Colorectal cancers with microsatellite instability display unique miRNA profiles

Francesc Balaguer¹,²*, Leticia Moreira²*, Jose Juan Lozano³, Alexander Link¹,⁴, Georgina Ramirez², Yan Shen¹, Miriam Cuatrecasas², Mildred Arnold¹, Stephen J. Meltzer⁵, Sapna Syngal⁶, Elena Stoffel⁶, Rodrigo Jover⁷, Xavier Llor⁸, Antoni Castells², C. Richard Boland¹, Meritxell Gironella²§, Ajay Goel¹§.

¹Department of Internal Medicine, Division of Gastroenterology, Charles A. Sammons Cancer Center and Baylor Research Institute, Baylor University Medical Center, Dallas, USA.
²Department of Gastroenterology, Institut de Malalties Digestives i Metabòliques, Hospital Clinic, Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBEREHD), IDIBAPS, University of Barcelona, Barcelona, Catalonia, Spain.
³Plataforma de Bioinformática, CIBEREHD, Barcelona, Catalonia, Spain.
⁴Department of Gastroenterology, Hepatology and Infectious Diseases, Otto-von-Guericke University, Magdeburg, Germany
⁵Division of Gastroenterology, Johns Hopkins Medical Institute, Baltimore, MD, USA
⁶Department of Internal Medicine, Brigham and Women’s Hospital, Boston, Massachusetts, USA
⁷Department of Gastroenterology, Hospital Universitario de Alicante, Alicante, Spain.
⁸Section of Digestive Diseases and Nutrition, University of Illinois at Chicago, Chicago, IL, USA

* Both authors contributed equally
§ Shared senior authorship

Running title: miRNA expression in MSI CRC

Keywords: Colorectal cancer; microRNA; Lynch syndrome; microsatellite instability.

Grant support:The present work was supported by grants from Fundación Mutua Madrileña, Instituto de Salud Carlos III/FEDER (PI10-02888) and Fundació Olga Torres to MG, and from the National Cancer Institute (R01 CA72851 and CA129286), National Institutes of Health, and funds from the Baylor Research Institute to CRB and AG. AC is supported by grants from Ministerio de Ciencia e Innovación (SAF2010-19273), and Agència de Gestió d’Ajuts Universitaris i de Recerca (2009 SGR 849). CIBEREHD is funded by the Instituto de Salud Carlos III. FB was supported by a grant from Fundación Alfonso Martín Escudero and from Societat Catalana de Digestologia. LM was supported by a grant Josep Font from the Hospital Clinic of Barcelona. The work was carried out (in part) at the Esther Koplowitz Centre, Barcelona.

Address for correspondence: Ajay Goel, Ph.D. (ajay.goel@baylorhealth.edu), Gastrointestinal Cancer Research Laboratory, Baylor University Medical Center, 3500 Gaston Avenue, Suite H-250, Dallas, TX75246, USA. Phone: 214-820-2692; Fax: 214-818-9292. Meritxell Gironella, Ph.D. (meritxell.gironella@ciberehd.org), Gastrointestinal and Pancreatic Oncology Group. Esther Koplowitz Centre; Rosselló, 149-153; 08036 Barcelona, Catalonia, Spain. Phone: +34-93-227-5400; ext.4183; Fax: +34-93-312-9405.

Conflict of interest: There is no conflict of interest to disclose for all authors.

Abbreviations: colorectal cancer, CRC; CpG island Methylator Phenotype, CIMP; false discovery rate, FDR; fold change, FC; formalin-fixed paraffin-embedded, FFPE; microRNA, miRNA; microsatellite instability, MSI; microsatellite stability, MSS; mismatch repair, MMR; normal colonic, N-C; overall error rate, OER.

Author Contributions: study concept and design (FB, AG, MG, LM); acquisition of data (FB, LM, RG, YS, MG); analysis and interpretation of data (FB, LM, J-JL, AG, MG); drafting of the manuscript (FB, LM, J-JL, MG, AG); critical revision of the manuscript for important intellectual content (CRB, AC, SS and SJM); statistical analysis (J-JL, FB, LM); provision of samples (MC, MA, SS, ES, RJ, XL).
Translational Relevance

Accumulating evidence in the last few years clearly indicates that microRNAs (miRNAs) have critical functions across various biological processes, particularly, in the pathogenesis of human cancers. Since miRNA expression patterns are unique for different cancers, there is a growing interest in exploiting the diagnostic potential of these small noncoding RNAs in the classification of various cancers and their subtypes. Colorectal cancer (CRC) is a heterogeneous disease. There are currently several types of colorectal cancer recognized by histopathological methods; however, the disease classification at a molecular level is currently an area of intensive investigation. In this study, we have used miRNA expression profiling approach for the molecular classification of familial and sporadic forms of CRC. We have identified that miRNA expression profiles not only distinguished between tumor and normal tissues, but more importantly, a subset of miRNAs successfully discriminated between Lynch syndrome, sporadic MSI and MSS CRCs. These findings have clinical and translational relevance in the diagnosis, prognosis and guiding treatment strategies for various subtypes of colorectal cancer patients.
Abstract

**Purpose:** microRNAs (miRNAs) are small non-coding transcripts that play an important role in carcinogenesis. miRNA expression profiles have been shown to discriminate between different types of cancers. The aim of this study was to analyze the global miRNA signatures in various groups of colorectal cancers (CRCs) based on the presence of microsatellite instability (MSI).

**Experimental Design:** We analyzed genome-wide miRNA expression profiles in 54 CRCs (22 with Lynch syndrome, 13 sporadic MSI due to MLH1 methylation, 19 without MSI [MSS]) and 20 normal colonic tissues by miRNA microarrays. Using an independent set of MSI samples (13 with Lynch syndrome and 20 with sporadic MSI) we developed a miRNA-based predictor to differentiate both types of MSI by quantitative reverse transcriptase PCR.

**Results:** We found that the expression of a subset of 9 miRNAs significantly discriminates between tumor and normal colonic mucosa (overall error rate (OER) =0.04). More importantly, Lynch syndrome tumors displayed a unique miRNA profile compared with sporadic MSI tumors; miR-622, miR-1238 and miR-192* were the most differentially expressed miRNAs between these two groups. We developed a miRNA-based predictor able to differentiate the type of MSI in an independent set of samples.

**Conclusions:** CRC tissues show distinct miRNA expression profiles compared to normal colonic mucosa. The discovery of unique miRNA expression profiles that can successfully discriminate between Lynch syndrome, sporadic MSI and sporadic MSS CRCs provides novel insights into the role of miRNAs in colorectal carcinogenesis which may contribute to the diagnosis, prognosis and treatment of this disease.
**Introduction**

Colorectal cancer (CRC) is one of the most common tumors in Western countries and the second leading cause of cancer-related deaths(1). From a molecular standpoint, CRC is a complex and heterogeneous disease caused by the accumulation of genetic and epigenetic events(2-4). Based on the evidence that tumors with similar molecular characteristics arise and behave similarly, the molecular classification of CRC has been greatly developed over the last decade(5, 6). The main goal of classification is to empirically understand the pathogenesis and predict the biological behavior of each tumor, which may have diagnostic, prognostic and therapeutic implications.

Based on the presence or absence of microsatellite instability (MSI), the hallmark of DNA mismatch repair (MMR) deficiency, CRC may be classified into 3 groups: Lynch syndrome, sporadic MSI and microsatellite stable (MSS) tumors. Lynch syndrome, which account for 3% of all CRCs is caused by a germline mutation in one of the MMR genes (*MLH1*, *MSH2*, *MSH6* and *PMS2*)(7). Tumors from Lynch syndrome patients are typically characterized by MSI and the absence of the protein corresponding to the mutated gene, and are associated with a better prognosis than MSS tumors. On the other hand, the majority of CRCs with MSI arise through biallelic somatic methylation of the *MLH1* promoter in older patients with no family history of CRC (so called sporadic MSI)(8, 9). This form of CRC, which accounts for ~12% of all CRCs, arises through a process that involves the CpG island Methylator Phenotype (CIMP), is usually associated with *BRAF* mutations (never present in Lynch syndrome) and is associated with a reduced mortality(10, 11). Finally, MSS tumors account for 85% of all CRCs and are often characterized by chromosomal instability, aneuploidy and a worse prognosis than MSI tumors.

MicroRNAs (miRNAs) are small non-coding RNA molecules of ~18-22 nucleotides that negatively regulate gene expression by inhibiting translation or inducing messenger RNA (mRNA) degradation(12). Since their discovery, miRNAs have been implicated in various cellular processes including apoptosis, differentiation and cell proliferation, and play a key role in carcinogenesis(13-15). Altered miRNA expression has been reported in most tumors, including CRC, and specific miRNAs dysregulated in certain types of cancers may act as biomarkers for diagnosis and outcome in that cancer type(16). Besides their potential as diagnostic and prognostic tools, one of the most interesting biological features of miRNA compared to mRNA is that they are present in different tissues in a stable form, and are remarkably protected from endogenous degradation, thus making it feasible to analyze their expression in archived materials(17, 18). Finally, understanding the miRNA regulation is critical to gain insight into the different colorectal carcinogenesis pathways and their specific role as potential therapeutic targets.

The miRNA profiles of CRC have been analyzed in several studies(16, 19-22), however, only a few have specifically analyzed miRNA signatures in different subtypes of CRC based on the presence of MSI(16, 23-25). Although the current evidence suggests that the miRNA profile can distinguish between MSI and MSS tumors,
most studies have been limited to a modest number of samples. In addition, the majority of studies have used arrays with a limited number of miRNAs, and more importantly, none have validated their results in an independent set of samples. Another issue is that the nature of MSI in the tumor (i.e., Lynch syndrome or sporadic MSI) has not been described in prior studies, and consequently, the miRNA signature in Lynch syndrome tumors remains unknown. In this study we have addressed the issues raised above by analyzing global miRNA signatures in well-characterized CRCs based on the presence of MSI and validated our results in an independent set of samples.
**Patients and Methods**

**Patient selection**

A total of 87 CRCs available as formalin-fixed paraffin-embedded (FFPE) tissue were divided into training and test sets. The training set was used for miRNA microarray profiling and included 54 CRCs and 20 normal colonic (N-C) tissues. CRC tissues were divided in 3 groups. The first group, Lynch syndrome (n=22), was comprised of tumors with MMR deficiency (loss of MLH1/MSH2/MSH6/PMS2 protein expression and/or MSI). These tumors were collected either from carriers of a germline mutation in one of the MMR genes (n=13; 7 in MLH1, 5 in MSH2 and 1 in MSH6) or from patients fulfilling the Amsterdam criteria but without an identified germline mutation (n=9; 6 with loss of MLH1 and PMS2, and 3 with loss of MSH2 and MSH6). The second group, sporadic MSI (n=13), included tumors with loss of MLH1 protein expression from non-familial CRC cases associated with somatic MLH1 promoter methylation. The third group, MSS (n=19), were mismatch repair-proficient tumors. N-C tissues were obtained from individuals undergoing colonic surgery for reasons other than cancer (i.e. diverticulosis), showing microscopically normal mucosa. All samples from the training set used for microarray analysis were carefully selected from the same institution (Baylor University Medical Center at Dallas).

The test set included an independent collection of Lynch syndrome (n=13; 4 with a germline mutation in MLH1, 5 in MSH2, 3 in MSH6, and 1 EpCam deletion) and sporadic MSI (n=20) tumors. This set of tumors was used to develop a miRNA-based predictor to distinguish both types of MSI based on the microarray results from the training set. These samples were obtained from different institutions (Lynch syndrome tumors from Brigham and Women’s Hospital, Boston and Hospital Universitario de Alicante; sporadic MSI from Hospital Universitario de Alicante and Hospital Clinic of Barcelona). The clinico-pathological features of the samples included in the study are detailed in Table 1. Informed written consent was obtained from all patients and the project was approved by the institutional review board of all participating institutions.

**Mismatch repair deficiency analysis**

Mismatch repair deficiency was evaluated in all cases by MSI analysis and immunohistochemistry for the 4 MMR proteins (MLH1/MSH2/MSH6/PMS2), and 100% concordance was observed in both analyses. MSI testing was performed using the five markers of the original Bethesda panel (BAT25, BAT26, D2S123, D5S346 and D17S250)(26). Since mononucleotide sequences have been shown to perform better in identifying MSI-high tumors, we confirmed the MSI results using five quasi-monomorphic mononucleotide markers (BAT25, BAT26, NR21, NR24 and NR27) as recently described(27). MSI was defined as the presence of >2 unstable markers for the Bethesda panel, and >3 unstable markers for the mononucleotide penta panel. Tumors with stability at all loci were labeled as MSS. Immunohistochemistry for the 4 MMR proteins was performed as previously described(28).
Germline mutational analysis of the MMR genes was performed by Myriad Genetics, Inc. (Salt Lake City, UT). Tumor MLH1 promoter methylation was analyzed by bisulfite pyrosequencing as previously described(29). Primers sequences are available on request.

RNA extraction
Total RNA from 10μm thick microdissected FFPE tissue cuts was isolated using the RecoverAll™ Total Nucleic Acid Isolation Kit for FFPE tissues (Ambion Inc, Austin, TX) according to manufacturer instructions.

miRNA microarray experiments

RNA processing
Global miRNA expression profiles were analyzed using the MicroRNA Expression Profiling Assay based on BeadArray™v.2 (Illumina Inc., San Diego, CA), which contains 1,146 probes including 743 validated miRNAs. The miRNA microarray analysis was carried out in collaboration with the Genomics Platform CICbioGUNE (Center for Cooperative Research in Biosciences, Derio, Spain). The assay was performed following manufacturer's instructions (Illumina, Inc., San Diego, CA, USA), as previously described(30, 31).

Microarray data normalization
Data were extracted using Bead Studio data analysis software and transformed to a log base 2 scale. Microarray data from 74 samples (20 N-C, 22 Lynch, 13 sporadic MSI and 19 MSS) were quantile-normalized using the Lumi bioconductor package(32). Next, we employed a conservative probe-filtering step, excluding those probes not reaching a detection p value <0.05 in 90% of the samples, which resulted in the selection of a total of 891 probes out of the original 1146 set. Fold changes (FC) in miRNA expression in the microarray analyses were calculated based on the difference of the group median values ($2^{\log_{base 2} \text{difference}}$). All microarray data discussed in this publication have been deposited in NCBI’s Gene Expression Omnibus (GEO; accession number GSE30454).

Differential miRNA expression assessment and prediction
We first used LIMMA (Linear Models for Microarray Data)(33) to identify miRNAs differentially expressed among the four groups included in the study (N-C, Lynch syndrome, sporadic MSI and MSS) within the filtered 891-probe set. LIMMA uses linear models and empirical Bayes paired moderated t-statistics and F-statistics. Since the MicroRNA Expression Profiling Assay from Illumina includes 403 non-validated probes, these were not considered for further analyses. False discovery rates (FDR) were determined using the Benjamini-Hochberg procedure. The top most significant 50 miRNAs using F-statistics were used on the 74 sample set to perform a correspondence analysis as implemented in the bga (between group analysis) function included in the made4 package(34). Bga plots are based on principal component analysis (PCA), where pc1 corresponds to the x axis and pc2 corresponds to the y axis. This method is capable of visualizing high dimensional data (such as multiple...
expression measurements) in a 2D graph in which the areas delimited by the ellipses represent 95% of the estimated binormal distribution of the sample scores on the first and second axes(35).

Predictability of the most discriminant miRNAs with statistical significance was further explored using the nearest shrunken centroid classifier implemented in the PAM package(36) to identify the minimal set of miRNAs capable of discriminating between the following groups: tumor tissue (CRC) vs N-C, Lynch vs sporadic MSI, and sporadic MSI vs MSS. To estimate the classification accuracy of the miRNA signature in the training set (area under the receiving operating curves (AUROC) with 95% confidence interval (CI)), a 10-fold crossvalidation was conducted by selecting the threshold associated with the lowest error rate and filtering the noisiest miRNAs(37). For each of these groups, the PAM classifier was then used to perform a multidimensional scaling analysis on the basis of between sample Euclidean distances as implemented by the isoMDS function in R. This method is capable of visualizing high dimensional data (such as multiple expression measurements) on a 3D graph in which the distances between samples are kept as unchanged as possible(37).

Validation of microarray data by TaqMan qRT-PCR
We first performed a technical validation of the microarray data by analyzing the expression of a subset of 10 miRNAs in a randomly selected subset of samples from the training set (8 N-C, 7 Lynch syndrome, 7 sporadic MSI and 8 MSS). The selection of miRNAs for validation was rigorously based on the following criteria: log base 2 intensity >8, FDR <5%, fold change (FC) and selection in either LIMMA or PAM analyses. Following these criteria, we selected those miRNAs that were commonly found in the different comparisons. Expression of miRNAs was analyzed using the TaqMan miRNA Assay (Applied Biosystems Inc., Foster City, CA) as previously described. The expression of miR-16 was used as endogenous control(38). All experiments were performed in triplicate. In order to normalize the miRNA expression levels from different experiments, the miRNA expression in each sample was calculated by comparing the normalized Ct of the sample with a normalized Ct of a technical replicate common in all experiments (\(\Delta\Delta C_t = \Delta C_t \) of sample - \(\Delta C_t \) of technical replicate). Fold change was calculated based on the 2^{-\Delta\Delta C_t} method.

We next evaluated the expression of the same 10 target miRNAs in an independent set of MSI tumor samples (n=33), including Lynch syndrome and sporadic-MSI tumors (test set). Based on this analysis we developed a miRNA-based predictor model to differentiate the type of MSI (Lynch syndrome vs sporadic MSI). The method implements a forward stepwise cross-validated procedure to find the optimal prediction model. We specified the Linear Discriminant Analysis method as classification rule, and different candidate models were evaluated with 10-fold cross-validation and 1000 random split using the subset of 13 Lynch samples and 20 sporadic MSI samples. All these algorithms are included in the mipp. seq function from MiPP package(39). The performance of the resulting model was evaluated in the set of MSI tumors from the training set used for technical validation (n=14). We next combined the MSI cases from both training and test set cohorts to assess the performance of the predictor model to discriminate the type of MSI based on the \(\Delta\Delta C_t \) value. We constructed receiving operating curves and determined the AUROC (95%CI).
**In situ hybridization**

*In situ* detection of miR-622 on FFPE colonic tissues (5 primary CRC and 5 normal colonic mucosae) was performed as previously described(40). Positive controls (RNU6B, Exiqon) and no probe controls were included for each hybridization procedure.

**Statistical Analysis**

Quantitative variables were analyzed using Student's test. Qualitative variables were analyzed using either the Chi Square Test or the Fisher's test. A two sided p-value of < 0.05 was regarded as significant. Clinical data was analyzed using Graph Pad Prism 4.0 (San Diego, CA) statistical software.
Results
Overview
In this study we performed global miRNA microarray profiling on a large collection of tumor and N-C tissues categorized by the presence MMR deficiency with the aim of recognizing the most significant differences in miRNA expression. This is the first study to investigate the miRNA expression profile in Lynch syndrome, the most common form of hereditary CRC, and compare it with the sporadic form of MMR deficiency, which is caused by somatic inactivation of MLH1 by methylation of its promoter. The study was conducted in three steps: a) miRNA microarray profiling in a training set (n=74) comprised of 4 well-defined groups: N-C tissue, Lynch syndrome tumors, sporadic MSI tumors and MSS tumors; b) technical validation of the most significant results by qRT-PCR in an randomly selected subset of samples from the training set (n=30); and c) development of a predictor to differentiate the type of MSI (Lynch syndrome vs sporadic MSI tumors) using an independent set of samples (n=33). Clinico-pathological characteristics of all patients included in this study are summarized in Table 1. There were no clinical differences between the training set and the test set.

A miRNA expression signature discriminates normal colonic mucosa from tumor tissue
We first used linear effects models (LIMMA) to determine the miRNAs differentially expressed between the four groups included in the study, identifying 692 probes with an adjusted F <0.05 (Supplementary Table 1). Expression profiles of the 50 most significant miRNAs are depicted in Figure 1A. Between group analysis (bga) plot was then performed to visually represent the distance/separation between the 4 different groups according to the expression of the 50 most significant miRNAs. As depicted in Figure 1B, N-C tissues and tumor tissues appeared as 2 clearly separated groups, and within tumor samples, sporadic MSI, MSS and Lynch syndrome tumors were also visibly different.

We identified 499 probes differentially expressed between normal colonic mucosa and tumor tissue (FDR<0.05) (Supplementary Table 1). To identify the minimal set of miRNAs capable of predicting tumor tissues, PAM was performed comparing tumor vs N-C tissue resulting in the identification of 9 miRNAs (all of them present in the LIMMA list) with an overall error rate of 0.04 (AUROC, 0.99 (95%CI, 0.98-1)) (Supplementary Table 2a). In particular, upregulation in tumor tissue of miR-1238, miR-938, miR-622, and downregulation of miR-133b, miR-490-3p, miR-138 and miR-1 were among the most significantly dysregulated miRNAs. Overall, the miRNA microarray data resulted in the identification of a set of miRNAs capable of discriminating tumor vs N-C mucosa with a high degree of accuracy.

We next analyzed the specific miRNA profile for each tumor type compared to N-C mucosa, and found that a subset of 176, 46 and 55 probes were exclusively and significantly dysregulated in sporadic MSI, MSS and Lynch syndrome tumors, respectively (Figure 1C and Supplementary Table 3).
Tumors from patients with suspected Lynch syndrome show a similar miRNA profile compared with the proven ones

The Lynch syndrome group in our study involved tumor tissues from patients with an identified germline mutation in one of the DNA MMR genes (i.e. Lynch-mutated), and tumor tissues with MMR deficiency belonging to patients that fulfilled the Amsterdam criteria but had a negative genetic test (i.e. Lynch-like). From a clinical standpoint, both groups are considered to have the same disease, and it is assumed that the underlying genetic mutation remains undetected by current analytical methods in the latter group. To explore the similarities between these two subgroups in miRNA expression, we performed an unsupervised hierarchical clustering analysis, and the dendrogram revealed a lack of clustering between these 2 subgroups (Figure 2A). Notably, none of the probes of the array showed a significant difference between the two groups. A multidimensional scaling plot showed that both subgroups are grouped together, in concordance with the clinical phenotype (Figure 2B). Overall, these findings indicate that Lynch syndrome-like patients with unidentified germline mutation have a similar miRNA profile as seen in mutated cases, suggesting the presence of a common molecular basis.

A miRNA expression signature discriminates Lynch syndrome from sporadic MSI tumors

We next evaluated the ability of microarray data to predict the molecular type of CRC based on the type of MMR deficiency. Lynch syndrome accounts for about 3% of all MSI CRC and is caused by germline mutations in DNA MMR genes, whereas the most frequent cause of MSI involves CIMP, associated with somatic methylation of the MLH1 gene. We identified 418 probes differentially (FDR<0.05) expressed between these two groups (Supplementary Table 1). To explore the possibility of distinguishing both types of MSI based on the miRNA microarray signature, we employed PAM prediction (Figure 2C), identifying a set of 31 miRNAs (29 downregulated and 2 upregulated) able to predict the type of MSI with an overall error rate of 0.11 (AUROC, 0.94 (95%CI, 0.84-1)) (Supplementary Table 2b). The most up- and down-regulated miRNAs in Lynch syndrome tumors compared to sporadic MSI were miR-30a*, miR-16-2*, 362-5p and miR-1238, and miR-622, respectively.

Multidimensional scaling was next used to plot Lynch syndrome and sporadic MSI samples based on the PAM-derived signature, and there was a remarkable separation between them (Figure 2D). When we performed a sub-analysis comparing only those Lynch syndrome tumors with MLH1mutations versus sporadic MSI tumors, we obtained the same different miRNA patterns found by analyzing all Lynch syndrome tumors together, demonstrating that the miRNA profiles do not exclusively depend on MLH1 mutations. Overall, these results suggest that Lynch syndrome and sporadic MSI CRCs can be distinguished based on the miRNA expression profile.

The miRNA expression signature discriminates between sporadic MSI and MSS tumors
We identified 353 probes differentially expressed between sporadic MSI and MSS tumors (FDR<0.05) (Supplementary Table 1). The analysis of miRNA expression profiles using PAM revealed a signature of 59 miRNAs capable of predicting the presence of MSI with an overall error rate of 0.25 (AUROC, 0.73 (95%CI, 0.53-0.93)) (Supplementary Table 2c). The most up- and down-regulated miRNAs in sporadic MSI compared to MSS tumors included miR-938, miR-615-5p, miR-1184, miR-551a, miR-622 and miR-17-5p, miR-192* and miR-337-3p, respectively. Using the PAM cross-validation procedure, all but 4 tumors were correctly assigned, and although both groups were separated in the multidimensional scaling plot (Supplementary Figure), the spatial differential distribution was not as clean as in the previous comparisons.

Technical validation of miRNA expression

We employed TaqMan qRT-PCR to confirm the expression differences of target miRNAs identified by microarray in a randomly selected subset of samples from the training set (technical validation). Selected target miRNAs for qRT-PCR experiments included 10 miRNAs that were selected among LIMMA or PAM analyses: miR-1238, miR-192*, miR-362-5p, miR-938, miR-622, miR-133b, miR-16-2*, miR-30a*, miR-183 and miR-486-5p. The results from these analyses are shown in Supplementary Table 4. Overall, we were able to validate most of the microarray results.

In this study we have revealed and validated several miRNAs that are differentially expressed in CRC tissues compared to normal colonic mucosa. We found that miR-1238 and miR-622 are consistently overexpressed in CRC, and we could successfully validate previously known dysregulated miRNAs in CRC (i.e. miR-133b and miR-30a*) (19). To further validate the microarray results, we performed in situ hybridization using 5’-DIG-labeled LNA probes for miR-622 in several normal colonic mucosa and CRC tissues to further investigate the pattern of expression of this miRNA. In normal colonic mucosa, miR-622 was expressed only in the colonic epithelial cells throughout the colonic crypts, with a gradient of miRNA expression decreasing from the bottom to the top of the crypts (Figure 3). CRC samples evaluated showed a marked increase in the expression of this miRNA, consistent with our observation that miR-622 is overexpressed in the majority of CRCs.

Differentiation of MMR-deficient tumors based on miRNA analysis

We next aimed to develop a predictor able to differentiate the type of mismatch repair deficiency based on miRNA analysis using an independent set of samples with sporadic-MSI (n=20) and Lynch syndrome tumors (n=13). For this purpose we analyzed by TaqMan qRT-PCR the expression of the 10 target miRNAs evaluated in the training set in an independent set of MSI tumors (Supplementary Table 4). Statistical analyses showed that the combination of the expression of 3 miRNAs (miR-622, miR-362-5p, miR-486-5p), all of them present in the PAM classifier identified in the microarray analysis, could differentiate the 2 types of MSI with high accuracy (AUROC, 0.77; 95%CI, 0.57-0.98). These results are of great significance since we could successfully validate the microarray results in an independent cohort of MSI CRC samples and develop a miRNA-based predictor to differentiate both types of MSI.
Discussion

In this study we performed miRNA profiling by microarrays in a group of CRCs categorized by the presence and type of MSI. Our results show that miRNAs can be used to discriminate between normal versus tumor tissue, and more importantly within tumor subtypes. We identified the miRNA signature in Lynch syndrome tumors and compare it to the sporadic form of MSI, which is caused by somatic methylation of \textit{MLH1}, showing that each type of MMR deficiency is associated with a unique miRNA signature. In addition, we demonstrate that tumors from patients with suspected Lynch syndrome who have an unidentified germline mutation display a similar miRNA expression profile as those in which a mutation has been identified, suggesting a shared molecular basis. Finally, using an independent cohort of MSI tumor tissues, we developed a miRNA-based predictor able to differentiate the type of MSI with high accuracy, reinforcing the robustness of this approach.

In agreement with previous reports, our findings confirm that numerous miRNAs are aberrantly expressed in CRC relative to normal tissues. Although several groups have profiled miRNAs in CRC tissues using different platforms(16, 19, 22, 24, 25)we have used the most comprehensive commercial platform available including 1,146 probes with 743 validated human miRNAs. Despite methodological differences between our study and previous reports, we found concordant expression of previously reported miRNAs altered in cancers (i.e. down-regulation of miR-9, miR-129, miR-137, miR-34b, miR-133b, miR-124 and upregulation of miR-183, miR-31, miR-182), and described several other miRNAs that are significantly deregulated in tumor tissues (i.e. miR-1238, miR-938, miR-622, miR-1290, miR-490-3p). More importantly, qRT-PCR experiments confirmed the most significant microarray results that can discriminate between tumor and N-C tissue (i.e.miR-1238, miR-622 or miR-938). These results illuminate novel miRNAs that are involved in the pathogenesis of CRC, which provide potential diagnostic and prognostic markers.

miRNA patterns from Lynch syndrome tumors and CRCs with sporadic MSI are significantly different. Although these two conditions share the same unique molecular mechanism of tumor development (i.e. MSI), the underlying cause is quite different. Lynch syndrome is an autosomal dominant disorder caused by germline mutations in one of four MMR genes (\textit{MLH1, MSH2, MSH6, PMS2}) and accounts for a minority of MMR deficient tumors (~20%)(7). Sporadic MSI CRCs, which account for ~80% of MSI cases, are caused by somatic inactivation of the \textit{MLH1} gene through biallelic methylation of its promoter in the setting of the CIMP. Tumors with CIMP are characterized by altered patterns of DNA methylation, with concordant hypermethylation of multiple tumor suppressor genes, although the cause of this alteration remains unknown(11). Consistent with the distinct genetic and epigenetic backgrounds, we found that both types of MSI can be distinguished by the miRNA profile. Microarray data revealed a set of 31 miRNAs that could be used as classifiers with high accuracy (AUROC, 0.94). In addition, using an independent set of MSI tumors (including Lynch syndrome and sporadic MSI) we found that the expression of 3 miRNAs identified in the microarray analysis (miR-622, miR-362-5p and miR-486-5p) could accurately classify the type of MSI, although these results will require further independent validation in the future. The explanation for the differential miRNA expression between these 2 groups of
tumors deserves a more in depth analysis. Since sporadic MSI tumors are consistently associated with the CIMP phenotype, it is plausible to suggest that this phenotype could explain the observed differences. In addition, Melo et al.(41) recently showed that somatic frameshift mutations in one of the miRNA processing genes (TARBP2) could explain the miRNA disruption in Lynch syndrome and sporadic MSI tumors. It is worth mentioning that Lynch syndrome tumors from MLH1, MSH2 and MSH6 mutation carriers shared the same miRNA profile, suggesting a common mechanism of miRNA dysregulation. Overall, our results could shed light on the molecular mechanism underlying the sporadic MSI and Lynch tumors, and contribute to the generation of biomarkers to improve diagnosis and prognosis in these two forms of CRC.

This study also demonstrates that MMR deficient tumors from patients that fulfill the Amsterdam criteria in families with an unidentified germline mutation have a similar miRNA profile as those in whom the mutation has been found. In clinical practice, it is usually assumed that the underlying genetic mutation has not been detected by current methods, but it remains possible that these tumors have a unique pathogenesis. Our results support the hypothesis that these are all Lynch syndrome tumors, and that the germline mutations have been missed because of technical limitations in the gene analysis, since the global miRNA signatures resembles those in tumors from patients with known germline mutations in the MMR genes. This data suggest that the somatic miRNA profile could be used to predict the presence of a germline mutation in the MMR genes, which could have a significant impact in the genetic counseling of these patients. However, further studies are needed to validate this hypothesis.

Several studies have analyzed and compared the miRNA profiles in MSI compared to MSS tumors(16, 23-25). It is noteworthy, however, that the consistency of the results regarding the miRNAs that distinguish both types of tumors has been poor. There are several possible explanations for these discrepancies. First, although most CRCs with MSI are a consequence of somatic methylation of MLH1 promoter, the presence of Lynch syndrome tumors in the MSI group would have distorted the results. One publication(23)attempted to validate a previously published profile(24)and only 3/8 miRNAs provided confirmatory results using qRT-PCR. Second, biological differences other than the presence of MSI could also explain the differences between studies, and finally, technical differences (array characteristics, tissues source) cannot be ruled out.

There are several potential pitfalls from this study. First, colonic tissues used in this study came from different institutions and times. Although we made an earnest effort to collect the most recent tumors, differential processing procedures could have biased the results. Second, the number of tumors analyzed may not be enough to generalize our conclusions, and future studies may be needed to validate them. Finally, although most of the studies of miRNA expression in cancer tissues have been performed on frozen samples, we used FFPE tissue samples due to their ready availability. Several studies have shown that miRNAs are well preserved in FFPE tissues, and there is an excellent correlation between miRNA expression in fresh frozen and FFPE tissues(17, 18). In spite of these concerns, our microarray data validated several miRNAs shown
previously to be differentially expressed between healthy and CRC tissue and between MSI and MSS tumors. Comparative miRNA expression profiles between our study and previous literature is summarized in **Supplementary Table 5**. These findings are of considerable significance since they come from different populations, and analyses were performed using different technologies, which indicate the potential biological relevance of these miRNAs in the pathogenesis of CRC. For example, our results are quite consistent with the data obtained by Lanza et al(24), in which miRNA profiling was performed in 23 MSS and 16 MSI fresh frozen tissues using a custom array, and with the study from Sarver et al(25), where the miRNA profiling was evaluated in 12 MSI and 68 MSS tumors using Illumina microarray technology.

In summary, this study describes the miRNA signature in CRCs from Lynch syndrome patients and demonstrates a unique expression signature compared with sporadic MSI tumors caused by somatic methylation of the MLH1 promoter. In addition, we have discovered that the tumor miRNA profiles from patients with ‘suspected’ as well as ‘definitive’ Lynch syndrome showed a similar profile, suggesting common molecular pathogenesis for both categories of Lynch syndrome patients. Finally, using a comprehensive platform and a large number of samples, we have identified several miRNAs dysregulated between tumor and N-C tissue, and within molecular subtypes of CRC based on the presence of MSI. These miRNAs are likely to insight into the pathogenesis of CRC, but in a more immediate fashion, they may be used to classify tumors for diagnostic purposes – particularly in the case of a Lynch syndrome family without an identified germline mutation – and may be useful in the future for the design of individualized treatment strategies.
Figure legends

Figure 1. Differential miRNA expression between normal colonic mucosa and tumor tissue. A) This heat map shows the expression of the 50 most significant miRNAs identified by LIMMA in the four groups (Lynch syndrome, sporadic MSI, MSS, and N-C). Rows represent miRNAs and columns represent individual samples. B) The bga plot shows sample clustering based on the expression of the 50 most significant miRNAs. The four groups clearly clustered in distinct sets based on this group of miRNAs. C) The Venn diagram shows significantly dysregulated miRNAs among the 3 tumor subtypes (sporadic MSI, MSS and Lynch syndrome) compared to normal colonic mucosa (N-C).

Figure 2. miRNAs differentially expressed between sporadic MSI and Lynch syndrome tumors. A) Unsupervised hierarchical clustering of Lynch syndrome tumor samples based on the microarray results. Tumor tissues from patients with a germline mutation in one of the DNA mismatch repair genes (Lynch-mutated, in blue) and those with a negative genetic testing (Lynch-like, in red) are shown. According to miRNAs expression, there is no clustering between these two subgroups. B) The multidimensional scaling plot incorporates Lynch-mutated (blue) vs Lynch-like (red) subgroups, revealing that both are grouped together and they are undistinguishable based on the miRNA profile. C) The heat map represents expression profiles of the 31 miRNAs that best discriminate between sporadic MSI and Lynch syndrome, identified by PAM prediction. D) The multidimensional scaling plot includes Lynch syndrome (blue) and sporadic MSI tumor (red) samples, showing a significant separation between both groups. Fails: misclassified.

Figure 3. In situ hybridization (ISH) analysis of miR-622 in normal colorectal mucosa and CRC. miR-622, positive control (U6) and negative control (no probe) ISH analyses were performed in normal colorectal mucosa (A) and a group of CRCs (B). Staining for miR-622 was observed in the epithelium throughout the colonic crypt, but there was no staining in the stromal cells. There was a marked overexpression of mir-622 in the CRC samples compared to the normal colorectal mucosa.

Figure 4. Performance of the miRNA-based predictor to distinguish the type of MSI. A) Receiver operating curve of the miRNA-based predictor (miR-622, miR-362-5p, miR-486-5p) to distinguish the presence of Lynch syndrome among MSI tumors. S1, training set (n=14); S2, test set (n=33). B) Discriminant probability plot. The graphical representation shows the LOO-CV probabilities (0.0 to 1.0) of each tumor for being sporadic MSI (red dots and triangles) or Lynch syndrome (blue dots and triangles). Dots indicate samples from the training set (set 1) and triangles from the test set (set 2).
References


Table 1. Clinico-pathological characteristics of patients included in the study.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>N-C patients (n=20)</th>
<th>Sporadic MSI (n=33)</th>
<th>Lynch syndrome (n=35)</th>
<th>p value&lt;sup&gt;a&lt;/sup&gt;</th>
<th>MSS (n=19)</th>
<th>p value&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (+standard deviation)</td>
<td>64.4 (15.8)</td>
<td>65.3 (12.6)</td>
<td>47.5 (11.7)</td>
<td>&lt;0.0001</td>
<td>67.1 (12.0)</td>
<td>0.629</td>
</tr>
<tr>
<td>Sex, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Males</td>
<td>10 (50)</td>
<td>16 (48.5)</td>
<td>19 (54.3)</td>
<td>0.808</td>
<td>8 (42.1)</td>
<td>0.775</td>
</tr>
<tr>
<td>- Females</td>
<td>10 (50)</td>
<td>17 (51.5)</td>
<td>16 (45.7)</td>
<td></td>
<td>11 (57.9)</td>
<td></td>
</tr>
<tr>
<td>Tumor location*, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Proximal</td>
<td>26 (78.8)</td>
<td>20 (57.1)</td>
<td>0.146</td>
<td>5 (26.3)</td>
<td>&lt;0.0001</td>
<td></td>
</tr>
<tr>
<td>- Distal</td>
<td>5 (15.2)</td>
<td>11 (31.4)</td>
<td></td>
<td>14 (73.7)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Unknown</td>
<td>2 (6)</td>
<td>4 (11.4)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tumor stage, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- I</td>
<td>2 (6)</td>
<td>8 (22.9)</td>
<td>0.065</td>
<td>3 (15.8)</td>
<td>0.334</td>
<td></td>
</tr>
<tr>
<td>- II</td>
<td>15 (45.5)</td>
<td>11 (31.4)</td>
<td></td>
<td>9 (47.4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- III</td>
<td>7 (21.2)</td>
<td>9 (25.7)</td>
<td></td>
<td>6 (31.6)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- IV</td>
<td>7 (21.2)</td>
<td>2 (6.7)</td>
<td></td>
<td>1 (5.2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Unknown</td>
<td>2 (6)</td>
<td>5 (14.3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MMR protein expression, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Loss of MLH1</td>
<td>31 (93)</td>
<td>17 (48.6)</td>
<td>0.0001</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>- Loss of MSH2</td>
<td>-</td>
<td>14 (40)</td>
<td></td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>- Loss of MSH6</td>
<td>-</td>
<td>4 (11.4)</td>
<td></td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>- Normal expression</td>
<td>2 (7)</td>
<td>-</td>
<td></td>
<td>19 (100)</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

* Relative to the splenic flexure.
<sup>a</sup> p-value for the comparison of sporadic MSI vs Lynch syndrome
<sup>b</sup> p-value for the comparison of sporadic MSI vs MSS
Figure 2.

A

B

C

D

Lynch syndrome Fails Sporadic MSI

Lynch-mutated
Lynch-like

+5
-5

Lynch syndrome Sporadic MSI Fails

hsa-miR-421
hsa-miR-130b
hsa-miR-29b-1
hsa-miR-139-5p
hsa-miR-362-5p
hsa-miR-193a*
hsa-miR-214*
hsa-miR-432
hsa-miR-151-3.1
hsa-miR-92b*
hsa-miR-128
hsa-miR-486-5p
hsa-miR-30a*
hsa-miR-642
hsa-miR-16-2
hsa-miR-501-3p
hsa-miR-361-3p
hsa-miR-502-3p, hsa-miR-500*
hsa-miR-337-3p
hsa-miR-130b*
hsa-miR-598
hsa-miR-29b-2
hsa-miR-411
hsa-miR-142-5p
hsa-miR-269*
hsa-miR-628-3p
hsa-miR-34a*
hsa-miR-200c*
hsa-miR-187
hsa-miR-1238

D1
D2
D3

Lynch syndrome Sporadic MSI Fails
Figure 3.

A) Normal colonic mucosa

B) Adenocarcinoma
Figure 4.

A

B

LOO-CV probabilities

Sensitivity

Specificity

-2

0.0

0.2

0.4

0.6

0.8

1.0

-2

0.0

0.2

0.4

0.6

0.8

1.0

Lynch (test 1)

MSII (test 1)

Lynch (test 2)

MSII (test 2)
Colorectal cancers with microsatellite instability display unique miRNA profiles

Francesc Balaguer, Leticia Moreira, Juan Jose Lozano, et al.

Clin Cancer Res  Published OnlineFirst August 15, 2011.