PORTLAND PRESS

Research Article

Comprehensive analysis of long noncoding RNA and mRNA in five colorectal cancer tissues and five normal tissues

Zhen-Xu Zhou[1,2,*], Xiao-Ming Chen[3,4,*], Yu-Qi Zhang[3,4], Liu Peng[3,4], Xiang-Yang Xue[5] and ⬤ Guo-Xin Li[1]

[1]Department of General Surgery, Nanfang Hospital, Southern Medical University, Guangzhou 510515, Guangdong Province, China; [2]Department of Hernia and Abdominal Wall Surgery, The First Affiliated Hospital of Wenzhou Medical University, Wenzhou 325000, Zhejiang Province, China; [3]Institute of Glycobiological Engineering/School of Laboratory Medicine and Life Sciences, Wenzhou Medical University, Wenzhou 325035, Zhejiang Province, China; [4]Key Laboratory of Laboratory Medicine, Ministry of Education of China, School of Laboratory Medicine and Life Sciences, Wenzhou Medical University, Wenzhou 325035, Zhejiang Province, China; [5]Department of Microbiology and Immunology, Institute of Molecular Virology and Immunology, Institute of Tropical Medicine, Wenzhou Medical University, Wenzhou 325035, Zhejiang Province, China

Correspondence: Xiang-Yang Xue (wzxxy001@163.com) or Guo-Xin Li (hengp97@126.com)

OPEN ACCESS

The present study investigated the role of abnormally expressed mRNA and long noncoding RNA (lncRNA) in the development of colorectal cancer (CRC). We used lncRNA sequencing to analyze the transcriptome (mRNA and lncRNA) of five pairs of CRC tissues and adjacent normal tissues. The total expression of mRNAs and lncRNAs in each sample was determined using the R package and the gene expression was calculated using normalized FPKM. The structural features and expression of all detected lncRNAs were compared with those of mRNAs. Differentially expressed mRNAs were selected to perform Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses. The functional analysis of differentially expressed lncRNAs was performed by analyzing the GO and KEGG enrichment of predicted *cis*-regulated target genes. A total of 18.2×10^8 reads were obtained by sequencing, in which the clean reads reached $\geq 94.67\%$, with a total of 245.2 G bases. The number of mRNAs and lncRNAs differentially expressed in CRC tissues and normal tissues were 113 and 6, respectively. Further predictive analysis of target genes of lncRNAs revealed that six lncRNA genes had potential *cis*-regulatory effects on 13 differentially expressed mRNA genes and co-expressed with 53 mRNAs. Up-regulated CTD-2256P15.4 and RP11-229P13.23 were the most important lncRNAs in these CRC tissues and involved in cell proliferation and pathway in cancer. In conclusion, our study provides evidence regarding the mRNA and lncRNA transcription in CRC tissues, as well as new insights into the lncRNAs and mRNAs involved in the development of CRC.

Introduction

A transcriptome is a collection of all transcripts in a cell, including messenger RNA (mRNA) and noncoding RNA (ncRNA), at a particular developmental stage or under certain physiological conditions. Transcriptome research is the basis for research on the function and structure of genes. Millions of reads can be obtained by high-throughput sequencing; sequence analysis can be performed by sequence alignment of these reads with reference genome based on existing gene annotation results. The differential expression of genes in different tissues, new transcript predictions, alternative splicing, single-nucleotide polymorphisms, and gene fusion can be analyzed by transcriptome sequencing [1–3]. Transcriptome sequencing technology is a highly flexible platform with a wide range of applications for high throughput, rapid, and high sensitivity analyses [4].

Long noncoding RNA (LncRNA) is a class of regulatory ncRNAs (length >200 nucleotides) found in multicellular animals in recent years [5]. LncRNA regulates chromosomal remodeling, genomic imprinting, histone modification, DNA methylation and RNA metabolism, as well as gene expression by *cis* and *trans* interactions with target molecules (e.g. DNA, RNA and proteins) [6]. The *cis* interactions affect the expression of neighboring genes based on gene location, while *trans* interactions refer to the ability of lncRNA to regulate the expression of encoding genes with the same expression pattern [7]. LncRNA forms an RNA regulatory network by regulating the expression of protein-coding genes, and participates in the physiological and pathological processes of the organism. Therefore, the change in the expression level of lncRNA is closely related to the occurrence and development of various diseases [8,9]. Annotation of eukaryotic transcriptomes by binding sequencing and biological information methods can reveal the functions of lncRNA in epigenetic modifications, transcriptional regulation and post-transcriptional regulation [10,11].

An increasing body of evidence shows that lncRNA is abnormally expressed in various types of cancer, regulating tumorigenesis, growth, apoptosis, invasion and metastasis, metabolism and other processes, which are closely related to the occurrence and development of various types of cancer [12,13]. The dysregulation of certain lncRNA target genes is also associated with the staging and prognosis of some tumors [14,15] characterized by tolerance to chemotherapy and resistance to targeted treatment [16,17].

The abnormal expression of lncRNA can be used as a potential biomarker for the diagnosis, prognosis and targeted therapy of cancer [18,19]. Transcriptome analysis can reveal abnormal expression or mutation of tens of thousands of lncRNAs related to tumorigenesis, metastasis, and tumor staging in various types of cancer [20].

Therefore, in the present study, lncRNA sequencing analysis was performed on five pairs of colorectal cancer (CRC) tissues and adjacent normal tissues. Functionalized enrichment analysis was performed on differentially expressed mRNAs and lncRNAs to determine whether their functional transcription was related to the development of CRC. Our data also contribute to a comprehensive understanding of the role of lncRNA in this process.

Materials and methods
Samples
Five CRC tissues and five adjacent normal tissues were collected from five patients with CRC who underwent tumorectomy at the Department of General Surgery, Nanfang Hospital, Southern Medical University (Guangzhou, Guangdong, China) between March 2015 and May 2016. These patients did not receive radiotherapy or chemotherapy prior to surgery and the tumor stage was confirmed using pathological methods according to the seventh edition of the Tumor-Node-Metastasis/American Joint Committee on Cancer classification. The patients included three males and two females, aged >70 years, with low or moderate tumor differentiation, and Tumor-Node-Metastasis stage II or IV (Table 1). Human cytomegalovirus testing and detection of related genes were performed as previously described [21]. Written informed consent was provided by all patients prior to the genetic research, and the research protocol was approved by the Ethics Review Committee of Nanfang Hospital, Southern Medical University.

Sequencing and data processing
Total RNA was extracted and the ribosomal RNA was completely removed. Following the breakage of RNA into fragments, the sequencing adaptor was ligated and polymerase chain reaction amplification was performed to construct a sequencing library. Subsequently, sequencing was performed using an Illumina X10 with sequencing mode PE150. During pre-processing of the raw data, adaptor and low-quality sequences were removed through Cutadapt [22]. Data quality control was performed using FastQC [23]. Valid data were mapped reads to the reference genome using Bowtie 2 [24], TopHat 2 [25], and the mapped reads of each sample were assembled by StringTie [26]. A comprehensive transcriptome was reconstructed and the expression levels of all transcripts were estimated by StringTie and Ballgown [27]. The gene position information of the sequence was simultaneously counting according to the genome annotation file.

Identification of mRNA and lncRNA
The comparison and visualization of the transcriptome were performed using the R package Ballgown ($P < 0.01$). The gene expression was calculated using normalized fragments per kilobase of exon model per million mapped reads (FPKM) values to avoid the effects of some false positives and extremely high-expression genes on low-abundance genes. The mRNA was analyzed at the gene and transcript levels. The known mRNA and transcripts <200 bp were removed, and the remaining transcripts were subjected to lncRNA prediction through coding potential assessment using Coding Potential Calculator (CPC) and Coding-Non-Coding Index (CNCI) software. These transcripts with

Table 1 Clinicopathological features of colorectal cancer patients in the present study

No.	Gender[a]	Age	Location[b]	Differentiation[c]	T stage[d]	Pathology[e]	Tumor size	Number of lymph nodes	Metastasis[f]	TNM stage	HCMV(+)	UL47	UL56	UL77
1	1	75	0	2	3	1	6	0	0	II	1	1	1	1
2	1	73	1	3	3	2	5	0	0	II	1	1	1	1
3	2	86	3	2	3	1	4	4	1	IV	1	1	1	1
4	2	67	3	2	3	2	7	0	0	II	1	1	1	1
5	1	73	1	2	3	1	4	0	0	II	1	1	1	1

a: 1 is male, 2 is female; b: 0 is the rectum, 1 is the sigmoid colon, 2 is the right half colon, and 3 is the left half colon; c: 1 is high, 2 is medium, 3 is low; d: T1 is mucosa and submucosa; T2 is intrinsic muscular layer; T3 is subserosal/full-layer/outer serosa; e: 1 is ulcer, 2 is bulges; f: 1 is transferred, 0 is not transferred; HCMV(+): human cytomegalovirus, and 1 is positive; UL47, UL56, UL77 are HCMV genes, and 1 is positive.
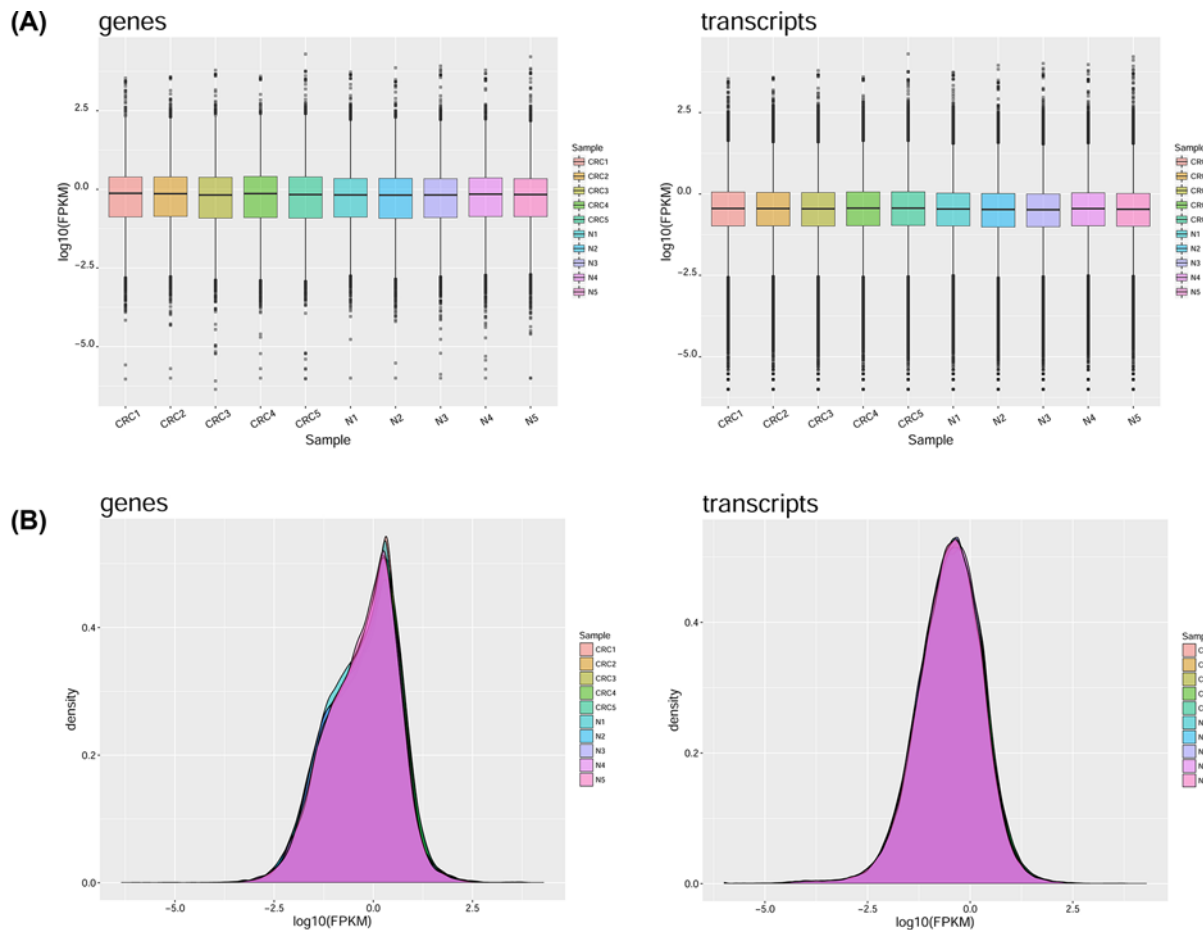
PORTLAND PRESS

Figure 1. Gene expression analysis results of RNA sequencing
(**A**) The distribution of gene expression was shown using a box plot: the abscissa is the sample name, the ordinate is log10 (FPKM), and the box plot of each region corresponds to five statistics (top to bottom are the maximum, upper quartile, middle value, lower quartile, and minimum). (**B**) Distribution of FPKM density. The expression density of each sample conforms to the normal distribution, and the expression trends of biological replicate samples are consistent. The abscissa is log10 (FPKM) and the ordinate is the density of the gene.

coding potential (i.e. CPC score < -1 and CNCI score <0) were classified as novel mRNAs and filtered. The lncRNA sequence was finally obtained and genome mapping was performed to display the distribution of lncRNA candidates in the chromosome using the Circos software. Genome mapping was displayed according to different classification of lncRNAs in different samples.

Comparison of mRNA and lncRNA

The structural features (i.e. length distribution, number of exons and length of open reading frame [ORF]) and expression of lncRNAs were compared and analyzed. The differences between lncRNAs and mRNAs were compared according to the ORF lengths of translation. The ORF analysis is mainly based on the six-frame translation principle of nucleic acids.

Functional analysis of differentially expressed mRNAs and lncRNAs

The expression levels of mRNAs and lncRNAs were determined by calculating the FPKM [28] using StringTie. The differentially expressed mRNAs and lncRNAs who met the threshold of significant difference ($P < 0.05$) were selected using the R package-Ballgown [27]. Gene Ontology (GO) is an internationally standardized gene function classification system that provides a dynamically updated vocabulary to fully describe the properties of genes and gene products in organisms. The basic unit of GO is term (term, node), and each term corresponds to an attribute. The GO function
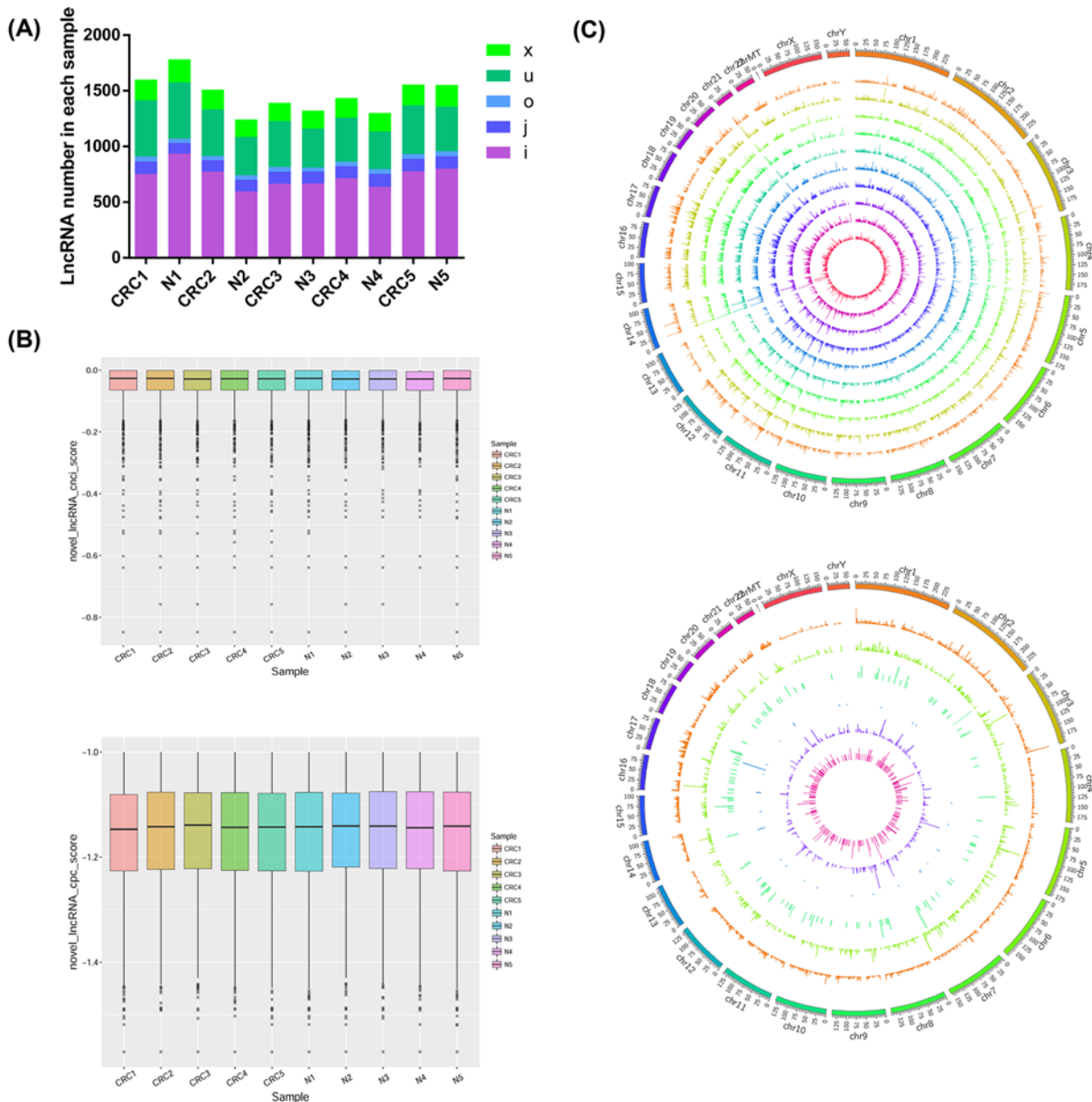
Figure 2. LncRNA screening analysis

(**A**) A variety of lncRNA expression analyses. x, exonic overlap with reference on the opposite strand; u, unknown, intergenic transcript; o, generic exonic overlap with a reference transcript; j, potentially novel isoform (fragment); at least one splice junction is shared with a reference transcript; i, a transfrag located entirely within a reference intron. (**B**) The score statistic box plot of lncRNA CPC and CNCL in each sample. (**C**) Macroscopic statistics of the expression levels of lncRNAs in different chromosomes.

significant enrichment analysis mapped all significant differentially expressed genes to the various terms of the GO database to calculate the number of genes per term. Subsequently, a hypergeometric test was applied to identify the GO terms enriched with differentially expressed genes versus the entire genome. *In vivo*, different genes coordinate with each other to perform their biological functions; pathway-based analysis assists in further understanding these biological functions of genes. Kyoto Encyclopedia of Genes and Genomes (KEGG) is the main public database for pathway analysis. The pathway significant enrichment analysis was also performed using a hypergeometric test.
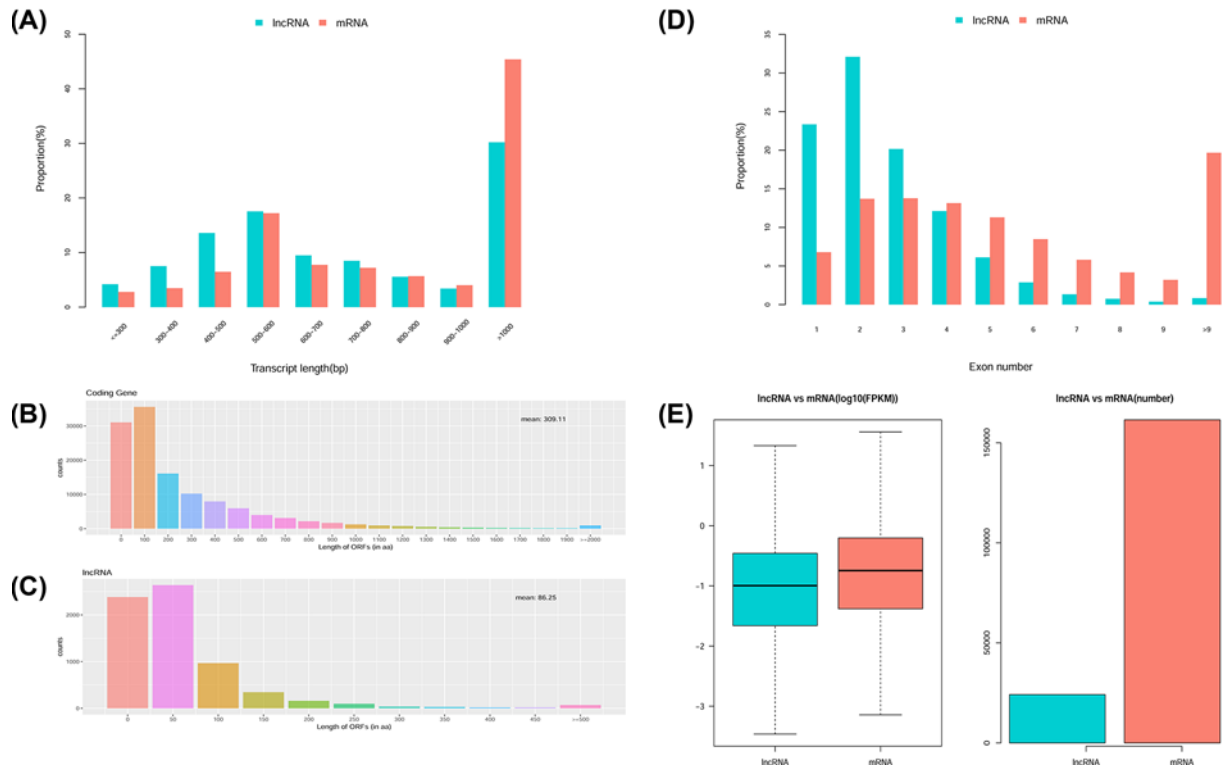
Figure 3. Structural comparison between lncRNAs and mRNAs
(**A**) mRNA versus lncRNA length distribution. (**B**) mRNA ORF length distribution. (**C**) LncRNA ORF length distribution. (**D**) mRNA versus lncRNA exon number. (**E**) mRNA versus lncRNA expression level and number.

Prediction and functional analysis of differentially expressed lncRNA target genes

The *cis*-regulated target genes of lncRNAs are mainly predicted through their positional relationship. The lncRNA *cis*-regulation is defined as differentially expressed mRNA upstream and downstream (100 kbp range) of the differentially expressed lncRNA in the chromosome. GO and KEGG enrichment analyses were performed on the targeted mRNAs for lncRNAs using the BLAST2GO [29]. Statistical significance was denoted by a *P*-value <0.05. Pearson correlation analyses of differentially expressed lncRNAs and mRNAs were performed using the Co-LncRNA online software (http://bio-bigdata.hrbmu.edu.cn/Co-LncRNA/). LncRNA-mRNA pairs with correlation coefficient ≥0.8 and *P* < 0.05 were selected for the KEGG enrichment analysis. The lncRNA–mRNA network was performed using the Cytoscape software (version 3.6.1).

Results
Reads and mapping results of lncRNA sequencing

A total of 18.2×10^8 reads were obtained by sequencing, in which the clean reads reached ≥94.67%, with a total of 245.2 G bases. The Q30 values in all samples reached >94.63% (Table 2). This sequencing amount and quality provided useful raw data for subsequent data assembly. The distribution of gene expression values, reflected in the FPKM box plot, indicated the repeatability of the design samples. The values of log10 (FPKM) were used to express the density of the expression values in different samples, which showed that the expression trends in different samples were consistent with normal distribution (Figure 1).

LncRNA classification and comparison

The lncRNAs predicted by CPC and CNCI were classified into five class codes: (1) potentially novel isoform (fragment), in which at least one splice junction is shared with a reference transcript; (2) a transfrag located entirely within a reference intron; (3) generic exonic overlap with a reference transcript; (4) unknown, intergenic transcript; and

Table 2 Sequence statistics and quality control

Sample	Raw data		Valid data		Valid ratio (reads)	Q20 (%)	Q30 (%)	GC content (%)
	Read	Base	Read	Base				
CRC1	1.80E+08	27.00G	1.70E+08	25.71G	95.2	99.65	94.95	48
N1	1.50E+08	22.16G	1.40E+08	21.34G	96.29	99.66	94.96	50
CRC2	1.70E+08	25.41G	1.60E+08	24.31G	95.68	99.77	95.57	52.5
N2	1.90E+08	28.16G	1.80E+08	26.87G	95.44	99.74	95.51	49
CRC3	1.80E+08	26.34G	1.70E+08	25.21G	95.72	99.77	95.67	50
N3	1.80E+08	27.03G	1.70E+08	25.79G	95.42	99.73	95.12	49
CRC4	1.70E+08	25.96G	1.60E+08	24.72G	95.22	99.67	94.9	49
N4	1.70E+08	25.08G	1.60E+08	23.75G	94.67	99.64	94.63	50
CRC5	1.50E+08	23.11G	1.50E+08	22.13G	95.75	99.74	95.32	51
N5	1.80E+08	26.93G	1.70E+08	25.37G	94.22	99.79	95.6	52

Table 3 Differentially expressed lncRNAs and differentially expressed target mRNAs

LncRNA					mRNA		
Gene_id	Gene_name	Known/novel	Cis/trans	Cis location	Gene_id	Gene_name	Description
MSTRG.20579	CTD-2537I9.12	Known	Cis	100K	MSTRG.20574	ZNF865	Zinc finger protein 865
		Known	Cis	10K	MSTRG.20578	U2AF2	U2 small nuclear RNA auxiliary factor 2
		Known	Cis	1K	MSTRG.20578	U2AF2	U2 small nuclear RNA auxiliary factor 2
		Known	Cis	100K	MSTRG.20577	ZNF580	Zinc finger protein 580
MSTRG.11978	RP11-731F5.2	Known	Cis	100K	MSTRG.11978	IGHG4	Immunoglobulin heavy constant gamma 4 (G4m marker)
		Known	Cis	10K	MSTRG.11978	IGHG2	Immunoglobulin heavy constant gamma 2 (G2m marker)
		Known	Cis	100K	MSTRG.11987	IGHA1	Immunoglobulin heavy constant alpha 1
		Known	Cis	100K	MSTRG.11984	IGHG1	Immunoglobulin heavy constant gamma 1 (G1m marker)
MSTRG.30320	CTD-2256P15.4	Known	Cis	100K	MSTRG.30319	CCT5	Chaperonin containing TCP1, subunit 5 (epsilon)
		Known	Cis	10K	MSTRG.30319	CCT5	Chaperonin containing TCP1, subunit 5 (epsilon)
		Known	Cis	100K	MSTRG.30322	CMBL	Carboxymethylenebutenolidase homolog (Pseudomonas)
MSTRG.39692	RP11-229P13.23	Known	Cis	100K	MSTRG.39695	NPDC1	Neural proliferation, differentiation and control, 1
		Known	Cis	100K	MSTRG.39703	DPP7	Dipeptidyl-peptidase 7
		Known	Cis	100K	MSTRG.39696	ENTPD2	Ectonucleoside triphosphate diphosphohydrolase 2
		Known	Cis	10K	MSTRG.39691	SAPCD2	Suppressor APC domain containing 2

(5) exonic overlap with reference on the opposite strand. The distribution of various lncRNAs was similar in each sample (Figure 2A). The statistical results of FPKM expression in each sample showed that the expression levels of lncRNA in each sample were also similar, reflecting the good sample repeatability (Figure 2B). The Circos software was used to map the lncRNA obtained by screening and visually display the distribution of lncRNA candidates in the chromosome. The expression levels of lncRNA were similar in different samples, while the distribution of lncRNAs of different class codes in the whole genome varied (Figure 2C).

Comparison of mRNAs and lncRNA

The length distribution of lncRNAs identified by sequencing was similar to that of mRNAs, mainly between 500–600 bp and >1000 bp (Figure 3A). The length of the ORF in the lncRNA sequence was mainly distributed between 50 and 250 amino acids, and the average ORF length was 86.25 amino acids. Notably, the length of the ORF in the mRNA
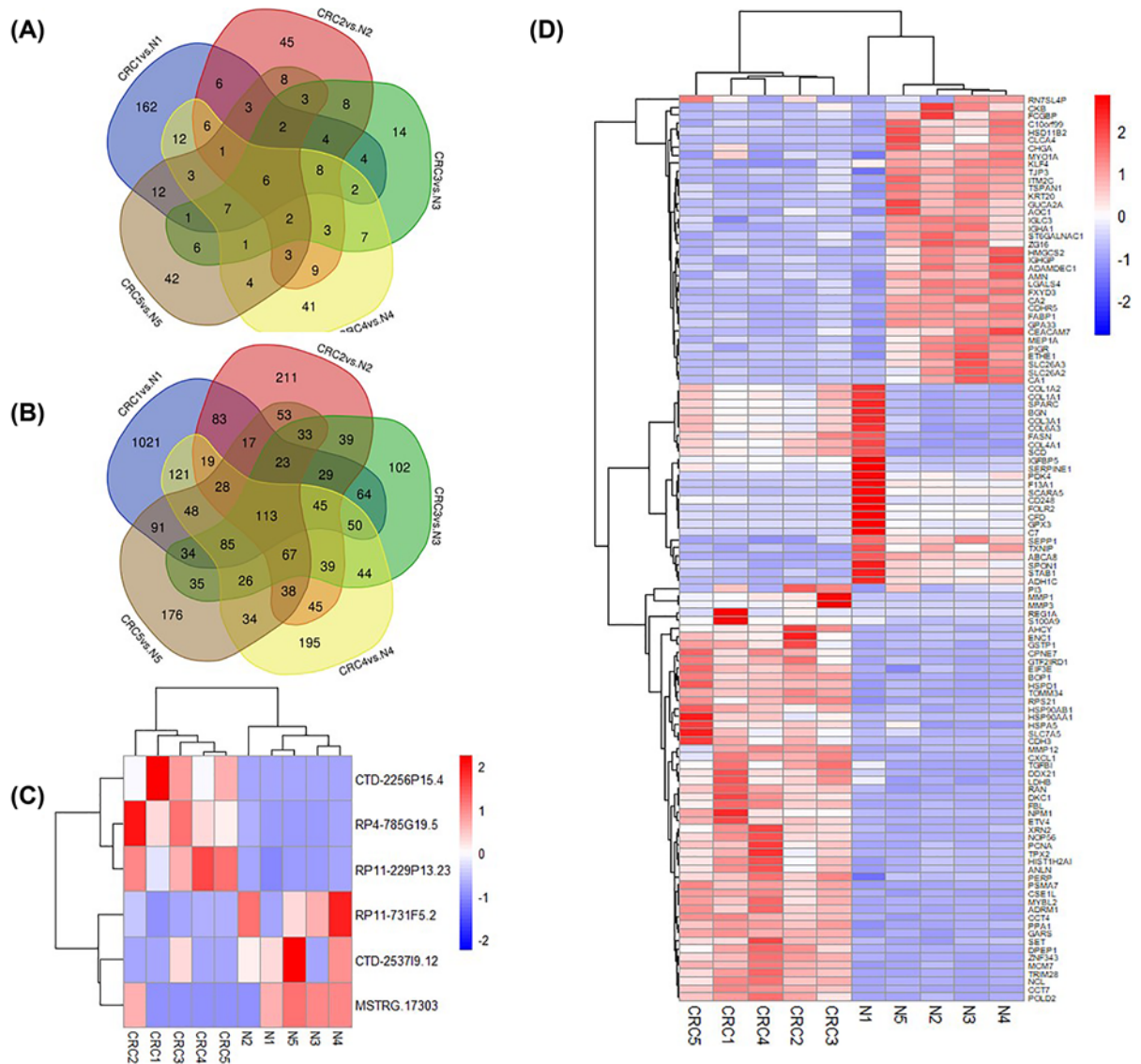
Figure 4. Differentially expressed lncRNAs and mRNAs in CRC tissues
Venn plot of differentially expressed lncRNAs (**A**) and mRNAs (**B**) in five CRC samples. Cluster analysis of these lncRNAs (**C**) and mRNAs (**D**).

sequence was markedly longer, with an average of 309.11 amino acids and mainly distributed between 100 and 1000 amino acids (Figure 3B,C). The number of exons in the lncRNA sequence was mainly between 1 and 6, while the distribution of mRNA exons demonstrated two peaks (i.e. 2–6 and >9) (Figure 3D). In terms of expression levels and numbers, those of mRNAs were higher than those of lncRNAs, especially in the latter (Figure 3E).

Differentially expressed lncRNAs and mRNAs

Difference analysis was performed on the assembled and quantified genes using the Ballgown package in R language. Finally, the number of differentially expressed lncRNAs detected in five pairs of samples was 239, 117, 78, 115 and 104, respectively. Of those, six lncRNAs (i.e. CTD-2256P15.4, RP4-785G19.5, RP11-229P13.23, RP11-731F5.2, CTD-2537I9.12 and MSTRG.17303) were abnormally expressed in all five CRC samples (Figure 4A) (Supplementary Table S1). The number of differentially expressed mRNAs detected in the five CRC samples was 1871, 882, 828, 997 and 901, respectively, and the number of intersections of the five data sets was 113 (Figure 4B) (Supplementary Table S2). Cluster analysis of differentially expressed lncRNAs and mRNAs was performed using the pheatmap package.
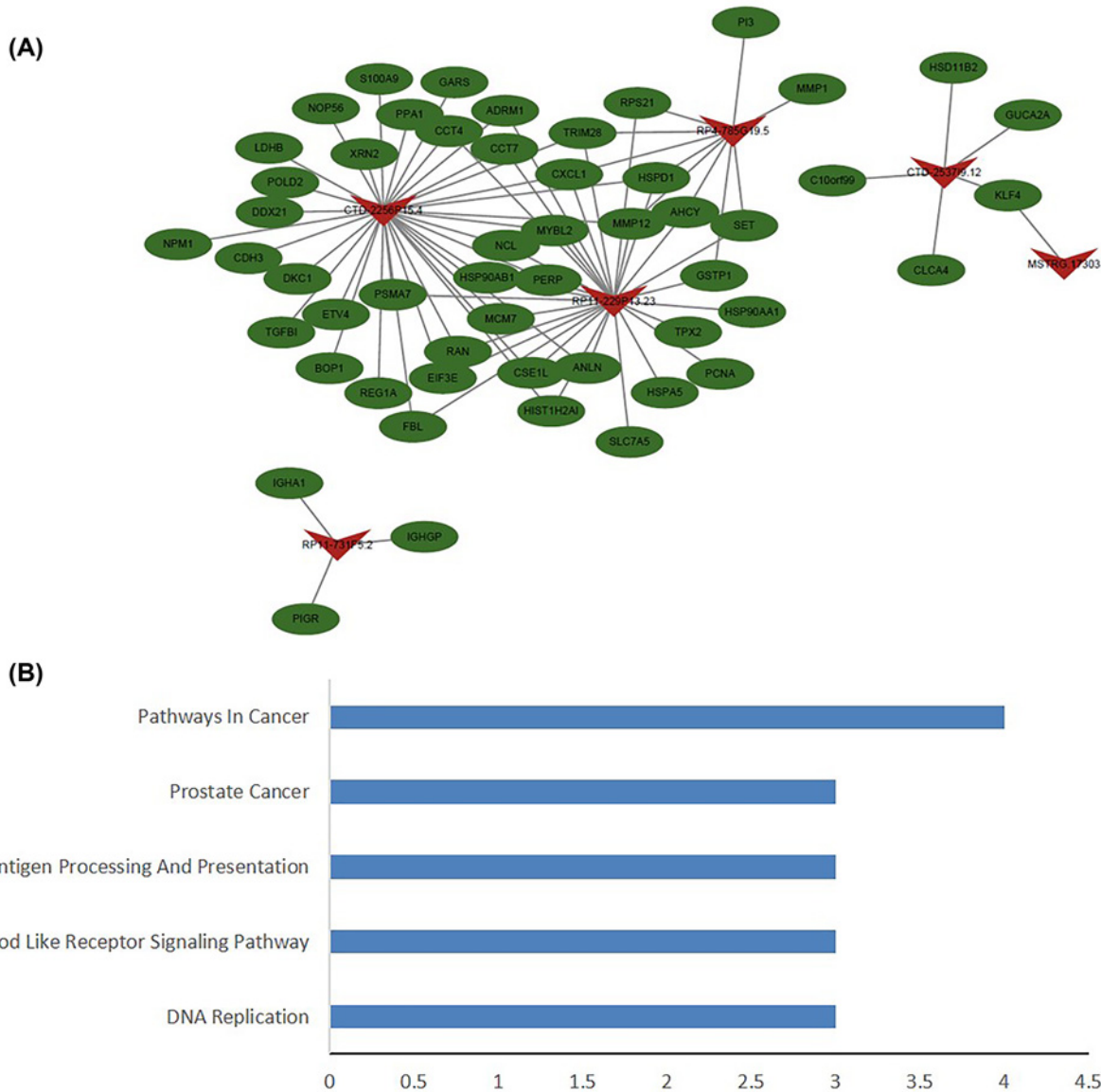
Figure 5. Functional analysis of differentially expressed lncRNAs in CRC tissues
(**A**) The lncRNA–mRNA network was performed using the Pearson correlation analysis (correlation coefficient ≥ 0.8 and $P < 0.05$) and the Cytoscape software. Green denotes mRNA and red denotes lncRNA. (**B**) The KEGG pathway enriched by mRNAs co-expressed with lncRNAs.

The results showed that lncRNAs/mRNAs exhibited a large difference in overall expression patterns between CRC and non-CRC samples, while the expression patterns in the same group were similar (Figure 4C,D).

Functional analysis of lncRNA target mRNA

A total of 15 differentially expressed mRNAs ($P < 0.05$) were screened by predicting the lncRNA *cis*-regulated target genes within 100 kbp upstream and downstream of the lncRNA (Table 3). The expression of these mRNAs was not consistent in the five CRC tissues. A total of 81 lncRNA–mRNA relationship pairs were obtained from the lncRNA–mRNA co-expression analysis. The lncRNAs with the highest correlation with mRNA expression were CTD-2256P.15.4 and RP11-229P13.23 (Figure 5A). The KEGG enrichment analysis showed that the main signaling pathways ($P < 0.05$) were DNA replication, nod-like receptor signaling pathway, antigen processing and presentation, prostate cancer and pathway in cancer (Figure 5B).

Discussion

LncRNA is structurally similar to mRNA, without the function of encoding protein [30,31]. It can participate in the regulation of gene expression at the genetic level, transcription level, post-transcription level etc., and is closely related to the occurrence and development of various diseases [11]. The present study compared five CRC tissues and adjacent normal tissues through lncRNA sequencing. Bioinformatics comparison of the identified lncRNAs with mRNAs involved transcript length, exon number, ORF length, transcript number and expression levels. It was found that the transcript length, exon number, ORF length, number of transcripts and expression levels of lncRNAs were smaller/lower than those of mRNAs. This is consistent with the characteristics of lncRNAs reported in other studies [32,33].

The GO enrichment analysis of mRNAs with abnormal expression in CRC samples showed that differentially expressed mRNAs were mainly located in vesicles, exosomes, and the intracellular lumen, involving in protein localization and cytokine interaction. The peroxisome proliferator-activated receptors signaling pathway plays an important role in the regulation of cell differentiation, development, metabolism and tumorigenesis in higher organisms [34,35]. The present study found four abnormal genes in this signaling pathway (i.e. low expression of fatty acid binding protein 1 [FABP1] and 3-hydroxy-3-methylglutaryl-CoA synthase 2 [HMGCS2]; high expression of matrix metallopeptidase 1 [MMP1] and stearoyl-CoA desaturase [SCD]). FABP1 was found to block cytotoxicity by binding to harmful molecules, such as heme, fatty acids [36]. HCMV infection can cause increased mRNA level of MMP1 [37], and MMP1 can decompose the extracellular matrix and participate in tissue remodeling and metastasis [38]. SCD is an important metabolic control factor; inhibition of its expression enhances the therapeutic effects on many metabolic diseases [39]. This suggests that CRC cells may contribute to tumorigenesis by decreasing the expression of FABP1 and HMGCS2 and increasing that of MMP1 and SCD. The present study also found that collagen type I alpha 1 chain (COL1A1), COL1A2, COL3A1 and COL6A3, which are involved in the extracellular matrix-receptor interaction signaling pathway, are highly expressed in CRC tissues. COL1A1 and COL1A2 encode the alpha1 and alpha2 chains of type I collagen, respectively. It has been reported that the translocation of the COL1A1 gene is related to dermatofibrosarcoma protuberance, which is caused by the unregulated expression of growth factors [40]. This is consistent with the phenotype of patients with CRC. In addition, eight genes (i.e. calreticulin [CALR], protein disulfide isomerase family A member 3 [PDIA3], heat shock protein family A (Hsp70) member 5 [HSPA5], HSPA8, heat shock protein 90 alpha family class A member 1 [HSP90AA1] and HSP90AB1), involved in the antigen processing and presentation signaling pathway, were highly expressed in CRC. This indicated that protein processing was highly active in CRC tissue.

The results of the present study are inconsistent with those of other previous studies including data from The Cancer Genome Atlas. LncRNA GAS5, which is lowly expressed in CRC tumor tissues and plays a tumor suppressive effect [41–43], was highly expressed in CRC samples in the present study without reaching a significant difference in all pairs of samples. It has been reported that SATB2 antisense RNA 1 (SATB2-AS1) is down-regulated and associated with poor survival of patients with CRC [44]. However, this lncRNA was not abnormal in the present study. Up-regulated HOX transcript antisense RNA (HOTAIR), and oncogene colorectal neoplasia differentially expressed (CRNDE) [42], as well as down-regulated long intergenic non-protein coding RNA, p53 induced transcript (LINC-PINT), maternally expressed 3 (MEG3), PR-lncRNA-1 [43,45], PGM5-AS1, and B3GALT5-AS1 [46] were not detected in the present study. The specific reasons for this observation warrant further investigation.

The regulation of target genes by lncRNA can be divided into *cis* and *trans* according to the position relationship between lncRNAs and target genes in the genome [7]. Specifically, *cis* regulation mainly relies on *cis*-acting elements, which are sequences involved in the flanking sequence of genes and can affect gene expression. These elements (e.g. promoters, enhancers, regulatory sequences, inducing elements, etc.) are involved in the regulation of gene expression in the nucleus, and are usually transcribed into ncRNAs [6]. In the present study, a large number of lncRNA *cis*-acting target genes were predicted using software; however, not all target genes were abnormally expressed in CRC samples. This implies that the target genes are *trans*-regulated by lncRNAs or also regulated by other factors. However, differentially expressed mRNAs that co-expressed with differentially expressed lncRNAs in the present study were mostly involved in tumor-related signaling pathways and biological processes related to cell proliferation. These findings indicated that these lncRNAs may play an important role in the occurrence and development of CRC. This hypothesis should be further studied and verified.

In conclusion, we screened 113 differentially expressed mRNAs and six differentially expressed lncRNAs in five groups of CRC and adjacent normal tissues through lncRNA sequencing. Among those, 13 mRNAs were *cis*-regulated by lncRNAs and 53 mRNAs were co-expressed with these six lncRNAs. Finally, we performed GO function and KEGG enrichment analyses on these differentially expressed mRNAs. The results showed that these genes were involved

PORTLAND
PRESS

in the occurrence and development of tumors by improving tumor cell activity, migration ability, and resistance to harmful factors.

Competing Interests

The authors declare that there are no competing interests associated with the manuscript.

Funding

Author Contribution

Zhen-Xu Zhou and Xiao-Ming Chen contributed equally to this work in the design and preparation of this study and should be considered as co-first authors. Xiang-Yang Xue and Guo-Xin Li conceived and designed the research. Zhen-Xu Zhou and Xiao-Ming Chen collected and analyzed the data, and wrote the paper. Yu-Qi Zhang and Liu Peng performed the literature search and were responsible for data visualization. Xiang-Yang Xue and Guo-Xin Li revised the paper critically for intelligent content. All authors have read and approved the final manuscript.

Abbreviations

AS, alternative splicing; CNCI, Coding-Non-Coding Index; CPC, Coding Potential Calculator; CRC, colorectal cancer; FPKM, fragments per kilobase of exon model per million mapped reads; GO, Gene Ontology; KEGG , Kyoto Encyclopedia of Genes and Genomes; LncRNA , long noncoding RNA; PPAR, peroxisome proliferator-activated receptor; SNP, single-nucleotide polymorphism.

References

1 Wilhelm, B.T., Marguerat, S., Watt, S., Schubert, F., Wood, V., Goodhead, I. et al. (2008) Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. *Nature* **453**, 1239–1243, https://doi.org/10.1038/nature07002

2 Alagna, F., D'Agostino, N., Torchia, L., Servili, M., Rao, R., Pietrella, M. et al. (2009) Comparative 454 pyrosequencing of transcripts from two olive genotypes during fruit development. *BMC Genomics* **10**, 399, https://doi.org/10.1186/1471-2164-10-399

3 Wang, E.T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C. et al. (2008) Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**, 470–476, https://doi.org/10.1038/nature07509

4 Edwards, D. and Batley, J. (2010) Plant genome sequencing: applications for crop improvement. *Plant Biotechnol. J.* **8**, 2–9, https://doi.org/10.1111/j.1467-7652.2009.00459.x

5 Soshnev, A.A., Ishimoto, H., McAllister, B.F., Li, X., Wehling, M.D., Kitamoto, T. et al. (2011) A conserved long noncoding RNA affects sleep behavior in Drosophila. *Genetics* **189**, 455–468, https://doi.org/10.1534/genetics.111.131706

6 Wang, K.C. and Chang, H.Y. (2011) Molecular mechanisms of long noncoding RNAs. *Mol. Cell* **43**, 904–914, https://doi.org/10.1016/j.molcel.2011.08.018

7 Mercer, T.R., Dinger, M.E. and Mattick, J.S. (2009) Long non-coding RNAs: insights into functions. *Nat. Rev. Genet.* **10**, 155–159, https://doi.org/10.1038/nrg2521

8 Nagano, T. and Fraser, P. (2011) No-nonsense functions for long noncoding RNAs. *Cell* **145**, 178–181, https://doi.org/10.1016/j.cell.2011.03.014

9 Zhang, J., Sun, P., Gan, L., Bai, W., Wang, Z., Li, D. et al. (2017) Genome-wide analysis of long noncoding RNA profiling in PRRSV-infected PAM cells by RNA sequencing. *Sci. Rep.* **7**, 4952, https://doi.org/10.1038/s41598-017-05279-z

10 Yan, B., Wang, Z.H. and Guo, J.T. (2012) The research strategies for probing the function of long noncoding RNAs. *Genomics* **99**, 76–80, https://doi.org/10.1016/j.ygeno.2011.12.002

11 Kurokawa, R., Rosenfeld, M.G. and Glass, C.K. (2009) Transcriptional regulation through noncoding RNAs and epigenetic modifications. *RNA Biol.* **6**, 233–236, https://doi.org/10.4161/rna.6.3.8329

12 Jiang, C., Li, X., Zhao, H. and Liu, H. (2016) Long non-coding RNAs: potential new biomarkers for predicting tumor invasion and metastasis. *Mol. Cancer* **15**, 62, https://doi.org/10.1186/s12943-016-0545-z

13 Evans, J.R., Feng, F.Y. and Chinnaiyan, A.M. (2016) The bright side of dark matter: lncRNAs in cancer. *J. Clin. Invest.* **126**, 2775–2782, https://doi.org/10.1172/JCI84421

14 Flett, K., Youngster, I., Huang, J., McAdam, A., Sandora, T.J., Rennick, M. et al. (2012) Hand, foot, and mouth disease caused by coxsackievirus a6. *Emerg. Infect. Dis.* **18**, 1702–1704

15 Flynn, R.A. and Chang, H.Y. (2014) Long noncoding RNAs in cell-fate programming and reprogramming. *Cell Stem Cell* **14**, 752–761, https://doi.org/10.1016/j.stem.2014.05.014

16 Rossi, M.N. and Antonangeli, F. (2014) LncRNAs: New Players in Apoptosis Control. *Int. J. Cell Biol.* **2014**, 473857

17 Liu, B., Pan, C.F., He, Z.C., Wang, J., Wang, P.L., Ma, T. et al. (2016) Long Noncoding RNA-LET Suppresses Tumor Growth and EMT in Lung Adenocarcinoma. *Biomed. Res. Int.* **2016**, 4693471

18 Du, L., Duan, W., Jiang, X., Zhao, L., Li, J., Wang, R. et al. (2018) Cell-free lncRNA expression signatures in urine serve as novel non-invasive biomarkers for diagnosis and recurrence prediction of bladder cancer. *J. Cell. Mol. Med.* **22**, 2838–2845, https://doi.org/10.1111/jcmm.13578

19 Duan, W., Du, L., Jiang, X., Wang, R., Yan, S., Xie, Y. et al. (2016) Identification of a serum circulating lncRNA panel for the diagnosis and recurrence prediction of bladder cancer. *Oncotarget* **7**, 78850–78858

20 Bhan, A., Soleimani, M. and Mandal, S.S. (2017) Long Noncoding RNA and Cancer: A New Paradigm. *Cancer Res.* **77**, 3965–3981, https://doi.org/10.1158/0008-5472.CAN-16-2634

21 Zhang, L., Guo, G., Xu, J., Sun, X., Chen, W., Jin, J. et al. (2017) Human cytomegalovirus detection in gastric cancer and its possible association with lymphatic metastasis. *Diagn. Microbiol. Infect. Dis.* **88**, 62–68, https://doi.org/10.1016/j.diagmicrobio.2017.02.001

22 Martin, M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *Embnet J.* **17**, 3

23 Brown, J., Pirrung, M. and McCue, L.A. (2017) FQC Dashboard: integrates FastQC results into a web-based, interactive, and extensible FASTQ quality control tool. *Bioinformatics* **33**, 3137–3139, https://doi.org/10.1093/bioinformatics/btx373

24 Langmead, B. and Salzberg, S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359, https://doi.org/10.1038/nmeth.1923

25 Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R. and Salzberg, S.L. (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36, https://doi.org/10.1186/gb-2013-14-4-r36

26 Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.C., Mendell, J.T. and Salzberg, S.L. (2015) StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **33**, 290–295, https://doi.org/10.1038/nbt.3122

27 Frazee, A.C., Pertea, G., Jaffe, A.E., Langmead, B., Salzberg, S.L. and Leek, J.T. (2015) Ballgown bridges the gap between transcriptome assembly and expression analysis. *Nat. Biotechnol.* **33**, 243–246, https://doi.org/10.1038/nbt.3172

28 Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J. et al. (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28**, 511–515, https://doi.org/10.1038/nbt.1621

29 Conesa, A., Gotz, S., Garcia-Gomez, J.M., Terol, J., Talon, M. and Robles, M. (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**, 3674–3676, https://doi.org/10.1093/bioinformatics/bti610

30 Khalil, M.A., Azhar, A., Anwar, N., Aminullah, Najm ud, D. and Wali, R. (2009) Aetiology, maternal and foetal outcome in 60 cases of obstetrical acute renal failure. *J. Ayub Medical College, Abbottabad : JAMC* **21**, 46–49

31 Wilusz, J.E., Sunwoo, H. and Spector, D.L. (2009) Long noncoding RNAs: functional surprises from the RNA world. *Genes Dev.* **23**, 1494–1504

32 Cabili, M.N., Trapnell, C., Goff, L., Koziol, M., Tazon-Vega, B., Regev, A. et al. (2011) Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.* **25**, 1915–1927

33 Pauli, A., Valen, E., Lin, M.F., Garber, M., Vastenhouw, N.L., Levin, J.Z. et al. (2012) Systematic identification of long noncoding RNAs expressed during zebrafish embryogenesis. *Genome Res.* **22**, 577–591, https://doi.org/10.1101/gr.133009.111

34 Berger, J. and Moller, D.E. (2002) The mechanisms of action of PPARs. *Annu. Rev. Med.* **53**, 409–435, https://doi.org/10.1146/annurev.med.53.082901.104018

35 Feige, J.N., Gelman, L., Michalik, L., Desvergne, B. and Wahli, W. (2006) From molecular action to physiological outputs: peroxisome proliferator-activated receptors are nuclear receptors at the crossroads of key cellular functions. *Prog. Lipid Res.* **45**, 120–159, https://doi.org/10.1016/j.plipres.2005.12.002

36 Wang, G., Bonkovsky, H.L., de Lemos, A. and Burczynski, F.J. (2015) Recent insights into the biological functions of liver fatty acid binding protein 1. *J. Lipid Res.* **56**, 2238–2247, https://doi.org/10.1194/jlr.R056705

37 Botero, J.E., Contreras, A. and Parra, B. (2008) Effects of cytomegalovirus infection on the mRNA expression of collagens and matrix metalloproteinases in gingival fibroblasts. *J. Periodont Res.* **43**, 649–657, https://doi.org/10.1111/j.1600-0765.2007.01053.x

38 Gross, J. and Lapiere, C.M. (1962) Collagenolytic activity in amphibian tissues: a tissue culture assay. *PNAS* **48**, 1014–1022, https://doi.org/10.1073/pnas.48.6.1014

39 Flowers, M.T. and Ntambi, J.M. (2009) Stearoyl-CoA desaturase and its relation to high-carbohydrate diets and obesity. *Biochim. Biophys. Acta* **1791**, 85–91

40 Dalgleish, T., Rolfe, J., Golden, A.M., Dunn, B.D. and Barnard, P.J. (2008) Reduced autobiographical memory specificity and posttraumatic stress: exploring the contributions of impaired executive control and affect regulation. *J. Abnorm. Psychol.* **117**, 236–241, https://doi.org/10.1037/0021-843X.117.1.236

41 Ni, W., Yao, S., Zhou, Y., Liu, Y., Huang, P., Zhou, A. et al. (2019) Long noncoding RNA GAS5 inhibits progression of colorectal cancer by interacting with and triggering YAP phosphorylation and degradation and is negatively regulated by the m 6 A reader YTHDF3. *Mol. Cancer* **18**, 143, https://doi.org/10.1186/s12943-019-1079-y

42 Han, D., Wang, M., Ma, N., Xu, Y., Jiang, Y. and Gao, X. (2015) Long noncoding RNAs: novel players in colorectal cancer. *Cancer Lett.* **361**, 13–21, https://doi.org/10.1016/j.canlet.2015.03.002

43 Huarte, M. (2015) The emerging role of lncRNAs in cancer. *Nat. Med.* **21**, 1253–1261, https://doi.org/10.1038/nm.3981

44 Xu, M., Xu, X., Pan, B., Chen, X., Lin, K., Zeng, K. et al. (2019) LncRNA SATB2-AS1 inhibits tumor metastasis and affects the tumor immune cell microenvironment in colorectal cancer by regulating SATB2. *Mol. Cancer* **18**, 135, https://doi.org/10.1186/s12943-019-1063-6

45 Wu, M., Li, W., Huang, F., Sun, J., Li, K.P., Shi, J. et al. (2019) Comprehensive Analysis of the Expression Profiles of Long Non-Coding RNAs with Associated ceRNA Network Involved in the Colon Cancer Staging and Progression. *Sci. Rep.* **9**, 16910, https://doi.org/10.1038/s41598-019-52883-2

46 Zhang, Q., Ding, Z., Wan, L., Tong, W., Mao, J., Li, L. et al. (2020) Comprehensive analysis of the long noncoding RNA expression profile and construction of the lncRNA-mRNA co-expression network in colorectal cancer. *Cancer Biol. Ther.* **21**, 157–169