International Journal of Surgery xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

International Journal of Surgery

journal homepage: www.elsevier.com/locate/ijsu

Integrative systematic review meta-analysis and bioinformatics identifies MicroRNA-21 and its target genes as biomarkers for colorectal adenocarcinoma

Narjes Saheb Sharif-Askari[a], Fatemeh Saheb Sharif-Askari[a], Salman Yousuf Guraya[c], Riyad Bendardaf[b,c], Rifat Hamoudi[a,c,*]

[a] Sharjah Institute for Medical Research, College of Medicine, University of Sharjah, Sharjah, United Arab Emirates
[b] Oncology Unit, University Hospital Sharjah, Sharjah, United Arab Emirates
[c] Department of Clinical Sciences, College of Medicine, University of Sharjah, Sharjah, United Arab Emirates

ARTICLE INFO

Keywords:
Tissue/serum microRNA-21
Biomarkers
Colorectal cancer
Bioinformatics

ABSTRACT

Background: Advanced colorectal has poor survival and are difficult to treat. Therefore, there is an urgent need for biomarkers to diagnose this cancer at earlier manageable stages. Micro-RNAs (miRNAs) are amongst the most significant biomarkers that have shown promise in improving management and early detection of different types of cancers. However, since MiRNAs are non-coding, the main limitation of using them as biomarkers is that they do not have associated phenotype and therefore difficult to validate using other techniques. This makes it difficult to understand the mechanism of miRNA is disease initiation and progression, therefore any methodology that can provide semantics to miRNA expression would enhance the understanding of the role of miRNA in disease.

Methods: Here we report an integrative meta-analysis and bioinformatics methodology that showed microRNA-21 and its associated target mRNA to be the most significant predictive biomarkers for colorectal adenoma and adenocarcinoma. After drawing key inferences by meta-analysis, the authors then developed a bioinformatics method to identify mir-21 gene targeting in a specific tissue using two different bioinformatics approaches; absolute GSEA (Gene Set Enrichment Analysis) and LIMMA (Linear Models for MicroArray data) to identify differentially expressed genes of miRNA-21.

Results: Results from GSEA intersection with mir-21 gene targets was a subset of longer gene list that was obtained from the GEO2R intersect. In our study, both of longer GEO2R gene target list and the more focused GSEA list established the fact that mir-21 target numerous functional pathways that are mostly interconnected. Our three steps bioinformatics approach identified *ABCB1, HPGD, BCL2, TIAM1, TLR3,* and *PDCD4* as common targets for mir-21 in both of adenoma as well as adenocarcinoma suggesting they are biomarkers for early CRC.

Conclusions: The approach in this study proposed combining the big data from the scientific literature together with novel bioinformatics to bring about a methodology that can be used to first identify which microRNAs are involved in a specific disease, and then to identify a panel of biomarkers derived from the microRNAs target genes, and from these target genes the functional significance of these microRNAs can be inferred providing better clinical value for the surgeon.

1. Introduction

Colorectal cancer (CRC) is the third most common cancer worldwide and the fourth most common cause of mortality [1,2]. Compared to early stage, response to treatment and survival of *advanced stage* colorectal cancers remains poor, with 5-year survival rates of patients dropping from 50% to 10% in more advanced cases [3]. Surgical tumour resection remains the cornerstone of curative therapy for locally advanced colorectal carcinoma, with no curable treatments for the metastatic tumours that are unable to be surgically removed and in which the chemotherapy and radiation have proven to be less effective [4].

The inner epithelial lining of colorectal tissue is derived from endodermal cell of origin and turn in to adenocarcinoma upon continuous

* Corresponding author. Sharjah Institute for Medical Research, College of Medicine, University of Sharjah, Sharjah, United Arab Emirates.
E-mail address: rhamoudi@sharjah.ac.ae (R. Hamoudi).

https://doi.org/10.1016/j.ijsu.2019.11.017
Received 30 August 2019; Received in revised form 12 November 2019; Accepted 14 November 2019

accumulation of genetic and epigenetic abnormalities. Quantitative analysis of CRC has estimated a large window of around 10 years between occurrence of first mutation in stem cells transforming to malignant cells, and an additional 5 years thereafter for these neoplastic cells to obtain metastatic abilities [5]. In addition, 30–40% of patients develop recurrent disease following post-curative treatment [6,7].

Micro-RNAs (miRNAs), small non-coding RNAs of size 19–25 nucleotides [8], are amongst the most significant biomarkers that have shown promise in improving management and early detection of different types of cancer owing to their short structure making them amenable to degradation [9]. These oncogenes are highly expressed in cancer tissue and are secreted, embedded in exosomes, into body fluids such as serum and urine [10].

Amongst the many miRNAs, the role of mir-21 has been associated with different cancers, including CRC [11,12]. When compared with normal tissue, expression of mir-21 is found to be significantly higher in colorectal cancer tissue. Nevertheless, this expression is associated with inconsistent cancer prognosis and survival reports [13]. Although role of different miRNAs such as mir-21 have been examined in different systematic reviews and meta-analysis, these investigations focused on reporting or combing the outcomes and less attention has been paid to variabilities that were present in data and methodology used. Additionally, in the current literature there is an uncertainty about the function and gene targeting capability of miRNAs. For instance, there is scarce information on how a higher expression of certain miRNA could lead to a better or worse outcome. However, since miRNAs are non-coding, the main limitation of using them as biomarkers is that they do not have associated phenotype and therefore difficult to validate using other techniques. This makes it difficult to understand the mechanism of miRNA in disease initiation and progression, therefore any methodology that can provide semantics to miRNA expression would enhance the understanding of the role of miRNA in disease.

Although prognostic value of different miRNAs including mir-21 have been highlighted through previous systematic reviews and meta-analyses, however these investigations did not identify the downstream gene targets for these non-coding biomarkers and therefore the predictive accuracy of mir-21 as a biomarker will vary significantly in a heterogeneous disease such as CRC.

In this study an integrative meta-analysis followed by bioinformatics methodology were applied to identify predictive biomarkers for CRC. The systematic review and meta-analysis were conducted to quantitatively determine the prognostic significance of serum mir-21 expression signatures in CRC. This was followed by applying *in silico* bioinformatics methodology to determine how miRNA mechanistically leads to disease. This research further investigated what happens when they become deregulated, and to investigate different levels of gene expression of mir-21 gene targets across different tissues of normal, adenoma and adenocarcinoma. The findings of this research can potentially be applied to study the role of different miRNAs in other cancers.

2. Methods

A systematic review was conducted exploring the prognostic biomarkers for CRC, and from this review mir-21 was identified as a valuable biomarker. Owing to non-coding nature of microRNA a strategy was devised to identify the mir-21 gene targets in CRC. In the first part of this investigation a mir-21 systematic review and meta-analysis was conducted, and in the second part gene target analysis for mir-21 was carried out using bioinformatics (Fig. 1).

2.1. Meta-analysis

This systematic review and meta-analysis was conducted using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) [14] and Assessing the methodological quality of systematic reviews (AMSTAR) guidelines to determine the prognostic significance of serum microRNA-21 expression signatures in CRC.

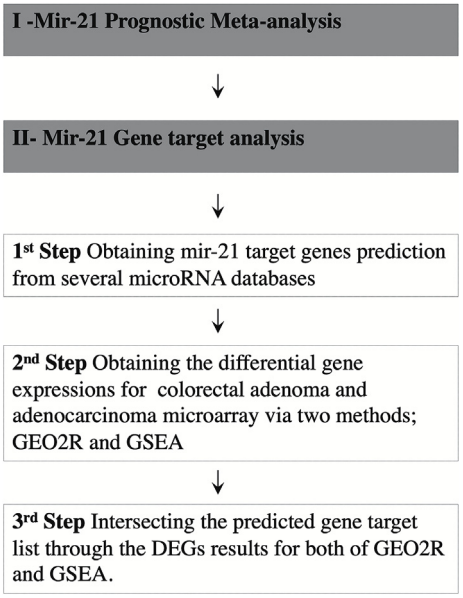

Fig. 1. Overall flow chart of the first step (meta-analysis) and second step (gene target analysis) of mir-21 investigation.

The MeSH terms serum/tissue, microRNA-21, prognosis, and colorectal cancer was used in the databases of Medline, Wiley online library, Cochrane library, Taylor and Francis Online, CINAHL, Springer, Proquest, ISI Web of knowledge, ScienceDirect, and Emerald for full-text English original research studies published during 2010–2017. During the search process, the following inclusion criteria were considered if the studies [1]; were conducted on the patients with CRC [2]; were performed on both the control and experimental cohorts of patients [3]; measured the expression of miR-21in serum or tissue [4]; investigated the association between miR-21 expression levels and prognosis of cancer or survival from cancer with hazard ratio (HR), and [5] Hazard ratio adjusted using multivariate cox regression analysis. Studies that investigated the prognostic significance of expression of miR-21in serum or tissue of the patients without CRC with normal subjects were excluded from the search. The review and editorial articles, letters to editors, brief communication, short communications and personal opinions and commentaries were also excluded.

2.1.1. Quality assurance

Two independent reviewers (N.S.A. and F.S.A.) objectively reviewed the selected studies and reached consensus by matching the inclusion criteria and MeSH terms. Data were extracted and organized according to the following criteria: first author, year of publication, country of origin, location of cancer, sample size, nature of sample whether serum or tissue, method of testing circulating miR-21, methods of normalization, HR of circulating miR-21 for overall survival (OS), and disease-free survival (DFS) as well as their corresponding 95% confidential interval (CI) to calculate the standard error (SE) as shown in (Table 1).

2.1.2. Statistical analysis

Meta-analysis was conducted by the Forest plot that graphically presents the consistency and reliability of results of the selected studies. The Forest plot was designed by following the steps recommended by Neyeloff et al. [26], where the effect size of each study is computed as an outcome and the pooled effect summary is calculated to observe the heterogeneity across studies. The Q test is the tool used for verifying heterogeneity with the null hypothesis that all studies are identical. In addition, the I squared (I^2) statistic is used to ensure the quantity of heterogeneity in percentages and the consistency data from the selected

Table 1
Characteristics of the studies selected for this systematic review and meta-analysis.

	Study	Origin	Disease	Specimen	N	Stage	Internal control	miR-21 assay	Cut-off	Survival analysis	Follow-up, months	HR (CI)
1	Schetter 2008 [15]	HK/USA	Colon	Tissue	103 (84 113)	I-IV	U6	qRT-PCR	Highest tertile	OS	HK 84.6, US 68	USA: HR, 2.70 (1.30–5.50) HK: HR, 2.40 (1.40–4.10)
2	Nielsen 2011 [16]	Denmark	Colorectal	Tissue	129 CC, (67 RC)	II	–	FISH		DFS	60	CC: HR, 1.26 (1.04–1.53) RC:HR, 0.97 (0.82–1.15)
3	Shibuya 2010 [17]	Japan	Colorectal	Tissue	156	Dukes A-D	U6	qRT-PCR	Mean	DFS, OS	44 (2–84)	DFS HR 0.40 (0.19–0.84) OS HR 0.51 (0.28–0.96)
4	Kjaer-Frifeldt 2012 [18]	Denmark	Colon	Tissue	520	II	–	FISH	Tertiles	DFS, OS	84	DFS HR, 1.41 (1.19–1.67) OS HR, 1.08 (0.97–1.22)
5	Toiyama 2013 [19]	Japan	Colorectal	T/Serum	166/188	I-IV	Tissue: mir-16, Serum: Cel-mir-39	qRT-PCR	Youden's Index	OS	60	Tissue: HR, 0.59 (0.21–1.63) Serum: HR, 4.12 (1.01–15.4)
6	Bovell 2013 [20]	USA	Colorectal	Tissue	55	IV	U6	qRT-PCR	NR	OS	Black 228, White 180	HR, 3.25 (1.37–7.72)
7	Chen 2013 [21]	Taiwan	Colorectal	Tissue	195	I-IV	U6	qRT-PCR	Mean	OS	60	HR, 2.56 (1.43–4.57)
8	Menendez 2013 [22]	Spain	Colorectal	Serum	102	I-IV	mir-16	qRT-PCR	Relative expression > 1	DFS, OS	23 (0–36)	DFS HR, 0.51 (0.25–1.06) OS HR, 0.50 (0.25–1.02)
9	Oue 2014 [23]	Japan/ Germany	Colon	Tissue	87 Japan, 145 Germany	I-IV	U48	qRT-PCR	Highest tertile	OS	72	Japan, HR, 3.13 (1.20–8.17) Germany, HR, 2.65 (1.06–6.66)
10	Kang 2015 [24]	Korea	Colorectal	Tissue	173 CC, 104 RC	II-III	–	ISH	Dichotomize	DFS, OS	80	DFS CC:HR, 2.45 (1.05–5.72) RC:HR, 1.65 (0.65–4.16) OS CC:HR, 0.43 (0.14–1.27) RC:HR, 2.05 (0.56–7.51)
11	Fukushima 2015 [25]	Japan	Colorectal	Tissue	306	I-IV	U6	qRT-PCR	Mean	DFS, OS	48	DFS HR, 2.94, (1.68–5.14) OS HR, 2.88, (1.70–5.08)

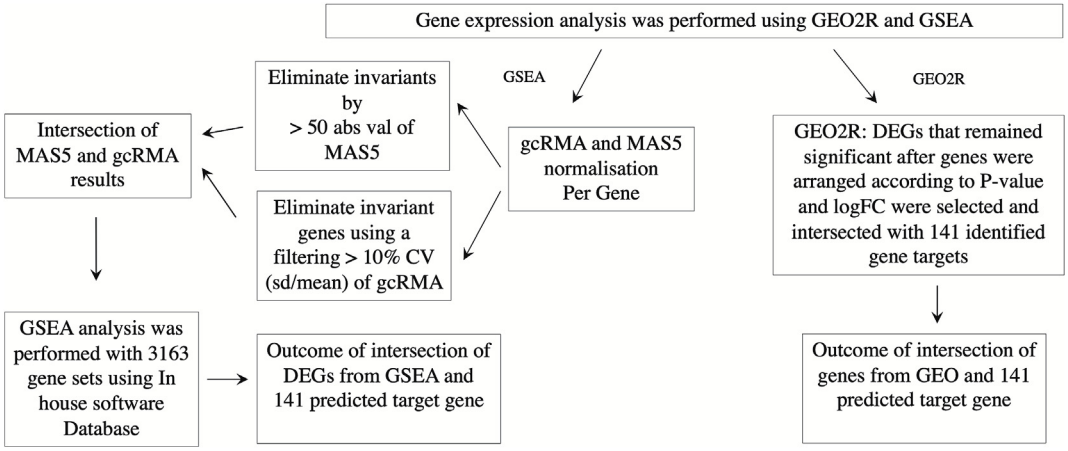

Fig. 2. Process of gene target identification and transcriptomic analysis.

studies [27]. After carefully analyzing the heterogeneity, next step was to apply appropriate effect summary model fixed effects or random effects model. The selected studies were analyzed through Review Manger 5.3 software developed by the Cochrane Library [28] and the level of significance was considered as 5% ($p < 0.05$). All HRs and 95% CI were calculated following Tierney's method and the log HR and standard error (SE) were used for the aggregation of survival results. Generally, an observed HR of > 1 implied worse survival [29].

2.2. Patients sample selection from publicly available data

The publicly available microarray datasets for colorectal adenocarcinoma (accession number: GSE20916, GSE21510, and GSE23878) and colorectal adenoma (accession number: GSE8671 and GSE14580) were obtained from National Center for Biotechnology Information Gene Expression Omnibus (NCIB GEO, http://www.ncbi.nlm.nih.gov/geo).

In this study we referred to non-cancerous tissue as that which is adjacent to adenocarcinoma area, and we referred to non-adenomatous tissue as that which is adjacent to adenoma area. From the colorectal adenocarcinoma datasets, a total of 249 different patient samples were identified, 73 normal colon (taken from non-cancerous areas of CRC patients) and 176 adenocarcinomas patients. From the adenoma datasets, a total of 70 different patient samples were identified, 32 adenoma tissue and 38 normal colons (taken from non-adenomatous tissue of patients with adenomas).

2.3. The target genes of the miRNA identified through bioinformatics

After drawing key inferences by meta-analysis, the authors then developed a bioinformatics method to identify mir-21 gene targeting in a specific tissue using two different bioinformatics analysis approaches; GSEA (Gene Set Enrichment Analysis) and LIMMA (Linear Models for MicroArray data) to identify differentially expressed genes of miRNA-21.

Firstly, a list of computationally predicted and experimentally validated gene targets were obtained from four different microRNA target databases. To obtain the gene targets for the miRNA identified from the meta-analysis, TargetScan, MiRDB, and Miranda databases were accessed. The top 25% target genes were retrieved from each database and the intersection results of these three databases were added to the experimentally validated gene targets provided by the fourth database, mirTarBase. These gene targets were then further validated by transcriptomics analysis using microarray database source.

Secondly, gene expression analysis was carried out. Here two alternatives tools were proposed for identification of differentially expressed genes (DEGs) in colorectal adenocarcinoma.

The first tool utilized the GEO2R publicly available source to generate an overall gene target list for mir-21, whereas the second tool utilized a modified gene set enrichment analysis (GSEA) method designed to generate more focused list of gene targets.

2.3.1. GEO2R

For colorectal adenocarcinoma dataset (GSE20916), the differentially expressed genes (DEGs) between CRC and normal colon from non-cancerous areas of CRC patients were analyzed separately using the interactive and publicly available web tool GEO2R (https://www.ncbi.nlm.nih.gov/geo/info/geo2r.html). GEO2R performs comparisons on selected cases using the GEOquery and limma R packages from the Bioconductor open source software. Correction for false positive results was performed using the Benjamini & Hochberg false discovery rate method which is selected by default in GEO2R. The output results of GEO2R analysis was arranged according to P-value (smallest to largest) and logFC (largest to smallest) and all DEGs that remained significant were selected.

2.3.2. GSEA

For comparison of GEO2R results, we repeated colorectal gene expression analysis using previously published method of GSEA [30]. Briefly, data were first normalized using MAS5 and gcRMA normalization methods, filtered using two non-specific filtering methods to eliminate the non-expressed and non-variant genes, and then results of two filtering methods were intersected to obtain a common set of variant genes. Normalized and filtered genes were used to analyse the following 3163 gene sets; hallmark gene sets, C2-curated gene set, and C5-biological process and C5-molecular function from Gene Ontology (GO). The list of genes that were repeated more than three times in the said gene sets were selected and used in transcriptomic analysis following the approach of Hamoudi et al. [30]. This GSEA analyses were performed using in-house scripts written in R [30]. Compared to the standard GSEA identifying either the up or down regulated genes, our GSEA method isolated both the up and down gene regulations by including additional step of absolute GSEA [31]. Process of gene target identification and transcriptomic analysis displayed in Fig. 2.

Thirdly, for the transcriptomic validation, DEGs lists obtained from two methods of GEO2R and GSEA analysis were intersected with the mir-21 gene target list obtained earlier from four different miRNA gene target databases.

3. Results

3.1. Meta-analysis

During initial search, 430 articles of CRC were retrieved, later,

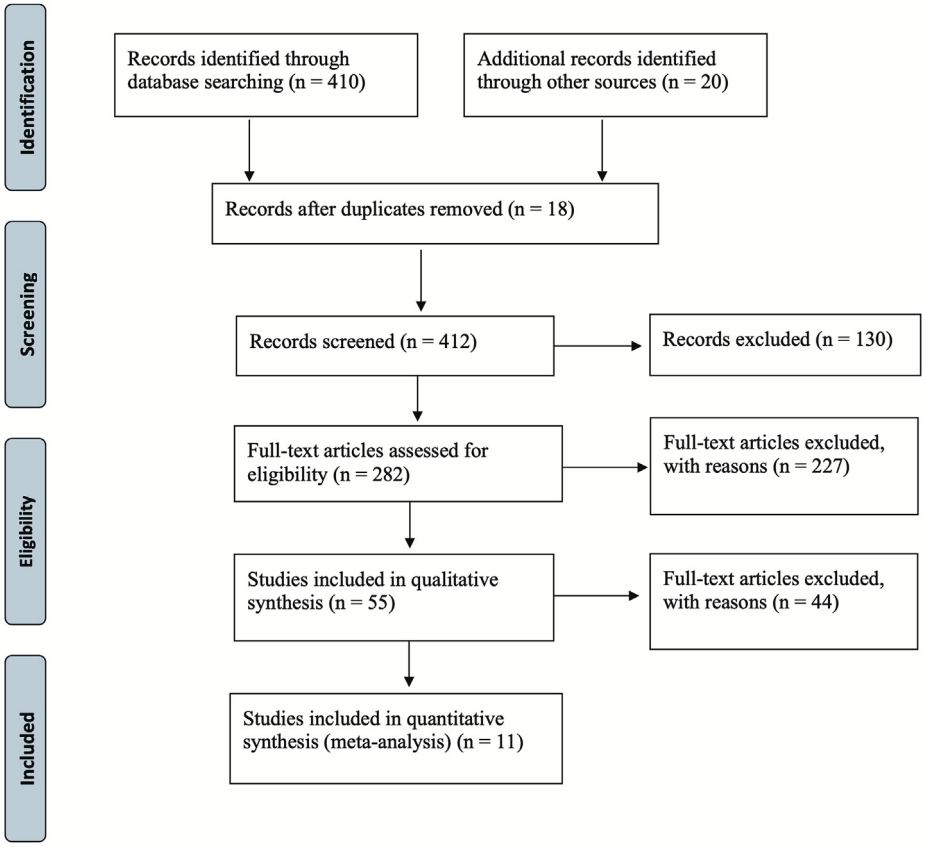

Fig. 3. Flow diagram showing the selection of studies in this systematic review and meta-analysis.

during the review of titles and abstracts, 375 studies were excluded as they were found to be irrelevant. Only 55 studies were found to be relevant as they empirically explored the prognostic significance of serum microRNA-21 expression signatures as effective biomarker in colorectal carcinomas. During the full text analysis of these 55 relevant studies, 44 were furtherer excluded due to incomplete data that could not fulfil the required criteria. Finally, a total of 11 relevant studies were selected for this meta-analysis. Fig. 3 shows the flow chart for process of study selection, while Table 1 shows the characteristics of the studies selected for this systematic review and meta-analysis.

The pooled HRs for the two outcomes of OS and DFS were measured separately using the random model effect. The results of the analysis for patients with CRC indicated that while high value of micoRNA-21 was significantly linked to worse OS with the pooled HR of 1.75 (95% CI 1.23–2.51, *p* value of 0.001), this overexpression and DFS outcome link was not significant with pooled HR of 1.21 (95% CI 0.91–1.60, *p* value of 0.19), and only showed a trend toward worse relapse rate. The forest plots for OS and DFS are presented in Figs. 4 and 5, respectively.

Then a subgroup analysis was carried out to compare the prognostic value for mir-21 expression values obtained from tissue versus serum. In this meta-analysis most of specimens were obtained from tissue (12 out of 14 cohorts) and only two studies analyzed serum levels of mir-21. Serum expression of mir-21 showed a trend towards worse OS in CRC, but this association was not significant. In contrast to serum expression, increase in tissue expression of mir-21 was significantly associated with worse OS in CRC (pooled HR of 1.88; 95% CI 1.30–2.74 with a *p* value of 0.0009).

Lastly, a subgroup was carried out to pool HR for all studies that have used the qrt-PCR method of mir-21 analysis, excluding the few studies that have measured the mir-21 level using FISH or microarray techniques. The result of this subgroup analysis displayed that compared to earlier analysis conducted using all the 21 studies, filtering only studies with qrt-PCR (11 studies) maintained the significant association between mir-21 and worse OS, and decreased the level of heterogeneity from I^2 = 78% ($P < 0.001$) *to* I^2 = 63% ($P = 0.01$). Results for all subgroup analysis are demonstrated in Table 2.

3.2. Bioinformatics gene targeting method

A list of 141 computationally predicted and experimentally validated gene targets were obtained from TargetScan, MiRDB, Miranda, and mirTarBase microRNA target databases (Fig. 6).

Primarily, in order to compare the two methods of GEO2R and GSEA, the GSE20916 dataset of 24 normal colons from non-cancerous areas of CRC patients and 36 adenocarcinomas cases were selected and analyzed. In GEO2R analysis of this set, 9110 DEGs remained significant after all genes were arranged according to *p*-value and logFC and were intersected with 141 identified gene targets, resulting in a list of 85 genes. Next, GSEA analysis was performed with 3163 gene sets using in house software Database. In GSEA analysis, 615 DEGs were selected and intersected with 141 identified gene targets, resulting in a list of 17 genes. Result of GSEA intersect with mir-21 gene targets was a subset of longer gene list that was obtained from the GEO2R intersect.

Because the GEO2R method could not be used for more than one dataset and generated a longer gene target list, the remaining of analyses were conducted using the GSEA. In two separate analysis the CRC adenocarcinoma tissues were compared to the adjacent non-cancerous areas and the CRC adenoma tissues were compared to the adjacent non-adenomatous areas. Firstly, 140 adenocarcinoma and 49 normal adjacent tissues were obtained from two datasets of GSE21510, and GSE23878 and analyzed using GSEA. In this analysis, 904 DEGs were selected and intersected with 141 identified gene targets, resulting in a list of 17 genes. Following, combing the results of first GSEA using GSE20916 dataset (24 normal non-cancerous and 36 adenocarcinoma) and the second GSEA using two datasets of GSE21510 and GSE23878 (49 normal non-cancerous and 140 adenocarcinoma) generated a list of

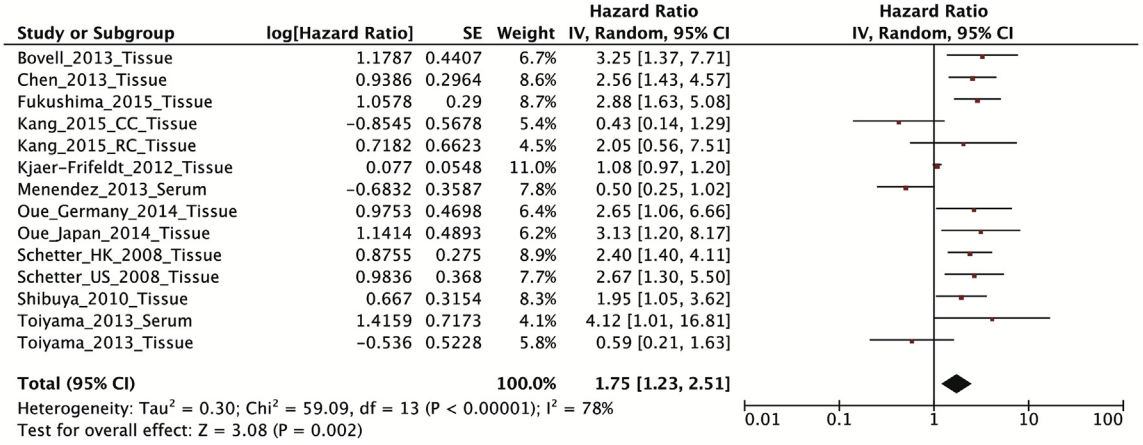

Fig. 4. Forest plot for overall survival (OS) of all the colorectal adenocarcinomas.

27 unique gene targets. Following we wanted to use the selected 27 gene set for differentiating the adenoma from normal non-adenomatous areas cases using two datasets of GSE8671 and GSE14580, and also to differentiate the Duke I colorectal adenocarcinoma from the remaining cases staged II to IV using two datasets of GSE21510, and GSE23878.

In the earlier step, literature review and meta-analysis revealed that the mir-21 expression level was increased with progression of colorectal cancer, and it is expected that this microRNA regulated gene expression system by reducing the translation of 114 identified direct target genes. However, the GSEA revealed that among the differential expressed direct gene targets, not all were downregulated by mir-21 but we focused our results on the differentially down regulated target genes.

Running the GSEA using the identified microRNA 27 gene set, revealed that *BCL2, TLR3, PDCD4, RASGRP1, ABCB1,* and *TIAM1* were down regulated in colorectal adenoma versus normal comparison (Fig. 7a). Next, running the GSEA for adenocarcinoma versus normal analyses revealed that *CLU, HPGD, TLR3, TIAM1, PDCD4, ABCB1, BCL2*, and *SMAD7* were down regulated in adenocarcinoma compared to normal (Fig. 7b). Intersecting the downregulated genes in adenocarcinoma and adenoma showed that 6 genes were commonly targeted by mir-21. Finally, the GSEA for Duke I CRC versus Duke cases staged II to IV did not reveal any differentially expressed gene (*P=0.42*).

4. Discussion

4.1. Meta-analysis

Initially, a meta-analysis was carried out for better understanding of existing link between high mir-21 expression and patient's outcomes of relapse, i.e. DFS and overall survival. By this mir-21 meta-analysis the authors identified a large variation in microRNA assay methods ranging from source and type of specimen to quantification techniques. These method discrepancies are often overlooked by other meta-analyses, which tend to focus more on combining studies results and show the need for standardization of microRNA measurement. In this analysis the differences in sociodemographic and clinical characteristics such as disease stage were adjusted by including only studies with multivariate cox regression analysis. The pooled result showed that high mir-21 expression in CRC studies was significantly associated with poor OS, while DFS outcome varied greatly and did not show an identifiable link. Additionally, the prognostic value of miR-21 was compared between serum and tissue and was found that although the tissue expression from both CRC was connected to significantly poor OS, serum overexpression showed only a trend toward a poor OS, but this result was not significant because HRs varied greatly. The non-uniform multivariable adjusted outcomes with matching mir-21 source of either tissue or serum could have been influenced by presence of methodological variations. For instance, in CRC tissue subgroup, studies varied by kind of tissue, formalin-fixed, paraffin-embedded versus fresh frozen, and type miRNA measurements techniques (qrt-pcr, FISH, or microarray). Among studies following the qrt-pcr method, results could have varied because of non-standard technique, different machines, different endogenous control, variable cut-off value for miRNA, and different normalization method.

The measurement of miRNAs expression and their prognostic value could be affected by the normalization method. In our meta-analysis, the following endogenous control were used for tissue derived mir-21 qrt-PCR method: RNU6B (U6 snRNA) [25] and mir-16 [19]. To normalize the circulating mir-21, studies also used different methods ranging from exogenous control (cel-mir-39) [19] or endogenous control of mir-16. Although, lack of consensus exists regarding the selection of optimal normalizers of miRNA expression levels from both tissue and serum specimen, best results were obtained when first miRNA was normalized with respect to average value of more than one internal or endogenous controls, and second when exogenous control was used beside the endogenous one. Vandesompele et al. have argued that

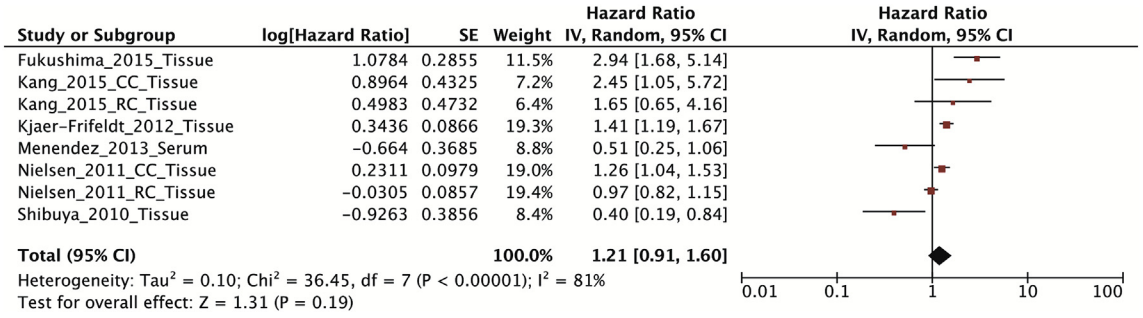

Fig. 5. Forest plot for disease-free survival (DFS) of all the colorectal adenocarcinomas studies.

Table 2
Results of subgroup analysis.

Outcome	Group	Number of cohorts	Model	HR (95% CI)	*P* value	Heterogeneity (Higgins I2 statistic)
DFS	All	8	Random	1.21 (0.91–1.60)	0.19	Q = 36.45, df = 7 (P= < 0.001), I^2 = 81%
OS	All	14	Random	1.75 (1.23–2.51)	< 0.001	Q = 59.09, df = 13 (P < 0.001), I^2 = 78%
	Colorectal blood	2	Random	1.32 (0.17, 1.0.21)	0.79	Q = 6.85, df = 1 (P = 0.009), I^2 = 85%
	Colorectal tissue	12	Random	1.88 (1.30–2.74)	0.0009	Q = 50.10, df = 11 (P= < 0.001), I^2 = 78%
	qrt-PCR	11	Random	2.05 (1.41–2.97)	< 0.001	Q = 26.88, df = 10 (P = 0.003), I^2 = 63%

normalization using single gene lead to relatively large errors in significant proportion of sample tested, and have demonstrated that reliable normalization could be obtained by using geometric mean, rather than arithmetic mean, of carefully selected internal controls [32]. For miRNA assay, whereas internal control is used for control of sample quality, external controls are also needed to decrease technical variability as an spike-in method [33]. In this meta-analysis no study applied both endogenous and exogenous controls for normalization of mir-21 level.

4.2. Using bioinformatics to identify mir-21 target genes

MicroRNAs are involved in the posttranscriptional regulations via binding to the untranslated 3' part of messenger RNA. This information together with several other parameters are considered in microRNA databases such as TargetScan, MiRDB, and Miranda to predict gene targets for a specific microRNA. However, these databases utilize different algorithms and mathematical models and hence they predict different gene targets and scores for a named microRNA. Some of these computational predictions, even the highly scored targets, could be false positive and can only be confirmed experimentally, a laborious task and at present only few targets have been experimentally validated [34]. There is no evidence for superiority of any of these microRNA databases, therefore, in this analysis gene target prediction errors were reduced by taking intersection of named three computationally assessed databases. This result was added to gene target list that were obtained from a fourth experimentally validated database, mirTarBase and the final list were searched among differentially expressed genes that were obtained using two independent methods of GEO2R and GSEA.

Subsequently, the differentially expressed genes in colorectal adenocarcinoma and adenoma were obtained using GEO2R and were validated with a second method, GSEA. Compared to GSEA, GEO2R method provides a simple publicly available tool that doesn't require additional skills such as knowledge of R language. The GEO2R, utilizes a limma package to generate an arbitrary list of differentially expressed genes, while GSEA genes are arranged according to known pathways or functions. Interestingly our result of GSEA intersect with mir-21 gene targets was a subset of longer gene list that was obtained from the GEO2R intersect. This approach could be applied for studying any other microRNAs association with a heterogeneous disease by using GEO2R publicly available method to get a broad but an inclusive gene targets, or apply our more advanced GSEA method, which when compared with the standard GSEA method isolate both the up as well as down gene regulations. This will help to cope with heterogenous data and could be used to decipher the molecular signature from such data.

Previously, isolated studies linked the increase in mir-21 level to increase in tumour size, proliferation, invasion, metastasis, and resistance to chemotherapy, and showed that mir-21 exert its carcinogenic effect by modulation of tumour suppressor and inflammatory networks such as NF-κB, PI3K/AKT, BCR signaling and TGF-β [35]. Additionally, mir-21 expression was reported to increase in stepwise manner during progression and transition of healthy colorectal mucosa between stages of precancerous colorectal adenoma and advanced carcinoma [19].

In this analysis we compared the dysregulation of mir-21 target genes during the transition of normal colorectal mucosa to adenoma and carcinoma. The results revealed that mir-21 down-regulation of the gene targets such as *PDCD4* starts from very early stage of adenoma where compared to normal tissues these gene targets were found to be significantly downregulated. These downregulations continue throughout the transition from adenoma to different stages of carcinoma with no significant changes in gene regulation were found from comparing first, the adenoma versus adenocarcinoma, and second comparing the Duke I colorectal adenocarcinoma from the remaining cases staged II to IV. Between the CRC gene targets of mir-21 the 6 genes that were down regulated in both of colorectal adenoma and adenocarcinoma. Table 3 displays gene targets studied by previous investigations. Although the link between these gene targets and colorectal cancer have been experimentally validated previously, the interaction between mir21 and their target genes have been reported only for *PDCD4* [36] and thereby further vivo and invitro validations required for the rest of identified gene targets to better understand the role of mir-21 in induction and progression of colorectal cancer.

The current clinical evidence is less clear about the adequate margin for colorectal tumour resection, particularly in setting of early colorectal adenocarcinoma. And although identification of mi21 level by itself might not be enough to guide the adequate surgical margin, using the method described in this study could provide a list of mir-21 targets that could be used not only to cross validate the mir21 level but also as a marker of early cancer.

To make the treatment more effective, colorectal cancer should be detected at an earlier stage where it is easily operable. However, even after complete resection of the colorectal tumour there is a need for active surveillance to avoid any missed adenomas and tumours in addition to detecting any recurrent adenoma and adenocarcinoma at a curable stage and before metastasis [4]. Although colonoscopy is considered as the gold standard for detecting colorectal cancer, it has been reported to have high missed-rate of up to 60% particularly for detection of flat polyps [47,48]. Hence, there is need for additional measures besides colonoscopy especially for early colorectal cancers.

Previously, Yamamichi et al. [49] analyzed miR-21 expression patterns in different stages of CRC, and discovered that higher miR-21

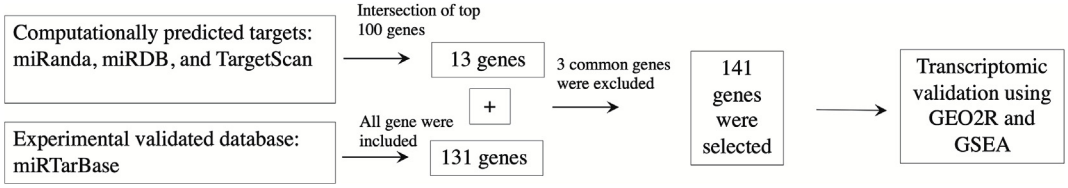

Fig. 6. Mir-21 gene target identification process.

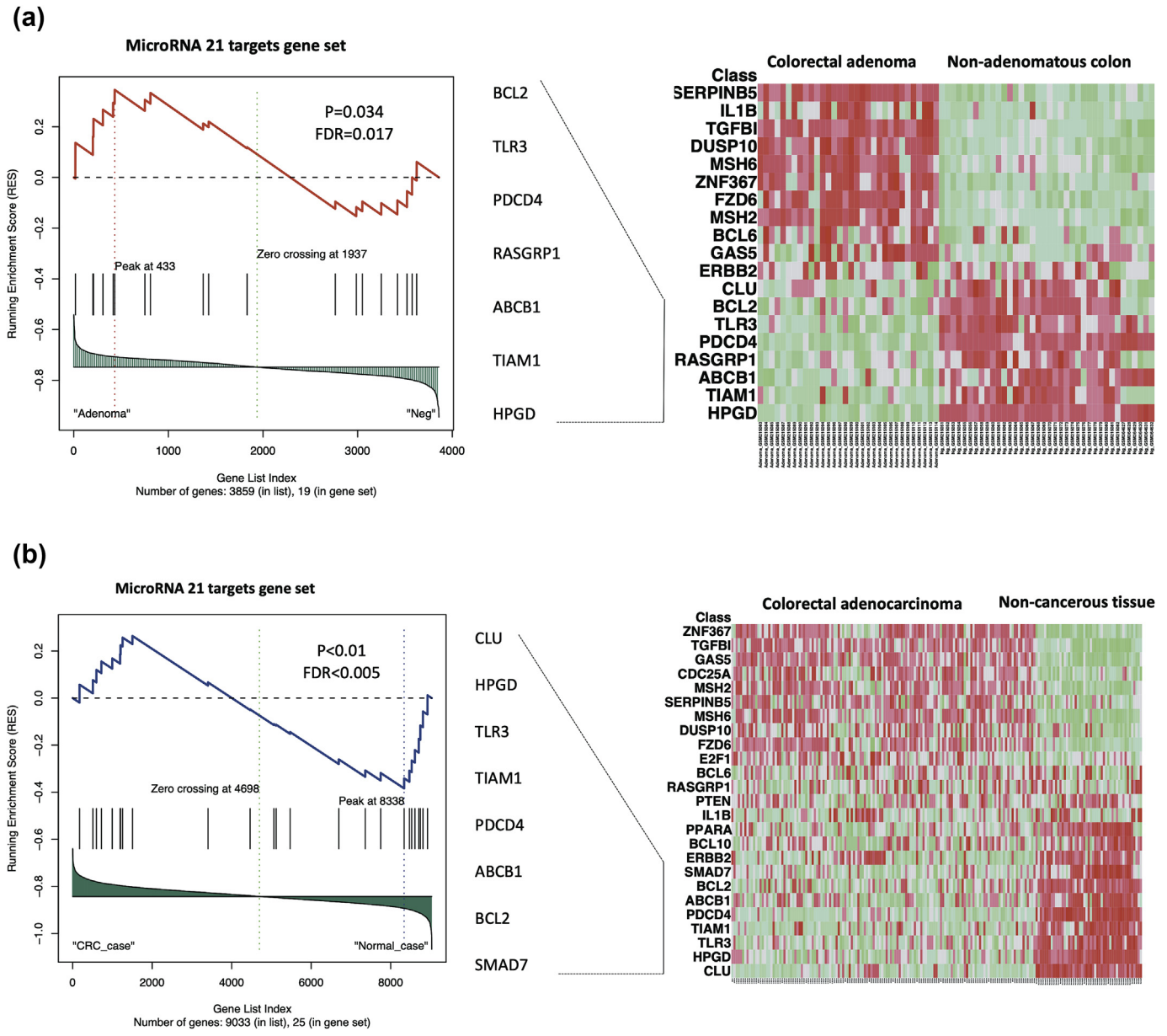

Fig. 7. GESA of mir-21 target genes. Heatmaps showing differential mRNA expression in normal colorectal versus a) adenoma and b) adenocarcinoma. The significant downregulated genes are shown.

Table 3
Expression of the main gene targets in colorectal adenocarcinoma disease.

Gene targets	References	Methods	Gene regulations
ABCB1	[37,38]	qrtPCR	Downregulated
HPGD	[39,40]	WB and NB	Downregulated
BCL2	[41,42]	IHC	Downregulated
TIAM1	[43]	qrtPCR, WB, Reporter assay, IHC	Downregulated
TLR3	[44,45]	IHC (Image analysis software)	Downregulated
PDCD4	[46]	IHC and WB	Downregulated

IHC, immunohistochemistry; NB, Northern blot; qRT-PCR, Real-Time Quantitative Reverse Transcription Polymerase Chain Reaction; WB, Western blot.

expression in pre-cancerous adenomas but not in non-cancerous polyps. In the current meta-analysis and bioinformatics, we have added that mir-21 and it is target genes could be used to detect adenoma from that of normal colorectal tissue as well as to detect the adenocarcinoma from healthy colorectal tissues.

In this meta-analysis study, the circulating levels of mir-21 were available only from two separate reports that showed a trend towards worse prognosis, however, this would benefit from further confirmation on larger number of studies. Taken together, the expression level of circulating mir21 and its gene targets (ABCB1, HPGD, BCL2, TIAM1, TLR3, and PDCD4) could potentially help in lowering the missed rates during colonoscopy of adenoma and adenocarcinoma as well as providing a more accurate delineation of surgical margins through immunohistochemistry of the target proteins on colorectal biopsies taken during colonoscopy. Having a panel of biomarkers can also identify cases where the tumour size or lesion is small which may lead the gastroenterologist to carry out more focused examination of the patient. In addition, the time and cost of measuring miRNAs expression is decreasing through the use of better technologies as well as faster experimental protocol [50,51]. Finally, a combination of mir-21 and its target genes could be useful in long term follow-up and active surveillance of colorectal patients' post-surgery".

5. Conclusion

This meta-analysis identified mir-21 as biomarker for CRC. In addition, our current three steps bioinformatics approach identified *ABCB1, HPGD, BCL2, TIAM1, TLR3,* and *PDCD4* as common targets for mir-21 in both of adenoma as well as adenocarcinoma. The approach in this study proposed combining the big data from the scientific literature together with novel bioinformatics to bring about a methodology that can be used to first identify which microRNAs are involved in a specific disease, and then to identify a panel of biomarkers derived from the microRNAs target genes, and from these target genes the functional significance of these microRNAs can be inferred.

Ethical approval

This is a systematic meta-analysis and bioinformatics and our institution does not need ethical approval for reviews.

Sources of funding

R.H. is funded by Al-Jalila Foundation (grant no: AJF201741), University of Sharjah grant (grant no: 1901090258) and Boehringer Ingelheim (grant no: 120102). R.B. is funded by Al-Jalila Foundation (grant no: AJF2018112).

Authors' contributions

All authors have contributed to the study and to the preparation of the manuscript. Study design and data collection: N.S.S, F.S.S and R.H.; Analysis and wrote the paper: F.S.S. and N.S.S.; Analysis and revised the paper, S.Y.G, R.B., R.H.

Unique Identifying number (UIN)

Name of the registry: researchregistery.com.
Unique Identifying number or registration ID: reviewregistry738.
Hyperlink to the registration (must be publicly accessible):
https://www.researchregistry.com/browse-the-registry#registryofsystematicreviewsmeta-analyses/registryofsystematicreviewsmeta-analysesdetails/5d651e82cc6e900011f809e1/

Guarantor

Rifat Hamoudi.

Provenance and peer review

Not commissioned, externally peer-reviewed.

Data statement

This is a summary and bioinformatics study. Data used for meta-analysis and gene expression analysis was extracted from previously published papers and publicly available databases.

Declaration of competing interests

None declared.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ijsu.2019.11.017.

References

[1] H. Marusawa, B.J. Jenkins, Inflammation and gastrointestinal cancer: an overview, Cancer Lett. 345 (2) (2014) 153–156.
[2] R.L. Siegel, K.D. Miller, A. Jemal, Cancer statistics, 2017, CA A Cancer J. Clin. 67 (1) (2017) 7–30 Epub 2017/01/06.
[3] E.J. Kuipers, W.M. Grady, D. Lieberman, T. Seufferlein, J.J. Sung, P.G. Boelens, et al., Colorectal cancer, Nat. Rev. Dis. Prim. 1 (2015) 15065.
[4] J.D. Vogel, C. Eskicioglu, M.R. Weiser, D.L. Feingold, S.R. Steele, The American society of colon and rectal surgeons clinical practice guidelines for the treatment of colon cancer, Dis. Colon Rectum 60 (10) (2017) 999–1017 Epub 2017/09/12.
[5] S. Jones, W.D. Chen, G. Parmigiani, F. Diehl, N. Beerenwinkel, T. Antal, et al., Comparative lesion sequencing provides insights into tumor evolution, Proc. Natl. Acad. Sci. U. S. A 105 (11) (2008) 4283–4288 Epub 2008/03/14.
[6] M. Jeffery, B.E. Hickey, P.N. Hider, Follow-up strategies for patients treated for non-metastatic colorectal cancer, Cochrane Database Syst. Rev. (1) (2007) Cd002200Epub 2007/01/27.
[7] D.J. Sargent, H.S. Wieand, D.G. Haller, R. Gray, J.K. Benedetti, M. Buyse, et al., Disease-free survival versus overall survival as a primary end point for adjuvant colon cancer studies: individual patient data from 20,898 patients on 18 randomized trials, official journal of the American Society of Clinical Oncology, J. Clin. Oncol. 23 (34) (2005) 8664–8670 Epub 2005/11/02.
[8] Bartel DP. MicroRNAs. Cell.116(2):281-297.
[9] J.R. Kapp, T. Diss, J. Spicer, M. Gandy, I. Schrijver, L.J. Jennings, et al., Variation in pre-PCR processing of FFPE samples leads to discrepancies in < em > BRAF < / em > and < em > EGFR < /em > mutation detection: a diagnostic RING trial, J. Clin. Pathol. (2014).
[10] R. Nedaeinia, M. Manian, M.H. Jazayeri, M. Ranjbar, R. Salehi, M. Sharifi, et al., Circulating exosomes and exosomal microRNAs as biomarkers in gastrointestinal cancer, Cancer Gene Ther. 24 (2) (2017) 48–56 Epub 2016/12/17.
[11] I.A. Asangani, S.A.K. Rasheed, D.A. Nikolova, J.H. Leupold, N.H. Colburn, S. Post, et al., MicroRNA-21 (miR-21) post-transcriptionally downregulates tumor suppressor Pdcd4 and stimulates invasion, intravasation and metastasis in colorectal cancer, Oncogene 27 (2007) 2128.
[12] B. Qian, D. Katsaros, L. Lu, M. Preti, A. Durando, R. Arisio, et al., High miR-21 expression in breast cancer associated with poor disease-free survival in early stage disease and high TGF-β1, Breast Canc. Res. Treat. 117 (1) (2009) 131–140.
[13] A. Horiuchi, H. Iinuma, T. Akahane, R. Shimada, T. Watanabe, Prognostic significance of PDCD4 expression and association with microRNA-21 in each Dukes' stage of colorectal cancer patients, Oncol. Rep. 27 (5) (2012) 1384–1392 Epub 2012/01/24.
[14] D. Moher, A. Liberati, J. Tetzlaff, D.G. Altman, P.G. The, Preferred reporting Items for systematic reviews and meta-analyses: the PRISMA statement, PLoS Med. 6 (7) (2009) e1000097.
[15] A.J. Schetter, S. Leung, J.J. Sohn, et al., Microrna expression profiles associated with prognosis and therapeutic outcome in colon adenocarcinoma, J. Am. Med. Assoc. 299 (4) (2008) 425–436.
[16] B.S. Nielsen, S. Jørgensen, Fog JU, R. Søkilde, I.J. Christensen, U. Hansen, et al., High levels of microRNA-21 in the stroma of colorectal cancers predict short disease-free survival in stage II colon cancer patients, Clin. Exp. Metastasis 28 (1) (2011) 27–38.
[17] H. Shibuya, H. Iinuma, R. Shimada, A. Horiuchi, T. Watanabe, Clinicopathological and prognostic value of MicroRNA-21 and MicroRNA-155 in colorectal cancer, Oncology 79 (3–4) (2010) 313–320.
[18] S. Kjaer-Frifeldt, T.F. Hansen, B.S. Nielsen, S. Joergensen, J. Lindebjerg, F.B. Soerensen, et al., The prognostic importance of miR-21 in stage II colon cancer: a population-based study, Br. J. Canc. 107 (2012) 1169.
[19] Y. Toiyama, M. Takahashi, K. Hur, T. Nagasaka, K. Tanaka, Y. Inoue, et al., Serum miR-21 as a diagnostic and prognostic biomarker in colorectal cancer, J. Natl. Cancer Inst. 105 (12) (2013) 849–859 Epub 2013/05/25.
[20] L.C. Bovell, C. Shanmugam, B.-D.K. Putcha, V.R. Katkoori, B. Zhang, S. Bae, et al., The prognostic value of MicroRNAs varies with patient race/ethnicity and stage of colorectal cancer, Clin. Cancer Res. 19 (14) (2013) 3955.
[21] T.H. Chen, S.W. Chang, C.C. Huang, K.L. Wang, K.T. Yeh, C.N. Liu, et al., The prognostic significance of APC gene mutation and miR-21 expression in advanced-stage colorectal cancer, Colorectal Dis. 15 (11) (2013) 1367–1374.
[22] P. Menéndez, D. Padilla, P. Villarejo, T. Palomino, P. Nieto, J.M. Menéndez, et al., Prognostic implications of serum microRNA-21 in colorectal cancer, J. Surg. Oncol. 108 (6) (2013) 369–373.
[23] N. Oue, K. Anami, J. Schetter Aaron, M. Moehler, H. Okayama, A. Khan Mohammed, et al., High miR-21 expression from FFPE tissues is associated with poor survival and response to adjuvant chemotherapy in colon cancer, Int. J. Cancer 134 (8) (2013) 1926–1934.
[24] W.K. Kang, J.K. Lee, S.T. Oh, S.H. Lee, C.K. Jung, Stromal expression of miR-21 in T3-4a colorectal cancer is an independent predictor of early tumor relapse, BMC Gastroenterol. 15 (1) (2015) 2.
[25] Y. Fukushima, H. Iinuma, M. Tsukamoto, K. Matsuda, Y. Hashiguchi, Clinical significance of microRNA-21 as a biomarker in each Dukes' stage of colorectal cancer, Oncol. Rep. 33 (2) (2015) 573–582 Epub 2014/11/26.
[26] J.L. Neyeloff, S.C. Fuchs, L.B. Moreira, Meta-analyses and Forest plots using a microsoft excel spreadsheet: step-by-step guide focusing on descriptive data analysis, BMC Res. Notes 5 (2012) 52 Epub 2012/01/24.
[27] J. Higgins, S.G. Thompson, J.J. Deeks, D.G. Altman, Measuring inconsistency in meta-analyses, Br. Med. J. 327 (7414) (2003) 557–560.
[28] J.P. Higgins, S. Green, Cochrane Handbook for Systematic Reviews of Interventions,

John Wiley & Sons, 2011.
[29] Y. Wang, X. Gao, F. Wei, X. Zhang, J. Yu, H. Zhao, et al., Diagnostic and prognostic value of circulating miR-21 for cancer: a systematic review and meta-analysis, Gene 533 (1) (2014) 389–397.
[30] R.A. Hamoudi, A. Appert, H. Ye, A. Ruskone-Fourmestraux, B. Streubel, A. Chott, et al., Differential expression of NF-κB target genes in MALT lymphoma with and without chromosome translocation: insights into molecular mechanism, Leukemia 24 (2010) 1487.
[31] V. Saxena, D. Orgill, I. Kohane, Absolute enrichment: gene set enrichment analysis for homeostatic systems, Nucleic Acids Res. 34 (22) (2006) e151–e.
[32] J. Vandesompele, K. De Preter, F. Pattyn, B. Poppe, N. Van Roy, A. De Paepe, et al., Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes, Genome Biol. 3 (7) (2002) research0034.1-research.11.
[33] H. Schwarzenbach, A.M. da Silva, G. Calin, K. Pantel, Data normalization strategies for MicroRNA quantification, Clin. Chem. 61 (11) (2015) 1333.
[34] M.S. Peter, T. Zsofia, M. Viktor, F. Andras, R. Karoly, I. Peter, MicroRNA target prediction: problems and possible solutions, Curr. Bioinform. 5 (1) (2010) 81–88.
[35] B. Xiong, Y. Cheng, L. Ma, C. Zhang, MiR-21 regulates biological behavior through the PTEN/PI-3 K/Akt signaling pathway in human colorectal cancer cells, Int. J. Oncol. 42 (1) (2013) 219–228 Epub 2012/11/24.
[36] K.H. Chang, N. Miller, E.A.H. Kheirelseid, H. Ingoldsby, E. Hennessy, C.E. Curran, et al., MicroRNA-21 and PDCD4 expression in colorectal cancer, Eur. J. Surg. Oncol. 37 (7) (2011) 597–603.
[37] V. Andersen, U. Vogel, S. Godiksen, F.B. Frenzel, M. Sæbø, J. Hamfjord, et al., Low ABCB1 gene expression is an early event in colorectal carcinogenesis, PLoS One 8 (8) (2013) e72119-e.
[38] J. Yu, W.D. Shannon, M.A. Watson, H.L. McLeod, Gene expression profiling of the irinotecan pathway in colorectal cancer, an official journal of the American Association for Cancer Research, Clin. Cancer Res. 11 (5) (2005) 2053–2062 Epub 2005/03/10.
[39] M.G. Backlund, J.R. Mann, V.R. Holla, F.G. Buchanan, H.-H. Tai, E.S. Musiek, et al., 15-Hydroxyprostaglandin Dehydrogenase Is Down-Regulated in Colorectal Cancer, 280 2005, pp. 3217–3223 5.
[40] S.-J. Myung, R.M. Rerko, M. Yan, P. Platzer, K. Guda, A. Dotson, et al., 15-Hydroxyprostaglandin dehydrogenase is an < em > in vivo < /em > suppressor of colon tumorigenesis, Proceedings of the National Academy of Sciences, 103 2006, p. 12098 32.
[41] M. Ilyas, X.P. Hao, K. Wilkinson, I.P.M. Tomlinson, A.M. Abbasi, A. Forbes, et al., Loss of Bcl-2 expression correlates with tumour recurrence in colorectal cancer, Gut 43 (3) (1998) 383.
[42] L. Poincloux, X. Durando, J.F. Seitz, E. Thivat, V.J. Bardou, M.H. Giovannini, et al., Loss of Bcl-2 expression in colon cancer: a prognostic factor for recurrence in stage II colon cancer, Surg. Oncol. 18 (4) (2009) 357–365 Epub 2008/11/26.
[43] Z. Diamantopoulou, G. White, M.Z.H. Fadlullah, M. Dreger, K. Pickering, J. Maltas, et al., TIAM1 antagonizes TAZ/YAP both in the destruction complex in the cytoplasm and in the nucleus to inhibit invasion of intestinal epithelial cells, Cancer Cell 31 (5) (2017) 621–634 e6.
[44] I. Niedzielska, Z. Niedzielski, M. Tkacz, T. Orawczyk, K. Ziaja, J. Starzewski, et al., Toll-like receptors and the tendency of normal mucous membrane to transform to polyp or colorectal cancer, an official journal of the Polish Physiological Society, J. Physiol. Pharmacol. 60 (Suppl 1) (2009) 65–71 Epub 2009/07/23.
[45] L. Xiang, S. Wang, X. Jin, W. Duan, X. Ding, C. Zheng, Expression of BMP2, TLR3, TLR4 and COX2 in colorectal polyps, adenoma and adenocarcinoma, Mol. Med. Rep. 6 (5) (2012) 973–976 Epub 2012/08/28.
[46] G. Mudduluru, F. Medved, R. Grobholz, C. Jost, A. Gruber, J.H. Leupold, et al., Loss of programmed cell death 4 expression marks adenoma-carcinoma transition, correlates inversely with phosphorylated protein kinase B, and is an independent prognostic factor in resected colorectal cancer, Cancer 110 (8) (2007) 1697–1707 Epub 2007/09/13.
[47] T. Kaltenbach, S. Friedland, R. Soetikno, A randomised tandem colonoscopy trial of narrow band imaging versus white light examination to compare neoplasia miss rates, Gut 57 (10) (2008) 1406–1412 Epub 2008/06/05.
[48] C.J. Kahi, J.C. Anderson, I. Waxman, W.R. Kessler, T.F. Imperiale, X. Li, et al., High-definition chromocolonoscopy vs. high-definition white light colonoscopy for average-risk colorectal cancer screening, Am. J. Gastroenterol. 105 (6) (2010) 1301–1307 Epub 2010/02/25.
[49] N. Yamamichi, R. Shimomura, Inada K-i, K. Sakurai, T. Haraguchi, Y. Ozaki, et al., Locked nucleic acid < em > In situ < /em > hybridization analysis of miR-21 expression during colorectal cancer development, Clin. Cancer Res. 15 (12) (2009) 4009.
[50] X.-Y. Qiu, L.-Y. Zhu, C.-S. Zhu, J.-X. Ma, T. Hou, X.-M. Wu, et al., Highly effective and low-cost MicroRNA detection with CRISPR-cas9, ACS Synth. Biol. 7 (3) (2018) 807–813.
[51] A. Rogers, N. Sparkes, D. Zhou, V. Brenner, G. Speight, A cost-effective high-throughput MicroRNA profiling service based on array and non-enzymatic labeling technologies, Biotechniques 45 (5) (2008) 588–589.