

Identification of potential long non-coding RNA biomarkers associated with the progression of colon cancer

Jingwen Li^{1,*}, Weinan Xue^{1,*}, Junli Lv^{2,*}, Peng Han¹, Yanlong Liu¹ and Binbin Cui¹

¹Department of Colorectal Surgery, Harbin Medical University Cancer Hospital, Harbin 150081, China

²Department of Science, Harbin Medical University Cancer Hospital, Harbin 150081, China

*These authors contributed equally to this work

Correspondence to: Binbin Cui, **email:** binbincuihrb@126.com
Yanlong Liu, **email:** yanlliu@yahoo.com

Keywords: *biomarkers, colon cancer, disease progression, long non-coding RNA*

Received: March 29, 2017

Accepted: May 03, 2017

Published: May 17, 2017

Copyright: Li et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License 3.0 (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

ABSTRACT

Increasing evidence has suggested that dysregulated lncRNA expression played important roles in the development and progression of human cancers. Although prognostic roles of lncRNAs have been recognized for colon cancer (CC) patients, the search for novel lncRNA biomarkers potentially involved in CC progression is an urgent and still largely unmet medical need. In this study, we evaluated the lncRNA expression changes during the progression of CC by analyzing two cohorts of previously published expression profiles of CC patients and identified hundreds of differentially expressed lncRNAs. Then we identified eight lncRNAs that closely associated with the progression of CC patients from a large number of significantly altered lncRNAs using random forest supervised classification algorithm. Finally, an SVM-based lncRNA risk classifier was developed to discriminate high-risk CC patients from persons with early-stage and validated in both the training dataset and testing dataset by survival analysis and five-fold cross-validation strategy. Our pathway enrichment analysis based on protein-coding genes that are co-expressed with lncRNAs, suggested that variation in expression of eight lncRNAs biomarkers might affect critical pathways involved in CC progression. With further validation, these eight lncRNAs might have significant implications for the clinical management of CC patients with early stage and improve our understanding of cancer progression.

INTRODUCTION

Colon cancer (CC) is one of the most common types of human cancers and is the major cause of cancer-related death worldwide [1]. Surgery followed by adjuvant therapy (such as chemotherapy and radiation therapy) is the most common option for CC patients. However, the fact that early-stage CC patients with an initial treatment with surgery are known to still have a recurrence rate of 20%–30% and are likely to progress into malignancy [2]. Systemic treatment of early-stage patients with high-risk for disease progression attempts to prevent disease progression and improve patients' outcome. Therefore, it is an urgent

need to identify high-risk patients who are likely to progress into malignancy at an early stage of cancer development.

With recent advances in high-throughput technologies (such as RNA deep sequencing), transcriptomes of many organisms have been surveyed which identified thousands of long transcripts (> 200 bp) that have no significant protein-coding capacity and thus termed long noncoding RNAs (lncRNAs) [3]. Similar to mRNAs and microRNAs, increasing evidence has suggested that lncRNA are the key regulators of transcriptional and translational output and therefore make effects in many biological and pathological processes (such as dosage compensation, genomic

imprinting, cell differentiation and organogenesis) at the epigenetic level, transcriptional and post-transcriptional level [4, 5]. There is generally believed that lncRNAs are emerging as a novel hallmark of cancer beyond ten underlying principles shared by all cancers [6]. A large number of lncRNAs are found to be differentially expressed in various types of human cancer. Recently many studies have demonstrated their intriguing possibilities of application for diagnostics, prognostics and therapeutics in different cancers, including breast cancer, lung cancer, ovarian cancer, multiple myeloma and so on [7–23]. Although prognostic roles of lncRNAs have been recognized for CC patients, the search for novel lncRNA biomarkers potentially involved in CC progression is an urgent and still largely unmet medical need.

The aim of the present study was to evaluate lncRNA expression changes during the progression of CC by analyzing two cohorts of previously published expression profiles of CC patients, and tried to identify specific lncRNAs biomarkers closely associated with the progression of CC patients from a large number of significantly altered lncRNAs.

RESULTS

Identification of altered lncRNAs in the progression of LC

The GSE37892 patient dataset was chosen as a training dataset for the detection of lncRNA biomarkers. We first compared lncRNA expression profiles among patients with early-stage and those with advanced-stage in the training dataset, and identified 422 differentially expressed lncRNAs (FDR-adjusted $p \leq 0.01$) using the SAM analysis. Of them, 269 lncRNAs are up-regulated and 153 lncRNAs are down-regulated in patients with advanced-stage compared with early-stage patients.

Hierarchical clustering of these 422 lncRNAs based on centered Pearson correlation clearly separated advanced-stage patients from early-stage patients (Figure 1A). Only 33 patients (10 advanced-stage patients and 22 early-stage patients) were misclassified by the clustering analysis. Cluster 1 consisted of 61 patients, including 51 early-stage patients and 10 advanced-stage patients, whereas cluster 2 consisted of 69 patients, including 22 early-stage patients and 47 advanced-stage patients, which achieved a high prediction accuracy of 74.6%. The statistical result suggested that two patient clusters grouped by these 422 lncRNAs were significantly correlated with disease progression status ($p < 0.001$, Chi-square test). The Kaplan-Meier curves and log-rank test showed that metastasis-free survival (MFS) was significantly

different between these two patient clusters ($p = 0.044$, log-rank test) (Figure 1B).

Identification of optimal lncRNA biomarkers significantly associated with the progression of CC from the training dataset

In order to identify optimal lncRNA biomarkers for clinical use, a random forest supervised classification algorithm was used to narrow down the number of lncRNAs using these differentially expressed lncRNAs as features and their expression levels as feature values. Finally, eight lncRNAs mostly related to the prognostic classification were selected as optimal biomarkers among 422 differentially expressed lncRNAs according to the permutation important score (Table 1). Hierarchical clustering analysis with selected eight lncRNA biomarkers clearly separated the 130 patients of training dataset into two clusters (Figure 2A). Cluster 1 consisted of 80 patients, including 56 advanced-stage patients and 24 early-stage patients, whereas cluster 2 consisted of 50 patients, including 49 early-stage patients and only one advanced-stage patient, which achieved a high prediction accuracy of 80.8%. The Kaplan-Meier curves and log-rank test showed that MFS was significantly different between these two patient clusters grouped by these eight lncRNA biomarkers ($p < 0.001$, log-rank test) (Figure 2B). These results revealed the better predictive performance of eight lncRNA biomarkers for identifying patients at high risk of CC progression

SVM-based lncRNA risk classifiers distinguished advanced-stage and early-stage patients

Subsequently, these eight lncRNA biomarkers were integrated into a risk classifiers using SVM to identify patients at high risk for progression. SVM analysis with a 5-fold cross-validation procedure was performed to evaluate the predictive performance of the SVM-based lncRNA risk classifiers. Results of performance evaluation showed that the SVM-based lncRNA risk classifiers were able to correctly classify 107 out of 130 patients, achieving an overall predictive accuracy of 85.4% with a sensitivity of 89.5% and a specificity of 76.7%. The AUC of ROC analysis is 0.908 (Figure 3A). The Kaplan-Meier analysis demonstrated a significant difference in MFS between predicted early-stage-like group and advanced-stage-like group (Figure 3B). Patients in predicted early-stage-like group have significantly higher MFS than those in the predicted advanced-stage-like group ($p = 0.004$, log-rank test). The three-year and five-year MFS rates of the predicted advanced-stage-like group were 59.2% and 56.4%, respectively, whereas the corresponding rates in the predicted early-stage-like group were 86.3% and 81.5%, respectively.

Further validation of optimal lncRNA biomarkers in another independent testing dataset

To further confirm the predictive performance of optimal lncRNA biomarkers, eight lncRNA biomarkers were tested using another completely independent dataset of 55 patients from Smith's study. We first clustered CC

patients in the testing dataset according to the expression value of eight lncRNA biomarkers. Results with unsupervised hierarchical clustering analysis were similar to that observed in the training dataset. Two distinctive patient clusters were obtained by hierarchical clustering analysis (Figure 4A). Cluster 1 consisted of 24 patients, including 13 advanced-stage patients and 11 early-stage patients, whereas cluster 2 consisted of 31 patients,

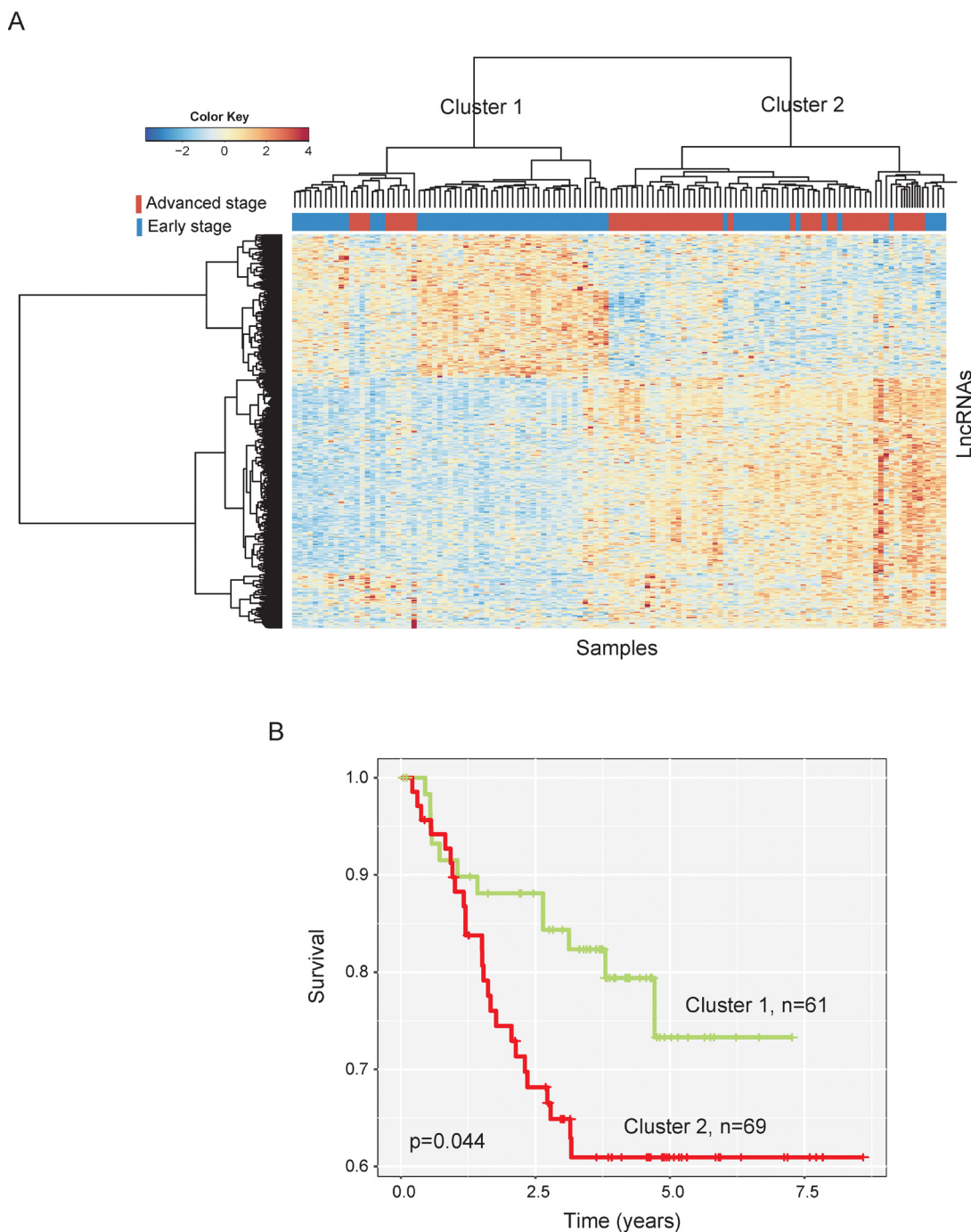


Figure 1: The heatmap and survival analysis of hierarchical clustering of 130 CC patients of the training dataset based on differentially expressed lncRNAs. (A) Heatmap of differentially expressed lncRNAs between early-stage and advanced-stage breast cancers. (B) Kaplan-Meier survival curves of metastasis-free survival between two clusters.

Table 1: Eight lncRNA biomarkers associated with the progression of colon cancer

Ensembl id	Gene name	Chromosomal location	FC	FDR
ENSG00000232560	LINC01549	Chr21: 17,438,890-17,449,185 (+)	0.827	0
ENSG00000232656	IDI2-AS1	Chr10: 1,022,666-1,044,201 (+)	0.75	0
ENSG00000214188	ST7-OT4	Chr7: 116,953,899-117,098,806 (+)	0.929	0.002
LOC100505942	LOC100505942	Chr16: 67,517,706-67,528,745(-)	1.157	0
ENSG00000244791	RP11-65D17.1	Chr8: 126,325,495-126,329,535(+)	0.862	0
ENSG00000228391	AC011995.3	Chr2: 2,870,558-2,871,231(-)	0.83	0
ENSG00000264859	DSG2-AS1	Chr 18: 31,542,146-31,556,911(-)	0.791	0
ENSG00000240567	RP11-3P17.4	Chr3: 161,426,427-161,448,242 (+)	0.469	0

including 8 early-stage patients and 23 advanced-stage patients, which achieved an overall prediction accuracy of 61.8%. Then the SVM-based lncRNA risk classifiers composing of eight lncRNA biomarkers in combination with 5-fold cross validation was validated in the testing dataset for determining patients at high-risk of disease progression. As expected, ROC analysis suggested that the AUC is 0.63 with a predictive accuracy of 63.6%, sensitivity of 65.4% and specificity of 33.3% (Figure 4B).

Functional analysis of eight lncRNA biomarkers

To better understand the functional roles of eight lncRNA biomarkers, we examined the expression correlation between their expression values and those of the mRNAs in the training dataset and identified protein-coding genes correlated with the eight lncRNA biomarkers. Finally, 1117 protein-coding genes were found to be positively correlated with at least one of the eight lncRNAs, and 1065 protein-coding genes were negatively correlated with at least one of the eight lncRNAs (top 1% as our criterion). Results of enrichment analysis suggested that 1117 protein coding genes positively correlated with lncRNAs mainly involved the following pathways: citrate cycle (TCA cycle), apoptosis, mRNA surveillance pathway, RIG-I-like receptor signaling pathway, spliceosome, Hepatitis C, AMPK signaling pathway and adipocytokine signaling pathway, and 1065 protein-coding genes negatively correlated with lncRNAs clustered most significantly in ECM-receptor interaction, focal adhesion, amoebiasis, pathways in cancer, protein digestion and absorption, renin secretion, notch signaling pathway, colorectal cancer, melanoma, PI3K-Akt signaling pathway, signaling pathways regulating pluripotency of stem cells and osteoclast differentiation (Figure 5).

DISCUSSION

Despite recent significant advances in treatment and management of colon cancer patients, cancer recurrence following initial treatment with surgery remains a considerable problem. Almost half of all colorectal cancer patients will develop recurrent disease [24]. For those

patients with stage III, the recurrence rate can exceed 50% and adjuvant chemotherapy has been shown to significantly reduce the risk of recurrence. Despite patients with stage I and II have more favorable outcomes with survival rates of 75% to 95% at 5 years, some patients with stage II faced an increased risk of recurrence approaching that of stage III colon cancer patients and chemotherapy is advised for those patients with stage II disease and reduce the risk of disease recurrence [25, 26]. However, adjuvant therapy, such as chemotherapy, has a wide range of side effects that substantially affect the patients' life quality. Therefore, it is an urgent need to identify patients with early-stage colon cancer at high risk who might benefit from more aggressive therapy. Current clinic-pathologic information, such as TNM staging system, is inadequate for recurrence prediction of colon cancer patients with early-stage [25]. With the development in high-throughput gene expression and sequencing technologies, molecular profiles have been shown to be a powerful tool in diagnosis and prognosis of colon cancer, such as mRNAs and miRNAs [27, 28]. A recent study of ncRNAs has identified lncRNAs as a novel ncRNA class [3]. It is now evident that dysregulation of lncRNAs has been observed in the development and progression of many cancers, including colon cancer [10, 29–31]. Therefore, we attempt to identify potential lncRNA biomarkers of CC progression which could help determine patient with early-stage at the high-risk of disease recurrence and would allow selecting those patients for more aggressive treatment.

In this study, we performed a comprehensive analysis of lncRNA expression profiles of early-stage and advanced-stage CC patients and found altered lncRNA expression pattern during the progression of CC. Comparative analysis of lncRNA expression alternations between early-stage and advanced-stage CC patients improves our understanding of CC progression. Previous studies often focused on mRNAs and miRNAs, and identified altered mRNA/miRNA expression patterns during the progression of CC [27, 28, 32–35]. However, it is shown that lncRNAs are typically expressed in more cell-type-, and tissue-specific manner than mRNAs or miRNAs, thus having great advantages and priorities

as diagnostic and prognostic markers [36]. Although lncRNA expression profiles have been investigated between CC patients and normal tissues and between cancer subtypes [37, 38], this is the first report that indicates the existence of stage-specific lncRNA expression pattern between early-stage and advanced-stage CC patients. Then we used bioinformatics methods to identify eight lncRNAs from a large number of significantly altered lncRNAs as potential biomarkers of CC progression. By performing hierarchical clustering, survival analysis and 5-fold cross-validation strategy, the predictive value of these eight lncRNAs was validated in the training dataset and testing dataset.

Although more and more lncRNAs have been discovered during the past years, a small fraction of lncRNAs has been functionally characterized. Functional study of individual lncRNAs still remains challenging [39]. Since current knowledge suggests that lncRNAs function by regulating or interacting with its partner molecules, identifying the function of protein-coding genes that are co-expressed with lncRNAs has been shown to be an effective way to better characterize the potential functions of novel lncRNAs. Our pathway enrichment analysis based on protein-coding genes that are co-expressed with lncRNAs, suggested that variation in expression of eight lncRNAs biomarkers might affect

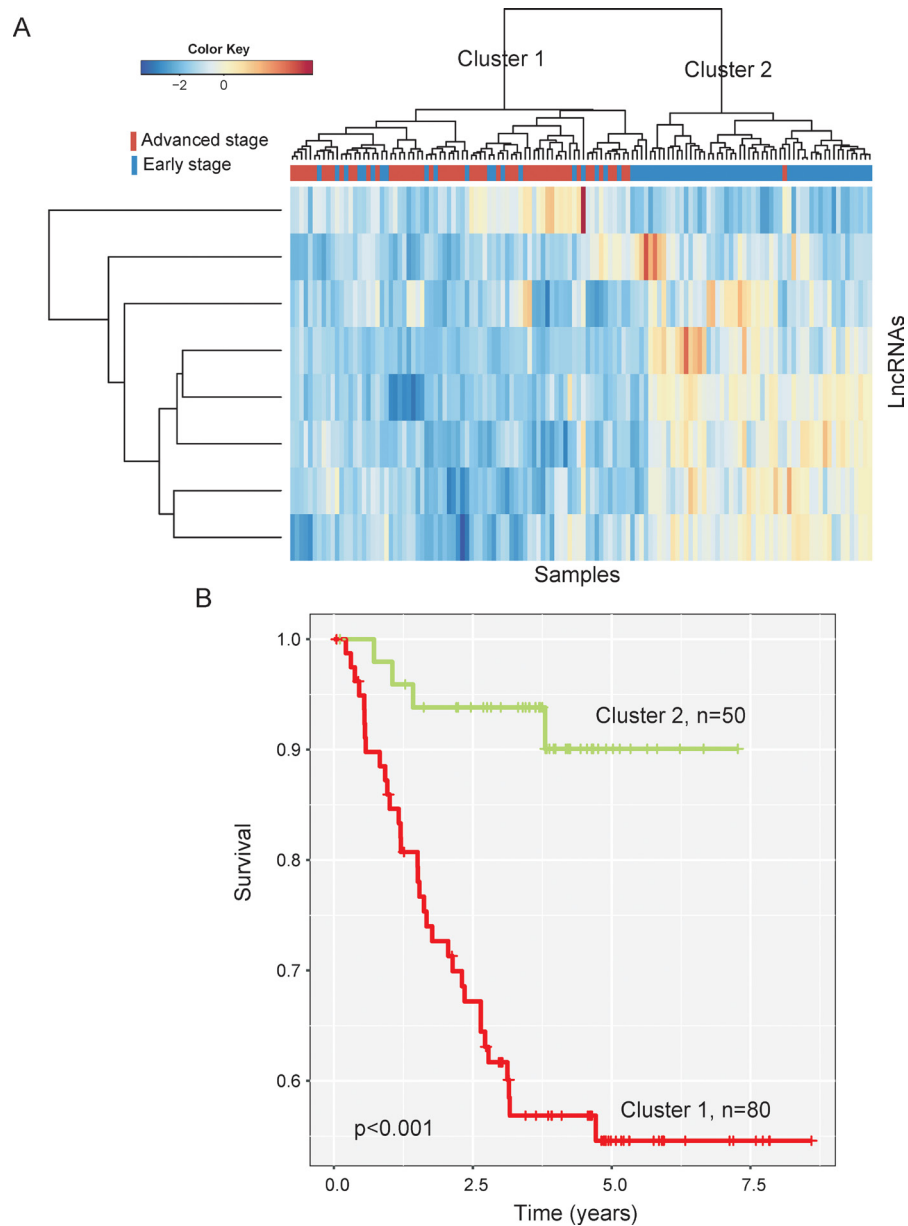


Figure 2: The heatmap and survival analysis of hierarchical clustering of 130 CC patients of the training dataset based on selected eight lncRNA biomarkers. (A) Heatmap of selected eight lncRNA biomarkers. (B) Kaplan-Meier survival curves of metastasis-free survival between two clusters.

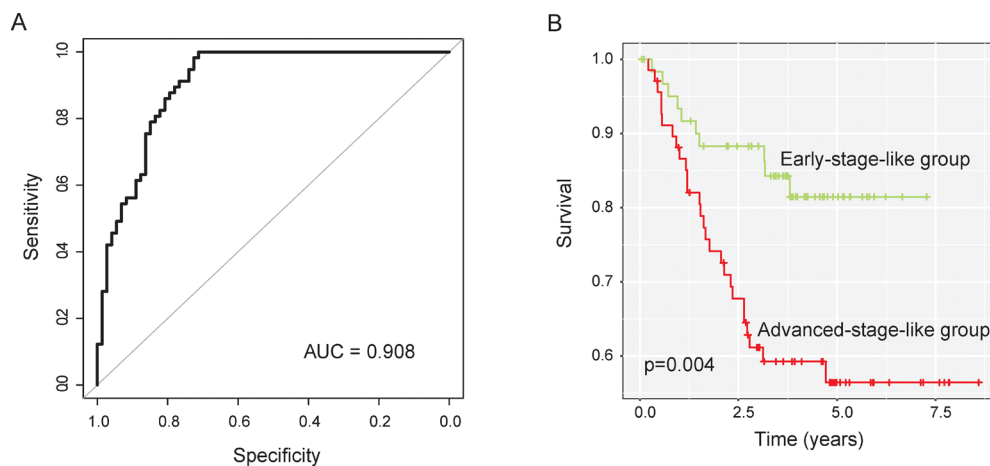


Figure 3: Performance evaluation of SVM-based lncRNA risk classifiers in the training dataset. (A) Receiver operating characteristic (ROC) curves for SVM-based lncRNA risk classifiers in distinguishing advanced-stage and early-stage patients (B) Kaplan-Meier survival curves of metastasis-free survival between two early-stage-like group and advanced-stage-like group.

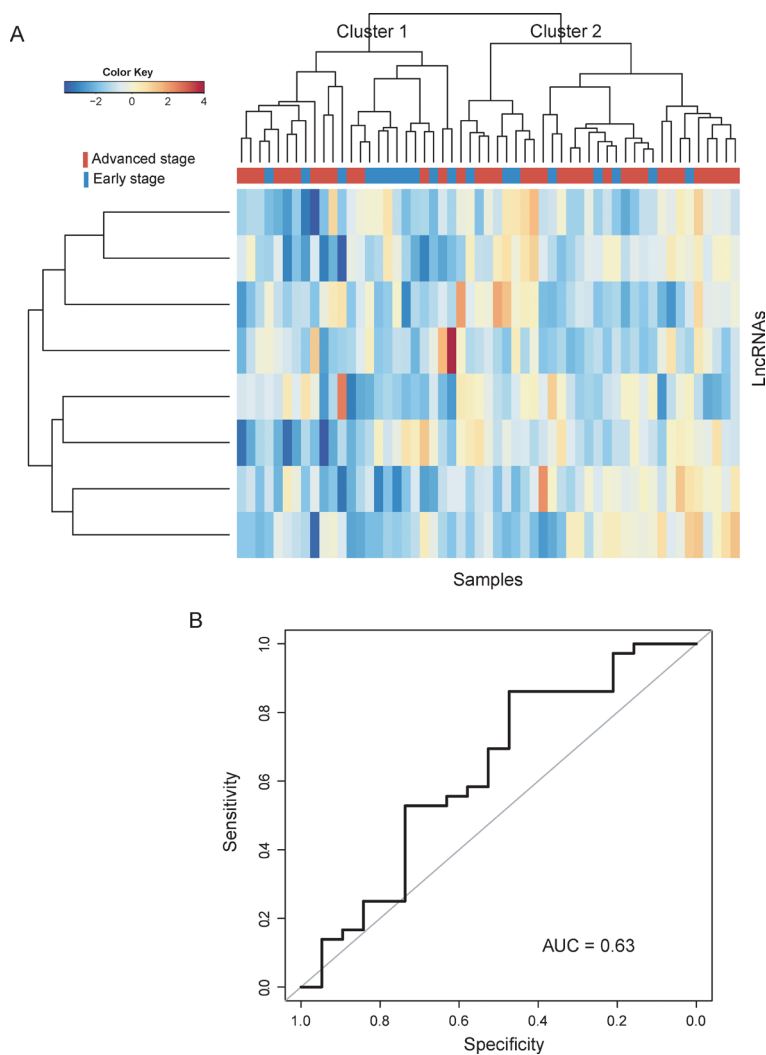


Figure 4: Performance validation of SVM-based lncRNA risk classifiers in the testing dataset. (A) Heatmap of selected eight lncRNA biomarkers. (B) Receiver operating characteristic (ROC) curves for SVM-based lncRNA risk classifiers in distinguishing advanced-stage and early-stage patients.

critical pathways involved in CC progression. For example, apoptosis-related genes were associated with high disease recurrence rates and have been used to identify stage II and III colon cancer patients with high risk of recurrence [40]. Perturbations of the AKT signaling pathway have been implicated in human cancers, including colon cancer. Overexpression of AKT was observed as an early event during colon tumorigenesis [41]. Focal adhesion kinase (FAK) expression has been correlated with worse prognosis in several tumors and may be involved in cancer radio- and chemosensitivity [42]. A component of focal adhesion, TRIM15, has been found to function as a tumor suppressor in colon cancer [43]. It is well known that Notch signaling is an important molecular pathway involved in the determination of cell fate. Genes of Notch signaling pathway such as Notch1, Notch2, HES1, DLL1, and JAG1 have been reported to be associated with the pathological tumor characteristics and degree of differentiation in colorectal cancer [44, 45].

In conclusion, our study has shown that the lncRNA expression pattern is altered in advanced-stage CC patients compared with early-stage CC patients. We identified eight lncRNAs as potential biomarkers capable of identifying patients with early-stage at high risk for progression to advanced-stage using differentially expressed analysis and random forest supervised classification algorithm methods. Finally, an SVM-based lncRNA risk classifier was developed to discriminate high-risk CC patients from persons with early-stage and

validated in both the training dataset and testing dataset. With further validation, these eight lncRNAs might have important implications for the clinical management of CC patients with early-stage.

MATERIALS AND METHODS

Patient datasets

Two patient datasets of colon cancer were collected from the Gene Expression Omnibus (GEO) database (<https://www.ncbi.nlm.nih.gov/geo/>). A total of 185 patients with CC was used in this study, including 130 CC patients (including 73 stage-II patients and 57 stage-III patients) in GSE37892 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE37892>) from Laibe's study and 55 CC patients (including 4 stage-I, 15 stage-II, 19 stage-III and 17 stage-IV patients) in GSE17537 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE17537>) from Smith's study.

Acquisition and analysis of lncRNA expression profiles

Raw gene expression data profiled from Affymetrix Human Genome U133 Plus 2.0 Array (HG-U133_Plus_2.0) in the three patient cohorts were processed and normalized using the Robust Multichip Average (RMA) algorithm for background adjustment and log-

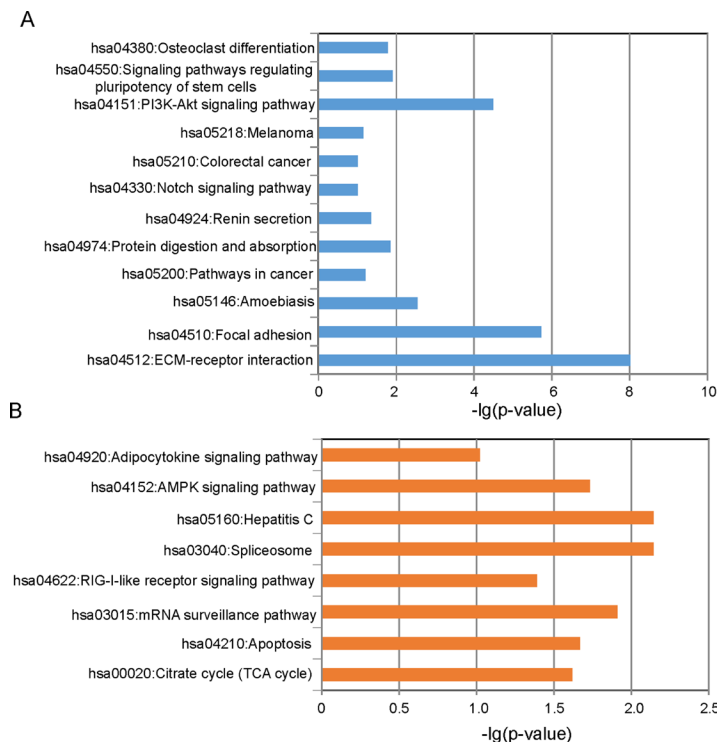


Figure 5: Functional enrichment analysis. (A) Enriched KEGG pathways of protein coding genes positively correlated with lncRNAs. (B) Enriched KEGG pathways of protein coding genes negatively correlated with lncRNAs.

transformed (base 2). LncRNA expression profiles of CC patients in this study were obtained by repurposing the probes in the HG-U133_Plus_2.0 array according to Zhang's study [46]. Briefly, probes (probe sets) representing lncRNAs were obtained by remapping their RefSeq IDs and Ensembl IDs to the annotation of lncRNAs from GENCODE.

Significance analysis of microarrays (SAM) was applied to identify differentially expressed lncRNAs between patients with early-stage (stage I/II) and those with advanced-stage (stage III/IV). Those lncRNAs with $FDR \leq 0.01$ (Benjamini and Hochberg's multiple-test adjustment) from SAM analysis were identified as differentially expressed lncRNAs. Unsupervised hierarchical clustering was used to investigate the effectiveness of lncRNA biomarkers in distinguishing early-stage and advanced-stage patients, and the Chi-square test was used to test the significance of the association between tumor status and lncRNA biomarkers.

Statistical analysis

To obtain optimal lncRNA biomarkers significantly associated with tumor progression, a random forest supervised classification algorithm was used to narrow down the number of lncRNAs by several iterative steps, in which one-third of the least important lncRNAs were discarded at each step according to their importance score using R packages "randomForest" as previously described [47]. Finally, optimal lncRNA biomarkers were integrated to construct a predictive risk classifier using support vector machine (SVM) with the sigmoid kernel. An unbiased performance estimate of SVM-based lncRNA risk classifier in distinguishing early-stage and advanced-stage patients was evaluated using 5-fold cross-validation strategy followed by establishing receiver operation characteristic (ROC) curve. The area under the receiver operating curve (AUC) was produced by plotting true positive rates (sensitivity) against false positive rates (1-specificity) to indicate the prediction performance. Kaplan-Meier survival curves and log-rank tests were used to assess the differences in survival time between the predicted early-stage-like group and advanced-stage-like patients using R packages "survival".

Functional enrichment analysis

Functional enrichment analysis of Kyoto encyclopedia of genes and genomes (KEGG) for mRNAs co-expressed with lncRNA biomarkers was performed to infer potential biological pathways of lncRNA biomarkers using DAVID Bioinformatics Tool (version 6.7) [48] limited to KEGG pathway categories. The biological pathways with a p -value of < 0.1 and an enrichment score of > 2.0 using the whole human genome as background were considered as a significantly enriched pathway.

ACKNOWLEDGMENTS AND FUNDING

This work was supported by the National Natural Science Foundation of China (Grant No. 81402367), Science & technology Bureau of Harbin (Grant No. 2014RFQJ and 2015RAXYJ063), Hospital elite fund (Grant No. ZY2015-06).

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

REFERENCES

1. Jemal A, Center MM, DeSantis C, Ward EM. Global patterns of cancer incidence and mortality rates and trends. *Cancer Epidemiol Biomarkers Prev.* 2010; 19:1893–1907.
2. Markle B, May EJ, Majumdar AP. Do nutraceuticals play a role in the prevention and treatment of colorectal cancer? *Cancer Metastasis Rev.* 2010; 29:395–404.
3. Moran VA, Perera RJ, Khalil AM. Emerging functional and mechanistic paradigms of mammalian long non-coding RNAs. *Nucleic Acids Res.* 2012; 40:6391–6400.
4. Cao J. The functional role of long non-coding RNAs and epigenetics. *Biol Proced Online.* 2014; 16:11.
5. Mercer TR, Dinger ME, Mattick JS. Long non-coding RNAs: insights into functions. *Nat Rev Genet.* 2009; 10:155–159.
6. Gutschner T, Diederichs S. The hallmarks of cancer: a long non-coding RNA point of view. *RNA Biol.* 2012; 9:703–719.
7. Zhou M, Zhao H, Xu W, Bao S, Cheng L, Sun J. Discovery and validation of immune-associated long non-coding RNA biomarkers associated with clinically molecular subtype and prognosis in diffuse large B cell lymphoma. *Mol Cancer.* 2017; 16:16.
8. Zhou M, Zhong L, Xu W, Sun Y, Zhang Z, Zhao H, Yang L, Sun J. Discovery of potential prognostic long non-coding RNA biomarkers for predicting the risk of tumor recurrence of breast cancer patients. *Sci Rep.* 2016; 6:31038.
9. Zhou M, Xu W, Yue X, Zhao H, Wang Z, Shi H, Cheng L, Sun J. Relapse-related long non-coding RNA signature to improve prognosis prediction of lung adenocarcinoma. *Oncotarget.* 2016; 7:29720–38. <https://doi.org/10.18632/oncotarget.8825>.
10. Zhang F, Ren C, Zhao H, Yang L, Su F, Zhou MM, Han J, Sobie EA, Walsh MJ. Identification of novel prognostic indicators for triple-negative breast cancer patients through integrative analysis of cancer genomics data and protein interactome data. *Oncotarget.* 2016; 7:71620–34. <https://doi.org/10.18632/oncotarget.12287>.
11. Zhou M, Sun Y, Sun Y, Xu W, Zhang Z, Zhao H, Zhong Z, Sun J. Comprehensive analysis of lncRNA expression profiles reveals a novel lncRNA signature to discriminate nonequivalent outcomes in patients with ovarian cancer. *Oncotarget.* 2016; 7:32433–48. <https://doi.org/10.18632/oncotarget.8653>.

12. Zhou M, Diao Z, Yue X, Chen Y, Zhao H, Cheng L, Sun J. Construction and analysis of dysregulated lncRNA-associated ceRNA network identified novel lncRNA biomarkers for early diagnosis of human pancreatic cancer. *Oncotarget*. 2016; 7:56383–94. <https://doi.org/10.18632/oncotarget.10891>.
13. Sun J, Cheng L, Shi H, Zhang Z, Zhao H, Wang Z, Zhou M. A potential panel of six-long non-coding RNA signature to improve survival prediction of diffuse large-B-cell lymphoma. *Sci Rep*. 2016; 6:27842.
14. Zhou M, Zhang Z, Zhao H, Bao S, Cheng L, Sun J. An immune-related six-lncRNA signature to improve prognosis prediction of glioblastoma multiforme. *Mol Neurobiol*. 2017 May 19. [Epub ahead of print]. <https://doi.org/10.1007/s12035-017-0572-9>.
15. Zhou M, Zhao H, Wang Z, Cheng L, Yang L, Shi H, Yang H, Sun J. Identification and validation of potential prognostic lncRNA biomarkers for predicting survival in patients with multiple myeloma. *J Exp Clin Cancer Res*. 2015; 34:102.
16. Zhou M, Guo M, He D, Wang X, Cui Y, Yang H, Hao D, Sun J. A potential signature of eight long non-coding RNAs predicts survival in patients with non-small cell lung cancer. *J Transl Med*. 2015; 13:231.
17. Sun J, Chen X, Wang Z, Guo M, Shi H, Wang X, Cheng L, Zhou M. A potential prognostic long non-coding RNA signature to predict metastasis-free survival of breast cancer patients. *Sci Rep*. 2015; 5:16553.
18. Zeng JH, Liang L, He RQ, Tang RX, Cai XY, Chen JQ, Luo DZ, Chen G. Comprehensive investigation of a novel differentially expressed lncRNA expression profile signature to assess the survival of patients with colorectal adenocarcinoma. *Oncotarget*. 2017; 8:16811–28. <https://doi.org/10.18632/oncotarget.15161>.
19. Wang W, Yang F, Zhang L, Chen J, Zhao Z, Wang H, Wu F, Liang T, Yan X, Li J, Lan Q, Wang J, Zhao J. LncRNA profile study reveals four-lncRNA signature associated with the prognosis of patients with anaplastic gliomas. *Oncotarget*. 2016; 7:77225–36. <https://doi.org/10.18632/oncotarget.12624>.
20. Li J, Chen Z, Tian L, Zhou C, He MY, Gao Y, Wang S, Zhou F, Shi S, Feng X, Sun N, Liu Z, Skogerboe G, et al. LncRNA profile study reveals a three-lncRNA signature associated with the survival of patients with oesophageal squamous cell carcinoma. *Gut*. 2014; 63:1700–1710.
21. Fan ZY, Liu W, Yan C, Zhu ZL, Xu W, Li JF, Su L, Li C, Zhu ZG, Liu B, Yan M. Identification of a five-lncRNA signature for the diagnosis and prognosis of gastric cancer. *Tumour Biol*. 2016; 37:13265–13277.
22. Cao W, Liu JN, Liu Z, Wang X, Han ZG, Ji T, Chen WT, Zou X. A three-lncRNA signature derived from the Atlas of ncRNA in cancer (TANRIC) database predicts the survival of patients with head and neck squamous cell carcinoma. *Oral Oncol*. 2017; 65:94–101.
23. Bao Z, Zhang W, Dong D. A potential prognostic lncRNA signature for predicting survival in patients with bladder urothelial carcinoma. *Oncotarget*. 2017; 8:10485–97. <https://doi.org/10.18632/oncotarget.14441>.
24. Young PE, Womeldorph CM, Johnson EK, Maykel JA, Brucher B, Stojadinovic A, Avital I, Nissan A, Steele SR. Early detection of colorectal cancer recurrence in patients undergoing surgery with curative intent: current status and challenges. *J Cancer*. 2014; 5:262–271.
25. Tsikitis VL, Larson DW, Huebner M, Lohse CM, Thompson PA. Predictors of recurrence free survival for patients with stage II and III colon cancer. *BMC Cancer*. 2014; 14:336.
26. van Eeghen EE, Bakker SD, van Bochove A, Loffeld RJ. High Risk Stage 2 and Stage 3 Colon Cancer, Predictors of Recurrence and Effect of Adjuvant Therapy in a Nonselected Population. *Int Sch Res Notices*. 2015; 2015:790186.
27. Huerta S. Recent advances in the molecular diagnosis and prognosis of colorectal cancer. *Expert Rev Mol Diagn*. 2008; 8:277–288.
28. Slaby O, Svoboda M, Michalek J, Vyzula R. MicroRNAs in colorectal cancer: translation of molecular biology into clinical application. *Mol Cancer*. 2009; 8:102.
29. Xu MD, Qi P, Du X. Long non-coding RNAs in colorectal cancer: implications for pathogenesis and clinical application. *Mod Pathol*. 2014; 27:1310–1320.
30. Luo J, Xu L, Jiang Y, Zhuo D, Zhang S, Wu L, Xu H, Huang Y. Expression profile of long non-coding RNAs in colorectal cancer: A microarray analysis. *Oncol Rep*. 2016; 35:2035–2044.
31. Chen X, Liu B, Yang R, Guo Y, Li F, Wang L, Hu H. Integrated analysis of long non-coding RNAs in human colorectal cancer. *Oncotarget*. 2016; 7:23897–908. <https://doi.org/10.18632/oncotarget.8192>.
32. Shivapurkar N, Weiner LM, Marshall JL, Madhavan S, Deslattes Mays A, Juhl H, Wellstein A. Recurrence of early stage colon cancer predicted by expression pattern of circulating microRNAs. *PLoS One*. 2014; 9:e84686.
33. Marisa L, de Reynies A, Duval A, Selves J, Gaub MP, Vescovo L, Etienne-Grimaldi MC, Schiappa R, Guenot D, Ayadi M, Kirzin S, Chazal M, Flejou JF, et al. Gene expression classification of colon cancer into molecular subtypes: characterization, validation, and prognostic value. *PLoS Med*. 2013; 10:e1001453.
34. Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. *Nature*. 2012; 487:330–37.
35. Hofslis E, Sjursen W, Prestvik WS, Johansen J, Rye M, Trano G, Wasmuth HH, Hatlevoll I, Thommesen L. Identification of serum microRNA profiles in colon cancer. *Br J Cancer*. 2013; 108:1712–1719.
36. Cheetham SW, Gruhl F, Mattick JS, Dinger ME. Long noncoding RNAs and the genetics of cancer. *Br J Cancer*. 2013; 108:2419–2425.
37. Chen H, Xu J, Hong J, Tang R, Zhang X, Fang JY. Long noncoding RNA profiles identify five distinct molecular subtypes of colorectal cancer with clinical relevance. *Mol Oncol*. 2014; 8:1393–1403.

38. Yang L, Xu L, Wang Q, Wang M, An G. Dysregulation of long non-coding RNA profiles in human colorectal cancer and its association with overall survival. *Oncol Lett.* 2016; 12:4068–4074.
39. Da Sacco L, Baldassarre A, Masotti A. Bioinformatics tools and novel challenges in long non-coding RNAs (lncRNAs) functional analysis. *Int J Mol Sci.* 2012; 13:97–114.
40. Belt EJ, Stockmann HB, Delis-van Diemen PM, Bril H, Tijssen M, van Essen HF, Heymans MW, Belien JA, Carvalho B, Cillessen SA, Meijer GA. Expression of apoptosis regulating proteins identifies stage II and III colon cancer patients with high risk of recurrence. *J Surg Oncol.* 2014; 109:255–265.
41. Altomare DA, Testa JR. Perturbations of the AKT signaling pathway in human cancer. *Oncogene.* 2005; 24:7455–7464.
42. Gomez Del Pulgar T, Cebrian A, Fernandez-Acenero MJ, Borrero-Palacios A, Del Puerto-Nevaldo L, Martinez-Useros J, Marin-Arango JP, Carames C, Vega-Bravo R, Rodriguez-Remirez M, Cruz-Ramos M, Manzarbeitia F, Garcia-Foncillas J. Focal adhesion kinase: predictor of tumour response and risk factor for recurrence after neoadjuvant chemoradiation in rectal cancer. *J Cell Mol Med.* 2016; 20:1729–1736.
43. Lee OH, Lee J, Lee KH, Woo YM, Kang JH, Yoon HG, Bae SK, Songyang Z, Oh SH, Choi Y. Role of the focal adhesion protein TRIM15 in colon cancer development. *Biochim Biophys Acta.* 2015; 1853:409–421.
44. Qiao L, Wong BC. Role of Notch signaling in colorectal cancer. *Carcinogenesis.* 2009; 30:1979–1986.
45. Jin HY, Zhang HY, Wang X, Xu J, Ding Y. Expression and clinical significance of Notch signaling genes in colorectal cancer. *Tumour Biol.* 2012; 33:817–824.
46. Zhang X, Sun S, Pu JK, Tsang AC, Lee D, Man VO, Lui WM, Wong ST, Leung GK. Long non-coding RNA expression profiles predict clinical phenotypes in glioma. *Neurobiol Dis.* 2012; 48:1–8.
47. Sorensen KP, Thomassen M, Tan Q, Bak M, Cold S, Burton M, Larsen MJ, Kruse TA. Long non-coding RNA expression profiles predict metastasis in lymph node-negative breast cancer independently of traditional prognostic markers. *Breast Cancer Res.* 2015; 17:55.
48. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc.* 2009; 4:44–57.