncRNAs and mRNAs. $|PCC| \geq 0.8$ and P value < 0.01 were considered as statistically significantly relevant. In order to distinguish the TF regulatory network in the stepwise process encompassing CC progression, the interactions were predicted starting from a TF by searching the conserved TF binding sites within a putative promoter area 2000 bp upstream and 500 bp downstream of the transcriptional initiation site of co-expressed genes.

Validation by quantitative real-time PCR (qRT-PCR)

The expression levels of five lncRNAs were examined by qRT-PCR in 225 colorectal tissue samples. qRT-PCR was performed with a QuantStudioTM 6 Flex Real Time PCR System (Thermo Fisher, Rochester, NY, USA). Primers are listed in Additional file 1 and glyceraldehyde 3-phosphate dehydrogenase (GAPDH) was used as the endogenous control. Each qRT-PCR reaction (in 10 μL) contained 5 μL 2X SYBR Green PCR Mix (ABI, USA), 0.4 μM each of the forward and reverse primers, and 5 ng of template cDNA. The PCR program was as follows: initial denaturing at 95 °C for 5 min followed by 40 cycles at 94 °C for 30 s and 60 °C for 1 min and fluorescence acquisition at 60 °C for 1 min. PCR amplifications were performed in triplicate for each sample.

Statistical analysis

Computer-based calculations were conducted using SPSS version 20.0 (SPSS Inc., Chicago, IL, USA) and MedCalc version 19.1 (MedCalc Software bvba, Ostend, Belgium). The differences in the expression of selected lncRNAs among multiple groups were compared using ANOVA. Further pairwise comparisons were performed with the least significance difference (LSD) test. The differences in the expression of lncRNAs between the pMMR CC group and the dMMR CC group were compared with the

nonparametric Mann–Whitney U test. The relationship between lncRNAs and their target coding genes were determined by Spearman correlation analysis. Receiver-operating characteristic (ROC) curves and the area under the ROC curve (AUC) were employed to evaluate the diagnostic accuracy of candidate lncRNAs. The data were presented as the mean with standard deviation (SD) for normally distributed data and the median with the interquartile range for skewed data. |Fold change| ≥ 5 was used as the significant threshold to screen differentially expressed lncRNA and mRNA. All P values were two-sided and FDR was calculated for multiple comparisons. Differences with $P < 0.05$ were considered as statistically significant.

Results

Basic characteristics of the microarray data

The total RNA extracted from pMMR colonic tissue samples, consisting of five NC, five LGIN, five HGIN, and four pMMR CC samples, were hybridized to a transcriptome microarray. The quality control results for RNA/microarray are summarized in Additional file 1 and Additional file 2, respectively. Cluster analysis of normalized data indicated that the samples from each group were well-separated when all DEGs were considered (Fig. 1), suggesting the microarray data were reliable for further bioinformatics analysis.

Dysregulated lncRNAs and mRNAs in the colonic adenoma-carcinoma sequence

A total of 5783 lncRNAs and 4483 mRNAs were differentially expressed among the NC, LGIN, HGIN, and pMMR CC samples. In order to generate dynamic and significant trending models of DEGs related to colon mucosa malignant transformation, STC analysis was used to further identify the clusters of DEGs with similar expression patterns. As shown in Fig. 2A, the differential expression profile consisted of 26 differentially expressed lncRNA clusters, among which 15 clusters (profile No. 0, 1, 2, 3, 4, 5, 9, 12, 17, 20, 21, 22, 23, 24, and 25) showed significant expression trends ($P < 0.05$). Moreover, the lncRNA expression levels in clusters 21, 22, 24, and 25 were stable or gradually elevated, while the lncRNA expression in clusters 0, 1, 3, 4, 9, and 12 showed a decreasing trend. Cluster 23 contained 207 lncRNAs that gradually increased from NC to adenoma and were suppressed in CC (Fig. 2B). However, lncRNA expression in cluster 3 was opposite that in cluster 23. As shown in Fig. 2A, a total of nine mRNA expression clusters, specifically clusters 2, 4, 5, 13, 20, 21, 22, 24, and 25, showed significant expression trends ($P < 0.05$), among which clusters 21, 22, 24, and 25 showed a stable or gradual elevation. The results suggested that adenoma is an intermediate step from normal tissue to CC, and certain lncRNAs may play important roles during the dynamic process in colonic mucosal protruding lesions.

Functional and pathway enrichment analysis

DEGs from significant expression trends were subjected to GO and KEGG pathway analyses to identify their potential functions and mechanisms in the dynamic process for colonic mucosal protruding lesions (Fig. 3).

The top 20 GO terms included multicellular organismal development, cell division, cell adhesion, cell cycle, cell differentiation, mitosis, proteolysis, and angiogenesis. KEGG enrichment analysis revealed the DEGs were mainly involved in several signaling pathways, including metabolism, cancer, cell cycle, PI3K-Akt signaling, and transcriptional misregulation in cancer (Fig. 3B).

Construction of the lncRNA-mRNA co-expression networks

To discover the significant molecular mechanisms for the lncRNAs associated with tumorigenesis in colonic ACS, lncRNAs from elevated clusters (21, 22, 24, 25) or reduced clusters (0, 1, 3, 4, 9, 12) and mRNAs with significant GO terms and KEGG enrichment pathways were selected to construct elevated-lncRNA-mRNA co-expression networks (Co-expression network E) and decreased-lncRNAs-mRNA co-expression networks (Co-expression network D), respectively. The top 10 lncRNA /mRNA in terms

by degree are listed in Tables 1 and 3. Co-expression network E contained 714 nodes and 1711 edges (Fig. 4A), among which the hub nodes with the highest degrees were NOTCH4 (mRNA, Degree = 29), INHBA (mRNA, Degree = 26), TSTA3 (mRNA, Degree = 29), lnc-NPRL3-1:1 (lncRNA, Degree = 22), NR_024431 (lncRNA, Degree = 21), and ENST00000564984 (lncRNA, Degree = 17). Co-expression network D contained 598 nodes and 1908 edges (Fig. 4B) and the DEGs with the highest degrees were AQP8 (mRNA, Degree = 47), STRADB (mRNA, Degree = 47), ENST00000435912 (lncRNA, Degree = 15), NR_024431 (lncRNA, Degree = 21), and ENST00000564984 (lncRNA, Degree = 17) (Tables 1 and 3).

Table 1
The top 10 lncRNAs with the highest degree in co-expression network E/D

Gene	NC	LGIN	HGIN	CC	p-value	FDR	Type	Degree
lncRNAs of elevated profiles								
lnc-NPRL3-1:1	1625.59	2834.43	2901.7	2777.36	0.0037928	0.0253	sense	22
NR_024431	7.42	18.65	19	18.6	0.0035591	0.0242	intergenic	21
ENST00000564984	9.17	18.63	19.15	18.83	0.002729	0.0204	antisense	17
lnc-CLEC3A-9:1	10892.39	20059.59	19686.96	19862.6	0.0007135	0.00894	sense	13
NR_029374	17.3	41.96	42.92	39.14	0.0064708	0.0358	antisense	12
lnc-MKI67IP-8:1	25.14	43.39	43.53	40.98	0.0044777	0.0282	antisense	12
NONHSAT064233	60.28	118.49	113.3	273.65	1.68e-005	0.000976	antisense	11
lnc-DLX2-9:1	13.74	32.64	32.78	30.29	0.0038192	0.0254	antisense	11
ENST00000545920	202.78	259.09	260.38	421.62	0.0019042	0.0162	bidirectional	10
lnc-MAP3K9-9:1	406.91	663.73	501	1392.28	0.0010763	0.0114	antisense	10
lncRNAs of descending profiles								
ENST00000435912	35.42	19.61	19.43	19.98	0.0083754	0.0425	antisense	15
lnc-PDZK1-1:1	1838.67	149.47	119.91	220.16	0.0004298	0.00669	antisense	14
lnc-MRPL54-2:1	200.94	123.77	123.05	125.75	0.0106357	0.0495	sense	14
ENST00000364025	7.73	3.84	3.78	3.92	0.0026283	0.0199	sense	14
lnc-CCNB3-2:1	617.57	101.05	95.87	115.2	4.63e-005	0.00179	intergenic	14
lnc-CCDC14-2:2	68.74	7.14	5.92	10.93	5.71e-005	0.00201	sense	13
NR_003064	29637.97	1140.93	703.58	1579.91	6e-007	0.000143	intergenic	13
lnc-SRSF5-1:1	8.49	3.84	3.77	4.14	1.61e-005	0.000956	intergenic	13
NR_110552	472.62	9.51	10.64	39.1	0.0001868	0.00406	antisense	13
lnc-ATF6B-1:4	41.99	5.78	5.88	8.03	3.64e-005	0.00154	sense	13

Table 3
The top 10 mRNAs with the highest degree in co-expression networks

co-expression network E		co-expression network D		A-co-expression network E		A-co-expression network D	
NOTCH4	29	AQP8	47	NOTCH4	35	CAMK2B	61
INHBA	26	STRADB	47	MAD2L1	34	STRADB	60
TSTA3	24	GRIA3	46	E2F7	34	LTK	57
RFC3	24	SCN7A	46	CCNE1	32	PDZK1	56
POLR3K	24	AGTR1	46	INHBA	32	FGF18	55
CCNE1	24	EEF2K	46	RNASEH2A	30	SEMA3E	54
ZYX	23	ALPI	46	RFC3	29	SCIN	54
ALYREF	23	CAMK2B	44	CLSPN	29	SST	53
BRCA2	22	CNTN1	44	NME1	29	IGF1	53
MAD2L1	21	IGF1	44	ERCC6L	28	BMP3	52

Angiogenesis is a crucial step in tumor growth and progression. The angiogenic switch has been observed at the adenoma stage in ACS [20]. We constructed angiogenesis-related lncRNA-mRNA co-expression networks (A-co-expression networks E and D) with DEGs from significantly elevated/decreased clusters as well as GO terms and KEGG pathways related to angiogenesis. A-co-expression network E consisted of 649 nodes and 2299 edges (Fig. 4C) and the top hub genes were NOTCH4 (mRNA, Degree = 35), MAD2L1 (mRNA, Degree = 34), E2F7 (mRNA, Degree = 34), lnc-ZBTB20-2:1 (lncRNA, Degree = 22), lnc-CENPH-2:1 (lncRNA, Degree = 20), and lnc-PAQR4-2:1 (lncRNA, Degree = 19) (Tables 2 and 3). The top hub genes in A-co-expression network D (Fig. 4D) were CAMK2B (mRNA, Degree = 61), STRADB (mRNA, Degree = 60), LTK (mRNA, Degree = 57), ENST00000513255 (lncRNA, Degree = 12), NR_024605 (lncRNA, Degree = 11), and NONHSAT057082 (lncRNA, Degree = 11) (Tables 2 and 3). All the DEGs primarily participated in angiogenesis as the key genes in each network.

Table 2
The top 10 lncRNAs with the highest degree in A-co-expression network E/D

Gene	NC	LGIN	HGIN	CC	p-value	FDR	Type	Degree
lncRNAs of elevated profiles								
lnc-ZBTB20-2:1	32.95	48.38	41.53	94.12	0.0001576	0.00367	sense	22
lnc-CENPH-2:1	718.1	1438.07	1182.1	3286.29	0.0003228	0.0056	sense	20
lnc-PAQR4-2:1	450.98	753.46	692.77	1489.34	0.0032726	0.0229	sense	19
lnc-SRGN-1:2	3.78	6.76	5.07	16.83	0.0059561	0.034	intergenic	19
NR_045669	94.62	135.86	126.36	246.74	8.6e-006	0.000666	antisense	18
lnc-GRASP-2:2	5	13.63	12.86	40.9	0.0012449	0.0125	sense	18
NR_036480	90.08	160.88	152.83	351.49	0.0001743	0.00391	antisense	18
lnc-BDKRB1-3:1	1199.21	1623.17	1474.82	2825.3	0.0005168	0.00739	sense	18
lnc-TXNDC3-1:1	146.23	288.55	263.33	624.69	0.001605	0.0146	intergenic	17
ENST00000455309	140.37	210.07	197.67	369.84	0.0093281	0.0455	antisense	17
lncRNAs of descending profiles								
ENST00000513255	193.19	58.84	55.95	67.08	< 1e-07	5.34e-005	antisense	12
NONHSAT057082	28.11	6.37	6.24	7.91	2.31e-005	0.00117	_sense	11
lnc-SRSF5-1:1	8.49	3.84	3.77	4.14	1.61e-005	0.000956	intergenic	11
lnc-CCDC14-2:2	68.74	7.14	5.92	10.93	5.71e-005	0.00201	sense	11
ENST00000617921	173.45	13.12	9.11	22.58	1.7e-006	0.00027	intergenic	11
lnc-RAG1-5:1	427.82	137.02	132.72	154.63	0.000499	0.00724	sense	10
lnc-TMTC3-5:4	84.52	4.05	4.46	8.36	< 1e-07	< 1e-07	intergenic	10
lnc-SLC10A6-5:1	32.91	12.69	12.13	13.98	0.0044594	0.0281	sense	10
lnc-GCNT1-6:1	24.59	4.32	4.44	5.47	< 1e-07	< 1e-07	sense	10
NR_126337	101.45	35.03	33.64	38.33	0.0048936	0.0299	intergenic	10

Table 4
Key transcription factors (TFs) in regulatory networks

TFs	Degree	TFs	Degree
TF-DEGs network		TF-angiogenesis -DEGs network	
LUN-1	218	LUN-1	73
Tel-2	22	Tel-2	18
1-Oct	22	MEF-2	13
RSRFC4	20	POU3F2	9
POU3F2	20	1-Oct	9
MEF-2	20	RSRFC4	7
STE11	19	PPARalpha	5
Evi-1	18	MADS-A	4
STAT5B	16	Brachyury	4
IRF-1	12	STAT5A	3

Construction of TF regulatory networks

To investigate the mechanism of gene regulation at the transcriptional level, we constructed two interaction networks (TF-DEG network and TF-angiogenesis-DEG network) between TFs and DEGs that originated from the co-expression network or A-co-expression network. The TF-DEG network contained 99 regulation models and 402 nodes involving 85 TFs (Fig. 4B), among which three TFs, namely LUN-1 (TF, Degree = 218), Tel-2 (TF, Degree = 22), and 1-Oct (TF, Degree = 20), regulated over 20 nodes (Table 3). LUN-1 had the highest degree and might play more important roles in the TF regulatory networks. The lncRNAs regulated by the most TFs were NR_103548, lnc-ARRDC3-1:16, ENST00000513626, and ENST00000516496 (regulated by six TFs). In the TF-angiogenesis-DEG network, due to the important role of LUN-1 (TF, Degree = 73) and Tel-2 (TF, Degree = 18) in the progression of CC, a subnetwork was constructed showing their effects on angiogenesis functions (Fig. 3, Table 3).

qRT-PCR verification for the candidate genes

According to the probe signal value, type, DEG fold change, degree value, and protein coding ability, five LncRNAs were selected from the networks above for validation in the expanded clinical samples (70 NC, 58 LIGN, 27 HIGN, 62 pMMR CC, eight dMMRCC samples). LncRNAs ENST00000545920, ENST00000521815, ENST00000609220, and ENST00000603052 presented a sequentially increasing trend in expression with tumor progression (Fig. 6A-D), whereas NR_026543 showed a sequentially decreasing trend (Fig. 6E). No statistically significant differences were observed in the expression of NR_026543 and ENST00000521815 among LGIN, HGIN, and pMMR CC. However, the expression levels for each group were significantly elevated or decreased compared with that of the NC samples, suggesting that these genes could be used to distinguish normal from lesion tissue. Furthermore, the expression of ENST00000603052, ENST00000521815, and ENST00000609220 was significantly different in pMMR CC and dMMR CC tissue samples (Fig. 6. F-J), indicating these three genes could distinguish pMMR CC from dMMR CC.

Diagnostic efficacy of selected lncRNAs

The diagnostic efficacy of the above selected five lncRNAs was evaluated as potential biomarkers for pMMR CC diagnosis compared to carcinoembryonic antigen (CEA), a traditional serum tumor biomarker. The specificities of ENST00000521815,

ENST00000603052, ENST00000609220, NR_026543, ENST00000545920, and CEA were 0.981, 1, 0.962, 0.932, 0.981, and 0.904, and the sensitivities were 0.931, 0.864, 0.746, 0.846, 0.576, and 0.593, respectively. ROC analysis showed significantly higher AUC for lncRNAs ENST00000521815, ENST00000603052, ENST00000609220, and NR_026543 compared with CEA (0.785) (P = 0.0001, 0.0002, 0.0028, and 0.0011, respectively) (Fig. 7. A-D). No statistically significant difference in the AUC was observed between ENST00000545920 (0.794) and CEA (P = 0.8734) (Fig. 7. E).

Discussion

Among the human transcriptome, more than 90% are non-coding RNAs (ncRNAs), including microRNAs (miRNAs), lncRNAs, and circular RNAs (circRNAs) [21]. As a relatively new field, accumulated studies have uncovered the crucial roles of lncRNAs in oncogenicity [22, 23]. With the improvement of microarray and next-generation sequencing technology, large-scale aberrantly expressed lncRNAs have been discovered in multiple cancers, including breast cancer [24], gastric cancer [25], CRC [26], lung cancer [27], and pancreatic cancer [28].

A large proportion of colon adenomas are able to transform into adenocarcinoma through a dynamic process with the accumulation of multiple mutations and preternatural activation of signaling pathways [13]. Altered lncRNA expression patterns for microsatellite-stable colorectal cancer indicated that several lncRNAs were continuously upregulated/downregulated during the adenoma-adenocarcinoma sequence [29]. Unlike previous studies that focused on one or more lncRNAs in CRC [30–32], we first conducted a systematic and comprehensive analysis using microarrays to reveal the differentially expressed profiles of lncRNAs and mRNAs in the malignant evolutionary process for pMMR CC. Thousands of dysregulated transcripts were identified in the ACS, including 5783 lncRNAs and 4483 mRNAs. Subsequently, STC analysis revealed that adenoma is an intermediate step from normal tissue to CC and certain lncRNAs participate in this continuous process. As the dysregulation can be detected already in adenomas, these lncRNAs can be used as biomarkers for the screening and early detection of CC.

GO annotation and KEGG pathway analysis were performed to determine the functions underlying the differentially expressed lncRNAs in ACS. GO annotation suggested that aberrant transcripts were predominantly enriched in multicellular organismal development, cell division, cell adhesion, cell cycle, cell differentiation, mitosis, proteolysis, and angiogenesis. KEGG enrichment analysis revealed that the dysregulated transcripts mainly participated in metabolic pathways, pathways in cancer, cell cycle, PI3K-Akt signaling pathway, and purine metabolism. Cell cycle involvement was identified in both analyses. The uncontrolled tumor cell cycle is an essential feature of cancer and is caused by aberrant cell cycle proteins including cyclin-dependent kinases, Aurora kinases, and Polo-like kinases. Exploiting key molecules in the cell cycle may provide a new perspective for cancer therapy [33].

To further explore the underlying functions of lncRNAs in ACS, co-expression of lncRNAs with their associated coding genes was performed based on the degree of correlation. Among the top 10 elevated and decreased lncRNAs according to the degree, NR_029374 with 12 degrees was upregulated, consistent with the results obtained by Sun et.al [34], who reported that NR_029374 was upregulated in CC and correlated with poor overall survival. NR_029374 promoted the migration, invasion, and metastasis of CC through the miR-30-5p/SOX9 axis and this finding was supported by previous results [35]. NR_029374 was upregulated in hepatocellular carcinoma (HCC), markedly promoting HCC proliferation, migration, and angiogenesis [35].

Angiogenesis plays a key role in tumorigenesis and development by delivering nutrients and evacuating metabolic wastes [36]. Recently, several lncRNAs and mRNAs were demonstrated to participate in tumor angiogenesis. Upregulation of the lncRNA FLANC promoted angiogenesis by upregulating and prolonging the half-life of phosphorylated STAT3/VEGFA in CRC [37]. In the present study, we discovered that NOTCH4 (mRNA, Degree = 35), a member of the Notch family of transmembrane receptors, had the highest degree. Wu et al. reported the association of overexpressed NOTCH4 with CRC survival [38]. The NOTCH4 expression level was especially higher in non-small cell lung cancer tissues than in the NC tissues, related to the tumor size and TNM stage [39]. Active NOTCH4 might inhibit endothelial sprouting in vitro and vivo [40]. Further study is required to determine the expression level of NOTCH4 in a large population of pMMR CC patients and the underlying interactions with lncRNAs in angiogenesis. Several lncRNAs have demonstrated interactions with TFs in tumorigenesis and progression. For instance,

lncRNA SNHG15 obstructed the ubiquitination and degradation of Slug, a fast-turnover TF critical for controlling cancer cell invasion and metastasis, thereby promoting CC progression [41]. The transcription of lncRNA LINC01503 was activated by TF TP63, resulting in shorter survival times for patients with esophageal squamous cell carcinoma [42]. In the current study, the lncRNA–TF network analysis revealed that LUN-1 had the highest degree both in the TF-DEG network and the TF-angiogenesis-DEG network.

qRT-PCR analysis showed the upregulation of ENST00000545920, ENST00000521815, ENST00000609220, and ENST00000603052 and the downregulation of NR_026543 with tumor progression. Furthermore, ENST00000603052, ENST00000521815, and ENST00000609220 can be used to identify pMMR CC and dMMR CC. Previous studies revealed the upregulation of ENST00000545920, also known as lnc-SNHG1, in CRC [41], non-small cell lung cancer [43], gastric cancer [44], and glioma [45] to promote cancer cell growth by interacting with EZH2 in the nucleus and miR-154-5p in the cytoplasm [46]. Similar results have been observed in other tumors, including ENST00000521815, also known as CASC19, which possesses an oncogenic function through targeting miR-140-5p/CEMIP in CRC progression [47]. The downregulation of NR_026543 was associated with liver metastasis and poor prognosis for CC by binding with the miR-203 promoter [48]. ENST00000609220 was found to be highly expressed in CRC tissue, promoting CRC growth and metastasis through the miR-206/YAP1 axis [49]. To date, there is little knowledge on the function of ENST00000603052 in cancer.

Ultimately, the diagnostic efficacy of the identified lncRNAs was evaluated against serum CEA. Most of the selected lncRNAs achieved a slightly higher AUC in distinguishing CC from normal tissue. Further study is needed to validate these lncRNAs as biomarkers for the early diagnosis of CC by measuring their expression in serum samples to assess the consistency in tissue samples. Monitoring expression changes in these genes may also provide a new strategy to assess disease progression.

There are certain limitations in the present study. First, the sample size was relatively small, thereby limiting result reliability. We increased the number of clinical samples to confirm the function of the lncRNAs. Second, we only constructed expression profiles for colon tissue. Additional studies are required to evaluate tissues from other tumors. Last, compared with high-throughput sequencing, microarray is incapable of identifying novel lncRNAs.

Conclusions

In conclusion, we verified a series of differentially expressed lncRNAs and mRNAs in a normal mucosa-adenoma-pMMR CC sequence in the current study. The potential roles of these RNAs were predicted through bioinformatics analyses. The findings may provide novel insights for the diagnosis and therapeutic strategy of pMMR CC. Further studies are required to provide robust validation evidence for the functions of lncRNAs in CC.

Abbreviations

pMMR CC: DNA mismatch repair proficient colon cancer

lncRNAs: Long non-coding RNAs

LGIN: Low-grade intraepithelial neoplasia

HGIN: High-grade intraepithelial neoplasia

NC: Normal control

GO: Gene Ontology

KEGG: Kyoto Encyclopedia of Genes and Genomes

qRT-PCR: Quantitative real-time polymerase chain reaction

ROC: Receiver-operating characteristic

AUC: Area under the ROC curve

CRC: Colorectal cancer

CC: Colon cancer

CIN: Chromosomal instability

MSI: Microsatellite instability

CIMP: CpG island methylator phenotype

dMMR: Defective mismatch repair

MSI-H: High-level MSI

ACS: Adenoma–carcinoma sequence

APC: Adenomatous polyposis coli

STC: Series test of cluster

DEGs: Differential genes expression

RVM: Random variance model

FDR: False discovery rate

TF: Transcription factor

PCC: Pearson's correlation coefficient

GAPDH: Glyceraldehyde 3-phosphate dehydrogenase

LSD: Least significance difference

SD: Standard deviation

CEA: Carcinoembryonic antigen

ncRNAs: Non-coding RNAs

miRNAs: microRNAs

circRNAs: circular RNAs

Declarations

Ethics approval and consent to participate

Approval was obtained from the ethics committee of Beijing Shijitan Hospital (No: 2018-59). Written informed consent was obtained from all individual participants included in the study.

Consent for publication

Not applicable

Availability of data and materials

All data generated or analyzed during this study are included in this published article and its supplementary information files.

Competing interests

The authors declare that they have no competing interests.

Funding

This study was founded by the Digestive Medical Coordinated Development Center of Beijing Hospitals Authority under Grant No. XXZ015, Medical and Health Public Foundation of Beijing under Grant No. YWJKJJHKYJJ-B17262-067, and the Science and Technology Development Project of China State Railway Group under Grant No. N2019Z004. The funding bodies had no role in the study design, data collection, data interpretation, or writing of this manuscript.

Authors' contributions

Study concept and design: JW, QL, WKL; recruitment of patient: QL, WKL, JG, acquisition of data: QL, WKL, YW, JG; statistical analysis and drafting of manuscript: WKL, QL; material preparation, acquisition of tissue samples: WKL, QL, JG, YQL, NL, KL, YDW, CG , WL, GJ, NW, CW, HL; critical review of manuscript: JW, KL, YDW. All authors read and approved the final manuscript.

Acknowledgements

Not applicable

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians* 2018, **68**(6):394-424.10.3322/caac.21492
2. Bailey CE, Hu CY, You YN, Bednarski BK, Rodriguez-Bigas MA, Skibber JM, et al. Increasing disparities in the age-related incidences of colon and rectal cancers in the United States, 1975-2010. *JAMA surgery* 2015, **150**(1):17-22.10.1001/jamasurg.2014.1756
3. Chen W, Zheng R, Baade PD, Zhang S, Zeng H, Bray F, et al. Cancer statistics in China, 2015. *CA: a cancer journal for clinicians* 2016, **66**(2):115-132.10.3322/caac.21338
4. Dienstmann R, Vermeulen L, Guinney J, Kopetz S, Tejpar S, Tabernero J. Consensus molecular subtypes and the evolution of precision medicine in colorectal cancer. *Nature reviews Cancer* 2017, **17**(2):79-92.10.1038/nrc.2016.126
5. Fearon ER. Molecular genetics of colorectal cancer. *Annual review of pathology* 2011, **6**:479-507.10.1146/annurev-pathol-011110-130235
6. Fearon ER, Vogelstein B. A genetic model for colorectal tumorigenesis. *Cell* 1990, **61**(5):759-767.10.1016/0092-8674(90)90186-i
7. Carethers JM, Jung BH. Genetics and Genetic Biomarkers in Sporadic Colorectal Cancer. *Gastroenterology* 2015, **149**(5):1177-1190.e1173.10.1053/j.gastro.2015.06.047

8. Grady WM, Carethers JM. Genomic and epigenetic instability in colorectal cancer pathogenesis. *Gastroenterology* 2008, **135**(4):1079-1099.10.1053/j.gastro.2008.07.076
9. Walther A, Houlston R, Tomlinson I. Association between chromosomal instability and prognosis in colorectal cancer: a meta-analysis. *Gut* 2008, **57**(7):941-950.10.1136/gut.2007.135004
10. Benson AB, 3rd, Venook AP, Cederquist L, Chan E, Chen YJ, Cooper HS, et al. Colon Cancer, Version 1.2017, NCCN Clinical Practice Guidelines in Oncology. *Journal of the National Comprehensive Cancer Network : JNCCN* 2017, **15**(3):370-398.10.6004/jnccn.2017.0036
11. Sargent DJ, Marsoni S, Monges G, Thibodeau SN, Labianca R, Hamilton SR, et al. Defective mismatch repair as a predictive marker for lack of efficacy of fluorouracil-based adjuvant therapy in colon cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2010, **28**(20):3219-3226.10.1200/jco.2009.27.1825
12. Leslie A, Carey FA, Pratt NR, Steele RJ. The colorectal adenoma-carcinoma sequence. *The British journal of surgery* 2002, **89**(7):845-860.10.1046/j.1365-2168.2002.02120.x
13. Wu Z, Liu Z, Ge W, Shou J, You L, Pan H, et al. Analysis of potential genes and pathways associated with the colorectal normal mucosa-adenoma-carcinoma sequence. *Cancer medicine* 2018, **7**(6):2555-2566.10.1002/cam4.1484
14. Brenner H, Kloor M, Pox CP. Colorectal cancer. *Lancet (London, England)* 2014, **383**(9927):1490-1502.10.1016/s0140-6736(13)61649-9
15. Liu Z, Ren L, Tian J, Liu N, Hu Y, Zhang P. Comprehensive Analysis of Long Noncoding RNAs and Messenger RNAs Expression Profiles in Patients with Marjolin Ulcer. *Medical science monitor : international medical journal of experimental and clinical research* 2018, **24**:7828-7840.10.12659/msm.911177
16. Xu M, Xu X, Pan B, Chen X, Lin K, Zeng K, et al. LncRNA SATB2-AS1 inhibits tumor metastasis and affects the tumor immune cell microenvironment in colorectal cancer by regulating SATB2. *Molecular cancer* 2019, **18**(1):135.10.1186/s12943-019-1063-6
17. Wang CJ, Zhu CC, Xu J, Wang M, Zhao WY, Liu Q, et al. The lncRNA UCA1 promotes proliferation, migration, immune escape and inhibits apoptosis in gastric cancer by sponging anti-tumor miRNAs. *Molecular cancer* 2019, **18**(1):115.10.1186/s12943-019-1032-0
18. Li Q, Li N, Lao Y, Lin W, Jiang G, Wei N, et al. Variable Levels of Long Noncoding RNA Expression in DNA Mismatch Repair-Proficient Early-Stage Colon Cancer. *Digestive diseases and sciences* 2017, **62**(5):1235-1245.10.1007/s10620-017-4465-6
19. Yoav, Benjamini, Yosef, B HJJotRSSS. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. 1995
20. Staton CA, Chetwood AS, Cameron IC, Cross SS, Brown NJ, Reed MW. The angiogenic switch occurs at the adenoma stage of the adenoma carcinoma sequence in colorectal cancer. *Gut* 2007, **56**(10):1426-1432.10.1136/gut.2007.125286
21. Yu X, Zheng H, Tse G, Chan MT, Wu WK. Long non-coding RNAs in melanoma. *Cell proliferation* 2018, **51**(4):e12457.10.1111/cpr.12457
22. Fang Y, Fullwood MJ. Roles, Functions, and Mechanisms of Long Non-coding RNAs in Cancer. *Genomics, proteomics & bioinformatics* 2016, **14**(1):42-54.10.1016/j.gpb.2015.09.006
23. Sanchez Calle A, Kawamura Y, Yamamoto Y, Takeshita F, Ochiya T. Emerging roles of long non-coding RNA in cancer. *Cancer science* 2018, **109**(7):2093-2100.10.1111/cas.13642
24. Jiang YZ, Liu YR, Xu XE, Jin X, Hu X, Yu KD, et al. Transcriptome Analysis of Triple-Negative Breast Cancer Reveals an Integrated mRNA-lncRNA Signature with Predictive and Prognostic Value. *Cancer research* 2016, **76**(8):2105-2114.10.1158/0008-5472.Can-15-3284
25. Zhang K, Shi H, Xi H, Wu X, Cui J, Gao Y, et al. Genome-Wide lncRNA Microarray Profiling Identifies Novel Circulating lncRNAs for Detection of Gastric Cancer. *Theranostics* 2017, **7**(1):213-227.10.7150/thno.16044
26. Yamada A, Yu P, Lin W, Okugawa Y, Boland CR, Goel A. A RNA-Sequencing approach for the identification of novel long non-coding RNA biomarkers in colorectal cancer. *Scientific reports* 2018, **8**(1):575.10.1038/s41598-017-18407-6

27. Peng Z, Wang J, Shan B, Yuan F, Li B, Dong Y, et al. Genome-wide analyses of long noncoding RNA expression profiles in lung adenocarcinoma. *Scientific reports* 2017, **7**(1):15331.10.1038/s41598-017-15712-y
28. Müller S, Raulefs S, Bruns P, Afonso-Grunz F, Plötner A, Thermann R, et al. Next-generation sequencing reveals novel differentially regulated mRNAs, lncRNAs, miRNAs, sdRNAs and a piRNA in pancreatic cancer. *Molecular cancer* 2015, **14**:94.10.1186/s12943-015-0358-5
29. Bartley AN, Yao H, Barkoh BA, Ivan C, Mishra BM, Rashid A, et al. Complex patterns of altered MicroRNA expression during the adenoma-adenocarcinoma sequence for microsatellite-stable colorectal cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research* 2011, **17**(23):7283-7293.10.1158/1078-0432.Ccr-11-1452
30. Lan Z, Yao X, Sun K, Li A, Liu S, Wang X. The Interaction Between lncRNA SNHG6 and hnRNPA1 Contributes to the Growth of Colorectal Cancer by Enhancing Aerobic Glycolysis Through the Regulation of Alternative Splicing of PKM. *Frontiers in oncology* 2020, **10**:363.10.3389/fonc.2020.00363
31. Wu H, Wei M, Jiang X, Tan J, Xu W, Fan X, et al. lncRNA PVT1 Promotes Tumorigenesis of Colorectal Cancer by Stabilizing miR-16-5p and Interacting with the VEGFA/VEGFR1/AKT Axis. *Molecular therapy Nucleic acids* 2020, **20**:438-450.10.1016/j.omtn.2020.03.006
32. Yao X, Lan Z, Lai Q, Li A, Liu S, Wang X. LncRNA SNHG6 plays an oncogenic role in colorectal cancer and can be used as a prognostic biomarker for solid tumors. *Journal of cellular physiology* 2020.10.1002/jcp.29672
33. Otto T, Sicinski P. Cell cycle proteins as promising targets in cancer therapy. *Nature reviews Cancer* 2017, **17**(2):93-115.10.1038/nrc.2016.138
34. Sun T, Liu Z, Zhang R, Ma S, Lin T, Li Y, et al. Long Non-Coding RNA LEF1-AS1 Promotes Migration, Invasion and Metastasis of Colon Cancer Cells Through miR-30-5p/SOX9 Axis. *OncoTargets and therapy* 2020, **13**:2957-2972.10.2147/ott.S232839
35. Dong H, Jian P, Yu M, Wang L. Silencing of long noncoding RNA LEF1-AS1 prevents the progression of hepatocellular carcinoma via the crosstalk with microRNA-136-5p/WNK1. *Journal of cellular physiology* 2020.10.1002/jcp.29503
36. Folkman J. Role of angiogenesis in tumor growth and metastasis. *Seminars in oncology* 2002, **29**(6 Suppl 16):15-18.10.1053/sonc.2002.37263
37. Pichler M, Rodriguez-Aguayo C, Nam SY, Dragomir MP, Bayraktar R, Anfossi S, et al. Therapeutic potential of FLANC, a novel primate-specific long non-coding RNA in colorectal cancer. *Gut* 2020.10.1136/gutjnl-2019-318903
38. Wu G, Chen Z, Li J, Ye F, Chen G, Fan Q, et al. NOTCH4 Is a Novel Prognostic Marker that Correlates with Colorectal Cancer Progression and Prognosis. *Journal of Cancer* 2018, **9**(13):2374-2379.10.7150/jca.26359
39. Wang Y, Yang R, Wang X, Ci H, Zhou L, Zhu B, et al. Evaluation of the correlation of vasculogenic mimicry, Notch4, DLL4, and KAI1/CD82 in the prediction of metastasis and prognosis in non-small cell lung cancer. *Medicine* 2018, **97**(52):e13817.10.1097/md.0000000000013817
40. Leong KG, Hu X, Li L, Noseda M, Larrivée B, Hull C, et al. Activated Notch4 inhibits angiogenesis: role of beta 1-integrin activation. *Molecular and cellular biology* 2002, **22**(8):2830-2841.10.1128/mcb.22.8.2830-2841.2002
41. Jiang H, Li T, Qu Y, Wang X, Li B, Song J, et al. Long non-coding RNA SNHG15 interacts with and stabilizes transcription factor Slug and promotes colon cancer progression. *Cancer letters* 2018, **425**:78-87.10.1016/j.canlet.2018.03.038
42. Xie JJ, Jiang YY, Jiang Y, Li CQ, Lim MC, An O, et al. Super-Enhancer-Driven Long Non-Coding RNA LINC01503, Regulated by TP63, Is Over-Expressed and Oncogenic in Squamous Cell Carcinoma. *Gastroenterology* 2018, **154**(8):2137-2151.e2131.10.1053/j.gastro.2018.02.018
43. Lu Q, Shan S, Li Y, Zhu D, Jin W, Ren T. Long noncoding RNA SNHG1 promotes non-small cell lung cancer progression by up-regulating MTDH via sponging miR-145-5p. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* 2018, **32**(7):3957-3967.10.1096/fj.201701237RR
44. Guo W, Huang J, Lei P, Guo L, Li X. LncRNA SNHG1 promoted HGC-27 cell growth and migration via the miR-140/ADAM10 axis. *International journal of biological macromolecules* 2019, **122**:817-823.10.1016/j.ijbiomac.2018.10.214

45. Li H, Xue Y, Ma J, Shao L, Wang D, Zheng J, et al. SNHG1 promotes malignant biological behaviors of glioma cells via microRNA-154-5p/miR-376b-3p- FOXP2- KDM5B participating positive feedback loop. *Journal of experimental & clinical cancer research : CR* 2019, **38**(1):59.10.1186/s13046-019-1063-9
46. Xu M, Chen X, Lin K, Zeng K, Liu X, Pan B, et al. The long noncoding RNA SNHG1 regulates colorectal cancer cell growth through interactions with EZH2 and miR-154-5p. *Molecular cancer* 2018, **17**(1):141.10.1186/s12943-018-0894-x
47. Wang XD, Lu J, Lin YS, Gao C, Qi F. Functional role of long non-coding RNA CASC19/miR-140-5p/CEMIP axis in colorectal cancer progression in vitro. *World journal of gastroenterology* 2019, **25**(14):1697-1714.10.3748/wjg.v25.i14.1697
48. Wang L, Wei Z, Wu K, Dai W, Zhang C, Peng J, et al. Long noncoding RNA B3GALT5-AS1 suppresses colon cancer liver metastasis via repressing microRNA-203. *Aging* 2018, **10**(12):3662-3682.10.18632/aging.101628
49. Dong X, Yang Z, Yang H, Li D, Qiu X. Long Non-coding RNA MIR4435-2HG Promotes Colorectal Cancer Proliferation and Metastasis Through miR-206/YAP1 Axis. *Frontiers in oncology* 2020, **10**:160.10.3389/fonc.2020.00160

Figures

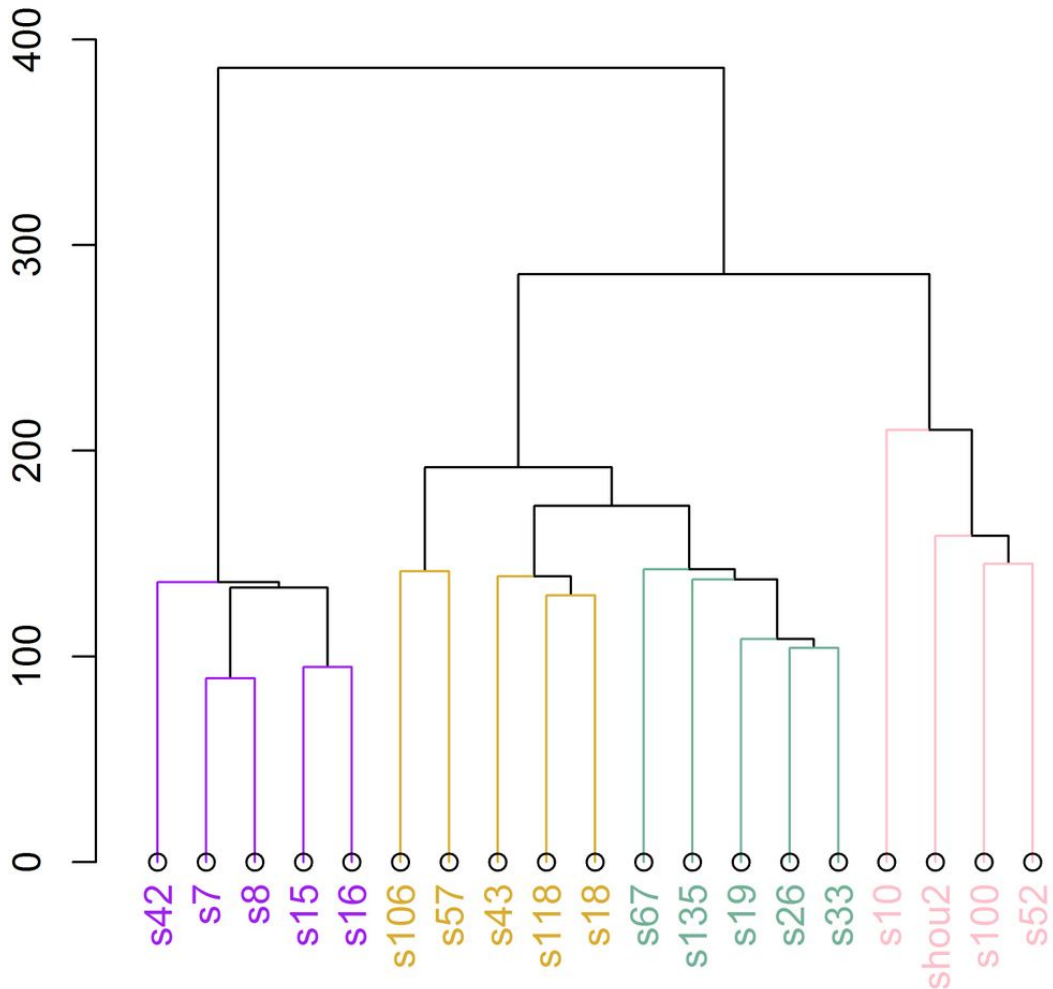

Figure 1

Cluster of samples. Purple represents NC group, yellow represents LGIN group, green represents HGIN group, and pink represents CC group.

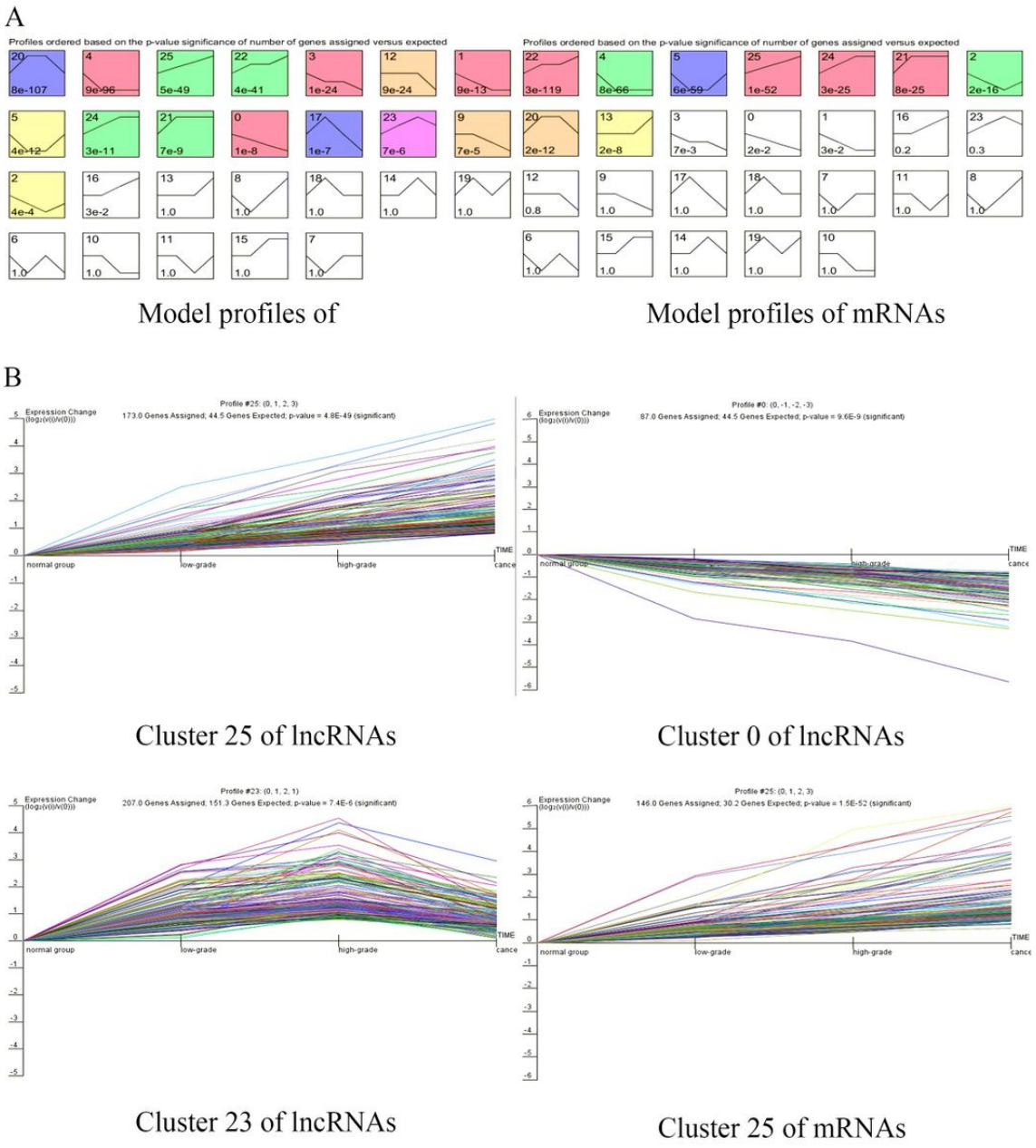

Figure 2

STC analysis of the dysregulated DEGs in multistage colonic mucosa tissues. (A) Twenty-six clusters for the expression pattern of DEGs, 15 lncRNA expression patterns, and nine mRNA expression patterns showed significant P values (colored boxes, P < 0.05/26). Each box represents a model expression profile, the lower number in the profile box is the P value and the upper number is the model profile number. (B) Cluster 25 (173 lncRNAs) gradually increased, cluster 0 (87 lncRNAs) showed a decreasing trend, cluster 23 contained 207 lncRNAs that gradually increased from NC to adenoma, and then decreased in CC, and cluster 25 (146 mRNAs) gradually increased. Horizontal axis represents different stages of colonic mucosal protruding lesions and the vertical axis shows the time series for the lncRNA/mRNA expression levels. r normalized with Log transformation.

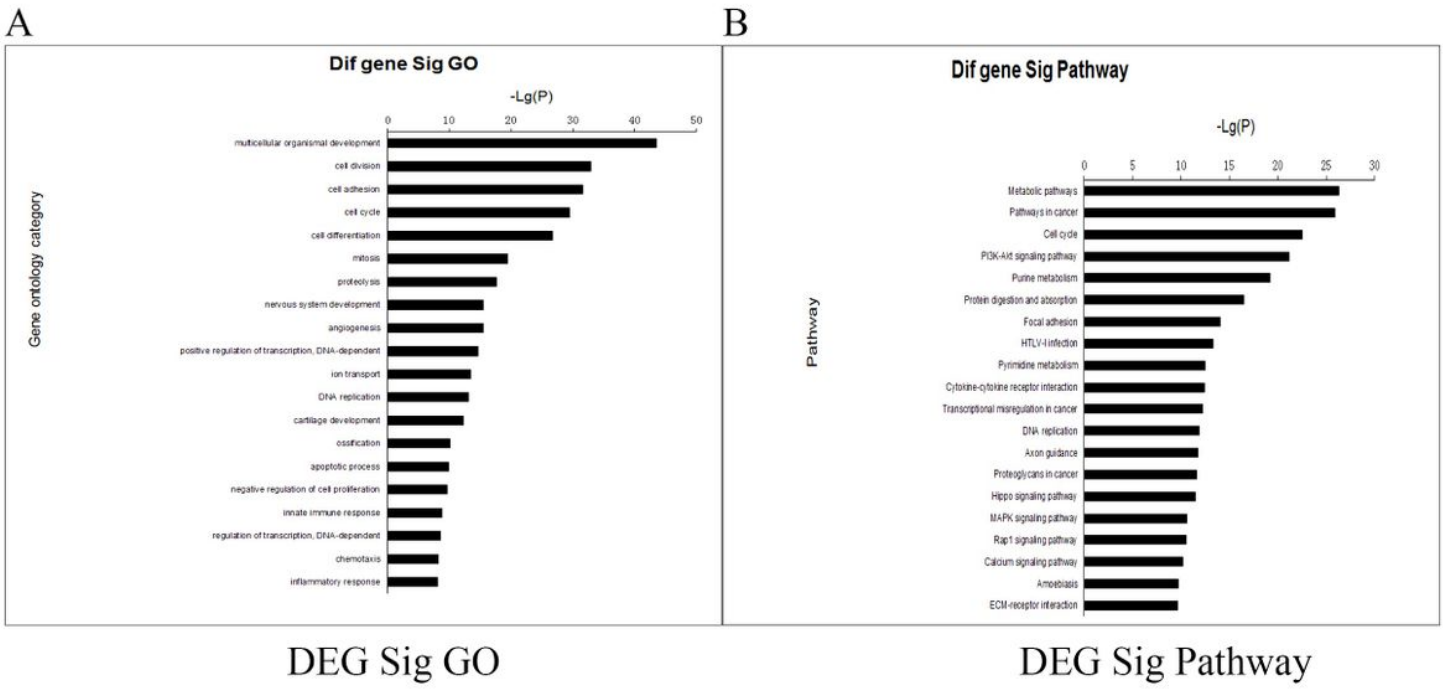

Figure 3

Overview of Go annotation and KEGG enrichment pathways. (A) Top 20 GO terms for DEGs. (B) Top 20 pathways corresponding to DEGs. P value < 0.05 and FDR < 0.05.

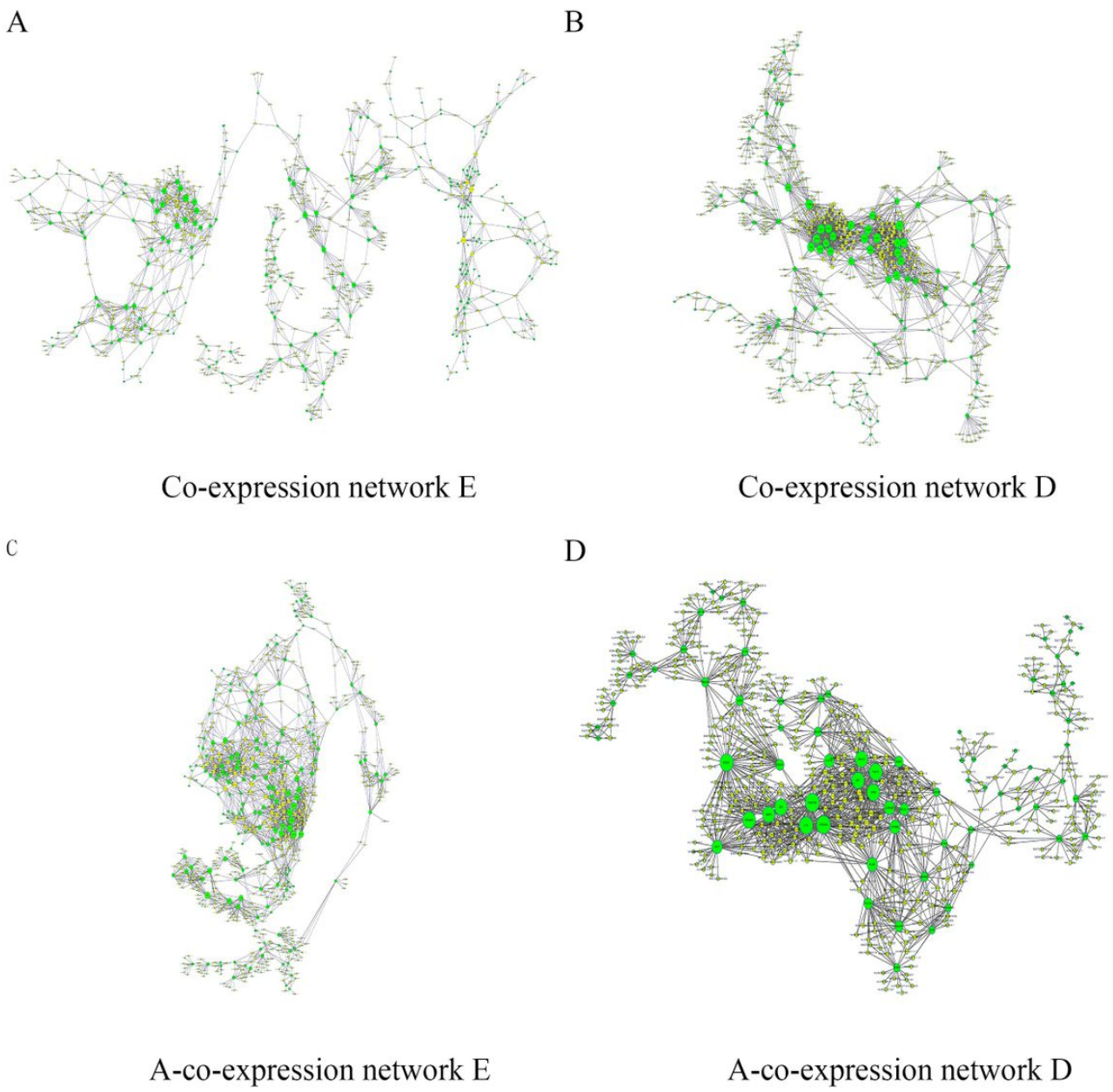

Figure 4

lncRNA-mRNA co-expression networks. (A-B) lncRNAs with elevated or decreased profiles and mRNAs for significant terms from GO and KEGG enrichment analysis were respectively selected to construct co-expression networks E and D in colonic ACS. (C-D) lncRNAs with significantly elevated/decreased model profiles and angiogenesis-related GO and KEGG items were respectively selected to construct A-co-expression networks E and D in colonic ACS. Nodes represent DEGs (green for mRNAs and yellow for lncRNAs). Lines represent interactions and node size represents degree value.

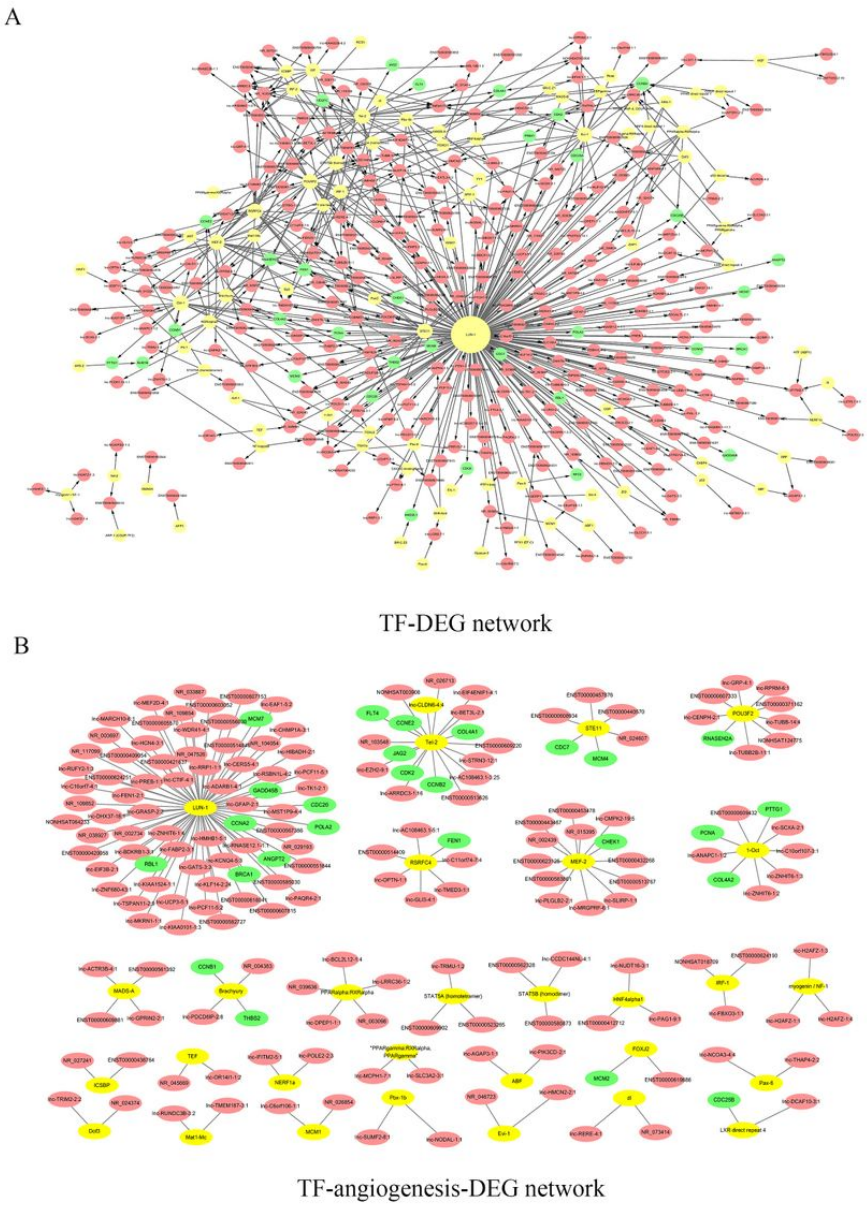

Figure 5

Transcription factor (TF) regulatory network. (A) Interaction networks between TFs and DEGs that originated from significant model profiles. (B) Interaction networks between TFs and DEGs that originated from angiogenesis-related model profiles. The yellow-colored nodes represent TFs, the red-colored nodes represent lncRNAs, and the green-colored nodes represent mRNAs.

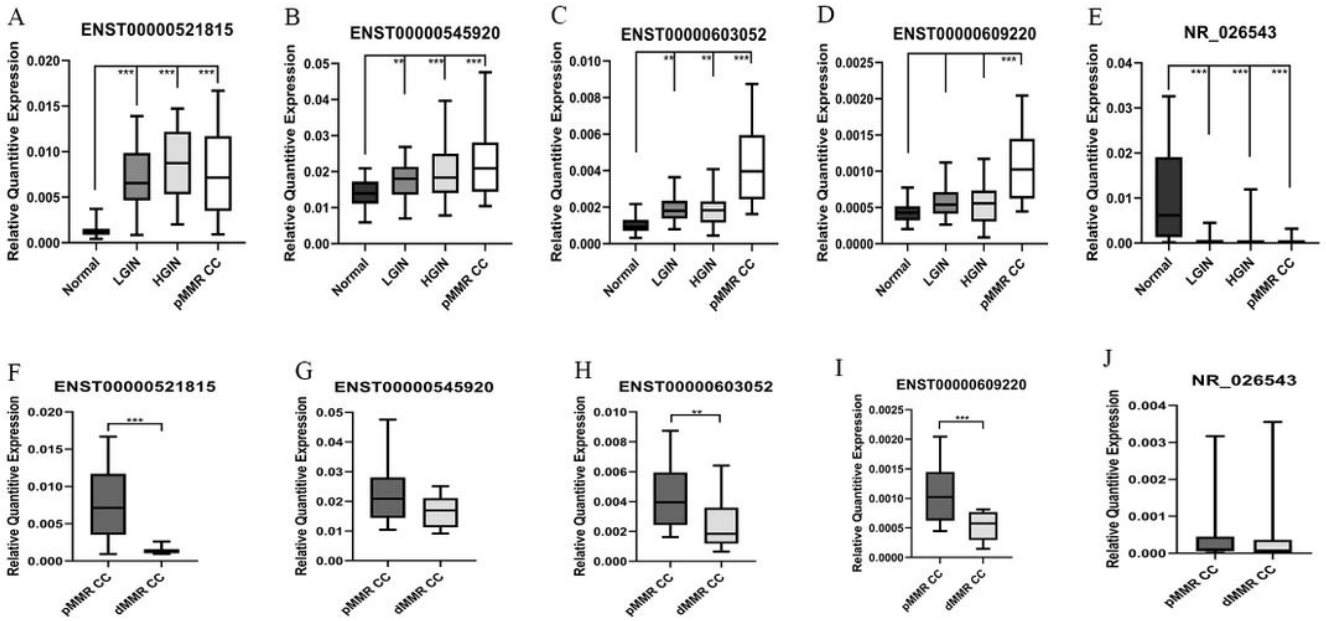

Figure 6

qRT-PCR verification of the candidate genes. ENST00000545920 (A), ENST00000521815 (B), ENST00000609220 (C), and ENST00000603052 (D) presented a sequentially ascending trend in expression with tumor progression, whereas NR_026543 (E) showed a sequentially decreasing trend. The expression of ENST00000521815 (F), ENST00000603052 (H), and ENST00000609220 (I) showed statistically significant differences between pMMR CC and dMMR CC.

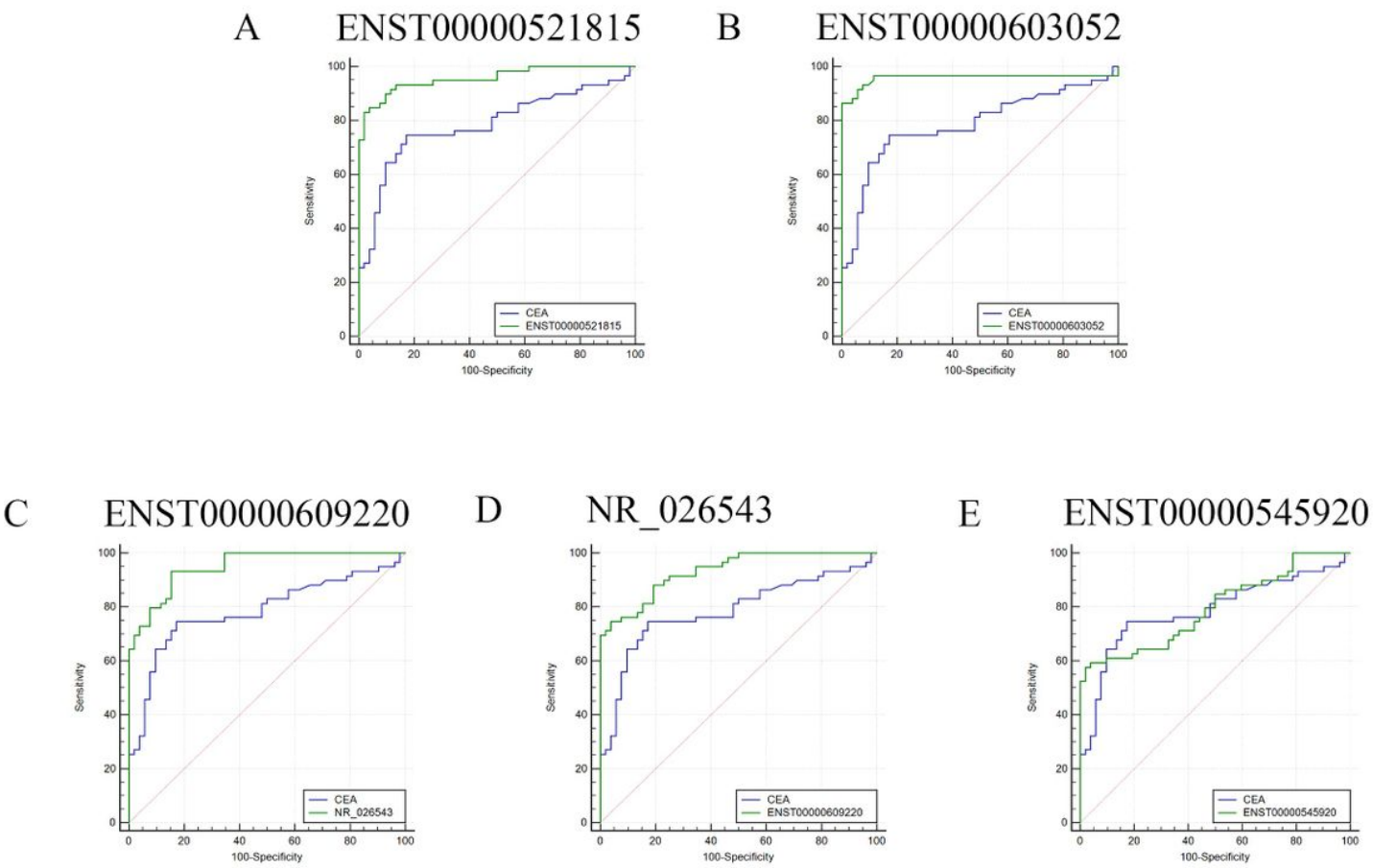

Figure 7

Diagnostic efficacy of selected lncRNAs. Compared with CEA, ENST00000521815 (A), ENST00000603052 (B), ENST00000609220 (C), and NR_026543 (D) had significantly larger AUCs (AUCs 0.957, 0.958, 0.933, 0.949 vs. 0.785, P=0.0001, 0.0002, 0.0028, and 0.0011, respectively). However, there were no statistically significant differences between the AUC of ENST00000545920 (0.794) and that of CEA (P=0.8734) (E). The ROC curve for CEA is shown in blue, while the ROC curves for the validation of the lncRNAs are shown in green.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Additionalfile2.docx
- Additionalfile1.docx