Identification of colorectal cancer associated biomarkers: an integrated analysis of miRNA expression

André Fonseca[1], Sara Ventura Ramalhete[1,2], André Mestre[1,2], Ricardo Pires das Neves[3,4], Ana Marreiros[1,2], Pedro Castelo-Branco[1,2,5,*], Vânia Palma Roberto[1,2,6,*]

[1]Faculty of Medicine and Biomedical Sciences (FMCB), University of Algarve, Campus de Gambelas, Faro 8005-139, Portugal
[2]Algarve Biomedical Center Research Institute (ABC-RI), Faro 8005-139, Portugal
[3]CNC, Center for Neuroscience and Cell Biology, CIBB - Centre for Innovative Biomedicine and Biotechnology, University of Coimbra, Coimbra 3004-517, Portugal
[4]IIIUC-Institute of Interdisciplinary Research, University of Coimbra, Coimbra 3030-789, Portugal
[5]Champalimaud Research Program, Champalimaud Center for the Unknown, Lisbon 1400-038, Portugal
[6]Centre of Marine Sciences (CCMAR), University of Algarve, Faro 8005-139, Portugal
*Co-senior author

ABSTRACT

Colorectal cancer is one of the leading causes of cancer-related deaths worldwide. This complex disease still holds severe problems concerning diagnosis due to the high invasiveness nature of colonoscopy and the low accuracy of the alternative diagnostic methods. Additionally, patient heterogeneity even within the same stage is not properly reflected in the current stratification system. This scenario highlights the need for new biomarkers to improve non-invasive screenings and clinical management of patients.

MicroRNAs (miRNAs) have emerged as good candidate biomarkers in cancer as they are stable molecules, easily measurable and detected in body fluids thus allowing for non-invasive diagnosis and/or prognosis.

In this study, we performed an integrated analysis first using 4 different datasets (discovery cohorts) to identify miRNAs associated with colorectal cancer development, unveil their role in this disease by identifying putative targets and regulatory networks and investigate their ability to serve as biomarkers. We have identified 26 differentially expressed miRNAs which interact with frequently deregulated genes known to participate in commonly altered pathways in colorectal cancer. Most of these miRNAs have high diagnostic power, and their prognostic potential is evidenced by panels of 5 miRNAs able to predict the outcome of stage II and III colorectal cancer patients. Notably, 8 miRNAs were validated in three additional independent cohorts (validation cohorts) including a plasma cohort thus reinforcing the value of miRNAs as non-invasive biomarkers.

INTRODUCTION

Colorectal cancer (CRC) is the third most diagnosed cancer worldwide with over 1.84 million novel cases reported in 2018, preceded only by lung, and breast cancers [1, 2]. Responsible for more than 880.729

deaths (9.2% of all tumour deaths) CRC is the second deadliest cancer worldwide representing a major public health problem [1]. Approximately 30% more prominent in men than women, CRC incidence is generally low for individuals younger than 50 years old in both genders and strongly increases with age [3–5].

Currently, colonoscopy is the gold-standard test for CRC diagnosis however, due to its highly invasive nature, patients usually tend to avoid it [6–9]. This issue, associated with the fact that patients with CRC are often asymptomatic in the early stages of the disease, leads CRC to go undetected until later stages of disease when cancer has already grown or spread to other tissues [10]. As the disease progresses, patient survival rates significantly decline stressing that early identification of CRC is of particular importance [11, 12]. Therefore, in order to increase test-adherence more appealing techniques are in need. Alternative non-invasive tests, such as the faecal occult blood tests (FOBTs) and serum biomarkers as Carcinoembryonic antigen (CEA) and the Carbohydrate antigen 19-9 (CA19-9), have been proposed for CRC but, their lack of specificity and sensitivity has limited their use [3, 6, 8, 13–15]. Furthermore, CEA and CA19-9 usually only present elevated serum concentrations in cases of advanced CRC, thus limiting their diagnostic value in earlier cases [14, 15].

Accurate assessment of CRC patient prognosis and optimal treatment selection is also challenging and mainly based on histological characteristics and pathological staging [16]. According to these features, patients are stratified into one of the tumour–node–metastasis (TNM) stages that range from I (the less advanced stage of the disease) to IV (the most advanced stage), and treatment is then selected accordingly [5, 17, 18]. Nevertheless, even after stratification, optimal individual therapy and survival times are very heterogeneous for patients within the same stage, significantly hampering a quality assessment of prognosis [16]. This scenario highlights the necessity for a more precise clinical management system for CRC, especially between stages II and III where it's most difficult to properly discern patients and thus to provide accurate treatment [19]. In this sense, the identification of novel, more efficient and less invasive biomarkers may help to improve the effectiveness of both diagnosis and prognosis of CRC patients.

As a complex disease, CRC arises from the progressive accumulation of both genetic and epigenetic alterations over the course of 10 to 20 years that ultimately result in the transformation of the normal glandular epithelium into early and then advance cancer [20–22]. Some of the most known specific genetic alterations in CRC are mutations in key genes that deregulate several cellular homeostatic functions from proliferation, differentiation, adhesion, migration, and cell death to DNA stability and repair [16, 23]. Mutations occurring in the genes *APC, TP53, KRAS, TGFβ* and *PIK3CA* are some of the most observed in CRC and have often been proposed as genetic biomarkers [23, 24]. Epigenetic modifications, such as DNA methylation, post-translational histone modifications and microRNAs post-transcriptional regulation also have a key role in CRC development and complement the molecular landscape of this disease [25, 26]. Epigenetics broadly consists in regulating gene expression without altering the DNA sequence and occurs in normal tissues, being of extreme importance during embryonic development, imprinting and tissue differentiation [25]. Nevertheless, when disturbed, epigenetic mechanisms can alter cellular homeostasis favouring tumour development [25–27].

Since the discovery of microRNAs (miRNAs) in patients with chronic lymphocytic leukaemia in 2002, their role in regulating post-transcriptional gene expression in cancer development, progression and metastasis has been well-established in the field [28–37]. MiRNAs are small non-coding RNAs, of approximately 20-22 nucleotides, that are processed from larger transcripts and exert a fine-tuning regulation by binding to the 3'-untranslated region (UTR) of their targets messenger RNA (mRNA) in the cytoplasm [25, 38, 39]. One single miRNA can regulate hundreds of target genes primarily by inducing translational repression, but they can also lead to mRNA cleavage and consequent decay [38, 39]. In both situations, the interaction of miRNAs with their target mRNAs results in a lower expression of the protein levels [38, 39]. They are highly stable molecules, with cell and tissue relative high specificity and easily measurable in different biological samples from tissues to saliva, serum and faeces [38, 40–43]. Due to their features, miRNAs have emerged as good non-invasive biomarker candidates for the detection of pre-cancerous/cancerous lesions as well as therapeutic targets [41–47]. Of notice, technological advances allowing large-scale miRNA profiling and miRNA regulatory network analysis have been crucial to understanding the role of miRNAs in pathological contexts [48–50]. Accordingly, several therapeutical strategies relying on mimics and antimiRs are currently in clinical trials as is the case of MRX34 (miR-34 mimic), in a study for the treatment of patients with advanced solid tumours including CRC [41, 51]. Although several miRNAs have been reported as promising cancer biomarkers [37, 42, 43, 45, 52–57] only a few are available to clinicians as the miRNA panel ThyraMIR® (Interpace Biosciences, Inc.®) for thyroid cancer diagnosis. Regarding CRC, despite the advances in the field and the number of miRNAs identified as potential biomarkers, their progress to the clinic has been compromised in part by contradictory results of independent studies and lack of consistency between biomarker panels [13, 58].

In this sense, here we propose to identify and scrutinize panels of miRNAs with the ability to behave as diagnostic and/or prognostic tools for CRC by compiling data from different studies to assure a bona fide data analysis and more concise results. This approach revealed novel miRNA panels able to accurately distinguish malignant from normal tissue, with high sensitivity and specificity, and to predict patient outcome, particularly in stages II and III. Our findings highlight promising candidate biomarkers to improve colorectal cancer clinical management.

RESULTS

MicroRNAs are promising candidates for CRC screening and management. Since one biomarker is not enough to diagnose disease, a group of markers should increase sensitivity and specificity [59–61]. In that sense, here we aim to discover miRNAs associated with CRC development and/or progression and evaluate their combined diagnostic and prognostic values as panels of miRNAs.

MiRNA expression dataset selection

In order to perform an integrated analysis, we combined data from 4 datasets (discovery cohorts), obtained through Illumina platforms: the TCGA colon and cancer cohorts (TCGA-COAD/READ, from now on here referred as the TCGA cohort or dataset) and 3 GEO datasets that fitted our inclusion criteria (Table 1). TCGA is the most complete dataset used in our study, with numerous multi-omics data and patient clinical history and therefore was the dataset used for the prognostic analysis. The TCGA also comprised the largest number of miRNAs with a total number of 2164, followed by GSE33125 and GSE30454 with 878 each and GSE18392 with 510 miRNAs (Table 1 and Figure 1). In total, our bioinformatics analyses were carried out on 588 samples of which 69 were normal tissue and 519 tumour samples of colorectal carcinoma. Table 1 summarizes a detailed description of the discovery 4 datasets.

Identification of differently expressed miRNAs in CRC

To normalize our analysis, the 3 datasets from GEO were transformed and expressed as $\log_2$ values to exhibit the same scale value as the TCGA dataset. From there all datasets were processed equally. We started by applying listwise deletion and outlier removal followed by a process of statistical analysis. All steps of this analysis resulted in the elimination of miRNAs as resumed in Figure 1. Proportionally, the GSE33125 dataset reported the biggest miRNA loss, with almost 86% of all miRNAs being excluded, and a total of 125 miRNAs remaining for

further analysis. GSE30454, GSE18392 and TCGA exhibited losses close to 37%, 12% and 9% respectively, totalizing 459, 451 and 449 miRNAs remaining in each dataset (Figure 1). We then proceeded to identify which miRNAs were present in at least 2 of the 3 GEO datasets while simultaneously being present in the TCGA. A total of 42 miRNAs fulfilled our criteria. From these, 31 miRNAs were simultaneously found in 3 datasets: 6 in TCGA, GSE33125 and GSE18392, 23 in TCGA, GSE30454 and GSE18392 and 2 in TCGA, GSE30454 and GSE33125 (Figure 2). Notably, 11 miRNAs were found in all 4 datasets at the same time (Figure 2).

Finally, we calculated the $\log_2$FC between tumour and normal tissue samples for the identified 42 miRNAs. Since this study only involves human samples, the name of each miRNA from now on will not include the prefix related to the species (hsa-).

From these, 26 miRNAs presented lower expression levels in colorectal cancer tissues when compared to normal tissues and exhibited the same expression pattern in all the datasets where they were present: miR-139-5p, miR-125a-5p, miR-193a-5p, miR-484, miR-486-5p, miR-28-3p, miR-342-3p, miR-129-5p, miR-299-5p, miR-296-5p, miR-326, miR-324-5p, miR-339-5p, miR-133b, miR-127-3p, miR-331-3p, miR-324-3p, miR-423-3p, miR-490-3p, miR-502-3p, miR-574-3p, miR-330-3p, miR-320d, miR-320b, miR-375 and miR-320a (Figure 3). The remaining 16 miRNAs (miR-491-5p, miR-654-5p, miR-485-3p, miR-632, miR-76, miR-769-3p, miR-28-5p, miR-140-5p, miR-338-3p, miR-107, miR-199b-5p, miR-429, miR-501-5p, miR-582-5p, miR-142-3p and miR-1), exhibited opposite expression values in at least one dataset in which they were present and thus no regulation pattern could be perceived (Figure 3). Therefore, these miRNAs were excluded from the subsequent analyses.

Our results also point out different magnitudes of the $\log_2$(FC) values across datasets. While the TCGA dataset $\log_2$(FC) values ranged from less than -2 to almost 2, for downregulated and upregulated miRNAs respectively, in the remaining datasets the $\log_2$(FC) values were never below -1 or higher than 1 (Figure 3). Interestingly, the miRNA with the highest deregulation pattern overall (miR-129-5p) and the most downregulated in TCGA with a $\log_2$(FC) value of -2, is reported as a frequently downregulated miRNA in CRC, supporting our findings [62, 63].

### Gene target analysis of selected miRNAs

To better understand the role of the 26 downregulated miRNAs in CRC, we used the MirTarBase to identify functional miRNA-target interactions (MTIs). Our

Table 1. Detailed patient information for colorectal cancer datasets used in miRNA expression analysis.

Datasets	TCGA	GSE33125	GSE18392	GSE30454
Number of miRNAs available	2164	878	510	878
Patient nationality	USA	Italy	USA	Spain
Platform	Illumina HiSeq miRNA Seq	Illumina Human v2 MicroRNA expression beadchip	Illumina Human v1 MicroRNA expression beadchip	Illumina Human v2 MicroRNA expression beadchip
Tissue type	Tumour and normal tissue	Paired tumour and normal tissue	Tumour and healthy tissue	Tumour and healthy tissue
Number of samples				
Tumour	340 (97%)	9 (50%)	116 (80%)	20 (27%)
Normal	11 (3%)	9 (50%)	29 (20%)	54 (73%)
Age (mean ± SD[1] years)				
Tumour sample patients	65 ± 13	-	-	59 ± 16
Normal sample patients	68 ± 19	-	-	64 ± 16
Gender				
Female	163 (46%)	-	-	43 (58%)
Male	185 (53%)	-	-	31 (42%)
Unknown	3 (1%)	18 (100%)	145 (100%)	-
Tumour Site				
Colon	259 (74%)	18 (100%)	145 (100%)	74 (100%)
Rectum	92 (26%)	-	-	-

1 – standard deviation.
"-"– inexistent or unspecified.
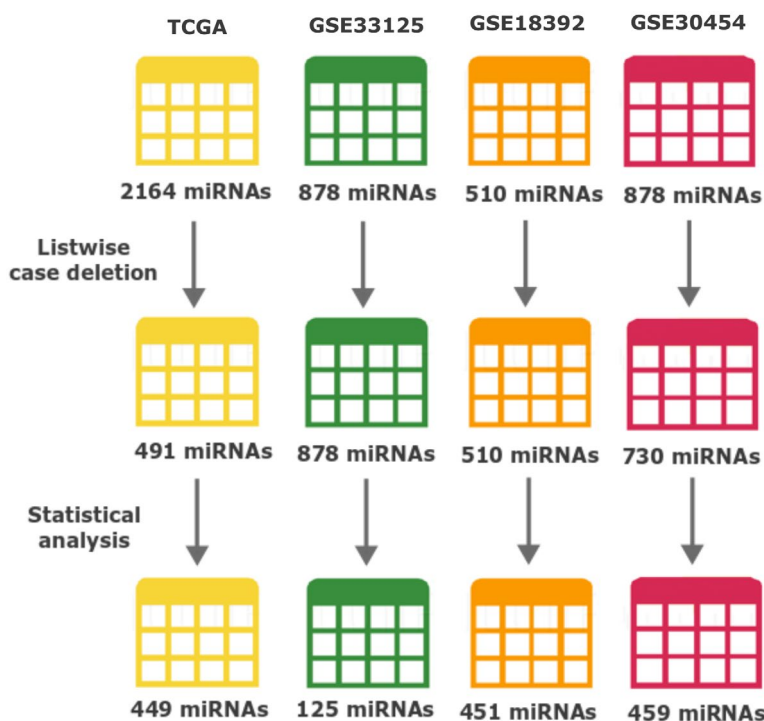
Figure 1. Number of miRNAs throughout data processing and statistical analysis. The number of miRNAs after each major event (Listwise case deletion and Statistical analysis) is shown below each dataset. Different colours represent: yellow TCGA, green GSE33215, orange GSE18392 and reddish-purple GSE30454. From the top, the numbers of miRNAs represent: the initially number of miRNAs present in each dataset before any processing step; the number of miRNAs after removing the ones with more than 50% information missing (modified listwise case deletion); and the number of miRNAs after performing all the statistics and used to perform the identification of differently expressed miRNAs across datasets.

analysis identified a total of 309 different human target genes previously experimentally validated (either by report assay, western blot or qPCR) in a total of 384 papers (Supplementary Table 1). Among the target genes identified, *ERBB2* was the most prevalent, described in 12 different papers as regulated by miR-125a-5p, miR-331-3p, miR-375 and miR-193a-5p. *MCL1* was the second most frequent gene being reported in 9 papers as modulated by miR-133b, miR-320a, miR-139-5p and miR-125a-5p. *IGF1R* and *SP1* genes followed, both cited in 8 occasions (Supplementary Table 1).

Then, to understand which of the 26 miRNAs could be playing a role in the development of CRC, the 309 targets identified in the MTIs were screened in the CoReCG database which highlighted 173 genes previously associated to CRC (Figure 4 and Supplementary Figure 1 and Supplementary Table 2). From our set of 26 miRNAs, all were found to interact with at least one gene involved in CRC except for miR-502-3p. Among the remaining 25 miRNAs, miR-125a-5p a well-known deregulated miRNA in CRC [37] presented the highest number of interactions with CRC associated genes, interacting with a total of 44 genes among which were *AKT1*, *EGFR*, *TP53*, *VEGFA*, *SMAD2* and *SMAD4* (Figure 4 and Supplementary Table 2). MiR-133b which has also been reported as showing an altered expression in

CRC [57], interacted with 38 genes including *AKT1*, *EGFR*, *MET*, *FGFR1* among other frequently altered genes in CRC. MiR-320a showed a total of 35 interactions, proceeded by miR-375 with 34, miR-139-5p with 24 and miR-129-5p with 22 interactions. Altogether these miRNAs were the ones with the highest number of interactions with commonly altered genes in CRC suggesting their potential contribution to the development of this disease (Figure 4, Supplementary Figure 1 and Supplementary Table 2). Besides the genes above mentioned other genes such as *APC*, *PIK3CA*, *PTEN*, *RUNX3* and *CTNNB1*, are some additional examples of the well-known altered genes in CRC targeted by some of the 25 miRNAs identified [25, 64–69] (Figure 4 and Supplementary Table 2). MiR-125a-5p, miR-375 and miR-28-3p were found to interact with *TP53*. Simultaneously, miR-125a-5p was also found to regulate both *SMAD2* and *SMAD4*. *SMAD4* was also a target gene of miR-574-3p. *PTEN* was a target of miR-486-5p and miR-320a, which also regulates *CTNNB1*. MiR-375, besides interacting with *TP53* also regulates *CTNBB1*, *RUNX3* and *PIK3CA*. *PIK3CA* was also found to be targeted by miR-139-5p. Finally, our results demonstrate that *APC* is a target of miR-129-5p (Figure 4, Supplementary Figure 1 and Supplementary Table 2). These results suggest that miRNAs may play a role in the development or progression of CRC through the regulation of key genes found frequently altered or mutated in CRC.
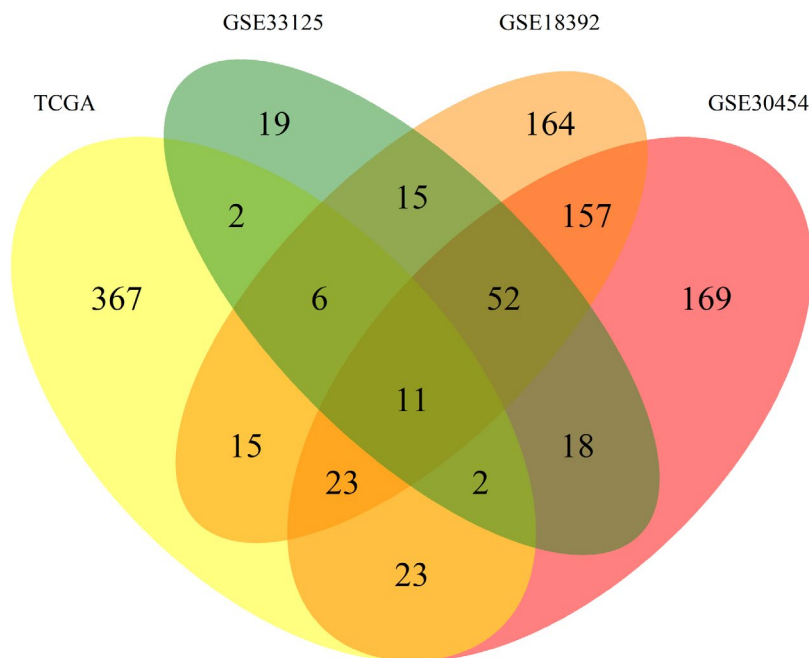
Figure 2. Venn diagram of the differently expressed miRNAs between datasets. Allocation of the 1043 differently expressed miRNAs found between the 4 datasets used in this work. Each dataset is represented by a colour, TCGA (yellow), GSE30545 (reddish purple), GSE33215 (green) and GSE18392 (orange). The number in each overlay of datasets represents the common miRNAs between those datasets.
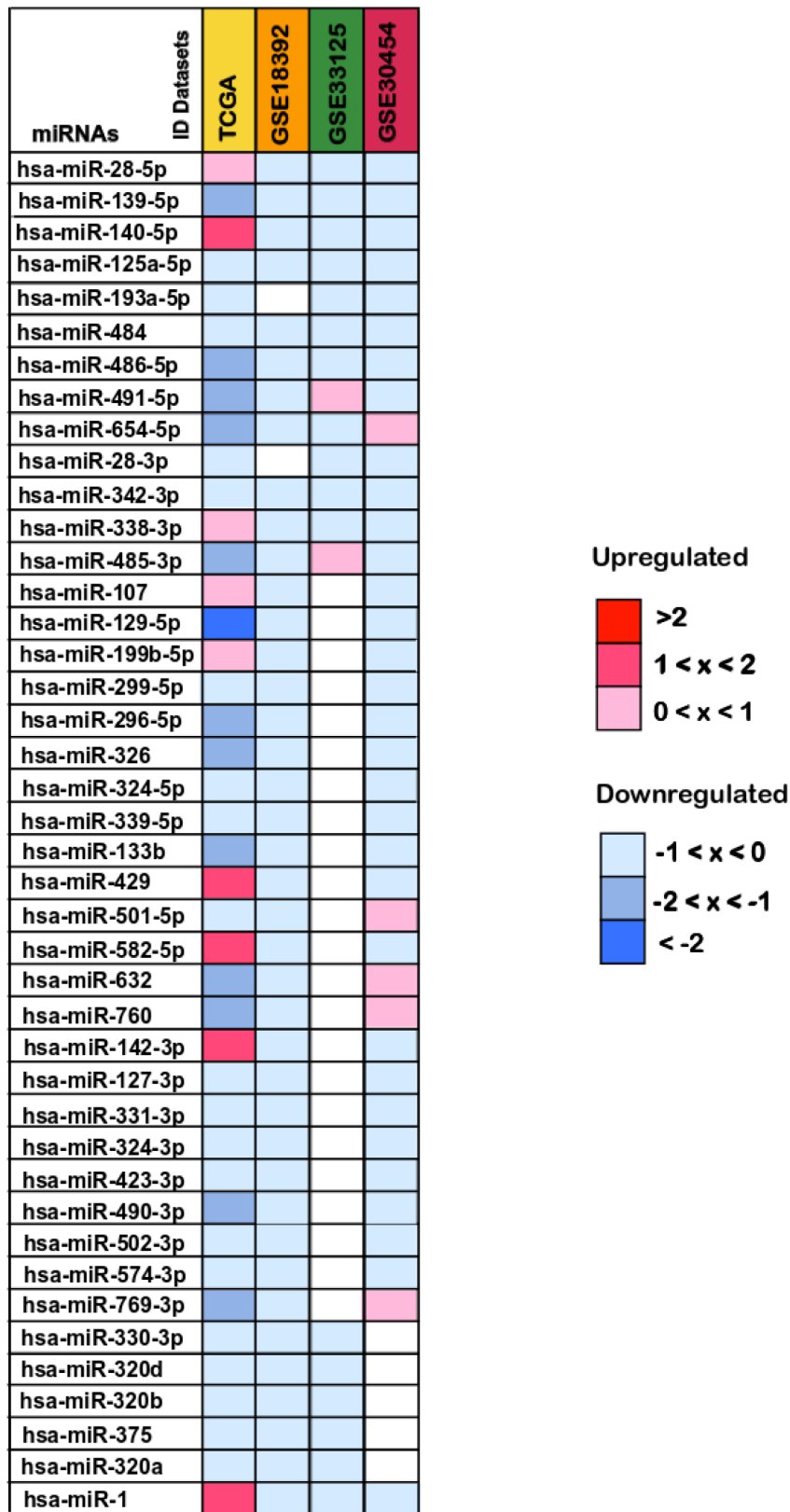
Figure 3. Differentially expressed miRNAs between colorectal cancer and normal tissue samples. The log$_2$(FC) values calculated for each dataset are reported with red scale boxes for upregulated miRNAs and blue scale boxes for the downregulated miRNAs. White boxes represent the inexistence of the miRNA on the dataset. Only the miRNAs differentially expressed in at least 2 datasets while simultaneously being present in the TCGA dataset are displayed.

To get more insights into the miRNA regulation in the specific context of CRC, we investigated the 173 genes expression patterns and its association with the 25 miRNAs expression profile, in the TCGA dataset. We analysed 172 genes since information for PKM was missing. We found 76 upregulated genes and 96 downregulated genes in tumour samples (Figure 5 and Supplementary Table 3). *NTRK3*, *SFRP1*, *ABCG2* and *WNT1* with log$_2$(FC) values of approximately -1.28, -1.22, -1.05 and -1.03 respectively (Figure 5) were the genes with the lowest expression values. Oppositely *MMP13* (log$_2$(FC) ≈ 2.80) and *WT1* (log$_2$(FC) ≈ 2.16)
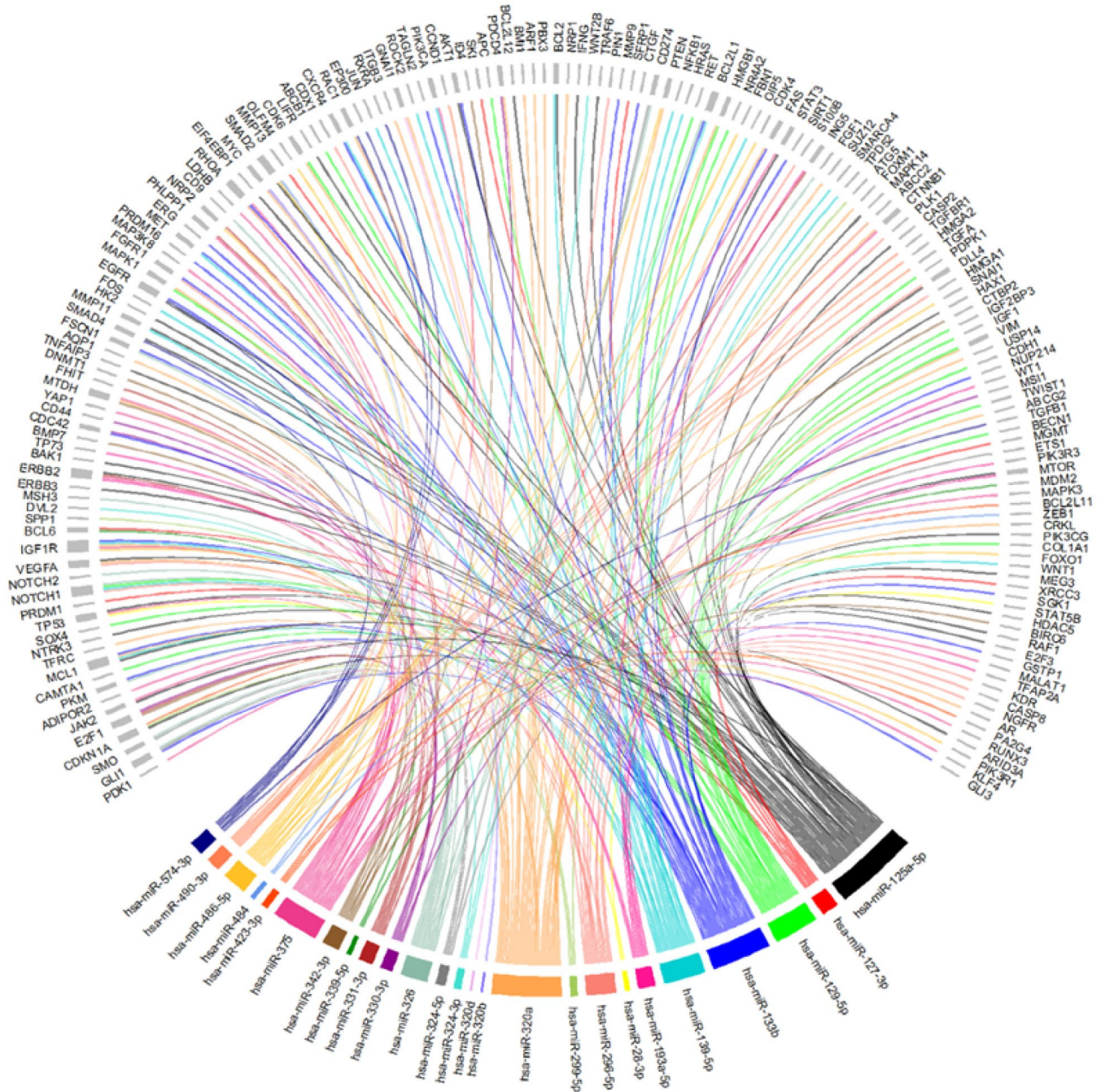
Figure 4. Chord-dendrogram of the interactions between differently expressed miRNAs and altered genes in CRC. The interactions between the 25 differently expressed miRNAs identified and the 173 experimentally validated target genes were obtained through MirTarBase. Target genes' role in CRC is according to CoReCG database. All targets are represented in grey, while each miRNA is represented by one colour. MiRNA:target interactions are represented by a line of the same colour of the respective miRNA. The size of rectangle next to the name of the miRNAs and target genes is proportional to the number of interactions they perform.

were the most upregulated genes followed by *ABCC2* ($\log_2$(FC) $\approx$ 0.85) and *TFAP2A* ($\log_2$(FC) $\approx$ 0.71) (Figure 5 and Supplementary Table 3). Correlation analysis between the expression of the 25 miRNA expression levels and their targets revealed that 104 miRNA-target genes interactions were statistically significant. From the 42 negative correlations, 10 had a *p-value* < 0.0001 while for the 62 positive interactions a *p-value* < 0.0001 was found in 25 associations. These 35 associations also presented the highest correlation coefficients, ranging from 0.215 to 0.478 in absolute value (Supplementary Table 4). The strongest negative interaction identified was miR-375:*YAP1* (correlation coefficient = -0.478; *p-value* = 4.99×10^{-20}), followed by miR-484:*ZEB1* (correlation coefficient = -0.381; *p-value* = 1.17×10^{-12}) and by miR-375:*MAP3K8* (correlation coefficient = -0.307; *p-value* = 1.17×10^{-8}). On the other hand, the strongest positive interaction was identified between miR-375:*KLF4* (correlation coefficient = 0.427; *p-value* = 7.62×10^{-16}), followed by the interactions miR-133b:GLI3 (correlation coefficient = 0,417; *p-value* = 5.22×10^{-14}) and miR-133b:*FGFR1* (correlation coefficient = 0,386; *p-value* = 4.42×10^{-12}). MiR-133 also interacted with *ERG* (correlation coefficient = 0.366; *p-value* = 6.89×10^{-11}) and *GLI1* (correlation coefficient = 0.311; *p-value* = 4.14×10^{-8}). MiR-125a-5p which we found to target the highest number of CRC altered genes, displayed interactions with *MEG3* (correlation coefficient = 0.373; *p-value* = 5.86×10^{-12}) and *HDAC5* (correlation coefficient = 0.302; *p-value* = 3.87×10^{-8}). MiR-129-5p presented an interaction with *IGF1* (correlation coefficient = 0.312; *p-value* = 1.70×10^{-8}) (Supplementary Table 4).

Altogether, these results suggest that miR-125a-5p, miR-133b, miR-375 and miR-129-5p, which were among the miRNAs with highest number of CRC-related targets, may play a key role in the altered gene expression observed in CRC.

Pathway analysis of the miRNAs target genes

To further understand the role of the 25 differentially expressed miRNAs in CRC development we investigated the molecular mechanisms of their 173 target genes. An enrichment pathway analysis using KEEG at DAVID database revealed a total of 104 pathways (*p-value* $\leq$ 0.05, Supplementary Table 5). The results pointed to the key role of miRNAs in cancer since pathways such as "Pathways in Cancer" came up on the top of our list (*p-value* = 6.44×10^{-49}; FDR of 7.98×10^{-46}), followed by "Prostate cancer" in the 3rd position (*p-value* = 1.53×10^{-28}; FDR = 1.9×10^{-25}), "Pancreatic cancer" in the 4th position (*p-value* = 3.2×10^{-28}; FDR = 4.0×10^{-25}) and "MicroRNAs in Cancer" in the 8th position (*p-value* = 9.3×10^{-23}; FDR 1.2×10^{-19}) (Supplementary Table 5), evidenced the

enrolment of these miRNAs in cancer. The "Colorectal cancer" pathway (*p-value* = 1.4×10^{-22}; FDR =1.7×10^{-19}, Supplementary Table 5) was in the 10th position, stressing that the 25 miRNAs might play a role in CRC.

Furthermore, several pathways usually found altered in CRC came up in our analysis such as "PI3K-Akt signalling pathway" (*p-value* = 1.2×10^{-21}; FDR = 1.9×10^{-20}), "RAS signalling pathway" (*p-value* = 8.1×10^{-11}; FDR = 4.5×10^{-10}), "TP53 signalling pathway" (*p-value* = 7.1×10^{-7}; FDR = 2.3×10^{-6}), "TGF-beta signalling pathway" (*p-value* = 7.6×10^{-7}; FDR = 2.4×10^{-6}), "WNT signalling pathway" (*p-value* = 8.5×10^{-8} and FDR = 3.0×10^{-7} and "MAPK signalling pathway*"* (*p-value* = 2.39×10^{-08}; FDR = 2.96×10^{-05}). Additionally, cellular functions often aberrant in colorectal cancer such as "Apoptosis" (*p-value* = 3.4×10^{-7}; FDR = 1.2×10^{-6}) and "Cell cycle" (*p-value* = 9.9×10^{-7}; FDR = 3.0×10^{-6}) were found in our analysis (Supplementary Table 5). Together these results support the association between the 25 deregulated miRNAs and several alterations often reported in CRC.

We then aimed to understand in which frequently altered pathways in CRC (TP53, PI3K-AKT, RAS, TGF-Beta, WNT and MAPK signalling pathways) each miRNA was intervening [70–75]. For that, we crossed the miRNA-target genes found to be involved in CRC with the genes intervening in each of the signalling pathways. Our results show that most of the miRNAs were involved in at least one of these pathways, except for miR-320d (Figure 6). More interestingly, we found that several miRNAs seem to intervene in more than one pathway, and 5 miRNAs (miR-125a-5p, miR-320a, miR-375, miR-133b and miR-490-3p) interact with all the six pathways. Not surprisingly, apart from miR-490-3p these miRNAs represent the ones targeting the highest number of genes involved in CRC. Therefore, by targeting multiple genes that are constituents in these pathways it's plausible that altered expression of these miRNAs may contribute to the aberration of frequently altered pathways in CRC (Figures 4–6). These results support the hypothesis that aberrant miRNA expression may contribute for the occurrence of several alteration events observed during CRC.

Identification of potential diagnostic biomarkers for colorectal cancer

Given the potential key role in CRC of the 25 miRNAs identified in this study, we then investigated their diagnostic value. For that, Receiver Operating Characteristic (ROC) curve analysis was performed in all the datasets for the 25 differently expressed miRNAs and the Area Under the Curve (AUC) values were obtained in order to determine their ability to
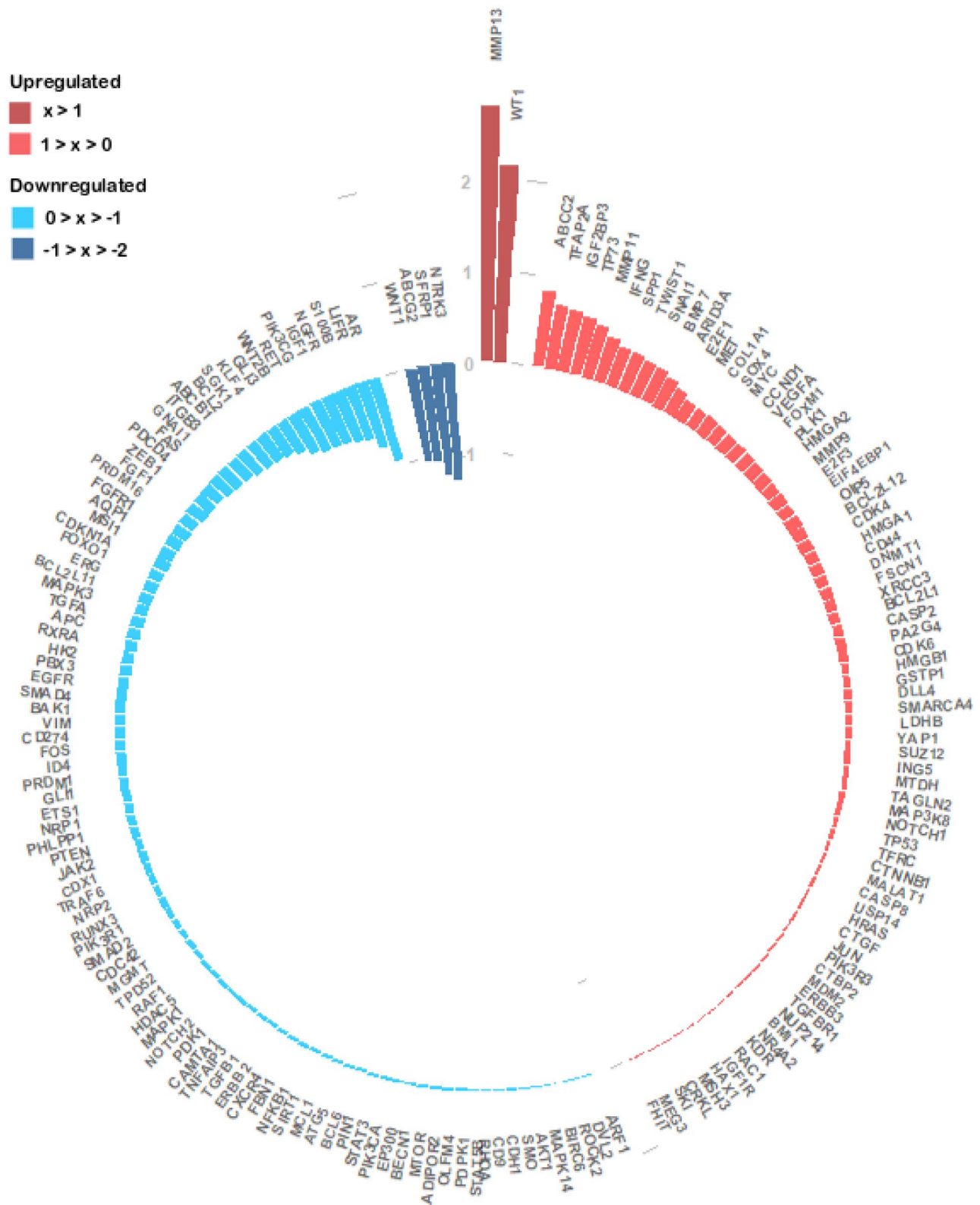
Figure 5. Circular barplot evidencing the expression of the 172 experimentally validated miRNA target genes. The expression values of the 172 genes were obtained from the TCGA colon and rectal cohorts and the $\log_2$(FC) values between the primary tumour and normal tissue samples were calculated. The $\log_2$(FC) value of each gene is given by the length and tone of each coloured bar in accordance with each gene regulation status (red colours – upregulated; blue colours- downregulated).

distinguish between colorectal cancer tissue and normal tissue. The diagnostic value of each miRNA was here characterized according to the classification by Greiner and colleagues as: less accurate ($0.5 < \text{AUC} \leq 0.7$), moderately accurate ($0.7 < \text{AUC} \leq 0.9$), highly accurate ($0.9 < \text{AUC} < 1$) and perfect diagnostic biomarker (AUC =1) [76]. Our analysis revealed that 23 out of the 25 miRNAs could be considered highly accurate ($0.9 < \text{AUC} < 1$) or better (AUC =1) in at least one of the cohorts they were present, except for miR-296-5p and miR-127-3p which were found to be moderately accurate in all the datasets they were present (Figure 7 and Supplementary Table 6).

Besides, the miRNAs miR-28-3p, miR-193a-5p, miR-299-5p, miR-320a, miR-320b, miR-324-3p, miR-326, miR-339-5p, miR-486-5p and miR-574-3p had perfect AUC values of diagnostic biomarker (AUC =1) in the TCGA dataset, and miR-125a-5p was considered perfect diagnostic tool (AUC =1) in the GEO GSE33125 dataset. Our results indicate that these 11 miRNAs could help to differentiate tumour from normal tissue with 100% sensibility and 100% specificity (Figure 7 and Supplementary Table 6). Thus, these deregulated miRNAs can be powerful diagnostic tools for an early CRC detection, an observation also corroborated by other groups [36, 56, 57, 77–79].

Figure 6. KEGG pathway analysis – interaction between the 25 differently expressed miRNAs and the frequently altered pathways in colorectal cancer. The 173 miRNA-target genes were used to perform the enrichment pathway analysis using KEEG at DAVID database. The miRNA-target genes found to be involved in CRC were crossed with the genes intervening in each signalling pathway in order to identify the miRNAs affecting each pathway. The miRNAs interaction with each KEGG Pathway is reported in pink.
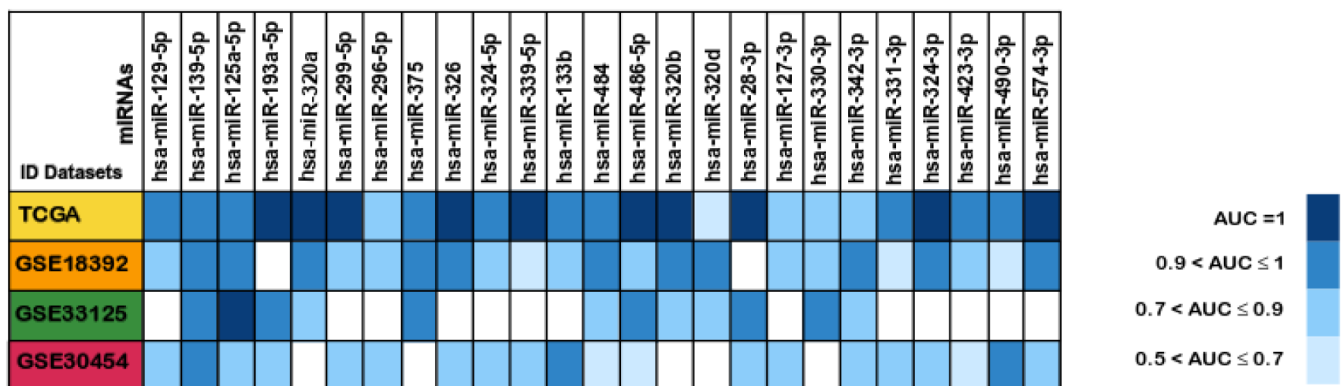
Figure 7. Areas under the ROC curve (AUC) of the 25 differentially expressed miRNAs between colorectal cancer and normal tissues samples. The miRNAs AUC values in each of the datasets TCGA (Yellow), GSE18392 (orange), GSE33125 (Green) and GSE30454 (reddish purple) are reported as blue scale boxes. MiRNAs with AUC = 1 were considered perfect diagnostic biomarkers, 0.9 < AUC < 1 highly accurate, 0.7 < AUC ≤ 0.9 moderately accurate and 0.5 < AUC ≤ 0.7 less accurate [76].

Identification of potential prognostic biomarkers for colorectal cancer

The current TNM staging system lacks the ability to accurately predict patient outcome, leading to suboptimal clinical management [17]. In this sense, we proceeded to investigate if the 25 differentially expressed miRNAs could differentiate patients within each stage of disease. To perform these analyses, Kaplan-Meier curves for Overall Survival (OS) and Recurrence Free Survival (RFS) were plotted using the median expression of each miRNA to form two distinct groups ("higher expression" group and "lower expression" group) with a minimum of 30 patients. Our analysis revealed that only miR-133b (*p-value* = 0.047) and miR-129-5p (*p-value* = 0.021), for stages III and IV respectively, could be potential solo OS biomarkers (Figure 8A, 8B). While a lower expression of miR-133b was indicative of a better outcome for patients in stage III, higher levels of miR-129-5p in stage IV patients seems to relate to extended overall survival (Figure 8A, 8B). This highlights the potential of miR-133b and miR-129-5p as individual OS biomarkers in CRC stages III and IV, respectively. On the other hand, none of the selected 25 miRNAs was able to perform as a good predictive biomarker for RFS. Nevertheless, it is highly unlikely that a single miRNA could provide a strong and precise prognostic tool. Instead, the combination of a group of miRNAs should be a more powerful clinical tool for patient management.

In this context, we analysed the potential of the 25 miRNAs to distinguish patients with different outcomes in combinations of up to 5 miRNAs. A maximum of 5 miRNAs was set in order to assure a minimum of 30 patients when performing the analysis, as panels with more miRNAs reduced the number of patients to less than 30. These analyses were performed across all stages of CRC, however the results from stages I and IV were excluded due to the small number of patients (below 30). Nevertheless, with this approach we could identify one panel of 5 miRNAs able to predict the outcome of patients in stage II regarding OS. In this panel, composed by miR-320b - miR-326 - miR-331-3p - miR-339-5p - miR-484 (*p-value* = 0.032; hazard ratio (HR) = 5.23), the group of patients with lower miRNA expression had a better outcome with ~5 times more probability to be alive than the patients in the group with higher miRNAs expression (Figure 9A and Supplementary Table 7). Moreover, by combining less than 5 miRNAs we could also identify 27 panels of 4 miRNAs, 54 panels of 3 miRNAs and 6 panels of 2 miRNAs capable of predicting the overall survival of stage II patients in CRC (Supplementary Table 7).

Regarding RFS analysis, we could identify two panels of 5 miRNAs with prognostic value. The best panel was the one composed by miR-320a - miR-324-5p - miR-324-3p - miR-423-3p - miR-484 in which the group of patients with higher combined expression had a better outcome than the lower expression group (*p-value* = 0.043) (Figure 9B and Supplementary Table 8). Likewise, for the panel miR-320a - miR-320b - miR-324-3p - miR-423-3p - miR-484 the group of patients with higher miRNAs expression had a better outcome than the lower expression group (*p-value* = 0.049) (Supplementary Table 8). Additionally, we identified 1 panel of 4 miRNAs, 3 panels of 3 miRNAs and 3 panels of 2 miRNAs capable of predicting RFS for stage II patients in CRC (Supplementary Table 8).

Regarding Stage III patients, we have identified one panel with potential OS value: miR-324-5p - miR-324-3p - miR-331-3p - miR-484 - miR-486-5p (*p-value* = 0.020; HR = 8.150), indicating that at any given time, the patients with higher miRNAs expression had approximately 8 times more probability to be alive than the patients with lower expression levels (Figure 10A and Supplementary Table 9). Regarding RFS, one panel of 5 miRNAs showed potential prognostic value: miR-299-5p - miR-324-5p - miR-324-3p - miR-331-3p - miR-484 with a *p-value* = 0.030. With this panel of miRNAs, the patients presenting higher miRNAs expression have a better outcome than the patients with lower expression values (Figure 10B and Supplementary Table 10).

Moreover, for stage III patients we could identify a total of 65 remaining panels for OS (19 panels of 4 miRNAs, 31 panels of 3 miRNAs and 15 panels of 2 miRNAs; Supplementary Table 9) and 27 panels for RFS (10 panels of 4 miRNAs, 14 panels of 3 miRNAs and 3 panels of 2 miRNAs; Supplementary Table 10). Altogether these results further evidence the power of panels of miRNAs to predict patient outcome regarding both overall and recurrence free survivals.

Validation analysis

In order to validate our findings and further confirm the utility of the selected 25 miRNAs as potential non-invasive biomarkers for colorectal cancer we have inspected three additional and independent cohorts of CRC patients (validation cohorts). These included two cohorts with CRC tissue samples (GSE115513 and GSE41655 datasets) and one cohort with CRC plasma samples (GSE71008 dataset) plus the respective normal samples (Table 2). From these, the GSE41655 dataset showed the highest degree of similarity with the results described so far, with 23 out of the 25 miRNAs being downregulated in this cohort as well (Figure 11A). Although, since the total number of samples was
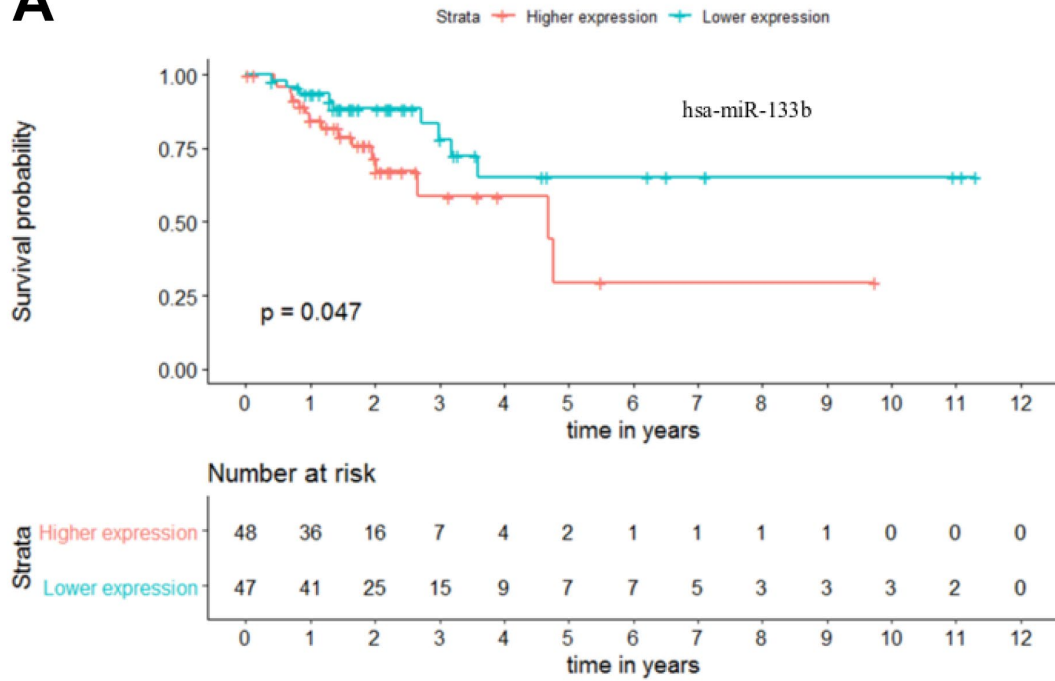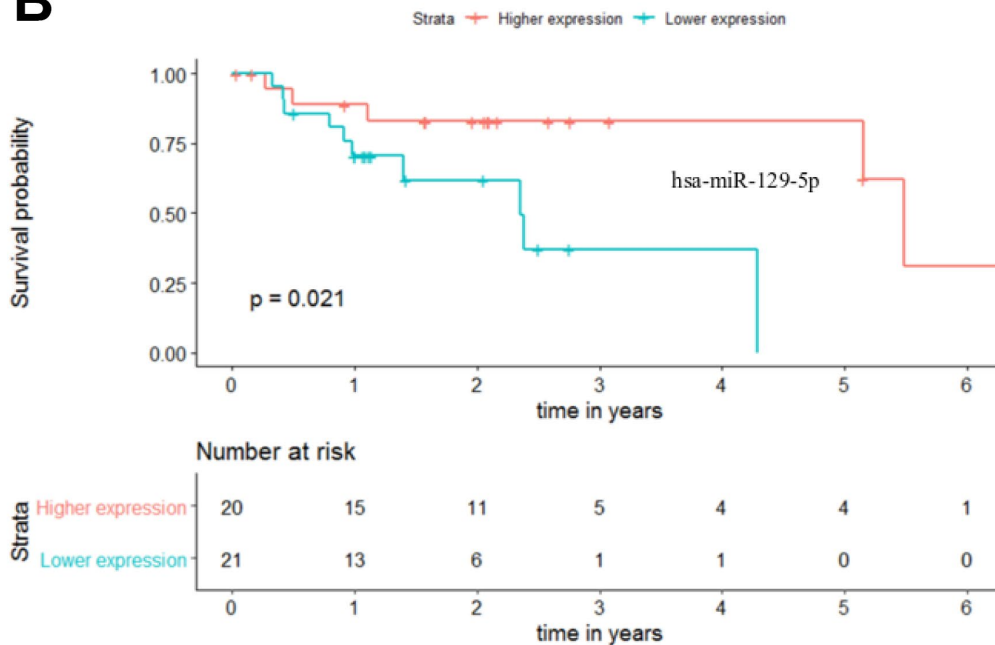
Figure 8. Individual miRNA overall survival (OS) prognosis for stage III and stage IV patients. (**A**) Stage III OS Kaplan-Meier curve for miR-133b (*p-value* = 0.047, Log rank; HR = 2.28) (**B**) Stage IV OS Kaplan-Meier curve for miR-129-5p (*p-value* = 0.021, Log rank; HR = 4.23). Time is represented in years. Higher (in red) and Lower (in blue) expression groups represent the patients with miRNA expression above and below miRNAs' median expression, respectively. Censored data is represented by small plus signs in each group. The number of patients at risk for each group and per time point is shown in the table below each graph. HR, hazard ratio.
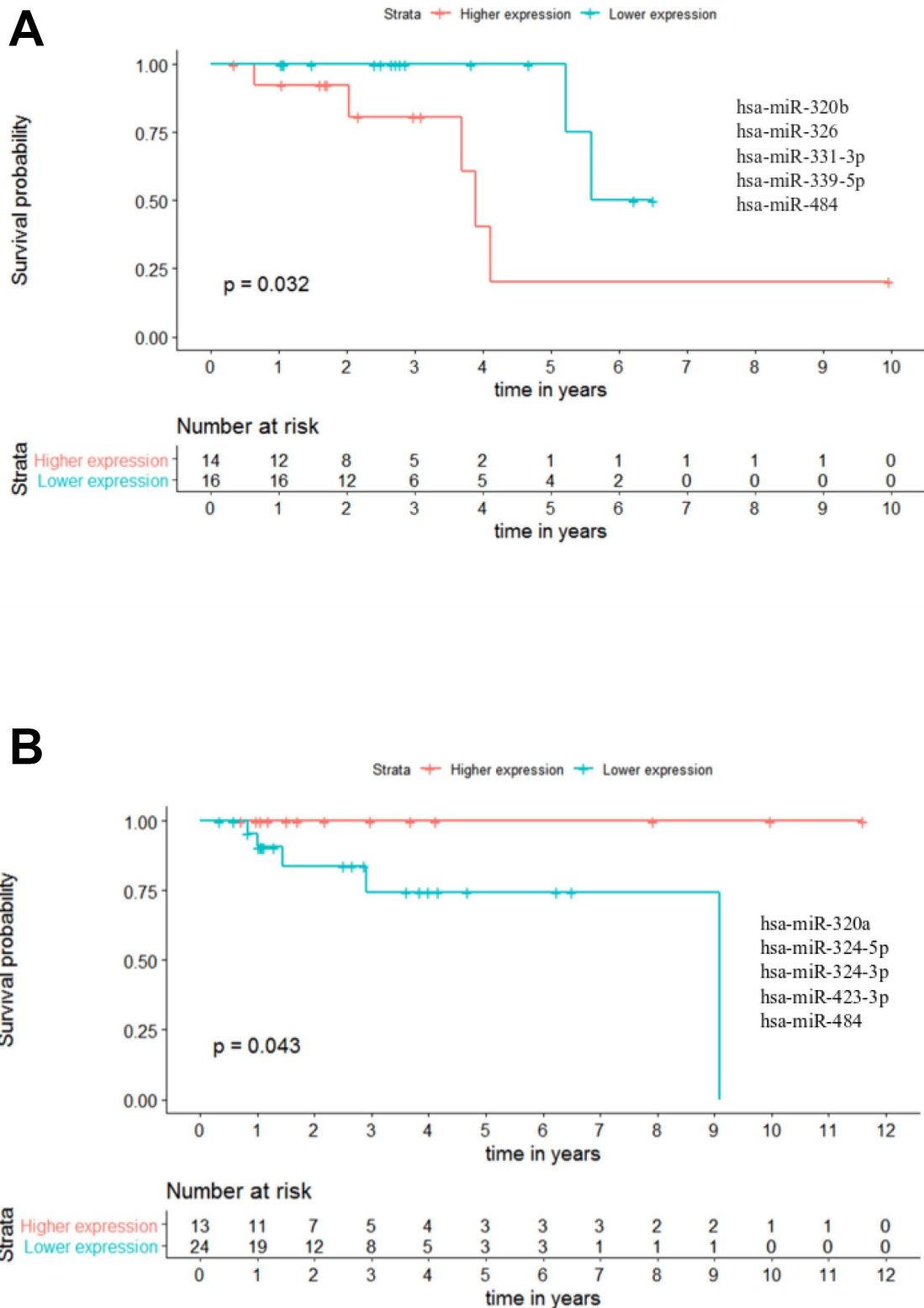
Figure 9. Panels of 5 miRNAs for overall survival (OS) and recurrence free survival (RFS) prognosis of stage II patients. (**A**) Stage II OS Kaplan-Meier curve based on miR-320b - miR-326 - miR-331-3p - miR-339-5p - miR-484 (*p-value* = 0.032, Log rank test; HR= 5.23) (**B**) Stage II recurrence free survival RFS Kaplan-Meier curve based on miR-320a - miR-324-5p - miR-324-3p - miR-423-3p - miR-484 (*p-value* = 0.043, Log rank test). Time is represented in years. Higher (in red) and Lower (in blue) expression groups represent the group of patients with miRNA expression above and below miRNAs median expression, respectively. Censored data is represented by small plus signs in each group. The number of patients at risk for each group and per time point is shown in the table below each graph. HR, hazard ratio.
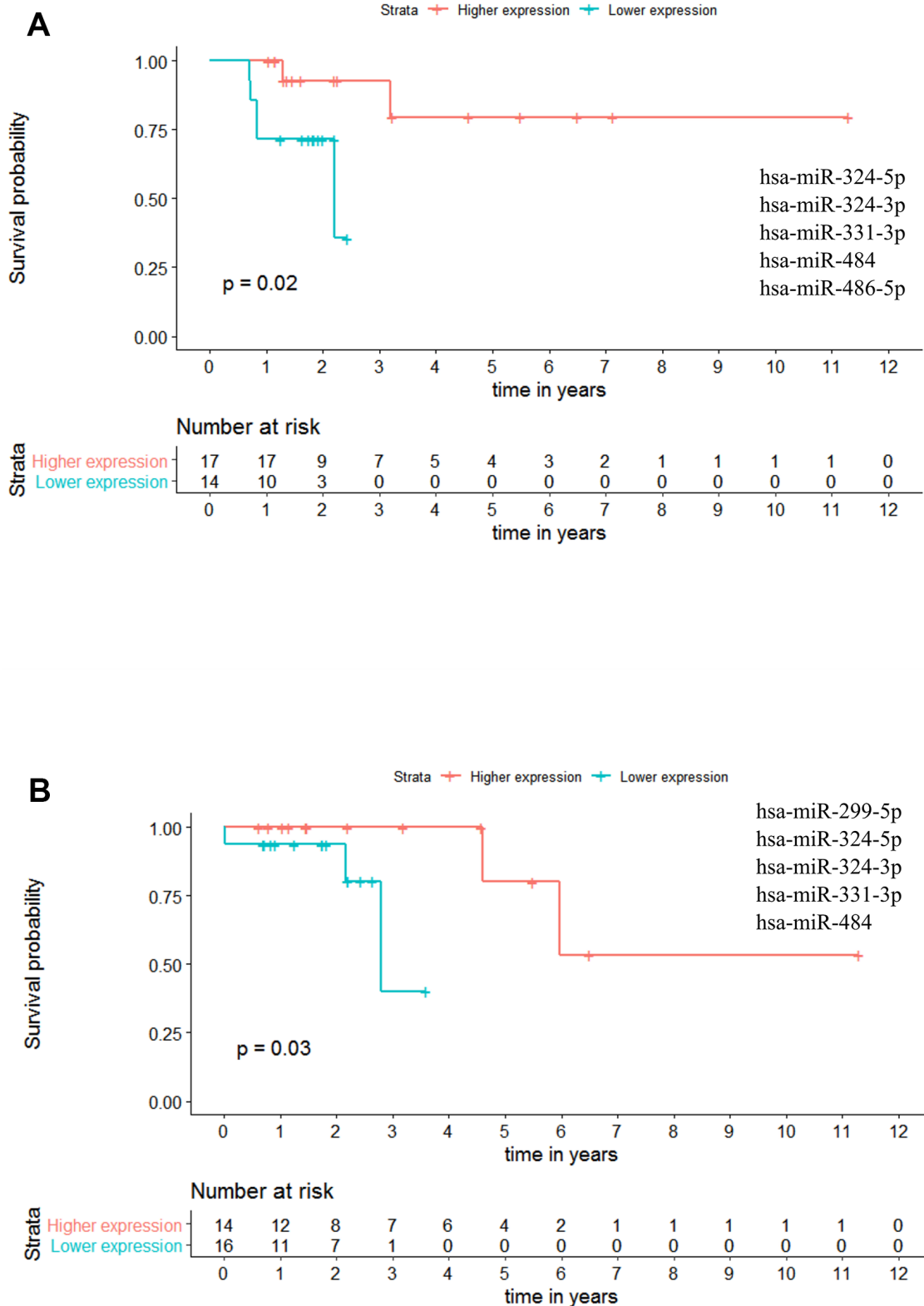
Figure 10. Panels of 5 miRNAs for overall survival (OS) and recurrence free survival (RFS) prognosis of stage III patients.
(**A**) Stage III OS Kaplan-Meier curve based on miR-324-5p - miR-324-3p - miR-331-3p - miR-484 - miR-486-5p (*p-value* = 0.020, Log rank test; HR= 8.15). (**B**) Stage III recurrence free survival RFS Kaplan-Meier curve based on miR-299-5p - miR-324-5p - miR-324-3p - miR-331-3p - miR-484 (p = 0.030, Log rank test). Time is represented in years. Higher (in red) and Lower (in blue) expression groups represent the group of patients with miRNA expression above and below miRNAs median expression, respectively. Censored data is represented by small plus signs in each group. The number of patients at risk for each group and per time point is shown in the table below each graph. HR, hazard ratio.

Table 2. Detailed patient information for the three colorectal cancer datasets used to corroborate our analysis.

Datasets	GSE115513	GSE41655	GSE71008
Number of miRNAs available	2006	851	421
Patient nationality	USA	China	USA
Platform	Agilent-046064 Unrestricted Human miRNA V19.0 Microarray	Agilent-021827 Human miRNA Microarray	Illumina Genome Analyzer
Tissue type	Paired tumour and normal tissue	Tumour and normal tissue	Plasma microvesicles
Number of samples			
Tumour	750 (54%)	33 (69%)	100 (67%)*
Normal	649 (46%)	15 (31%)	50 (33%)*
Age (mean ± SD[1] years)			
Tumour sample patients	65 ± 11	63 ± 11	55 ± 18
Normal sample patients	65 ± 11	48 ±14	54 ± 14
Gender			
Female	667 (48%)	16 (33%)	74 (49%)
Male	732 (52%)	32 (67%)	76 (51%)
Tumour Site			
Colon	792 (57%)	-	-
Rectum	607 (43%)	-	-

1 – standard deviation.

"-"– unspecified.

"*"– Tumour = cancerous patient samples; Normal = healthy individual samples.

relatively low (48 total samples) we have also analysed a larger cohort with a more significant number of cases (750 carcinomas and 649 normal mucosa samples). This analysis revealed that 12 out of the initial 25 miRNAs were also downregulated in the GSE115513 dataset (Figure 11A). Since miR-326 had no expression values available in this dataset for the normal mucosa samples we were unable to analyse its fold-change (Figure 11A).

Furthermore, and toward identifying miRNAs that could be used as non-invasive diagnostic and prognostic tools, we have investigated if our results could be extrapolated to a plasma context. Therefore, the expression profile of the 25 downregulated miRNAs was assessed in a cohort with plasma samples from CRC patients (GSE71008 dataset; Figure 11A). MiR-133b, miR-296-5p and miR-299-5p were not present in the GSE71008 dataset and therefore their regulation status was not evaluated. For the remaining 22 miRNAs, 12 were found to be downregulated and thus presenting the same expression profile as in the tumour tissue (Figure 11A).

Interestingly, comparing the outcomes from the three validation datasets we were able to identify 8 miRNAs commonly downregulated: miR-139-5p, miR-28-3, miR-330-3p, miR-331-3p, miR-375, miR-423-3p, miR-486-5p and miR-574-3p, further corroborating part of our results from the set of 'discovery' cohorts (Figure 11A).

Finally, we proceed to validate the diagnostic and prognostic ability of the 8 miRNAs commonly downregulated in all three validation cohorts. Regarding diagnostic ability, the 8 miRNAs did not present accuracy values as high as the ones obtained in the set of 'discovery' cohorts, with most of the miRNAs showing AUCs in the range of 0.5 and 0.7 in all three datasets and only miR-375 reaching an AUC above 0.9 in the GSE41655 dataset (Figure 11B and Supplementary Table 11).

When considering prognostic ability, none of the validation datasets provided information regarding the OS or RFS of patients. Therefore, to infer the prognostic potential of these 8 miRNAs we had to use the patient information available in the TCGA dataset. Importantly, we found that the panel of 2 miRNAs containing miR-486-5p - miR-375 (*p-value* = 0.032; HR = 2.890) and the panels of 3 miRNAs containing miR-486-5p - miR-330-3p - miR-375 (*p-value* = 0.025; HR = 4.010) or miR-486-5p - miR-331-3p - miR-375 (*p-value* = 0.031; HR = 3.911) could predict survival outcomes of stage III patients (Supplementary Table 12 and Figure 11C).

DISCUSSION

MiRNAs are key regulators of diverse biological and physiological processes and thus their deregulation is

strongly associated with pathological contexts [38, 39, 42]. The role of these small RNAs in cancer, including in CRC, has been one of the most studied over the years [45, 54, 80–82]. MiRNAs are stable, easily measurable, detectable in body fluids allowing for a non-invasive evaluation and, they are targetable molecules [41, 42, 83]. Due to their characteristics, miRNAs are promising biomarkers for both the detection and treatment of diseases, including CRC.

From the major hurdles in cancer management, late detection of malignancy, poor patient stratification and consequent suboptimal treatment are probably the ones that most impact on patient survival and recurrence of disease (excluding metastasis). Here we propose to identify miRNAs able to accurately distinguish malignant from healthy tissue and deliver panels of miRNAs as prognostic biomarkers for better patient stratification, particularly for CRC stages II and III. For that, we started by an integrated analysis using 4 colorectal cancer miRNA expression datasets (discovery cohorts) to identify miRNAs which expression was altered upon disea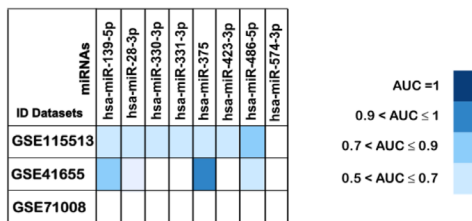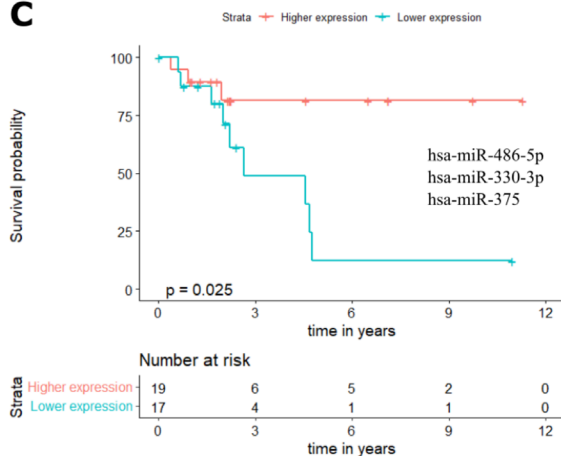se initiation and progression. Most miRNA expression changes in CRC seem to occur during normal mucosa to tumour transformation and are maintained throughout metastatic transition [15], which is similar to what we observed. We have unveiled 26 deregulated miRNAs in tumour tissue, exhibiting the same downregulation profile in at least 3 of the 4 datasets studied. The same regulation pattern was described in previous studies in multiple human cancers, including CRC, pancreatic, prostate and breast cancer [84]. An explanation for the massive miRNA downregulation in cancer was reported previously by Sun et al. 2016 and others. The authors found that ERK suppresses pre-miRNA export from the nucleus into the cytoplasm through phosphorylation of exportin-5 (EXPO5) [85]. Accordingly, in another study, pre-miRNAs are retained in the nucleus of cancer cell lines implying that the function of the nuclear-cytoplasmatic export machinery is compromised in tumour cells [86]. Even though, other mechanisms such as deletions or

A

miRNAs / ID Datasets	GSE115513	GSE41655	GSE71008
hsa-miR-125a-5p			
hsa-miR-127-3p	0<x<1		
hsa-miR-129-5p	0<x<1	<-2	
hsa-miR-133b			
hsa-miR-139-5p	-1<x<0		-1<x<0
hsa-miR-193a-5p			
hsa-miR-28-3p			
hsa-miR-296-5p	0<x<1	-1<x<0	
hsa-miR-299-5p			
hsa-miR-320a		0<x<1	
hsa-miR-320b		0<x<1	
hsa-miR-320d		0<x<1	
hsa-miR-324-3p			
hsa-miR-324-5p	0<x<1	0<x<1	
hsa-miR-326		0<x<1	
hsa-miR-330-3p	-1<x<0		
hsa-miR-331-3p			
hsa-miR-339-5p	>2	<-2	
hsa-miR-342-3p	0<x<1		
hsa-miR-375			
hsa-miR-423-3p			
hsa-miR-484			
hsa-miR-486-5p	-1<x<0		
hsa-miR-490-3p	-1<x<0		
hsa-miR-574-3p	-1<x<0		

Legend:
- >2 (red)
- 1 < x < 2
- 0 < x < 1
- -1 < x < 0
- -2 < x < -1
- < -2 (blue)

B

ID Datasets / miRNAs	hsa-miR-139-5p	hsa-miR-28-3p	hsa-miR-330-3p	hsa-miR-331-3p	hsa-miR-375	hsa-miR-423-3p	hsa-miR-486-5p	hsa-miR-574-3p
GSE115513	0.5<AUC≤0.7		0.5<AUC≤0.7	0.5<AUC≤0.7	0.5<AUC≤0.7	0.5<AUC≤0.7	0.7<AUC≤0.9	0.5<AUC≤0.7
GSE41655	0.7<AUC≤0.9				0.9<AUC≤1	0.5<AUC≤0.7	0.5<AUC≤0.7	
GSE71008								

Legend:
- AUC =1
- 0.9 < AUC ≤ 1
- 0.7 < AUC ≤ 0.9
- 0.5 < AUC ≤ 0.7

C

Strata — Higher expression — Lower expression

hsa-miR-486-5p
hsa-miR-330-3p
hsa-miR-375

p = 0.025

Number at risk

Strata	0	3	6	9	12
Higher expression	19	6	5	2	0
Lower expression	17	4	1	1	0

time in years

Figure 11. Validation analysis for the 25 downregulated miRNAs in CRC. (**A**) The log$_2$(FC) values calculated for each dataset are reported with red scale boxes for upregulated miRNAs and blue scale boxes for the downregulated miRNAs. White boxes represent the inexistence of the miRNA on the dataset. (**B**) The miRNAs AUC values in each of the datasets GSE115513, GSE41655 and GSE71008 are reported as blue scale

mutations, transcriptional and epigenetic regulation, seem to contribute to this phenomenon [87, 88].

We then hypothesize that the 26 altered miRNAs may regulate key genes in CRC further contributing to tumour progression. Also, miRNA function is known to be better understood through their target genes and regulatory networks [39]. Thus, we used MirTarBase to unveil the downstream targets of the miRNAs in study, since it relies on experimentally validated miRNA:target interactions [89]. We found that 25 out of the 26 miRNAs targeted a total of 173 genes previously described to be altered in CRC thus supporting our first analysis and the relevance of these deregulated miRNAs in colorectal cancer [90]. From the 312 miRNA:target interactions found, 104 corresponded to significant correlations between the expressions of the intervenient miRNA and target gene. Importantly, 22 out of the 25 miRNAs with interactions from the first analysis were maintained after the correlation analysis corroborating our approach. More than half of these statistically significant interactions were positive correlations, meaning that the expressions of miRNA and target gene evolve in the same direction. Although we should expect a decrease in the target levels upon an increase of the miRNA (or the other way around), it is well known that miRNAs are fine-tuning regulators of gene expression by primarily blocking protein translation rather than mRNA cleavage [20, 38, 39, 67, 69, 91]. Besides, the MirTarBase interactions were validated in a variety of backgrounds and not only in CRC contexts [20, 67, 69, 89, 91]. Therefore, future studies should confirm that the miRNA:target associations here identified also occur in colorectal cancer settings. Nevertheless, all miRNAs were found to intervene in signalling pathways usually deregulated in CRC such as the ones of WNT, TGFβ, TP53, PI3K-AKT, Ras and MAPK. Various studies have already demonstrated the key role of miRNAs in regulating these pathways [4, 20, 67, 69, 91–95] thus reinforcing our findings. Furthermore, Reid and colleagues showed that many miRNAs deregulated in CRC were computationally mapped to targets involved in pathways related to tumour progression [96]. Overall, these reports and our results suggest that miRNAs may contribute to the deregulation of important genes and cellular pathways that initiate and sustain the carcinogenic process. If this is the case, the miRNAs here identified should help detect precancerous and/or cancerous lesions and, in fact, we found they all have diagnostic value as 23 out of 25 are considered highly accurate or better biomarkers (AUC >0.9) in at least one of the datasets they were present. Among those, 11 miRNAs could differentiate tumour from normal tissue with 100% sensibility and 100% specificity. These same miRNAs had already been reported as potential biomarkers in CRC in several independent studies corroborating our findings [30–37]. Importantly, the potential biomarkers here identified surpass by far the sensibilities and specificities of the ones used nowadays in a clinical setting such as CEA and CA-19 [14, 15, 97]. Both proteins are found significantly increased in the serum of CRC patients although, usually in late stages of disease thus questioning their value for early detection of CRC [14, 15]. On the other hand, miRNAs have been considered promising non-invasive biomarkers for disease since the discovery that they can also be found in circulation [98–100]. In that sense, we propose that these 11 miRNAs which could accurately identify tumour tissue, might preserve their diagnostic biomarker status in blood samples and this should be addressed in future studies. Supporting our premise, the majority of the diagnostic miRNAs here described were already found circulating in blood [33, 99, 101–104] and in fact 4 out of these 11 miRNAs present the same expression profile in the plasma cohort here analysed.

Moreover, optimal clinical management of CRC regarding risk stratification of patients is still an unmet need. The current TNM staging system lacks the ability to accurately predict patient outcome, even in patients within the same stage, leading to suboptimal treatment administration [17, 19]. Interestingly, the majority of the 25 differentially expressed miRNAs had predictive value in panels with 2 to 5 miRNAs. Principally, these panels of miRNAs were able to predict the outcome of patients in stages II and III, the two disease stages for which risk stratification and therapeutic options are more challenging [17, 19]. Interestingly, some miRNAs were more associated with prognosis of either stage II or III, being part of a higher number of panels in a specific stage. For stage II OS we could evidence the role of miR-331-3p, miR-320b, miR-342-3p and miR-324-5p participating in an average of 15 (out of 87) panels. On the other hand, miR-324-3p, miR-375 and miR-486-5p participated in about 20 (out of 66) panels for stage III OS prognosis. These findings highlight some miRNAs that might specifically contribute to the prognosis of a specific CRC stage. In fact, all of them were previously described as tumour suppressor miRNAs by targeting genes that primarily control cell proliferation and/or apoptosis [105–109]. These reports are in agreement with the downregulation profile we observe for the same miRNAs in the present study. Additionally, the significant correlations for the miRNA:target interactions we have identified are reported by others as crucial for the function of these miRNAs. In CRC cell lines, miR-320b was found to target *MYC* to suppress cell proliferation [105] while miR-331-3p besides inhibiting proliferation additionally stimulates apoptosis by targeting *HER2* through activation of the PI3K/Akt and ERK1/2 pathways [106]. Inhibition of cell growth in

cervical and gastric cancers by miR-331-3p was also reported via targeting *NRP2* and *E2F1* respectively [107, 108]. Interestingly, our analysis depicted several target genes belonging to the same signalling pathways but targeted by distinct miRNAs, suggesting that different miRNAs might work in synergy to activate or repress specific signalling cascades. Corroborating our findings, Ferretti and colleagues reported that miR-324-5p, miR-125b and miR-326 work together to activate Hedgehog signalling by targeting *GLI1* and Smoothened [109], the same targets we predicted with our analysis. Thus, we propose that in CRC the downregulation of these miRNAs might lead to high levels of Hedgehog downstream genes, contributing to tumour cell proliferation.

Moreover, miR-324-3p regulates *WNT2B*, one of the WNT ligands able to activate the WNT pathway and suppresses migration and invasion in nasopharyngeal carcinoma [110]. The interaction of miR-486-5p and components of insulin growth factor (IGF) signalling, is another example of regulations depicted in our analysis and reported in the literature. By targeting *IGF1*, *IGF1R* and *PIK3R1* this miRNA induces cell cycle arrest and reduces migration of lung cancer cells while acting as a tumour suppressor in a xenograft mouse model [111]. In a similar model for CRC, miR-375 could significantly reduce tumour growth *in vivo* and its pro-apoptotic role in colorectal cancer was associated with the targeting of *YAP1* [112]. These studies further corroborate our analysis and strengthen our findings, which highlight combinations of miRNAs that might help clinical decisions in the future. Two of these combinations, one for stage II and one for stage III, are composed of panels of 5 miRNAs as prognostic biomarkers for overall survival, sharing 2 miRNAs (miR-331-3p and miR-484). The targeting of *ZEB1* and *SMAD2* by miR-484 was predicted in our analysis and is reported to suppress proliferation and epithelial-mesenchymal transition in cervical cancer [113]. While for stage II the lower expression of the combined 5 miRNAs was indicative of better prognosis, for stage III the opposite was true. This pattern is maintained for most of the miRNA's combinations regarding stages II and III OS. This suggests that the expression of each miRNA in cancer must be meticulously controlled throughout tumour progression since small changes in miRNAs levels seem to have a high impact on patient outcome. The strict transcriptional regulation of miRNAs as well as feedback loops altering the expression of some miRNAs has been reported in several contexts [39, 111, 112, 114–116].

Some of the miRNAs in study also showed a predictive value regarding recurrence-free survival. Of those, miR-320d, miR-324-3p, miR-324-5p and miR-331-3p were shared in more miRNAs panels (average of 9/28)

for stage III RFS prognosis, supporting the validity of our findings. For each of II and III stages of disease, we report a panel of 5 miRNAs with the ability to predict patient outcome, where higher miRNA levels are indicative of better prognosis. Corroborating our data, the predictive value of these miRNAs was previously described in cancer [59–61, 117].

Notably, we were able to validate the results of about one third (8 miRNAs) out of the 25 downregulated miRNAs simultaneously in three new cohorts (validation cohorts) regarding both their regulation status and biomarker potential. One of the validation cohorts is composed by plasma samples further corroborating our results and highlighting the potential of these miRNAs as non-invasive CRC biomarkers. Interestingly, among these 8 miRNAs were miR-486-5p, miR-375 and miR-331-3p which had shown to provide great prognostic results in several miRNAs panels for stages II and III in the 'discovery' analysis. These results were further outlined when combinations with only the 8 miRNAs were used for prognostic analysis and still, we were able to generate 3 miRNAs panels able to predict survival in stage III CRC patients.

Here we highlight promising candidate biomarkers for diagnosis and prognosis of colorectal cancer. We focused on delivering panels of miRNAs able to stratify the risk of patients with stages II and III of CRC while evidencing miRNAs potential as non-invasive biomarkers. The combination of a group of biomarkers for cancer management is considered a reliable asset for the clinic [59–61] and has in fact became a thyroid cancer diagnostic commercial tool. Finally, we spot possible therapeutical targets for CRC since most of the downregulated miRNAs we have found act as tumour suppressors and manipulation of their levels should arrest cancer progression. MiRNA-based therapy is thus a promising avenue for future treatment options in cancer [41].

MATERIALS AND METHODS

Colorectal cancer miRNA expression data collection

MiRNA profiling data in CRC was collected from two repositories: The Cancer Genome Atlas (TCGA) database and the Gene Expression Omnibus DataSets portal (GEO DataSets).

TCGA CRC miRNA expression data was obtained from the miRNA mature strand expression RNAseq – IlluminaHiseq dataset for both TCGA Colon Cancer (COAD) and Rectal Cancer (READ;) cohorts (https://tcga.xenahubs.net) using the University of California, Santa Cruz cancer (UCSC) Xena

Public Data Hubs (https://xena.ucsc.edu/pubichubs/) [118, 119].

GEO selection of the CRC datasets publicly available on NCBI (https://www.ncbi.nlm.nih.gov/geo/) was carried out by searching the terms in the advanced research tool "(microRNA expression) AND (Illumina) AND (colorectal cancer) AND "Homo sapiens"[porgn:__txid9606]" and "(microRNA expression) AND (colorectal cancer) AND "Homo sapiens"[porgn:__txid9606]". Only the datasets containing the miRNA expression data for both colorectal cancer tissue samples and healthy tissue counterpart were included in our study. Exclusion criteria for the discovery cohorts were: i) datasets containing information only regarding cancer patients or pathological tissue samples; ii) datasets containing information on miRNAs expression levels for plasma samples only; iii) datasets that studied a set of specifically determined miRNAs. Four GEO datasets fulfil our criteria: GSE30454, GSE54088, GSE18392 and GSE33125. Solely GEO datasets containing miRNA expression obtained through Illumina platforms were used in the present work.

To identify the differently expressed miRNA in CRC, the data matrix of each of the 5 selected datasets was downloaded, transformed into a. csv file, and imported to R.

Data processing – preparing data for analysis

The miRBase database (miRBase 21, http://www.mirbase.org/) was used to standardize miRNA nomenclature [120]. Missing data treatment was performed using a modified listwise case deletion, in which variable elimination was performed when more than half of the information for each miRNA was missing in either normal or colorectal cancer samples [121]. This technique represented a more conservative approach to listwise case deletion, as the last would result in the omission of an extensive number of data which could lead to having insufficient data to perform the analysis. This step allowed us to further reduce the noise in our results. Missing data treatment was proceeded by outlier detection, performed using the Box plot technique, and for each miRNA, outlier values in both tumour and normal groups were eliminated [122, 123]. After data processing, the GSE54088 dataset was removed from the analysis due to its low number of miRNAs (only 6) available for further analysis.

Power consideration and statistical analysis

To assure that the miRNA expression values between normal and cancer patient samples were large enough to be considered statistically significant, a series of statistical tests were conducted. First, normal distribution assessment through Shapiro-Wilk test was performed. If miRNAs were normally distributed ($p\text{-}value > 0.05$) we applied a Wilcoxon-Mann-Whitney test, otherwise, a Levene's test followed by a two-sample t-test was used to assess statistical differences between CRC and normal tissue samples [124–128]. A $p\text{-}value$ of 0.05 was considered statistically significant in all tests. A False Discovery Rate (FDR) equal to or below 5% was considered statistically significant for both t-test and Mann-Whitney [129, 130].

Identification of differently expressed miRNAs in colorectal cancer

MiRNA differential expression in CRC, for each dataset, was determined by the fold change (FC) values between colorectal cancer tumour and normal samples. After calculated, the FC values were then transformed in base-2 logarithm of FC ($\log_2(FC)$). The differently expressed miRNAs obtained for each dataset were then merged using a Venn diagram to identify miRNAs that presented statistical differences in at least 3 of our datasets.

Identification of miRNAs target genes and functional analysis

MirTarBase version 7.0 was used to retrieve the target genes of the differentially expressed miRNAs until the 15th of December 2019. MirTarBase is a manually collected dataset of microRNA-target interactions (MTIs) experimentally validated that holds more than four hundred and twenty thousand MTIs [89]. Only functional MTIs catalogued in *Homo sapiens* were considered in our analysis.

The miRNA-target genes reported to be involved in colorectal carcinoma development were obtained from the Colorectal Cancer Gene Database (CoReCG) (http://lms.snu.edu.in/corecg/) until the 22nd of December of 2019 [131]. Visualization of the miRNA-target interaction analysis was obtained thought a Chord-diagram performed using the *circlize* R package.

Functional and pathway enrichment analyses were performed using Kyoto Encyclopedia of Genes and Genomes (KEEG) (https://www.genome.jp/kegg/) available at The Database for Annotation, Visualization and Integrated Discovery (DAVID) Bioinformatics Resources 6.8 (https://david.ncifcrf.gov/home.jsp). A $p\text{-}value$ <0.05 was considered to indicate statistical significance.

Gene expression analysis

Gene expression data was obtained from the colon and rectal gene expression RNAseq - IlluminaHiSeq dataset for both TCGA Colon Cancer (COAD) and Rectal Cancer (READ) cohorts as described above for the miRNAs. [118, 119].

Sample matching was performed to identify correspondent samples between the gene expression dataset and the miRNA expression dataset. Only matched samples were used in our analysis. FC values between colorectal cancer tumour and normal samples were $\log_2$ transformed. Visual display of the $\log_2$(FC) expression values was obtained using a circular bar plot.

MiRNA-target gene interaction analysis

Gene expression normality was assessed using the Shapiro-Wilk test. Pearson's correlation coefficient was utilized to estimate the correlation between miRNA expression levels and gene expression levels if both miRNA and gene followed normality, otherwise, the Spearman test was used [132]. A *p-value* of 0.05 was considered statistically significant, and correlations were classified according to Mukaka 2012 [132].

Biomarker potential – diagnostic and prognostic value

Clinical data (time of overall survival and of recurrence-free survival) was retrieved for each stage of disease (I, II, III and IV) from the TCGA, the only database with clinical information available.

To access miRNAs diagnostic value, Receiver Operating Characteristic (ROC) curves were generated using the *roc* function found in the *pROC* package, and miRNA's ability to behave as potential biomarkers were determined by the Area under the curve (AUC) and stratified according to the classification provided in Greiner et al., 2000 [76]. The AUC represents the overall performance of the classifier summarized over all possible cut-off values, with an ideal ROC curve hugging the top left corner. Thus, the higher the AUC the better the classifier [76, 133].

For prognosis value, we considered statistically significant a *p-value* <0.05 for the log-rank test when performing the Cox Proportional-Hazard regression Model [134, 135]. To perform this analysis, two distinct groups of patients were constructed based on the tumour samples miRNAs medians. Patients with miRNA expression values

above the respective tumour samples miRNA median were allocated in the "Higher expression" group, while patients with miRNA expression values below the median were allocated in the "Lower median" group. The Cox Proportional-Hazard regressions were conducted separately for each of the four stages of disease (I, II, III, and IV) [136]. Hazard ratios (HR) which represented the instantaneous risk of dying from CRC over the course of life were obtained as the ratio of the hazard rates between the "Higher expression" and "Lower expression" groups. Wilcox regression Models were performed and graphical representations (Kaplan-Meier's) were plotted [137].

Validation cohorts and analysis

To validate our results, the methylation status, diagnostic and prognostic potential of the selected 25-downregulated miRNAs was further analysed in 3 additional datasets: two generated from platforms other than Illumina (GSE115513 and GSE41655 datasets) and a plasma dataset (GSE71008) generated from an Illumina platform. To obtain similar results we have transformed the raw data and expressed it as log2 values in order to obtain the same scale value as the previous datasets used in this work.

Statistical software

Analyses were performed using R studio version 3.6.4 [138] and supplemented by the R packages: car, outliers, dplyr, pROC, ggplot2, survival, survminer, VennDiagram and miRBaseVersions.db.

AUTHOR CONTRIBUTIONS

Conceptualization, VR, PCB and AM; Methodology, VR, AM and AF; Data Curation, AF, AM, AM and SR; Writing – Original Draft Preparation, IF and VR; Writing – Review and Editing, VR, PCB and RPN; Supervision, VR, PCB and AM; Funding Acquisition, VR and PCB.

CONFLICTS OF INTEREST

FUNDING

REFERENCES

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018; 68:394–424.
https://doi.org/10.3322/caac.21492
PMID:30207593

2. Van der Jeught K, Xu HC, Li YJ, Lu XB, Ji G. Drug resistance and new therapies in colorectal cancer. World J Gastroenterol. 2018; 24:3834–48.
https://doi.org/10.3748/wjg.v24.i34.3834
PMID:30228778

3. Levin B, Lieberman DA, McFarland B, Smith RA, Brooks D, Andrews KS, Dash C, Giardiello FM, Glick S, Levin TR, Pickhardt P, Rex DK, Thorson A, Winawer SJ, and American Cancer Society Colorectal Cancer Advisory Group, and US Multi-Society Task Force, and American College of Radiology Colon Cancer Committee. Screening and surveillance for the early detection of colorectal cancer and adenomatous polyps, 2008: a joint guideline from the American Cancer Society, the US Multi-Society Task Force on Colorectal Cancer, and the American College of Radiology. CA Cancer J Clin. 2008; 58:130–60.
https://doi.org/10.3322/CA.2007.0018
PMID:18322143

4. Falzone L, Scola L, Zanghì A, Biondi A, Di Cataldo A, Libra M, Candido S. Integrated analysis of colorectal cancer microRNA datasets: identification of microRNAs associated with tumor development. Aging (Albany NY). 2018; 10:1000–14.
https://doi.org/10.18632/aging.101444
PMID:29779016

5. Brenner H, Kloor M, Pox CP. Colorectal cancer. Lancet. 2014; 383:1490–502.
https://doi.org/10.1016/S0140-6736(13)61649-9
PMID:24225001

6. Strul H, Arber N. Screening techniques for prevention and early detection of colorectal cancer in the average-risk population. Gastrointest Cancer Res. 2007; 1:98–106.
PMID:19262715

7. Ciombor KK, Wu C, Goldberg RM. Recent therapeutic advances in the treatment of colorectal cancer. Annu Rev Med. 2015; 66:83–95.
https://doi.org/10.1146/annurev-med-051513-102539
PMID:25341011

8. Simon K. Colorectal cancer development and advances in screening. Clin Interv Aging. 2016; 11:967–76.
https://doi.org/10.2147/CIA.S109285
PMID:27486317

9. U.S. Preventive Services Task Force. Screening for colorectal cancer: recommendation and rationale. Ann Intern Med. 2002; 137:129–31.
https://doi.org/10.7326/0003-4819-137-2-200207160-00014 PMID:12118971

10. Levine JS, Ahnen DJ. Clinical practice. Adenomatous polyps of the colon. N Engl J Med. 2006; 355:2551–57.
https://doi.org/10.1056/NEJMcp063038
PMID:17167138

11. O'Connell JB, Maggard MA, Ko CY. Colon cancer survival rates with the new American Joint Committee on Cancer sixth edition staging. J Natl Cancer Inst. 2004; 96:1420–25.
https://doi.org/10.1093/jnci/djh275
PMID:15467030

12. Haggar FA, Boushey RP. Colorectal cancer epidemiology: incidence, mortality, survival, and risk factors. Clin Colon Rectal Surg. 2009; 22:191–97.
https://doi.org/10.1055/s-0029-1242458
PMID:21037809

13. Cappellani A, Di Vita M, Zanghi A, Veroux P, Cavallaro A, Lo Menzo E, Cacopardo B, Canzonieri V, Murabito P, Tirelli U, Berretta M. Biological and clinical markers in colorectal cancer: state of the art. Front Biosci (Schol Ed). 2010; 2:422–31.
https://doi.org/10.2741/s75 PMID:20036958

14. Vukobrat-Bijedic Z, Husic-Selimovic A, Sofic A, Bijedic N, Bjelogrlic I, Gogov B, Mehmedovic A. Cancer Antigens (CEA and CA 19-9) as Markers of Advanced Stage of Colorectal Carcinoma. Med Arch. 2013; 67:397–401.
https://doi.org/10.5455/medarh.2013.67.397-401
PMID:25568506

15. Swiderska M, Choromańska B, Dąbrowska E, Konarzewska-Duchnowska E, Choromańska K, Szczurko G, Myśliwiec P, Dadan J, Ladny JR, Zwierz K. The diagnostics of colorectal cancer. Contemp Oncol (Pozn). 2014; 18:1–6.
https://doi.org/10.5114/wo.2013.39995
PMID:24876814

16. Zamani M, Hosseini SV, Mokarram P. Epigenetic biomarkers in colorectal cancer: premises and prospects. Biomarkers. 2018; 23:105–14.
https://doi.org/10.1080/1354750X.2016.1252961
PMID:27788596

17. Li J, Yi CH, Hu YT, Li JS, Yuan Y, Zhang SZ, Zheng S, Ding KF. TNM Staging of Colorectal Cancer Should be Reconsidered According to Weighting of the T Stage: Verification Based on a 25-Year Follow-Up. Medicine (Baltimore). 2016; 95:e2711.
https://doi.org/10.1097/MD.0000000000002711
PMID:26871810

18. Brierley JD, Gospodarowicz MK, Wittekind C. Essential TNM. In: Sobin LH, Sobin LH, Gospodarowicz MK, Wittekind C, Brierley JD, editors. TNM Online. 2017.
https://doi.org/10.1002/9780471420194.tnmc66

19. Bender U, Rho YS, Barrera I, Aghajanyan S, Acoba J, Kavan P. Adjuvant therapy for stages II and III colon cancer: risk stratification, treatment duration, and future directions. Curr Oncol. 2019 (Suppl 1); 26:S43–52.
https://doi.org/10.3747/co.26.5605 PMID:31819709

20. Bardhan K, Liu K. Epigenetics and colorectal cancer pathogenesis. Cancers (Basel). 2013; 5:676–713.
https://doi.org/10.3390/cancers5020676
PMID:24216997

21. Armaghany T, Wilson JD, Chu Q, Mills G. Genetic alterations in colorectal cancer. Gastrointest Cancer Res. 2012; 5:19–27.
PMID:22574233

22. Shen L, Toyota M, Kondo Y, Lin E, Zhang L, Guo Y, Hernandez NS, Chen X, Ahmed S, Konishi K, Hamilton SR, Issa JP. Integrated genetic and epigenetic analysis identifies three different subclasses of colon cancer. Proc Natl Acad Sci USA. 2007; 104:18654–59.
https://doi.org/10.1073/pnas.0704652104
PMID:18003927

23. Khare S, Verma M. Epigenetics of colon cancer. Methods Mol Biol. 2012; 863:177–85.
https://doi.org/10.1007/978-1-61779-612-8_10
PMID:22359293

24. Kinzler KW, Vogelstein B. Lessons from hereditary colorectal cancer. Cell. 1996; 87:159–70.
https://doi.org/10.1016/s0092-8674(00)81333-1
PMID:8861899

25. Sharma S, Kelly TK, Jones PA. Epigenetics in cancer. Carcinogenesis. 2010; 31:27–36.
https://doi.org/10.1093/carcin/bgp220
PMID:19752007

26. Grønbaek K, Hother C, Jones PA. Epigenetic changes in cancer. APMIS. 2007; 115:1039–59.
https://doi.org/10.1111/j.1600-0463.2007.apm_636.xml.x PMID:18042143

27. Fedoriw A, Mugford J, Magnuson T. Genomic imprinting and epigenetic control of development. Cold Spring Harb Perspect Biol. 2012; 4:a008136.
https://doi.org/10.1101/cshperspect.a008136
PMID:22687277

28. Calin GA, Dumitru CD, Shimizu M, Bichi R, Zupo S, Noch E, Aldler H, Rattan S, Keating M, Rai K, Rassenti L, Kipps T, Negrini M, et al. Frequent deletions and down-regulation of micro- RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. Proc Natl Acad Sci USA. 2002; 99:15524–29.

https://doi.org/10.1073/pnas.242606799
PMID:12434020

29. Li M, Marin-Muller C, Bharadwaj U, Chow KH, Yao Q, Chen C. MicroRNAs: control and loss of control in human physiology and disease. World J Surg. 2009; 33:667–84.
https://doi.org/10.1007/s00268-008-9836-x
PMID:19030926

30. Zhang P, Ji DB, Han HB, Shi YF, Du CZ, Gu J. Downregulation of miR-193a-5p correlates with lymph node metastasis and poor prognosis in colorectal cancer. World J Gastroenterol. 2014; 20:12241–48.
https://doi.org/10.3748/wjg.v20.i34.12241
PMID:25232258

31. Chen B, Xia Z, Deng YN, Yang Y, Zhang P, Zhu H, Xu N, Liang S. Emerging microRNA biomarkers for colorectal cancer diagnosis and prognosis. Open Biol. 2019; 9:180212.
https://doi.org/10.1098/rsob.180212
PMID:30958116

32. Fateh A, Feizi MA, Safaralizadeh R, Azarbarzin S. Importance of miR-299-5p in colorectal cancer. Ann Gastroenterol. 2017; 30:322–26.
https://doi.org/10.20524/aog.2017.0139
PMID:28469363

33. Moya L, Meijer J, Schubert S, Matin F, Batra J. Assessment of miR-98-5p, miR-152-3p, miR-326 and miR-4289 Expression as Biomarker for Prostate Cancer Diagnosis. Int J Mol Sci. 2019; 20:1154.
https://doi.org/10.3390/ijms20051154
PMID:30845775

34. Zhou C, Liu G, Wang L, Lu Y, Yuan L, Zheng L, Chen F, Peng F, Li X. MiR-339-5p regulates the growth, colony formation and metastasis of colorectal cancer cells by targeting PRL-1. PLoS One. 2013; 8:e63142.
https://doi.org/10.1371/journal.pone.0063142
PMID:23696794

35. Tian F, Wang J, Ouyang T, Lu N, Lu J, Shen Y, Bai Y, Xie X, Ge Q. MiR-486-5p Serves as a Good Biomarker in Nonsmall Cell Lung Cancer and Suppresses Cell Growth With the Involvement of a Target PIK3R1. Front Genet. 2019; 10:688.
https://doi.org/10.3389/fgene.2019.00688
PMID:31402930

36. Tang Y, Zhao Y, Song X, Song X, Niu L, Xie L. Tumor-derived exosomal miRNA-320d as a biomarker for metastatic colorectal cancer. J Clin Lab Anal. 2019; 33:e23004.
https://doi.org/10.1002/jcla.23004
PMID:31420913

37. Tang L, Zhou L, Wu S, Shi X, Jiang G, Niu S, Ding D. miR-125a-5p inhibits colorectal cancer cell epithelial-

mesenchymal transition, invasion and migration by targeting TAZ. Onco Targets Ther. 2019; 12:3481–89.
https://doi.org/10.2147/OTT.S191247
PMID:31190857

38. Treiber T, Treiber N, Meister G. Regulation of microRNA biogenesis and its crosstalk with other cellular pathways. Nat Rev Mol Cell Biol. 2019; 20:5–20.
https://doi.org/10.1038/s41580-018-0059-1
PMID:30228348

39. Bartel DP. Metazoan MicroRNAs. Cell. 2018; 173:20–51.
https://doi.org/10.1016/j.cell.2018.03.006
PMID:29570994

40. Hanna J, Hossain GS, Kocerha J. The Potential for microRNA Therapeutics and Clinical Research. Front Genet. 2019; 10:478.
https://doi.org/10.3389/fgene.2019.00478
PMID:31156715

41. Rupaimoole R, Slack FJ. MicroRNA therapeutics: towards a new era for the management of cancer and other diseases. Nat Rev Drug Discov. 2017; 16:203–22.
https://doi.org/10.1038/nrd.2016.246
PMID:28209991

42. Condrat CE, Thompson DC, Barbu MG, Bugnar OL, Boboc A, Cretoiu D, Suciu N, Cretoiu SM, Voinea SC. miRNAs as Biomarkers in Disease: Latest Findings Regarding Their Role in Diagnosis and Prognosis. Cells. 2020; 9:276.
https://doi.org/10.3390/cells9020276
PMID:31979244

43. Hur K. MicroRNAs: promising biomarkers for diagnosis and therapeutic targets in human colorectal cancer metastasis. BMB Rep. 2015; 48:217–22.
https://doi.org/10.5483/bmbrep.2015.48.4.007
PMID:25603797

44. Cortez MA, Bueso-Ramos C, Ferdin J, Lopez-Berestein G, Sood AK, Calin GA. MicroRNAs in body fluids--the mix of hormones and biomarkers. Nat Rev Clin Oncol. 2011; 8:467–77.
https://doi.org/10.1038/nrclinonc.2011.76
PMID:21647195

45. Carter JV, Galbraith NJ, Yang D, Burton JF, Walker SP, Galandiuk S. Blood-based microRNAs as biomarkers for the diagnosis of colorectal cancer: a systematic review and meta-analysis. Br J Cancer. 2017; 116:762–74.
https://doi.org/10.1038/bjc.2017.12 PMID:28152545

46. Yan S, Han B, Gao S, Wang X, Wang Z, Wang F, Zhang J, Xu D, Sun B. Exosome-encapsulated microRNAs as circulating biomarkers for colorectal cancer. Oncotarget. 2017; 8:60149–58.
https://doi.org/10.18632/oncotarget.18557
PMID:28947960

47. Dumache R. Early Diagnosis of Oral Squamous Cell Carcinoma by Salivary microRNAs. Clin Lab. 2017; 63:1771–76.
https://doi.org/10.7754/Clin.Lab.2017.170607
PMID:29226639

48. Buermans HP, den Dunnen JT. Next generation sequencing technology: Advances and applications. Biochim Biophys Acta. 2014; 1842:1932–41.
https://doi.org/10.1016/j.bbadis.2014.06.015
PMID:24995601

49. Zhang J, Chiodini R, Badr A, Zhang G. The impact of next-generation sequencing on genomics. J Genet Genomics. 2011; 38:95–109.
https://doi.org/10.1016/j.jgg.2011.02.003
PMID:21477781

50. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet. 2009; 10:57–63.
https://doi.org/10.1038/nrg2484
PMID:19015660

51. Beg MS, Brenner AJ, Sachdev J, Borad M, Kang YK, Stoudemire J, Smith S, Bader AG, Kim S, Hong DS. Phase I study of MRX34, a liposomal miR-34a mimic, administered twice weekly in patients with advanced solid tumors. Invest New Drugs. 2017; 35:180–88.
https://doi.org/10.1007/s10637-016-0407-y
PMID:27917453

52. Kanaan Z, Rai SN, Eichenberger MR, Roberts H, Keskey B, Pan J, Galandiuk S. Plasma miR-21: a potential diagnostic marker of colorectal cancer. Ann Surg. 2012; 256:544–51.
https://doi.org/10.1097/SLA.0b013e318265bd6f
PMID:22868372

53. Kanaan Z, Roberts H, Eichenberger MR, Billeter A, Ocheretner G, Pan J, Rai SN, Jorden J, Williford A, Galandiuk S. A plasma microRNA panel for detection of colorectal adenomas: a step toward more precise screening for colorectal cancer. Ann Surg. 2013; 258:400–08.
https://doi.org/10.1097/SLA.0b013e3182a15bcc
PMID:24022433

54. Masuda T, Hayashi N, Kuroda Y, Ito S, Eguchi H, Mimori K. MicroRNAs as Biomarkers in Colorectal Cancer. Cancers (Basel). 2017; 9:124.
https://doi.org/10.3390/cancers9090124
PMID:28902152

55. Naidu S, Magee P, Garofalo M. MiRNA-based therapeutic intervention of cancer. J Hematol Oncol. 2015; 8:68.
https://doi.org/10.1186/s13045-015-0162-0
PMID:26062952

56. Liu X, Xu X, Pan B, He B, Chen X, Zeng K, Xu M, Pan Y, Sun H, Xu T, Hu X, Wang S. Circulating miR-1290 and miR-320d as Novel Diagnostic Biomarkers of Human Colorectal Cancer. J Cancer. 2019; 10:43–50.
https://doi.org/10.7150/jca.26723
PMID:30662524

57. Yu J, Xu J, Zhao J, Zhang R. Serum miR-133b is a potential novel prognostic biomarker for colorectal cancer. Int J Clin Exp Pathol. 2017; 10:11673–78.
PMID:31966526

58. Okugawa Y, Grady WM, Goel A. Epigenetic Alterations in Colorectal Cancer: Emerging Biomarkers. Gastroenterology. 2015; 149:1204–25.e12.
https://doi.org/10.1053/j.gastro.2015.07.011
PMID:26216839

59. Hong HC, Chuang CH, Huang WC, Weng SL, Chen CH, Chang KH, Liao KW, Huang HD. A panel of eight microRNAs is a good predictive parameter for triple-negative breast cancer relapse. Theranostics. 2020; 10:8771–89.
https://doi.org/10.7150/thno.46142 PMID:32754277

60. O'Brien SJ, Netz U, Hallion J, Bishop C, Stephen V, Burton J, Paas M, Feagins K, Pan J, Rai SN, Galandiuk S. Circulating plasma microRNAs in colorectal neoplasia: A pilot study in assessing response to therapy. Transl Oncol. 2021; 14:100962.
https://doi.org/10.1016/j.tranon.2020.100962
PMID:33285367

61. Fredsøe J, Rasmussen AK, Mouritzen P, Borre M, Ørntoft T, Sørensen KD. A five-microRNA model (pCaP) for predicting prostate cancer aggressiveness using cell-free urine. Int J Cancer. 2019; 145:2558–67.
https://doi.org/10.1002/ijc.32296 PMID:30903800

62. Li M, Tian L, Wang L, Yao H, Zhang J, Lu J, Sun Y, Gao X, Xiao H, Liu M. Down-regulation of miR-129-5p inhibits growth and induces apoptosis in laryngeal squamous cell carcinoma by targeting APC. PLoS One. 2013; 8:e77829.
https://doi.org/10.1371/journal.pone.0077829
PMID:24194897

63. Zhang Y, An J, Lv W, Lou T, Liu Y, Kang W. miRNA-129-5p suppresses cell proliferation and invasion in lung cancer by targeting microspherule protein 1, E-cadherin and vimentin. Oncol Lett. 2016; 12:5163–69.
https://doi.org/10.3892/ol.2016.5372
PMID:28105223

64. Ashton-Rickardt PG, Dunlop MG, Nakamura Y, Morris RG, Purdie CA, Steel CM, Evans HJ, Bird CC, Wyllie AH. High frequency of APC loss in sporadic colorectal carcinoma due to breaks clustered in 5q21-22. Oncogene. 1989; 4:1169–74.
PMID:2797819

65. Chang PY, Chen JS, Chang SC, Wang MC, Chang NC, Wen YH, Tsai WS, Liu WH, Liu HL, Lu JJ. Acquired somatic TP53 or PIK3CA mutations are potential predictors of when polyps evolve into colorectal cancer. Oncotarget. 2017; 8:72352–62.
https://doi.org/10.18632/oncotarget.20376
PMID:29069792

66. Jiang BH, Liu LZ. PI3K/PTEN signaling in angiogenesis and tumorigenesis. Adv Cancer Res. 2009; 102:19–65.
https://doi.org/10.1016/S0065-230X(09)02002-8
PMID:19595306

67. Binefa G, Rodríguez-Moranta F, Teule A, Medina-Hayas M. Colorectal cancer: from prevention to personalized medicine. World J Gastroenterol. 2014; 20:6786–808.
https://doi.org/10.3748/wjg.v20.i22.6786
PMID:24944469

68. Stoffel EM, Koeppe E, Everett J, Ulintz P, Kiel M, Osborne J, Williams L, Hanson K, Gruber SB, Rozek LS. Germline Genetic Features of Young Individuals With Colorectal Cancer. Gastroenterology. 2018; 154:897–905.e1.
https://doi.org/10.1053/j.gastro.2017.11.004
PMID:29146522

69. Garraway LA, Lander ES. Lessons from the cancer genome. Cell. 2013; 153:17–37.
https://doi.org/10.1016/j.cell.2013.03.002
PMID:23540688

70. Li XL, Zhou J, Chen ZR, Chng WJ. P53 mutations in colorectal cancer - molecular pathogenesis and pharmacological reactivation. World J Gastroenterol. 2015; 21:84–93.
https://doi.org/10.3748/wjg.v21.i1.84 PMID:25574081

71. Slattery ML, Mullany LE, Wolff RK, Sakoda LC, Samowitz WS, Herrick JS. The p53-signaling pathway and colorectal cancer: Interactions between downstream p53 target genes and miRNAs. Genomics. 2019; 111:762–71.
https://doi.org/10.1016/j.ygeno.2018.05.006
PMID:29860032

72. Zenonos K, Kyprianou K. RAS signaling pathways, mutations and their role in colorectal cancer. World J Gastrointest Oncol. 2013; 5:97–101.
https://doi.org/10.4251/wjgo.v5.i5.97 PMID:23799159

73. Hamada T, Nowak JA, Ogino S. PIK3CA mutation and colorectal cancer precision medicine. Oncotarget. 2017; 8:22305–06.
https://doi.org/10.18632/oncotarget.15724
PMID:28423591

74. Zhao M, Mishra L, Deng CX. The role of TGF-β/SMAD4 signaling in cancer. Int J Biol Sci. 2018; 14:111–23.
https://doi.org/10.7150/ijbs.23230
PMID:29483830

75. Xu Y, Pasche B. TGF-beta signaling alterations and susceptibility to colorectal cancer. Hum Mol Genet. 2007; 16:R14–20.
https://doi.org/10.1093/hmg/ddl486 PMID:17613544

76. Greiner M, Pfeiffer D, Smith RD. Principles and practical application of the receiver-operating characteristic analysis for diagnostic tests. Prev Vet Med. 2000; 45:23–41.
https://doi.org/10.1016/s0167-5877(00)00115-x PMID:10802332

77. Wang Q, Huang Z, Ni S, Xiao X, Xu Q, Wang L, Huang D, Tan C, Sheng W, Du X. Plasma miR-601 and miR-760 are novel biomarkers for the early detection of colorectal cancer. PLoS One. 2012; 7:e44398.
https://doi.org/10.1371/journal.pone.0044398 PMID:22970209

78. Liu X, He B, Xu T, Pan Y, Hu X, Chen X, Wang S. MiR-490-3p Functions As a Tumor Suppressor by Inhibiting Oncogene VDAC1 Expression in Colorectal Cancer. J Cancer. 2018; 9:1218–30.
https://doi.org/10.7150/jca.23662 PMID:29675103

79. Tao K, Yang J, Guo Z, Hu Y, Sheng H, Gao H, Yu H. Prognostic value of miR-221-3p, miR-342-3p and miR-491-5p expression in colon cancer. Am J Transl Res. 2014; 6:391–401.
PMID:25075256

80. Xuan Y, Yang H, Zhao L, Lau WB, Lau B, Ren N, Hu Y, Yi T, Zhao X, Zhou S, Wei Y. MicroRNAs in colorectal cancer: small molecules with big functions. Cancer Lett. 2015; 360:89–105.
https://doi.org/10.1016/j.canlet.2014.11.051 PMID:25524553

81. Wang YN, Chen ZH, Chen WC. Novel circulating microRNAs expression profile in colon cancer: a pilot study. Eur J Med Res. 2017; 22:51.
https://doi.org/10.1186/s40001-017-0294-5 PMID:29187262

82. Huang Z, Huang D, Ni S, Peng Z, Sheng W, Du X. Plasma microRNAs are promising novel biomarkers for early detection of colorectal cancer. Int J Cancer. 2010; 127:118–26.
https://doi.org/10.1002/ijc.25007 PMID:19876917

83. Fesler A, Jiang J, Zhai H, Ju J. Circulating microRNA testing for the early diagnosis and follow-up of colorectal cancer patients. Mol Diagn Ther. 2014; 18:303–08.
https://doi.org/10.1007/s40291-014-0089-0 PMID:24566942

84. Lu J, Getz G, Miska EA, Alvarez-Saavedra E, Lamb J, Peck D, Sweet-Cordero A, Ebert BL, Mak RH, Ferrando AA, Downing JR, Jacks T, Horvitz HR, Golub TR. MicroRNA expression profiles classify human cancers. Nature. 2005; 435:834–38.
https://doi.org/10.1038/nature03702 PMID:15944708

85. Sun HL, Cui R, Zhou J, Teng KY, Hsiao YH, Nakanishi K, Fassan M, Luo Z, Shi G, Tili E, Kutay H, Lovat F, Vicentini C, et al. ERK Activation Globally Downregulates miRNAs through Phosphorylating Exportin-5. Cancer Cell. 2016; 30:723–36.
https://doi.org/10.1016/j.ccell.2016.10.001 PMID:27846390

86. Lee EJ, Baek M, Gusev Y, Brackett DJ, Nuovo GJ, Schmittgen TD. Systematic evaluation of microRNA processing patterns in tissues, cell lines, and tumors. RNA. 2008; 14:35–42.
https://doi.org/10.1261/rna.804508 PMID:18025253

87. Qu Y, Shi B, Hou P. Activated ERK: An Emerging Player in miRNA Downregulation. Trends Cancer. 2017; 3:163–65.
https://doi.org/10.1016/j.trecan.2017.01.002 PMID:28718429

88. Peng Y, Croce CM. The role of MicroRNAs in human cancer. Signal Transduct Target Ther. 2016; 1:15004.
https://doi.org/10.1038/sigtrans.2015.4 PMID:29263891

89. Chou CH, Shrestha S, Yang CD, Chang NW, Lin YL, Liao KW, Huang WC, Sun TH, Tu SJ, Lee WH, Chiew MY, Tai CS, Wei TY, et al. miRTarBase update 2018: a resource for experimentally validated microRNA-target interactions. Nucleic Acids Res. 2018; 46:D296–302.
https://doi.org/10.1093/nar/gkx1067 PMID:29126174

90. Shen Y, Han X, Wang J, Wang S, Yang H, Lu SH, Shi Y. Prognostic impact of mutation profiling in patients with stage II and III colon cancer. Sci Rep. 2016; 6:24310.
https://doi.org/10.1038/srep24310 PMID:27074743

91. Gao ZJ, Yuan WD, Yuan JQ, Yuan K, Wang Y. miR-486-5p functions as an oncogene by targeting PTEN in non-small cell lung cancer. Pathol Res Pract. 2018; 214:700–05.
https://doi.org/10.1016/j.prp.2018.03.013 PMID:29567332

92. Bronisz A, Godlewski J, Wallace JA, Merchant AS, Nowicki MO, Mathsyaraja H, Srinivasan R, Trimboli AJ, Martin CK, Li F, Yu L, Fernandez SA, Pécot T, et al. Reprogramming of the tumour microenvironment by stromal PTEN-regulated miR-320. Nat Cell Biol. 2011; 14:159–67.
https://doi.org/10.1038/ncb2396 PMID:22179046

93. Slattery ML, Mullany LE, Sakoda LC, Wolff RK, Stevens JR, Samowitz WS, Herrick JS. The PI3K/AKT signaling

pathway: Associations of miRNAs with dysregulated gene expression in colorectal cancer. Mol Carcinog. 2018; 57:243–61.
https://doi.org/10.1002/mc.22752 PMID:29068474

94. Balacescu O, Sur D, Cainap C, Visan S, Cruceriu D, Manzat-Saplacan R, Muresan MS, Balacescu L, Lisencu C, Irimie A. The Impact of miRNA in Colorectal Cancer Progression and Its Liver Metastases. Int J Mol Sci. 2018; 19:3711.
https://doi.org/10.3390/ijms19123711 PMID:30469518

95. Rokavec M, Li H, Jiang L, Hermeking H. The p53/microRNA connection in gastrointestinal cancer. Clin Exp Gastroenterol. 2014; 7:395–413.
https://doi.org/10.2147/CEG.S43738 PMID:25328413

96. Reid JF, Sokolova V, Zoni E, Lampis A, Pizzamiglio S, Bertan C, Zanutto S, Perrone F, Camerini T, Gallino G, Verderio P, Leo E, Pilotti S, et al. miRNA profiling in colorectal cancer highlights miR-1 involvement in MET-dependent proliferation. Mol Cancer Res. 2012; 10:504–15.
https://doi.org/10.1158/1541-7786.MCR-11-0342 PMID:22343615

97. Pamudurthy V, Bissonnette M, Konda V. Biomarkers in Colorectal Cancer Screening. J Gastrointest Dig Syst. 2016; 06:2–9.
https://doi.org/10.4172/2161-069X.1000389

98. Chakraborty C, Das S. Profiling cell-free and circulating miRNA: a clinical diagnostic tool for different cancers. Tumour Biol. 2016; 37:5705–14.
https://doi.org/10.1007/s13277-016-4907-3 PMID:26831657

99. Marcuello M, Vymetalkova V, Neves RPL, Duran-Sanchon S, Vedeld HM, Tham E, van Dalum G, Flügen G, Garcia-Barberan V, Fijneman RJ, Castells A, Vodicka P, Lind GE, et al. Circulating biomarkers for early detection and clinical management of colorectal cancer. Mol Aspects Med. 2019; 69:107–22.
https://doi.org/10.1016/j.mam.2019.06.002

100. Mansueto G, Benincasa G, Della Mura N, Nicoletti GF, Napoli C. Epigenetic-sensitive liquid biomarkers and personalised therapy in advanced heart failure: a focus on cell-free DNA and microRNAs. J Clin Pathol. 2020; 73:535–43.
https://doi.org/10.1136/jclinpath-2019-206404 PMID:32616540

101. Hayes CN, Chayama K. MicroRNAs as Biomarkers for Liver Disease and Hepatocellular Carcinoma. Int J Mol Sci. 2016; 17:280.
https://doi.org/10.3390/ijms17030280 PMID:26927063

102. Szejniuk WM, Robles AI, McCulloch T, Falkmer UG, Røe OD. Epigenetic predictive biomarkers for response or outcome to platinum-based chemotherapy in non-small cell lung cancer, current state-of-art. Pharmacogenomics J. 2019; 19:5–14.
https://doi.org/10.1038/s41397-018-0029-1 PMID:30190521

103. Rapado-González Ó, Álvarez-Castro A, López-López R, Iglesias-Canle J, Suárez-Cunqueiro MM, Muinelo-Romay L. Circulating microRNAs as Promising Biomarkers in Colorectal Cancer. Cancers (Basel). 2019; 11:898.
https://doi.org/10.3390/cancers11070898 PMID:31252648

104. Schou JV, Johansen JS, Nielsen D, Rossi S. Circulating microRNAs as Prognostic and Predictive Biomarkers in Patients with Colorectal Cancer. Noncoding RNA. 2016; 2:5.
https://doi.org/10.3390/ncrna2020005 PMID:29657263

105. Wang H, Cao F, Li X, Miao H, E J, Xing J, Fu CG. miR-320b suppresses cell proliferation by targeting c-Myc in human colorectal cancer cells. BMC Cancer. 2015; 15:748.
https://doi.org/10.1186/s12885-015-1728-5 PMID:26487644

106. Zhao D, Sui Y, Zheng X. MiR-331-3p inhibits proliferation and promotes apoptosis by targeting HER2 through the PI3K/Akt and ERK1/2 pathways in colorectal cancer. Oncol Rep. 2016; 35:1075–82.
https://doi.org/10.3892/or.2015.4450 PMID:26718987

107. Fujii T, Shimada K, Asano A, Tatsumi Y, Yamaguchi N, Yamazaki M, Konishi N. MicroRNA-331-3p Suppresses Cervical Cancer Cell Proliferation and E6/E7 Expression by Targeting NRP2. Int J Mol Sci. 2016; 17:1351.
https://doi.org/10.3390/ijms17081351 PMID:27548144

108. Guo X, Guo L, Ji J, Zhang J, Zhang J, Chen X, Cai Q, Li J, Gu Q, Liu B, Zhu Z, Yu Y. miRNA-331-3p directly targets E2F1 and induces growth arrest in human gastric cancer. Biochem Biophys Res Commun. 2010; 398:1–6.
https://doi.org/10.1016/j.bbrc.2010.05.082 PMID:20510161

109. Ferretti E, De Smaele E, Miele E, Laneve P, Po A, Pelloni M, Paganelli A, Di Marcotullio L, Caffarelli E, Screpanti I, Bozzoni I, Gulino A. Concerted microRNA control of Hedgehog signalling in cerebellar neuronal progenitor and tumour cells. EMBO J. 2008; 27:2616–27.
https://doi.org/10.1038/emboj.2008.172 PMID:18756266

110. Liu C, Li G, Yang N, Su Z, Zhang S, Deng T, Ren S, Lu S, Tian Y, Liu Y, Qiu Y. miR-324-3p suppresses migration

and invasion by targeting WNT2B in nasopharyngeal carcinoma. Cancer Cell Int. 2017; 17:2. https://doi.org/10.1186/s12935-016-0372-8 PMID:28053597

111. Peng Y, Dai Y, Hitchcock C, Yang X, Kassis ES, Liu L, Luo Z, Sun HL, Cui R, Wei H, Kim T, Lee TJ, Jeon YJ, et al. Insulin growth factor signaling is regulated by microRNA-486, an underexpressed microRNA in lung cancer. Proc Natl Acad Sci USA. 2013; 110:15043–48. https://doi.org/10.1073/pnas.1307107110 PMID:23980150

112. Christensen LL, Holm A, Rantala J, Kallioniemi O, Rasmussen MH, Ostenfeld MS, Dagnaes-Hansen F, Øster B, Schepeler T, Tobiasen H, Thorsen K, Sieber OM, Gibbs P, et al. Functional screening identifies miRNAs influencing apoptosis and proliferation in colorectal cancer. PLoS One. 2014; 9:e96767. https://doi.org/10.1371/journal.pone.0096767 PMID:24892549

113. Hu Y, Xie H, Liu Y, Liu W, Liu M, Tang H. miR-484 suppresses proliferation and epithelial-mesenchymal transition by targeting ZEB1 and SMAD2 in cervical cancer cells. Cancer Cell Int. 2017; 17:36. https://doi.org/10.1186/s12935-017-0407-9 PMID:28286418

114. Roberto VP, Gavaia P, Nunes MJ, Rodrigues E, Cancela ML, Tiago DM. Evidences for a New Role of miR-214 in Chondrogenesis. Sci Rep. 2018; 8:3704. https://doi.org/10.1038/s41598-018-21735-w PMID:29487295

115. Berezikov E. Evolution of microRNA diversity and regulation in animals. Nat Rev Genet. 2011; 12:846–60. https://doi.org/10.1038/nrg3079 PMID:22094948

116. Zhou K, Liu M, Cao Y. New Insight into microRNA Functions in Cancer: Oncogene-microRNA-Tumor Suppressor Gene Network. Front Mol Biosci. 2017; 4:46. https://doi.org/10.3389/fmolb.2017.00046 PMID:28736730

117. Gu J, Zhang J, Zheng L, Ajani JA, Wu X, Ye Y. Serum miR-331-3p predicts tumor recurrence in esophageal adenocarcinoma. Sci Rep. 2018; 8:14006. https://doi.org/10.1038/s41598-018-32282-9 PMID:30228315

118. Colaprico A, Silva TC, Olsen C, Garofano L, Cava C, Garolini D, Sabedot TS, Malta TM, Pagnotta SM, Castiglioni I, Ceccarelli M, Bontempi G, Noushmehr H. TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. Nucleic Acids Res. 2016; 44:e71. https://doi.org/10.1093/nar/gkv1507 PMID:26704973

119. Tomczak K, Czerwińska P, Wiznerowicz M. The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. Contemp Oncol (Pozn). 2015; 19:A68–77. https://doi.org/10.5114/wo.2014.47136 PMID:25691825

120. Kozomara A, Griffiths-Jones S. miRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Res. 2014; 42:D68–73. https://doi.org/10.1093/nar/gkt1181 PMID:24275495

121. Kang H. The prevention and handling of the missing data. Korean J Anesthesiol. 2013; 64:402–06. https://doi.org/10.4097/kjae.2013.64.5.402 PMID:23741561

122. Aguinis H, Gottfredson RK, Joo H. Best-Practice Recommendations for Defining, Identifying, and Handling Outliers. Organ Res Methods. 2013; 16:270–301. https://doi.org/10.1177/1094428112470848

123. Rousseeuw PJ, Hubert M. Robust statistics for outlier detection. Wiley Interdiscip Rev Data Min Knowl Discov. 2011; 1:73–79. https://doi.org/10.1002/widm.2

124. Ghasemi A, Zahediasl S. Normality tests for statistical analysis: a guide for non-statisticians. Int J Endocrinol Metab. 2012; 10:486–89. https://doi.org/10.5812/ijem.3505 PMID:23843808

125. Davis RB, Mukamal KJ. Hypothesis testing: means. Circulation. 2006; 114:1078–82. https://doi.org/10.1161/CIRCULATIONAHA.105.586461 PMID:16952996

126. Lamb TJ, Graham AL, Petrie A. T testing the immune system. Immunity. 2008; 28:288–92. https://doi.org/10.1016/j.immuni.2008.02.003 PMID:18341999

127. Gastwirth JL, Gel YR, Miao W. The Impact of Levene's Test of Equality of Variances on Statistical Theory and Practice. Stat Sci. 2009; 24: 343–60. https://doi.org/10.1214/09-STS301

128. Hart A. Mann-Whitney test is not just a test of medians: differences in spread can be important. BMJ. 2001; 323:391. https://doi.org/10.1136/bmj.323.7309.391

129. Reiner A, Yekutieli D, Benjamini Y. Identifying differentially expressed genes using false discovery rate controlling procedures. Bioinformatics. 2003; 19:368–75. https://doi.org/10.1093/bioinformatics/btf877 PMID:12584122

130. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful

Approach to Multiple Testing. J R Stat Soc B. 1995; 57:289–300.
https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

131. Agarwal R, Kumar B, Jayadev M, Raghav D, Singh A. CoReCG: a comprehensive database of genes associated with colon-rectal cancer. Database (Oxford). 2016; 2016:baw059.
https://doi.org/10.1093/database/baw059
PMID:27114494

132. Mukaka MM. Statistics corner: A guide to appropriate use of correlation coefficient in medical research. Malawi Med J. 2012; 24:69–71.
PMID:23638278

133. Hajian-Tilaki K. Receiver Operating Characteristic (ROC) Curve Analysis for Medical Diagnostic Test Evaluation. Caspian J Intern Med. 2013; 4:627–35.
PMID:24009950

134. In J, Lee DK. Survival analysis: part II - applied clinical data analysis. Korean J Anesthesiol. 2019; 72:441–57.
https://doi.org/10.4097/kja.19183
PMID:31096731

135. Bakhshi E, Ali Akbari Khoei R, Azarkeivan A, Kooshesh M, Biglarian A. Survival analysis of thalassemia major patients using Cox, Gompertz proportional hazard and Weibull accelerated failure time models. Med J Islam Repub Iran. 2017; 31:97.
https://doi.org/10.14196/mjiri.31.97
PMID:29951398

136. Goerdten J, Carrière I, Muniz-Terrera G. Comparison of Cox proportional hazards regression and generalized Cox regression models applied in dementia risk prediction. Alzheimers Dement (N Y). 2020; 6:e12041.
https://doi.org/10.1002/trc2.12041 PMID:32548239

137. Jager KJ, van Dijk PC, Zoccali C, Dekker FW. The analysis of survival data: the Kaplan-Meier method. Kidney Int. 2008; 74:560–65.
https://doi.org/10.1038/ki.2008.217
PMID:18596735

138. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2020. https://www.R-project.org/.

Supplementary Figure

Please browse Full Text version to see the data of Supplementary Figure 1.

Supplementary Tables

Please browse Full Text version to see the data of Supplementary Tables 1–5.

Supplementary Table 1. List of functional microRNA-target interactions (MTIs) experimentally validated for the 26 downregulated miRNAs retrieved from the MirTarBase dataset.

Supplementary Table 2. List of microRNA-target interactions (MTIs) experimentally validated for the 26 downregulated miRNAs comprising only the 172 genes reported to be involved in colorectal carcinoma development obtained from the Colorectal Cancer Gene Database (CoReCG).

Supplementary Table 3. CRC regulation status for the 172 miRNAs-target genes found to be involved in colorectal carcinoma development obtained from the Colorectal Cancer Gene Database (CoReCG).

Supplementary Table 4. Correlation coefficients between the 25 miRNAs and their 172 experimentally validated target-genes obtained from the Colorectal Cancer Gene Database (CoReCG).

Supplementary Table 5. MiRNA functional enrichment for the 25 downregulated miRNAs, performed using the Kyoto Encyclopedia of Genes and Genomes (KEEG) available at The Database for Annotation, Visualization and Integrated Discovery (DAVID) Bioinformatics Resources.

Supplementary Table 6. Area under the ROC curves (AUC) for the 25 downregulated miRNAs according to each dataset.

MiRNA name	TCGA	GSE30454	GSE33125	GSE18392
hsa-miR-125a-5p	0.998	0.795	1.000	0.942
hsa-miR-127-3p	0.880	0.839		0.794
hsa-miR-129-5p	0.904	0.792		0.801
hsa-miR-133b	0.999	0.975		0.750
hsa-miR-139-5p	0.965	0.944	0.944	0.993
hsa-miR-193a-5p	1.000	0.793	0.986	
hsa-miR-28-3p	1.000	0.753	0.905	
hsa-miR-296-5p	0.875	0.859		0.882
hsa-miR-299-5p	1.000	0.842		0.764
hsa-miR-320a	1.000		0.864	0.929
hsa-miR-320b	1.000		0.864	0.929
hsa-miR-320d	0.695		0.864	0.929
hsa-miR-324-3p	1.000	0.739		0.940
hsa-miR-324-5p	0.934	0.754		0.841
hsa-miR-326	1.000	0.819		0.986
hsa-miR-330-3p	0.890		0.903	0.885
hsa-miR-331-3p	0.941	0.855		0.696
hsa-miR-339-5p	1.000	0.741		0.621
hsa-miR-342-3p	0.853	0.843	0.875	0.945
hsa-miR-375	0.917		0.972	0.970
hsa-miR-423-3p	0.990	0.677		0.849
hsa-miR-484	0.995	0.681	0.889	0.968
hsa-miR-486-5p	1.000	0.680	0.952	0.875
hsa-miR-490-3p	0.912	0.970		0.690
hsa-miR-574-3p	1.000	0.848		0.950

Supplementary Table 7. Stage II overall survival (OS) results for the panels of miRNAs containing combinations of 2, 3, 4 and 5 simultaneous miRNAs from the 25 selected miRNAs.

Stage 2 overall survival (OS) analysis				
Combinations 2 by 2				
miRNA combinations	**p-value**	**Number of individuals in the Higher expression group**	**Number of individuals in the Lower expression group**	**Hazard ratio**
hsa-miR-342-3p - hsa-miR-331-3p	0.000183	25	29	4.15E-10
hsa-miR-342-3p - hsa-miR-574-3p	0.0045	25	26	2.97E-10
hsa-miR-342-3p - hsa-miR-326	0.007241	26	29	0.132538444
hsa-miR-127-3p - hsa-miR-331-3p	0.020103	28	29	0.124275996
hsa-miR-127-3p - hsa-miR-342-3p	0.026718	26	27	0.187519247
hsa-miR-342-3p - hsa-miR-324-5p	0.040052	30	34	0.257492788
Combinations 3 by 3				
miRNA combinations	**p-value**	**Number of individuals in the Higher expression group**	**Number of individuals in the Lower expression group**	**Hazard ratio**
hsa- miR- 342- 3p - hsa-miR- 331- 3p - hsa-miR- 324- 5p	0.000371	18	23	1.91E-10
hsa- miR- 342- 3p - hsa- miR- 331- 3p - hsa- miR- 484	0.000644	17	23	2.50E-10
hsa- miR- 342- 3p - hsa- miR- 331- 3p - hsa- miR- 320b	0.000797	16	19	1.67E-10
hsa- miR- 342- 3p - hsa- miR- 331- 3p - hsa- miR- 324- 3p	0.002392	19	20	1.63E-10
hsa- miR- 342- 3p - hsa- miR- 326 - hsa- miR- 331- 3p	0.00311	19	17	2.73E-10
hsa- miR- 342- 3p - hsa- miR- 331- 3p - hsa- miR- 339- 5p	0.0036	15	18	2.02E-10
hsa- miR- 320a - hsa- miR- 342- 3p - hsa- miR- 331- 3p	0.004512	16	17	2.03E-10
hsa- miR- 330- 3p - hsa- miR- 342- 3p - hsa- miR- 331- 3p	0.004909	17	16	2.08E-10
hsa- miR- 342- 3p - hsa- miR- 326 - hsa- miR- 324- 5p	0.006234	14	19	0.082704817
hsa- miR- 127- 3p - hsa- miR- 331- 3p - hsa- miR- 320b	0.007548	15	18	5.07E-10
hsa- miR- 342- 3p - hsa- miR- 331- 3p - hsa- miR- 574- 3p	0.00761	14	19	1.53E-10
hsa- miR- 342- 3p - hsa- miR- 331- 3p - hsa- miR- 193a- 5p	0.008237	14	17	1.63E-10
hsa- miR- 127- 3p - hsa- miR- 331- 3p - hsa- miR- 320d	0.008371	16	14	6.19E-10
hsa- miR- 127- 3p - hsa- miR- 326 - hsa- miR- 320b	0.008834	15	17	3.02E-10
hsa- miR- 342- 3p - hsa- miR- 331- 3p - hsa- miR- 28- 3p	0.009528	13	17	1.79E-10
hsa- miR- 342- 3p - hsa- miR- 331- 3p - hsa- miR- 423- 3p	0.010063	16	19	1.83E-10
hsa- miR- 342- 3p - hsa- miR- 484 - hsa- miR- 574- 3p	0.011389	14	18	2.04E-10
hsa- miR- 342- 3p - hsa- miR- 324- 3p - hsa- miR- 486- 5p	0.012555	17	15	2.17E-10
hsa- miR- 342- 3p - hsa- miR- 324- 5p - hsa- miR- 574- 3p	0.012883	17	21	3.52E-10
hsa- miR- 331- 3p - hsa- miR- 299- 5p - hsa- miR- 320d	0.015158	14	18	5.15E-10
hsa- miR- 127- 3p - hsa- miR- 320a - hsa- miR- 331- 3p	0.015451	16	17	5.20E-10
hsa- miR- 125a- 5p - hsa- miR- 342- 3p - hsa- miR- 574- 3p	0.017208	17	16	3.72E-10
hsa- miR- 342- 3p - hsa- miR- 326 - hsa- miR- 320b	0.017705	17	17	0.104040078
hsa- miR- 127- 3p - hsa- miR- 574- 3p - hsa- miR- 320b	0.019981	16	19	4.28E-10
hsa- miR- 342- 3p - hsa- miR- 574- 3p - hsa- miR- 320b	0.021539	13	18	1.16E-10
hsa- miR- 125a- 5p - hsa- miR- 342- 3p - hsa- miR- 331- 3p	0.02189	14	17	1.18E-10
hsa- miR- 129- 5p - hsa- miR- 133b - hsa- miR- 574- 3p	0.025347	15	15	2.74E-10
hsa- miR- 326 - hsa- miR- 574- 3p - hsa- miR- 320b	0.027324	19	22	0.166390667
hsa- miR- 342- 3p - hsa- miR- 484 - hsa- miR- 28- 3p	0.027724	15	19	0.114500372
hsa- miR- 342- 3p - hsa- miR- 339- 5p - hsa- miR- 574- 3p	0.02844	17	17	3.33E-10
hsa- miR- 342- 3p - hsa- miR- 324- 3p - hsa- miR- 484	0.029009	20	21	0.116029538
hsa- miR- 342- 3p - hsa- miR- 324- 5p - hsa- miR- 324- 3p	0.029462	20	24	0.129069723
hsa- miR- 342- 3p - hsa- miR- 331- 3p - hsa- miR- 486- 5p	0.029685	16	15	1.41E-10
hsa- miR- 331- 3p - hsa- miR- 299- 5p - hsa- miR- 320b	0.030283	18	21	6.25E-10
hsa- miR- 326 - hsa- miR- 331- 3p - hsa- miR- 320b	0.03173	23	28	0.278381887
hsa- miR- 342- 3p - hsa- miR- 324- 3p - hsa- miR- 320b	0.033438	17	22	0.129436651
hsa- miR- 127- 3p - hsa- miR- 331- 3p - hsa- miR- 486- 5p	0.033689	16	16	0.132743995
hsa- miR- 342- 3p - hsa- miR- 326 - hsa- miR- 193a- 5p	0.034183	17	19	0.126647253
hsa- miR- 127- 3p - hsa- miR- 342- 3p - hsa- miR- 299- 5p	0.034542	19	22	0.141023701
hsa- miR- 331- 3p - hsa- miR- 484 - hsa- miR- 299- 5p	0.036181	23	20	6.93E-10
hsa- miR- 296- 5p - hsa- miR- 423- 3p - hsa- miR- 484	0.036411	16	16	1791989256
hsa- miR- 342- 3p - hsa- miR- 326 - hsa- miR- 324- 3p	0.03694	17	18	0.12921111
hsa- miR- 127- 3p - hsa- miR- 320a - hsa- miR- 324- 5p	0.037925	16	19	5.73E-10
hsa- miR- 127- 3p - hsa- miR- 574- 3p - hsa- miR- 193a- 5p	0.038784	19	22	6.70E-10
hsa- miR- 299- 5p - hsa- miR- 574- 3p - hsa- miR- 193a- 5p	0.039263	22	22	7.03E-10
hsa- miR- 127- 3p - hsa- miR- 331- 3p - hsa- miR- 484	0.042952	18	18	7.30E-10
hsa- miR- 127- 3p - hsa- miR- 324- 5p - hsa- miR- 320b	0.042952	17	18	7.30E-10
hsa- miR- 127- 3p - hsa- miR- 331- 3p - hsa- miR- 324- 5p	0.042963	18	19	0.150536555
hsa- miR- 127- 3p - hsa- miR- 330- 3p - hsa- miR- 331- 3p	0.045563	13	18	0.142419981
hsa- miR- 127- 3p - hsa- miR- 326 - hsa- miR- 193a- 5p	0.045818	23	23	6.06E-10

hsa- miR- 139- 5p - hsa- miR- 342- 3p - hsa- miR- 574- 3p	0.047428	20	18	1.20E-09
hsa- miR- 299- 5p - hsa- miR- 574- 3p - hsa- miR- 320b	0.04939	20	21	4.59E-10
hsa- miR- 330- 3p - hsa- miR- 342- 3p - hsa- miR- 326	0.049694	22	21	0.2024995
hsa- miR- 127- 3p - hsa- miR- 324- 5p - hsa- miR- 486- 5p	0.049795	15	16	0.142344939

Combinations 4 by 4

miRNA combinations	p-value	Number of individuals in the Higher expression group	Number of individuals in the Lower expression group	Hazard ratio
hsa-miR-342-3p - hsa-miR-331-3p - hsa-miR-324-5p - hsa-miR-320b	0.000535	12	18	1.50E-10
hsa-miR-342-3p - hsa-miR-331-3p - hsa-miR-324-5p - hsa-miR-484	0.001118	13	20	1.19E-10
hsa-miR-342-3p - hsa-miR-331-3p - hsa-miR-484 - hsa-miR-320b	0.001827	12	18	7.11E-11
hsa-miR-342-3p - hsa-miR-331-3p - hsa-miR-324-5p - hsa-miR-324-3p	0.00253	13	17	1.68E-10
hsa-miR-342-3p - hsa-miR-331-3p - hsa-miR-324-3p - hsa-miR-484	0.004245	14	16	1.14E-10
hsa-miR-320a - hsa-miR-342-3p - hsa-miR-331-3p - hsa-miR-484	0.005893	13	17	1.33E-10
hsa-miR-326 - hsa-miR-331-3p - hsa-miR-324-5p - hsa-miR-320b	0.021147	18	22	0.2287878
hsa-miR-326 - hsa-miR-331-3p - hsa-miR-574-3p - hsa-miR-320b	0.021749	14	19	0.161899927
hsa-miR-326 - hsa-miR-324-5p - hsa-miR-574-3p - hsa-miR-320b	0.023646	13	19	0.164324
hsa-miR-342-3p - hsa-miR-324-5p - hsa-miR-324-3p - hsa-miR-484	0.026451	16	20	0.113930394
hsa-miR-331-3p - hsa-miR-324-5p - hsa-miR-299-5p - hsa-miR-320b	0.028684	13	17	5.04E-10
hsa-miR-326 - hsa-miR-331-3p - hsa-miR-339-5p - hsa-miR-320b	0.029346	15	18	0.188073968
hsa-miR-127-3p - hsa-miR-299-5p - hsa-miR-193a-5p - hsa-miR-320b	0.031653	12	18	0.128504824
hsa-miR-127-3p - hsa-miR-324-5p - hsa-miR-484 - hsa-miR-299-5p	0.035638	16	17	7.99E-10
hsa-miR-331-3p - hsa-miR-299-5p - hsa-miR-193a-5p - hsa-miR-320b	0.035737	14	17	5.35E-10
hsa-miR-331-3p - hsa-miR-324-5p - hsa-miR-484 - hsa-miR-299-5p	0.035855	17	16	6.87E-10
hsa-miR-326 - hsa-miR-331-3p - hsa-miR-320b - hsa-miR-320d	0.03676	14	21	0.257376989
hsa-miR-127-3p - hsa-miR-330-3p - hsa-miR-299-5p - hsa-miR-193a-5p	0.037357	16	16	5.67E-10
hsa-miR-127-3p - hsa-miR-324-5p - hsa-miR-299-5p - hsa-miR-193a-5p	0.040035	15	18	7.07E-10
hsa-miR-127-3p - hsa-miR-299-5p - hsa-miR-574-3p - hsa-miR-320b	0.040135	15	17	4.06E-10
hsa-miR-331-3p - hsa-miR-484 - hsa-miR-299-5p - hsa-miR-320b	0.040575	15	15	5.89E-10
hsa-miR-127-3p - hsa-miR-326 - hsa-miR-299-5p - hsa-miR-193a-5p	0.040887	19	20	5.82E-10
hsa-miR-127-3p - hsa-miR-331-3p - hsa-miR-484 - hsa-miR-299-5p	0.04118	17	16	7.16E-10
hsa-miR-342-3p - hsa-miR-324-5p - hsa-miR-324-3p - hsa-miR-320b	0.041366	13	20	0.139955123
hsa-miR-127-3p - hsa-miR-299-5p - hsa-miR-574-3p - hsa-miR-193a-5p	0.043999	18	19	7.11E-10
hsa-miR-125a-5p - hsa-miR-324-5p - hsa-miR-574-3p - hsa-miR-320b	0.047546	13	18	0.198082863
hsa-miR-331-3p - hsa-miR-484 - hsa-miR-299-5p - hsa-miR-574-3p	0.048803	17	16	6.34E-10

Combinations 5 by 5

miRNA combinations	p-value	Number of individuals in the Higher expression group	Number of individuals in the Lower expression group	Hazard ratio
hsa-miR-326-hsa-miR-331-3p-hsa-miR-339-5p-hsa-miR-484-hsa-miR-320b	0.031802	14	16	0.191311093

Supplementary Table 8. Stage II recurrence-free survival (RFS) results for the panels of miRNAs containing combinations of 2, 3, 4 and 5 simultaneous miRNAs from the 25 selected miRNAs.

Stage 2 recurrence-free survival (RFS) analysis				
Combinations 2 by 2				
miRNA combinations	**p-value**	**Number of individuals in the Higher expression group**	**Number of individuals in the Lower expression group**	**Hazard ratio**
hsa-miR-129-5p - hsa-miR-423-3p	0.012472997	20	23	9.843039717
hsa-miR-129-5p - hsa-miR-486-5p	0.03761568	17	20	1223790672
hsa-miR-331-3p - hsa-miR-320d	0.039159223	28	30	0.331858185
Combinations 3 by 3				
miRNA combinations	**p-value**	**Number of individuals in the Higher expression group**	**Number of individuals in the Lower expression group**	**Hazard ratio**
hsa-miR-129-5p - hsa-miR-324-3p - hsa-miR-423-3p	0.012135384	16	14	2891271744
hsa-miR-129-5p - hsa-miR-324-5p - hsa-miR-423-3p	0.018478048	16	14	3747635084
hsa-miR-423-3p - hsa-miR-484 - hsa-miR-299-5p	0.042449644	17	17	4.611534615
Combinations 4 by 4				
miRNA combinations	**p-value**	**Number of individuals in the Higher expression group**	**Number of individuals in the Lower expression group**	**Hazard ratio**
hsa-miR-320a - hsa-miR-324-3p - hsa-miR-423-3p - hsa-miR-484	0.025165735	15	25	1495885403
Combinations 5 by 5				
miRNA combinations	**p-value**	**Number of individuals in the Higher expression group**	**Number of individuals in the Lower expression group**	**Hazard ratio**
hsa-miR-320a - hsa-miR-324-5p - hsa-miR-324-3p - hsa-miR-423-3p - hsa-miR-484	0.042737763	13	24	1516837907
hsa-miR-320a - hsa-miR-324-3p - hsa-miR-423-3p - hsa-miR-484 - hsa-miR-320b	0.049294563	14	22	1725069778

Supplementary Table 9. Stage III overall survival (OS) results for the panels of miRNAs containing combinations of 2, 3, 4 and 5 simultaneous miRNAs from the 25 selected miRNAs.

	Stage 3 overall survival (OS) analysis			
	Combinations 2 by 2			
miRNA combinations	**p-value**	**Number of individuals in the Higher expression group**	**Number of individuals in the Lower expression group**	**Hazard ratio**
hsa-miR-133b - hsa-miR-574-3p	0.012378	21	23	0.232222735
hsa-miR-375 - hsa-miR-324-3p	0.012638	29	31	4.407768211
hsa-miR-129-5p - hsa-miR-324-3p	0.014643	20	20	5.499440216
hsa-miR-129-5p - hsa-miR-486-5p	0.019913	25	25	5.066314309
hsa-miR-133b - hsa-miR-320b	0.021693	21	21	0.200765726
hsa-miR-129-5p - hsa-miR-375	0.021949	27	29	3.964313324
hsa-miR-127-3p - hsa-miR-375	0.028557	23	26	3.316305702
hsa-miR-127-3p - hsa-miR-486-5p	0.029109	27	28	2.967651866
hsa-miR-375 - hsa-miR-486-5p	0.031743	30	32	2.889364248
hsa-miR-324-3p - hsa-miR-486-5p	0.033261	30	31	2.924417356
hsa-miR-127-3p - hsa-miR-324-3p	0.034171	25	26	3.252558631
hsa-miR-133b - hsa-miR-423-3p	0.036065	16	14	0.141561917
hsa-miR-125a-5p - hsa-miR-133b	0.03707	29	26	0.343839238
hsa-miR-125a-5p - hsa-miR-574-3p	0.040105	25	27	0.373019785
hsa-miR-125a-5p - hsa-miR-320b	0.042205	26	27	0.288138938
	Combinations 3 by 3			
miRNA combinations	**p-value**	**Number of individuals in the Higher expression group**	**Number of individuals in the Lower expression group**	**Hazard ratio**
hsa-miR-127-3p - hsa-miR-324-3p - hsa-miR-484	0.006047	15	16	7.109617001
hsa-miR-139-5p - hsa-miR-133b - hsa-miR-574-3p	0.00666	13	19	0.183225311
hsa-miR-129-5p - hsa-miR-375 - hsa-miR-330-3p	0.00815	15	15	10.62715282
hsa-miR-375 - hsa-miR-324-3p - hsa-miR-484	0.009241	19	25	9.673416356
hsa-miR-375 - hsa-miR-324-3p - hsa-miR-486-5p	0.010604	18	22	5.785941523
hsa-miR-375 - hsa-miR-324-3p - hsa-miR-299-5p	0.010849	17	20	9.195008842
hsa-miR-375 - hsa-miR-324-5p - hsa-miR-324-3p	0.015211	18	21	8.538590986
hsa-miR-375 - hsa-miR-326 - hsa-miR-324-3p	0.016561	19	22	8.384076984
hsa-miR-320a - hsa-miR-375 - hsa-miR-324-3p	0.019791	15	16	8.828596729
hsa-miR-324-5p - hsa-miR-324-3p - hsa-miR-486-5p	0.020048	23	22	4.372611975
hsa-miR-127-3p - hsa-miR-324-3p - hsa-miR-486-5p	0.020823	17	16	4.108012764
hsa-miR-324-3p - hsa-miR-484 - hsa-miR-486-5p	0.020911	21	25	4.149880583
hsa-miR-324-3p - hsa-miR-484 - hsa-miR-299-5p	0.02108	20	24	4.256428455
hsa-miR-324-3p - hsa-miR-486-5p - hsa-miR-299-5p	0.021618	17	20	4.987332351
hsa-miR-375 - hsa-miR-326 - hsa-miR-486-5p	0.022499	23	19	4.086629092
hsa-miR-375 - hsa-miR-331-3p - hsa-miR-484	0.023042	18	23	5.228028533
hsa-miR-127-3p - hsa-miR-324-3p - hsa-miR-299-5p	0.024471	20	19	5.035989942
hsa-miR-375 - hsa-miR-330-3p - hsa-miR-486-5p	0.025423	19	17	4.009611391
hsa-miR-331-3p - hsa-miR-324-3p - hsa-miR-299-5p	0.026531	22	26	3.557455347
hsa-miR-375 - hsa-miR-331-3p - hsa-miR-486-5p	0.031352	19	17	3.911348665
hsa-miR-375 - hsa-miR-324-3p - hsa-miR-339-5p	0.031628	16	18	5.059477829
hsa-miR-125a-5p - hsa-miR-574-3p - hsa-miR-320b	0.034116	14	17	0.255000527
hsa-miR-129-5p - hsa-miR-127-3p - hsa-miR-486-5p	0.035293	13	17	6.802202805
hsa-miR-127-3p - hsa-miR-484 - hsa-miR-486-5p	0.035369	14	17	3.820791631
hsa-miR-324-3p - hsa-miR-486-5p - hsa-miR-28-3p	0.036437	15	15	4.63958342
hsa-miR-375 - hsa-miR-324-3p - hsa-miR-28-3p	0.038203	16	16	7.255695673
hsa-miR-129-5p - hsa-miR-375 - hsa-miR-486-5p	0.041162	17	16	4.405278216
hsa-miR-331-3p - hsa-miR-486-5p - hsa-miR-299-5p	0.04183	19	20	3.542986862
hsa-miR-127-3p - hsa-miR-324-5p - hsa-miR-324-3p	0.042172	18	13	5.080609763
hsa-miR-129-5p - hsa-miR-375 - hsa-miR-326	0.042493	17	18	6.766177524
hsa-miR-330-3p - hsa-miR-324-3p - hsa-miR-299-5p	0.046184	18	21	4.411783414
	Combinations 4 by 4			
miRNA combinations	**p-value**	**Number of individuals in the Higher expression group**	**Number of individuals in the Lower expression group**	**Hazard ratio**
hsa.miR.375-hsa.miR.324.3p-hsa.miR.484-hsa.miR.486.5p	0.003367	12	18	1846827734
hsa.miR.375-hsa.miR.324.3p-hsa.miR.484-hsa.miR.299.5p	0.008283	11	19	1378705155

miRNA combinations	p-value	Higher expression group	Lower expression group	Hazard ratio
hsa.miR.375-hsa.miR.331.3p-hsa.miR.324.3p-hsa.miR.484	0.008617	14	21	10.24584142
hsa.miR.331.3p-hsa.miR.324.3p-hsa.miR.484-hsa.miR.299.5p	0.008625	17	21	5.463558923
hsa.miR.331.3p-hsa.miR.324.3p-hsa.miR.486.5p-hsa.miR.299.5p	0.00879	15	16	6.031759481
hsa.miR.375-hsa.miR.326-hsa.miR.324.3p-hsa.miR.484	0.009778	13	18	1596889310
hsa.miR.127.3p-hsa.miR.331.3p-hsa.miR.324.3p-hsa.miR.299.5p	0.013565	15	16	5.858868629
hsa.miR.331.3p-hsa.miR.484-hsa.miR.486.5p-hsa.miR.299.5p	0.014746	14	17	5.734555866
hsa.miR.331.3p-hsa.miR.324.3p-hsa.miR.484-hsa.miR.486.5p	0.015847	19	19	4.546265704
hsa.miR.330.3p-hsa.miR.331.3p-hsa.miR.324.3p-hsa.miR.299.5p	0.019236	15	18	5.80803277
hsa.miR.324.3p-hsa.miR.484-hsa.miR.299.5p-hsa.miR.320b	0.024586	14	16	5.804163282
hsa.miR.375-hsa.miR.324.5p-hsa.miR.324.3p-hsa.miR.484	0.024595	14	17	1639582946
hsa.miR.375-hsa.miR.324.3p-hsa.miR.339.5p-hsa.miR.484	0.024757	13	17	8.740073882
hsa.miR.330.3p-hsa.miR.324.3p-hsa.miR.484-hsa.miR.299.5p	0.028156	13	17	7.652459638
hsa.miR.375-hsa.miR.324.3p-hsa.miR.484-hsa.miR.320b	0.031191	13	17	7.651464677
hsa.miR.375-hsa.miR.326-hsa.miR.331.3p-hsa.miR.324.3p	0.034775	13	18	7.145246862
hsa.miR.330.3p-hsa.miR.331.3p-hsa.miR.484-hsa.miR.486.5p	0.039314	16	14	4.011014862
hsa.miR.324.3p-hsa.miR.339.5p-hsa.miR.484-hsa.miR.299.5p	0.039558	15	15	5.098768587
hsa.miR.324.5p-hsa.miR.324.3p-hsa.miR.484-hsa.miR.486.5p	0.041095	18	17	5.134269181
Combinations 5 by 5				
miRNA combinations	**p-value**	**Number of individuals in the Higher expression group**	**Number of individuals in the Lower expression group**	**Hazard ratio**
hsa-miR-331-3p - hsa-miR-324-5p - hsa-miR-324-3p - hsa-miR-484 - hsa-miR-486-5p	0.02038	17	14	8.81519013

Supplementary Table 10. Stage III recurrence-free Survival (RFS) results for the panels of miRNAs containing combinations of 2, 3, 4 and 5 simultaneous miRNAs from the 25 selected miRNAs.

miRNA combinations	p-value	Number of individuals in the Higher expression group	Number of individuals in the Lower expression group	Hazard ratio
Stage 3 recurrence-free survival (RFS) analysis				
Combinations 2 by 2				
hsa-miR-127-3p - hsa-miR-320d	0.006346	19	24	1753902556
hsa-miR-324-3p - hsa-miR-320d	0.012169	21	21	5.7592911
hsa-miR-342-3p - hsa-miR-320d	0.046197	18	17	6.702539172
Combinations 3 by 3				
hsa-miR-127-3p - hsa-miR-326 - hsa-miR-320d	0.002417	16	15	2807731991
hsa-miR-127-3p - hsa-miR-330-3p - hsa-miR-320d	0.002723	12	19	2489265019
hsa-miR-326 - hsa-miR-299-5p - hsa-miR-320d	0.012646	15	15	3478257329
hsa-miR-330-3p - hsa-miR-299-5p - hsa-miR-320d	0.013823	16	16	10.5674672
hsa-miR-127-3p - hsa-miR-299-5p - hsa-miR-320d	0.016057	14	16	1933160272
hsa-miR-331-3p - hsa-miR-324-3p - hsa-miR-320d	0.019325	17	19	5.644369439
hsa-miR-296-5p - hsa-miR-486-5p - hsa-miR-574-3p	0.026561	13	17	3.32E-10
hsa-miR-125a-5p - hsa-miR-324-3p - hsa-miR-484	0.028803	15	17	2033531460
hsa-miR-127-3p - hsa-miR-28-3p - hsa-miR-320d	0.029091	13	18	1571292250
hsa-miR-125a-5p - hsa-miR-331-3p - hsa-miR-484	0.033347	14	16	8.256186753
hsa-miR-127-3p - hsa-miR-331-3p - hsa-miR-324-3p	0.035378	16	18	3.974019274
hsa-miR-375 - hsa-miR-324-5p - hsa-miR-339-5p	0.041514	20	20	0.151915197
hsa-miR-375 - hsa-miR-339-5p - hsa-miR-423-3p	0.043281	19	22	0.313552875
hsa-miR-375 - hsa-miR-423-3p - hsa-miR-320b	0.045428	17	18	0.269099457
Combinations 4 by 4				
hsa-miR-331-3p- hsa-miR-324-3p- hsa-miR-484- hsa-miR-299-5p	0.013791	15	19	10.59870794
hsa-miR-331-3p- hsa-miR-324-3p- hsa-miR-339-5p- hsa-miR-484	0.013997	16	17	1736443836
hsa-miR-375- hsa-miR-324-5p- hsa-miR-339-5p- hsa-miR-423-3p	0.016913	15	16	0.117756367
hsa-miR-324-5p- hsa-miR-423-3p- hsa-miR-574-3p- hsa-miR-320b	0.021091	17	13	5.59E-10
hsa-miR-324-5p- hsa-miR-339-5p- hsa-miR-423-3p- hsa-miR-574-3p	0.024983	16	14	4.64E-10
hsa-miR-324-5p- hsa-miR-324-3p- hsa-miR-484- hsa-miR-299-5p	0.027121	16	16	1244231980
hsa-miR-324-5p- hsa-miR-339-5p- hsa-miR-423-3p- hsa-miR-486-5p	0.028999	16	17	0.132653626
hsa-miR-320a- hsa-miR-331-3p- hsa-miR-324-3p- hsa-miR-484	0.037118	14	16	9.067631054
hsa-miR-324-5p- hsa-miR-423-3p- hsa-miR-486-5p- hsa-miR-320b	0.038805	15	15	0.139366947
hsa-miR-331-3p- hsa-miR-484- hsa-miR-299-5p- hsa-miR-574-3p	0.039364	14	17	1735697193
Combinations 5 by 5				
hsa-miR-331-3p - hsa-miR-324-5p - hsa-miR-324-3p - hsa-miR-484 - hsa-miR-299-5p	0.029939	14	16	1202659728

Supplementary Table 11. Area under the ROC curves (AUC) for the 25 downregulated miRNAs according to each new dataset.

MIRNA name	GSE115513	GSE41655	GSE71008
hsa-miR-125a-5p	0.537	0.541	0.7017
hsa-miR-127-3p	0.603	0.755	0.5232
hsa-miR-129-5p	0.512	0.546	0.5568
hsa-miR-133b	0.500	0.990	
hsa-miR-139-5p	0.602	0.818	0.6873
hsa-miR-193a-5p	0.498	0.748	0.5599
hsa-miR-28-3p	0.511	0.499	0.6416
hsa-miR-296-5p	0.516	0.589	
hsa-miR-299-5p	0.688	0.567	
hsa-miR-320a	0.533	0.688	0.5922
hsa-miR-320b	0.555	0.891	0.5315
hsa-miR-320d	0.537	0.546	0.5737
hsa-miR-324-3p	0.563	0.791	0.5714
hsa-miR-324-5p	0.583	0.796	0.5346
hsa-miR-326		0.539	0.5086
hsa-miR-330-3p	0.665	0.792	0.5769
hsa-miR-331-3p	0.577	0.560	0.5539
hsa-miR-339-5p	0.787	0.519	0.5094
hsa-miR-342-3p	0.540	0.884	0.5191
hsa-miR-375	0.568	0.923	0.5238
hsa-miR-423-3p	0.623	0.471	0.6176
hsa-miR-484	0.531	0.835	0.5273
hsa-miR-486-5p	0.737	0.682	0.4702
hsa-miR-490-3p	0.778	0.422	0.545
hsa-miR-574-3p	0.520	0.710	0.5996

Supplementary Table 12. Stage III overall survival (OS) results for the panels of miRNAs containing combinations of 2 and 3 simultaneous miRNAs from the selected 8 miRNAs.

Stage 3 overall survival (OS) analysis				
Combinations 2 by 2				
miRNA combinations	*p-value*	**Number of individuals in the Higher expression group**	**Number of individuals in the Lower expression group**	**Hazard ratio**
hsa-miR-486-5p - hsa-miR-375	0.031743	30	32	2.889364
Combinations 3 by 3				
miRNA combinations	*p-value*	**Number of individuals in the Higher expression group**	**Number of individuals in the Lower expression group**	**Hazard ratio**
hsa-miR-486-5p - hsa-miR-330-3p - hsa-miR-375	0.025423	19	17	4.009611
hsa-miR-486-5p - hsa-miR-331-3p - hsa-miR-375	0.031352	19	17	3.911349