Received: 8 July 2022 | Accepted: 25 January 2023

DOI: 10.1002/iub.2710

SPECIAL ISSUE

IUBMB LIFE WILEY

A hybrid approach to assess the structural impact of long noncoding RNA mutations uncovers key *NEAT1* interactions in colorectal cancer

Efe Aydın[1†] | **Ester Saus**[2,3†] | **Uciel Chorostecki**[2,3] | **Toni Gabaldón**[2,3,4,5]

[1]Department of Laboratory Medicine, Division of Clinical Genetics, Lund University, Lund, Sweden

[2]Institute for Research in Biomedicine (IRB Barcelona), The Barcelona Institute of Science and Technology, Barcelona, Spain

[3]Barcelona Supercomputing Centre (BSC-CNS). Plaça Eusebi Güell, Barcelona, Spain

[4]Catalan Institution for Research and Advanced Studies (ICREA), Barcelona, Spain

[5]Centro de Investigación Biomédica En Red de Enfermedades Infecciosas (CIBERINFEC), Barcelona, Spain

Correspondence

Toni Gabaldón, Institute for Research in Biomedicine (IRB Barcelona), The Barcelona Institute of Science and Technology, Baldiri Reixac, 10, 08028 Barcelona, Spain.
Email: toni.gabaldon@bsc.es

Uciel Chorostecki, Institute for Research in Biomedicine (IRB Barcelona), The Barcelona Institute of Science and Technology, Baldiri Reixac, 10, 08028 Barcelona, Spain.
Email: uciel.chorostecki@bsc.es

Funding information

H2020 European Research Council, Grant/Award Number: 724173

Abstract

Long noncoding RNAs (lncRNAs) are emerging players in cancer and they entail potential as prognostic biomarkers or therapeutic targets. Earlier studies have identified somatic mutations in lncRNAs that are associated with tumor relapse after therapy, but the underlying mechanisms behind these associations remain unknown. Given the relevance of secondary structure for the function of some lncRNAs, some of these mutations may have a functional impact through structural disturbance. Here, we examined the potential structural and functional impact of a novel A > G point mutation in *NEAT1* that has been recurrently observed in tumors of colorectal cancer patients experiencing relapse after treatment. Here, we used the nextPARS structural probing approach to provide first empirical evidence that this mutation alters *NEAT1* structure. We further evaluated the potential effects of this structural alteration using computational tools and found that this mutation likely alters the binding propensities of several *NEAT1*-interacting miRNAs. Differential expression analysis on these miRNA networks shows upregulation of *Vimentin*, consistent with previous findings. We propose a hybrid pipeline that can be used to explore the potential functional effects of lncRNA somatic mutations.

Abbreviations: BLACAT1, BLACAT1 Overlapping LEMD1 Locus; CCAT2, Colon Cancer Associated Transcript 2; CRC, Colorectal cancer; DDX5, DEAD-Box Helicase 5; FENDRR, Adjacent Non-Coding Developmental Regulatory RNA; FOXF1, Forkhead Box F1; lncRNAs, Long noncoding RNAs; LUCAT1, lung cancer associated transcript 1; NEAT1, Nuclear Enriched Abundant Transcript 1; PCAT1, Prostate Cancer Associated Transcript 1; SNHG7, Small Nucleolar RNA Host Gene 7.

[†]Efe Aydın and Ester Saus contributed equally to this work.

[Correction added on 7 April 2023, after first online publication. Section headings under Method section have been modified].

1 | INTRODUCTION

Colorectal cancer (CRC) is the third most common cancer worldwide, with 1.8 million new cases and 880,000 deaths each year.[1] CRC is a progressive and heterogeneous disease. However, two molecular pathways account for the majority of cases: The chromosomal instability pathway, which accounts for 70–75% of the cases and is triggered by the accumulation of mutations in various oncogenes and tumor suppressor genes, and the serrated pathway, accounting for the remaining 25–30% and initiated by specific mutations on the *BRAF* oncogene.[1] Treatment of CRC is based on a combination of surgery, chemotherapy, and radiotherapy, and its success correlates negatively with the level of the advance of the malignancy at diagnosis, and with other parameters such as the location of the tumor and the condition of the patient.[2] Tumor relapse after therapy significantly reduces the survival rates of CRC patients. Relapse affects 30–40% of treated patients and may manifest on local, regional, and distant tissues.[3–5] Better predicting the likelihood of relapse is important to guide therapy and improve patient care, and there is a growing interest in finding new prognostic biomarkers for CRC. In addition, identifying molecules involved in the onset and progression of CRC may pave the way for novel, more directed therapeutic strategies.

Long, noncoding RNAs (lncRNAs) are transcripts longer than 200 nt that do not encode proteins.[6] Contrary to most protein-coding genes, lncRNAs do not exhibit strong evolutionary conservation,[7] making it harder to find homologs and study their functional roles in model organisms. Synteny could be an alternative to study lncRNA conservation, particularly the poorly conserved at the sequence level.[8] In addition, lncRNAs are generally expressed at low levels, follow tissue-specific or cell-type specific expression patterns,[9] and undergo frequent alternative splicing.[10] These features hamper their study and the identification of disease-associated lncRNAs. Accumulating evidence supports an important role of lncRNAs in CRC,[11] and an important therapeutic, diagnostic, or prognostic potential of lncRNAs has been proposed.[12] LncRNAs with altered expression in CRC and with potential roles in the progression of the disease include *AK123657*, *BX649059*, *CCAT2*, *HOTAIR*, and *MALAT1*, among many others.[13–17] However, the roles and mechanisms of action for most of these lncRNA remain poorly understood.

LncRNAs can act through diverse molecular mechanisms, often resulting in altered expression of other genes through their regulation at the epigenetic, transcriptional and post-transcriptional levels.[18] Therefore, gaining knowledge of the lncRNA interactome—that is, all biomolecules that interact with a given lncRNA—is key for elucidating the mechanism of action of a lncRNA of interest. One of the most prominent types of lncRNA-interacting molecules are microRNAs (miRNAs). Several studies have shown competitive binding of oncogenic lncRNAs to tumor suppressor miRNAs in CRC. For example, lncRNA *SNHG7* sponging miR-216b is related to liver metastasis of CRC.[19] A recent study demonstrated the suppression of miR-145 maturation through the lncRNA *CCAT2*, a regulatory mechanism that is thought to be associated with CRC stem cell proliferation and differentiation.[20] In addition to miRNAs, lncRNAs can also interact directly with messenger RNAs (mRNAs). Through these interactions, lncRNAs may affect mRNA splicing, editing, and stability, thereby affecting translation of the encoded protein.[21,22] It is also possible for lncRNAs to form triple helices by interacting with dsDNA.[23] For example, lncRNA *FENDRR* binds to the promoter regions of *FOXF1* and *PITX2* genes, forming a triple helix and thereby exposing binding sites for the polycomb repressive complex, which in turn affects target genes through an epigenetic control mechanism.[24,25] Apart from nucleic acids, lncRNAs are also known to interact with proteins.[23] For instance, the lncRNA *NEAT1* indirectly activates the Wnt/β-catenin signaling pathway via binding to DDX5 and promotes CRC progression.[26]

As stated previously, lncRNAs generally exhibit low evolutionary sequence conservation, which prevents the use of functional annotation approaches relying on sequence comparisons. This drawback puts an extra emphasis on lncRNA secondary/tertiary structures, which are thought to play a significant role in the function of lncRNAs.[27] Indeed, earlier studies have shown that nucleotide substitutions that are predicted to be involved in secondary structure formation are subject to stronger selection constraints.[28]

Numerous computational algorithms have been developed to predict the secondary structures of RNAs, including, among many others MFold,[29] RNAfold,[30] RNAStructure,[31] and RNAShapes.[32] However, evaluations on these in silico methods show an average accuracy of 38% in secondary structure topology,[33] indicating a significant need of improvement to reach confident results. This is particularly problematic when trying to predict the secondary structure of long RNA molecules, as shown by a recent comparison of computational methods.[34] Recent experimental approaches based on enzymatic probing followed by sequencing such as

nextPARS[35,36] allow to obtain complementary information and better determine lncRNA secondary structures.

In this study, we assessed the structural impact of recurrent lncRNA mutations in CRC tumors identified in a recent study that used a novel lncRNA enrichment approach.[37] The use of target-enrichment RNA-Seq enabled detection and genotyping of lncRNAs, despite their low expression levels. Such approach opens new avenues for the functional annotation studies of cancer variants within the noncoding genome, which stands for the majority of known cancer variants.[38] In addition, stage II of CRC represents a crossroads in clinical decision-making, since the potential use of adjuvant chemotherapy is decided on this phase.[39] That study identified 379 lncRNAs that were differentially expressed between tumor and adjacent healthy colonic tissue derived from stage II CRC patients, and identified putative somatic mutations detected in tumors but not in paired healthy tissues. Some of them were recurrently observed. Here we built up from those results and developed a hybrid approach to assess the potential structural and functional impact of such mutations, based on experimental structural probing using nextPARS and a set of computational tools. After prioritizing candidate mutations with low population frequency, and enriched in tumors with a clinical history of relapse, we selected a novel *NEAT1* mutation. The experimental structural characterization of mutant and wild type variants of the *NEAT1* transcript, identified significant structural changes around the mutated nucleotide that were undetectable by a computational-only approach. Such structural changes were not observed in control mutations with high population frequency and hence likely non-deleterious. Using a set of computational tools, we further evaluated the potential effects of this mutation from a functional point of view. We found that tumors carrying this mutation showed large transcriptional differences with tumors lacking it and found dozens of genes differentially expressed when the mutation was present. Using an interactome oriented approach and stringent analytical criteria, we uncovered potential miRNA interactions that could be affected in the presence of this mutation, thereby providing a mechanistic hypothesis for the association of this mutation with CRC and relapse.

Altogether, our results provide, to our knowledge, the first empirical evidence that a lncRNA mutation associated with tumor relapse has a structural impact. More generally, we propose a hybrid approach that can serve to prioritize lncRNAs for further experimental characterization, to generate testable hypotheses of lncRNA mechanisms of action and, more generally, to address the low functional annotation rate of lncRNAs, which is drastically outpaced by the discovery of novel signatures.

2 | METHODS

2.1 | Data collection

We used the sequence and clinical data derived from a lncRNA target-enrichment study in CRC.[37]

SNV selection

A total of 5,715 lncRNA mutations found in tumors but not in paired healthy tissue (i.e., putative somatic) were reported in table S7 of aforementioned study.[37] Using these data, we selected mutations that were recurrently occurring on at least three relapse samples. To filter out potentially wrongly imputed somatic mutations, we removed mutations with population frequencies higher than 0.001 on gnomAD 3.1.2.[40] As controls, we included seven other mutations that were reported with high recurrence in this dataset but failed to pass the allele frequency filtering. These mutations were found in the following lncRNAs; *BLACAT1*, *LUCAT1*, *LINC01811*, and *PCAT1*.

2.2 | nextPARS

We used the nextPARS technique[35] to experimentally probe the structure of the following lncRNAs: *BLACAT1*, *NEAT1*, *LUCAT1*, *LINC01811*, and *PCAT1*. We selected a fragment of 300 nucleotides containing the somatic mutation of interest per each gene, except for *BLACAT1*, for which two fragments were chosen (BLACAT1_1 of 1,000 bases, and BLACAT1_2 of 300 bases) to cover the three mutations present in that gene. (Dataset S1).

Two independent nextPARS experiments were run for this study: (i) lncRNAs fragments without the mutations, and (ii) lncRNAs fragments containing the somatic mutations of interest. All lncRNA fragments were produced as described before.[35,41] Briefly, PCRs were used to amplify and linearize the different fragments from each corresponding plasmid preparation (pUC57 vector with cloned regions of interest, GenScript Biotech, Netherlands). Primer sequences, amplicon sizes, mutations, and PCR conditions per each fragment are shown in Tables S1 and S2, respectively. After confirmation by Sanger sequencing that the amplified fragments were correct, all fragments were in vitro transcribed using the T7

RiboMax Large-scale RNA production system (Promega) and size-selected and purified using Novex-TBE Urea gels (Life Technologies). Final purified RNAs were quality controlled by means of Agilent 2,100 Bioanalyzer with the RNA 6000 Pico LabChip Kit (Agilent) and the Qubit Fluorometer with the Qubit RNA BR (Broad-Range) Assay Kit (ThermoFisher Scientific).

NextPARS was used to probe the secondary structure of the RNA molecules at 23°C as described before,[35] with a starting material of 2 μg of polyA+ RNA mixed with 20 ng of each lncRNA fragment in each reaction. 0.03 U of RNase V1 (Ambion) and 200 U of S1 nuclease (Fermentas) were used to digest the corresponding samples. After confirming the good quality of the final digested samples, TruSeq Small RNA Sample Preparation Kit (Illumina) was used to prepare the libraries following a modified protocol previously described.[35] Final libraries were sequenced in single-reads of 50 nucleotides in Illumina HiSeq2500 sequencers at the Genomics Unit of the CRG (CRG-CNAG). The computation of the nextPARS scores was obtained following the protocol described in.[36] The code to obtain the structural profile from nextPARS experiments is available on Github at https://github.com/Gabaldonlab/nextPARS_docker. NextPARS scores were converted to SHAPE-like normalized reactivities using the nextPARS2SHAPE v1.0 script (https://github.com/Gabaldonlab/MutiFolds/blob/master/scripts/nextPARS2SHAPE.py).

2.3 | Interactome analysis

RNAInter v4.0 was used to extract interactome information for *NEAT1*.[42] Our search query was filtered to show only results from RNA–RNA, RNA–DNA interactions of *NEAT1* with strong experimental evidence. A list of interaction detection methods that were identified as strong experimental evidence can be seen on the RNAInter website (www.rnainter.org).

IntaRNA 2.0 was used for predicting miRNA-lncRNA binding sites.[43] IntaRNA is a widely used in silico method for RNA–RNA interaction prediction that incorporates seed constraints and accessibility of interacting subsequences.[43] Energy scores generated from these predictions are based on the sum of the free energy from hybridization and the required free energy of site accessibility. Default values were slightly altered following the assessment done on intaRNA seed and interaction constraints by Raden et al.[44] We ran the algorithm with the following settings; maximal interaction length of 60, maximum loop length of 8, minimum number of base pairs in seed as 7, and minimal unpaired probability of 0.001. Our SHAPE-like reactivity scores generated from the nextPARS experiment were integrated within the target SHAPE-Input parameters of the tool to increase prediction accuracy. Galaxy Europe server was used for the data analysis and implementation of the IntaRNA algorithm.[45] The remaining RNAs were targeted via LncTar algorithm to measure whether our candidate mutation was altering the conditional probability of these interactions.[46] LncTar was selected for method consistency since it also incorporates free energy minimization to generate results. LongTarget was used to predict if there is any triplex formation propensity alteration due to the given *NEAT1* mutation.[47] All triplex formation rules were included within the command and only TFO1 results were extracted to provide higher confidence levels. The following command was used to generate results:

```
./LongTarget-f1 DNA.fa-f2 lncRNA.fa-r 0
```

To assess possible RNA-Protein interactions, we used ENCODE RNA Binding Proteins track on UCSC Genome Browser to see if any protein interactions were reported on the mutation site from RIP-chip, TillingArray or RIP-seq experiments done by ENCODE researchers.[48] Binding propensities of mutant and wild structures were compared using PRIdictor.[49]

Protein, miRNA, other-RNA, and DNA sequences were downloaded from NCBI (GRCh38.p14), miRBase (Release 22.1), Ensembl (Release 105), and NCBI (GRCh38.p14) respectively.[50–52] As we are only interested in the neighboring regions of our candidate mutation, we only included the same sequences we used for the nextPARS spike-ins to ensure consistency throughout the analytical workflow.

Differential expression analysis

We used DESeq2 v1.22.2 for differential gene expression analysis.[53] Briefly, using Salmon we obtained gene-level abundance estimates and we used tximport v1.10.1 with txOut = T option to import it in R v3.5.1. We performed differential expression analysis using the deseq function.[53] We extracted the results after filtering out genes with an adjusted *p*-value lower than .01 and log2 fold-change ranging between −2 and 2. Differentially expressed transcripts were compared against the candidate genes that were generated from LncTar and LongTarget applications. To assess the effect of miRNA binding, we searched for gene regulation relations in CRC from the emiRIT portal and extracted entries that

included our selected miRNAs after intaRNA execution.[54] Interacting genes of these miRNAs were extracted and compared to our differentially expressed genes.

2.4 | Enrichment analysis

ShinyGO 0.76 was used to analyze the Gene Ontology (GO) Enrichment in Biological Processes, Cellular Components, Molecular Functions, and KEGG pathways. Minimum pathway size was selected as 2 and a cut-off value of 0.01 was used for FDR.[55]

For the Gene Set Enrichment Analysis (GSEA), fsgea package of Bioconductor was used.[56,57] Following MSigDB Collections were selected for analysis; Hallmark gene sets, Curated Gene Sets: Canonical Pathways, and Oncogenic Gene Sets.[57] FDR value of 0.01 was used for selection constraint.

2.5 | RNA secondary structure visualization

To obtain the secondary structure of *NEAT1* wild type (wt) and *NEAT1* mutated (mut) molecules, we use the RNAfold software (version 2.4.13), using pseudo energy restraints. Residues for which there was no nextPARS data were assigned a reactivity of 999, as sug-gested by the RNAfold manual. The RNA arc diagram was done using R-chie (version 2.0.8).[30]

2.6 | Data availability

The raw sequencing reads produced in this project have been deposited in the Short Read Archive of the European Nucleotide Archive under Bioproject ID PRJNA838569. (Reviewer's access https://dataview.ncbi.nlm.nih.gov/object/PRJNA838569?reviewer=qjf34nj0ud7jdbah98padqvshf).

3 | RESULTS

3.1 | Selection of candidate mutations affecting lncRNAs

The RNA-Seq of the aforementioned study included matched samples taken from tumor and adjacent healthy tissues from 35 stage II CRC patients, and identified 5,714 lncRNA mutations. We sorted these mutations based on their prevalence in relapse samples, and selected the top 5 lncRNAs that harbor the mutations most prevalent in tumors of relapse patients. We further evaluated their allele frequencies in the global human population reported in gnomAD v3.1.2. Two mutations passed this filtering step; an A > G point mutation on chr11:65441347 *NEAT1* and a T > C point mutation on chrX:74277823 of *FTX*. According to the original study, *NEAT1* mutation was observed on 11 samples, while the FTX one was seen on 3. *NEAT1* is also a lncRNA that is more thoroughly studied with numerous publications demonstrating its carcinogenic effects. Thereby, we decided to keep our focus on *NEAT1* mutation on this study but we think this other recurrent mutation occurring on *FTX* is also worth studying further, considering the accumulating evidence of its association with CRC.[58,59]

To assess whether the selected recurrent *NEAT1* mutation had been reported in other cancer patients, we looked at dbSNP, which includes information from additional databases besides gnomAD.[60] No reports of this mutation was found. We further checked the existence of the mutation on the Noncoding Variants dataset of COSMIC, but we found only one liver cancer sample harboring this mutation with the id number CHC2113.[61]

As controls for the structural analyses, we selected seven recurrent mutations from four lncRNAs from the same study that presented a high frequency in gnomAD, therefore likely representing nondeleterious variants (Table 1) [Correction added on 7 April 2023, after first online publication. The sentence citing Table 1 has been changed from COSMIC to gnomAD.].

Amongst these lncRNAs; *BLACAT1*, *PCAT1*, and *LUCAT1* have also been previously associated with CRC progression.[62–64] To our knowledge, only *LINC01811* has not been previously associated with CRC from a functional standpoint. However, due to the presence of *LINC01811* mutation on four relapse samples, we also included this lncRNA in our analysis.

3.2 | Assessing the impact of somatic mutations in lncRNA structure using nextPARS

To determine whether our selected mutations can affect the structure of *BLACAT1*, *NEAT1*, *LUCAT1*, *LINC01811*, *and PCAT1*, we designed RNA probing experiments using nextPARS that compared wild type and specifically mutated versions of each transcript (see Materials and Methods). This allowed us to study changes in the secondary structure of our candidate lncRNAs at a single-nucleotide resolution. Each nextPARS experiment was performed in duplicates which allowed us to validate the robustness of our experiment, which was high (Average replicate correlation values for *BLACAT1*_1,

Location	lncRNA	Mutation	*n* Relapse	gnomAD AF
chr1:205435419	*BLACAT1*	A > G	6	0.689
chr1:205435109	*BLACAT1*	G > A	5	0.651
chr1:205435750	*BLACAT1*	T > C	4	0.722
chr1:205436836	*BLACAT1*	A > G	4	0.764
chr5:91313192	*LUCAT1*	T > C	3	0.764
chr3:34394859	*LINC01811*	A > G	4	0.948
chr8:127339505	*PCAT1*	G > C	4	0.982

TABLE 1 Recurrent SNPs reported on the CRC study which we further selected for the nextPARS experiments as controls. Columns indicate, in this order, the location of the mutation, the corresponding lncRNA, base substitution information, number of relapse samples that carry the corresponding mutation, and the allele frequencies of the mutations based on gnomAD.

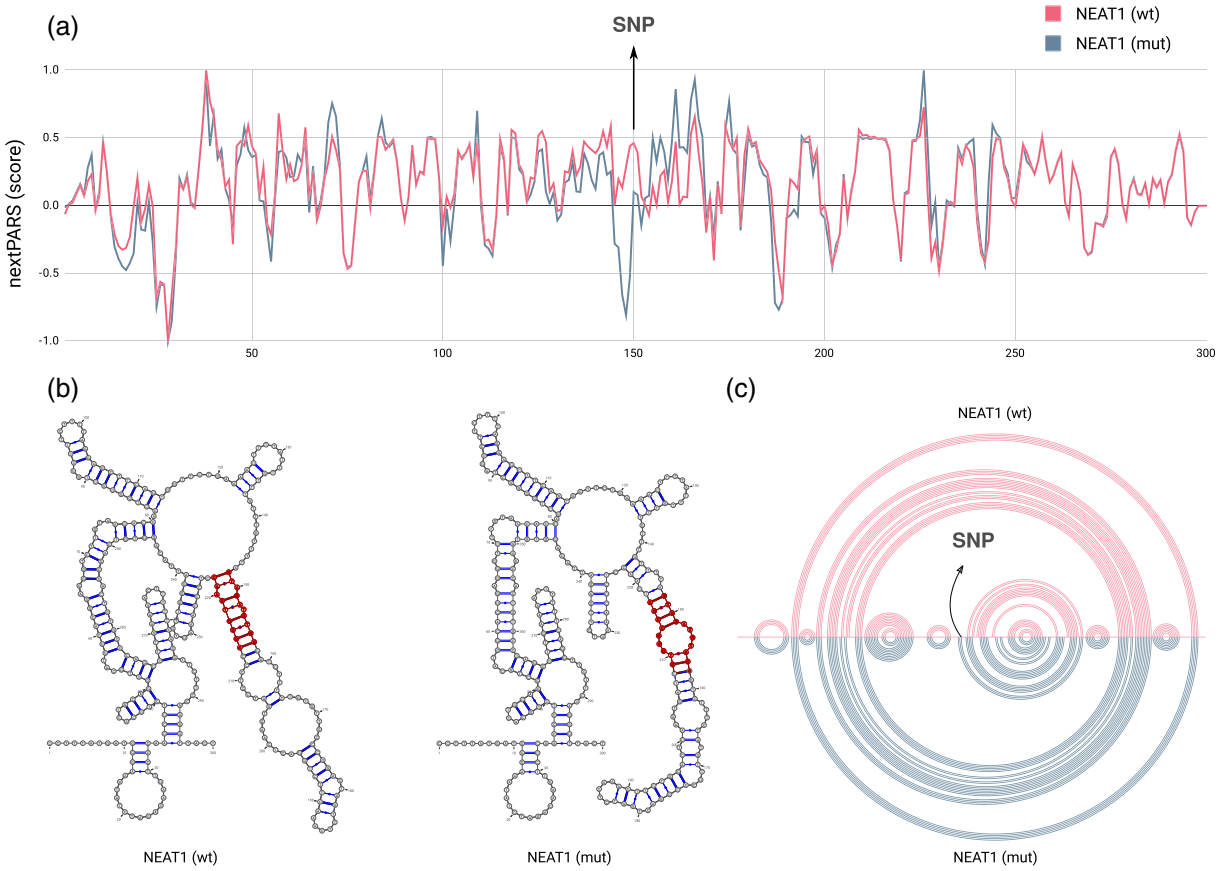

FIGURE 1 Visualization of nextPARS results showing the structural impact of the *NEAT1* mutation on its corresponding base. (a) Normalized nextPARS scores for wild type (wt) and mutant (mut) molecules of *NEAT1*. The mutation site A > G chr11:65441347) found on base 150 indicates a paired structural preference while a strong unpaired structural preference score was obtained for the wt molecule. The rest of the regions display generally similar structural patterns for both structures. (b) Secondary structure representation of *NEAT1* (wt) and *NEAT1* (mut) using nextPARS data as a constraint. (c) RNA arc diagram representing the base-pairing of the two secondary structure sequences of *NEAT1* (wt) and *NEAT1* (mut) using nextPARS data as a constraint.

BLACAT1_2, *NEAT1*, *LUCAT1*, *LINC01811*, and *PCAT1* were found to be 0.98, 0.97, 0.98, 0.97, 0.98, and 0.98 respectively), even higher than the correlations reported in the original nextPARS publication.[35] A heatmap showing high agreement between replicates can be found in Figure S1.

We next compared the structural preference profiles using SHAPE-like normalized reactivities for each residue (see Materials and Methods, Dataset S2). Our results indicate that the introduction of the A > G chr11:65441347 mutation in *NEAT1* generated a different structural preference, particularly in a region around the mutated residue (Figure 1), shifting from an unpaired to a paired state. Gibbs Free Energy difference for this structural change was found to be −0.9 when the wild type structure was considered as reference. In contrast, there

were no structural preference alterations observed for the other transcripts (Dataset S2). The lack of structural effects of the mutations in *BLACAT1*, *LUCAT1*, *LINC01811*, and *PCAT1* is consistent with the relatively high allele frequencies of these mutations in the overall population. These two observations make it less likely for these SNPs to drive significant changes in the tumor environment. Thereby, we continued our downstream analysis with *NEAT1* to assess the potential functional effects of this structural alteration. Of note, the detected structural change in *NEAT1* was not apparent in a computational-only structural assessment (Figure S2), reinforcing the need to combine experimental structural probing when assessing structural impact of mutations.

3.3 | Inferring potential functional consequences through interactome analysis

A total of 349 *NEAT1* interacting molecules supported by strong experimental evidence were retrieved from the RNAInter Database,[42] of which 322 were miRNAs, 14 mRNAs, one lncRNA, and 12 DNAs. To assess whether these interactions might be affected in the presence of the studied somatic mutation, we compared the interactions propensities with the wild type and mutant versions of *NEAT1*, using a set of computational tools, specific for each class of interacting molecules. RNAInter results included multiple molecules for some miRNA targets, four of the mRNA interactions were corresponding to HIV-1 and one of the DNA interactions included *PLOR2H*, for which we could not retrieve a sequence. Filtering these cases and removing redundancy resulted in 338 miRNAs, 10mRNAs, 1 lncRNA, and 11 DNAs (Dataset S3).

We evaluated the effect of the somatic mutation for 338 miRNA-*NEAT1* interactions with IntaRNA, using SHAPE-like reactivities from the nextPARS experiment to strengthen the computational prediction accuracy (see Materials and Methods). IntaRNA successfully predicted 328 and 330 *NEAT1* interacting miRNAs for the wild type and mutated molecules, respectively (Dataset S4). All of the interacting miRNA predictions found for *NEAT1* (wt) molecules overlapped with the *NEAT1* (mut) ones. Only hsa-miR-219b-5p and hsa-miR-744-3p interactions were predicted to interact with *NEAT1* (mut), but not *NEAT1* (wt). The number of predicted binding sites was compared between the *NEAT1* (wt) and *NEAT1* (mut) structures to check whether our candidate mutation can add or remove a binding site from the interaction. An identical number of binding sites were found for 302 *NEAT1*-miRNA interactions on both wt and mut molecules, and a differential effect was predicted for the remaining 28. These 28 interactions were further assessed in terms of the effects on the structure of free energy caused by the presence of the mutation using intaRNA (Dataset S4).

For the remaining 11 non-miRNA RNA interactions, we used lncTAR with normalized free energy cut off value of −0.1 as suggested by the authors.[46] LncTAR successfully predicted six *NEAT1* interactions for these 11 molecules in the regions surrounding our mutation of interest (Dataset S4). Importantly, three of these interactions (*CD36-NEAT*1, *SFPQ-NEAT1*, and *LSINCT5--NEAT1*) resulted in altered free energy values, suggesting potential disruption of the interaction strength. Of note, *CD36* and *SFPQ* have been involved in regulation, having tumor-promoting effects in CRC.[65–67] *LSINCT1* was also reported to be associated with negative prognosis in several cancers.[68–70]

For the 11 triplex-forming interactions of *NEAT1* with DNA molecules, we used LongTarget (see Material and Methods) to assess whether our candidate mutation could affect the binding propensity of triplex-forming oligonucleotides (TFO). To minimize the number of false negatives, we included the usage of all triplex formation rules within the algorithm. After generating the TFO sorted results, we extracted only class 1 (TFO1) predictions to increase confidence and compared the number of predicted TFO1 sites between the wild type and mutant structures. LongTarget successfully predicted TFO1 sites for all of these molecules. Out of these 11 predictions, four included identical amounts of TFO1 sites, while seven resulted in predicted alterations (Table 2): *FOLH1*, *MAPK15*, *RNF40*, *RPS24*, *SAP18*, *SP3*, *ZSCAN22* (Dataset S4).

To assess possible protein interactions occurring on the mutated region, we used ENCODE RNA Binding Protein Tracks and observed that an ELAVL1 (NP_001410.2) binding was reported.[48] We used PRIdictor with default settings to predict differential binding sites.[49] No differential binding sites were predicted for the wild and mutant sequences.

3.4 | Transcriptional effects of *NEAT1* somatic mutation

To evaluate whether the presence of the somatic mutation of interest had a measurable transcriptomic effect, we performed differential expression analyses using the data provided in the original publication.[37] We compared differences between tumor samples carrying the NEAT1

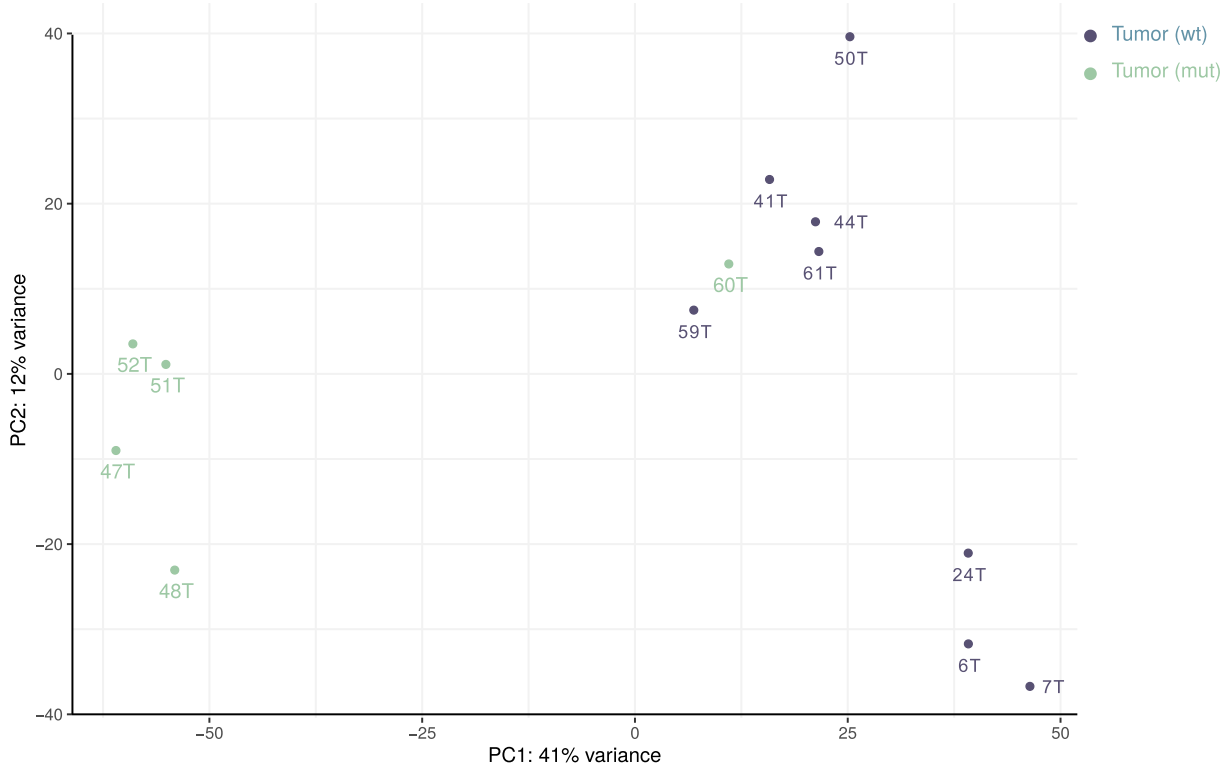

FIGURE 2 PCA plots generated using RNA-Seq data of tumors carrying NEAT1 (wt) or NEAT1 (mut) alleles.

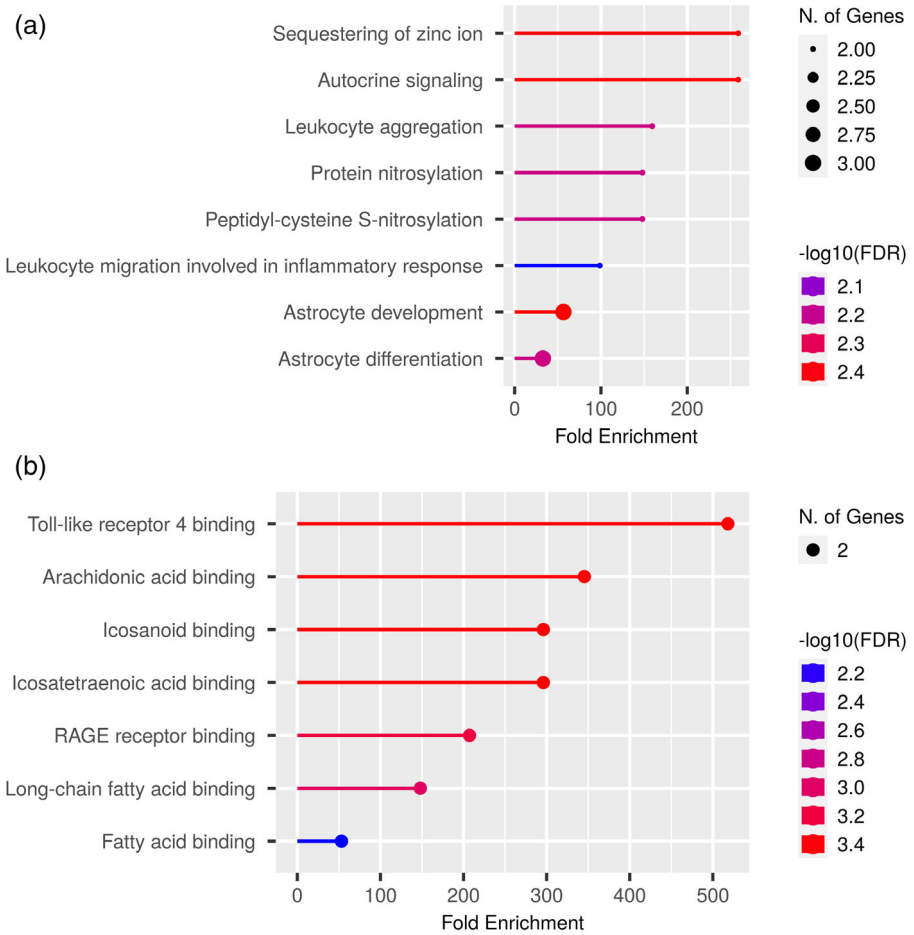

FIGURE 3 GO Enrichment results on upregulated genes for (a) Biological Processes (b) Molecular Functions. Legend represents the absolute number of genes in their corresponding pathways and -log10(FDR) values.

somatic mutation of interest and those without the alteration. Our results (Figure 2) suggest a strong transcriptional effect of the NEAT1 mutation, with most mutation-carrying tumors clustering together.

To investigate the transcriptional signature of the four clustered NEAT1 (mut) samples we performed a differential expression analysis by comparing their expression levels with the eight NEAT1(wt) tumors (see Materials and Methods). We identified 322 differentially expressed genes (Dataset S5). We performed functional enrichment analysis (see Materials and Methods, Figure 3), and found that genes upregulated in NEAT1 (mut) tumors are enriched in the following Gene Ontology (GO) Biological Process pathways: Sequestering of zinc ion, Autocrine signaling, Leukocyte aggregation, Protein nitrosylation, Peptidyl−cysteine S − nitrosylation, Leukocyte migration involved in inflammatory response, Astrocyte development and Astrocyte differentiation. Upregulated genes were also seen to be enriched in the following GO Molecular Functions; Toll−like receptor 4 binding, Arachidonic acid binding, Icosanoid binding, Icosatetraenoic acid binding, RAGE receptor binding, Long−chain fatty acid binding, and Fatty acid binding. No significant GO enrichment was found for downregulated genes.

We next compared differentially expressed genes against the candidate interacting genes derived from the previous analyses. To determine the candidate genes for miRNA interactions, we used emiRIT[54] and extracted 1829 entries found within CRC gene regulation. These entries were compared against our 28 miRNAs that were predicted to have altered lncRNA binding and the interacting genes from the intersecting ones were selected for differential expression analysis. Candidate genes derived from other RNA and DNA interactions are selected from lncTar and LongTarget results respectively. There was only one gene common within both sets: Vimentin (VIM), which is upregulated by 2.11 log fold change with a p adjusted value of .002. We think this finding is valuable because upregulation of VIM has been reported to be an indicator of poor prognosis in CRC.[71–73] VIM was found to be interacting with miR-200c[74] and miR-378,[75] which are both parts of the NEAT1 interactome. Our intaRNA predictions generated an extra binding site on the NEAT1 (mut) structure for miR-200a-5p and miR-378a-5p. However, intaRNA also generated a conflicting result for miR-378e, where an extra binding site was predicted for NEAT1 (wt) instead.

4 | DISCUSSION

Over the last decade, many lncRNAs have been discovered, and the evidence supporting their role in human diseases, including cancer, has been growing. Today, lncRNAs are seen as potential biomarkers and therapeutic targets in cancer.[76] Although their aberrant expression levels in cancers suggest a possible role in the disease, only a limited number of lncRNAs have been well characterized from a functional standpoint. Low evolutionary conservation and a complex interactome are two defining features of lncRNAs, which hamper their functional annotation. This study aimed to target these two factors via a hybrid approach to determine potential structural implications of lncRNA somatic mutations and their possible consequences for molecular interactions. Our approach aims to minimize false discovery rates by looking for multiple evidence levels and using stringent constraints.

Single nucleotide polymorphisms (SNP) and somatic mutations are known to affect the structure and expression levels of lncRNAs.[77] For this reason, identifying driver mutations on lncRNAs is crucial and pipelines such as ExInAtor began to emerge to address this issue[78]. Nevertheless, lncRNAs mostly fulfill their function through interacting with other molecules, and these interactions can be highly dependent on the lncRNA secondary structures.[23,79] Subtle changes in these interactions could lead to important outcomes. For these reasons, lncRNAs are sometimes mentioned as fine-tuners of gene regulation.[80,81] Here, we predict the potential effects of lncRNA somatic mutations associated with CRC relapse. In particular, we wanted to test the hypothesis that these somatic mutations were acting through alterations of the secondary structures of their corresponding lncRNAs.

A novel mutation along with seven control variants with high allele frequencies recurrently observed on CRC samples were tested for structural effects using the nextPARS experimental approach. Mutations found in *BLACAT1*, *LUCAT1*, *LINC01811*, and *PCAT1* carried high allele frequencies in the population and we did not observe any structural alteration. Considering both the high allele frequencies shown for these mutations on gnomAD and our nextPARS results, we conclude these variants represent non-deleterious polymorphisms and likely result from errors in the detection of somatic mutations. For *NEAT1*, however, our results showed that a novel A > G mutation (not previously reported in gnomAD), occurring on chr11:65441347, affects the structural preference from an unpaired to a paired state. Importantly, this effect was not detected when comparing the secondary structure information using only in-silico computational predictions, which underscores the utility of experimental probing approaches such as nextPARS.

NEAT1 has been previously shown to be associated with CRC through several mechanisms of action. *NEAT1*

is responsible for indirect activation of the Wnt/β-catenin signaling pathway via binding to DDX5 and KDM5A/Cul4A/Wnt axis.[26,82] In addition, *NEAT1* regulates several tumorigenesis-associated pathways via sponging multiple miRNAs.[83] Studies show that *NEAT1* is overexpressed in CRC -especially in metastatic cases- and high expression of this lncRNA is associated with poor prognosis.[84,85] We studied the 322 *NEAT1*-miRNA, 14 *NEAT1*-mRNA, 1 *NEAT1*-lncRNA, and 12 *NEAT1*-DNA interactions reported in the RNAInter database and based on strong experimental evidence as well as a protein interaction with ELAVL1 that is seen on ENCODE RNA Binding Proteins tracks. Our computational workflow (see Figure S3), described in the methods section, predicted that the somatic mutation results in binding propensity alterations for 28 miRNA, 5 mRNA, 1 RNA, and 7 DNA interactions. Consisting with these interactions having a regulatory effect, a re-analysis of available RNAseq data showed large differences in the transcriptional profiles of tumors carrying that *NEAT1* somatic mutation as compared to those lacking it. We assessed genes that were differentially expressed in tumors depending on the *NEAT1* mutation, and crossed these data with potential targets of miRNAs whose interactions with NEAT1 was predicted to be altered. This procedure solely identified the Vimentin-coding gene *VIM*, which was upregulated in tumors with *NEAT1* mutation and which interacts with *NEAT1*-interacting miRNAs *miR-200c*[74] and *miR-378*.[75] Vimentin is an epithelial-mesenchymal transition (EMT) marker which is aberrantly expressed in CRC and correlated with poor prognosis.[86,88] A recent study found that knockdown of *ZFAS1* lncRNA was associated with better prognosis in CRC and reduced Vimentin levels.[86] Authors showed that *ZFAS1* regulates epithelial-mesenchymal transition in CRC through reciprocal interaction with the *miR-200* family. Our intaRNA runs predict an extra binding site on *NEAT1*/*miR-200a-5p* for the mutated molecule. To our knowledge, there is no published work assessing the potential existence of *NEAT1/miR-200/VIM* axis in CRC. Two other altered binding sites were found for vimentin interacting miRNAs. One extra binding site for *miR-378a-5p* and one fewer binding site for *miR-378 e* were predicted for *NEAT1* (mut). These conflicting results made it harder for us to interpret the effect of the mutation for *NEAT1/miR-378* family interactions. Nevertheless, *miR-378* has been previously demonstrated to have a tumor suppressive role in CRC and the knockdown of *miR-378* has been associated with increased vimentin levels.[75] Similar expression patterns were also found on transgenic mice for the specific knockdown of *miR-378-5p*.[87] In addition, knockdown experiments on CRC cell lines indicate a strong coexpression pattern between *NEAT1* and *VIM* expressions.[84]

Finally, we observe an enrichment of the Toll-like Receptor 4 (TLR4) binding for the upregulated genes on *NEAT1* (mut) samples (see Figure 3b). Increased *TLR4* expression was previously observed in CRC patients as well as in cell lines.[89] Doan et al suggested that TLR4 regulates inflammation-mediated CRC prognosis by playing a role in the activation of PI3K/Akt pathway.[89] Links between TLR4 activation and CRC prognosis have also been supported through additional mechanisms that are associated with carcinogenesis such as increased Cox-2 expression and EGFR signaling.[90] As a summary TLR4 has a well-documented physiological role in CRC prognosis, making it a potential therapeutic target.

The upregulated genes that were involved in TLR4 Binding (GO:0035662) were *S100A9* and *S100A8*. *NEAT1* has been shown to regulate *S100A9* expression through miR-196a-5p sponging in rosacea.[91] Our intaRNA runs did not predict an alteration on the number of binding sites for miR-196a-5p between *NEAT1* (wt) and *NEAT1* (mut) molecules. We also checked the other possible miRNA interactions for *S100A9* but none was reported on RNAInter under strong experimental evidence constraint. Nevertheless, *S100A9* itself is considered a strong diagnostic biomarker for CRC[92] and we think that the 3.94 fold upregulation observed on *NEAT1* (mut) samples is an interesting finding which may indicate the existence of novel regulatory mechanisms between *NEAT1* and *S100A9*.

We did not predict any altered binding for ELAVL1, which is a protein that was previously demonstrated to bind the *NEAT1* region where our mutation occurs. However, several articles associate ELAVL1 association with CRC[93,94] and considering that many lncRNAs perform their functions through protein interactions[95] we think this interaction is also worth examining further.

In summary, we suspected that lncRNA mutations found to be enriched in CRC tumors from patients experiencing relapse and that showed low allelic frequencies in the overall population could lead to clinically relevant functional outcomes. We hypothesized that such mutations may trigger structural alterations on the RNA molecule to affect its interaction network. To our knowledge, our study represents the first empirical demonstration that a tumor-associated somatic mutation enriched in CRC samples with relapse after surgery induces a structural change on its corresponding lncRNA, which suggests that structural alterations may mediate functional changes and pathogenic impact. Downstream computational analysis on this novel *NEAT1* mutation revealed the potential effects of this alteration through multiple mechanisms. Although we implemented conservative selection constraints on every level of our analytical workflow, we acknowledge that our findings should

be experimentally validated. Our primary goal in this study was to demonstrate the strength of using a secondary structure-oriented approach on lncRNA somatic mutations on characterizing lncRNA behavior. We also underline the immense potential of structure-altering mutations on the fine-tuning role of lncRNAs through their complex interactome relationships. Altogether, we propose that our hybrid approach could be used to prioritize disease-associated lncRNAs and their mutations, and to generate testable hypotheses on possible mechanisms of action.

ORCID

Efe Aydın https://orcid.org/0000-0002-0380-5838
Toni Gabaldón https://orcid.org/0000-0003-0019-1735

REFERENCES

1. Wild CP, Weiderpass E, Stewart BW, editors. World cancer report: Cancer research for cancer prevention. Lyon, France: International Agency for Research on Cancer, 2020 Available from: http://publications.iarc.fr/586.
2. Duineveld LAM, van Asselt KM, Bemelman WA, et al. Symptomatic and asymptomatic colon cancer recurrence: A multicenter cohort study. Ann Fam Med. 2016;14(3):215–220. https://doi.org/10.1370/afm.1919.
3. Mekenkamp LJ, Koopman M, Teerenstra S, et al. Clinicopathological features and outcome in advanced colorectal cancer patients with synchronous vs metachronous metastases. Br J Cancer. 2010;103(2):159–164.
4. Jeffery M, Hickey BE, Hider PN. Follow-up strategies for patients treated for nonmetastatic colorectal cancer. Cochrane Database Syst Rev. 2019;9(1):CD002200.
5. Sargent DJ, Wieand HS, Haller DG, et al. Disease-free survival versus overall survival as a primary end point for adjuvant colon cancer studies: Individual patient data from 20,898 patients on 18 randomized trials. J Clin Oncol. 2005;23(34): 8664–8670.
6. Statello L, Guo CJ, Chen LL, Guo CJ, Chen LL, Huarte M. Gene regulation by long non-coding RNAs and its biological functions. Nat Rev Mol Cell Biol. 2021;22:96–118.
7. Ponting CP, Oliver PL, Reik W. Evolution and functions of long noncoding RNAs. Cell. 2009;136:629–641.
8. Pegueroles C, Iraola-Guzmán S, Chorostecki U, Ksiezopolska E, Saus E, Gabaldón T. Transcriptomic analyses reveal groups of co-expressed, syntenic lncRNAs in four species of the genus Caenorhabditis. RNA Biol. 2019;16:320–329. https://doi.org/10.1080/15476286.2019.1572438.
9. Jiang C, Li Y, Zhao Z, et al. Identifying and functionally characterizing tissue-specific and ubiquitously expressed human lncRNAs. Oncotarget. 2016;7(6):7120–7133. https://doi.org/10.18632/oncotarget.6859.
10. Cabili MN, Trapnell C, Goff L, et al. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genes Dev. 2011;25(18):1915–1927.
11. Gutschner T, Diederichs S. The hallmarks of cancer: A long non-coding RNA point of view. RNA Biol. 2012;9:703–719. [PMC free article] [PubMed] [Google Scholar].
12. Saus E, Brunet-Vega A, Iraola-Guzmán S, Pegueroles C, Gabaldón T, Pericay C. Long non-coding RNAs as potential novel prognostic biomarkers in colorectal cancer. Front Genet. 2016;7(54). https://doi.org/10.3389/fgene.2016.00054.
13. Ragusa M, Barbagallo C, Statello L, et al. Non-coding landscapes of colorectal cancer. World J Gastroenterol. 2015;21(41): 11709–11739. https://doi.org/10.3748/wjg.v21.i41.11709.
14. Hu Y, Chen HY, Yu CY, et al. A long non-coding RNA signature to improve prognosis prediction of colorectal cancer. Oncotarget. 2014;5(8):2230–2242. https://doi.org/10.18632/oncotarget.1895.
15. Ling H, Spizzo R, Atlasi Y, et al. CCAT2, a novel noncoding RNA mapping to 8q24, underlies metastatic progression and chromosomal instability in colon cancer. Genome Res. 2013; 23(9):1446–1461. https://doi.org/10.1101/gr.152942.112.
16. Gupta RA, Shah N, Wang KC, et al. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. Nature. 2010;464(7291):1071–1076. https://doi.org/10.1038/nature08975.
17. Garen A, Song X. Regulatory roles of tumor-suppressor proteins and noncoding RNA in cancer and normal cell functions. Int J Cancer. 2008;122(8):1687–1689. https://doi.org/10.1002/ijc.23285.
18. Kung JTY, Colognori D, Lee JT. Long noncoding RNAs: Past, present, and future. Genetics. 2013;193(3):651–669. https://doi.org/10.1534/genetics.112.146704.
19. Shan Y, Ma J, Pan Y, Hu J, Liu B, Jia L. LncRNA SNHG7 sponges miR-216b to promote proliferation and liver metastasis of colorectal cancer through upregulating GALNT1. Cell Death Dis. 2018;9:722. https://doi.org/10.1038/s41419-018-0759-7.
20. Yu Y, Nangia-Makker P, Farhana L, Majumdar APN. A novel mechanism of lncRNA and miRNA interaction: CCAT2 regulates miR-145 expression by suppressing its maturation process in colon cancer cells. Mol Cancer. 2017;16:155. https://doi.org/10.1186/s12943-017-0725-5.
21. Romero-Barrios N, Legascue MF, Benhamed M, Federico Ariel F, Crespi M. Splicing regulation by long non-coding RNAs. Nucleic Acids Res. 2018;46:2169–2184. [CrossRef] [PubMed].
22. Karapetyan AE, Coen Buiting C, Kuiper RA, Marcel W, Coolen MW. Regulatory roles for long ncRNA and mRNA. Cancer. 2013;5:462–490. [CrossRef] [PubMed].
23. Kazimierczyk M, Kasprowicz MK, Kasprzyk ME, Wrzesinski J. Human long noncoding RNA interactome: Detection, characterization and function. Int J Mol Sci. 2020;21(3):1027. https://doi.org/10.3390/ijms21031027.
24. Grote P, Herrmann BG. The long non-coding RNA Fendrr links epigenetic control mechanisms to gene regulatory networks in mammalian embryogenesis. RNA Biol. 2013;10:1579–1585.
25. Xu T, Huang M, Xia R, et al. Decreased expression of the long non-coding of FENDRR is associated with poor prognosis in gastric cancer and FENDRR regulates gastric cancer cell metastasis by a_ecting fibronectin1 expression. J Hematol Oncol. 2014;7:63 [CrossRef] [PubMed].
26. Zhang M, Weng W, Zhang Q, et al. The lncRNA NEAT1 activates Wnt/β-catenin signaling and promotes colorectal cancer progression via interacting with DDX5. J Hematol Oncol. 2018; 11(1):113. https://doi.org/10.1186/s13045-018-0656-7.

27. Necsulea A, Kaessmann H. Evolutionary dynamics of coding and non-coding transcriptomes. Nat Rev Genet. 2014;15:734–748. [CrossRef] [PubMed].
28. Pegueroles C, Gabaldón T. Secondary structure impacts patterns of selection in human lncRNAs. BMC Biol. 2016;14:60. https://doi.org/10.1186/s12915-016-0283-0.
29. Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. 2003;31:3406–3415.
30. Gruber AR, Bernhart SH, Lorenz R. The ViennaRNA web services. Methods Mol Biol. 2015;1269:307–326.
31. Bellaousov S, Reuter JS, Seetin MG, Mathews DH. RNAstructure: Web servers for RNA secondary structure prediction and analysis. Nucleic Acids Res. 2013;41:W471–W474.
32. Janssen S, Giegerich R. The RNA shapes studio. Bioinformatics. 2015;31:423–425.
33. Zhao Y, Wang J, Zeng C, Xiao Y. Evaluation of RNA secondary structure prediction for both base-pairing and topology. Biophys Rep. 2018;4:123–132. https://doi.org/10.1007/s41048-018-0058-y.
34. Bugnon LA, Edera AA, Prochetto S, et al. Secondary structure prediction of long noncoding RNA: Review and experimental comparison of existing approaches. Brief Bioinform. 2022;23: bbac205. https://doi.org/10.1093/bib/bbac205.
35. Saus E, Willis JR, Pryszcz LP, et al. nextPARS: Parallel probing of RNA structures in Illumina. RNA (New York, N.Y.). 2018; 24(4):609–619. https://doi.org/10.1261/rna.063073.117.
36. Chorostecki U, Willis JR, Saus E, Gabaldon T. Profiling of RNA structure at single-nucleotide resolution using nextPARS. Methods Mol Biol (Clifton, N.J.). 2021;2284:51–62. https://doi.org/10.1007/978-1-0716-1307-8_4.
37. Iraola-Guzmán S, Brunet-Vega A, Pegueroles C, et al. Target enrichment enables the discovery of lncRNAs with somatic mutations or altered expression in paraffin-embedded colorectal cancer samples. Cancer. 2020;12(10):2844. https://doi.org/10.3390/cancers12102844.
38. Cuykendall TN, Rubin MA, Khurana E. Non-coding genetic variation in cancer. Curr Opin Syst Biol. 2017;1:9–15. https://doi.org/10.1016/j.coisb.2016.12.017.
39. O'Connell MJ, Campbell ME, Goldberg RM, et al. Survival following recurrence in stage II and III colon cancer: Findings from the ACCENT data set. J Clin Oncol Off J Am Soc Clin Oncol. 2008; 26(14):2336–2341. https://doi.org/10.1200/JCO.2007.15.8261.
40. Karczewski KJ, Francioli LC, Tiao G, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature. 2020;581:434–443.
41. Chorostecki U, Saus E, Gabaldón T. Structural characterization of NORAD reveals a stabilizing role of spacers and two new repeat units. Comput Struct Biotechnol J. 2021;19:3245–3254. https://doi.org/10.1016/j.csbj.2021.05.045.
42. Kang J, Tang Q, He J, et al. RNAInter v4.0: RNA interactome repository with redefined confidence scoring system and improved accessibility. Nucleic Acids Res. 2022;50(D1):D326–D332. https://doi.org/10.1093/nar/gkab997.
43. Mann M, Wright PR, Backofen R. IntaRNA 2.0: Enhanced and customizable prediction of RNA-RNA interactions. Nucleic Acids Res. 2017;45(W1):W435–W439. https://doi.org/10.1093/nar/gkx279 PMID: 28472523; PMCID: PMC5570192.
44. Raden M, Müller T, Mautner S, Gelhausen R, Backofen R. The impact of various seed, accessibility and interaction constraints on sRNA target prediction—A systematic assessment. BMC Bioinform. 2020;21:15. https://doi.org/10.1186/s12859-019-3143-4.
45. Afgan E, Baker D, Batut B, et al. The galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. Nucleic Acids Res. 2018;46(W1):W537–W544. https://doi.org/10.1093/nar/gky379.
46. Li J, Ma W, Zeng P, et al. LncTar: A tool for predicting the RNA targets of long noncoding RNAs. Brief Bioinform. 2015; 16(5):806–812. https://doi.org/10.1093/bib/bbu048.
47. He S, Zhang H, Liu H, Zhu H. LongTarget: A tool to predict lncRNA DNA-binding motifs and binding sites via Hoogsteen base-pairing analysis. Bioinformatics (Oxford, England). 2015;31(2):178–186. https://doi.org/10.1093/bioinformatics/btu643.
48. Rosenbloom KR, Sloan CA, Malladi VS, et al. ENCODE data in the UCSC genome browser: Year 5 update. Nucleic Acids Res. 2013;41(Database issue):D56–D63.
49. Tuvshinjargal N, Lee W, Park B, Han K. PRIdictor: Protein-RNA interaction predictor. Biosystems. 2016;139:17–22. https://doi.org/10.1016/j.biosystems.2015.10.004.
50. Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ. miRBase: microRNA sequences, targets and gene nomenclature. Nucleic Acids Res. 2006;34(Database issue): D140–D144. https://doi.org/10.1093/nar/gkj112.
51. Cunningham F, Allen JE, Allen J, et al. Ensembl 2022. Nucleic Acids Res. 2022;50(1):D988–D995 PubMed PMID: 34791404. https://doi.org/10.1093/nar/gkab1049.
52. NCBI Resource Coordinators. Database resources of the National Center for biotechnology information. Nucleic Acids Res. 2016;44(D1):D7–D19. https://doi.org/10.1093/nar/gkv1290.
53. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15:550. https://doi.org/10.1186/s13059-014-0550-8.
54. Roychowdhury D, Gupta S, Qin X, Arighi CN, Vijay-Shanker K. emiRIT: A text-mining-based resource for microRNA information. Database J Biol Databases Curation. 2021; 2021:baab031. https://doi.org/10.1093/database/baab031.
55. Ge SX, Jung D, Yao R, Shiny GO. A graphical gene-set enrichment tool for animals and plants. Bioinformatics. 2020;36(8): 2628–2629. https://doi.org/10.1093/bioinformatics/btz931.
56. Korotkevich G, Sukhov V, Sergushichev A. Fast gene set enrichment analysis. bioRxiv. 2019. https://doi.org/10.1101/060012 http://biorxiv.org/content/early/2016/06/20/060012.
57. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005;102:15545–15550.
58. Pan L, Du M, Liu H, et al. LncRNA FTX promotes the malignant progression of colorectal cancer by regulating the miR-214-5p-JAG1 axis. Ann Transl Med. 2021;9(17):1369.
59. Chen GQ, Liao ZM, Liu J, Li F, Huang D, Zhou YD. LncRNA FTX promotes colorectal cancer cells migration and invasion by miRNA-590-5p/RBPJ axis. Biochem Genet. 2021;59(2):560–573. https://doi.org/10.1007/s10528-020-10017-8.
60. Sherry ST, Ward M, Sirotkin K. dbSNP—Database for single nucleotide polymorphisms and other classes of minor genetic variation. Genome Res. 1999;9:677–679.

61. Tate JG, Bamford S, Jubb HC, et al. COSMIC: The catalogue of somatic mutations in cancer. Nucl Acids Res. 2019;47(D1): D941–D947. https://doi.org/10.1093/nar/gky1015.
62. Bu C, Wang B, Zhang D, Mao Z, Pu C. Overexpression of the long non-coding RNA BLACAT1 promotes cell proliferation and invasion in colorectal cancer. Transl Cancer Res. 2019;8(1): 35–43. https://doi.org/10.21037/tcr.2018.12.26.
63. Ge X, Chen Y, Liao X, et al. Overexpression of long noncoding RNA PCAT-1 is a novel biomarker of poor prognosis in patients with colorectal cancer. Med Oncol (Northwood, London, England). 2013;30(2):588. https://doi.org/10.1007/s12032-013-0588-6.
64. Wu R, Li L, Bai Y, et al. The long noncoding RNA LUCAT1 promotes colorectal cancer cell proliferation by antagonizing Nucleolin to regulate MYC expression. Cell Death Dis. 2020; 11(10):908. https://doi.org/10.1038/s41419-020-03095-4.
65. Ji Q, Zhang L, Liu X, et al. Long non-coding RNA MALAT1 promotes tumour growth and metastasis in colorectal cancer through binding to SFPQ and releasing oncogene PTBP2 from SFPQ/PTBP2 complex. Br J Cancer. 2014;111(4):736–748. https://doi.org/10.1038/bjc.2014.383.
66. Drury J, Rychahou PG, He D, et al. Inhibition of fatty acid synthase upregulates expression of CD36 to sustain proliferation of colorectal cancer cells. Front Oncol. 2020;10(1185):31. https://doi.org/10.3389/fonc.2020.01185.
67. Fang Y, Shen ZY, Zhan YZ, et al. CD36 inhibits β-catenin/c-myc-mediated glycolysis through ubiquitination of GPC4 to repress colorectal tumorigenesis. Nat Commun. 2019;10:3981. https://doi.org/10.1038/s41467-019-11662-3.
68. Long X, Li L, Zhou Q, et al. Long non-coding RNA LSINCT5 promotes ovarian cancer cell proliferation, migration and invasion by disrupting the CXCL12/CXCR4 signalling axis. Oncol Lett. 2018;15(5):7200–7206. https://doi.org/10.3892/ol.2018.8241.
69. Zhu X, Li Y, Zhao S, Zhao S. LSINCT5 activates Wnt/β-catenin signaling by interacting with NCYM to promote bladder cancer progression. Biochem Biophys Res Commun. 2018;502(3):299–306. https://doi.org/10.1016/j.bbrc.2018.05.076.
70. Qi P, Lin W-r, Zhang M, et al. E2F1 induces LSINCT5 transcriptional activity and promotes gastric cancer progression by affecting the epithelial-mesenchymal transition. Cancer Manag Res. 2018;10:2563–2571. https://doi.org/10.2147/CMAR.S171652.
71. Al-Maghrabi J. Vimentin immunoexpression is associated with higher tumor grade, metastasis, and shorter survival in colorectal cancer. Int J Clin Exp Pathol. 2020;13,3:493–500.
72. Wang Q, Zhu G, Lin C, et al. Vimentin affects colorectal cancer proliferation, invasion, and migration via regulated by activator protein 1. J Cell Physiol. 2021;236(11):7591–7604. https://doi.org/10.1002/jcp.30402.
73. Niknami Z, Muhammadnejad A, Ebrahimi A, Harsani Z, Shirkoohi R. Significance of E-cadherin and vimentin as epithelial-mesenchymal transition markers in colorectal carcinoma prognosis. EXCLI J. 2020;19:917–926. https://doi.org/10.17179/excli2020-1946.
74. Hur K, Toiyama Y, Takahashi M, et al. MicroRNA-200c modulates epithelial-to-mesenchymal transition (EMT) in human colorectal cancer metastasis. Gut. 2013;62(9):1315–1326. https://doi.org/10.1136/gutjnl-2011-301846.
75. Zhang G-J, Zhou H, Xiao H-x, Li Y, Zhou T. MiR-378 is an independent prognostic factor and inhibits cell growth and invasion in colorectal cancer. BMC Cancer. 2014;14(109):20. https://doi.org/10.1186/1471-2407-14-109.
76. Chandra Gupta S, Nandan Tripathi Y. Potential of long non-coding RNAs in cancer patients: From biomarkers to therapeutic targets. Int J Cancer. 2017;140(9):1955–1967. https://doi.org/10.1002/i.
77. Minotti L, Agnoletto C, Baldassari F, Corrà F, Volinia S. SNPs and somatic mutation on long non-coding RNA: New frontier in the cancer studies? High-Throughput. 2018;7(4):34. https://doi.org/10.3390/ht7040034.
78. Lanzós A, Carlevaro-Fita J, Mularoni L, et al. Discovery of cancer driver long noncoding RNAs across 1112 tumour genomes: New candidates and distinguishing features. Sci Rep. 2017;7: 41544. https://doi.org/10.1038/srep41544.
79. Somarowthu S, Legiewicz M, Chillón I, Marcia M, Liu F, Pyle AM. HOTAIR forms an intricate and modular secondary structure. Mol Cell. 2015;58(2):353–361. https://doi.org/10.1016/j.molcel.2015.03.006.
80. Sanchez Calle A, Kawamura Y, Yamamoto Y, Takeshita F, Ochiya T. Emerging roles of long non-coding RNA in cancer. Cancer Sci. 2018;109(7):2093–2100. https://doi.org/10.1111/cas.13642.
81. Adriaens C, Marine J-C. NEAT1-containing paraspeckles: Central hubs in stress response and tumor formation. Cell Cycle (Georgetown, Tex). 2017;16(2):137–138. https://doi.org/10.1080/15384101.2016.1235847.
82. Shen X, Ye Z, Wu W, et al. lncRNA NEAT1 facilitates the progression of colorectal cancer via the KDM5A/Cul4A and Wnt signaling pathway. Int J Oncol. 2021;59(1):51. https://doi.org/10.3892/ijo.2021.5231.
83. Li K, Yao T, Zhang Y, Li W, Wang Z. NEAT1 as a competing endogenous RNA in tumorigenesis of various cancers: Role, mechanism and therapeutic potential. Int J Biol Sci. 2021; 17,13:3428–3440. https://doi.org/10.7150/ijbs.62728.
84. Yu H-M, Wang C, Yuan Z, Chen G-L, Ye T, Yang B-W. LncRNA NEAT1 promotes the tumorigenesis of colorectal cancer by sponging miR-193a-3p. Cell Prolif. 2019;52(1):e12526. https://doi.org/10.1111/cpr.12526.
85. Azizidoost S, Ghaedrahmati F, Anbiyaee O, Ahmad Ali R, Cheraghzadeh M, Farzaneh M. Emerging roles for lncRNA-NEAT1 in colorectal cancer. Cancer Cell Int. 2022;22:209. https://doi.org/10.1186/s12935-022-02627-6.
86. O'Brien SJ, Fiechter C, Burton J, et al. Long non-coding RNA ZFAS1 is a major regulator of epithelial-mesenchymal transition through miR-200/ZEB1/E-cadherin, vimentin signaling in colon adenocarcinoma. Cell Death Discov. 2021;7:61. https://doi.org/10.1038/s41420-021-00427-x.
87. Li H, Chang L, du WW, et al. Anti-microRNA-378a enhances wound healing process by eupregulating integrin beta-3 and vimentin. Mol Therapy J Am Soc Gene Ther. 2014;22(10):1839–1850. https://doi.org/10.1038/mt.2014.115.
88. Toiyama Y, Yasuda H, Saigusa S, et al. Increased expression of slug and vimentin as novel predictive biomarkers for lymph node metastasis and poor prognosis in colorectal cancer. Carcinogenesis. 2013;34(11):2548–2557. https://doi.org/10.1093/carcin/bgt282.

89. Doan HQ, Bowen KA, Jackson LA, Evers BM. Toll-like receptor 4 activation increases Akt phosphorylation in colon cancer cells. Anticancer Res. 2009;29(7):2473–2478.
90. Fukata M, Chen A, Vamadevan AS, et al. Toll-like receptor-4 promotes the development of colitis-associated colorectal tumors. Gastroenterology. 2007;133(6):1869–1881. https://doi.org/10.1053/j.gastro.2007.09.008.
91. Wang L, Wang Y-J, Hao D, et al. Long non-coding RNA NEAT1 functions as a competing endogenous RNA to regulate S100A9 expression by sponging miR-196a-5p in rosacea. J Dermatol Sci. 2021;102(1):58–67. https://doi.org/10.1016/j.jdermsci.2021.02.005.
92. Zhou M, Li M, Liang X, et al. The significance of serum S100A9 and TNC levels as biomarkers in colorectal cancer. J Cancer. 2019;10,22:5315–5323. https://doi.org/10.7150/jca.31267.
93. Li K, Huang F, Li Y, et al. Stabilization of oncogenic transcripts by the IGF2BP3/ELAVL1 complex promotes tumorigenicity in colorectal cancer. Am J Cancer Res. 2020;10(8):2480–2494.
94. Gu C, Zhang M, Sun W, Dong C. Upregulation of miR-324-5p inhibits proliferation and invasion of colorectal cancer cells by targeting ELAVL1. Oncol Res. 2019;27(5):515–524. https://doi.org/10.3727/096504018X15166183598572.
95. Redis RS, Calin GA. The interplay between lnRNAs, SNPs, and protein complexes - what does it mean for cancer metabolism? Mol Cell Oncol. 2016;3(4):e1166308. https://doi.org/10.1080/23723556.2016.1166308.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Aydın E, Saus E, Chorostecki U, Gabaldón T. A hybrid approach to assess the structural impact of long noncoding RNA mutations uncovers key *NEAT1* interactions in colorectal cancer. IUBMB Life. 2023. https://doi.org/10.1002/iub.2710