

## Article

# Transcriptomic Profiling Reveals an Enhancer RNA Signature for Recurrence Prediction in Colorectal Cancer

Divya Sahu<sup>1,2</sup>, Chen-Ching Lin<sup>3</sup> and Ajay Goel<sup>1,\*</sup> 

<sup>1</sup> Department of Molecular Diagnostics and Experimental Therapeutics, Beckman Research Institute of City of Hope Comprehensive Center, Biomedical Research Center, Monrovia, CA 91016, USA

<sup>2</sup> Department of Genetics, University of Alabama at Birmingham, Birmingham, AL 35294, USA

<sup>3</sup> Institute of Biomedical Informatics, National Yang Ming Chiao Tung University, Taipei 11221, Taiwan

\* Correspondence: [ajgoel@coh.org](mailto:ajgoel@coh.org); Tel.: +1-626-218-3452

**Abstract:** Background: Colorectal cancer (CRC) is one of the most fatal malignancies worldwide, and this is in part due to high rates of tumor recurrence in these patients. Currently, TNM staging remains the gold standard for predicting prognosis and recurrence in CRC patients; however, this approach is inadequate for identifying high-risk patients with the highest likelihood of disease recurrence. Recent evidence has revealed that enhancer RNAs (eRNAs) represent a higher level of cellular regulation, and their expression is frequently dysregulated in several cancers, including CRC. However, the clinical significance of eRNAs as recurrence predictor biomarkers in CRC remains unexplored, which is the primary aim of this study. Results: We performed a systematic analysis of eRNA expression profiles in colon cancer (CC) and rectal cancer (RC) patients from the TCGA dataset. By using rigorous biomarker discovery approaches by splitting the entire dataset into a training and testing cohort, we identified a 22-eRNA panel in CC and a 19-eRNA panel in RC for predicting tumor recurrence. The Kaplan–Meier analysis showed that biomarker panels robustly stratified low and high-risk CC ( $p = 7.29 \times 10^{-5}$ ) and RC ( $p = 6.81 \times 10^{-3}$ ) patients with recurrence. Multivariate and LASSO Cox regression models indicated that both biomarker panels were independent predictors of recurrence and significantly superior to TNM staging in CC (HR = 11.89,  $p = 9.54 \times 10^{-4}$ ) and RC (HR = 3.91,  $p = 3.52 \times 10^{-2}$ ). Notably, the ROC curves demonstrated that both panels exhibited excellent recurrence prediction accuracy in CC (AUC = 0.833; 95% CI: 0.74–0.93) and RC (AUC = 0.834; 95% CI: 0.72–0.92) patients. Subsequently, a combination signature that included the eRNA panels and TNM staging achieved an even greater predictive accuracy in patients with CC (AUC = 0.85). Conclusions: Herein, we report a novel eRNA signature for predicting recurrence in patients with CRC. Further experimental validation in independent clinical cohorts, these biomarkers can potentially improve current risk stratification approaches for guiding precision oncology treatments in patients suffering from this lethal malignancy.

**Keywords:** colorectal cancer; enhancer RNA; recurrence prediction biomarker; transcriptome expression profiling; LASSO Cox regression



**Citation:** Sahu, D.; Lin, C.-C.; Goel, A. Transcriptomic Profiling Reveals an Enhancer RNA Signature for Recurrence Prediction in Colorectal Cancer. *Genes* **2023**, *14*, 137. <https://doi.org/10.3390/genes14010137>

Academic Editor: Paolo Cinelli

Received: 8 December 2022

Revised: 19 December 2022

Accepted: 21 December 2022

Published: 3 January 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Despite advances in the diagnosis and treatment of colorectal cancer (CRC), the disease still accounts for approximately 10% of global cancer cases [1]. According to the American Cancer Society, an estimated 52,580 deaths are expected from CRC in the United States in 2022 [2]. CRC is a highly heterogeneous disease [3]. Depending on the anatomical location where the disease initiates, it can be classified into colon cancer (CC) or rectal cancer (RC); however, both disease types are often grouped because they share similar molecular characteristics [4]. The five-year survival rate for CRC patients is approximately 65% for all stages [2]. However, only an estimated 38% of patients are diagnosed with localized disease, for which the five-year survival rate can be more than 90% [2]. However,

anatomically, the five-year survival rate for all stages in CC is 63% and for RC is 67% [2]. In the current clinical practice, clinicopathological features such as patient age, gender, primary site location, tumor grade, and tumor-node-metastasis (TNM) staging are used for predicting prognosis and recurrence in CRC patients [5]. However, these conventional clinical practices are not precise, resulting in inaccurate patient stratification [6,7]. Several protein-coding genes have been developed as molecular markers for CRC prognosis [8–11], yet the survival rate is unsatisfactory [12]. Therefore, there is imperative to discover novel recurrence-associated biomarkers that may contribute to improved risk stratification and identify high-risk CRC patients for appropriate clinical decision-making and treatments.

With the completion of the Human Genome Project in 2003, it is evident that protein-coding genes account for merely 1–2% of the human genome [13], while an astounding proportion of the genome serves as a noncoding component, which was previously characterized as a dark matter of the genome. Noncoding RNAs are a rich assortment of RNA molecules spanning from short RNAs (including miRNAs, piRNAs, and snoRNAs) to long RNAs (including lncRNAs and circular RNAs) [14]. Evidence suggests that these noncoding RNAs play a regulatory role in the biological processes and are involved in cancer development and progression [14,15]. Recently, enhancer RNAs (eRNAs) that are between 50–2000 nucleotide long have been discovered as one class of noncoding RNAs that are bidirectionally transcribed from the enhancer regions [16]. Several studies have reported dysregulation of eRNAs in cancers [17], which leads to the altered expression of oncogenes [18,19] and tumor suppressor genes [20]. Gu et al. [21] identified AP001056.1 enhancer RNA high expression was associated with good outcomes in patients with squamous cell carcinoma of the head and neck carcinoma. Likewise, eRNAs regulate active enhancers, promote gene expression and affect cell-specific transcriptional regulation [16,22]. In colon cancer, CRISPR loss-of-function screens analysis identified bromodomain and extra terminal (BET) member protein BRD4. Inhibition of BRD4 decreased tumor growth and increased differentiation of CIMP+ colon cancer tumors. JQ1, a BET inhibitor, reduced colon cancer proliferation. Further examining the RNA-Seq and ChIP-Seq data after JQ1 treatment led to the identification of colon cancer-associated transcript 1 (CCAT1) eRNA, which was significantly downregulated. CCAT1 enhancer regulates cMYC expression and was associated with poor outcomes in colon cancer patients [19]. The expression level of eRNAs within the enhancer regions positively correlates with the expression level of nearby protein-coding genes [16]. Sweeney et al. [23] demonstrated that Kallikrein-related peptidase 3 enhancer RNA (KLK3e) mediates KLK3 enhancer and KLK2 promoter interaction and promotes KLK2 transcription. In addition, KLK3e showed a positive correlation with KLK2 gene expression (neighbor target gene) in both primary and metastatic prostate tumors. This evidence suggests eRNAs can be considered as genome wide markers for active transcriptional regulation and prognostic markers predictive of patient cancer outcome.

In view of the emerging interest for eRNAs as a promising category of biomarkers, herein we developed an unbiased, systematic bioinformatics approach to identify a novel and robust eRNA panel that can predict tumor recurrence in patients with stage II and III colorectal cancer.

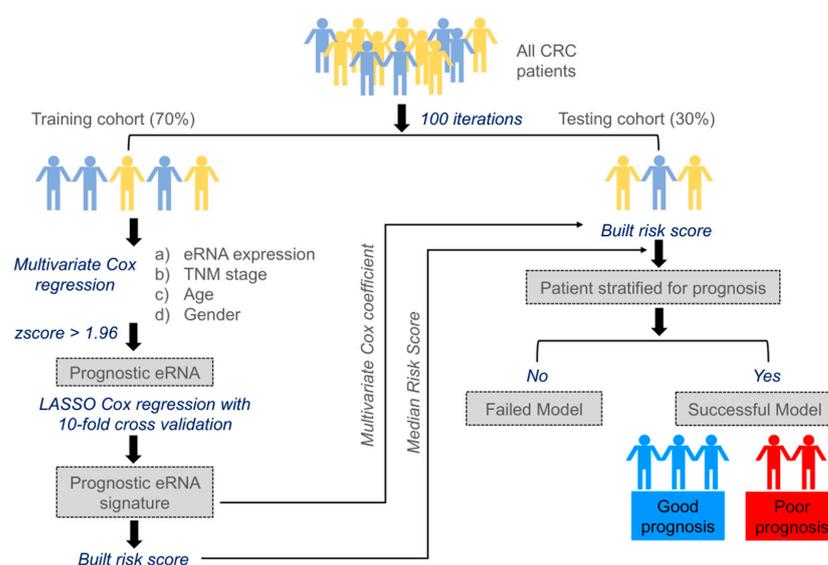
## 2. Results

### 2.1. Building Survival Model to Identify Clinically Relevant eRNAs

Through the following steps, we developed an eRNA-based survival model to identify clinically relevant eRNA for CC and RC (Figure 1).

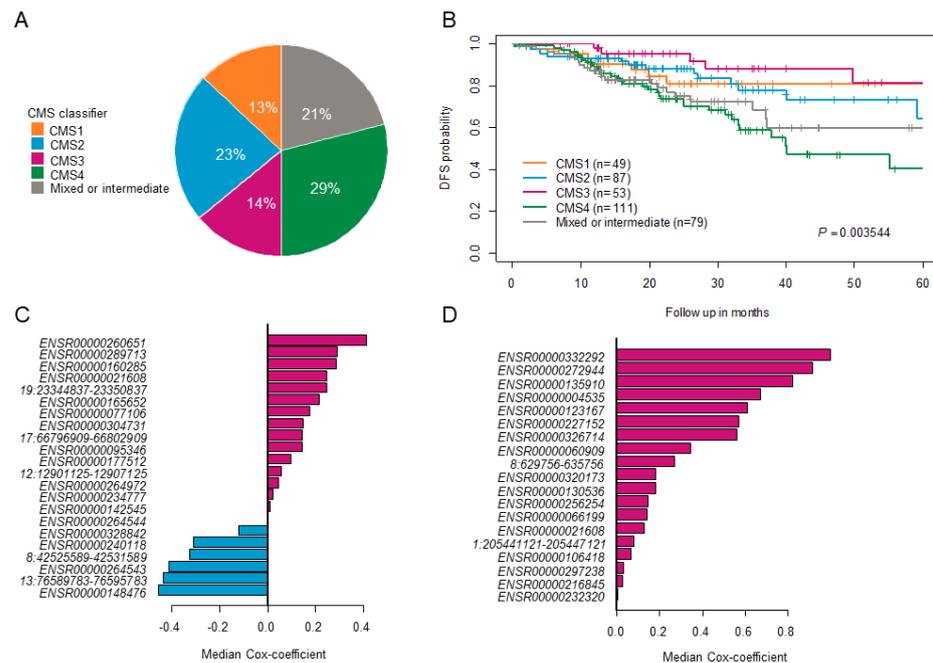
1. We randomly selected 70% of the entire cohort from The Cancer Genome Atlas (TCGA) as the training set and reserved the remaining 30% of the cohort as the testing set. To avoid deviation affecting the stability of the model, we maintained the distribution of disease-free and relapsed patients from the entire cohort in both training and testing sets.

- Multivariate Cox regression analysis was carried out training on 477 eRNAs for CC and 460 eRNAs for RC, along with controlling the effects from other clinical risk factors, including patient age, gender, and TNM stage.
- eRNAs significantly associated ( $p < 0.05$  &  $z\text{-score} > 1.96$ ) for predicting patient disease-free survival (DFS) were retained and termed as prognostic eRNAs and were further selected by LASSO Cox regularization with 10-fold cross-validation.
- Following LASSO Cox regularization and eRNA selection, a risk score formula was established. The risk score for each patient was calculated by a linear combination of expression and multivariate Cox coefficient of eRNAs.
- Patients were classified into low-risk or high-risk groups using the median risk score as the cut-off threshold from the training set. The coefficient for each eRNA and the cut-off value of the risk score from the training set was used to calculate the risk score and to stratify patients into individual risk groups in the testing set.
- Survival differences between the two groups were estimated using Kaplan–Meier curve and compared using the log-rank test.



**Figure 1.** Workflow showing steps involved in the identification of prognostic enhancer RNA signature.

The above-mentioned steps 1–6 were repeated 100 times to obtain 100 different eRNA sets. Only those sets that significantly stratified patients into two groups (high-risk groups exhibited poorer survival than low-risk groups) in the testing were termed successful survival models or failed models. The results obtained from one such trial are shown in Figure S1. For CC, we applied the model to the entire cohort ( $n = 379$ ), however, the model was generalized, and we found very few prognostic eRNAs signatures, which successfully stratified patients into two different risk groups within the testing cohort. Due to CC heterogeneity, we identified prognostic eRNAs signatures within CC subtypes based on Consensus Molecular Subtypes (CMS) classification. The entire CC cohort ( $n = 379$ ) was classified into 5 subtypes, with CMS1 comprising 13%, CMS2 23%, CMS3 14%, CMS4 29%, and mixed subtype 21% (Figure 2A). The proportion of patients classified into CMS subtypes was approximately similar to those described in Guinney et al. [24] study. We further checked the five-year DFS probability of each CMS subtype with CMS1: 89%; CMS2: 64%; CMS3: 81%; CMS4: 40%; and mixed subtype: 60%, as shown in Figure 2B. As CMS4 displayed the worst survival compared to other CMS subtypes, we focused on identifying a clinically relevant eRNA signature for this subtype. We found 22 prognostic eRNAs for the CC-CMS4 subtype and 19 prognostic eRNAs for RC that significantly distinguished patients into two risk groups in the testing set (Figure 2C,D).



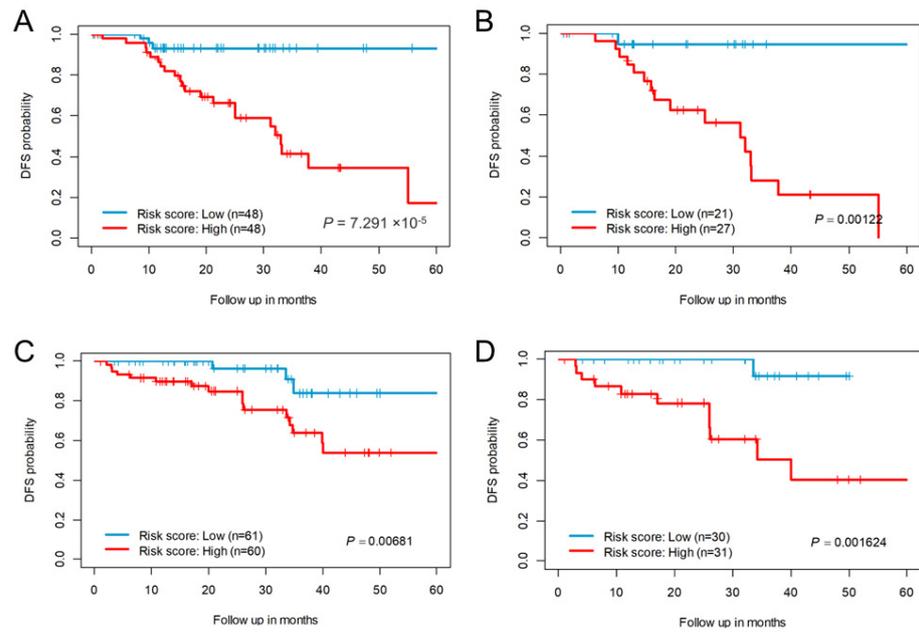
**Figure 2.** Identification of clinically relevant enhancer RNAs. **(A)** Pie-chart shows the percentage of colon cancer patients classified into various consensus molecular subtypes (CMS). **(B)** Kaplan–Meier plot shows survival estimates of disease-free survival in colon cancer patients stratified based on the CMS classification. **(C,D)** The bar graph shows prognostic eRNAs from the successful trials that significantly separated samples into two risk groups in the testing set and are ordered by their median cox-coefficient for disease-free survival. Positive scores are associated with shorter survival, and negative scores are associated with longer survival—22 prognostic eRNAs in colon cancer CMS4 subtype and 19 prognostic eRNAs in rectal cancer. ERNAs = enhancer RNAs.

### 2.2. Prognostic Association of eRNA Signature Risk Score with Disease-Free Survival in Patients

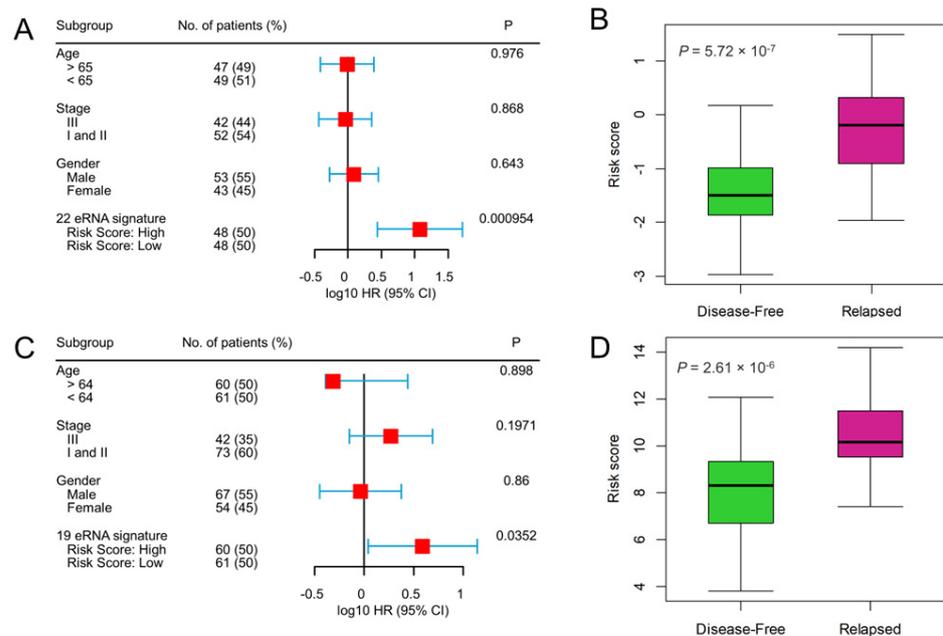
A risk score was constructed using the median Cox-coefficients from the successful survival models. For the CC-CMS4 subtype, compared with patients in the low-risk score group, those with high-risk scores exhibited poor DFS (Figure 3A). Only 17% of CC patients in the high-risk group were disease-free at 5 years, compared with 93% of the patients in the low-risk score group. We further randomly divided the entire cohort into 50% and used the same risk-scoring formula and cut-off thresholds of the whole cohort, wherein the 22-eRNA signature still significantly stratified patients into two risk groups for DFS (Figure 3B). For RC, 54% of the patients in the high-risk score group were disease-free at 5 years, compared with 84% of the low-risk groups (Figure 3C). Similarly, upon randomly splitting the RC cohort into 50% and using the same cut-off threshold, the 19-eRNA signature significantly stratified the patients into two risk groups (Figure 3D).

### 2.3. Recurrence Prediction by eRNA Signature Is Independent of Clinical Risk Factors

To evaluate the prognostic independence of eRNA signature against the known clinical risk factors, multivariate Cox regression analysis revealed that the 22-eRNA signature was an independent predictor of DFS in CC patients (Hazard Ratio [HR] = 11.89,  $p = 9.54 \times 10^{-4}$  Figure 4A). We also found that patients with relapsed tumors manifested significantly higher risk scores than disease-free patients ( $p = 5.72 \times 10^{-7}$ ; Figure 4B). Similarly, the 19-eRNA signature was independently associated with DFS in RC (HR = 3.91,  $p = 3.52 \times 10^{-2}$ ; Figure 4C), and their expression was significantly higher in the relapsed patient ( $p = 2.61 \times 10^{-6}$ ; Figure 4D).



**Figure 3.** Survival estimates of enhancer RNA signature for disease-free survival probability. (A) Kaplan–Meier plots show low- and high-risk score groups based on the median risk score of the colon cancer CM4 subtype cohort (B) Survival estimates of randomly selected 50% of the colon cancer CMS4 subtype cohort. Kaplan–Meier plots show low- and high-risk score groups based on the median risk score threshold value from the entire colon cancer CMS4 cohort. (C) Kaplan–Meier plots show low- and high-risk score groups based on the median risk score of the rectal cancer cohort (D) Survival estimates of randomly selected 50% of the rectal cancer cohort. Kaplan–Meier plots show low- and high-risk score groups based on the median risk score threshold value from the entire rectal cancer cohort. The *p* values were obtained using a Mantel log-rank test (two-sided). DFS = Disease-free survival.

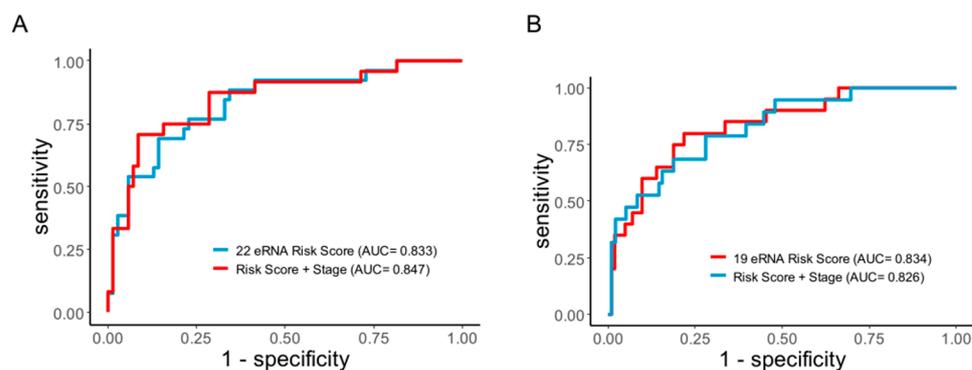


**Figure 4.** Multivariate Cox-regression analysis shows survival prediction by enhancer RNA signature is independent of clinical risk factors, and enhancer RNA signature risk score is up-regulated in the

relapsed patient. (A,B) colon cancer CMS4 subtype, (C,D) rectal cancer. A Forest plot of the eRNA signature shows that patients in high-risk score groups had poor outcomes and an independent predictor of disease-free survival after adjusting for the clinical risk factors. The hazard ratio and the confidence interval for colon and rectal cancer are represented on a log10 scale. All statistical tests were two-sided. CI = confidence interval, HR = hazard ratio, eRNA = enhancer RNA.

#### 2.4. The eRNA Signature Is a Better Predictor of Recurrence with High Sensitivity and Specificity

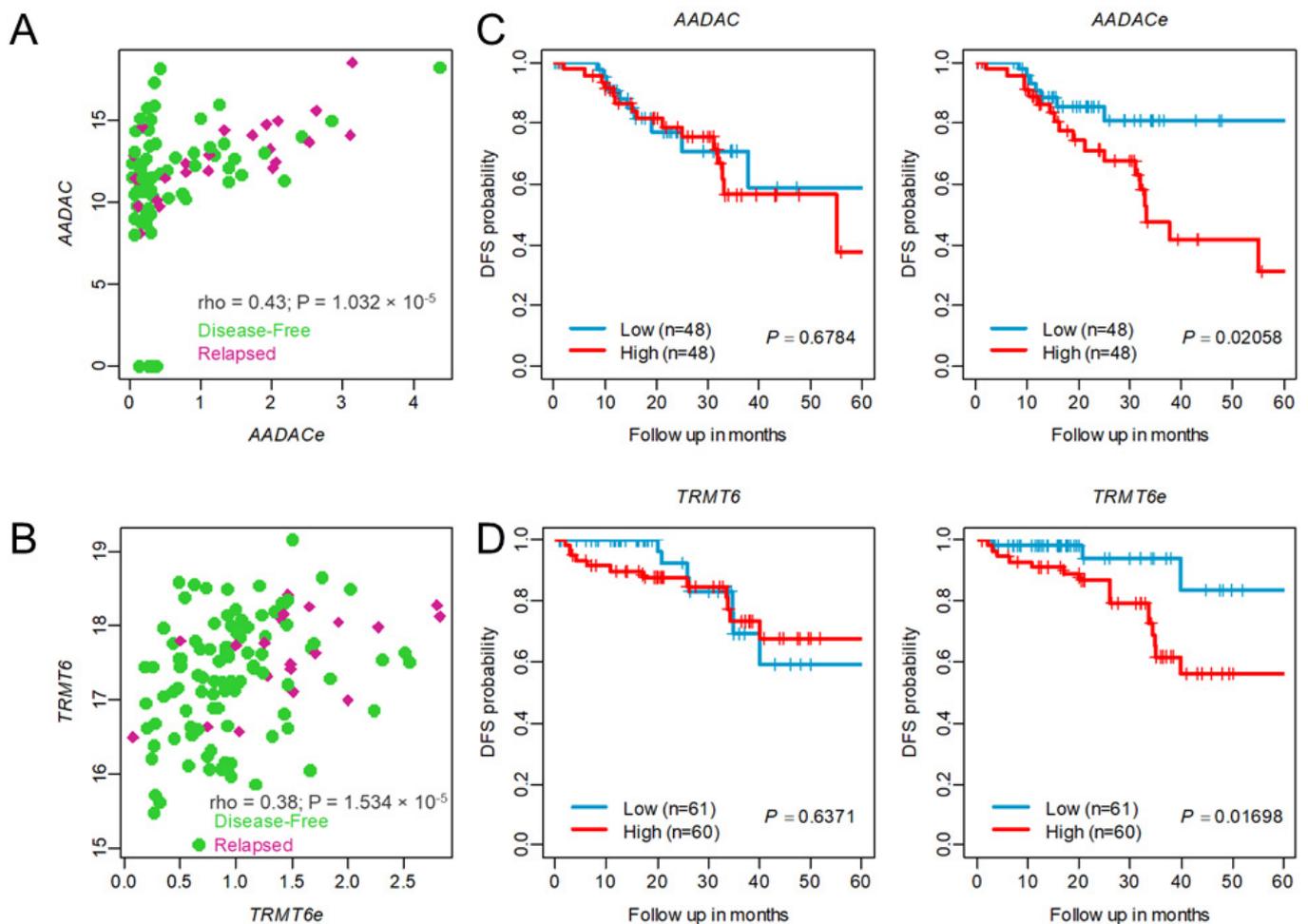
To confirm the prediction accuracy for DFS, we performed the Receiver Operating Characteristic Curve analysis (ROC) curve analysis for the eRNA signature. In both CC and RC, we observed that the eRNA signature exhibited excellent prediction power for relapse. For CC, the area under the curve (AUC) and 95% confidence interval (CI) for the 22-eRNA signature was 0.83 and 0.739–0.928 (Figure 5A), and for RC, the 19-eRNA signature achieved a remarkable prediction accuracy (AUC = 0.834, 95% CI = 0.737–0.931; Figure 5B). Subsequently, when the eRNA signature was combined with the TNM stage, it achieved an even higher prediction accuracy in CC (AUC = 0.847), suggesting its potential clinical utility in this malignancy. However, no increment in prediction accuracy was observed in RC when combined with the TNM stage.



**Figure 5.** Receiver operating characteristics (ROC) curve analyses of disease-free survival prediction by the enhancer RNA signature. ROC curve shows eRNA higher sensitivity and specificity for predicting disease-free survival prediction when combined with TNM staging (A) colon cancer CMS4 subtype (B) rectal cancer. eRNA = enhancer RNA; ROC = receiver operating characteristic.

#### 2.5. Higher Expression of eRNAs Associated with Tumor Recurrence Compared to Its Target Genes

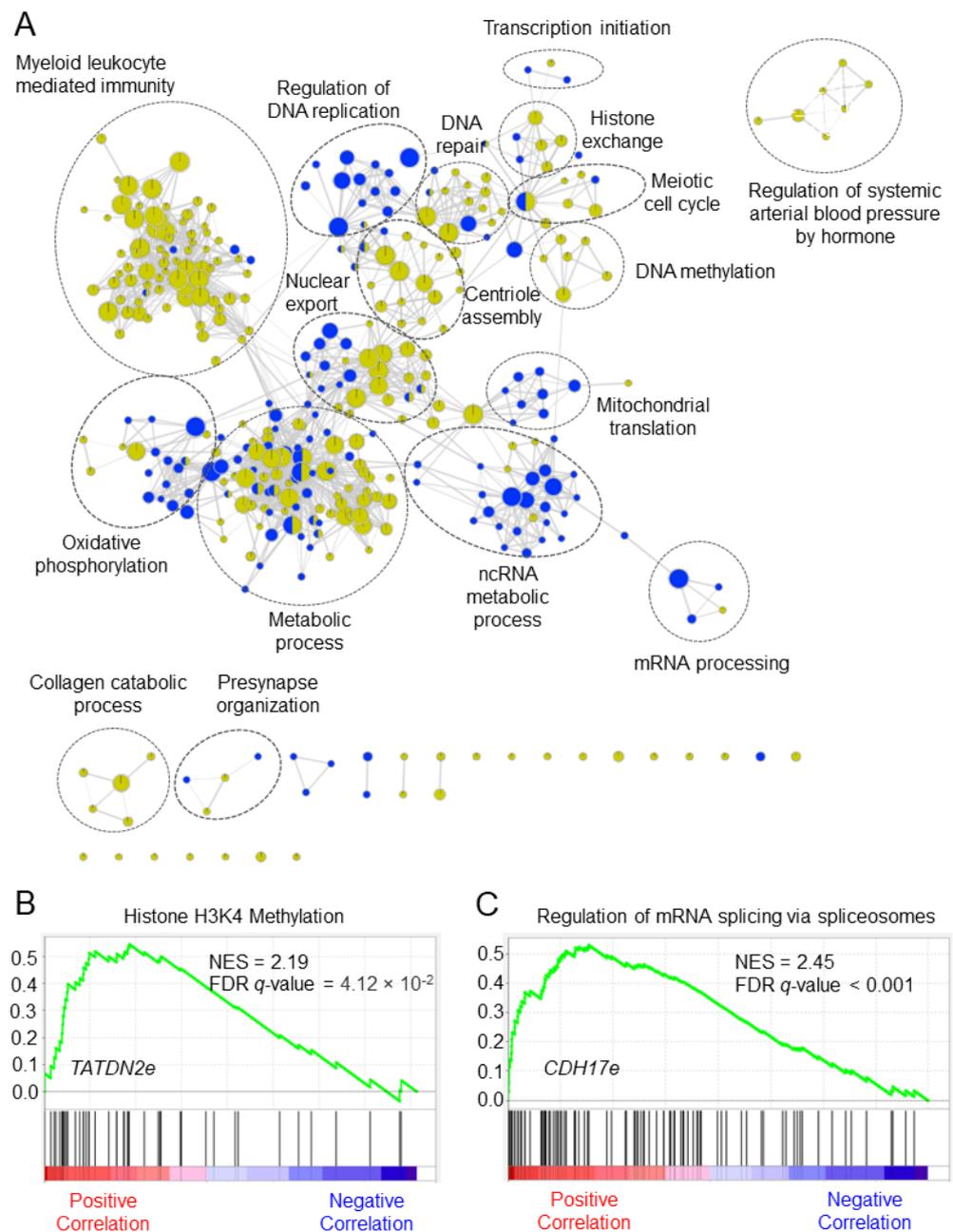
We performed the annotation of each prognostic eRNA (as suggested by Zhang et al. [25]) and examined the expression profiles of eRNAs and their neighboring target genes, which were highly positively correlated. For example, *AADAC*-associated eRNA (*AADACe*, ENSR00000160285) was highly associated with the *AADAC* gene in CC (SCC = 0.43,  $p = 1.03 \times 10^{-5}$ , as shown in Figure 6A). Similarly, *TRMT6*-associated eRNA (*TRMT6e*, ENSR00000106418) was highly correlated with the expression of the *TRMT6* gene in RC (SCC = 0.38,  $p = 1.53 \times 10^{-5}$ ) as shown in Figure 6B. Several studies have reported that eRNA expression has a better survival prediction potential than target genes. In support of the previous studies, we also observed that a high level of *AADACe* was associated with worse survival (log-rank test  $p = 0.02058$ , Figure 6C). Interestingly, the *AADAC* gene was not associated with CC patient survival, suggesting *AADACe* clinical utility regardless of *AADAC* expression in CC patients. Similar results were observed for *TRMT6*-associated eRNA (*TRMT6e*) in RC, as shown in Figure 6D. Further, we compared the prediction accuracy of the eRNA signature with other published CRC recurrence prediction gene signatures [26,27]. We extracted their gene list, performed multivariate Cox regression analysis, built risk scores from the CC cohort, and calculated the AUC of the ROC curve. As expected, the eRNA signature showed higher performance for DFS than other prognostic gene signatures (Figure S2).



**Figure 6.** Association between enhancer RNA and their neighbor protein-coding target genes. Spearman correlation coefficient (SCC) analysis shows a positive correlation between enhancer RNA and neighbor protein-coding genes from the (A) colon cancer CMS4 subtype cohort and (B) rectal cancer cohort. Dots in green represent disease-free patients, and red represents patients with relapsed tumors. (C,D) Kaplan–Meier plots show enhancer RNA higher expression is associated with tumor relapsed than its neighbor target genes in the colon cancer CMS4 subtype cohort and rectal cancer cohort. rho = spearman correlation coefficient.

## 2.6. Putative Biological Functions of eRNA Signature in Colorectal Cancer

The eRNAs have no protein-coding capacity; thus, we applied a guilt-by-association strategy to explore the putative potential biological functions of the eRNA signature in CRC, as shown in Figure 7A. We found biological functions related to metabolic process, DNA methylation, transcription initiation, cell cycle, and myeloid leukocyte mediated immunity for CC eRNA signature. Similarly, biological functions associated with the regulation of DNA replication, ncRNA metabolic process, mRNA processing, and oxidative phosphorylation were enriched for RC eRNA signature. The highest enriched biological functions for *TATDN2*-associated eRNA (*TATDN2e*, ENSR00000148476) in CC was histone H3K4 methylation, and *CDH17*-associated eRNA (*CDH17e*, ENSR00000227152) in RC was regulation of mRNA splicing by spliceosomes (Figure 7B,C).



**Figure 7.** Putative biological functions of the enhancer RNA signature in colorectal cancer. **(A)** The network shows overrepresented gene sets for the signature. Green nodes represent gene sets enriched in colon cancer, and blue nodes represent gene sets enriched in rectal cancer. Node size is proportional to the normalized enrichment score. Gene sets with similar biological functions tend to form clusters, these were manually identified, and gene ontology terms with the highest NES were labeled. **(B,C)** The enrichment plot shows the highest enriched function for *TATDN2*-associated eRNA (*TATDN2e*, ENSR00000148476) in colon cancer and *CDH17*-associated eRNA (*CDH17e*, ENSR00000227152). FDR = false discovery rate, NES = normalized enrichment score.

### 3. Discussion

The human eRNAs are a specific class of noncoding regulatory RNAs transcribed from the genomic enhancer regions and act in cis-direction to influence the transcription of the neighboring protein-coding genes. To the best of our knowledge, our study is the first to report the analysis and discovery of an RNA-seq-based prognostic eRNA signature in CRC. Cox proportional hazard regression is widely used to decipher the association of

molecular marker expression profiles with disease outcomes. Using the eRNA expression profile for the CC and RC cohort from the TCGA database, we identified a 22- and 19-eRNA panel for predicting prognosis in CC and RC, respectively. Using a machine learning approach for the random selection of eRNAs from the training set and further validation of the survival model in the testing set with 100 iterations increased the robustness of the identified prognostic eRNAs. A risk-scoring formula was applied to integrate predictive eRNA expression into a signature. Patients were stratified into a low-risk score group and a high-risk score group based on the median risk score of the entire cohort. Kaplan–Meier survival analysis with Mantel log-rank test evaluated the association of eRNA signature with patient DFS. We validated the prognostic association of the signature by randomly selecting 50% of the cohort and using the median risk score of the entire cohort to stratify the patient into two risk score groups. Multivariable Cox analysis was used to estimate the prognostic independence of the eRNA signature against the established clinical risk factors in CRC. ROC curve analysis demonstrated that the CC recurrence prediction accuracy for the eRNA signature was higher than the protein-coding genes signature.

The human CC is a highly heterogeneous disease; therefore, we identified consensus molecular subtypes using gene sets for consensus molecular subtype classification. As the CMS4 subtype was associated with the worst prognosis, we developed a prognostic signature for this subtype. In this study, we also incorporated stage 1 patients, as we hypothesized that these patients in CMS4 subtypes might have a higher risk of relapse than in other subtypes. We found that among 477 eRNAs from CC and 460 eRNAs from RC, only 412 eRNAs (78%) were in common, suggesting similarity in the eRNA landscape in the two-cancer type, however, we found only 1 prognostic eRNA *ENSR0000021608* similar between the two cancer types (Figure S3) suggesting prognostic specificity of eRNA in their respective cohort.

We would like to acknowledge the potential limitations of our study. First, our study's CC and RC cohorts were only from the TCGA. Therefore, this study used a random selection of prognostic eRNA signatures from the training set and further validated the successful stratification of the cohort into two risk score groups in the testing set. Second, the mutational status of KRAS and BRAF genes and CIMP and MSI status were not included as a covariate in the multivariate Cox analysis as information for these clinical risk factors were unavailable for these patient cohorts. Third, our study could not significantly stratify the patients into two risk score groups for overall survival because only a few deaths were observed (death vs. censored in CC; 14/82 and in RC; 5/116). Despite these drawbacks, significant stratification of patients into two risk score groups in the testing sets in various iterations and further validation in the entire cohort with sophisticated statistical analysis offers a high confidence level in the overall analyses.

#### 4. Conclusions

To conclude, we developed a signature consisting of 22 eRNAs in CC and 19 eRNAs in RC that can predict tumor relapse in these two malignancies. In addition, the expression level of the eRNA signature is sensitive and profiled using the RNA-seq platform. The expression of eRNA signature possesses a superior prediction power for recurrence compared to other clinical risk factors. The eRNA signature was independent in predicting patient tumor relapse. Thus, our study provides novel evidence for recurrence-associated eRNAs for experimental validation in clinical cohorts and deeper investigations to better understand the CRC etiology and progression while providing a clinical application for precise risk stratification to identify those groups of patients at higher risk for CRC recurrence.

#### 5. Methods

##### 5.1. CRC Patients Datasets

The expression profiles of eRNAs for CC and RC were downloaded from the eRNA data portal in the cancer (eRic) database [25]. Here, RNA-seq expression profiles from the TCGA database were mapped to eRNA regions, followed by calculating their expression

levels as reads per million (RPM) for each patient. For CC and RC, we retrieved 477 and 460 eRNA transcripts. The datasets were  $\log_2(\text{RPM} + 1)$  normalized. The clinical information for CRC patients was downloaded from the cBioPortal database [28,29]. Only CRC patients with TNM stages 1, 2, and 3 were included. This study included 96 CC patients with CMS4 subtype and 121 RC patients for which both RNA-seq and survival information were available. The demographics of CC and RC patients are shown in Table S1.

### 5.2. Consensus Molecular Subtypes (CMS) Classification

The CMS classification for each patient with CC was performed using gene sets with Synapse ID syn4983432 (5973 genes available) downloaded from the Sage-Bionetworks (<https://sagebionetworks.org/research-projects/colorectal-cancer-subtyping-consortium-crcsc/>) (accessed on 12 July 2020). The R package *CMS classifier* was used for CMS subtype calling using a random forest classifier function with a default posterior probability of 0.5 [24]. Based on the similarity of gene expression profiles in this method, all CC patients were classified into four subtypes CMS1—CMS4. All CC patients with mixed features or intra-tumoral heterogeneity were classified as mixed subtypes.

### 5.3. Identification of Prognostic eRNAs Associated with Tumor Relapse

We randomized 70% of the entire cohort as a training set and the rest 30% as a testing set. The distribution of disease-free and relapse-free patient populations in training and testing sets was comparable. Since clinical information, including patient age, gender, and tumor stage, can affect patient survival, we accordingly examined the prognostic potential of each eRNA using multivariate Cox regression analysis in the training set. We used  $z\text{score} > 1.96$  (which corresponds to  $p < 0.05$ ) as a cut-off threshold to select prognostic eRNAs associated with disease-free survival (DFS). We used the LASSO Cox regression with 10-fold cross-validation to identify a subset of prognostic eRNAs and establish an eRNA signature. The eRNA signature's prognostic potential was evaluated in the testing set. We repeated the above split cohort approach into a training set, and a testing set 100 times, which generated 100 different sets of eRNA signatures.

### 5.4. Risk Score Calculation

During each iteration, the risk score for each patient was calculated using a linear combination of expression and coefficients from multivariate Cox regression analysis as follows:

$$\text{Risk score} = \sum_{j=1}^n e_j \times \text{exp}_{ij},$$

where  $e_j$  is the multivariate coefficient of eRNA  $j$ ,  $\text{exp}_{ij}$  is the expression value of eRNA  $j$  in patient  $i$ , and  $n$  is the number of prognostic eRNAs (selected after lasso Cox analysis). To calculate the risk score of the prognostic eRNAs chosen from all the successful survival models (which significantly stratify the patients into two groups in the testing set), we used median Cox coefficients obtained from lasso Cox analysis.

### 5.5. Statistical Analysis

Kaplan–Meier survival analysis (low-risk score groups vs. high-risk score groups) with Mantel log-rank test was performed to assess differences between the two survival probabilities. Multivariable Cox proportional hazard regression was performed to determine the prognostic independence of risk scores and clinical risk factors. ROC curve and AUC analyses were performed to evaluate the sensitivity and specificity of eRNA risk scores for predicting DFS. Survival analysis was performed using the *survival* R package [30]. LASSO Cox proportional hazard regression analysis was performed using the *glmnet* R package [31]. ROC curve and AUC analysis were performed using the *pROC* R package [32], and plots were generated using the *ggplot2* R package [33].

### 5.6. Gene Set Enrichment Analysis

Spearman correlation coefficient (SCC) was calculated between prognostic eRNAs and whole genome protein-coding genes (PCGs) in CC and RC patients, respectively. To assess the putative biological functions of each eRNA, gene set enrichment analysis (GSEA) was performed using MSigDB (C5.bp.v7.1.symbols.gmt gene set collection; 7530 gene set available), ranked list of eRNAs correlated genes and their SCC, maximum gene set size of 5000, minimum gene set size of 15, weighted enrichment statistics, and 1000 permutations. GSEA analysis was performed using GSEA pre-ranked tool [34]. Over-represented gene sets (FDR  $q$ -value 0.05 and overlap coefficient of 0.5) were filtered and visualized using an Enrichment map-a Cytoscape plug-in [35].

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/genes14010137/s1>, Figure S1: Demonstration of identification of prognostic eRNA signatures in failed model and successful model in the testing set; Figure S2: Receiver operating characteristics curve analyses of colon cancer recurrence prediction by the enhancer RNA signature and protein-coding gene signature.; Figure S3: Spearman correlation analysis of the enhancer RNA signature (eRNA); Table S1: Clinical characteristics of TCGA colorectal cancer patients in the entire training and testing cohort.

**Author Contributions:** D.S. and A.G. conceived and designed the study. D.S. developed the computational methodology and carried out the analysis. D.S., C.-C.L. and A.G. helped with data analysis and interpreted the result. D.S. drafted the manuscript. D.S., C.-C.L. and A.G. revised the manuscript. A.G. supervised the study. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by CA184792, CA187956, CA227602, CA072851, and CA202797 grants from the National Cancer Institute, National Institutes of Health to A.G.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The eRNA expression dataset used in this study can be found in the eRIC database <https://hanlab.uth.edu/eRic/> (accessed on 15 May 2020). The clinical information for CC and RC can be downloaded from the cBioPortal database <https://www.cbioportal.org> (accessed on 6 May 2020).

**Acknowledgments:** We thank Hsuan-Cheng Huang at the Institute of Biomedical Informatics, National Yang Ming Chiao Tung University, Taipei, Taiwan, for helpful discussions.

**Conflicts of Interest:** The authors declare no conflict of interest.

### Abbreviations

CRC: Colorectal cancer; eRNA: Enhancer RNA; CC: Colon cancer; RC: Rectal cancer; LASSO: Least Absolute Shrinkage and Selection Operator; AUC: Area under the curve; CMS: Consensus molecular subtype; DFS: Disease-free survival; ROC: Receiver operating characteristics; CIMP: CpG island methylator phenotype; MSI: Microsatellite instability.

### References

1. Sung, H.; Ferlay, J.; Siegel, R.L.; Laversanne, M.; Soerjomataram, I.; Jemal, A.; Bray, F. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **2021**, *71*, 209–249. [[CrossRef](#)] [[PubMed](#)]
2. American Cancer Society. *Cancer Facts & Figures 2022*; American Cancer Society: Atlanta, GA, USA, 2022.
3. Blank, A.; Roberts, D.E.I.; Dawson, H.; Zlobec, I.; Lugli, A. Tumor heterogeneity in primary colorectal cancer and corresponding metastases. Does the apple fall far from the tree? *Front. Med.* **2018**, *5*, 234. [[CrossRef](#)] [[PubMed](#)]
4. Hong, T.S.; Clark, J.W.; Haigis, K.M. Cancers of the colon and rectum: Identical or fraternal twins? *Cancer Discov. AACR* **2012**, *2*, 117–121. [[CrossRef](#)]

5. Marks, K.; West, N.; Morris, E.; Quirke, P. Clinicopathological, genomic and immunological factors in colorectal cancer prognosis. *J. Br. Surg.* **2018**, *105*, e99–e109. [[CrossRef](#)]
6. Dienstmann, R.; Mason, M.J.; Sinicrope, F.A.; Phipps, A.I.; Tejpar, S.; Nesbakken, A.; Danielsen, S.A.; Sveen, A.; Buchanan, D.D.; Clendenning, M.; et al. Prediction of overall survival in stage II and III colon cancer beyond TNM system: A retrospective, pooled biomarker study. *Ann. Oncol.* **2017**, *28*, 1023–1031. [[CrossRef](#)] [[PubMed](#)]
7. Moccia, F.; Tolone, S.; Allaria, A.; Napolitano, V.; Ilaria, F.; Ottavia, M.; Cesaro, E.; Docimo, L.; Fei, L. Lymph node ratio versus TNM system as a prognostic factor in colorectal cancer staging. A Single Center experience. *Open Med.* **2019**, *14*, 523–531. [[CrossRef](#)] [[PubMed](#)]
8. Oh, S.C.; Park, Y.-Y.; Park, E.S.; Lim, J.Y.; Kim, S.M.; Kim, S.-B.; Kim, J.; Kim, S.C.; Chu, I.-S.; Smith, J.J. Prognostic gene expression signature associated with two molecularly distinct subtypes of colorectal cancer. *Gut* **2012**, *61*, 1291–1298. [[CrossRef](#)] [[PubMed](#)]
9. Kandimalla, R.; Ozawa, T.; Gao, F.; Wang, X.; Goel, A.; Nozawa, H.; Hata, K.; Nagata, H.; Okada, S.; Watanabe, T. Gene expression signature in surgical tissues and endoscopic biopsies identifies high-risk T1 colorectal cancers. *Gastroenterology* **2019**, *156*, 2338–2341.e2333. [[CrossRef](#)]
10. Ye, S.-B.; Cheng, Y.-K.; Hu, J.-C.; Gao, F.; Lan, P. Development and validation of an individualized gene expression-based signature to predict overall survival in metastatic colorectal cancer. *Ann. Transl. Med.* **2020**, *8*, 96. [[CrossRef](#)]
11. Yuan, Y.; Chen, J.; Wang, J.; Xu, M.; Zhang, Y.; Sun, P.; Liang, L. Development and clinical validation of a novel 4-gene prognostic signature predicting survival in colorectal cancer. *Front. Oncol.* **2020**, *10*, 595. [[CrossRef](#)]
12. Brouwer, N.P.; Bos, A.C.; Lemmens, V.E.; Tanis, P.J.; Huguen, N.; Nagtegaal, I.D.; de Wilt, J.H.; Verhoeven, R.H. An overview of 25 years of incidence, treatment and outcome of colorectal cancer patients. *Int. J. Cancer* **2018**, *143*, 2758–2766. [[CrossRef](#)] [[PubMed](#)]
13. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature* **2004**, *431*, 931–945. [[CrossRef](#)] [[PubMed](#)]
14. Mattick, J.S.; Makunin, I.V. Non-coding RNA. *Hum. Mol. Genet.* **2006**, *15*, R17–R29. [[CrossRef](#)] [[PubMed](#)]
15. Esteller, M. Non-coding RNAs in human disease. *Nat. Rev. Genet.* **2011**, *12*, 861–874. [[CrossRef](#)] [[PubMed](#)]
16. Kim, T.-K.; Hemberg, M.; Gray, J.M.; Costa, A.M.; Bear, D.M.; Wu, J.; Harmin, D.A.; Laptewicz, M.; Barbara-Haley, K.; Kuersten, S. Widespread transcription at neuronal activity-regulated enhancers. *Nature* **2010**, *465*, 182–187. [[CrossRef](#)]
17. Lee, J.-H.; Xiong, F.; Li, W. Enhancer RNAs in cancer: Regulation, mechanisms and therapeutic potential. *RNA Biol.* **2020**, *17*, 1550–1559. [[CrossRef](#)]
18. Xiang, J.F.; Yin, Q.F.; Chen, T.; Zhang, Y.; Zhang, X.O.; Wu, Z.; Zhang, S.; Wang, H.B.; Ge, J.; Lu, X.; et al. Human colorectal cancer-specific CCAT1-L lncRNA regulates long-range chromatin interactions at the MYC locus. *Cell Res.* **2014**, *24*, 513–531. [[CrossRef](#)]
19. McClelland, M.L.; Mesh, K.; Lorenzana, E.; Chopra, V.S.; Segal, E.; Watanabe, C.; Haley, B.; Mayba, O.; Yaylaoglu, M.; Gnad, F.; et al. CCAT1 is an enhancer-templated RNA that predicts BET sensitivity in colorectal cancer. *J. Clin. Investig.* **2016**, *126*, 639–652. [[CrossRef](#)]
20. Melo, C.A.; Drost, J.; Wijchers, P.J.; van de Werken, H.; de Wit, E.; Oude Vrielink, J.A.; Elkon, R.; Melo, S.A.; Léveillé, N.; Kalluri, R.; et al. eRNAs are required for p53-dependent enhancer activity and gene transcription. *Mol. Cell* **2013**, *49*, 524–535. [[CrossRef](#)] [[PubMed](#)]
21. Gu, X.; Wang, L.; Boldrup, L.; Coates, P.J.; Fahraeus, R.; Sgaramella, N.; Wilms, T.; Nylander, K. AP001056. 1, a prognosis-related enhancer RNA in squamous cell carcinoma of the head and neck. *Cancers* **2019**, *11*, 347. [[CrossRef](#)]
22. Sartorelli, V.; Laubert, S.M. Enhancer RNAs are an important regulatory layer of the epigenome. *Nat. Struct. Mol. Biol.* **2020**, *27*, 521–528. [[CrossRef](#)] [[PubMed](#)]
23. Hsieh, C.-L.; Fei, T.; Chen, Y.; Li, T.; Gao, Y.; Wang, X.; Sun, T.; Sweeney, C.J.; Lee, G.-S.M.; Chen, S.; et al. Enhancer RNAs participate in androgen receptor-driven looping that selectively enhances gene activation. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 7319–7324. [[CrossRef](#)] [[PubMed](#)]
24. Guinney, J.; Dienstmann, R.; Wang, X.; De Reynies, A.; Schlicker, A.; Sonesson, C.; Marisa, L.; Roepman, P.; Nyamundanda, G.; Angelino, P. The consensus molecular subtypes of colorectal cancer. *Nat. Med.* **2015**, *21*, 1350–1356. [[CrossRef](#)] [[PubMed](#)]
25. Zhang, Z.; Lee, J.H.; Ruan, H.; Ye, Y.; Krakowiak, J.; Hu, Q.; Xiang, Y.; Gong, J.; Zhou, B.; Wang, L.; et al. Transcriptional landscape and clinical utility of enhancer RNAs for eRNA-targeted therapy in cancer. *Nat. Commun.* **2019**, *10*, 4562. [[CrossRef](#)]
26. Tian, X.; Zhu, X.; Yan, T.; Yu, C.; Shen, C.; Hu, Y.; Hong, J.; Chen, H.; Fang, J.Y. Recurrence-associated gene signature optimizes recurrence-free survival prediction of colorectal cancer. *Mol. Oncol.* **2017**, *11*, 1544–1560. [[CrossRef](#)]
27. Xu, G.; Zhang, M.; Zhu, H.; Xu, J. A 15-gene signature for prediction of colon cancer recurrence and prognosis based on SVM. *Gene* **2017**, *604*, 33–40. [[CrossRef](#)]
28. Cerami, E.; Gao, J.; Dogrusoz, U.; Gross, B.E.; Sumer, S.O.; Aksoy, B.A.; Jacobsen, A.; Byrne, C.J.; Heuer, M.L.; Larsson, E. The cBio cancer genomics portal: An open platform for exploring multidimensional cancer genomics data. *Cancer Discov. AACR* **2012**, *2*, 401–404. [[CrossRef](#)]
29. Gao, J.; Aksoy, B.A.; Dogrusoz, U.; Dresdner, G.; Gross, B.; Sumer, S.O.; Sun, Y.; Jacobsen, A.; Sinha, R.; Larsson, E. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.* **2013**, *6*, p11. [[CrossRef](#)] [[PubMed](#)]
30. Therneau, T. A package for survival analysis in S. *R Package Version* **2015**, *2*, 38.
31. Friedman, J.; Hastie, T.; Tibshirani, R. Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* **2010**, *33*, 1–22. [[CrossRef](#)]

32. Robin, X.; Turck, N.; Hainard, A.; Tiberti, N.; Lisacek, F.; Sanchez, J.-C.; Müller, M. pROC: An open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinform.* **2011**, *12*, 77. [[CrossRef](#)] [[PubMed](#)]
33. Hadley, W. *Ggplot2: Elegant Graphics for Data Analysis*; Springer: Berlin/Heidelberg, Germany, 2016.
34. Subramanian, A.; Tamayo, P.; Mootha, V.K.; Mukherjee, S.; Ebert, B.L.; Gillette, M.A.; Paulovich, A.; Pomeroy, S.L.; Golub, T.R.; Lander, E.S. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 15545–15550. [[CrossRef](#)] [[PubMed](#)]
35. Merico, D.; Isserlin, R.; Stueker, O.; Emili, A.; Bader, G.D. Enrichment map: A network-based method for gene-set enrichment visualization and interpretation. *PLoS ONE* **2010**, *5*, e13984. [[CrossRef](#)] [[PubMed](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.