SCIENTIFIC REPORTS

OPEN

Meta-signature LncRNAs serve as novel biomarkers for colorectal cancer: integrated bioinformatics analysis, experimental validation and diagnostic evaluation

Meiyu Dai[1,*], Xiaoli Chen[1,*], Shanying Mo[1,*], Jinwan Li[1], Zhizhuo Huang[1], Shifeng Huang[2], Junyi Xu[2], Baoyu He[1], Yan Zou[1], Jingfan Chen[2] & Shengming Dai[1]

The aim of this study is to explore the differentially expressed lncRNAs, which may have potential biological function and diagnostic value in colorectal cancer (CRC). Through integrated data mining, we finally identified nine differentially expressed lncRNAs and their potential mRNA targets. After a series of bioinformatics analyses, we screened significant pathways and GO terms that are related to the up-regulated and down-regulated transcripts respectively. Meanwhile, the nine lncRNAs were validated in 30 paired tissues and cell lines by qRT-PCR and the results were basically consistent with the microarray data. We also tested the nine lncRNAs in the serum of 30 CRC patients matched with the CRC tissue, 30 non-cancer patients and 30 health controls. Finally, we found that BLACAT1 was significant for the diagnosis of CRC. The area under the curve (AUC), sensitivity and specificity were 0.858 (95% CI: 0.765– 0.951), 83.3% and 76.7% respectively between CRC patients and health controls. Moreover, BLACAT1 also had distinct value to discriminate CRC from other non-cancer diseases. The results indicated that the differentially expressed lncRNAs and their potential target transcripts could be considered as potential therapeutic targets for CRC patients. Meanwhile, lncRNA BLACAT1 might represent a new supplementary biomarker for the diagnosis of CRC.

Cancer is the biggest threat to human life and now has been a hotspot in academic research around the world. Statistical report indicates that there are 1,665,540 cases of new cancers and 585,720 cancer-related deaths every year in the United States[1]. Among them, colorectal cancer (CRC) is the second most common cancer in female and third in male with the occurrence of over one million new cases and half a million deaths by the global cancer statistics[2]. In China, the incidence of CRC ranks third and its mortality ranks fourth in various cancers[3]. Although considerable progress has been made in medical science, the incidence and mortality of CRC are still increasing year by year[4]. Therefore, it is essential to find novel biomarkers for the early diagnosis and targeted therapy for CRC patients.

Long non-coding RNAs (lncRNAs), following siRNA and microRNA, are another new class of transcripts that have been found pervasively associated with human diseases. LncRNA, with its sequence of longer than 200 bp, has the poly A tail and promoters in structure after splicing[5]. In the different processes of tissue differentiation and development, most lncRNAs have obviously specific expression. Meanwhile, they also have different expression patterns in different parts of the human tissue[6,7]. At present, several screening diagnosis methods for CRC are widely used, including digital rectal examination, imaging examination, endoscopy, fecal occult blood examination and tumor biomarkers detection. But they all have limitations in such aspects as high false negative rate, complex operation process, low sensitivity and specificity, and excessive dependence on the clinical pheno-type. With the rapid development of chip technology and high-throughput sequencing, the genetic biomarkers

[1]Medical Science Laboratory, the Fourth Affiliated Hospital of Guangxi Medical University, Liuzhou, Guangxi 545005, China. [2]Department of General Surgery, the Fourth Affiliated Hospital of Guangxi Medical University, Liuzhou, Guangxi 545005, China. *These authors contributed equally to this work. Correspondence and requests for materials should be addressed to J.C. (email: 3203008361@qq.com) or S.D. (email: daishm@sina.com)

in tumor have become increasingly valuable. A growing body of research indicate that lncRNAs play crucial roles in various human diseases, especially in cancers[8–11], some of which become classic research hotspots and are studied more comprehensively. For example, lncRNA HOTAIR is a poor prognostic indicator for breast cancer metastasis, pancreatic cancer, CRC and is also related to cell cycle progression in human glioma[12,13]. However, individual lncRNA in the diagnosis and prognosis of the disease is lack of specificity, so lncRNA expression signature or profile has been adopted by many researchers[10,14,15]. In recent years, there are many studies on lncRNAs in CRC. For instance, lncRNA CRNDE, is not only associated with the diagnosis and prognosis of CRC, but also promoted CRC cell proliferation and chemoresistance[16,17]. LncRNA H19 was an independent prognostic factor and mediated methotrexate resistance in colorectal cancer through Wnt/β-catenin signal pathway[18,19]. However, they are also differential expression genes and play an important role in other cancers like glioma[20,21], ovarian cancer[22,23] and hepatic carcinoma[24,25]. Therefore, more novel lncRNAs associated with CRC need to be found and detected in the future application.

The rapid development of modern gene chip technology and bioinformatics analysis makes it possible to identify and study more novel lncRNAs. Many researchers start to search for the significant differential expression genes through the data mining from public databases. Then, the functional enrichment analysis and experimental verification of target genes are carried out to provide reference for clinical target treatment, early diagnosis and prognosis evaluation[26–28]. Some evidence indicate that lncRNAs acting as competing endogenous RNAs (ceRNAs) are involved in a variety of tumor initiation and progression. Then systematic studies on lncRNA-associated ceRNA network also have been performed in various cancers such as ovarian cancer[29], breast cancer[30], and gastric cancer[31]. It could provide novel insights to understand the molecular mechanism and functions of lncRNAs in cancers.

In our study, a systematic expression profiling analysis was introduced to identify meta-signature lncRNAs between the CRC tissues and adjacent normal tissues. Meanwhile, these selected lncRNAs were also validated in tissues and cell lines. Then bioinformatics analysis was adopted to further study the target genes and functional annotation of the differentially expressed lncRNAs, which might contribute to the research on targeted therapy of CRC. Finally, we also detected the expression levels of lncRNAs in serum among CRC patients, non-cancer patients and health controls to find out novel biomarkers for the diagnosis and differentiation of CRC.

Results

Differentially expressed lncRNAs and mRNAs in the microarray datasets.
In our study, a total of eight datasets were included for comprehensive analysis (GSE77199, GSE76855, GSE32323, GSE62321, GSE8671, GSE20842, GSE39582, and GSE21815). After data preprocessing, we finally found 21,439 mRNAs and 1,299 lncRNAs in GSE77199 dataset, 12,319 mRNAs and 2,367 lncRNAs in GSE76855 dataset, 18,115 mRNAs and 1,944 lncRNAs in GSE32323 dataset, 8,455 mRNAs and 597 lncRNAs in GSE62321 dataset, 18,115 mRNAs and 1,944 lncRNAs in GSE8671 dataset, 18,228 mRNAs and 481 lncRNAs in GSE20842 dataset, 18,115 mRNAs and 1,944 lncRNAs in GSE39582 dataset and 18,545 mRNAs and 1,219 lncRNAs in GSE21815 dataset. With the standard of fold change ≥2.0 and p ≤ 0.05, difference analyses were performed in the eight datasets respectively. Then the method of robust rank aggregation (RRA) was used to integrally calculate the differentially expressed lncRNAs and mRNAs of the eight datasets. Finally, nine lncRNAs (including five up-regulated lncRNAs: UCA1, CRNDE, H19, ZFAS1, BLACAT1, and four down-regulated lncRNAs: LINC00675, DPP10-AS1, LOC344887, HAGLR) and 209 mRNAs (including 80 up-regulated mRNAs and 129 down-regulated mRNAs) were identified as the most significantly differential genes. Detailed information was listed in Supplementary Tables S1 and S2.

Target mRNAs selection and lncRNA-mRNA co-expression network.
In order to construct the lncRNA-mRNA co-expression network, we calculated the correlation coefficient between nine lncRNAs and 209 mRNAs selected from microarray data. Eventually, we found 140 mRNAs (48 up-regulated mRNAs and 92 down-regulated mRNAs) had strong correlation with the nine lncRNAs according to the Pearson's correlation coefficient analysis. Then we constructed the lncRNA-mRNA co-expression network by Cytoscape (version 3.3.0) to visualize the relevance between them (Fig. 1). The size of each node is proportional to the calculated functional connectivity of each lncRNA according to the counts of related target genes. Therefore, significant target genes such as TP53INP2, MMP28, C2orf88, LDHD, and NR3C2 etc. in the network were found to be regulated by multiple lncRNAs. In addition, some related studies show that these genes play a regulatory role in the human body function[17,32,33].

GO and KEGG pathway analyses.
In our study, the up-regulated and down-regulated mRNAs were performed with Gene ontology (GO) and KEGG pathway analyses separately. In the GO analysis, count ≥2 and p ≤ 0.05 were considered to be statistically significant. Usually, the top 10 objects were selected for concrete analysis. According to the proportion of the count, immune response (15%) and cell adhesion (15%) in the biological process, extracellular space (45%) and extracellular region (45%) in the cellular component, and chemokine activity (27%) in the molecular function were the major terms corresponding to the up-regulated mRNAs, while small molecule metabolic process (35%) in the biological process, extracellular exosome (62%) in the cellular component, and zinc ion binding (44%) in the molecular function were the major terms corresponding to the down-regulated mRNAs (Fig. 2). The p value and other detailed information were listed in Supplementary Table S3. In the KEGG pathway analysis, 5 pathways were associated with the up-regulated mRNAs and 11 pathways related to the down-regulated mRNAs (Fig. 3 and Supplementary Table S4).

qRT-PCR validation of the nine differentially expressed lncRNAs.
Validation between CRC tissues and adjacent normal tissues. All the nine lncRNAs, five up-regulated ones (UCA1, CRNDE, H19, ZFAS1 and BLACAT1) and four down-regulated ones (LINC00675, DPP10-AS1, LOC344887 and HAGLR), were
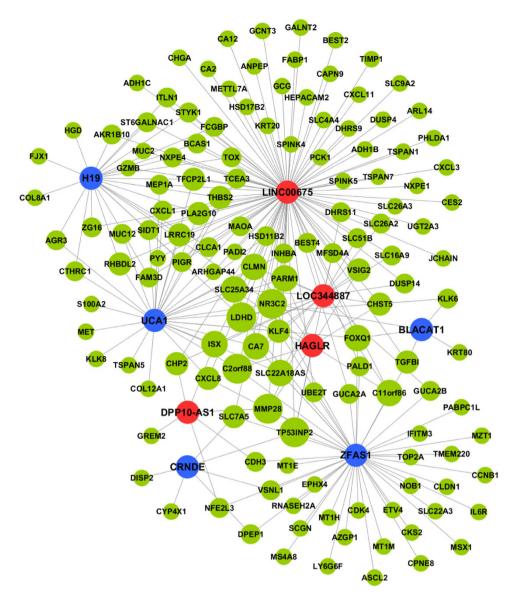
Figure 1. LncRNAs-mRNAs co-expression network. Blue color nodes represent up-regulation, red color nodes represent down-regulation and the others represent mRNA. The size of each node is proportional to the calculated functional connectivity of each lncRNA according to the counts of related target genes.

selected for further verification by qRT-PCR. The expression tendency of the five up-regulated lncRNAs (UCA1, CRNDE, H19, ZFAS1 and BLACAT1) and two down-regulated lncRNAs (DPP10-AS1 and HAGLR) were same as the results of microarray data ($p < 0.05$). In the two remaining down-regulated lncRNAs (LINC00675 and LOC344887), their expression levels were opposite to the microarray results ($p < 0.05$). All the detailed information was shown in Fig. 4.

Validation in cell lines. To further validate the lncRNA meta-signature, qRT-PCRs were also performed in all three CRC cell lines and normal human colorectal epithelial cell line (FHC). Expression levels of all the lncRNAs were almost consistent with lncRNA microarray results except for the undetectable DPP10-AS1 (Supplementary Figure S1). Moreover, we also detected the supernatant of cell culture, and their expression levels were basically consistent with the results of cells (Supplementary Figure S2).

Diagnostic value of lncRNA BLACAT1 in serum. Since part of lncRNAs were detectable in the supernatant of cell culture, so we established a training set in the serum among CRC patients, non-cancer patients and health controls. The results indicated that only lncRNA ZFAS1 and BLACAT1 were steadily detected among the three groups where they tended to be highly expressed. Then the two screened lncRNAs were further examined in the validation set. The results of the validation set and training + validation set revealed that the expression levels of BLACAT1 in CRC patients were higher than that in health controls (Fig. 5) but ZFAS1 had no significantly statistical difference (Supplementary Figure S3). Meanwhile, it was also found that BLACAT1 was significantly
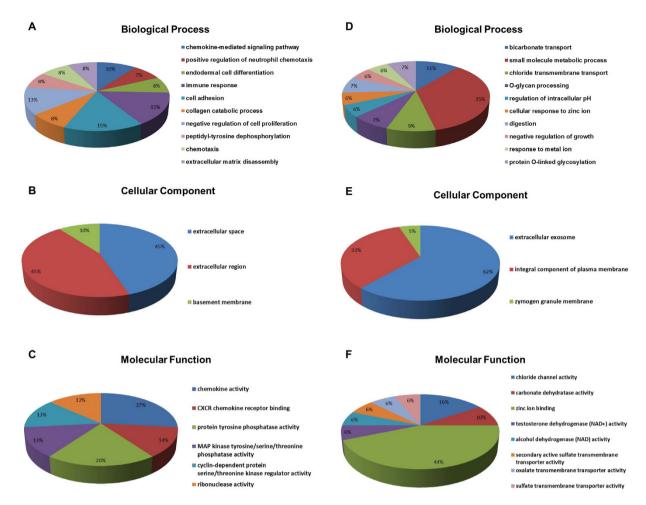
Figure 2. GO terms of the target genes. (**A**) GO analysis of up-regulated target genes according to biological process, (**B**) GO analysis of up-regulated target genes according to cellular component, (**C**) GO analysis of up-regulated target genes according to molecular function, (**D**) GO analysis of down-regulated target genes according to biological process, (**E**) GO analysis of down-regulated target genes according to cellular component, (**F**) GO analysis of down-regulated target genes according to molecular function.
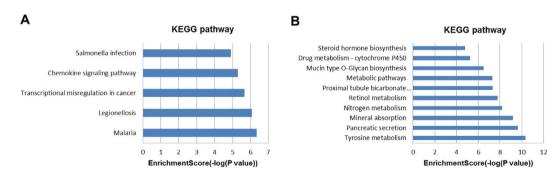

Figure 3. KEGG pathway analysis of target genes. (**A**) Pathways corresponding to the up-regulated target genes. (**B**) Pathways corresponding to the down-regulated target genes.

up-regulated between CRC patients and non-cancer patients (Fig. 5). In our study, we also evaluated the diagnostic accuracy of BLACAT1 among CRC patients, non-cancer patients and health controls. The area under the curve (AUC) were 0.858 (95% CI: 0.765–0.951) for BLACAT1 between CRC patients and health controls, and 0.800 (95% CI: 0.689–0.911) between CRC patients and non-cancer patients. The respective sensitivity and specificity were 83.3% and 76.7% for BLACAT1 with the cutoff value of 1.177 between CRC patients and health controls, 56.7% and 92.9% with the cutoff value of 5.035 between CRC patients and non-cancer patients (Fig. 6). The results indicated that lncRNA BLACAT1 was extremely valuable for the diagnosis of CRC.
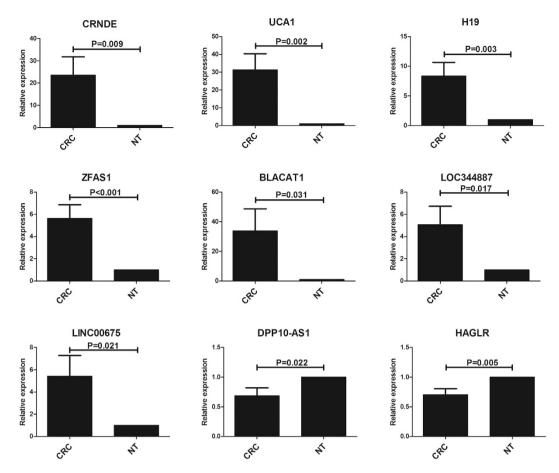
Figure 4. qRT-PCR analysis of nine lncRNAs expression in the CRC tissues and the adjacent non-cancerous tissue. CRC: colorectal cancer; NT, non-cancerous colorectal tissue.

Discussion

It has been proven in extensive researches on cancers that the differentially expressed lncRNA pattern was obviously associated with the pathogenesis of various malignant tumors[34,35]. In our study, nine lncRNAs meta-signature were incorporated for integrated analysis through a series of data mining and calculation. Among the nine lncRNAs, H19 and UCA1 were the star biomarkers in various types of cancers, such as breast cancer[36,37], gastric cancer[38,39], lung cancer[40,41], and colorectal cancer[42,43]. Meanwhile, both of them were closely associated with tumor growth, migration, invasion, proliferation, and prognosis. LncRNA CRNDE was mainly reported to affect the biological characteristics of gliomas[44,45]. However, there was only one article about CRC, which reported up-regulated CRNDE as an indicator to poor prognosis[16]. For lncRNA ZFAS1, there were two papers concerning CRC, among which one study reported that up-regulation predicted poor prognosis[46] while the other illuminated the interaction between ZFAS1 and Cyclin-dependent kinase 1[47]. LncRNA BLACAT1 and four down-regulated lncRNAs (LOC344887, LINC00675, DPP10-AS1, and HAGLR) were all novel biomarkers for the CRC and there were no detailed reports relevant to CRC for now. In recent years, more and more research indicated that mutations and abnormal expression of lncRNAs were closely related to various cancers. Some researchers have also summarized some of the common functions of the lncRNAs, important lncRNA-related diseases, common disease-associated lncRNAs and some classic lncRNA public databases to provide an important reference for clinical disease research[48]. For now, a plenty of lncRNAs have been found, but only a limited number of lncR-NAs that is related to diseases can be verified by experiments. Therefore, it has become a heated issue to develop some powerful computational models for lncRNA biomarker identification and prediction such as IRWRLDA[49], HGIMDA[50], FMLNCSIM[51], and LRLSLDA[52]. Based on rapid development of the computer and bioinformatics technology, our study could offer a new perspective and framework for the further research with respect to biological function and serological biomarkers of CRC.

In the present study, we finally identified nine lncRNAs and 140 mRNAs that were differentially expressed between CRC tissues and adjacent normal tissues through a series of integrated analysis. In order to explore the potential function of the differentially expressed transcripts, bioinformatics methods such as GO and KEGG pathway analyses were carried out in our research. A single lncRNA may target multiple transcripts, and a specific transcript may be regulated by multiple lncRNAs, which prompts lncRNAs to induce transformations in various GO processes and signal pathways. In our study, the results of GO analysis indicated that the up-regulated transcripts were principally enriched in cell functions, extracellular component and protease activity, and the down-regulated transcripts were mainly enriched in molecule metabolic processes, exosome and ion activity.
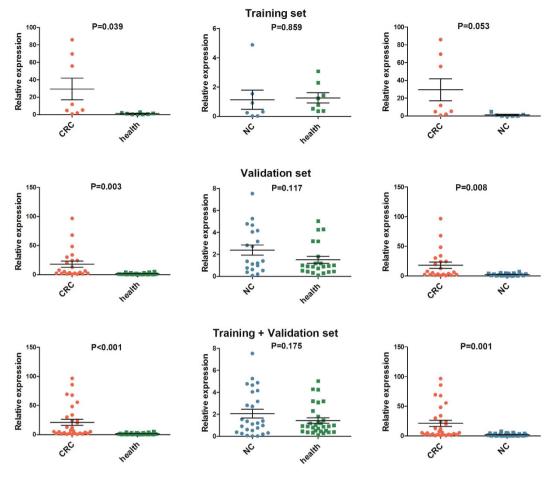
Figure 5. The expression levels of BLACAT1 in serum samples among CRC patients, non-cancer patients and health controls. CRC: colorectal cancer; NC: non-cancer patients.
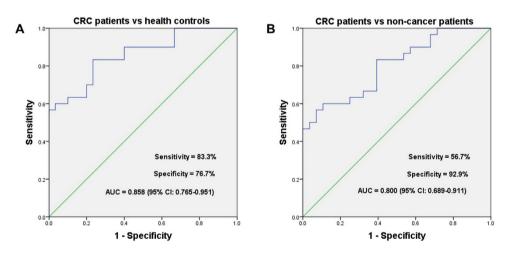

Figure 6. The ROC curve of LncRNA BLACAT1 in the diagnosis of CRC. (A) CRC patients vs health control; (**B**) CRC patients vs non-cancer patients.

Cell function processes consist of cell adhesion, differentiation, proliferation and so on. Dysregulation of all these cellular properties may be closely related to the onset and progression of CRC[53]. Molecule metabolic processes primarily contain small molecule metabolic process, one-carbon metabolic process, xenobiotic metabolic process, which may play an important role in the clinical application in the control, treatment, and prevention of CRC[54]. In short, each process has its value, which may be associated with the occurrence, development and prognosis of CRC.

KEGG pathway analysis demonstrated that 16 pathways were related to the dysregulated transcripts, of which 5 pathways were associated with the up-regulated transcripts and 11 pathways were associated with the down-regulated transcripts. Transcriptional misregulation in cancer and chemokine signaling pathway were the major enriched pathways for up-regulated transcripts, while metabolic pathways and pancreatic secretion were the primary pathways for down-regulated transcripts. These pathways needed to be further studied in that they were highly related to the cancer biology. According to the results of pathways enrichment analyses, we found that the meta-signature lncRNAs were the key regulatory factors for the tumor process and they might be potential therapeutic targets for CRC patients.

After validation between CRC tissues and paired adjacent normal tissues, the results of seven lncRNAs were consistent with the data screened from Gene Expression Omnibus (GEO) database, while the results of two lncRNAs were not. This may attribute to the limitations of data mining algorithm and different test methods mentioned in many related reports[55,56]. The qRT-PCR results of cell lines basically corresponded to microarray data. Meanwhile, we found part of the lncRNAs were detectable in the supernatant of cell culture and lncRNAs were reported to be stably expressed in serum samples[57,58]. Therefore, in addition to explore the function of target genes of the lncRNAs, we also expanded the scope of our study to find out whether the lncRNAs exhibited differential expression in serum among CRC patients, non-cancer patients and health controls. In this way, some novel biomarkers could be detected as a reference for the diagnosis and differentiation of CRC. In the preliminary serological test, we found two of the nine lncRNAs were detectable by qRT-PCR. Then we increased the sample size and tested the serum of other non-cancer patients. Finally, we found lncRNA BLACAT1 were significant for the diagnosis of CRC. Meanwhile, the BLACAT1 not only has an obviously diagnostic value for CRC, but it has distinctly differential value to separate CRC from other non-cancer diseases. In the co-expression network, we could find that the target genes of lncRNA BLACAT1 were mainly gene NR3C2, FOXQ1, TGFB1, KLK6 and KRT80. Meanwhile, gene NR3C2, FOXQ1 and TGFB1 interacted with multiple genes. It followed that lncRNA BLACAT1 might play an important role in the gene regulation network. However, there are still some limitations in our study, which rest on insufficient sample cases to perform subgroup analysis on clinical characteristics. In the further study, we will increase the sample size and make further study on the cell function and regulatory mechanism of BLACAT1 in CRC.

In summary, nine differentially expressed lncRNAs and 140 mRNAs strongly related to the lncRNAs were identified by data mining. After integrated GO analyses, KEGG signal pathways analyses and a series of experimental validation, we found that the differentially expressed lncRNAs and their potential target mRNAs influenced the pathogenesis of CRC and might be considered as therapeutic targets. Moreover, lncRNA BLACAT1 is found to be at obviously high expression in serum, which might represent a new supplementary biomarker for the diagnosis of CRC.

Materials and Methods

CRC gene expression datasets. All the CRC datasets were downloaded from GEO (http://www.ncbi.nlm.nih.gov/geo/). The selection criteria used in this study are as follows: (1) all specimens classified as tissues; cells, serum or plasma are not included; (2) all the included datasets must contain paired CRC tumors and adjacent non-cancerous tissues; (3) both lncRNAs and mRNAs should be included in all the gene expression profiles; (4) sample size should be greater than three pairs; (5) if there existed data overlapping, the largest sample size was selected. According to the above screening criteria, eight datasets were finally included in this study (GSE77199, GSE76855, GSE32323, GSE62321, GSE8671, GSE20842, GSE39582, and GSE21815).

Microarray data processing and mining. Series matrix files of each dataset were downloaded from GEO. Normalization of the eight datasets was performed by robust multi-array average (RMA) from Gene Expression Console (affymetrix). The normalization procedure such as log2-transformation was separately performed in each gene expression dataset. Then, we screened out the differentially expressed lncRNAs and mRNAs from each dataset on the basis of $p \leq 0.05$ and fold change ≥ 2. Due to the eight datasets from different technological platforms, we could integrally calculate them using the method of RRA, which is suitable and effective for identifying differentially expressed lncRNA meta-signature[59,60]. The new data frame results were constructed with the standard of adjusted p value < 0.05. The operation process can be performed by the Robust Rank Aggreg package in R software (version 3.2.3).

LncRNA functional analyses. *LncRNA-mRNA co-expression network.* The construction of gene co-expression network is mostly developed on basis of the correlation among gene expression profiles. In our study, the lncRNA-mRNA co-expression network construction was mainly divided into three steps: (1) calculate the Pearson's correlation coefficient between the differentially expressed lncRNAs and mRNAs; (2) identify the differentially expressed mRNAs which significantly correlate with the lncRNAs on basis of the correlation coefficient absolute value of 0.8 (0.8–1.0: extremely strong, 0.6–0.8: strong, 0.4–0.6: moderate, 0.2–0.4: weak, 0–0.2: weak or no correlation); (3) draw the lncRNA-mRNA co-expression network through the software of Cytoscape (version 3.3.0).

Gene ontology and pathway analyses. GO, as a representative term to describe the standard features of genes and their products, is widely used in the field of bioinformatics. In this way, the data of gene and gene products can be unified, processed, explained and shared. The relative terms of the GO are divided into three broad categories, including cellular component, molecular function and biological process. According to the GO analysis, we could find the relative terms of the differential gene and further identify differential gene between the samples regarding the change of certain gene function. Pathway analysis is also frequently used in function analysis of the microarray data. However, pathway analysis is different from GO analysis. It employs the resource of biological

Sample NO.	Gender	Age	Tumor site	Tumor morphology	Lymph node	Metastasis	TNM Stage
CRC-1	femal	78	colon	T4	N2	M0	IIIC
CRC-2	femal	54	rectum	T4	N2	M0	IIIC
CRC-3	femal	82	rectum	T4	N1	M0	IIIB
CRC-4	male	62	rectum	T3	N2	M0	IIIC
CRC-5	male	49	colon	T4	N0	M0	IIB
CRC-6	femal	74	rectum	T3	N0	M0	IIA
CRC-7	femal	65	colon	T4	N0	M0	IIB
CRC-8	male	51	rectum	T3	N0	M1	IV
CRC-9	female	73	rectum	T2	N0	M0	I
CRC-10	male	78	colon	T4	N0	M0	IIB
CRC-11	female	24	rectum	T2	N0	M0	I
CRC-12	female	60	colon	T3	N1b	M0	IIIB
CRC-13	male	66	colon	T3	N0	M0	IIA
CRC-14	male	52	colon	T4	N1	M0	IIIB
CRC-15	female	57	colon	T4	N1	M0	IIIB
CRC-16	male	62	rectum	T4	N0	M0	IIB
CRC-17	female	39	colon	T4	N0	M0	IIB
CRC-18	male	63	colon	T3	N1	M0	IIIB
CRC-19	male	55	rectum	T4a	N0	M0	IIB
CRC-20	male	69	colon	T4a	N0	M0	IIB
CRC-21	female	73	colon	T4	N1	M0	IIIB
CRC-22	male	51	colon	T2	N1	M0	IIIA
CRC-23	male	51	colon	T3	N0	M0	IIA
CRC-24	female	47	rectum	T3	N0	M0	IIA
CRC-25	male	87	colon	T4	N2	M1	IV
CRC-26	male	57	colon	T3	N1	M0	IIIB
CRC-27	female	67	colon	T4a	N0	M0	IIB
CRC-28	female	61	rectum	T3	N0	M0	IIA
CRC-29	male	54	rectum	T3	N2	M0	IIIC
CRC-30	male	47	colon	T3	N0	M0	IIA

Table 1. Clinical characteristics of 30 CRC patients.

pathways that have been clearly studied. In our study, the analyses of GO and KEGG pathway were performed with DAVID (Database for Annotation, Visualization, and Integrated Discovery, https://david.ncifcrf.gov/). The selection criteria of the GO terms and KEGG pathways were $p \leq 0.05$ and count ≥ 2.

Validation of the integrated-signature lncRNAs. *Tissue samples of CRC patients.* To validate the differentially expressed lncRNA-signature, we collected 30 pairs of cancer tissues and adjacent non-cancerous tissues of CRC patients. The patients' information was summarized in Table 1. There were no statistical difference in gender, age, tumor site and TNM stage. All the tissues were obtained from the pathologically confirmed CRC patients by the experienced surgeons at the Fourth Affiliated Hospital of Guangxi Medical University from 2014 to 2016. Informed consents were obtained from all patients and our experiments were also approved by the Institutional Review Boards of the Fourth Affiliated Hospital of Guangxi Medical University. In addition, all the experiment methods of our study were performed by following with relevant guidelines and regulations. The tissue samples were frozen in liquid nitrogen immediately after surgical resection and subsequently stored at −80 °C before use.

Cell culture. FHC and CRC cells (SW480, HT-29 and LS 174T) were obtained from Shanghai Institutes for Biological Sciences. The CRC cells were cultured in Dulbecco's modified Eagle's medium (DMEM) or Memorial Institute (RPMI) 1640 medium supplemented with 10% fetal bovine serum (FBS, Sigma Chemical Co., St. Louis, USA) and 100 units/ml antibiotics of penicillin-streptomycin (Invitrogen, Carlsbad, CA) in a humidified incubator with the atmosphere of 5% CO_2 at 37 °C. FHC were cultured in the base medium of DMEM: F12 medium supplemented with 10 ng/ml cholera toxin, 0.005 mg/ml insulin, 0.005 mg/ml transferrin, 100 ng/ml hydrocortisone, extra 10 mM 4-(2-hydroxyethyl)-1-piperazine-ëthanesulfonic acid (HEPES, for a final concentrate of 25 mM) and 10% FBS (final concentrate). The culture atmosphere and temperature were the same as what was set for the CRC cells. Meanwhile, we also collected the supernatant of cell culture after 24-hour steady growth.

Total RNA extraction and quantitative real-time PCR. The total RNA was extracted using the RNAiso Plus (TAKARA BIO INC.) and the concentration was detected by NanoDrop 2000. Total RNA 1 μg was added for synthesizing the first-strand complementary DNA (cDNA) according to the manufacturer's instructions of the PrimerScript™ RT reagent Kit with gDNA Eraser (TAKARA BIO INC.) The quantitative real-time PCR

Name	Forward	Reverse
CRNDE	5′-ACATGGAAAAATCAAAGTGCTCG-3′	5′-TAACCTTCTTCTGCGTGACAAC-3′
UCA1	5′-CTCTCCATTGGGTTCACCATTC-3′	5′-GCGGCAGGTCTTAAGAGATGAG-3′
H19	5′-CAACATCAAAGACACCATCGG-3′	5′-GAGACAGAAGGATGAAAAAGAAGAA-3′
ZFAS1	5′-GGCTTCATACGCTATTGTCCTG-3′	5′-CTTCCAACACCCGCATTCAT-3′
BLACAT1	5′-ATGACTGACTCCTGACCTTGGCAA-3′	5′-GAGGAAATGGGACTCATCGCC-3′
LOC344887	5′-TCCTCTTTACGGCACAACATTC-3′	5′-TTCCGAGGCTTCTCATTTTACC-3′
LINC00675	5′-ACCACAAGCACCAAAGTCCTAA-3′	5′-GGTCAGTGTCAAAGGGTAGATA-3′
DPP10-AS1	5′-ATCCAGCCCAGATTCTCCTACC-3′	5′-GCCTTCAGGTGGCTGTTTTGTA-3′
HAGLR	5′-GCCTGTGACTGTCCCTTGAATA-3′	5′-GATGTGTGTGGGTGTGTGTTTG-3′

Table 2. The primer sequences of the nine lncRNAs.

(qRT-PCR) was performed using the reagent of SYBR Premix Ex Taq™ II (TAKARA BIO INC.) and the 7500 Real-Time PCR Instrument from Life technologies. The primers of all the lncRNAs were obtained from TAKARA and the sequences were listed in Table 2. B2M gene was chosen as an internal control. The PCR thermal cycle process was: 95 °C for 10 min following 40 cycles of 95 °C for 15 s, 60 °C for 1 min, and the melting curve stage was set at the end of the cycles. The relative expression of each lncRNA was calculated on the basis of $2^{-\Delta\Delta Ct}$.

Serum collection and study design. *Serum samples of CRC patients, non-cancer patients and health controls.* In our study, a total of 90 serum samples were selected for detection, of which 30 samples are collected from CRC patients matched with tissues before surgery, 30 from patients with non-cancer diseases including ulcerative colitis (10 cases), proctitis (10 cases) and intestinal polyp (10 cases), and 30 from sex- and age-matched health controls. All the serum samples were recruited by the Fourth Affiliated Hospital of Guangxi Medical University from 2014 to 2016 and stored at −80 °C before use.

Study design. Serum test in our study mainly included three steps: (1) the serum expression levels of the nine lncRNAs were detected in eight CRC patients, eight non-cancer patients and eight health controls to confirm which lncRNAs were really detectable in the serum samples; (2) those detectable lncRNAs in the first step were tested in additional 22 CRC patients, 22 non-cancer patients and 22 health controls; (3) a diagnostic model was constructed to evaluate the diagnostic performance of the differentially expressed lncRNAs in serum samples among the three groups.

References

1. Siegel, R., Ma, J., Zou, Z. & Jemal, A. Cancer statistics, 2014. *CA: a cancer journal for clinicians* **64,** 9–29, doi: 10.3322/caac.21208 (2014).
2. Jemal, A. *et al.* Global cancer statistics. *CA: a cancer journal for clinicians* **61,** 69–90, doi: 10.3322/caac.20107 (2011).
3. Chen, W., Zheng, R., Zeng, H., Zhang, S. & He, J. Annual report on status of cancer in China, 2011. *Chinese journal of cancer research = Chung-kuo yen cheng yen chiu* **27,** 2–12, doi: 10.3978/j.issn.1000-9604.2015.01.06 (2015).
4. Zullig, L. L. *et al.* Colorectal Cancer Statistics From the Veterans Affairs Central Cancer Registry. *Clinical colorectal cancer*, doi: 10.1016/j.clcc.2016.04.005 (2016).
5. Wapinski, O. & Chang, H. Y. Long noncoding RNAs and human disease. *Trends in cell biology* **21,** 354–361, doi: 10.1016/j.tcb.2011.04.001 (2011).
6. Sun, M. & Kraus, W. L. From discovery to function: the expanding roles of long noncoding RNAs in physiology and disease. *Endocrine reviews* **36,** 25–64, doi: 10.1210/er.2014-1034 (2015).
7. Kanduri, C. Long noncoding RNAs: Lessons from genomic imprinting. *Biochimica et biophysica acta* **1859,** 102–111, doi: 10.1016/j.bbagrm.2015.05.006 (2016).
8. Zhu, X. *et al.* A long non-coding RNA signature to improve prognosis prediction of gastric cancer. *Molecular cancer* **15,** 60, doi: 10.1186/s12943-016-0544-0 (2016).
9. Van Grembergen, O. *et al.* Portraying breast cancers with long noncoding RNAs. *Science advances* **2,** e1600220, doi: 10.1126/sciadv.1600220 (2016).
10. Hu, Y. *et al.* A long non-coding RNA signature to improve prognosis prediction of colorectal cancer. *Oncotarget* **5,** 2230–2242, doi: 10.18632/oncotarget.1895 (2014).
11. Zhou, M. *et al.* A potential signature of eight long non-coding RNAs predicts survival in patients with non-small cell lung cancer. *Journal of translational medicine* **13,** 231, doi: 10.1186/s12967-015-0556-3 (2015).
12. Xue, Y. *et al.* Genome-wide analysis of long noncoding RNA signature in human colorectal cancer. *Gene* **556,** 227–234, doi: 10.1016/j.gene.2014.11.060 (2015).
13. Zhang, K. *et al.* Long non-coding RNA HOTAIR promotes glioblastoma cell cycle progression in an EZH2 dependent manner. *Oncotarget* **6,** 537–546, doi: 10.18632/oncotarget.2681 (2015).
14. Zhou, M. *et al.* Comprehensive analysis of lncRNA expression profiles reveals a novel lncRNA signature to discriminate nonequivalent outcomes in patients with ovarian cancer. *Oncotarget* **7,** 32433–32448, doi: 10.18632/oncotarget.8653 (2016).
15. Sun, J. *et al.* A potential prognostic long non-coding RNA signature to predict metastasis-free survival of breast cancer patients. *Scientific reports* **5,** 16553, doi: 10.1038/srep16553 (2015).
16. Liu, T., Zhang, X., Yang, Y. M., Du, L. T. & Wang, C. X. Increased expression of the long noncoding RNA CRNDE-h indicates a poor prognosis in colorectal cancer, and is positively correlated with IRX5 mRNA expression. *OncoTargets and therapy* **9,** 1437–1448, doi: 10.2147/OTT.S98268 (2016).
17. Han, P. *et al.* The lncRNA CRNDE promotes colorectal cancer cell proliferation and chemoresistance via miR-181a-5p-mediated regulation of Wnt/beta-catenin signaling. *Molecular cancer* **16,** 9, doi: 10.1186/s12943-017-0583-1 (2017).
18. Ohtsuka, M. *et al.* H19 Noncoding RNA, an Independent Prognostic Factor, Regulates Essential Rb-E2F and CDK8-beta-Catenin Signaling in Colorectal Cancer. *EBioMedicine* **13,** 113–124, doi: 10.1016/j.ebiom.2016.10.026 (2016).
19. Wu, K. F. *et al.* H19 mediates methotrexate resistance in colorectal cancer through activating Wnt/beta-catenin pathway. *Experimental cell research* **350,** 312–317, doi: 10.1016/j.yexcr.2016.12.003 (2017).

20. Jing, S. Y. *et al.* Expression of long non-coding RNA CRNDE in glioma and its correlation with tumor progression and patient survival. *European review for medical and pharmacological sciences* **20,** 3992–3996 (2016).

21. Zhang, T. *et al.* LncRNA H19 is overexpressed in glioma tissue, is negatively associated with patient survival, and promotes tumor growth through its derivative miR-675. *European review for medical and pharmacological sciences* **20,** 4891–4897 (2016).

22. Szafron, L. M. *et al.* The putative oncogene, CRNDE, is a negative prognostic factor in ovarian cancer patients. *Oncotarget* **6,** 43897–43910, doi: 10.18632/oncotarget.6016 (2015).

23. Zhu, Z. *et al.* Ectopic expressed long non-coding RNA H19 contributes to malignant cell behavior of ovarian cancer. *International journal of clinical and experimental pathology* **8,** 10082–10091 (2015).

24. Chen, Z. *et al.* LncRNA CRNDE promotes hepatic carcinoma cell proliferation, migration and invasion by suppressing miR-384. *American journal of cancer research* **6,** 2299–2309 (2016).

25. Yang, Z. *et al.* HULC and H19 Played Different Roles in Overall and Disease-Free Survival from Hepatocellular Carcinoma after Curative Hepatectomy: A Preliminary Analysis from Gene Expression Omnibus. *Disease markers* **2015,** 191029, doi: 10.1155/2015/191029 (2015).

26. Zhou, M. *et al.* Discovery and validation of immune-associated long non-coding RNA biomarkers associated with clinically molecular subtype and prognosis in diffuse large B cell lymphoma. *Molecular cancer* **16,** 16, doi: 10.1186/s12943-017-0580-4 (2017).

27. Zhou, M. *et al.* Construction and analysis of dysregulated lncRNA-associated ceRNA network identified novel lncRNA biomarkers for early diagnosis of human pancreatic cancer. *Oncotarget* **7,** 56383–56394, doi: 10.18632/oncotarget.10891 (2016).

28. Zhou, M. *et al.* Relapse-related long non-coding RNA signature to improve prognosis prediction of lung adenocarcinoma. *Oncotarget* **7,** 29720–29738, doi: 10.18632/oncotarget.8825 (2016).

29. Zhou, M. *et al.* Characterization of long non-coding RNA-associated ceRNA network to reveal potential prognostic lncRNA biomarkers in human ovarian cancer. *Oncotarget* **7,** 12598–12611, doi: 10.18632/oncotarget.7181 (2016).

30. Zhou, X., Liu, J. & Wang, W. Construction and investigation of breast-cancer-specific ceRNA network based on the mRNA and miRNA expression data. *IET systems biology* **8,** 96–103, doi: 10.1049/iet-syb.2013.0025 (2014).

31. Xia, T. *et al.* Long noncoding RNA associated-competing endogenous RNAs in gastric cancer. *Scientific reports* **4,** 6088, doi: 10.1038/srep06088 (2014).

32. Sala, D. & Zorzano, A. Is TP53INP2 a critical regulator of muscle mass? *Current opinion in clinical nutrition and metabolic care* **18,** 234–239, doi: 10.1097/MCO.0000000000000163 (2015).

33. Gan, Y. & Herzog, E. L. MMP28 and macrophage polarization: orchestrating the attack of the mac. *Journal of leukocyte biology* **95,** 1–3, doi: 10.1189/jlb.0813423 (2014).

34. Sharma Saha, S. *et al.* Identification of genetic variation in the lncRNA HOTAIR associated with HPV16-related cervical cancer pathogenesis. *Cellular oncology*, doi: 10.1007/s13402-016-0298-0 (2016).

35. She, K. *et al.* lncRNA-SNHG7 promotes the proliferation, migration and invasion and inhibits apoptosis of lung cancer cells by enhancing the FAIM2 expression. *Oncology reports* **36,** 2673–2680, doi: 10.3892/or.2016.5105 (2016).

36. Zhang, K. *et al.* Circulating lncRNA H19 in plasma as a novel biomarker for breast cancer. *Cancer biomarkers*: *section A of Disease markers* **17,** 187–194, doi: 10.3233/CBM-160630 (2016).

37. Lee, J. J., Kim, M. & Kim, H. P. Epigenetic regulation of long noncoding RNA UCA1 by SATB1 in breast cancer. *BMB reports* **49,** 578–583 (2016).

38. Hashad, D., Elbanna, A., Ibrahim, A. & Khedr, G. Evaluation of the Role of Circulating Long Non-Coding RNA H19 as a Promising Novel Biomarker in Plasma of Patients with Gastric Cancer. *Journal of clinical laboratory analysis* **30,** 1100–1105, doi: 10.1002/jcla.21987 (2016).

39. Shang, C., Guo, Y., Zhang, J. & Huang, B. Silence of long noncoding RNA UCA1 inhibits malignant proliferation and chemotherapy resistance to adriamycin in gastric cancer. *Cancer chemotherapy and pharmacology* **77,** 1061–1067, doi: 10.1007/s00280-016-3029-3 (2016).

40. Cui, J. *et al.* c-Myc-activated long non-coding RNA H19 downregulates miR-107 and promotes cell cycle progression of non-small cell lung cancer. *International journal of clinical and experimental pathology* **8,** 12400–12409 (2015).

41. Nie, W. *et al.* LncRNA-UCA1 exerts oncogenic functions in non-small cell lung cancer by targeting miR-193a-3p. *Cancer letters* **371,** 99–106, doi: 10.1016/j.canlet.2015.11.024 (2016).

42. Han, D. *et al.* Long noncoding RNA H19 indicates a poor prognosis of colorectal cancer and promotes tumor growth by recruiting and binding to eIF4A3. *Oncotarget* **7,** 22159–22173, doi: 10.18632/oncotarget.8063 (2016).

43. Bian, Z. *et al.* LncRNA-UCA1 enhances cell proliferation and 5-fluorouracil resistance in colorectal cancer by inhibiting miR-204-5p. *Scientific reports* **6,** 23892, doi: 10.1038/srep23892 (2016).

44. Zheng, J. *et al.* CRNDE affects the malignant biological characteristics of human glioma stem cells by negatively regulating miR-186. *Oncotarget* **6,** 25339–25355, doi: 10.18632/oncotarget.4509 (2015).

45. Wang, Y. *et al.* CRNDE, a long-noncoding RNA, promotes glioma cell growth and invasion through mTOR signaling. *Cancer letters* **367,** 122–128, doi: 10.1016/j.canlet.2015.03.027 (2015).

46. Wang, W. & Xing, C. Upregulation of long noncoding RNA ZFAS1 predicts poor prognosis and prompts invasion and metastasis in colorectal cancer. *Pathology, research and practice* **212,** 690–695, doi: 10.1016/j.prp.2016.05.003 (2016).

47. Thorenoor, N. *et al.* Long non-coding RNA ZFAS1 interacts with CDK1 and is involved in p53-dependent cell cycle control and apoptosis in colorectal cancer. *Oncotarget* **7,** 622–637, doi: 10.18632/oncotarget.5807 (2016).

48. Chen, X., Yan, C. C., Zhang, X. & You, Z. H. Long non-coding RNAs and complex diseases: from experimental results to computational models. *Briefings in bioinformatics*, doi: 10.1093/bib/bbw060 (2016).

49. Chen, X., You, Z. H., Yan, G. Y. & Gong, D. W. IRWRLDA: improved random walk with restart for lncRNA-disease association prediction. *Oncotarget* **7,** 57919–57931, doi: 10.18632/oncotarget.11141 (2016).

50. Chen, X. *et al.* HGIMDA: Heterogeneous graph inference for miRNA-disease association prediction. *Oncotarget* **7,** 65257–65269, doi: 10.18632/oncotarget.11251 (2016).

51. Chen, X., Huang, Y. A., Wang, X. S., You, Z. H. & Chan, K. C. FMLNCSIM: fuzzy measure-based lncRNA functional similarity calculation model. *Oncotarget* **7,** 45948–45958, doi: 10.18632/oncotarget.10008 (2016).

52. Chen, X. & Yan, G. Y. Novel human lncRNA-disease association inference based on lncRNA expression profiles. *Bioinformatics* **29,** 2617–2624, doi: 10.1093/bioinformatics/btt426 (2013).

53. Ciasca, G., Papi, M., Minelli, E., Palmieri, V. & De Spirito, M. Changes in cellular mechanical properties during onset or progression of colorectal cancer. *World journal of gastroenterology* **22,** 7203–7214, doi: 10.3748/wjg.v22.i32.7203 (2016).

54. Brown, D. G. *et al.* Metabolomics and metabolic pathway networks from human colorectal cancers, adjacent mucosa, and stool. *Cancer & metabolism* **4,** 11, doi: 10.1186/s40170-016-0151-y (2016).

55. Shi, K. Q. *et al.* Hepatocellular carcinoma associated microRNA expression signature: integrated bioinformatics analysis, experimental validation and clinical significance. *Oncotarget* **6,** 25093–25108, doi: 10.18632/oncotarget.4437 (2015).

56. Chen, D. *et al.* Genome-wide analysis of long noncoding RNA (lncRNA) expression in colorectal cancer tissues from patients with liver metastasis. *Cancer medicine* **5,** 1629–1639, doi: 10.1002/cam4.738 (2016).

57. Wang, C. *et al.* Long non-coding RNAs LOC285194, RP11-462C24.1 and Nbla12061 in serum provide a new approach for distinguishing patients with colorectal cancer from healthy controls. *Oncotarget*, doi: 10.18632/oncotarget.12220 (2016).

58. Weng, M. *et al.* Noncoding RNAs in the development, diagnosis, and prognosis of colorectal cancer. *Translational research*: *the journal of laboratory and clinical medicine*, doi: 10.1016/j.trsl.2016.10.001 (2016).

59. Vosa, U., Kolde, R., Vilo, J., Metspalu, A. & Annilo, T. Comprehensive meta-analysis of microRNA expression using a robust rank aggregation approach. *Methods in molecular biology* **1182,** 361–373, doi: 10.1007/978-1-4939-1062-5_28 (2014).
60. Kolde, R., Laur, S., Adler, P. & Vilo, J. Robust rank aggregation for gene list integration and meta-analysis. *Bioinformatics* **28,** 573–580, doi: 10.1093/bioinformatics/btr709 (2012).

Acknowledgements

Author Contributions

Shengming Dai and Jingfan Chen conceived and designed the experiments. Jinwan Li, Zhizhuo Huang, Shifeng Huang and Junyi Xu collected samples. Meiyu Dai, Xiaoli Chen and Shanying Mo performed the experiments. Meiyu Dai and Xiaoli Chen analyzed the data. Baoyu He and Yan Zou contributed reagents/materials/analysis tools. Meiyu Dai, Xiaoli Chen and Shanying Mo wrote the manuscript.

Additional Information

Supplementary information accompanies this paper at http://www.nature.com/srep

Competing Interests: The authors declare no competing financial interests.

How to cite this article: Dai, M. *et al.* Meta-signature LncRNAs serve as novel biomarkers for colorectal cancer: integrated bioinformatics analysis, experimental validation and diagnostic evaluation. *Sci. Rep.* **7**, 46572; doi: 10.1038/srep46572 (2017).

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.