

RESEARCH

Open Access



Establishment and validation of a prognostic model based on HRR-related lncRNAs in colon adenocarcinoma

Xingkui Tang^{1*†}, Yukun Lin^{2*†}, Jialin He¹, Xijun Luo¹, Junjie Liang¹, Xianjun Zhu¹ and Tao Li¹

Abstract

Background: Colon cancer (CRC) is the second leading cause of cancer-related death, and its 5-year survival rate is very low. Homologous recombination repair (HRR) is deficient in most colon cancer. Some long non-coding RNAs (lncRNAs) participate in tumorigenesis of colon cancer through the HRR pathway. We aim to establish a prognostic model based on the HRR-related lncRNAs, expecting to provide a new strategy for precision treatment development in colon cancer.

Methods: Pearson's correlation was used to identify the HRR-related prognostic lncRNAs in the TCGA-COAD cohort. The TCGA-COAD cohort was randomized into the training set and the testing set. LASSO Cox regression was used to establish the model which was analyzed in the training set and validated in the testing set and the entire TCGA-COAD cohort. Finally, we explored the potential biological function of our model.

Results: A prognostic model was established based on nineteen HRR-related lncRNAs in the training set. COAD patients were scored by the uniform formula and divided into high-risk and low-risk groups based on the median risk score. Patients with high-risk scores indicated poor prognosis in the training set, and the result was confirmed in the testing set and the entire TCGA-COAD cohort (all $p < 0.01$). Multivariable analysis suggested that our model was an independent factor for overall survival in COAD. The area under the curve (AUC) and C-index indicated that our model had better predictive efficiency than other indicators in the TCGA-COAD cohort. Functional enrichment analysis suggested that our model was associated with the MAPK pathway in COAD. Besides, our model was positively correlated with the HRD scores.

Conclusion: A new prognostic model was established based on nineteen HRR-related lncRNAs which had excellent predictive efficiency on the prognosis of COAD. This prognostic model may provide a new strategy for prognostic prediction of COAD patients.

Keywords: Colon cancer, Homologous combination repair, Long non-coding RNAs, Prognostic model, HRD score

Background

Colon cancer is the most common digestive tract malignancy, and colon adenocarcinoma (COAD) is the most frequent subtype of colon cancer. As the second leading cause of cancer-related death, the occurrence of colon cancer in patients aged less than 50 years keeps increasing [1]. Despite the achievements in shifting patterns of risk factors, cancer prevention, earlier diagnosis through screening, and better treatment modalities of

*Correspondence: kxtang2021@126.com; linyukun@mail.sysu.edu.cn

[†]Xingkui Tang and Yukun Lin have contributed equally to this work and share the first authorship.

¹ Department of General Surgery, Guangzhou Panyu Central Hospital, Guangzhou, China

² Zhongshan School of Medicine, Sun Yat-sen University, Guangzhou 510080, China



colon cancer [2], its morbidity is rapidly increasing and its 5-year survival rate is low [3]. Hence, it is essential to identify new predictors for prognosis in colon cancer and establish a prognostic stratification to improve adjuvant treatment selection helping patients avoid unnecessary toxicities without compromising outcomes.

The homologous recombination repair (HRR) pathway plays an important role in repairing highly cytotoxic double-stranded DNA breaks (DSB) and restarting stalled replication forks [4]. Numerous studies have investigated that harbor defects in components of the HRR pathway were an important cause of cancer incidence. However, cancer patients with DNA repair defects could benefit from treatments inducing double-stranded DNA breaks or replication fork arrest [5, 6]. Therefore, the deficiency of the HRR pathway in cancer has aroused the interest of researchers. In colon cancer, HRR is deficient in greater than 78% of tumors in an examination of a large cohort [7]. A study indicated that the mutations of HRR-related genes significantly increased the immune activities of COAD patients with microsatellite-stable (MSS), implying the feasibility of the HRR-mutation status as an immunotherapy response predictor in MSS COAD [8]. However, the function of HRR-related genes in colon cancer still needs to be further investigated.

Long non-coding RNA (lncRNA) is a module of RNA that is longer than 200nt in length and has limited or no capability of protein coding. Increasing studies have proved that lncRNAs were not artifacts of pervasive transcription of “junk DNA”; on the contrary, they played a crucial role in cell cycle controlling which was linked to cancer progression [9, 10]. Numerous lncRNAs have been reported to participate in the regulation of DSB repair by being involved in HRR [11]. Researchers have revealed that lncRNAs modulate DSB repair by five mechanisms [12]. Firstly, lncRNAs can modulate the transcription, translation, or posttranslational modifications of p53 resulting in modulating the activity of p53. Secondly, chromatin remodeling complexes can be recruited to the site of damage by lncRNAs. Thirdly, lncRNAs can act as decoys to sequester the negative regulators of DNA repair away from the damage sites. Besides, lncRNAs can interact with DNA repair proteins directly by acting as scaffolds. Finally, lncRNAs can also regulate the stability of DNA repair proteins by sponging miRNAs. Therefore, dysfunction of lncRNAs that participate in HRR would lead to pathological conditions such as cancer. In colon cancer, lncRNA HITT could attenuate ATM activation and HRR resulting in sensitized genotoxic treatment [13]. However, the function of a large number of HRR-related lncRNAs in colon cancer has not been elucidated.

Hence, in this study, we identified the prognostic HRR-related lncRNAs in COAD and established a prognostic

model based on the prognostic HRR-related lncRNAs by using the least absolute shrinkage and selection operator (LASSO). Besides, we discussed the biological function of our model by using functional enrichment analysis, expecting to help the development of therapy in COAD.

Methods

Data collection and identification of HRR-related prognostic lncRNAs

The transcriptome profile and corresponding clinical information of 438 COAD patients were downloaded from the TCGA database (<https://cancergenome.nih.gov/>). The forty-one HRR-related genes were obtained from the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (<https://www.kegg.jp/kegg/>), including SSBP1, RAD50, MRE11, NBN, ATM, BRCA1, BARD1, RBBP8, BRIP1, TOPBP1, ABRAXAS1, UIMC1, BABAM1, BABAM2, BRCC3, PALB2, BRCA2, SEM1, SYCP3, RPA1, RPA2, RPA3, RPA4, RAD51, RAD52, RAD51B, RAD51C, RAD51D, XRCC2, XRCC3, RAD54L, RAD54B, POLD1, POLD2, POLD3, POLD4, BLM, TOP3A, TOP3B, MUS81, and EME1. Pearson's correlation analysis was used to identify the HRR-related lncRNAs that significantly related to greater than or equal to one of the 41 HRR genes ($|\text{Pearson } R| > 0.3$ and $p < 0.01$). Univariable analysis was applied to screen the prognostic HRR-related lncRNAs with p -value cutoff = 0.05. Finally, we identified 33 HRR-related lncRNAs that had prognostic value in the TCGA-COAD cohort.

Establishment and validation of the prognostic model

Four hundred and thirty-eight COAD patients in the TCGA cohort were divided into two sets (training set and testing set) randomly. The training set was used to establish the prognostic model. LASSO Cox regression analysis [14, 15] was conducted to establish the prognostic model based on the prognostic HRR-related lncRNAs by using the R package “glmnet.” The following formula was used to calculate the risk score:

$$\text{risk score} = \sum_{i=1}^n \text{coef}(\text{lncRNA}_i) \times \text{expr}(\text{lncRNA}_i)$$

To validate the predictive effect of the model, the testing set and the entire TCGA cohort were used as internal validation. All patients were scored using the unified formula and the median score was used to divide the patients into high-risk and low-risk groups. Kaplan–Meier survival analysis and log-rank test were used to evaluate the prognostic relevance of the prognostic model. The receiver operating characteristic (ROC) curve and the concordance index (C-index) were used to evaluate the predictive effect of risk scores on the prognostic of COAD patients in the TCGA cohort.

To test whether the prognostic value of our model can be affected by clinical features such as age, gender, tumor grade, and TNM stage, COAD patients in the TCGA cohort were stratified by age (> 60 years and ≤ 60 years), gender (male and female), tumor grade (stages I–II and stages III–IV), and TNM stage (T1–T2 and T3–T4). Kaplan–Meier survival analysis was used to draw the survival curves, and the significant difference of OS between high-risk and low-risk groups in each subset was evaluated by log-rank test.

Functional enrichment analysis

Differentially expressed genes between the high-risk and low-risk groups in the entire TCGA cohort were identified by DESeq2 [16] with a false discovery rate (FDR) < 0.05 and $|FC| > 1.5$ as the threshold value. KEGG pathway analysis was used to explore the potential mechanism of the model in COAD and performed by the “ClusterProfiler” R package [17].

Homologous recombination deficiency (HRD) score analysis

HRD causes three characteristic genomic scar signatures. The loss of heterozygosity (LOH) was defined as the number of counts of chromosomal LOH regions shorter than the whole chromosome and longer than 15 Mb [18]. Large-scale state transitions (LST) were defined as chromosome breakpoint (change in copy number or allelic content) between adjacent regions each of at least 10 megabases obtained after smoothing and filtering shorter than 3Mb small-scale copy number variation [19]. Telomeric allelic imbalance (TAI) was defined as the number of regions with the allelic imbalance which extends to the sub-telomere but do not cross the centromere [20]. The HRD score is the sum of these scar signature scores [21]. Each of the genomic scar signatures was quantified by algorithms in R, and the HRD score of patients in the TCGA-COAD cohort was calculated as previously described [22]. Comparison of HRD scores between high-risk and low-risk groups in the entire TCGA-COAD cohort was performed.

Statistical analysis

The Mann–Whitney and chi-squared U test was used to investigate categorical and quantitative data differences between different datasets or groups, respectively. Two-tailed $p < 0.05$ was used to determine statistical significance. All the statistical analysis and visualization were performed with the R version 4.0.2 (Institute for Statistics and Mathematics, Vienna, Austria 4).

Results

Establishment of a prognostic model based on the HRR-related lncRNAs in COAD

The correlation between HRR genes and lncRNAs was displayed in the Sankey diagram (Fig. 1A, Supplementary Table 1). A total of 9216 lncRNAs were further used to select the HRR-related lncRNAs that were associated with prognosis in COAD, and 33 prognostic HRR-related lncRNAs were selected (Supplementary Table 2). The TCGA-COAD cohort was randomized into the training set and the testing set. In the training set, we established a nineteen HRR-related lncRNA prognostic model by using the LASSO Cox method, including lnc-C2orf74-2, LINC01578, LINC01023, lnc-DNAH10OS-6, lnc-CYP2A6-1, lnc-LUC7L-1, lnc-CHRNA3-1, lnc-BRF2-7, TNFRSF10A-AS1, DSCR9, lnc-PRPF18-1, HAND2-AS1, lnc-MOCS1-1, lnc-VPREB1-1, lnc-SOX15-1, SNHG7, lnc-NCBP2-AS2-1, lnc-TMEM243-1, and lnc-TMEM71-1 (Fig. 1B, C). Correlation between the nineteen lncRNAs and HRR genes showed that the selected nineteen lncRNAs were all strongly correlated with HRR genes (Fig. 1D, Supplementary Fig. 1). Besides, multivariate Cox analysis also indicated that all these nineteen HRR-related lncRNAs were independent factors on the prognosis in COAD (Fig. 1E). In addition, we also compared the expression of these nineteen HRR-related lncRNAs between the TCGA-COAD cohort and the normal samples in both the TCGA database and the Genotype-Tissue Expression (GTEx) database by using the Gene Expression Profiling Interactive Analysis (GEPIA) tool. The result showed that lnc-BRF2-7 and TNFRSF10A-AS1 were highly expressed in COAD compared to normal tissues (both $p < 0.05$), while the expression of lnc-PRPF18-1, HAND2-AS1, and lnc-MOCS1-1 was decreased in COAD compared to normal tissues (all $p < 0.05$). No significant difference was found in the expression of other HRR-related lncRNAs between COAD and normal tissues (Supplementary Fig. 2).

Evaluation and validation of the prognostic model

The risk scores of COAD patients in the training set were calculated by a uniform formula, and the training set was divided into high-risk and low-risk groups according to the median risk score (Fig. 2A). The survival status of each patient is shown in Fig. 2B, which suggested that patients' survival status deteriorated as the risk score increased. The expression standards of the 19 HRR-related lncRNAs for each patient are displayed in Fig. 2C. As shown in Fig. 2D, COAD patients with high-risk scores had significantly inferior overall survival compared with the COAD patients with low-risk scores ($p < 0.0001$). As validation, the same processes were performed in both the testing

set and the entire TCGA cohort. Distribution of HRR-related lncRNA model-based risk score in the testing set and the entire TCGA cohort is depicted in Fig. 2E and H. The survival status of each patient in the testing set and the entire TCGA cohort is described in Fig. 2F and I, and the expression standards of the 19 HRR-related lncRNAs for each patient are displayed in Fig. 2G and J, respectively. Unsurprisingly, the survival status of patients in the testing set and the entire TCGA cohort was also getting worse along with the risk score increased. Similarly, the expression patterns of the 19 HRR-related lncRNAs in the testing set and the entire TCGA cohort were both consistent with that in the training set. Besides, COAD patients with high-risk scores had shorter overall survival in both the testing set (Fig. 2K, $p = 0.0086$) and the entire TCGA cohort (Fig. 2L, $p < 0.0001$), which was also consistent with the result shown in the training set.

And then, COAD patients in the entire TCGA cohort were stratified into different subgroups depending on the age (< 60 years and ≥ 60 years), gender (female and male), tumor stage (stages I–II and stages III–IV), and T stage (T3–T4 stage). As shown in Fig. 3, patients with lower risk scores had a better prognosis in all the subgroups (all $p < 0.05$), which indicated that the prognostic value of the nineteen HRR-related lncRNAs cannot be impacted by the different clinical features.

Independence of the nineteen HRR-related lncRNA prognostic model

Together with the typical features such as age, gender, tumor stage, and TMN stage, our model was an independent protective factor for OS in COAD with a hazard ratio (HR) = 0.34 and 95% confidence interval (95% CI) = 0.20–0.55 (Fig. 4A, $p < 0.001$). Besides, tumor stage was an independent prognostic factor for OS in COAD (Fig. 4A, $HR = 1.18$, 95% $CI = 1.05$ – 1.27 , $p = 0.003$). ROC curve and C-index indicated that our model had the best predictive efficiency compared with age, gender, tumor stage, and TMN stage stratification system (Fig. 4B, C). To further evaluate the predictability of the nineteen HRR-related lncRNA prognostic model, three lncRNA-based published models were collected, including the autophagy-related lncRNA prognostic model reported by Zhou et al. [23], the immune-related lncRNA prognostic model reported by Lin et al. [24], and the genome instability-related lncRNA prognostic model

reported by Yun et al. [25]. The result showed that our model had the largest AUC compared to the other three published models, which suggested that our model had better efficiency on the prediction of 2-year overall survival in COAD (Supplementary Fig. 3).

Exploration of the biological function of the prognostic model

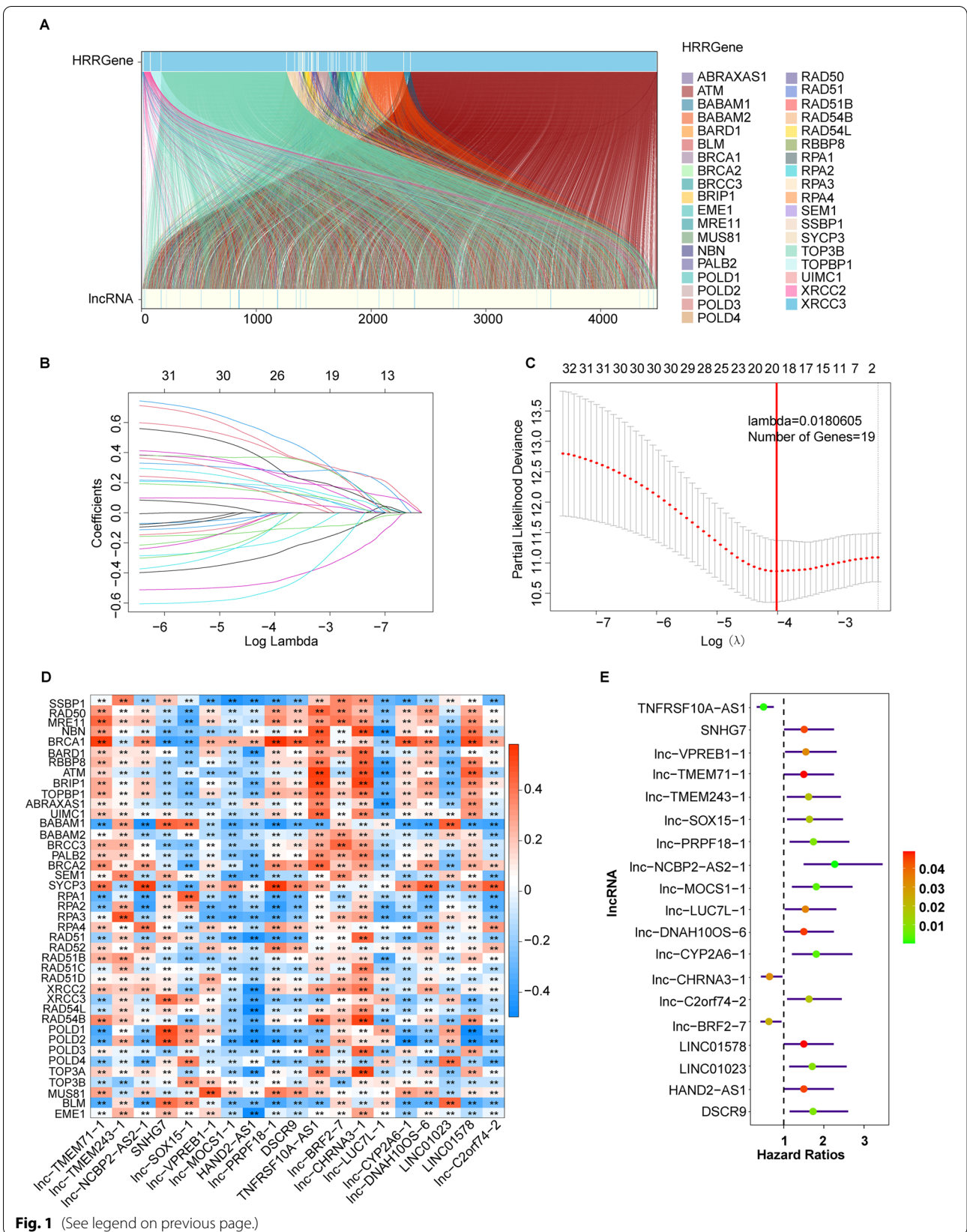
Differentially expressed gene analysis was performed between the high-risk and low-risk groups in the TCGA-COAD cohort. As displayed by the volcano plot, 27 upregulated genes and 721 downregulated genes were identified as the differentially expressed genes (DEGs) (Fig. 5A), and the composition of these DEGs is displayed as the bar diagram in Fig. 5B. KEGG pathway analysis was performed on the DEGs and the result showed that the DEGs were majorly enriched in MAPK signaling pathway, focal adhesion, ECM-receptor interaction, and proteoglycans in cancer and platelet activation (Fig. 5C). Besides, we predicted the target genes of the HRR-related lncRNAs using the starbase database [26] and constructed the HRR-related lncRNA interaction network by using the STRING database [27]. Pathway enrichment analysis on the metasplice website revealed that the HRR-related lncRNA interaction network was related to the nuclear chromosome segregation, DNA damage response, and CDC42 pathway (Supplementary Fig. 4). Additionally, we compared the HRD score between high-risk and low-risk groups in the TCGA-COAD cohort and the result showed that patients in the high-risk group had significantly higher HRD score than those in the low-risk group (Fig. 5D, $p < 0.0001$), which indicated that the nineteen HRR-related lncRNA model could be the potential predictor for the response of PARP inhibitors.

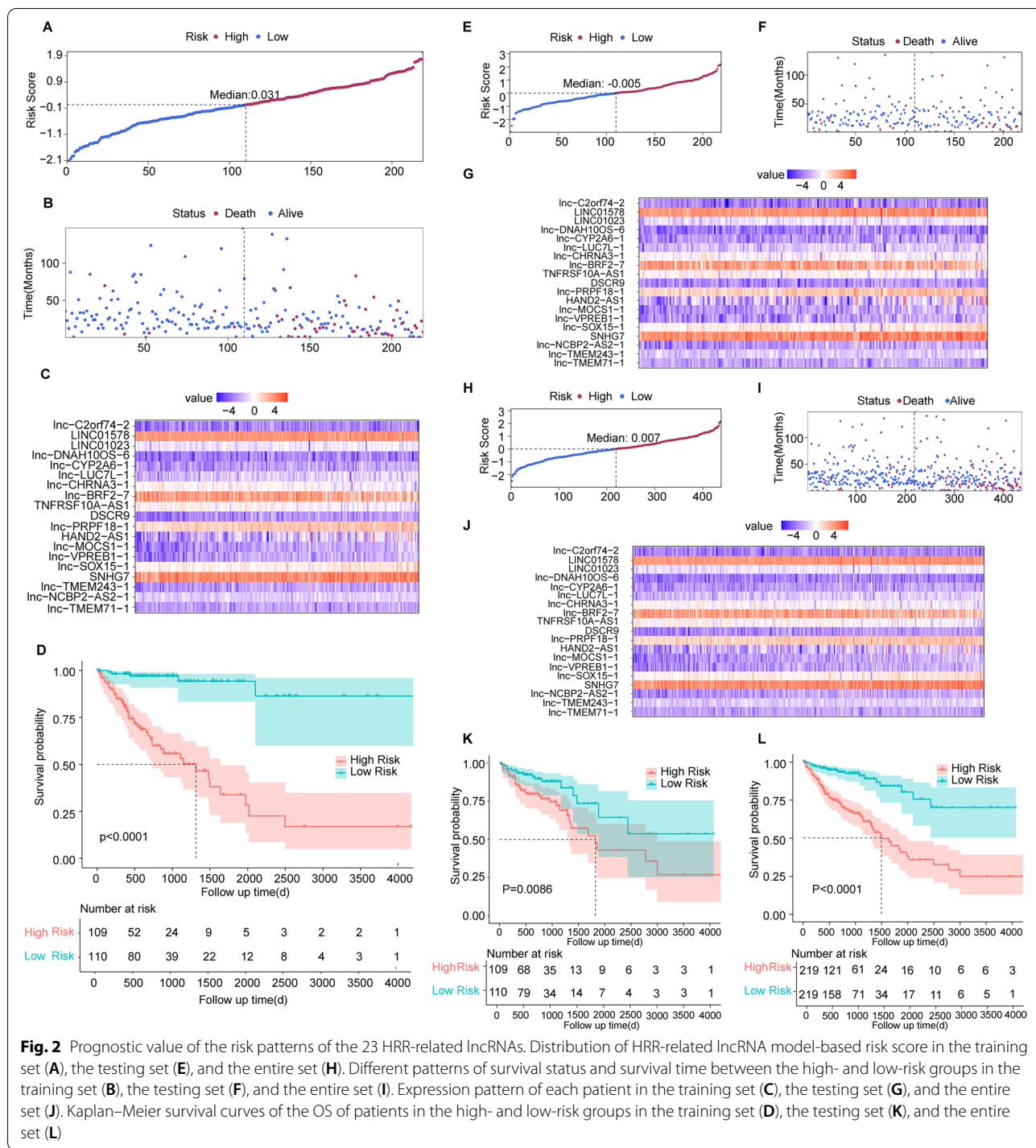
Discussion

lncRNAs are a kind of non-coding RNAs longer than 120nt in length, which have been illuminated as novel biomarkers and targets for the diagnosis and therapy in tumors [28–30]. In recent years, multiple lncRNAs have been reported to be regulators in the development of CRC and be associated with prognosis in CRC. For example, lncRNA LINC01224 regulated the proliferation and invasion of CRC through modulating the miR-485-5p/MYO6 axis [31]. Therefore, many lncRNA-based prognostic models have been developed, such

(See figure on next page.)

Fig. 1 Establishment of a prognostic model based on the HRR-related lncRNAs in the TCGA training set. **A** Sankey relational diagram for 41 HRR genes and HRR-related lncRNAs. **B** The LASSO coefficient profile of 33 prognostic HRR-related lncRNAs and perpendicular imaginary lines were drawn at the value chosen by 10-fold cross-validation. **C** The tuning parameters ($\log l$) of OS-related proteins were selected to cross-verify the error curve. According to the minimal criterion and 1-se criterion, perpendicular imaginary lines were drawn at the optimal value. **D** Heatmap for the correlations between 41 HRR genes and the 19 prognostic HRR-related lncRNAs. **E** Multivariable analysis of 19 HRR-related lncRNAs

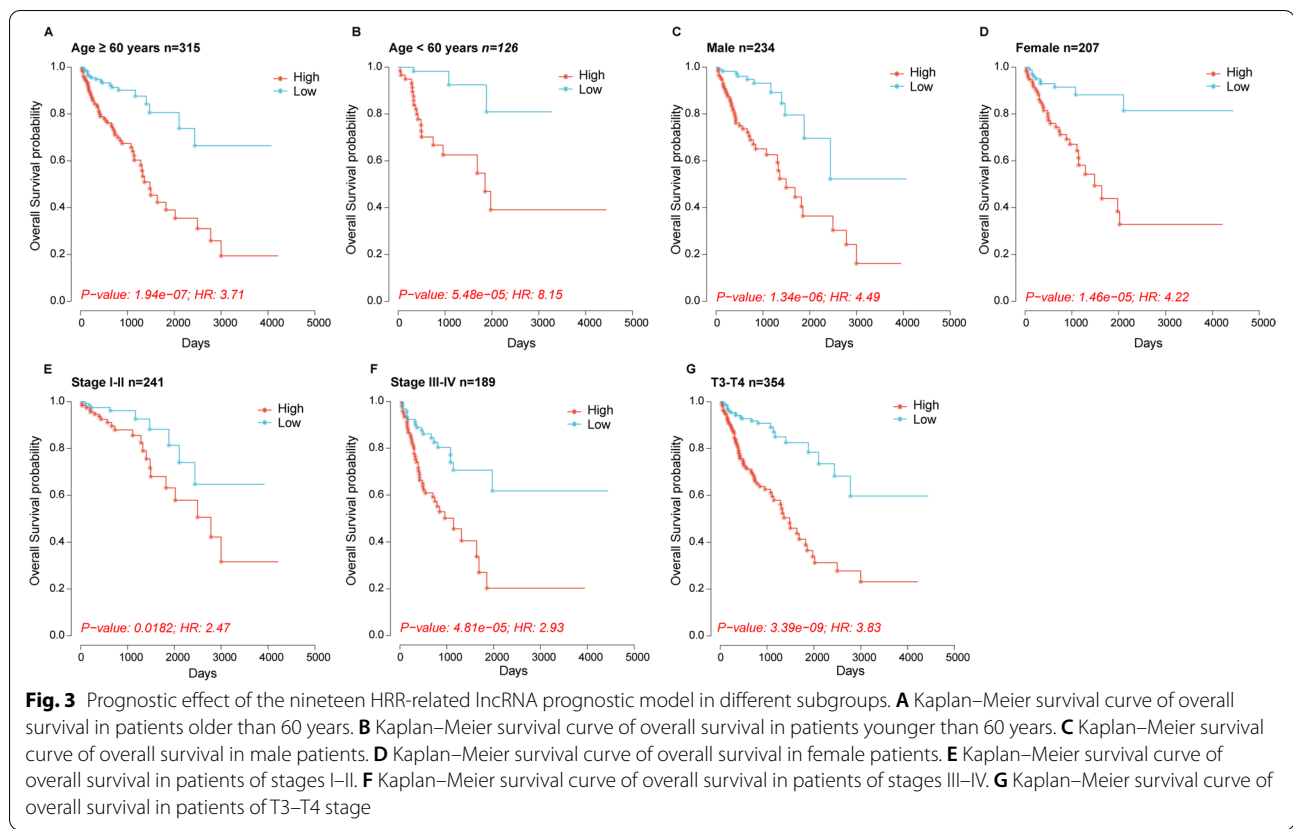




as a six lncRNA risk prognosis model reported by Gao et al. [32]. However, the role of a large number of lncRNAs in CRC remains to be illuminated. In this study, we selected nineteen lncRNAs that were correlated with HRR genes and associated with prognosis in COAD to establish a prognostic model. The model had better

predictive performance than other traditional indicators and recently published models in COAD.

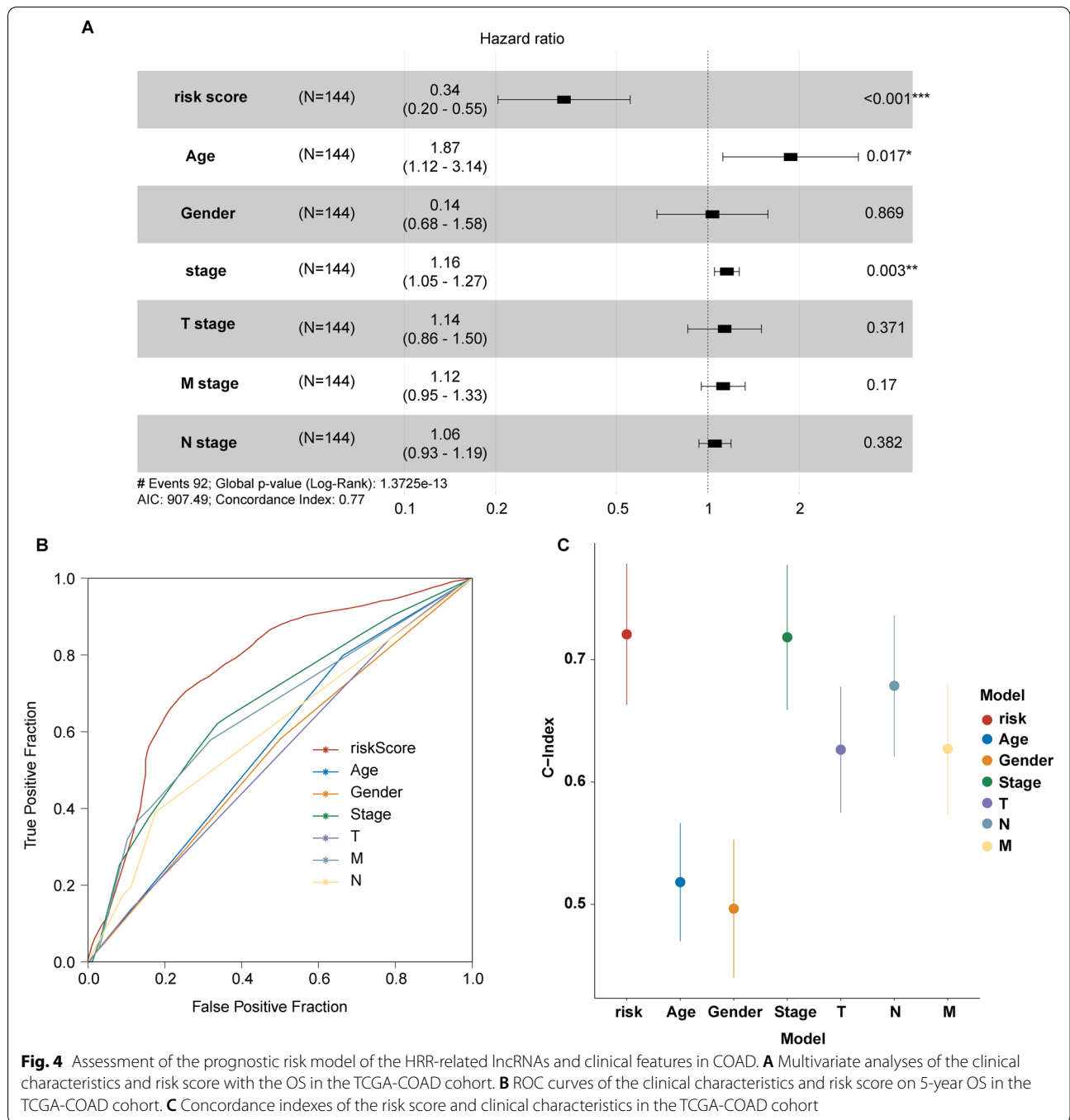
Among these nineteen HRR-related lncRNAs, six lncRNAs have been reported including LINC01578, LINC01023 [33], TNFRSF10A-AS1, DSCR9 [34, 35], HAND2-AS1, and SNHG7. LINC01578 has been



reported to be the driver in colon cancer metastasis [36]. TNFRSF10A-AS1 was related to autophagy in colorectal cancer and contributed to poor prognosis in colorectal cancer [37]. In addition, HAND2-AS1 is an important lncRNA that is significantly related to the regulation of cell proliferation, apoptosis, invasion, metastasis, and energy metabolism as well in numerous human tumors [38]. Besides, lncRNA SNHG7 is an important tumorigenesis promoter as well [39]. Additionally, other lncRNAs were unveiled for the first time. To evaluate the clinical significance of our model in COAD patients, we separated the COAD patients into high-risk and low-risk groups based on the median value of the risk score and compared the overall survival between the two groups. Unsurprisingly, patients in the high-risk group had worse prognosis than those in the low-risk group, suggesting that our model can be used to predict the prognosis of COAD patients. Moreover, to indicate whether the prognostic value of our model can be impacted by some clinical characteristics, COAD patients in the TCGA cohort were grouped by age, gender, tumor stage, and TMN stage. We found that our model had excellent prognostic effect in every subgroup. Besides, multivariate analysis also demonstrated that our model was an independent factor for prognosis in COAD. Furthermore, ROC curve

and C-index indicated that our model had the best performance on the 5-year OS prediction in COAD which suggested great dependability of our model in COAD. Additionally, we also compared the predictability of our model and three recently published models and the result also suggested that our model had better predictive efficiency than other published molecular models. Together, our findings illuminated that the nineteen HRR-related lncRNA prognostic model was a reliable predictor of prognosis in COAD.

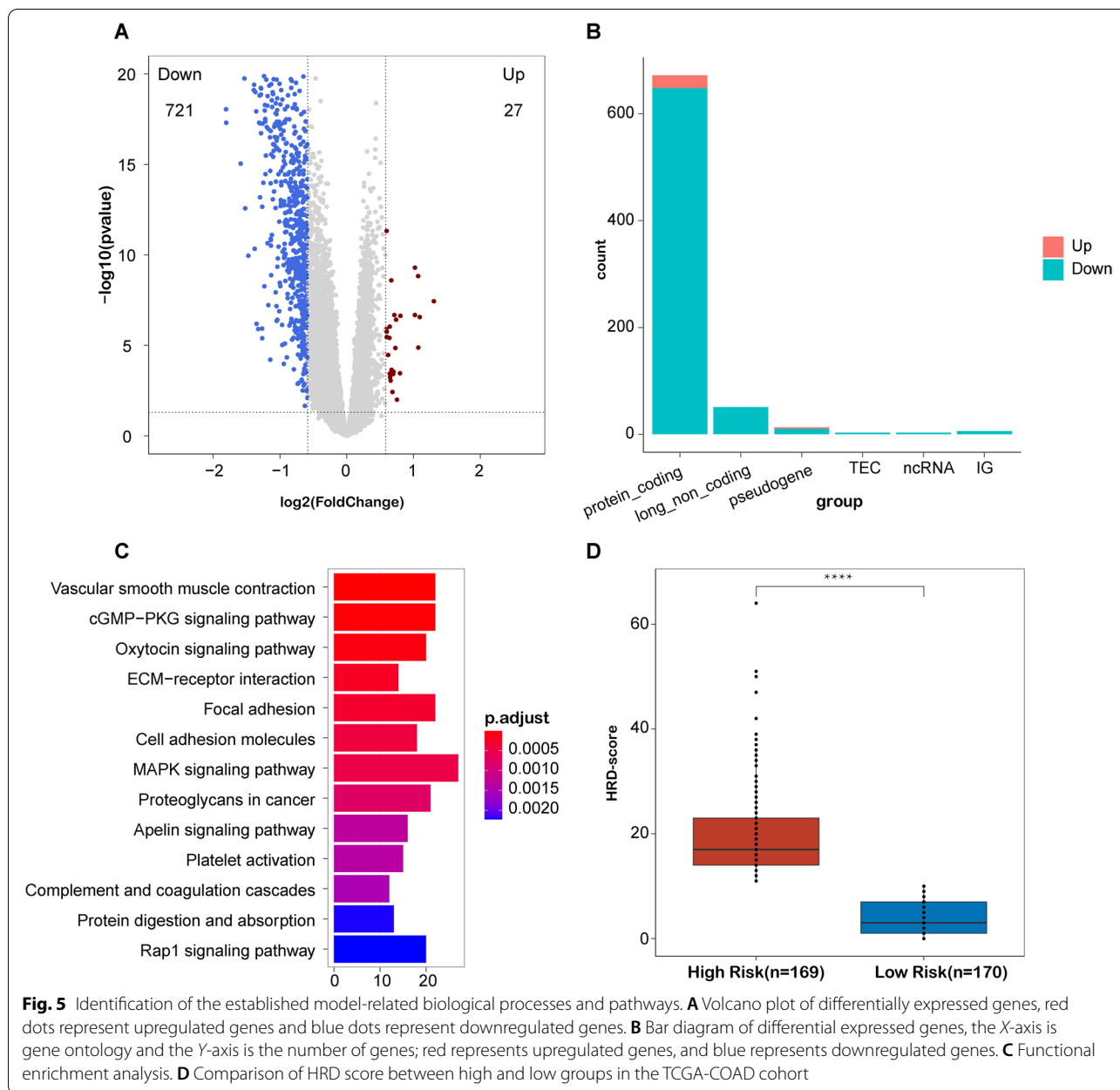
To further unveil the potential mechanism of this model in COAD, we identified the DEGs between high-risk and low-risk groups in the TCGA-COAD cohort, and the functional enrichment analysis was conducted to figure out the biological pathway of these DEGs. We indicated that our model was associated with the MAPK signaling pathway which was a famous signaling pathway that related to tumorigenesis and tumor metastasis [40]. In colon cancer, ERK/MAPK signaling pathway was involved in the effects of hepatocyte growth factor (HGF) of promoting proliferation and regulating cell cycle and apoptosis of cancer cells [41]. p38 α MAPK pathway also played a predominant role in colon cancer development and chemoresistance [42]. In recent years, FDA-approved inhibitors targeting the MAPK pathway have shown



promising clinical responses in patients with colorectal cancer [43]. Our findings suggested that the nineteen HRR-related lncRNA model might be the indicator for MAPK inhibitor response and provided new thought for novel biomarkers in COAD.

Numerous studies have illuminated that HRD existed in many types of cancer including ovarian cancer [44], breast cancer [45], lung cancer [46], and colon cancer

[7]. HRD score is an indicator to quantify the degree of HRD in tumors, and it has been employed on prediction of treatment response in triple-negative breast cancer [47, 48]. The presence of HRD results in irreparable DNA damage from platinum-containing drugs, which leads to cell death; therefore, tumors with HRD were more sensitive to PARP inhibitors such as Olaparib [49]. Therefore, we compared the HRD scores between high-risk



and low-risk groups in the TCGA-COAD cohort. We suggested that our model was positively correlated with the HRD score in the TCGA-COAD cohort. Our findings indicated that our model had the potential to be the predictor for PARP inhibitor response and provided new ideas for precision treatment in COAD.

In conclusion, we established a prognostic model containing nineteen HRR-related lncRNAs in COAD and its predictive efficiency of OS was excellent. Besides, our model also had a wonderful performance in the prediction of 5-year OS compared with other predictive factors

in the TCGA-COAD cohort. Additionally, we considered that our model could be the potential predictor and therapeutic target in the precision treatment of COAD.

However, the current study also has some limitations. We established and evaluated the model based on the published datasets and our study was more suitable for retrospective analysis. We preliminarily unveiled the mechanism of our model in COAD and further validation needs to be investigated. In the future, these shortcomings may be overcome through prospective experiments with a large number of patients.

Conclusion

We established a new prognostic model based on the HRR-related lncRNAs in COAD, and the model had excellent predictive efficiency. Besides, the model had better performance than other traditional indicators and recently published models. Besides, we preliminarily discussed the potential function of the model in COAD and hope to provide new ideas for the clinical treatment of COAD.

Abbreviations

CRC: Colon cancer; COAD: Colon adenocarcinoma; HRR: Homologous recombination repair; lncRNAs: Long non-coding RNAs; TCGA: The Cancer Genome Atlas; LASSO: Least absolute shrinkage and selection operator; ROC: Receiver operating characteristic curve; KEGG: Kyoto Encyclopedia of Genes and Genomes; C-index: Concordance index; HRD: Homologous recombination deficiency; LOH: Loss of heterozygosity; LST: Large-scale state transitions; TAL: Telomeric allelic imbalance.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12957-022-02534-0>.

Additional file 1: Supplementary Figure 1. Coexpression network between the nineteen HRR-related lncRNAs and HRR genes.

Additional file 2: Supplementary Figure 2. Comparison of the expression of the nineteen HRR-related lncRNAs between the COAD and normal tissues.

Additional file 3: Supplementary Figure 3. Comparison of ROC curve between the nineteen HRR-related lncRNAs prognostic model and other three published models in the TCGA-COAD cohort.

Additional file 4: Supplementary Figure 4. HRR-related lncRNAs interaction network construction and pathway enrichment analysis. (A) Interaction network of HRR-related lncRNAs and its target genes. (B) Pathway enrichment analysis.

Additional file 5: Supplementary tables.

Acknowledgements

We would like to acknowledge Guangzhou Panyu Central Hospital and Medical Simulation Center of Sun Yat-sen University for providing help during the project.

Authors' contributions

Study concept and design: TX, LY, HJ, and LX. Acquisition of the data: TX, LY, HJ, and LX. Analysis and interpretation of the data: TX, LY, HJ, LX, LJ, ZX, and LT. Drafting of the manuscript: TX, LY, HJ, and LX. Statistical analysis: TX, LY, HJ, LX, LJ, ZX, and LT. Administrative, technical, or material support: TX, LY, HJ, LX, LJ, ZX, and LT. Supervision: TX and LY. The authors read and approved the final manuscript.

Funding

The Science and Technology Program of Guangzhou under Grant No. 201904010070.

Availability of data and materials

The datasets generated and/or analyzed during the current study are available in the TCGA repository, <https://cancergenome.nih.gov/>.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 1 November 2021 Accepted: 18 February 2022

Published online: 09 March 2022

References

- Benson AB, Venook AP, Al-Hawary MM, Cederquist L, Chen YJ, Ciombor KK, et al. NCCN guidelines insights: colon cancer, version 2.2018. *J Natl Compr Canc Netw*. 2018;16(4):359–69.
- Edwards BK, Ward E, Kohler BA, Ehemam C, Zauber AG, Anderson RN, et al. Annual report to the nation on the status of cancer, 1975–2006, featuring colorectal cancer trends and impact of interventions (risk factors, screening, and treatment) to reduce future rates. *Cancer*. 2010;116(3):544–73.
- Azvolinsky A. Colorectal cancer: to stack or sequence therapy? *J Natl Cancer Inst*. 2015;107(5):djv138.
- del Rivero J, Kohn EC. PARP inhibitors: the cornerstone of DNA repair-targeted therapies. *Oncology (Williston Park)*. 2017;31(4):265–73.
- Jiang M, Jia K, Wang L, Li W, Chen B, Liu Y, et al. Alterations of DNA damage repair in cancer: from mechanisms to applications. *Ann Transl Med*. 2020;8(24):1685.
- D'Andrea AD. Mechanisms of PARP inhibitor sensitivity and resistance. *DNA Repair (Amst)*. 2018;71:172–6.
- Tennstedt P, Fresow R, Simon R, Marx A, Terracciano L, Petersen C, et al. RAD51 overexpression is a negative prognostic marker for colorectal adenocarcinoma. *Int J Cancer*. 2013;132(9):2118–26.
- Zhou P, Wu X, Chen H, Hu Y, Zhang H, Wu L, et al. The mutational pattern of homologous recombination-related (HRR) genes in Chinese colon cancer and its relevance to immunotherapy responses. *Aging (Albany NY)*. 2020;13(2):2365–78.
- Huarte M. The emerging role of lncRNAs in cancer. *Nat Med*. 2015;21(11):1253–61.
- Ling H, Vincent K, Pichler M, Fodde R, Berindan-Neagoe I, Slack FJ, et al. Junk DNA and the long non-coding RNA twist in cancer genetics. *Oncogene*. 2015;34(39):5003–11.
- Dianatpour A, Ghafouri-Fard S. The role of long non coding RNAs in the repair of DNA double strand breaks. *Int J Mol Cell Med*. 2017;6(1):1–12.
- Thapar R. Regulation of DNA double-strand break repair by non-coding RNAs. *Molecules*. 2018;23(11):2789.
- Zhao K, Wang X, Xue X, Li L, Hu Y. A long noncoding RNA sensitizes genotoxic treatment by attenuating ATM activation and homologous recombination repair in cancers. *PLoS Biol*. 2020;18(3):e3000666.
- Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw*. 2010;33(1):1–22.
- Simon N, Friedman J, Hastie T, Tibshirani R. Regularization paths for Cox's proportional hazards model via coordinate descent. *J Stat Softw*. 2011;39(5):1–13.
- Anders S, Huber W. Differential expression analysis for sequence count data. *Genome Biol*. 2010;11(10):R106.
- Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*. 2012;16(5):284–7.
- Abkevich V, Timms KM, Hennessy BT, Potter J, Carey MS, Meyer LA, et al. Patterns of genomic loss of heterozygosity predict homologous recombination repair defects in epithelial ovarian cancer. *Br J Cancer*. 2012;107(10):1776–82.
- Manie E, Popova T, Battistella A, Tarabeux J, Caux-Moncoutier V, Golmard L, et al. Genomic hallmarks of homologous recombination deficiency in invasive breast carcinomas. *Int J Cancer*. 2016;138(4):891–900.
- Birkbak NJ, Wang ZC, Kim JY, Eklund AC, Li Q, Tian R, et al. Telomeric allelic imbalance indicates defective DNA repair and sensitivity to DNA-damaging agents. *Cancer Discov*. 2012;2(4):366–75.
- Sztupinszki Z, Diossy M, Krzystanek M, Reiniger L, Csabai I, Favero F, et al. Migrating the SNP array-based homologous recombination deficiency measures to next generation sequencing data of breast cancer. *NPJ Breast Cancer*. 2018;4:16.

22. Marquard AM, Eklund AC, Joshi T, Krzystanek M, Favero F, Wang ZC, et al. Pan-cancer analysis of genomic scar signatures associated with homologous recombination deficiency suggests novel indications for existing cancer drugs. *Biomark Res.* 2015;3:9.
23. Zhou W, Zhang S, Li HB, Cai Z, Tang S, Chen LX, et al. Development of prognostic indicator based on autophagy-related lncRNA analysis in colon adenocarcinoma. *Biomed Res Int.* 2020;2020:9807918.
24. Lin Y, Pan X, Chen Z, Lin S, Chen S. Identification of an immune-related nine-lncRNA signature predictive of overall survival in colon cancer. *Front Genet.* 2020;11:318.
25. Yun D, Yang Z. Identification of a four-lncRNA prognostic signature for colon cancer based on genome instability. *J Oncol.* 2021;2021:7408893.
26. Li JH, Liu S, Zhou H, Qu LH, Yang JH. starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Res.* 2014;42(Database issue):D92–7.
27. von Mering C, Huynen M, Jaeggi D, Schmidt S, Bork P, Snel B. STRING: a database of predicted functional associations between proteins. *Nucleic Acids Res.* 2003;31(1):258–61.
28. Bhan A, Soleimani M, Mandal SS. Long noncoding RNA and cancer: a new paradigm. *Cancer Res.* 2017;77(15):3965–81.
29. Qu Y, Chen Y, Zhang L, Tian L. Construction of prognostic predictor by comprehensive analyzing alternative splicing events for colon adenocarcinoma. *World J Surg Oncol.* 2020;18(1):236.
30. Hu X, Zhang L, Tian J, Ma J. Long non-coding RNA PART1 predicts a poor prognosis and promotes the malignant progression of pancreatic cancer by sponging miR-122. *World J Surg Oncol.* 2021;19(1):122.
31. Gu J, Dong L, Wang Y, Nie W, Liu W, Zhao JA. LINC01224 promotes colorectal cancer progression through targeting miR-485-5p/MYO6 axis. *World J Surg Oncol.* 2021;19(1):281.
32. Gao M, Guo Y, Xiao Y, Shang X. Comprehensive analyses of correlation and survival reveal informative lncRNA prognostic signatures in colon cancer. *World J Surg Oncol.* 2021;19(1):104.
33. Yu M, Yu S, Gong W, Chen D, Guan J, Liu Y. Knockdown of linc01023 restrains glioma proliferation, migration and invasion by regulating IGF-1R/AKT pathway. *J Cancer.* 2019;10(13):2961–8.
34. Liu Z, Mi M, Li X, Zheng X, Wu G, Zhang L. lncRNA OSTN-AS1 may represent a novel immune-related prognostic marker for triple-negative breast cancer based on integrated analysis of a ceRNA network. *Front Genet.* 2019;10:850.
35. Zhuang H, Huang S, Zhou Z, Ma Z, Zhang Z, Zhang C, et al. A four prognosis-associated lncRNAs (PALnc) based risk score system reflects immune cell infiltration and predicts patient survival in pancreatic cancer. *Cancer Cell Int.* 2020;20:493.
36. Liu J, Zhan Y, Wang J, Wang J, Guo J, Kong D. Long noncoding RNA LINC01578 drives colon cancer metastasis through a positive feedback loop with the NF-kappaB/YY1 axis. *Mol Oncol.* 2020;14(12):3211–33.
37. Wei J, Ge X, Tang Y, Qian Y, Lu W, Jiang K, et al. An autophagy-related long noncoding RNA signature contributes to poor prognosis in colorectal cancer. *J Oncol.* 2020;2020:4728947.
38. Da CM, Cheng ZY, Gong CY, Nan W, Zhou KS, Zhao GH, et al. Role of HAND2-AS1 in human tumors. *Clin Chim Acta.* 2020;511:189–97.
39. Zhou Y, Tian B, Tang J, Wu J, Wang H, Wu Z, et al. SNHG7: a novel vital oncogenic lncRNA in human cancers. *Biomed Pharmacother.* 2020;124:109921.
40. Urošević J, Nebreda AR, Gomis RR. MAPK signaling control of colon cancer metastasis. *Cell Cycle.* 2014;13(17):2641–2.
41. Zhou G, Yang J, Song P. Correlation of ERK/MAPK signaling pathway with proliferation and apoptosis of colon cancer cells. *Oncol Lett.* 2019;17(2):2266–70.
42. Grossi V, Peserico A, Tezil T, Simone C. p38alpha MAPK pathway: a key factor in colorectal cancer therapy and chemoresistance. *World J Gastroenterol.* 2014;20(29):9744–58.
43. Shin MH, Kim J, Lim SA, Kim J, Lee KM. Current insights into combination therapies with MAPK inhibitors and immune checkpoint blockade. *Int J Mol Sci.* 2020;21(7):2531.
44. Konstantinopoulos PA, Ceccaldi R, Shapiro GI, D'Andrea AD. Homologous recombination deficiency: exploiting the fundamental vulnerability of ovarian cancer. *Cancer Discov.* 2015;5(11):1137–54.
45. Belli C, Duso BA, Ferraro E, Curigliano G. Homologous recombination deficiency in triple negative breast cancer. *Breast.* 2019;45:15–21.
46. Lok BH, Gardner EE, Schneeberger VE, Ni A, Desmeules P, Rekhman N, et al. PARP inhibitor activity correlates with SLFN11 expression and demonstrates synergy with temozolomide in small cell lung cancer. *Clin Cancer Res.* 2017;23(2):523–35.
47. Telli ML, Timms KM, Reid J, Hennessy B, Mills GB, Jensen KC, et al. Homologous recombination deficiency (HRD) score predicts response to platinum-containing neoadjuvant chemotherapy in patients with triple-negative breast cancer. *Clin Cancer Res.* 2016;22(15):3764–73.
48. Loibl S, Weber KE, Timms KM, Elkin EP, Hahnen E, Fasching PA, et al. Survival analysis of carboplatin added to an anthracycline/taxane-based neoadjuvant chemotherapy and HRD score as predictor of response-final results from GePARSixto. *Ann Oncol.* 2018;29(12):2341–7.
49. Ledermann JA, Pujade-Lauraine E. Olaparib as maintenance treatment for patients with platinum-sensitive relapsed ovarian cancer. *Ther Adv Med Oncol.* 2019;11:1758835919849753.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

