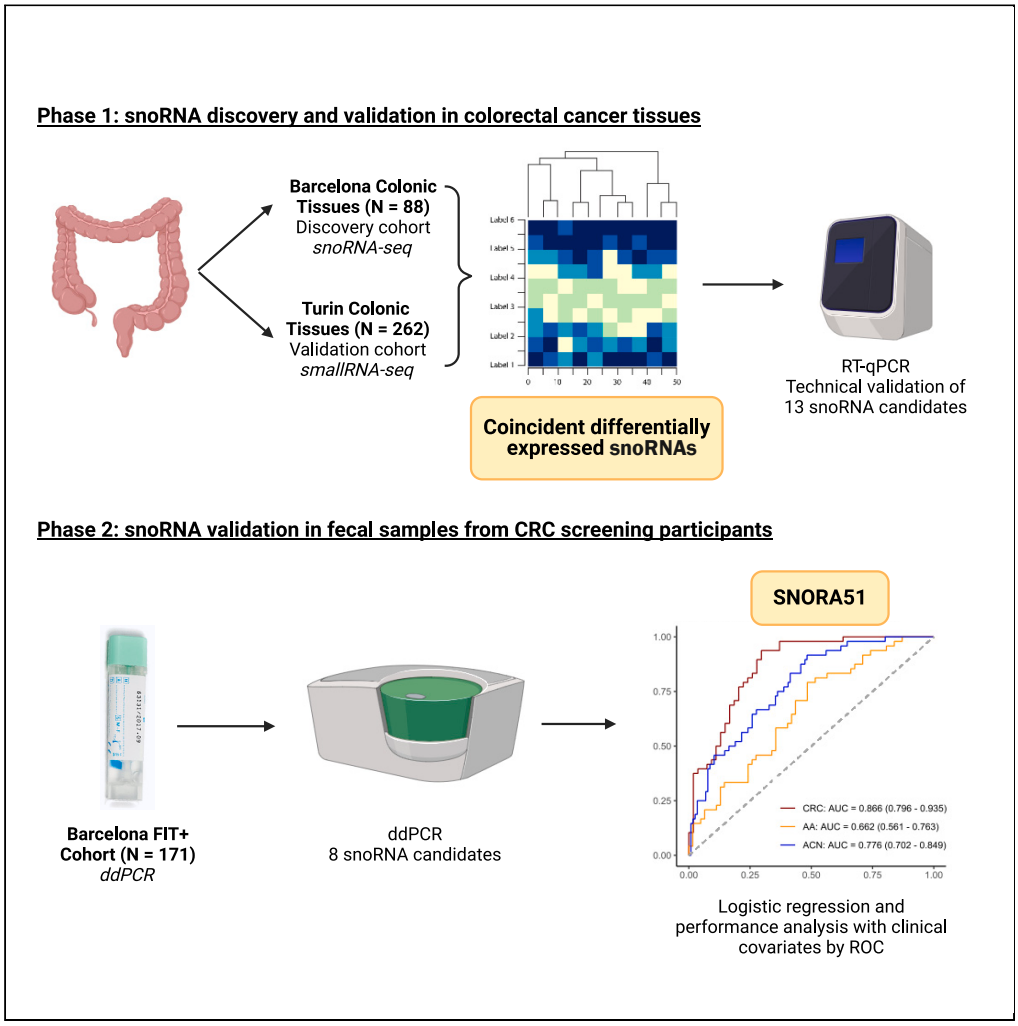
iScience

CellPress
OPEN ACCESS

Article

SnoRNA profiling in colorectal cancer and assessment of non-invasive biomarker capacity by ddPCR in fecal samples



Javier Gómez-Matas, Saray Duran-Sanchon, Juan-José Lozano, ..., Alessio Naccarati, Antoni Castells, Meritxell Gironella

meritxell.gironella@ciberehd.org

Highlights

snoRNA profiles in CRC or AA are different from those of normal mucosa

snoRNA profiles are stable between populations from different countries

snoRNAs can also be detected in fecal samples by ddPCR

SNORA51 can detect CRC cases among FIT+ participants in a CRC screening program

Gómez-Matas et al., iScience 27, 109283
March 15, 2024
https://doi.org/10.1016/j.isci.2024.109283

iScience

CellPress
OPEN ACCESS

Article

SnoRNA profiling in colorectal cancer and assessment of non-invasive biomarker capacity by ddPCR in fecal samples

Javier Gómez-Matas,[1] Saray Duran-Sanchon,[1] Juan-José Lozano,[2] Giulio Ferrero,[3] Sonia Tarallo,[4,5] Barbara Pardini,[4,5] Alessio Naccarati,[4,5] Antoni Castells,[1] and Meritxell Gironella[1,6,7,*]

SUMMARY

Small nucleolar RNAs (snoRNAs) have been identified dysregulated in several pathologies, and these alterations can be detected in tissues and in circulation. The main aim of this study was to analyze the whole snoRNome in advanced colorectal neoplasms and to identify new potential non-invasive snoRNA-based biomarkers in fecal samples by different analytical approaches. SNORA51, SNORD15B, SNORA54, SNORD12B, SNORD12C, SNORD72, SNORD89, and several members of SNORD115 and SNORD116 clusters were consistently deregulated in both tissue sets. After technical validation, SNORA51 and SNORD15B were detected in FIT+ samples. SNORA51 was significantly upregulated in FIT+ samples from CRC patients compared to healthy controls. This upregulation, together with the fecal hemoglobin concentration, was sufficient to identify, among FIT+ individuals, patients with CRC (AUC = 0.86) and individuals with advanced adenomas (AUC = 0.68). These findings portray snoRNAs as an alternative source of candidates for further studies and SNORA51 appears as a potential non-invasive biomarker for CRC detection.

INTRODUCTION

Colorectal cancer (CRC) is the third most common neoplasm in developed countries, and more than 500,000 cases are expected to be diagnosed each year in the European Union (EU). This neoplasm presents an elevated mortality rate, being the second most common cause of cancer-related death in the EU for both sexes.[1] This elevated mortality rate could potentially be reduced with an early diagnosis, being a cost-effective strategy in healthcare systems.[2,3] CRC stage at diagnosis is a crucial factor for patients' survival: for patients diagnosed in early stages, the 5-year survival is around 90%, whereas in more advanced or metastatic cases overall survival drops below 10%.[2] Currently, colonoscopy is the gold standard diagnostic tool for CRC.[4] However, this is an invasive procedure that carries potential complications, as well as an elevated cost for the healthcare system and low patient adherence, which, in time, hampers its application in population screening programs.[5,6]

To overcome this situation, several non-invasive methods based on the detection of occult blood in stool have been developed with remarkable results in the reduction of CRC-associated mortality[5,7]; among these, there are the guaiac fecal occult blood test, the fecal immunochemical test (FIT), and the multi-target stool DNA test. Besides their high potentialities, these non-invasive tests present several drawbacks, among which: (i) a limited sensitivity in the detection of pre-malignant lesions – advanced adenomas (AAs), (ii) a different performance depending on the location of the lesion (i.e., proximal-distal), and (iii) a relatively high number of false positive results which imply an excess of unnecessary colonoscopies.[8,9] Taking all aforementioned aspects into account, the ideal approach for large-population CRC screenings should be based on non-invasive tests, not representing a burden to the patients, with high sensitivity and specificity, able to detect pre-malignant lesions with a high risk of tumor development, highly reproducible, and, importantly, cost-effective.

In this regard, during the past few years, there has been increasing evidence that microRNAs (miRNAs) have the potential to function as CRC screening biomarkers, with several studies on both fecal[10,11] and plasma/serum samples.[12–14] However, little is known about other types of small non-coding RNAs in this setting. Small nucleolar RNAs (snoRNAs) are a family of small non-coding RNAs that perform a wide variety of functions inside and outside the cell. Generally, the size of snoRNAs ranges between 60 and 300 nucleotides long and their canonical function

[1]Gastrointestinal & Pancreatic Oncology Group, Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBERehd)/ Hospital Clínic Barcelona/ Fundació de Recerca Clínic Barcelona – Institut d'investigacions Biomèdiques August Pi i Sunyer, University of Barcelona, Barcelona, Catalonia, Spain
[2]Bioinformatics Platform, CIBERehd, Barcelona, Catalonia, Spain
[3]Department of Clinical and Biological Sciences, University of Turin, Turin, Italy
[4]Italian Institute for Genomic Medicine (IIGM), c/o IRCCS Candiolo, Turin, Italy
[5]Candiolo Cancer Institute, FPO-IRCCS, Candiolo, Turin, Italy
[6]Department of Experimental Pathology, Institute of Biomedical Research of Barcelona-Spanish National Research Council (IIBB-CSIC), Barcelona, Spain
[7]Lead contact
*Correspondence: meritxell.gironella@ciberehd.org
https://doi.org/10.1016/j.isci.2024.109283

Table 1. Clinicopathological characteristics of all the patients included in the study

	BCT cohort (NGS) (n = 44)		TCT cohort (NGS) (n = 131)		BFIT cohort (ddPCR) (n = 171)		
	AA (n = 21)	CRC (n = 23)	AA (n = 26)	CRC (n = 105)	Healthy (n = 52)	AA (n = 63)	CRC (n = 56)
Mean age (SD)	58 (6.0)	73 (11.6)	69.9 (10.9)	70.0 (10.5)	59.1 (5.7)	60.5 (5.9)	62.6 (6.6)
Gender no.							
Male	12	10	15	60	25	33	32
Female	9	13	11	45	27	30	24
Colorectal features							
TNM pathological stage							
I	–	7	–	26	–	–	17
II	–	9	–	33	–	–	13
III	–	6	–	35	–	–	14
IV	–	1	–	5	–	–	4
Unknown	–	0	–	6	–	–	8
Proximal/Distal							
Proximal	–	12	–	38 (2)	–	–	11
Distal	–	11	–	64 (2)	–	–	38
AA features							
Mean size (mm) (SD)	14 (3.8)	–	NA	–	–	14.6 (7.2)	–
no. AA Mean (SD)	3.9 (3.1)	–	NA	–	–	3.8 (3.3)	–
Stage 0	19	–	0	–	–	0	–
High - grade dysplasia no.	2	–	11	–	–	13	–
Villous component no.	4	–	0	–	–	18	–
Size $\geq$ 10 mm	21	–	12	–	–	7	–

BCT, Barcelona Colonic Tissue cohort; TCT, Turin Colonic Tissue cohort; BFIT, Barcelona FIT+ cohort, NA, not available. Numbers in parentheses in proximal/distal classification refers to the number of patients with lesions in both locations.

is the guidance of 2′-*O*-methylation (C/D-box snoRNAs) and pseudouridylation (H/ACA-box snoRNAs) of specific nucleotides when forming a complex with ribonucleoproteins.[15] Recently, it has been suggested that they may also perform other functions, such as regulating the alternative splicing of certain genes by targeting their mRNA[16] or generating snoRNA-derived miRNAs.[17,18] Their role in several malignancies, including cancer, has been previously described.[19,20] SnoRNAs can also be detected as free-circulating molecules, therefore they have been proposed as potential non-invasive biomarkers for several types of cancer, such as non-small cell lung cancer[21] and clear cell renal cell carcinoma.[22] Although these encouraging data exist, their behavior and use as biomarkers in the context of CRC have only been slightly explored. Several studies have observed a deregulation of some snoRNAs in CRC tissues[23–28] but there is still no evidence of their behavior in AAs.

The first aim of this study was to assess the whole snoRNA expression profile in CRC and AA tissues, followed by the identification of potential non-invasive snoRNA-based biomarkers in fecal samples by using different state-of-the-art technical approaches. We therefore performed a targeted RNA sequencing (RNA-seq) to profile snoRNAs in a set of paired tissue samples of CRC or AA and their matched healthy mucosa (CRC-Adj or AA-Adj, respectively) in a discovery setting approach. These results were compared to an external set of colonic tissue pairs, following validation of the most relevant results by reverse-transcription quantitative PCR (RT-qPCR). Finally, some snoRNA candidates were selected and analyzed in feces from FIT-positive participants of a CRC screening program by the ultrasensitive technique droplet digital PCR (ddPCR).

RESULTS

Genome-wide snoRNA profiling in colorectal neoplastic tissue by snoRNA-seq

A total of 44 pairs of CRC or AA tissue samples from the Barcelona Colonic Tissue (BCT) cohort (Table 1) were profiled using small RNA-seq specific for snoRNAs (snoRNA-seq). The obtained sequences resulted in 31,984 transcripts detected with at least one count in any sample, from which 1.25% were classified as snoRNAs. These transcripts were further filtered, only keeping those with a mean of counts higher than ten. In total, 252 snoRNA transcripts passed the filter and were kept for downstream analyses.

The exploratory analysis by principal components showed that snoRNA counts were able to distinguish CRC and AA tissues from their healthy matched mucosa and between them as well, forming well-defined clusters of samples. CRC tissues seemed to be more

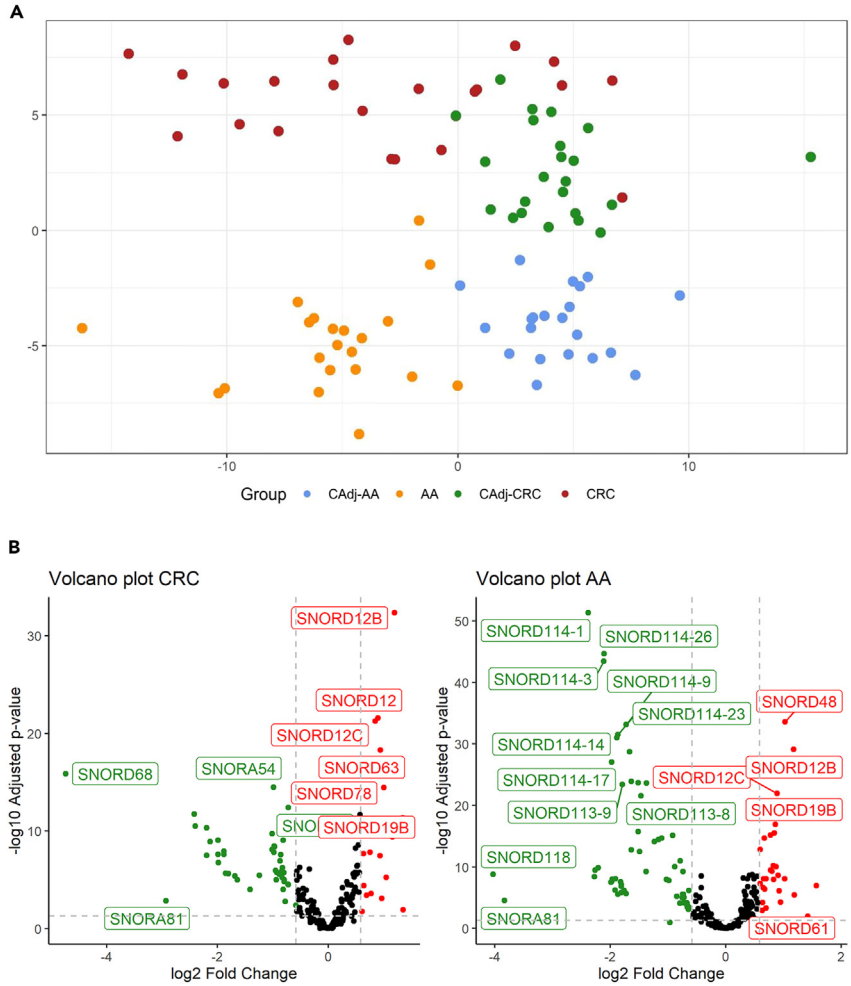


Figure 1. Descriptive analysis based on snoRNA expression of neoplasms from the Barcelona Colonic Tissue cohort

(A) Principal component analysis plot depicting tissue sample clustering based on the snoRNA expression profile. CRC: colorectal cancer; AA: advanced adenoma; CAdj: matched healthy adjacent tissue.

(B) Volcano plots of snoRNAs differentially expressed in CRC (left) and AA (right). The green and red colors indicate, respectively, downregulated and upregulated snoRNAs considering |FC| > 1.5 and FDR <0.05 as threshold values.

heterogeneous as compared to normal ones, whereas AA samples did not have such high variation and clustered closer and in a more regular manner (Figure 1A).

A paired analysis of the snoRNA expression of CRC and AA with respect to their matched adjacent mucosa was performed (Figure 1B). In total, 147 significantly deregulated snoRNAs were detected when comparing CRC samples to healthy adjacent tissues (FDR < 0.05), 68 of which were upregulated (FC > 0) whereas 81 downregulated (FC < 0). In AA, 183 deregulated snoRNAs were identified, 90 upregulated and 93 downregulated (Figure 2A). A total of 123 snoRNAs were commonly deregulated in both CRC and AA, but five of them presented an opposite trend of expression in CRC and AA. Regarding each individual group, 29 snoRNAs were uniquely deregulated in CRC (13 upregulated and 16 downregulated) and 65 in AA (37 upregulated and 28 downregulated) (Figure 2A, and Table S1).

Interestingly, after filtering for the top deregulated snoRNAs (|FC| > 1.5), their expression profiling was able to cluster samples into two main groups by using unsupervised hierarchical clustering. Concordantly, those clusters mostly belonged to advanced colorectal neoplasia (CRC or AA) and healthy adjacent tissues (Figure 2B). An overall trend toward snoRNA downregulation was found in comparison to their control counterparts. Interestingly, 40 out of 81 and 56 out of 93 were downregulated snoRNA transcripts, respectively, in CRC and in AA that constituted tandem repeats of the same snoRNA cluster (SNORD115, SNORD116, SNORD113, or SNORD114) (data not shown). After considering these tandem repeats as belonging to the same snoRNA, the overall situation regarding the number of deregulated snoRNAs shifted toward a general snoRNAs upregulation in colorectal neoplasms. Finally, the predictive capacity of these snoRNAs was individually measured with a logistic regression model for either CRC or AA against healthy tissues to select those with higher potential of being useful as diagnostic biomarkers (Table S1). Overall, the mean area under the curve (AUC) for significative snoRNAs in CRC was 0.756 $\pm$ 0.11, and 0.814 $\pm$ 0.13 in

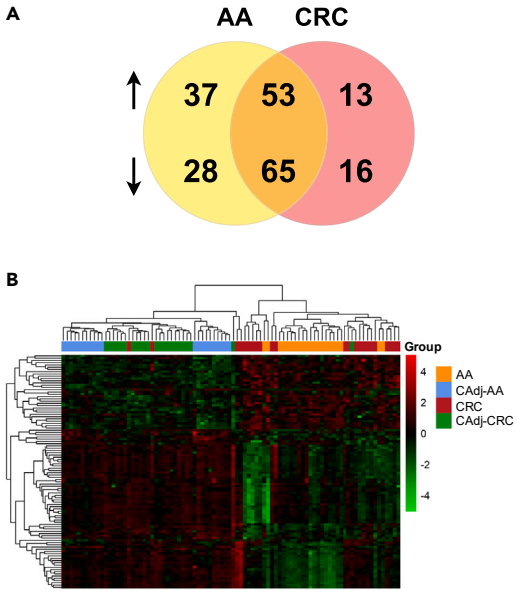


Figure 2. Overview of deregulated snoRNAs in neoplasms from the Barcelona Colonic Tissue cohort

(A) Venn diagram reporting the total number of deregulated snoRNAs in AA and CRC derived from snoRNA-seq results of 88 paired neoplastic and healthy adjacent colonic tissues (BCT). The number indicates the differentially expressed snoRNAs with significant FDR <0.05. The intersection represents the deregulated snoRNAs in common in both groups. Top numbers: upregulated snoRNAs. Bottom numbers: downregulated snoRNAs.

(B) Expression heatmap of the top 100 significantly deregulated snoRNAs for CRC or AA when compared with their normal matched mucosa (n = 88) (FDR <0.05 and |FC|>1.5). Green pixels indicate downregulation whereas red indicates upregulation.

AA. The top three snoRNAs with highest AUCs in CRC were SNORD12B, SNORA54, and SNORD12C, whereas in AA were SNORD113, SNORD114, and SNORD12C.

Validation of altered snoRNAs in colorectal neoplasia in an independent cohort

To validate previous findings, altered snoRNA profiles were compared with those from an independent cohort (Turin Colonic Tissue cohort [TCT]) composed by CRC and AA patients (Table 1). Comparing all deregulated snoRNA-labeled transcripts with a significance level (FDR <0.05) and commonly upregulated or downregulated in both cohorts, a total of 85 coincidences were found, 45 in AA and 63 in CRC (see Table S2). Nevertheless, it is worth mentioning that some top significantly deregulated snoRNAs in the BCT cohort (i.e., SNORD19B, SNORD12, or SNORD78) were not assessable in the TCT cohort since there were no detectable reads; this may be due to the different methodologies used in the NGS procedure for this second set of samples.

Taken together all previous results, we selected 14 deregulated snoRNAs to be confirmed in a subset of tissue samples from the BCT cohort by another technology (RT-qPCR): SNORA51, SNORA54, SNORD114, SNORD115, SNORD116, SNORD12, SNORD12B, SNORD12C, SNORD15B, SNORD19B, SNORD63, SNORD72, SNORD78, and SNORD89.

Quantification of selected snoRNAs in colonic tissue by RT-qPCR

The expression of the 14 previously selected snoRNAs was analyzed by RT-qPCR in the subset of BCT cohort (n = 60, 15 paired CRC-C-Adj, 15 paired AA-C-Adj) as a technical validation. Differences in mean expression between groups were compared. Upregulation of SNORD12C, SNORD12B, SNORD78, SNORA51, and SNORD15B and downregulation of SNORD115 were validated in both CRC and AA tissue samples. Moreover, SNORD19B, SNORD12, SNORD63, and SNORD72 showed upregulation in AA but not in CRC. Finally, SNORD114 was not detected and, therefore, it was discarded for further analyses (Figure 3).

SnoRNA detection in fecal samples by ddPCR

In order to see if any of the aforementioned snoRNAs could be detected in a non-invasive sample such as feces and could serve as a potential biomarker for CRC detection, some of the upregulated snoRNAs were selected (SNORD12B, SNORC12C, SNORD72, SNORD78, SNORD12, SNORD19B, SNORA51, and SNORD15B) as candidates to be analyzed by ultrasensitive ddPCR in fecal samples from FIT+ leftover of CRC screening participants.

Out of the eight selected snoRNAs, SNORD12B, SNORD12C, SNORD12, SNORD19B, SNORD78, and SNORD72 could not be detected in feces by ddPCR in preliminary tests and thus were discarded from further analyses. On the other hand, the expression of SNORA51 and SNORD15B was further tested in the whole Barcelona FIT cohort (56 CRC, 63 AA, 52 Controls). Importantly, SNORA51 was detected in 164 out of the 171 fecal samples and showed a significant overexpression in CRC patients in comparison to healthy control individuals, whereas no significant differences were found between AA patients and healthy controls (Figure 4A). On the other hand, fecal SNORD15B expression was detected in 162 out of the 171 fecal samples but did not show a significant upregulation in CRC patients (Figure 4B), and, unexpectedly, was downregulated in AA when compared to control samples in contrast with what we observed in colonic tissues. There were no differences in either CRC location of the lesion (Figure 4C) or pathological stage (Figure 4D).

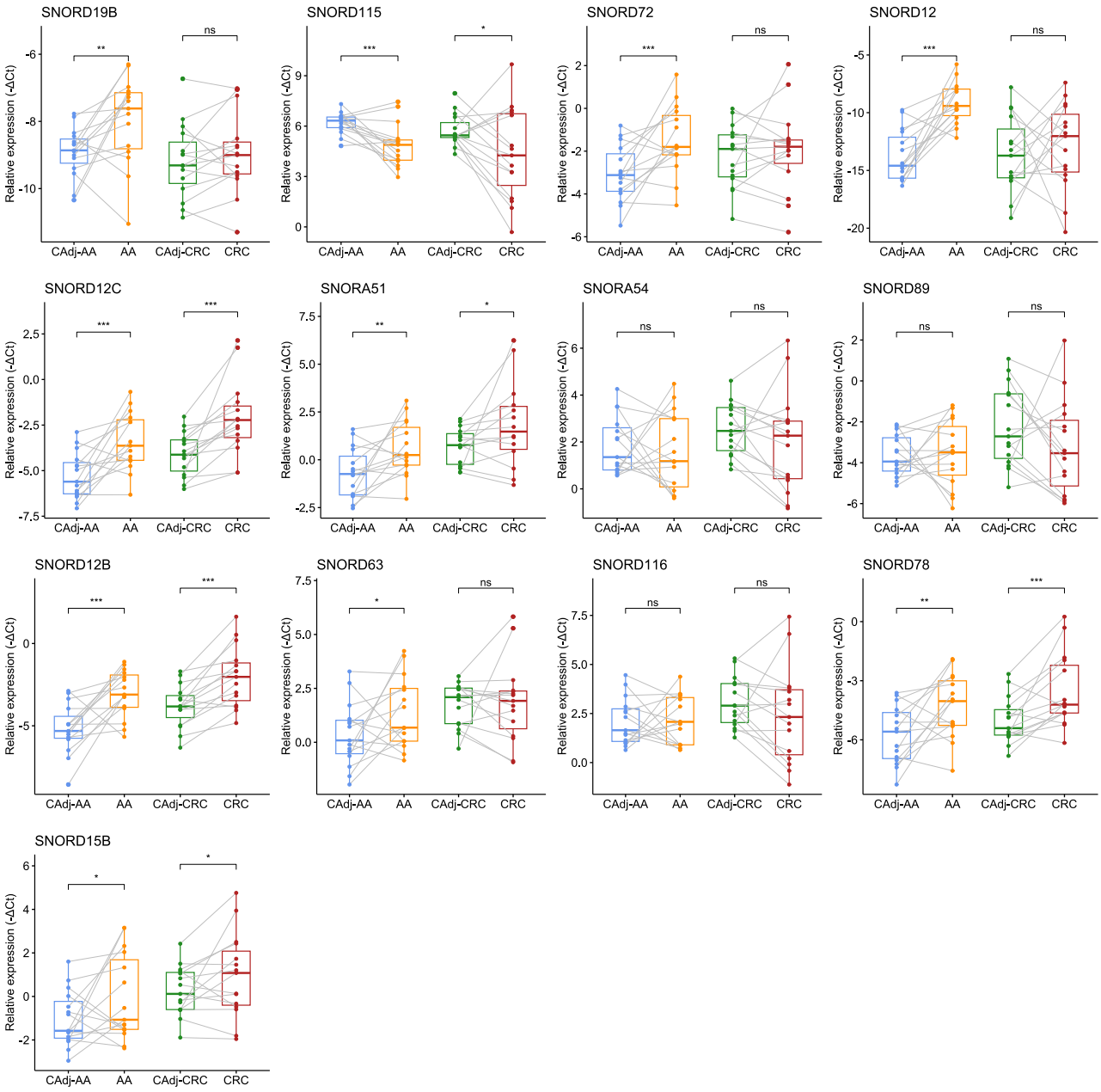


Figure 3. Expression levels of the selected snoRNAs as analyzed by RT-qPCR in the tissue subset from the BCT cohort for technical validation (n = 60)
Comparison between groups calculated by paired t test (ns: not significant: *p ≤ 0.1; **p ≤ 0.05; ***p ≤ 0.01).

Last, we measured the predictive capacity of SNORA51 by ROC analysis using a logistic regression model and adjusting for age and sex as covariates. The model was trained using 5 times 10-fold cross validation in a fraction of the main dataset. By using the concentration of this snoRNA as a predictor, we observed an AUC of 0.725 (CI 95% = 0.627–0.822) for CRC, 0.569 (CI 95% = 0.460–0.678) for AA, and 0.652 (CI 95% = 0.562–0.742) for the detection of advanced colorectal neoplasia (either CRC or AA) (Figure 4E, and Table 2). When adding fecal hemoglobin levels to the model, the discriminative capacity improved for all the three conditions (Figure 4F, and Table 2), reaching a specificity to detect CRC patients of 94% with a sensitivity of 70% among FIT-positive individuals participating in a CRC screening program.

DISCUSSION

In the last years, many efforts have been directed to study the expression profiles of miRNAs in different types of cancer, including CRC. However, there are also other types of small non-coding RNAs, such as snoRNAs, that have been little explored in this context. In this study, we

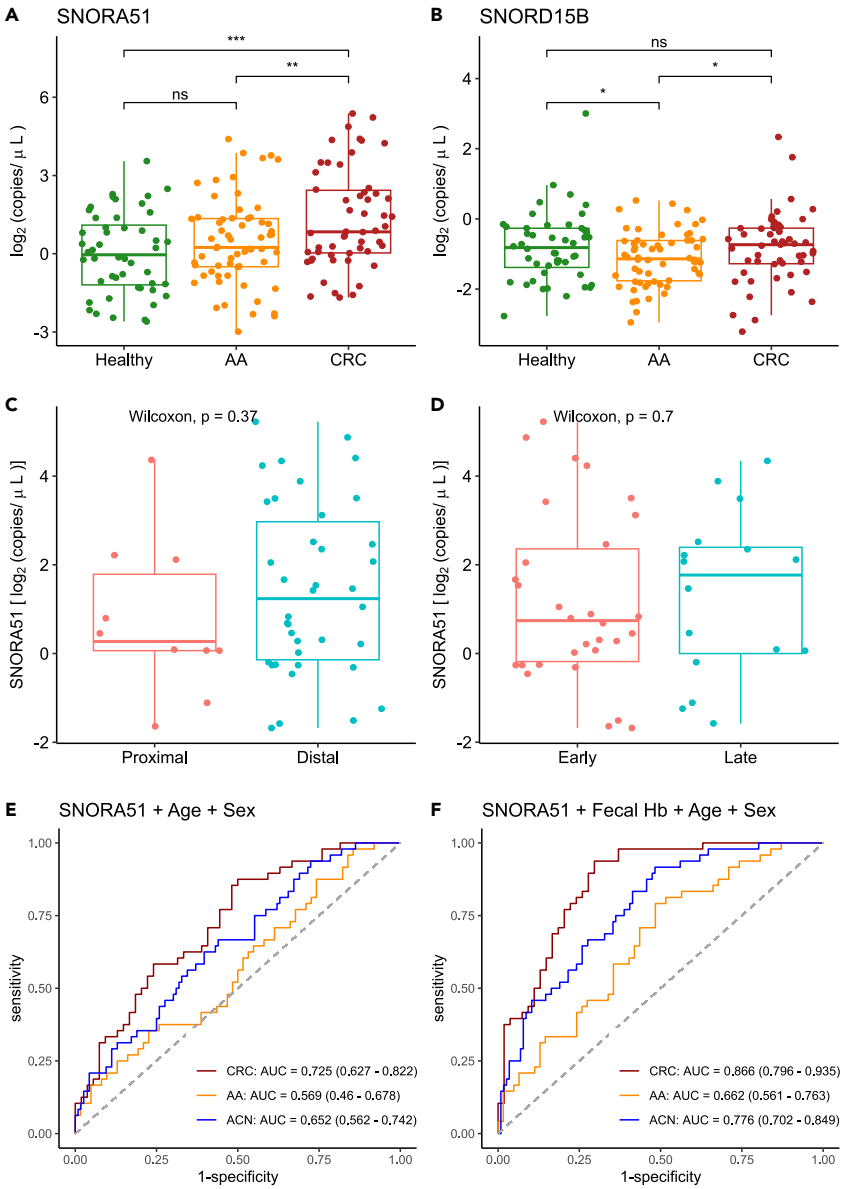


Figure 4. Evaluation of selected snoRNA candidates in fecal samples

(A and B) Levels of fecal SNORA51 and SNORD15B as obtained by ddPCR in FIT leftover from FIT-positive individuals (n = 171) participating in a CRC screening program. Differences in expression between groups were calculated by t test.

(C and D) Expression of fecal SNORA51 by CRC location and pathological stage. Differences in expression were calculated by Wilcoxon signed-rank test. Early stages: I and II, late stages: III and IV.

(E and F) ROC curves of the logistic regression model for SNORA51 and SNORA51 + fecal hemoglobin levels (fecal Hb) from FIT-positive samples, adjusted by age and sex. ACN, advanced colorectal neoplasia; CRC, colorectal cancer; AA, advanced adenoma; AUC, area under the ROC curve (ns, not significant; *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$).

profiled the whole snoRNome in CRC tumors by performing a targeted snoRNA-sequencing analysis in a set of paired colonic tissues from Spain (the BCT cohort). Moreover, in this analysis, we also included premalignant lesions such as AAs, in which snoRNAs were poorly characterized. Significantly altered snoRNAs in CRC and AA were also tested in an independent cohort of CRC and AA tissues from Italy. A substantial amount of snoRNAs were consistently found upregulated or downregulated in both cohorts. We therefore selected a set of significantly deregulated snoRNAs in both CRC and AA compared to their matched adjacent mucosa according to their fold change and diagnostic capability to be validated by RT-qPCR in a subset of tissue samples from the BCT cohort. Finally, a subset of snoRNAs was evaluated, for the first time, for their detectability in fecal samples from the FIT leftover of CRC screening participants using the ultrasensitive ddPCR technique.

Table 2. Performance characteristics of ddPCR levels of fecal SNORA51 to identify patients with CRC or AA

	AUC	AUC 95% CI	Sen	Spe	PPV	NPV
SNORA51 + Age + Sex						
CRC	0.725	(0.627–0.822)	0.648	0.938	0.921	0.703
AA	0.569	(0.46–0.678)	0.323	0.854	0.741	0.494
ACN	0.652	(0.562–0.742)	0.560	0.833	0.890	0.440
SNORA51 + Fecal Hb + Age + Sex						
CRC	0.866	(0.796–0.935)	0.704	0.938	0.927	0.738
AA	0.662	(0.561–0.763)	0.516	0.792	0.762	0.559
ACN	0.776	(0.702–0.849)	0.517	0.917	0.938	0.440

Adjusted by age and sex. Optimal cutpoint selection by Youden's J. AUC, Area under the curve; Spe, specificity; Sen, specificity; NPV, negative predictive value; PPV, positive predictive value.

Studies based on other types of small non-coding RNAs different from miRNAs, such as the snoRNAs, recently have gained more attention, mainly because these molecules could serve as a new source of biomarkers for several conditions. Previous studies have demonstrated the presence of deregulated snoRNAs in CRC and other pathologies. For instance, Xu et al.[23] found an upregulation of SNORD12B, SNORD12C, SNORD78, and SNORD19B in colorectal tissues that was also confirmed in our study. Okugawa et al.[27] found SNORA42, SNORD76, SNORD78, and ACA11 to be upregulated in CRC, with SNORD78 that was consistently upregulated also in the BCT cohort. The same authors also found an association of SNORA42 with poor prognosis and distant metastases, as well as promoting tumor growth in mice models. Moreover, SNORD15B and SNORA5C have been recently found upregulated in CRC tissues by Shen and colleagues[28]: interestingly, the authors performed a set of functional studies to uncover their biological role in this malignancy. The levels of these two snoRNAs were also associated with poor prognosis not only in CRC but also in other cancers. Hereby, we also found a slight upregulation of SNORD15B but not SNORA5C, which curiously was significantly downregulated in AA. Furthermore, Shang and colleagues[22] found SNORD63 and SNORD96A deregulated in tissue, plasma, and urinary sediment in the context of clear cell renal cell carcinoma: these deregulated snoRNAs were used to implement a model in body fluids capable of discriminating cases. SNORD78 and SNORD66 were reported to be upregulated in sputum of non-small cell lung cancer patients,[21] which established a model able to detect cases from healthy patients using the expression of these two snoRNAs.

Surprisingly, in our study, when comparing AA or CRC tissues with their healthy matched counterparts, we found a significant downregulation of most of the members of four snoRNA clusters. These snoRNAs (SNORD113, SNORD114, SNORD115, and SNORD116) are encoded in regions affected by parental imprinting, such as the 14q32 (DLK1-DIO3: SNORD113, SNORD114) and 15q11 (SNURF-SNRPN: SNORD115, SNORD116).[16] Interestingly, those snoRNAs are located in a specific region in which they form arrays of tandemly repeated copies of each other, giving place, for instance to 48 different SNORD115 and 29 different SNORD116. Changes in DNA methylation in these loci in CRC might alter the expression of these snoRNAs and explain the fact that they are found widely downregulated in CRC and AA compared to healthy mucosa.[29–32] Another possible explanation of these observations is the loss of the 15q11 and 14q32 regions in CRC, which is a common event in the majority of sporadic CRC.[32,33] Loss of 14q32 has also been reported in CRC and has been related to poor outcome and metastases.[34]

To further validate our findings, we compared the snoRNA-seq results from the BCT cohort to an external snoRNA dataset also obtained from CRC or AA tissues. This external dataset mainly came from the sequencing of the band used to study miRNA expression (18–30 nucleotides), which was smaller in size compared to the one used to profile snoRNAs in the BCT cohort. Overall, read counts related to snoRNAs in the second cohort were lower compared to the Barcelona discovery cohort, but data from literature suggested that small RNA-seq for miRNA profiling could be also useful as a surrogate of the snoRNA landscape without the need for a new sequencing experiment.[26] This might explain some of the differences found in deregulated snoRNAs between both cohorts and could tell why some of the tissue RT-qPCR-validated snoRNAs were not present in the external validation dataset. The same limitations may appear when comparing our data to other studies that analyze snoRNA expression based on miRNA-seq data: although the information per se is useful and most significant snoRNAs appear to be deregulated, we believe that it is more advisable to analyze each small non-coding RNA within the band where they are naturally located by size.[35]

Previous studies from our group and others[10,36] have demonstrated the presence of miRNAs in fecal samples coming from FIT leftover; however, to our knowledge, no other groups have assessed the presence of snoRNAs in fecal samples from CRC and AA patients, so far. The failure to detect some candidate snoRNAs in fecal samples coming from FIT leftover may rely on different aspects. On the one hand, there is not enough knowledge yet about the presence of this type of small non-coding RNAs in liquid biopsies, as well as about their biology and behavior. However, we have observed that snoRNA abundance in tissue is lower than that of miRNAs, and, consequently, the probability of being able to detect some levels in FIT leftover sample is even lower. On the other hand, we must consider the small amount of starting feces from the FIT remnant and the variability of preanalytical factors that may influence the results in these liquid biopsies. Despite these limitations, the fact that some snoRNAs could be reliably detected in FIT samples, being able to discriminate CRC patients from healthy subjects is no small feat.

SNORA51, upregulated in CRC and AA tissues, was also upregulated in FIT samples from CRC patients compared to healthy controls. This upregulation, together with the fecal hemoglobin concentration was sufficient to identify not only patients with CRC but also to some degree individuals with AA (AUCs for CRC and AA were 0.84 and 0.68, respectively). SNORA51 is an H/ACA-box snoRNA encoded as an intron of the NOP56 gene, a core component of the ribonucleoprotein complex for C/D-box snoRNAs.[37] This result points to SNORA51 as a potential non-invasive biomarker for CRC detection. To our knowledge, little is known about this particular snoRNA, thus making it an interesting candidate for further studies.

Among other snoRNAs that were not finally detected in FIT leftovers, we also selected SNORD15B because it was slightly upregulated in tumoral tissue from the Barcelona and Turin cohorts and because it was previously found upregulated in the study of Shen and colleagues.[28] In that study, authors found SNORD15B to be upregulated in CRC tissues and they studied its oncogenic potential in CRC cell lines, suggesting a role of this snoRNA in colorectal oncogenesis. However, although we could detect some fecal SNORD15B levels in FIT samples, we could not confirm a significant upregulation in CRC patients, and unexpectedly, it even showed a significant decrease in the AA group, in contrast to the increasing trend shown in tissue. Considering the overall low snoRNA count levels obtained in the NGS experiments, and the need of a preamplification step for most of them in the RT-qPCR tissue validation phase, we decided to establish an optimized protocol to analyze fecal snoRNA candidates in FIT samples with a more sensitive technique such as ddPCR. ddPCR is able to absolute quantify snoRNAs even at very low concentrations, without preamplification, and without the need of technical replicates or an internal normalizer. Thus, with this technique, we were able to detect the aforementioned fecal snoRNAs in FIT samples with concentrations ranging from 0.12 to 41.6 copies per microliter for SNORA51 and 0.1 to 8 copies per microliter for SNORD15B.

In conclusion, the present study highlights the deregulation of some snoRNAs in CRC and AA at a tissue and fecal level. This suggests that snoRNAs may play a role in colorectal carcinogenesis and could be potentially used as novel biomarkers for early detection of CRC. The fact of finding some of them in FIT leftovers opens the possibility of further exploring these snoRNAs in liquid biopsies in this context. Regardless, this study sets a step further into new small non-coding RNA species besides miRNAs as possible candidate for early detection and a better characterization of CRC which, in time, may be helpful in the application in the CRC screening setting.

Limitations of the study

In this article, we demonstrate how snoRNAs, and more specifically SNORA51, are present in fecal samples derived from FIT and are able to discriminate patients with CRC from those with benign polyps or healthy individuals. However, this study could have benefited from a wider and more varied cohort of patients, as well as the inclusion of FIT-negative individuals. Thus, increasing the similarity of the study population to an actual CRC screening. Likewise, we only tested a limited set of snoRNA candidates that presented characteristics we deemed biologically relevant in tissues. That would not discard the possibility of other snoRNAs having a potential use as biomarkers for CRC in fecal samples. Despite these limitations, we conclude that SNORA51 can be detected in the leftover from FIT and how it could detect patients with CRC from other individuals among FIT-positive samples.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
  - RNA extraction and snoRNA sequencing
  - SnoRNA expression analysis by RT-qPCR
  - SnoRNA quantification by ddPCR on fecal samples
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2024.109283.

ACKNOWLEDGMENTS

We are indebted to the Fundació de Recerca Clínic Barcelona-Institut d'investigacions Biomèdiques August Pi i Sunyer Biobank, integrated in the Spanish National Biobank Network, for samples procurement. We also appreciate the collaboration of staff from Endoscopy Unit-Hospital Clinic of Barcelona for helping us to collect the samples. This work was developed at the Centro Esther Koplowitz, Barcelona, Catalonia, Spain. The present work was supported by Instituto de Salud Carlos III (PI20/01696 to M.G.), co-funded by FEDER-European Union and Ministerio de Economía y Competitividad (SAF2014-54453-R to A.C.). J.G.-M. is a recipient of an iPFIS predoctoral contract from Instituto de

Salud Carlos III and co-funded by Fondo Social Europeo (FSE) (IFI20/00025). We also acknowledge the support of CERCA Programme and Agència de Gestió d'Ajuts Universitaris i de Recerca (2021SGR716), Generalitat de Catalunya. CIBEREHD (Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas) is funded by the Instituto de Salud Carlos III. This project has been partially funded from the European Union's Horizon 2020 research and innovation program under grant agreement no. 825410 (ONCOBIOME project to A.N., B.P., and S.T.). Part of the research leading to these results has received funding from AIRC under IG 2020 - ID. 24882 – P.I. Naccarati Alessio Gordon (to A.N.).

AUTHOR CONTRIBUTIONS

Conceptualization, J.G.-M. and M.G.; methodology, J.G.-M., S.D.-S., J.-J.L., G.F., B.P., and A.N.; data curation, J.G.-M. and J.-J.L.; investigation, J.G.-M., S.D.-S., and G.F.; formal analysis, J.G.-M. and G.F.; validation, J.G.-M.; resources, M.G., S.D.-S., B.P., and A.N.; writing – original draft, J.G.-M. and M.G.; review and editing, S.D.-S., G.F., S.T., B.P., A.N., and A.C.; supervision, M.G.; project administration, M.G.; funding acquisition, M.G. and A.C. All authors have read, edited, and approved the final manuscript.

DECLARATION OF INTERESTS

There are no conflicts of interest to disclose.

Received: September 18, 2023
Revised: January 26, 2024
Accepted: February 16, 2024
Published: February 20, 2024

REFERENCES

1. Sung, H., Ferlay, J., Siegel, R.L., Laversanne, M., Soerjomataram, I., Jemal, A., and Bray, F. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA. Cancer J. Clin. *71*, 209–249. https://doi.org/10.3322/CAAC.21660.
2. Siegel, R.L., Miller, K.D., Goding Sauer, A., Fedewa, S.A., Butterly, L.F., Anderson, J.C., Cercek, A., Smith, R.A., and Jemal, A. (2020). Colorectal cancer statistics, 2020. CA. CA. Cancer J. Clin. *70*, 145–164. https://doi.org/10.3322/caac.21601.
3. Ladabaum, U., Dominitz, J.A., Kahi, C., and Schoen, R.E. (2020). Strategies for Colorectal Cancer Screening. Gastroenterology *158*, 418–432. https://doi.org/10.1053/j.gastro.2019.06.043.
4. Binefa, G., Rodríguez-Moranta, F., Teule, A., and Medina-Hayas, M. (2014). Colorectal cancer: From prevention to personalized medicine. World J. Gastroenterol. *20*, 6786–6808.
5. Ladabaum, U. (2020). Cost-Effectiveness of Current Colorectal Cancer Screening Tests. Gastrointest. Endosc. Clin. N. Am. *30*, 479–497. https://doi.org/10.1016/j.giec.2020.02.005.
6. Quirke, P., Risio, M., Lambert, R., Von Karsa, L., and Vieth, M.; International Agency for Research on Cancer (2012). European guidelines for quality assurance in colorectal cancer screening and diagnosis. First Edition Quality assurance in pathology in colorectal cancer screening and diagnosis. Endoscopy *44*, SE116–SE130. https://doi.org/10.1055/s-0032-1309797.
7. Zou, J., Xiao, Z., Wu, Y., Yang, J., and Cui, N. (2022). Noninvasive fecal testing for colorectal cancer. Clin. Chim. Acta *524*, 123–131. https://doi.org/10.1016/j.cca.2021.10.030.
8. Khalid-de Bakker, C.A.J., Jonkers, D.M.A.E., Sanduleanu, S., De Bruïne, A.P., Meijer, G.A., Janssen, J.B.M.J., Van Engeland, M., Stockbrügger, R.W., and Masclee, A.A.M. (2011). Test performance of immunologic fecal occult blood testing and sigmoidoscopy compared with primary colonoscopy screening for colorectal advanced adenomas. Cancer Prev. Res. *4*, 1563–1571. https://doi.org/10.1158/1940-6207.CAPR-11-0076/35785/AM/TEST-PERFORMANCE-OF-IMMUNOLOGICAL-FECAL-OCCULT.
9. Wong, M.C.S., Ching, J.Y.L., Chan, V.C.W., Lam, T.Y.T., Luk, A.K.C., Ng, S.S.M., and Sung, J.J.Y. (2015). Factors Associated with False-Positive and False-Negative Fecal Immunochemical Test Results for Colorectal Cancer Screening. Gastrointest Endosc. *81*, 596–607. https://doi.org/10.1016/j.gie.2014.08.006.
10. Duran-Sanchon, S., Moreno, L., Augé, J.M., Serra-Burriel, M., Cuatrecasas, M., Moreira, L., Martín, A., Serradesanferm, A., Pozo, À., Costa, R., et al. (2020). Identification and Validation of MicroRNA Profiles in Fecal Samples for Detection of Colorectal Cancer. Gastroenterology *158*, 947–957.e4. https://doi.org/10.1053/j.gastro.2019.10.005.
11. Pardini, B., Ferrero, G., Tarallo, S., Gallo, G., Francavilla, A., Licheri, N., Trompetto, M., Clerico, G., Senore, C., Peyre, S., et al. (2023). A fecal miRNA signature by small RNA sequencing accurately distinguishes colorectal cancers: results from a multicentric study. Gastroenterology *165*, 582–599.e8. https://doi.org/10.1053/j.gastro.2023.05.037.
12. Zanutto, S., Ciniselli, C.M., Belfiore, A., Lecchi, M., Masci, E., Delconte, G., Primignani, M., Tosetti, G., Dal Fante, M., Fazzini, L., et al. (2020). Plasma miRNA-based signatures in CRC screening programs. Int. J. Cancer *146*, 1164–1173. https://doi.org/10.1002/IJC.32573.
13. Giráldez, M.D., Lozano, J.J., Ramírez, G., Hijona, E., Bujanda, L., Castells, A., and Gironella, M. (2013). Circulating MicroRNAs as biomarkers of colorectal cancer: Results from a genome-wide profiling and validation study. Clin. Gastroenterol. Hepatol. *11*, 681–688.e3. https://doi.org/10.1016/j.cgh.2012.12.009.
14. Marcuello, M., Duran-Sanchon, S., Moreno, L., Lozano, J.J., Bujanda, L., Castells, A., and Gironella, M. (2019). Analysis of A 6-Mirna Signature in Serum from Colorectal Cancer Screening Participants as Non-Invasive Biomarkers for Advanced Adenoma and Colorectal Cancer Detection. Cancers *11*, 1542. https://doi.org/10.3390/CANCERS11101542.
15. Dupuis-Sandoval, F., Poirier, M., and Scott, M.S. (2015). The emerging landscape of small nucleolar RNAs in cell biology. WIREs RNA *6*, 381–397. https://doi.org/10.1002/wrna.1284.
16. Cavaillé, J. (2017). Box C/D small nucleolar RNA genes and the Prader-Willi syndrome: a complex interplay. Wiley Interdiscip. Rev. RNA *8*, e1417. https://doi.org/10.1002/wrna.1417.
17. Brameier, M., Herwig, A., Reinhardt, R., Walter, L., and Gruber, J. (2011). Human box C/D snoRNAs with miRNA like functions: expanding the range of regulatory RNAs. Nucleic Acids Res. *39*, 675–686. https://doi.org/10.1093/nar/gkq776.
18. Scott, M.S., and Ono, M. (2011). From snoRNA to miRNA: Dual function regulatory non-coding RNAs. Biochimie *93*, 1987–1992. https://doi.org/10.1016/J.BIOCHI.2011.05.026.
19. Deogharia, M., and Majumder, M. (2018). Guide snoRNAs: Drivers or Passengers in Human Disease? Biology *8*, 1. https://doi.org/10.3390/biology8010001.
20. McMahon, M., Contreras, A., and Ruggero, D. (2015). Small RNAs with big implications: new insights into H/ACA snoRNA function and their role in human disease. Wiley Interdiscip. Rev. RNA *6*, 173–189. https://doi.org/10.1002/wrna.1266.
21. Su, J., Liao, J., Gao, L., Shen, J., Guarnera, M.A., Zhan, M., Fang, H., Stass, S.A., and Jiang, F. (2016). Analysis of small nucleolar RNAs in sputum for lung cancer diagnosis.

Oncotarget *7*, 5131–5142. https://doi.org/10.18632/oncotarget.4219.

22. Shang, X., Song, X., Wang, K., Yu, M., Ding, S., Dong, X., Xie, L., and Song, X. (2021). SNORD63 and SNORD96A as the non-invasive diagnostic biomarkers for clear cell renal cell carcinoma. Cancer Cell Int. *21*, 56. https://doi.org/10.1186/s12935-020-01744-4.
23. Xu, L., Ziegelbauer, J., Wang, R., Wu, W.W., Shen, R.F., Juhl, H., Zhang, Y., and Rosenberg, A. (2016). Distinct Profiles for Mitochondrial t-RNAs and Small Nucleolar RNAs in locally invasive and metastatic colorectal cancer. Clin. Cancer Res. *22*, 773–784. https://doi.org/10.1158/1078-0432.CCR-15-0737.
24. Koduru, S.V., Tiwari, A.K., Hazard, S.W., Mahajan, M., and Ravnic, D.J. (2017). Exploration of small RNA-seq data for small non-coding RNAs in Human Colorectal Cancer. J. Genomics *5*, 16–31. https://doi.org/10.7150/JGEN.18856.
25. Yang, X., Li, Y., Li, L., Liu, J., Wu, M., and Ye, M. (2017). SnoRNAs Are Involved in the Progression of Ulcerative Colitis and Colorectal Cancer. https://doi.org/10.1016/j.dld.2016.12.029.
26. Yoshida, K., Toden, S., Weng, W., Shigeyasu, K., Miyoshi, J., Turner, J., Nagasaka, T., Ma, Y., Takayama, T., Fujiwara, T., and Goel, A. (2017). SNORA21 – An Oncogenic Small Nucleolar RNA, with a Prognostic Biomarker Potential in Human Colorectal Cancer. EBioMedicine *22*, 68–77. https://doi.org/10.1016/j.ebiom.2017.07.009.
27. Okugawa, Y., Toiyama, Y., Toden, S., Mitoma, H., Nagasaka, T., Tanaka, K., Inoue, Y., Kusunoki, M., Boland, C.R., and Goel, A. (2017). Clinical significance of SNORA42 as an oncogene and a prognostic biomarker in colorectal cancer. Gut *66*, 107–117. https://doi.org/10.1136/gutjnl-2015-309359.
28. Shen, L., Lu, W., Huang, Y., He, J., Wang, Q., Zheng, X., and Wang, Z. (2022). SNORD15B and SNORA5C: Novel Diagnostic and Prognostic Biomarkers for Colorectal Cancer. BioMed Res. Int. *2022*, 8260800–8260810. https://doi.org/10.1155/2022/8260800.
29. Okugawa, Y., Grady, W.M., and Goel, A. (2015). Epigenetic Alterations in Colorectal Cancer: Emerging Biomarkers. Gastroenterology *149*, 1204–1225.e12. https://doi.org/10.1053/J.GASTRO.2015.07.011.
30. Benetatos, L., Hatzimichael, E., Dasoula, A., Dranitsaris, G., Tsiara, S., Syrrou, M., Georgiou, I., and Bourantas, K.L. (2010). CpG methylation analysis of the MEG3 and SNRPN imprinted genes in acute myeloid leukemia and myelodysplastic syndromes. Leuk. Res. *34*, 148–153. https://doi.org/10.1016/j.leukres.2009.06.019.
31. Sepulveda, J.L., Gutierrez-Pajares, J.L., Luna, A., Yao, Y., Tobias, J.W., Thomas, S., Woo, Y., Giorgi, F., Komissarova, E.V., Califano, A., et al. (2016). High-definition CpG methylation of novel genes in gastric carcinogenesis identified by next-generation sequencing. Mod. Pathol. *29*, 182–193. https://doi.org/10.1038/MODPATHOL.2015.144.
32. Carethers, J.M., and Jung, B.H. (2015). Genetics and Genetic Biomarkers in Sporadic Colorectal Cancer. Gastroenterology *149*, 1177–1190.e3. https://doi.org/10.1053/J.GASTRO.2015.06.047.
33. Müller, M.F., Ibrahim, A.E.K., and Arends, M.J. (2016). Molecular pathological classification of colorectal cancer. Virchows Arch. *469*, 125–134. https://doi.org/10.1007/S00428-016-1956-3.
34. Al-Mulla, F., AlFadhli, S., Al-Hakim, A.H., Going, J.J., and Bitar, M.S. (2006). Metastatic recurrence of early-stage colorectal cancer is linked to loss of heterozygosity on chromosomes 4 and 14q. J. Clin. Pathol. *59*, 624–630. https://doi.org/10.1136/JCP.2005.033167.
35. Krishnan, P., Ghosh, S., Wang, B., Heyns, M., Graham, K., Mackey, J.R., Kovalchuk, O., and Damaraju, S. (2016). Profiling of Small Nucleolar RNAs by Next Generation Sequencing: Potential New Players for Breast Cancer Prognosis. PLoS One *11*, e0162622. https://doi.org/10.1371/JOURNAL.PONE.0162622.
36. Chen, C.C., Chang, P.Y., Chang, Y.S., You, J.F., Chan, E.C., Chen, J.S., Tsai, W.S., Huang, Y.L., Fan, C.W., Hsu, H.C., and Chiang, J.M. (2023). MicroRNA-based signature for diagnosis and prognosis of colorectal cancer using residuum of fecal immunochemical test. Biomed. J. *46*, 144–153. https://doi.org/10.1016/j.bj.2022.01.011.
37. Lykke-Andersen, S., Ardal, B.K., Hollensen, A.K., Damgaard, C.K., and Jensen, T.H. (2018). Box C/D snoRNP Autoregulation by a cis-Acting snoRNA in the NOP56 Pre-mRNA. Mol. Cell *72*, 99–111.e5. https://doi.org/10.1016/j.molcel.2018.08.017.
38. Gagliardi, A., Francescato, G., Ferrero, G., Birolo, G., Tarallo, S., Francavilla, A., Piaggeschi, G., Di Battista, C., Gallo, G., Realis Luc, A., et al. (2023). The 8q24 region hosts miRNAs altered in biospecimens of colorectal and bladder cancer patients. Cancer Med. *12*, 5859–5873. https://doi.org/10.1002/CAM4.5375.
39. Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at arXiv *1*. https://doi.org/10.48550/arXiv.1303.3997.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Barcelona colorectal tissue samples (BCT cohort)	Hospital Clinic de Barcelona (Spain)	N/A
Turin colorectal tissue samples (TCT cohort)	Clinica S. Rita in Vercelli (Italy)	N/A
FIT-positive RNA samples (BFIT cohort)	Hospital Clinic de Barcelona (Spain)	N/A
Critical commercial assays		
miRNeasy Mini Kit	QIAGEN	Cat#217004
RNeasy MinElute Cleanup Kit	QIAGEN	Cat#74204
TruSeq small RNA Sample Prep Kit	Illumina	Cat#RS-200-0012
NEBNext Multiplex Small RNA Library Prep for Illumina	New England Biolabs	Cat#E7330L
QIAQuick PCR Purification Kit	QIAGEN	Cat#28104
TaqMan MicroRNA Reverse Transcription Kit	Life Technologies	Cat#4366596
TaqMan Preamp Master Mix	Life Technologies	Cat#4384267
TaqMan Universal Master mix no UNG II	Life Technologies	Cat#4440043
ddPCR Supermix for Probes	Bio-Rad	Cat#186-3010
Deposited data		
GENCODE release 29, GRCh38.p12	Genome Reference Consortium	https://www.gencodegenes.org/human/release_29.html
snoRNA-seq count data	This paper	https://doi.org/10.17632/v8m4vbbv5j.1
RT-qPCR data	This paper	Shared upon request by the Lead Contact
ddPCR data	This paper	Shared upon request by the Lead Contact
Oligonucleotides		
Custom TaqMan small RNA Assay SNORA51	Thermo Fisher	Cat#4398987; Assay ID: CTXGRDD
Custom TaqMan small RNA Assay SNORA54	Thermo Fisher	Cat#4398987; Assay ID: CTYMJXA
Custom TaqMan small RNA Assay SNORD114	Thermo Fisher	Cat#4398987; Assay ID: CTMFXCY
Custom TaqMan small RNA Assay SNORD115	Thermo Fisher	Cat#4398987; Assay ID: CTZTECJ
Custom TaqMan small RNA Assay SNORD116	Thermo Fisher	Cat#4398987; Assay ID: CTKA3R2
Custom TaqMan small RNA Assay SNORD12	Thermo Fisher	Cat#4398987; Assay ID: CTPRKGR
Custom TaqMan small RNA Assay SNORD12B	Thermo Fisher	Cat#4398987; Assay ID: CTNKRWU
Custom TaqMan small RNA Assay SNORD12C	Thermo Fisher	Cat#4398987; Assay ID: CTRWE2N
Custom TaqMan small RNA Assay SNORD19B	Thermo Fisher	Cat#4398987; Assay ID: CTEPTAJ
Custom TaqMan small RNA Assay SNORD63	Thermo Fisher	Cat#4398987; Assay ID: CTTZ9MK
Custom TaqMan small RNA Assay SNORD72	Thermo Fisher	Cat#4398987; Assay ID: CTWCWTF
Custom TaqMan small RNA Assay SNORD78	Thermo Fisher	Cat#4398987; Assay ID: CTU627H
Custom TaqMan small RNA Assay SNORD15B	Thermo Fisher	Cat#4398987; Assay ID: CTCE4MT
Custom TaqMan small RNA Assay SNORD89	Thermo Fisher	Cat#4398987; Assay ID: CTZTEG7
Software and algorithms		
Partek Flow	Partek Inc.	https://www.partek.com/
TrimGalore	Babraham Bioinformatics	https://github.com/FelixKrueger/TrimGalore
STAR	Alex Dobin	https://github.com/alexdobin/STAR

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Burrows-Wheeler Aligner	Li H. (2013)[39]	https://github.com/lh3/bwa
Partek Expectation Maximization Algorithm	Partek Inc.	https://www.partek.com/
R 4.0.4	R Core Team	https://www.R-project.org/

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Dr. Meritxell Gironella (meritxell.gironella@ciberehd.org).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- snoRNA-seq count data has been deposited at Mendeley Data and are publicly available as of the date of publication. Accession numbers are listed in thekey resources table. Clinical information data has been deposited at Mendeley Data. They are publicly available as of the date of publication. Accession numbers are listed in the key resources table. RT-qPCR and ddPCR data reported in this paper will be shared by the lead contact upon reasonable request.
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

For this study, a total of 521 samples from three independent cohorts of patients -two sets of colonic tissues and one set of fecal samples-were analyzed (Table 1). The Discovery Cohort called Barcelona Colonic Tissue cohort (BCT) consisted of 88 paired advanced colorectal neoplasia and healthy adjacent tissues obtained from 44 patients (23 CRC, 21 AA) from Hospital Clinic of Barcelona (Barcelona, Catalonia, Spain) during surgical intervention. An independent Validation Cohort, named Turin Colonic Tissue cohort (TCT), was made up of 262 paired colonic tissues from 131 patients (105 CRC, 26 AA) collected from Clinica S. Rita in Vercelli (Piemonte, Italy). Selected snoRNAs were further validated by RT-qPCR in a subset of paired tissues from the Barcelona colonic tissue cohort (sBCT) (15 CRC, 15 AA) and, finally, some snoRNA candidates from the previous results were assessed by ddPCR in a set of 171 stool samples from FIT-positive individuals participating in a CRC screening program named Barcelona FIT+ cohort (BFIT) (56 CRC, 63 AA, 52 Healthy controls). Controls were healthy individuals with FIT+ but no lesions at colonoscopy. Clinicopathological information from all these cohorts of patients is detailed in Table 1 and the overall workflow for this study is illustrated in Figure S1. Only biological sex was used in the analysis of these patients. No information about gender, ancestry, race, ethnicity, or socioeconomic status was collected for any of the participants. The study was approved by the institutional ethics committee of Hospital Clinic of Barcelona, and written informed consent was obtained from all participants in accordance with the Declaration of Helsinki.

METHOD DETAILS

RNA extraction and snoRNA sequencing

Total RNA from the BCT cohort was extracted using miRNeasy mini kit (QIAGEN, Hilden, Germany) followed by a step of RNeasy MinElute Cleanup kit (QIAGEN). RNA concentration was analyzed using a Nanodrop 1000 spectrophotometer (Nanodrop, Wilmington, DE, USA) and RNA purity was assessed with an Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA). Specific Small RNA libraries for snoRNA sequencing were generated with 1 μg of total tissue RNA using TruSeq small RNA Sample Prep kit (Illumina, San Diego, CA, USA), following manufacturer's protocol. The band of interest for sequencing (250bp-600bp) was size-selected with a polyacrylamide gel, as it represents the region where most small nucleolar RNAs are found. Samples were tagged with a specific adaptor before sequencing. Illumina adaptors were attached to the 3′ and 5′ ends of small RNAs, being the 3′ adaptor essential for cDNA synthesis. Generated libraries were analyzed using Bioanalyzer 2100 (Agilent Technologies). Sequencing was run in a HiSeq 2000 system (Illumina) with 2×75 base-pair reads to obtain >15M paired end reads per sample. Quality control was performed by spiking-in a small quantity of the PhiX reference genome and calculating the percentage of mismatches using GEM program. The average error percentage had to be lower than 2.5% on each sample for it to pass the test.

Small RNA-Seq FASTQ files were processed using Partek Flow software (Partek Inc., Chesterfield, MI, USA). Adapter sequences were trimmed with TrimGalore and first mapped using STAR using default settings. Unaligned reads from the first alignment were realigned using the Burrows-Wheeler Aligner using default settings. Both alignments were mapped to the reference genome of GENCODE release 29 (GRCh38.p12) and combined to quantify the expression of each gene using Partek's Expectation Maximization Algorithm.

For tissue samples from the TCT cohort, total RNA was extracted using QIAzol (Qiagen), according to the manufacturer's instructions. Small RNA-seq libraries were prepared by a slightly different approach than the Barcelona cohort and according to the protocol described by Gagliardi and colleagues.[38] Briefly, NEBNext Multiplex Small RNA Library Prep for Illumina (New England Biolabs, Ipswich, MA, USA) was used to process 6 μL of RNA as starting material following manufacturers' instructions. cDNA was purified with QIAQick PCR Purification Kit (QIAGEN) after amplification. Libraries were loaded on a Bioanalyzer 2100 system (Agilent Technologies) and finally run in a HiSeq 2000 Sequencing System (Illumina). Small RNA-Seq FASTQ files from BCT were processed using Partek Flow software (Partek Inc., Chesterfield, MI, USA). Adapter sequences were trimmed with TrimGalore and first mapped using STAR using default settings. Unaligned reads from the first alignment were realigned using the Burrows-Wheeler Aligner using default settings. Both alignments were mapped to the reference genome of GENCODE release 29 (GRCh38.p12) and combined to quantify the expression of each gene using Partek's Expectation Maximization Algorithm. TCT sequencing reads were processed using the same software and settings as BCT samples.

BCT and TCT cohort small RNA-seq data were pre-processed before further analyses. From all the genes that were sequenced, only those annotated as snoRNA by GENCODE, whose sum of counts was not zero, and mean expression was higher than ten raw counts, were kept for downstream analyses. Unpaired samples were excluded from the analysis.

SnoRNA expression analysis by RT-qPCR

In total, 60 tissue RNAs from a subset of the BCT cohort (15 CRC, 15 AA, and the corresponding paired normal tissues), called sBCT cohort, were used for the technical validation by RT-qPCR of the selected snoRNAs from the previous analyses. Reverse transcription (RT) was performed using TaqMan MicroRNA Reverse Transcription Kit (Life Technologies, ThermoFisher Scientific, Waltham, MA, USA) and TaqMan Custom small RNA assays (Life Technologies) for each of the selected snoRNAs (SNORA51, SNORA54, SNORD114, SNORD115, SNORD116, SNORD12, SNORD12B, SNORD12C, SNORD19B, SNORD63, SNORD72, SNORD78, SNORD15B, and SNORD89). SNORD12, SNORD12B, SNORD72, SNORD89, and SNORD114 were pre-amplified using TaqMan Preamp Master Mix (Life Technologies) and diluted in a 1:8 ratio before quantification. Expression levels of snoRNAs were assessed in 2 μL of cDNA using TaqMan Universal Master Mix no UNG II (Life Technologies) and specific TaqMan Real Time primers for each snoRNA (Life Technologies) as detailed above. qPCR was performed in triplicates and run on a ViiA 7 Real-Time PCR System (Applied Biosystems). Relative expression (-ΔCt) was calculated by using RNU6B as normalizer.

SnoRNA quantification by ddPCR on fecal samples

Fecal RNA was extracted from 500 μL of FIT leftover from the BFIT cohort (n = 171) as previously described in the study by Duran-Sanchon and colleagues using miRNeasy Mini Kit (QIAGEN).[10] The extracted RNA was then diluted to 5 ng/μL to quantify the expression of selected snoRNAs. RNA was reverse transcribed using TaqMan MicroRNA Reverse Transcription Kit (Life Technologies) and TaqMan Custom small RNA assays (Life Technologies) for each snoRNA (SNORD12C, SNORD12B, SNORA51, SNORD72, SNORD15B). SnoRNA expression was analyzed by using Custom TaqMan Real Time probes (Life Technologies) and BioRad Supermix (BioRad, Hercules, CA, USA) following manufacturers' protocol without diluting the cDNA. Droplets were generated in a QX100 Droplet Generator (BioRad) and absolute quantification was measured in a QX100 Droplet Reader System (BioRad) in copies/μL.

QUANTIFICATION AND STATISTICAL ANALYSIS

Paired differential expression analysis from this case-control sequencing data (n = 88) was performed using R version 4.0.4 and the package DESeq2. We defined statistical significance at a false discovery rate (FDR) lower than 0.05 and fold change (FC) 0 as a threshold to define upregulation (FC > 0) or downregulation (FC < 0). Log_2 fold change (L2FC) values lower than 0 were transformed using the formula −1/(2^L2FC) for ease of interpretation. To compare results between BCT cohort and TCT cohort, snoRNA were considered coincident when FDR <0.05 and FC were in the same direction in both cohorts (i.e., upregulation or downregulation in both cohorts). Differences between groups in RT-qPCR or ddPCR data were performed by t-test unless stated otherwise. All variables analyzed were tested for normality using Shapiro-Wilk test. Asterisks for statistical significance are displayed as ns: not significant; *p < 0.1; **p < 0.05; ***p < 0.01. SnoRNA expression data from the BCT and BFIT cohorts was analyzed using a logistic regression model adjusted by age and sex to determine CRC, AA or advanced colorectal neoplasia (ACN) as the main outcomes. A training dataset was sampled from the main dataset to construct a logistic regression model with 5 times 10-fold cross validation using the caret package. The R package pROC was used to compute the receiver operating characteristic analyses considering each snoRNA as a continuous variable with no missing values. Area under the ROC curve (AUC), sensitivities, specificities, positive predictive values, and negative predictive values were calculated for each model using the optimal threshold calculated by Youden's J statistic.