

## Article

# Molecular Classification of Colorectal Cancer by microRNA Profiling: Correlation with the Consensus Molecular Subtypes (CMS) and Validation of miR-30b Targets

Mateo Paz-Cabezas <sup>1,†</sup>, Tania Calvo-López <sup>1,†</sup>, Alejandro Romera-Lopez <sup>1</sup>, Daniel Tabas-Madrid <sup>2</sup>, Jesus Ogando <sup>2</sup>, María-Jesús Fernández-Aceñero <sup>3</sup> , Javier Sastre <sup>1</sup>, Alberto Pascual-Montano <sup>2</sup>, Santos Mañes <sup>2</sup> , Eduardo Díaz-Rubio <sup>1</sup>  and Beatriz Perez-Villamil <sup>1,\*</sup> 

<sup>1</sup> Genomics and Microarrays Laboratory, Medical Oncology Department, Instituto de Investigación Sanitaria San Carlos (IdiSSC), Hospital Clinico San Carlos, 28040 Madrid, Spain

<sup>2</sup> Immunology and Oncology Department, Centro Nacional de Biotecnología (CSIC), 28049 Madrid, Spain

<sup>3</sup> Surgical Pathology, Instituto de Investigación Sanitaria San Carlos (IdiSSC), Hospital Clinico San Carlos, 28040 Madrid, Spain

\* Correspondence: beatriz.perezvillamil@salud.madrid.org; Tel.: +34-91-330-3348

† These authors contributed equally to this work.

**Simple Summary:** Colorectal cancer is one of the most significant causes of cancer mortality worldwide. Patients stratification is central to improve clinical practice and the Consensus Molecular Subtypes (CMS) have been validated as a useful tool to predict both prognosis and treatment response. This is the first study describing that microRNA profiling can define colorectal cancer CMS subtypes as well as mRNA profiling. MicroRNAs small size facilitates its analysis in serum facilitating a real-time analysis of the disease course. Three microRNA subtypes are identified: miR-LS is associated with the low-stroma/CMS2-subtype; miR-MI with the mucinous-MSI/CMS1-subtype and miR-HS with the high-stroma/CMS4-subtype. MicroRNA novel subtypes and association to the CMS classification were externally validated using TCGA data. Analyzing both mRNAs and miRs in the same population enabled identification of miR target genes and altered biological pathways. A miR-mRNA interaction screening and regulatory network selected major miR targets and was functionally validated for the miR30b/SCL6A4 pair.

**Abstract:** Colorectal cancer consensus molecular subtypes (CMSs) are widely accepted and constitutes the basis for patient stratification to improve clinical practice. We aimed to find whether miRNAs could reproduce molecular subtypes, and to identify miRNA targets associated to the High-stroma/CMS4 subtype. The expression of 939 miRNAs was analyzed in tumors classified in CMS. TALASSO was used to find gene-miRNA interactions. A miR-mRNA regulatory network was constructed using Cytoscape. Candidate gene-miR interactions were validated in 293T cells. Hierarchical-Clustering identified three miRNA tumor subtypes (miR-LS; miR-MI; and miR-HS) which were significantly associated ( $p < 0.001$ ) to the reported mRNA subtypes. miR-LS correlated with the low-stroma/CMS2; miR-MI with the mucinous-MSI/CMS1 and miR-HS with high-stroma/CMS4. MicroRNA tumor subtypes and association to CMSs were validated with TCGA datasets. TALASSO identified 1462 interactions ( $p < 0.05$ ) out of 21,615 found between 176 miRs and 788 genes. Based on the regulatory network, 88 miR-mRNA interactions were selected as candidates. This network was functionally validated for the pair miR-30b/SLC6A6. We found that miR-30b overexpression silenced 3'-UTR-SLC6A6-driven luciferase expression in 293T-cells; mutation of the target sequence in the 3'-UTR-SLC6A6 prevented the miR-30b inhibitory effect. In conclusion CRC subtype classification using a miR-signature might facilitate a real-time analysis of the disease course and treatment response.

**Keywords:** colorectal cancer; microRNAs; microarray gene-expression profiling; molecular classification; prognostic factors



**Citation:** Paz-Cabezas, M.; Calvo-López, T.; Romera-Lopez, A.; Tabas-Madrid, D.; Ogando, J.; Fernández-Aceñero, M.-J.; Sastre, J.; Pascual-Montano, A.; Mañes, S.; Díaz-Rubio, E.; et al. Molecular Classification of Colorectal Cancer by microRNA Profiling: Correlation with the Consensus Molecular Subtypes (CMS) and Validation of miR-30b Targets. *Cancers* **2022**, *14*, 5175. <https://doi.org/10.3390/cancers14215175>

Academic Editor: Luis Franco

Received: 27 September 2022

Accepted: 19 October 2022

Published: 22 October 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Colorectal cancer (CRC) represents a major health problem being the third most frequent cancer and the second cause of cancer death worldwide [1]. CRC is traditionally classified according to clinical and morphological characteristics in TNM stages (American Joint Committee on Cancer). However, the phenotypic diversity of this disease and its clinical behavior are insufficiently explained by the simple histological grade classification and clinical factors in current use. Our group identified four tumor subtypes by transcriptional profiling [2] that largely overlaps in both, subtype distribution and clinic-biological interpretation with the four Consensus Molecular Subtypes (CMS) [3]. Recently relevant reports have confirmed the prognostic and predictive value of CMS subtypes in phase III clinical trials [4–6] supporting the use of the CMS classification as a useful tool for patient management. MicroRNAs (miRs) are noncoding small RNAs that regulate gene activity post-transcriptionally. In cancer, they can function as oncogenes or as tumor suppressors, and miR signatures can serve as promising biomarkers [7,8]. Previous attempts to associate miRs and CRC subtypes have identified members of the miR-200 family downregulated in the mesenchymal/CMS4 subtype [9,10]. However, no other associations between specific miRs and the other three tumor subtypes have been described. In this context, using unsupervised hierarchical clustering analysis, we have analyzed miR expression patterns in the CRC samples used in our previous molecular subtyping study [2] to investigate if miRs allowed CRC tumors classification as well as mRNAs. Since one miR can regulate multiple mRNAs, analyzing both mRNAs and miRs in the same population is an excellent strategy to determine miR target genes and identify altered biological pathways and regulatory networks. In this study we report the identification of three miR molecular subtypes that associate to the described CMSs. This can be an important advance, since it would allow the search of the relevant miRs in serum/plasma of patients and their classification, as other authors have reported for pancreatic adenocarcinoma [11], without the need to obtain biopsies or fragments of the tumor, facilitating real-time analysis of the course of the disease and of the response to the treatment. A just released report, develops a miR classifier using supervised analysis to predict four miR subtypes assigned from the four mRNA CMS subtypes [12]. Using *in silico* machine learning the study of Adam et al. [12] converts the four mRNA-CMS subtypes to four miR-subtypes. This procedure is different than ours. We used unsupervised analysis that does not constrain any subtype number or class.

## 2. Materials and Methods

### 2.1. Patients and RNA

For this study we have analyzed the same CRC patients' cohort used for our previous study, including RNA samples [2]. Tumor samples were taken from the Biobank of the Hospital Clinico San Carlos. The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Institutional Review Board and Ethics Committee of Hospital Clinico San Carlos. RNA was extracted from fresh frozen tumor samples using TRIZOL and the homogenizer Ultraturrax T8-S8N-5G. RNA quality was measured with Agilent Bioanalyzer 2100. Only tumors with an RNA Integrity Number (RIN)  $\geq 6.5$  were included in the analysis.

### 2.2. MicroRNA Expression Analysis and Tumor Classification

Agilent miR 21827 microarrays were used to analyze the expression of 939 miRs in 97 CRC tumor samples and 19 normal colon samples. Fluorescence was measured and quantile-normalized using Agilent scanner, Feature Extraction and GeneSpring software. 176 miRs were present in 90% of the samples and therefore considered for the following data analysis. Expression data was median centered and Average-linkage-hierarchical clustering (centered Pearson correlation) was carried out to perform unsupervised tumor classification considering the 176 expressed miRs in the 88 tumor samples from our previous study [2] (complete data set was submitted to ArrayExpress (E-MTAB-9288)). Then, Differential expression between miR subtypes was analyzed using one-way ANOVA, Student

Newman-Keuls (SNK) post hoc test and Benjamini-Hochberg multiple test correction. miRs were considered as differentially expressed only if global  $p < 0.05$  and fold change  $> 1.5$  considering any of the pairwise subtype comparison.

### 2.3. Identification of miRs Targets and Correlation with mRNA Expression

TALASSO software [13] was used to find miR-mRNA interactions between the 1722 genes selected from our previous study [2] and the 176 expressed miRs. In order to predict miR-target interactions, TALASSO analyzes miRs expression changes and down-regulation of their putative targets. As criteria to select the most relevant miR-transcript interactions, a class comparison analysis was carried out to find differentially expressed genes between groups.

Unpaired Student-*t*-test with Benjamini-Hochberg multiple correction was carried out between normal colon tissue and tumors from the low-stroma, high-stroma and Mucinous-MSI subtypes. Selected genes were considered as differentially expressed at  $p < 0.05$  and  $> 1.5$ -fold expression. Then, miRNA-mRNA predicted interactions were used to construct a regulatory network using Cytoscape software v3.6.1 [14]. Only the largest connected component was considered for each network. Centrality measures were determined using NetworkAnalyzer and CentiScaPe 2.2. Clusters with higher interconnections were unveiled using ClusterViz and EAGLE algorithm with default options (CliqueSize Threshold = 3, ComplexSize Threshold = 2). Two global centrality measurements, radiality and closeness centrality, were considered to rank the most relevant nodes, as they reflect not only the immediate connections of a node (the degree of each node) but the overall structure of the network. Combining two centrality measurers increase the reliability of this kind of approaches to predict the most relevant genes in an interaction network [15]. In our data, those two topological parameters predicted the same upmost central genes, considering that miRNA-mRNA interactions between the 20 upmost central nodes for each subtype were selected as putative candidates, along with the interactions between mRNA and miRNAs involved in the most relevant cluster for each subtype.

Potential microRNA-mRNA interaction candidates were annotated and scored using information from [16] with two different combined validated predicting scores (Weighted Scoring by Precision (WSP) and logistic regression score (LRS)). Previously experimentally validated interactions were determined using four different databases Tarbase (<http://www.microrna.gr/tarbase>), miRTarBase (<http://miRTarBase.mbc.nctu.edu.tw/>), miR-Walk (<http://mirwalk.umm.uni-heidelberg.de>) and miRecords (<http://miRecords.umn.edu/miRecords>), and also with significant Pearson correlation  $p$ -values from Starbase (<http://starbase.sysu.edu.cn/>). MicroRNA binding sites were predicted by five different algorithms Pita ([https://tools4mirs.org/software/target\\_prediction/pita/](https://tools4mirs.org/software/target_prediction/pita/)), FindTar (<http://bio.sz.tsinghua.edu.cn/findtar/>), Miranda (<https://www.mirbase.org>), rnaHybrid (<http://bibiserv.techfak.uni-bielefeld.de/rnahybrid/>) and TargetScan (<http://genes.mit.edu/targetscan>). MicroRNA candidate prioritization was assessed using an automated script, considering that the last accession date, for all databases accession dates, are 21 April 2016. Candidate gene-miR interactions were scored and biologically validated in HEK-293T cell line.

### 2.4. External Dataset Validation

TCGA data for miRNA and mRNA expression in CRC were downloaded from the repository using TCGA Biolinks [17] package, RNAseq using Illumina HiSeq platform was selected to obtain 285 samples with 20,531 features each. Normalized gene expression data for mRNA was classified in CMS subtypes using CMSclassifier R package 3 according to the nearest CMS criteria. Sample clustering: TCGA raw data for miRNA consisted of 444 samples and 1046 features. Expression data from RNAseq was processed using DESeq2 [18] to obtain normalized counts matrix. Prior to the unsupervised clustering of samples according to miRNA expression, we performed a 3D-PCA visualization to filter out those samples with an outlier expression pattern, following this criterion 5 samples

from the initial dataset were excluded. Afterwards, gene features with less than 10 counts in more than 90% of samples were filtered out, resulting in 336 features per sample. Hierarchical clustering on samples was performed using *hclust* function over  $\log_2$  transformed normalized expression matrix. Pearson correlation as distance measure and ward linkage as agglomeration method were chosen. In order to create the heatmap visualization, gene features were also classified using the same parameters. Finally, subtype association between miRNA and mRNA classification was addressed using Chi-square test ( $\chi^2$ ).

#### 2.5. MicroRNAs Differentially Expressed between Tumor-Epithelia and Tumor-Stroma

MicroRNAs expression data were downloaded from GSE35602 [19]. Differential expression between the epithelial and stromal components of the tumor was analyzed by T-Test and Benjamini-Hochberg Multiple Correction Test using GeneSpring Dx 14.9 software. Selected miRs were considered as differentially expressed at  $p < 0.05$  and  $>1.5$ -fold expression between tumor epithelia and tumor stroma.

#### 2.6. Evaluation of miRs-Subtypes Using miRaCL20 Classifier

MicroRNAs expression data (miRNA-Seq) from TCGA-COAD were classified using miRaCl classifier [12] available at Github/rsmadam/CMS-miRaCl. Subtype association between miRNA and mRNA classification was determined using Chi-square test ( $\chi^2$ ).

#### 2.7. Cell Lines, Transformation, Transfection and Luciferase Assay

Human HEK-293T cells were grown in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% fetal bovine serum (FBS), penicillin-streptomycin, L-Glutamine and NaPyr in a humidified incubator at 37 °C with 5% of CO<sub>2</sub>. HmiR0133-MR03 (hsa-miR-30b), HmiT070741-MT06 (FAP), HmiT017418a-MT06 (SLC6A6-A), HmiT017418b-MT06 (SLC6A6-B) and miR-Control plasmids from GeneCopoeia were used. XL1-Blue bacteria were transformed by thermal shock and DNA was extracted using Genomed kit (JETSTAR). HEK-293T cells were cultured in triplicate in 24-well plates (0.05 × 10<sup>6</sup> cells/well). They were transfected with miR-30b and miR-Control, using Lipofectamine 2000 (Invitrogen). Cells were selected with puromycin and miR-30b levels was checked by RT-PCR using Hs03303066\_pri (TaqManTMPri-miR Assays) oligonucleotide and U6 as control. HEK-293T-miR30b-expressing cells were transfected with SLC6A6-A, SLC6A6-B or FAP plasmids. Vectors of these plasmids include Firefly and Renilla luciferase reporter genes. After 12 and 24 h Firefly and Renilla luciferases activity were measured using Dual Luciferase Assay Kit (Promega Madison, WI, USA) in a Tecan Infinite 200 Luminometer. Luciferase intensity measurement was performed by triplicate per condition and analyzed as described [20].

#### 2.8. Site-Directed Mutagenesis

Predicted miR-30b interaction site at the SLC6A6 3'-UTR (2225-TGTTTAC-2231 nucleotides) was modified using QuikChange site-directed mutagenesis kit (Agilent Technologies, Palo Alto, CA, U.S.A). The oligonucleotide 5'-cctatgagaatctaattgtattacaaagcaggaaa gccgccggcc-3' (2207 to 2251 nucleotides) was designed using QuikChange Primer Design Tool. G2226T, T2228G and C2231A nucleotides were changed to destabilize the predicted interaction with miR-30b.

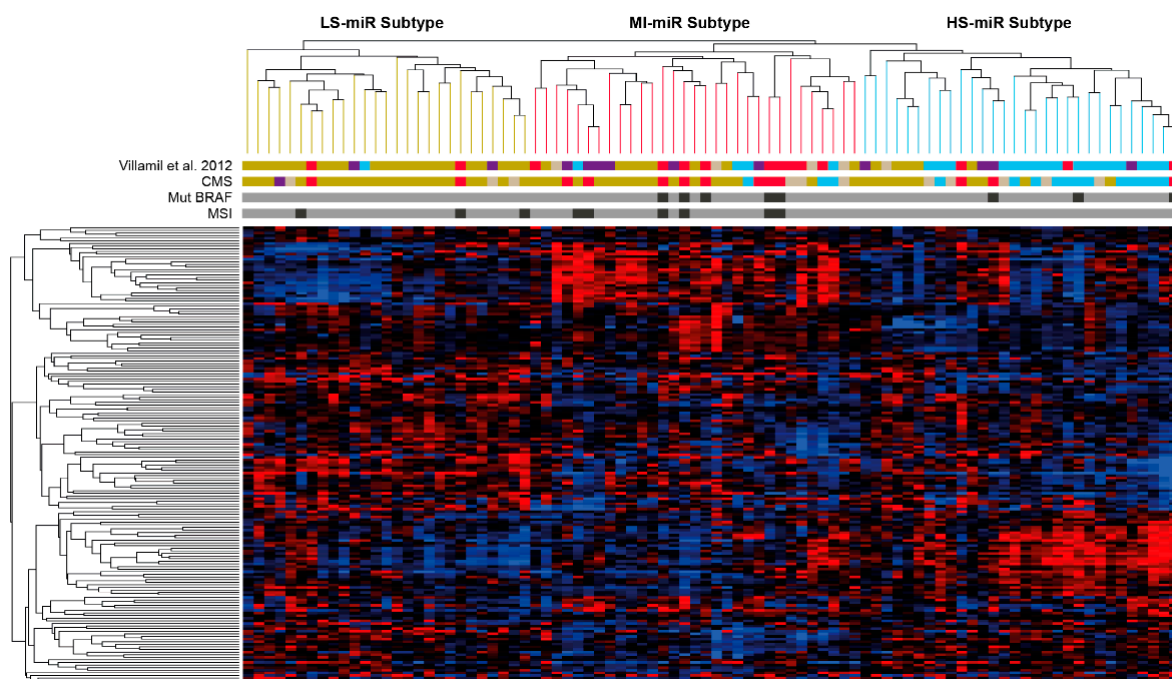
#### 2.9. Statistical Analysis

Luciferase analysis results were analyzed using Student's *t*-test to compare mutated vs control mir-30b. Subtype association was addressed using  $\chi^2$  Chi-square test. In order to compare the distribution of qualitative variables between groups Fisher exact test was applied (as all the variables presented less than 5 events in at least one of the categories) and "Mantel-Haenzel Test" for b-catenin linear categories. Mean comparison of quantitative variables between subtypes was performed using Kruskal-Wallis test. Statistical analysis was performed using GraphPad Prism 6 and R software.

### 3. Results

#### 3.1. Tumor Classification Based on miR Expression Patterns and Association to mRNA Subtypes

MicroRNAs arranged tumor samples in three clusters (Figure 1). There is a significant association ( $p < 0.001$ ) of the three miR subtypes with the four mRNA subtypes identified by us [2] as well as with the CMS subtypes [3] (Table 1). Supplementary Table S1 shows the classification of the 88 tumors from our previous study [2] using the SSP and RF [3]. miR-Cluster-1 contains 27 tumors showing a higher proportion of tumors belonging to the low-stroma-subtype, as well as the lowest proportion of stromal component in the tumors; consequently, we named this subtype miR-LS (miR-Low-Stroma). Additionally, miR-LS show a significant association with CMS2 whether random forest (RF) or single sample predictor (SSP) were used for sample classification. The highest proportion of tumors from the mucinous-MSI-subtype as well as from CMS1 are in miR-cluster-2 which contains 31 tumors; mucinous histology as well as microsatellite instability (MSI) are associated to this cluster, accordingly we term this cluster miR-MI (Mucinous, Instable). Cluster-3 with 30 tumors contains the highest proportion of tumors of the high-stroma-subtype as well as the highest proportion of stroma in the tumors; we term this cluster miR-HS (High Stroma). Like-wise, the highest proportion of tumors classified as CMS4 associate to miR-HS subtype.



**Figure 1.** Molecular classification of tumors and miRs. Centered Pearson correlation and average-linkage-hierarchical clustering of the 88 tumor samples and 176 miRs in three miR tumor subtypes (miR-LS pistachio-green lines; miR-MI red lines and miR-HS light blue lines). Villamil et al. subtypes, CMSs, BRAF mutations and MSI are specified below the tree. Low-stroma-subtype/CMS2: pistachio green bar; mucinous-MSI-subtype/CMS1: red bar; high-stroma-subtype/CMS4: light blue; immunoglobulin-related: purple bar; unclassified samples: beige bar. Black bar: BRAF mutated and MSI; grey bar: BRAF wt and MSS. Heatmap intensities: 3.099 (red) to  $-3.099$  (dark blue).

**Table 1.** Association of miRNA clusters to Villamil et al. subtypes, to CMS and to clinic-biological parameters.

|                                |              | miR-LS (n = 27) | miR-MI (n = 30) | miR-HS (n = 31) | p Value             |
|--------------------------------|--------------|-----------------|-----------------|-----------------|---------------------|
| Villamil et al., 2012 Subtypes | Low Stroma   | 22              | 7               | 6               | 0.000 $\chi^2$      |
|                                | Immunoglobul | 2               | 6               | 4               |                     |
|                                | High Stroma  | 1               | 4               | 17              |                     |
|                                | Mucinous-MSI | 2               | 9               | 3               |                     |
|                                | Unclassified | 0               | 4               | 1               |                     |
| RF                             | CMS1         | 1               | 7               | 2               | 0.000 $\chi^2$      |
|                                | CMS2         | 15              | 9               | 10              |                     |
|                                | CMS3         | 4               | 0               | 0               |                     |
|                                | CMS4         | 0               | 3               | 13              |                     |
| SSP                            | NA           | 7               | 11              | 6               | 0.001 $\chi^2$      |
|                                | CMS1         | 2               | 8               | 3               |                     |
|                                | CMS2         | 21              | 16              | 11              |                     |
|                                | CMS3         | 1               | 0               | 0               |                     |
| Microsatellite                 | CMS4         | 0               | 3               | 12              | 0.036 $\chi^2$      |
|                                | NA           | 3               | 3               | 5               |                     |
|                                | MSS          | 24              | 24              | 31              |                     |
| Histologic type                | MSI          | 3               | 6               | 0               | 0.144 $\chi^2$      |
|                                | Conventional | 26              | 24              | 28              |                     |
| BRAF                           | Mucinous     | 1               | 6               | 3               | 0.091 $\chi^2$      |
|                                | WT           | 27              | 25              | 28              |                     |
| FF Stroma                      | Mut          | 0               | 5               | 3               | 0.000 <sup>KW</sup> |
|                                | Range        | (5–28)          | (5–40)          | (8–65)          |                     |
| FFPE Stroma                    | Median       | 7.5             | 13.75           | 22.5            | 0.004 <sup>KW</sup> |
|                                | Range        | (5–20)          | (5–40)          | (5–60)          |                     |
|                                | Median       | 10              | 10              | 20              |                     |

RF: Random Forest, SSP: single sample predictor. FF: Fresh-Frozen, FFPE: Formalin Fixed Paraffin embedded. <sup>KW</sup>: Kruskal Wallis,  $\chi^2$ : Chi-Squared test.

### 3.2. External Dataset Validation

CRC data from TCGA were classified according to the CMS subtypes, resulting in the following subtype distribution for the 285 samples: CMS1 (59), CMS2 (144), CMS3 (33) and CMS4 (49). Hierarchical clustering of miRNA expression (Supplementary Figure S1) unveiled three different groups according to miR expression with the following correspondence with CMS subtypes determined by mRNA expression (Supplementary Table S2), this association presented a significant correlation ( $p < 0.0001$ ) and was performed in those 228 samples with mRNA and miRNA data.

### 3.3. Comparison between miR-LS, miR-HS and miR-MI with the miRCL20 Classifier Subtypes

Association between unsupervised miRNA subtypes (miR-LS, miR-HS, miR-MI) and miRaCl20 (CMS subtyping using miRNA data) was addressed in both TCGA data and Agilent CRC miRNA microarray dataset.

CMS distribution in TCGA data according to miRaCl20 supervised classifier resulted in 41 CMS1, 90 CMS2, 22 CMS3, 73 CMS4 and 2 unclassified for the total 228 samples. In the case of the microarray dataset samples were distributed: 23 CMS1, 44 CMS2, 9 CMS3 and 12 CMS4 for the 88 samples.

Association between the three miR subtypes (miR-LS, miR-HS, miR-MI) and miRaCl CMS subtypes (Supplementary Figures S2 and S3) is significant resulting pvalue of Chi-square test ( $\chi^2$ ) was  $< 2 \times 10^{-16}$  in both cases, with a wider consensus in high-stroma- and low-stroma- subtypes (CMS4 and CMS2).

### 3.4. Stromal or Epithelial Localization of the miRs Differentially Expressed between Subtypes

Stroma proportion is associated to miR-subtypes (Table 1) but our study was not designed to distinguish miR expression between the stromal or epithelial components

of the tumor. To find the contribution of stroma or epithelia to miR expression we took advantage of the study of Nishida et al. [19] in which miR expression was specifically analyzed using laser microdissection, in tumor stroma and in tumor epithelia. From the 176 miRs selected for tumor classification, 45 miRs were significantly differentially expressed at  $p < 0.05$  and  $FC > 1.5$  between tumor epithelia and tumor stroma.

The 176 miRs were also arranged in clusters. Among all, three of them showed the most significant miRs differentially expressed between clusters. Interestingly, two of these clusters contained miRs differentially expressed between tumor-stroma and tumor-epithelia as well (Supplementary Table S3).

MicroRNA-Cluster-A contains miRs that are down-regulated mainly in the miR-LS-subtype and up-regulated in the miR-MI-subtype (Figure 2A). It is worth noting that among the miRs of this cluster are viral miRs such as the human cytomegalovirus-encoded miR, hcmv-miR-UL70-3p and the Kaposi's sarcoma-associated herpesvirus miRs: kshv-miR-K12-3 and kshv-miR-K12-10b. Other relevant miRs of this cluster that have been shown to be involved in CRC progression are miR-572 [21], miR-1246 [22], and miR-494 [23]. This group of miRs does not show particularly a specific stromal or epithelial localization (Supplementary Table S3).

MicroRNA-Cluster-B contains miRs that are particularly inhibited in the miR-HS-subtype (Figure 2B). miR-141; miR-200a; miR-200b; miR-200c and miR-429 are in this cluster and belong to the miR-200 family. Other relevant miRs down-regulated in this cluster are miR-378 and miR-194. The miRs of this cluster are down-regulated in the stroma and up-regulated in the epithelia (Supplementary Table S3).

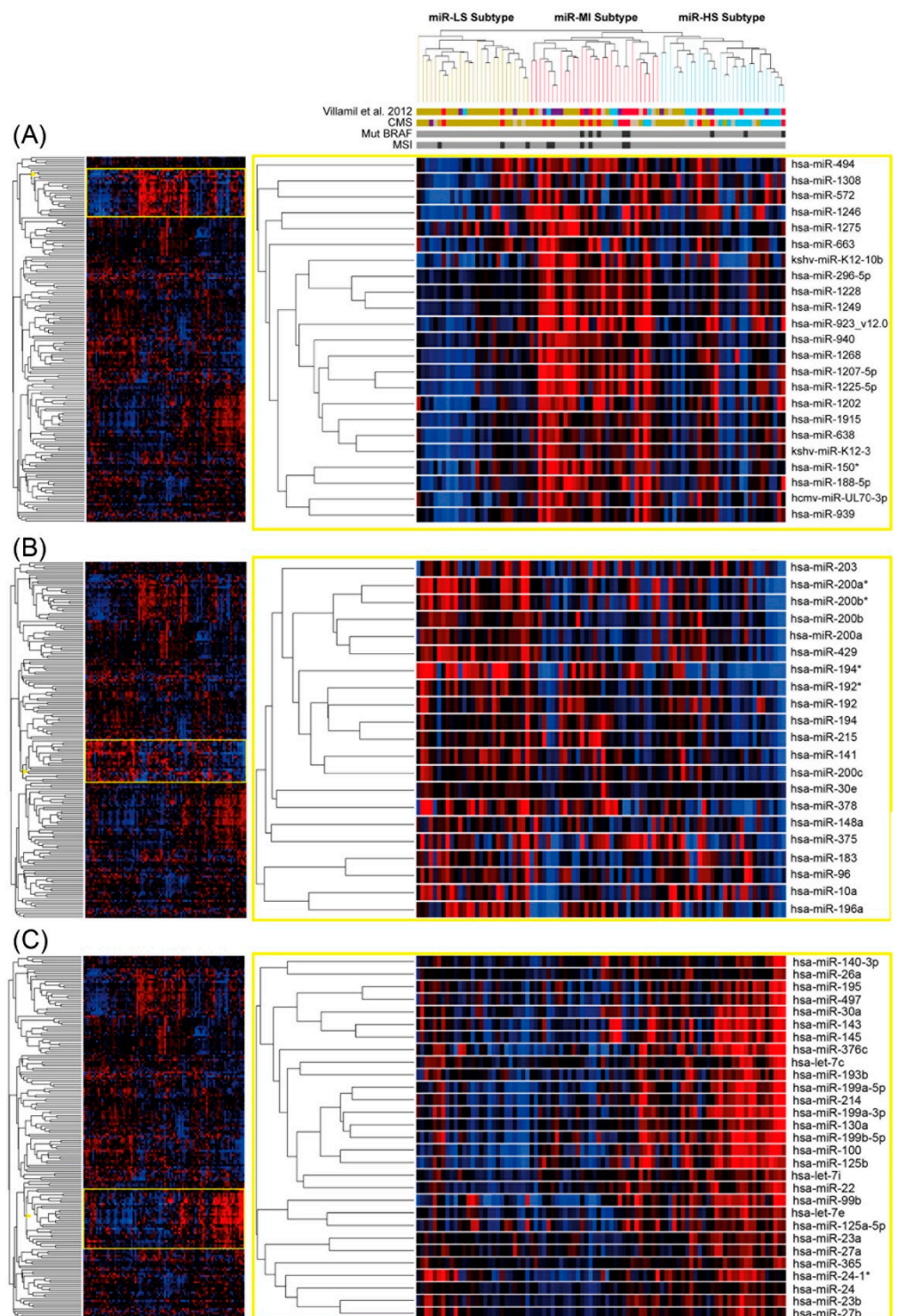
MicroRNA-Cluster-C contains miRs that are upregulated in the miR-HS-subtype (Figure 2C). Members of the miR-30 family and of the miR-100 family such as miR-100, miR-125 and miR-99 are in microRNA-Cluster-C. Other relevant miRs of this cluster are miR-143 and miR-145. These miRs are up-regulated in the stroma and down-regulated in the epithelia (Supplementary Table S3).

### 3.5. Identification of miRs Targets, Selection of Relevant Interactions Associated to Subtypes and Altered Pathways

TALASSO software [13] identified 1462 significant ( $p < 0.05$ ) interactions between 176 miRs and 788 genes out of the 21615 putative interactions (Supplementary Table S4). Out of the 788 genes showing significant miR interactions, 166 genes were differentially expressed in Low-stroma/subtype, 158 in High-stroma/subtype and 78 in Mucinous-MSI/subtype.

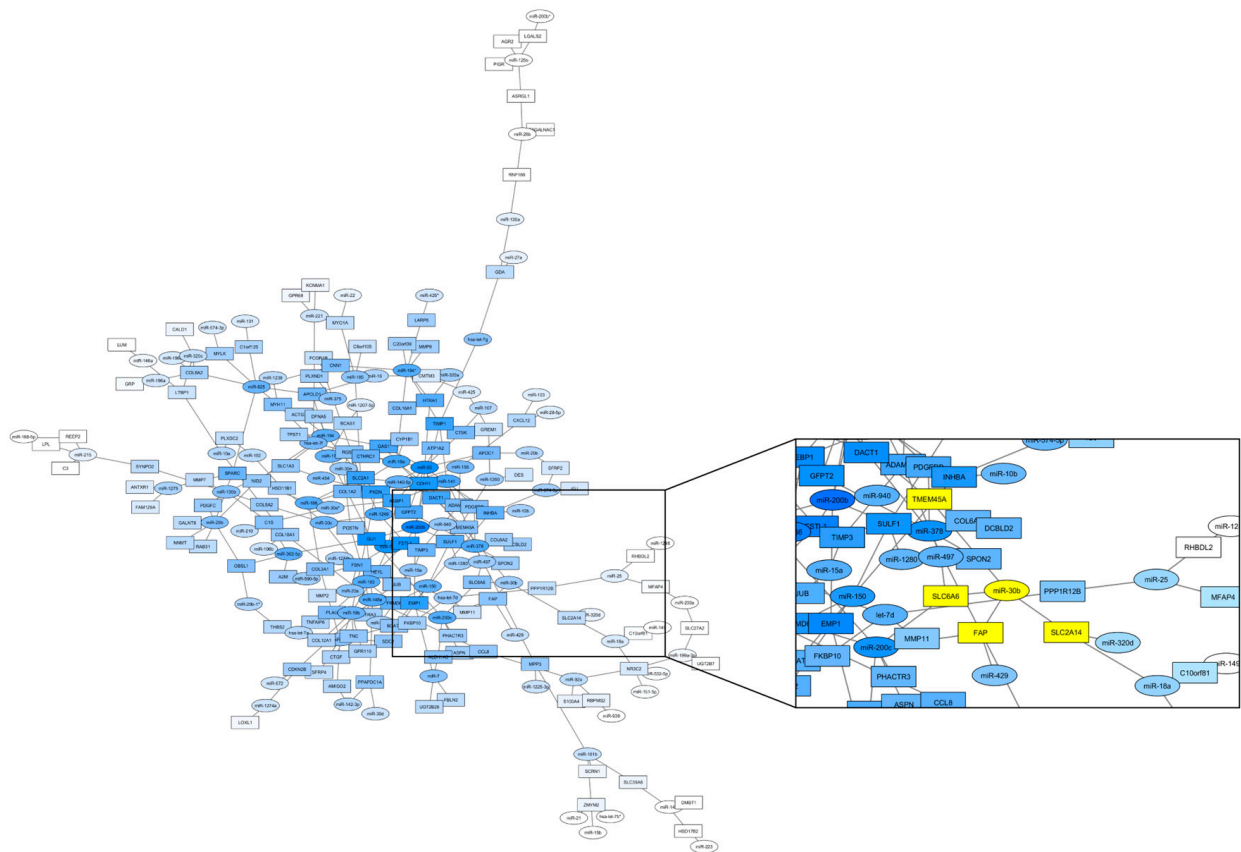
In order to identify relevant targets in miR-mRNA interaction patterns, three subtype specific network graphs were generated using those predicted interactions with differential expression (Supplementary Figure S4). MicroRNAs and mRNAs were represented as nodes, connected according to the in-silico predicted interactions ( $p < 0.05$ ), topological parameters and selecting criteria for the obtained networks are available in Supplementary Table S5.

A list of 88 mRNA-miR interactions was annotated and ranked (Supplementary Table S6). After discarding those interactions that were already biologically validated and taking in consideration the observed expression profiles between subtypes, network centrality values and annotated scores for each interaction, we decided to focus on studying miR-30b-FAP and miR-30b-SLC6A6 interactions as final candidates for biological validation. Moreover, miR-30b and their targets (FAP and SLC6A6) belong to the most connected cluster in miR-HS interaction network (Figure 3), being the subtype featuring the lowest survival.



**Figure 2.** MicroRNAs distribution between Subtypes. (A) miRNA-Cluster-A: miRs down-regulated in the miR-LS-subtype and up-regulated in the miR-MI-subtype (23 miRs located in the heatmap between the 7th and the 29th miRs). (B) miRNA-Cluster-B: miRs inhibited in the miR-HS-subtype (21 miRs located in the heatmap between the 89th and the 109th miRs). (C) miRNA-Cluster-C: miRs up-regulated in the miR-HS-subtype (29 miRs located in the heatmap between the 114th and the 142nd miRs). Heatmap intensities: 3.099 (red) to  $-3.099$  (dark blue).

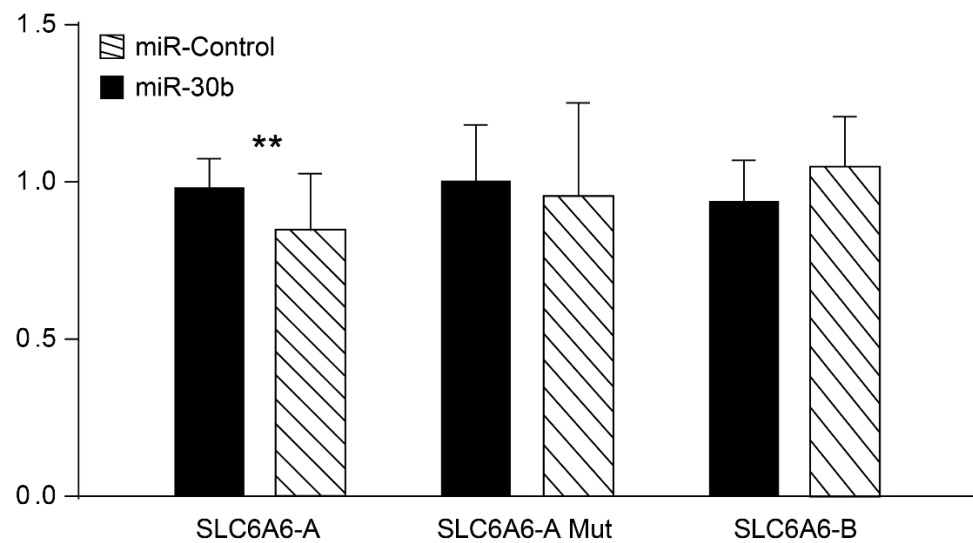




**Figure 3.** Regulatory Network of the High-Stroma/CMS4-miR-HS-subtype. Nodes reflect mRNAs (squares) and miRNAs (circles), while edges represent a predicted interaction between them. Grey intensity is mapped to each node's closeness centrality value, the lighter nodes being the most marginal nodes. Square details the interactions between miR-30b and their first neighbors (those genes with a predicted interaction) represented as striped squares.

### 3.6. SLC6A6 Up-Regulated in the High-Stroma/CMS4 Subtype Shows Specific Interaction with miR-30b In Vitro

The genes SLC6A6 and FAP, that are up-regulated in the High-stroma/CMS4 subtype, show in-silico interaction with miR-30b (Figure 3) which is down-regulated in tumors and in the stroma versus the epithelia component of the tumor (Supplementary Table S3). In order to validate in-silico predicted miR-transcript interactions, HEK-293T cells were transfected with a miR-30b expression plasmid and with reporter plasmids containing 3'UTR regions of the genes SLC6A6 and FAP. Since SLC6A6 3'UTR region is too long, two different reporter plasmids were used SLC6A6-A (between 2174 and 4573) carrying the putative miR-30b binding site (2225-TGTTTAC-2231) and SLC6A6-B (from 4353 to 6528 nucleotide). MicroRNA-30b significantly ( $p = 0.0038$ ) decrease luciferase activity of the SLC6A6-A reporter plasmid. No significant differences in luciferase activity were found when the putative binding site in SLC6A6-A is mutated or when plasmid SLC6A6-B lacking miR-30b predicted binding site is used (Figure 4). When using FAP 3'UTR reporter plasmid, no differences were found between miR-30b and miR-Control (not shown). These results indicate that miR-30b binds to SLC6A6 3'UTR region to decrease SLC6A6 3'UTR-driven reporter expression.



**Figure 4.** miR-30b target SLC6A6 expression. Luciferase reporter assays of HEK-293 cells transduced with pre-miR-30b or a control miR, and transfected with the 3'-UTR of SLC6A6-A wild-type, SLC6A6-A mutated in the target sequence, or SLC6A6-B lacking the predicted miR-30b binding site. Luciferase activity was normalized to that of control-transfected cells. Data shown as mean  $\pm$  SEM of triplicates per experiment ( $n = 3$ ). \*\*  $p < 0.005$ .

#### 4. Discussion

Tumor molecular classification using unsupervised analysis of gene expression is a powerful tool that has been widely applied to distinguish tumor subgroups with shared biological programs and similar clinical behavior [24]. In contrast, tumor subtyping using unsupervised analysis of miR expression has been barely employed. MicroRNAs are shown to regulate gene expression, and both, miR and mRNA expression patterns are altered in cancer [8]. Since miRs and mRNAs coordinately regulate pathways involved in CRC, our hypothesis was that miR profiling, could also classify CRC in molecular subtypes and that these miR-subtypes, would probably correlate with the described mRNA tumor subtypes [2,3]. In this report we describe the identification of three tumor subtypes based on miR expression patterns that correlates significantly with the four tumor subtypes previously discovered [2,3]. Subtype miR-LS is associated with low-stroma-subtype and CMS2, miR-MI is associated with mucinous-MSI-subtype and CMS1 and miR-HS is associated with high-stroma-subtype and CMS4. Consequently, it is feasible to classify colorectal tumors in the described molecular subtypes by miRs expression profiling. As expected, this correlation is also maintained when the CMS are assigned using miRNA data through miRaCl classifier. Despite of being two strategies for miR classification, they differ in their approach; whereas miRaCl classifier is a supervised method to determine the CMS subtype according to the miR data, our classification does not take any kind of previous assumption to segregate samples. Despite those differences, both classifications have a high correlation in high and low stroma subtypes, supporting the continuous flow of evidence of the role of stroma in the course of the colorectal disease.

This could be an important advance, since it would allow in future the search of these miRs in plasma and the classification of patients without the need to obtain tumor fragments, facilitating real-time analysis of the course of the disease and the response to treatment, as has been described for pancreatic adenocarcinoma [11].

MicroRNA-Cluster-A contains miRs that are up-regulated in the miR-MI-subtype and shows high expression of miRs belonging to the herpesvirus family [25,26]. It has been shown that viral miRs are able to modulate innate immune responses. MSI and CMS1 tumors are characterized by a higher level of tumor-infiltrating lymphocytes and activation of immune evasion pathways. MicroRNA-Cluster-B contains miRs down-regulated in the miR-HS-subtype. These miRs have been shown to be down-regulated in the stroma

compared to the epithelia component of the tumor [19] as well. MicroRNA-Cluster-B is mainly composed with members of the miR-200 family. It has been reported that methylation of the miR-200 promoter identifies tumors of the CMS4 [9,10] this agrees with the lower expression of these miRs in the miR-HS-subtype we found in our results. Other relevant miRs down-regulated in this cluster are miR-378 and miR-194. Low levels of miR-378 and miR-194 are implicated in the malignant phenotype of CRC and restoration of their expression inhibits EMT (Epithelial-Mesenchymal Transition) and prevent the migration and invasion of colon cancer cells [27,28].

MicroRNA-Cluster-C contains members of the miR-30 family and of the miR-100 family and are up-regulated in the miR-HS-subtype. Interestingly miRs of this cluster have been reported to be up-regulated in the stroma and down-regulated in the epithelia component of the tumor [19]. Overexpression of miR-100 and miR-125b has been associated with resistance to cetuximab treatment [29]. Other relevant miRs of this cluster are miR-143 and miR-145, these miRs are frequently reduced in colon cancer [30]. We found that when compared with normal colon tissue, tumor miR-143 and miR-145 levels are down regulated around three-fold in clusters miR-LS and miR-MI. However, miR-143 and miR-145 are up-regulated in miR-HS subtype when compared to the other miR clusters but still inhibited with respect to normal colon tissue. An elegant report shows that miR-143 and miR-145 are expressed in the intestinal mesenchyme [31]; this could explain the higher miR-143 and miR-145 in the miR-HS-subtype.

Appropriate integration of miR and mRNA expression profiles is essential to properly understand regulatory pathways and cellular dysfunction in CRC. Elucidating miR targets by bioinformatic analysis permits the identification of a panoply of miR-mRNA possible interactions that need to be ranked. Network analysis is an excellent tool for the selection of the most significant miR-mRNA interactions. Although relevant nodes were found in the three subtypes analyzed (low-stroma, high-stroma and Mucinous-MSI) for in vitro validation we focused on the high-stroma-subtype associated with a poor clinical outcome. The best scores within non-biologically-validated interactions were obtained between miR-30b which is down-regulated in the miR-HS subtype, and FAP or SLC6A6 genes, which are up-regulated in the high-stroma-subtype. High FAP and SLC6A6 levels are associated with worse prognosis in CRC [32,33]. We could not biologically validate miR-30b and FAP interaction; however, miR-30b has been shown to silence SLC6A6 expression. Since higher levels of SLC6A6 are associated with maintenance of stem-cells properties and with chemoresistance [33] restoring miR-30b could be a promising strategy for the treatment of CRC patients of the High-stroma/CMS4 subtype.

## 5. Conclusions

In summary, we show that miR profiles classify colorectal tumors with a straight correlation with the molecular subtypes identified by transcriptional profiling. miR-LS is associated with low-stroma/CMS2, miR-MI with the mucinous-MSI/CMS1 and miR-HS with high-stroma/CMS4 subtypes. Furthermore, the miR/mRNA network identified in High-stroma/CMS4 subtype was validated for the miR30b/SCL6A4 pair. Considering this, using miRs as a classifier provides a promising scenario, the classification of colorectal cancer patients by miR determination in plasma, allowing the classification of patients avoiding invasive procedures and allowing real-time analysis of the course of the disease and response to treatment by liquid biopsy.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/cancers14215175/s1>, Table S1: Association of Villamil et al. 2012 Subtypes and CMS; Table S2: miR vs CMS subtypes in TCGA; Table S3: miRs differentially expressed between clusters; Table S4: TALASSO Interaction between miRs and target genes; Table S5: Topological parameters of regulatory networks; Table S6: Final interactions; Figure S1: Hierarchical Clustering of miR expression from the TCGA dataset; Figure S2: Association between miR subtypes (miR-LS; miR-MI; miR-HS) and miRaCL20A (miRNA CMS); Figure S3: TCGA: Association between miR

subtypes (miR-LS; miR-MI; miR-HS) and miRaclCL20 (CMS); Figure S4: Regulatory networks of miRNA-mRNA interactions in each CRC tumor subtype.

**Author Contributions:** Conceptualization, S.M., E.D.-R. and B.P.-V.; methodology, T.C.-L., M.P.-C., A.R.-L. and J.O.; software, M.P.-C., D.T.-M. and A.P.-M.; validation, M.P.-C., T.C.-L. and M.-J.F.-A.; formal analysis, M.P.-C., T.C.-L. and M.-J.F.-A.; investigation, M.P.-C., T.C.-L., J.S. and B.P.-V.; resources, S.M., E.D.-R., J.S. and B.P.-V.; data curation, M.P.-C., D.T.-M. and A.P.-M.; writing, M.P.-C., T.C.-L. and B.P.-V.; writing-review and editing, T.C.-L., M.P.-C. and B.P.-V.; visualization, M.P.-C., T.C.-L. and B.P.-V.; supervision, S.M. and B.P.-V.; project administration, B.P.-V.; funding acquisition, S.M., E.D.-R. and B.P.-V. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was partially funded by: IMMUNOTHERCAN Comunidad de Madrid S2017/BMD-3733; Fundacion Mutua Madrileña AP151962014; Bayer Healthcare BPV.M01BAY; Fundacion Rodriguez-Pascual EDRFERP.2013; Fundacion 2000 Merck-Serono F01MSER13. PID2020-116303RB-I00/MCIN/AEI/10.13039/501100011033.

**Institutional Review Board Statement:** The study was approved by the Institutional Review Board of the Hospital and the Ethical Committee. The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Institutional Review Board and Ethics Committee of Hospital Clinico San Carlos. CEIC Hospital Clinico San Carlos n°: 17/241-E-BS.

**Informed Consent Statement:** The Bank of Tumors follows the rules established by the hospital including the patient consent approved by the Ethical Committee of the Hospital Clinico San Carlos.

**Data Availability Statement:** ArrayExpress E-MTAB-9288.

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

## References

- Sung, H.; Ferlay, J.; Siegel, R.L.; Laversanne, M.; Soerjomataram, I.; Jemal, A.; Bray, F. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA. Cancer J. Clin.* **2021**, *71*, 209–249. [[CrossRef](#)] [[PubMed](#)]
- Perez-Villamil, B.; Romera-Lopez, A.; Hernandez-Prieto, S.; Lopez-Campos, G.; Calles, A.; Lopez-Asenjo, J.; Sanz-Ortega, J.; Fernandez-Perez, C.; Sastre, J.; Alfonso, R.; et al. Colon cancer molecular subtypes identified by expression profiling and associated to stroma, mucinous type and different clinical behaviour. *BMC Cancer* **2012**, *12*, 260. [[CrossRef](#)]
- Guinney, J.; Dienstmann, R.; Wang, X.; de Reyniés, A.; Schlicker, A.; Soneson, C.; Marisa, L.; Roepman, P.; Nyamundanda, G.; Angelino, P.; et al. The consensus molecular subtypes of colorectal cancer. *Nat. Med.* **2015**, *21*, 1350–1356. [[CrossRef](#)] [[PubMed](#)]
- Hoorn, S.T.; de Back, T.R.; Sommeijer, D.W.; Vermeulen, L. Clinical Value of Consensus Molecular Subtypes in Colorectal Cancer: A Systematic Review and Meta-Analysis. *J. Natl. Cancer Inst.* **2021**, *144*, 503–516. [[CrossRef](#)]
- Mooi, J.K.; Wirapati, P.; Asher, R.; Lee, C.K.; Savas, P.; Price, T.J.; Townsend, A.; Hardingham, J.; Buchanan, D.; Williams, D.; et al. The prognostic impact of consensus molecular subtypes (CMS) and its predictive effects for bevacizumab benefit in metastatic colorectal cancer: Molecular analysis of the AGITG MAX clinical trial. *Ann. Oncol.* **2018**, *29*, 2240–2246. [[CrossRef](#)] [[PubMed](#)]
- Lenz, H.; Ou, F.; Venook, A.P.; Hochster, H.S.; Niedzwiecki, D.; Goldberg, R.M.; Mayer, R.J.; Bertagnolli, M.M.; Blanke, C.D.; Zemla, T.; et al. Impact of consensus molecular subtype on survival in patients with metastatic colorectal cancer: Results from CALGB/SWOG 80405 (Alliance). *J. Clin. Oncol.* **2019**, *37*, 1876–1885. [[CrossRef](#)] [[PubMed](#)]
- Al-Akhrass, H.; Christou, N. The clinical assessment of microRNA diagnostic, prognostic, and theranostic value in colorectal cancer. *Cancers* **2021**, *13*, 2916. [[CrossRef](#)] [[PubMed](#)]
- Rupaimoole, R.; Slack, F.J. MicroRNA therapeutics: Towards a new era for the management of cancer and other diseases. *Nat. Rev. Drug Discov.* **2017**, *16*, 203–221. [[CrossRef](#)] [[PubMed](#)]
- Fessler, E.; Jansen, M.; De Sousa E Melo, F.; Zhao, L.; Prasentyanti, P.R.; Rodermond, H.; Kandimalla, R.; Linnekamp, J.F.; Franitza, M.; van Hoof, S.R.; et al. A multidimensional network approach reveals microRNAs as determinants of the mesenchymal colorectal cancer subtype. *Oncogene* **2016**, *35*, 6026–6037. [[CrossRef](#)] [[PubMed](#)]
- Cantini, L.; Isella, C.; Petti, C.; Picco, G.; Chiola, S.; Ficarra, E.; Caselle, M.; Medico, E. MicroRNA-mRNA interactions underlying colorectal cancer molecular subtypes. *Nat. Commun.* **2015**, *6*, 8878. [[CrossRef](#)] [[PubMed](#)]
- Kandimalla, R.; Shimura, T.; Mallik, S.; Sonohara, F.; Tsai, S.; Evans, D.B.; Kim, S.C.; Baba, H.; Kodera, Y.; Von Hoff, D.; et al. Identification of Serum miRNA Signature and Establishment of a Nomogram for Risk Stratification in Patients with Pancreatic Ductal Adenocarcinoma. *Ann. Surg.* **2022**, *275*, E229–E237. [[CrossRef](#)] [[PubMed](#)]
- Adam, R.S.; Poel, D.; Ferreira, L.M.; Spronck, J.M.A.; de Back, T.R.; Torang, A.; Gomez, P.M.B.; ten Hoorn, S.; Markowitz, F.; Wang, X.; et al. Development of a miRNA-based classifier for detection of colorectal cancer molecular subtypes. *Mol. Oncol.* **2022**, *16*, 2693–2709. [[CrossRef](#)] [[PubMed](#)]

13. Muniategui, A.; Nogales-Cadenas, R.; Vázquez, M.; Aranguren, X.L.; Aguirre, X.; Luttun, A.; Prosper, F.; Pascual-Montano, A.; Rubio, A. Quantification of miRNA-mRNA interactions. *PLoS ONE* **2012**, *7*, e30766. [[CrossRef](#)] [[PubMed](#)]
14. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N.S.; Wang, J.T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A Software Environment for Integrated Models. *Genome Res.* **2003**, *13*, 2498–2504. [[CrossRef](#)] [[PubMed](#)]
15. Del Rio, G.; Koschützki, D.; Coello, G. How to identify essential genes from molecular networks? *BMC Syst. Biol.* **2009**, *3*, 102. [[CrossRef](#)]
16. Tabas-Madrid, D.; Muniategui, A.; Sánchez-Caballero, I.; Martínez-Herrera, D.J.; Sorzano, C.O.S.; Rubio, A.; Pascual-Montano, A. Improving miRNA-mRNA interaction predictions. *BMC Genomics* **2014**, *15* (Suppl. 10), S2. [[CrossRef](#)] [[PubMed](#)]
17. Colaprico, A.; Silva, T.C.; Olsen, C.; Garofano, L.; Cava, C.; Garolini, D.; Sabedot, T.S.; Malta, T.M.; Pagnotta, S.M.; Castiglioni, I.; et al. TCGAAbiolinks: An R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res.* **2016**, *44*, e71. [[CrossRef](#)]
18. Love, M.I.; Huber, W.; Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **2014**, *15*, 550. [[CrossRef](#)]
19. Nishida, N.; Magahara, M.; Sato, T.; Mimori, K.; Sudo, T.; Tanaka, F.; Shibata, K.; Ishii, H.; Sugihara, K.; Doki, Y.; et al. Microarray analysis of colorectal cancer stromal tissue reveals upregulation of two oncogenic miRNA clusters. *Clin. Cancer Res.* **2012**, *18*, 3054–3070. [[CrossRef](#)]
20. Ogando, J.; Tardáguila, M.; Díaz-Alderete, A.; Usategui, A.; Miranda-Ramos, V.; Martínez-Herrera, D.J.; de la Fuente, L.; García-León, M.J.; Moreno, M.C.; Escudero, S.; et al. Notch-regulated MIR-223 targets the aryl hydrocarbon receptor pathway and increases cytokine production in macrophages from rheumatoid arthritis patients. *Sci. Rep.* **2016**, *6*, 20223. [[CrossRef](#)]
21. Wang, S.; He, X.; Zhou, R.; Jia, G.; Qiao, Q. STAT3 induces colorectal carcinoma progression through a novel miR-572-MOAP-1 pathway. *Onco. Targets Ther.* **2018**, *11*, 3475–3484. [[CrossRef](#)]
22. Wang, S.; Zeng, Y.; Zhou, J.; Nie, S.; Peng, Q.; Gong, J.; Huo, J. MicroRNA-1246 promotes growth and metastasis of colorectal cancer cells involving CCNG2 reduction. *Mol. Med. Rep.* **2016**, *13*, 273–280. [[CrossRef](#)]
23. Sun, H.; Chen, X.; Ji, H.; Wu, T.; Lu, H.; Zhang, Y.; Li, H.; Li, Y. MiR-494 is an independent prognostic factor and promotes cell migration and invasion in colorectal cancer by directly targeting PTEN. *Int. J. Oncol.* **2014**, *45*, 2486–2494. [[CrossRef](#)] [[PubMed](#)]
24. Rhodes, D.R.; Chinnaiyan, A.M. Integrative analysis of the cancer transcriptome. *Nat. Genet.* **2005**, *37*, S31–S37. [[CrossRef](#)] [[PubMed](#)]
25. Wang, Y.; Lin, Y.; Guo, Y.; Pu, X.; Li, M. Functional dissection of human targets for KSHV-encoded miRNAs using network analysis. *Sci. Rep.* **2017**, *7*, 3159. [[CrossRef](#)] [[PubMed](#)]
26. Naqvi, A.R.; Shango, J.; Seal, A.; Shukla, D.; Nares, S. Viral miRNAs alter host cell miRNA profiles and modulate innate immune responses. *Front. Immunol.* **2018**, *9*, 433. [[CrossRef](#)]
27. Zeng, M.; Zhu, L.; Li, L.; Kang, C. miR-378 suppresses the proliferation, migration and invasion of colon cancer cells by inhibiting SDAD1. *Cell. Mol. Biol. Lett.* **2017**, *22*, 12. [[CrossRef](#)]
28. Chang, H.; Ye, S.; Pan, S.; Kuo, T.; Liu, B.C.; Chen, Y.; Huang, T. Overexpression of miR-194 Reverses HMGA2-driven Signatures in Colorectal Cancer. *Theranostics* **2017**, *7*, 3889–3900. [[CrossRef](#)] [[PubMed](#)]
29. Lu, Y.; Zhao, X.; Liu, Q.; Li, C.; Graves-Deal, R.; Cao, Z.; Singh, B.; Franklin, J.L.; Wang, J.; Hu, H.; et al. LncRNA MIR100HG-derived miR-100 and miR-125b mediate cetuximab resistance via Wnt/ $\beta$ -catenin signaling. *Nat. Med.* **2017**, *23*, 1331–1341. [[CrossRef](#)] [[PubMed](#)]
30. Michael, M.Z.; O'Connor, S.M.; van Holst Pellekaan, N.G.; Young, G.P.; James, R.J. Reduced Accumulation of Specific MicroRNAs in Colorectal Neoplasia. *Mol. Cancer Res.* **2003**, *1*, 882–891.
31. Chivukula, R.R.; Shi, G.; Acharya, A.; Mills, E.W.; Zeitels, L.R.; Anandam, J.L.; Abdelnaby, A.A.; Balck, G.C.; Mansour, J.C.; Yopp, A.C.; et al. An essential mesenchymal function for miR-143/145 in intestinal epithelial regeneration. *Cell* **2014**, *157*, 1104–1116. [[CrossRef](#)] [[PubMed](#)]
32. Liu, F.; Qi, L.; Liu, B.; Liu, J.; Zhang, H.; Che, D.; Cao, J.; Shen, J.; Geng, J.; Bi, Y.; et al. Fibroblast activation protein overexpression and clinical implications in solid tumours: A meta-analysis. *PLoS ONE* **2015**, *10*, e0116683. [[CrossRef](#)]
33. Yasunaga, M.; Matsumura, Y. Role of SLC6A6 in promoting the survival and multidrug resistance of colorectal cancer. *Sci. Rep.* **2014**, *4*, 4852. [[CrossRef](#)] [[PubMed](#)]