

RESEARCH ARTICLE

Oncogenic fusion transcript analysis identified ADAP1-NOC4L, potentially associated with metastatic colorectal cancer

Amin Talebi^{1,2} | Soodabeh Shahidsales³ | Mohsen Aliakbarian⁴ |
Masoud Pezeshki Rad⁵ | Mohammad Amin Kerachian^{1,2,6} 

¹Medical Genetics Research Center, Mashhad University of Medical Sciences, Mashhad, Iran

²Faculty of Medicine, Department of Medical Genetics, Mashhad University of Medical Sciences, Mashhad, Iran

³Cancer Research Center, Mashhad University of Medical Sciences, Mashhad, Iran

⁴Faculty of Medicine, Surgical Oncology Research Center, Mashhad University of Medical Sciences, Mashhad, Iran

⁵Faculty of Medicine, Department of Radiology, Mashhad University of Medical Sciences, Mashhad, Iran

⁶Cancer Genetics Research Unit, Reza Radiotherapy and Oncology Center, Mashhad, Iran

Correspondence

Soodabeh Shahidsales, Cancer Research Center, Mashhad University of Medical Sciences, Mashhad, Iran.

Email: sahidsales@mums.ac.ir

Mohammad Amin Kerachian, Department of Medical Genetics, Faculty of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran.

Email: amin.kerachian@mail.mcgill.ca

Funding information

Mashhad University of Medical Sciences, Grant/Award Number: 981199

Abstract

Purpose: Fusion transcripts are transcriptome-mediated alterations involved in tumorigenesis and are considered as diagnostic, prognostic, and therapeutic biomarkers. In metastatic colorectal carcinoma (mCRC), fusion transcripts are rarely reported. The main challenge is to identify driver chimeras with a significant role in cancer progression.

Methods: In the present study, 86 RNA sequencing data samples were analyzed to discover driver fusion transcripts. Functional assays included clonogenic cell survival, wound-healing, and transwell cell invasion. Quantitative expression analysis of epithelial-mesenchymal transition (EMT), apoptotic regulators, and metastatic markers were examined for the candidate fusion genes. Kaplan–Meier survival analysis was performed using patient overall survival (OS).

Results: A variety of driver fusions were identified. Fourteen fusion genes (51% of mCRC), each at least found in two mCRC samples, were determined as oncogenic fusion transcripts by in silico analysis of their functions. Among them, two recurrent chimeric transcripts confirmed by Sanger sequencing were selected. Positive expression of *ADAP1-NOC4L* was significantly associated with an increased risk of poor OS in mCRC patients. In vitro transforming potential for the chimera, resulting from the fusion of *ADAP1* and *NOC4L* was assessed. Overexpression of this fusion gene increased cell proliferation and enhanced migration and invasion of CRC cells. In addition, it significantly upregulated EMT and anti-apoptotic markers.

Conclusions: *ADAP1-NOC4L* transcript chimera, a driver chimera identified in this study, provides new insight into the underlying mechanisms involved in the development and spread of mCRC. It suggests the potential of RNA-based alterations as novel targets for personalized medicine in clinical practice.

KEYWORDS

functional study, fusion transcripts, metastatic colorectal cancer, RNA sequencing, transcript chimera

1 | INTRODUCTION

Colorectal cancer (CRC) is one of the primary causes of cancer-related morbidity and mortality around the world.^{1,2} The metastatic spread of tumor cells to the liver, the most common target of tumor cell dissemination, occurs eventually in most patients with primary CRC, with median overall survival (OS) of 5–20 months, if not treated.^{3,4}

In the last two decades, with the introduction of novel therapeutic methods such as anti-angiogenic and anti-Epidermal Growth Factor Receptor (EGFR),⁵ the resectability rates of patients' tumors with liver metastases have impressively improved.^{6,7} However, the problem that which patients with resectable disease and high-risk features benefit from adjuvant therapy still remain unknown.⁸ As a result, predictive biomarkers of chemotherapeutic efficacy are needed to select the appropriate metastatic colorectal cancer (mCRC) treatment. They could lead to better selection of patients for treatment options, as well as predicting tumors with a higher aggressiveness and those resistant to treatments.^{9–11} In addition, the discovery of additional genetic events should provide new cancer-related biomarkers important in early diagnosis, and prognosis. On the other hand, in order to generate more promising results in the decision-making process of mCRC patients, a personalized approach regarding molecular profiling is required.⁹ As we move closer to precision medicine, new molecular abnormalities are discovered as drivers for tumor initiation and development, potentially revealing novel therapeutic targets.¹² Recent advances over the last decade in tumor genomic testing have made it possible to extensively determine the mCRC molecular landscape.¹³

In this regard, fusion transcripts are transcriptome-mediated rearrangements which had been shown to play a role in the development of several malignancies.¹⁴ In many cases, the identification of fusion transcripts has diagnostic values (e.g., *FLI1/EWS* in Ewing Sarcoma) as determining a particular tumor subtype. They could also predict prognosis (e.g., the presence of gene fusions in embryonal rhabdomyosarcoma), or might be of therapeutic importance [e.g., *ALK* and *ROS1* in non-small cell lung cancer (*NSCLC*)].¹¹

In mCRC, several fusions including *NTRK*, *ALK*, *ROS*, *RET*, *BRAF* fusions with prognostic, predictive, or druggable targets potential have been defined, so far^{15–19} However, a comprehensive picture of effective fusion transcripts in the pathogenesis and development of mCRC, is not well understood.

In research practice, discovering chimeras involves two main aims and approaches: (1) Identifying targetable fusions associated with known therapeutic agents regardless of tumor type and (2) detailed molecular examination of a tumor to discover specific tumor alterations that require appropriately targeted treatments.²⁰ The critical issue in these two scenarios is the low frequency of discovered chimeras specially among solid tumors, which limits their pathogenetic and therapeutic relevance in trial studies and their applications in clinical settings.²¹ It should be noted that in some cases the presence of these fusions defines certain subtypes of the tumor that may benefit from specific treatments and should not be ignored. For example, *RET* rearrangements, in particular, characterize a subgroup of mCRC that is resistant to conventional anti-EGFR treatments but may respond to *RET* inhibitors.²²

Recurrent chimera appears to play a more significant role in disease pathogenesis; however, The critical aspect in these instances is that the driver chimeras must be distinguished from passengers, as they play essential roles in tumor development and progression.¹¹ The biological function of the genes involved, as well as the *in vitro* and *in vivo* characterization of chimera functions, could be determined as discriminating factors between driver and passenger chimera.¹⁴ A more comprehensive range of malignancies could be analyzed by using public databases, allowing to discover more significant recurrent molecular changes.

The present study aims to investigate recurrent driver fusion transcripts at the transcriptome level to discover potentially important chimeras in mCRC pathogenesis by analyzing RNA sequencing raw data from multiple public datasets. Here, we assess the driving function which is relevance of candidate cases at an *in-silico* level by evaluating functional domains in selected cases, followed by functional establishment *in vitro* on the candidate chimeras. In addition, the influence of candidate chimeras' expression on patient prognosis are also investigated.

2 | METHODS

2.1 | Data collection and samples

Paired-end RNA sequencing fastq files were downloaded from *Sequence Read Archive (SRA)* database (<https://www.ncbi.nlm.nih.gov/sra>) of the National Center of

Biotechnology Information (NCBI) with accession numbers: SRP060016, SRP095672, SRP029880, SRP078268, SRP222902. Eighty-six RNA sequencing data obtained from 24 paired liver metastasis, primary tumors, and normal specimens as well as seven primary tumors with seven liver metastases were considered for initial analysis. In addition, 17 fresh tissues (11 liver metastasis, and 6 primary tumors) with 23 formalin-fixed paraffin embedded (FFPE) tissues (12 liver metastasis and 11 primary tumors) from the archives of the Iranian biobank of Mashhad Sina and Imam Reza hospitals were included.

For FFPE tissues, briefly, sample areas with at least 90% tumor cells, without mucin-rich glands or prominent inflammatory cellular infiltration were selected under microscopic examination on each glass slide and the matched area of paraffin-embedded tissues were selected and re-embedded. Then, up to four sections of FFPE tissues were cut, each with a thickness of up to 10 micrometers. Ethical approval for this study was given by the ethics committee of the Mashhad University of Medical Sciences, Mashhad, Iran (Approval ID: IR.MUMS.MEDICAL.REC.1399.106).

2.2 | Chimeric transcript detection

For fusion transcript detection, we used a combination of criteria's followed by a filtering pipeline (Figure 1). Four computational gene fusion detection tools were applied for fusion transcript investigation and cross-validation, including Arriba,²³ CLC genomic workbench 20 with a plugin¹⁶,²⁴ SOAPfuse-v1.27,²⁵ and defuse-0.6.2,²⁶ which each has its own fusion detection algorithm and aligner. The pipelines were tuned to detect fusion transcripts in CLC genomic workbench 20 and SOAPfuse.

In order to find important potential protein-coding fusion transcripts that are overexpressed in metastatic cells and potentially might be important in the development of metastatic clones, we used a filtering pipeline. The fusion transcripts that are significantly upregulated in metastatic specimens in comparison to primary CRC specimens, were included. Other inclusion criteria were the alignment of both fusion partners to the protein-coding sequence and the fusion spanning reads with or without fusion crossing read, accepted greater than or equal to 5. Fusion transcripts that were out of the frame or also present in the normal colon tissue were excluded. In-frame fusion transcripts were selected from the initial candidates, which were also determined by using AGFusion software²⁷ or individually by checking each frame of the chimeric transcript separately.

We investigated the available literature and the fusion gene databases including TCGA,²⁸ FusionGDB,²⁹ ChimerDB 4.0,³⁰ and Mitelman³¹ databases to recognize

the novelty of the discovered fusions. Final candidate chimeras were evaluated to preserve driver domains by Oncofuse³² or a persistent open reading frame containing functional domains such as kinases. Human hg19/GRCH37 reference genome was considered for RNA-Seq reads alignment.

2.3 | Chimera validation

For fusion transcript validation in colon cancer cell lines and clinical tissue samples, quantitative reverse transcription polymerase chain reaction (RT-qPCR) was carried out using gene specific primers spanning the fusion junction region, described in Table S1. The PCR products were analyzed on a 2% agarose gel and confirmed using Sanger sequencing with PCR amplification primers.

Total RNA was isolated using AccuZol™ Total RNA extraction kit (Bioneer Corporation, South Korea) from fresh tissues and SW48 CRC cell lines, purchased from Pasteur Institute of Iran. To extract RNA from FFPE tissues, RNeasy FFPE Kit (QIAGEN) was used. Subsequently, 1 µg of total RNA was reverse transcribed with AccuPower® RocketScript™ RT PreMix Kit (Bioneer Corporation).

PCR reactions were conducted for 5 min at 95°C, and 40 cycles for 30s at 95°C, 30s at melting temperature (TM) according to each primer set, 30s at 72°C, and 10 min at 72°C. RT-qPCR was performed with the same method to validate the transfection status of the transfected vectors into the CRC cell line. Furthermore, RT-qPCR was performed for matrix metalloproteinase 9 (MMP9), N-cadherin, fibronectin 1 (FN1), Vimentin, Bcl2, and BAX genes with specific primers designed for gene transcripts to evaluate the expression level of epithelial to mesenchymal transition (EMT) and metastatic as well as regulators of apoptosis biomarkers, respectively (Table S2). Quantitative RT-PCR was performed by using the same cDNA referred above, and DNA was amplified using SYBRGreen, using LightCycler® 96 System (Roche, Germany). Relative expression levels of fusion transcripts were determined using the $2^{-\Delta\Delta Ct}$ method according to the MIQE guidelines.³³ GAPDH gene expression was used as a reference gene for data normalization.

2.4 | Construction of the ADAP1-NOC4L encoding vector

Coding sequences of ADAP1-NOC4L gene was used to construct vector for the chimera overexpression. After amplification in the cells containing this transcript using the forward primer 5'-ATGGCCAAGGAGCGGC-3' and reverse primer

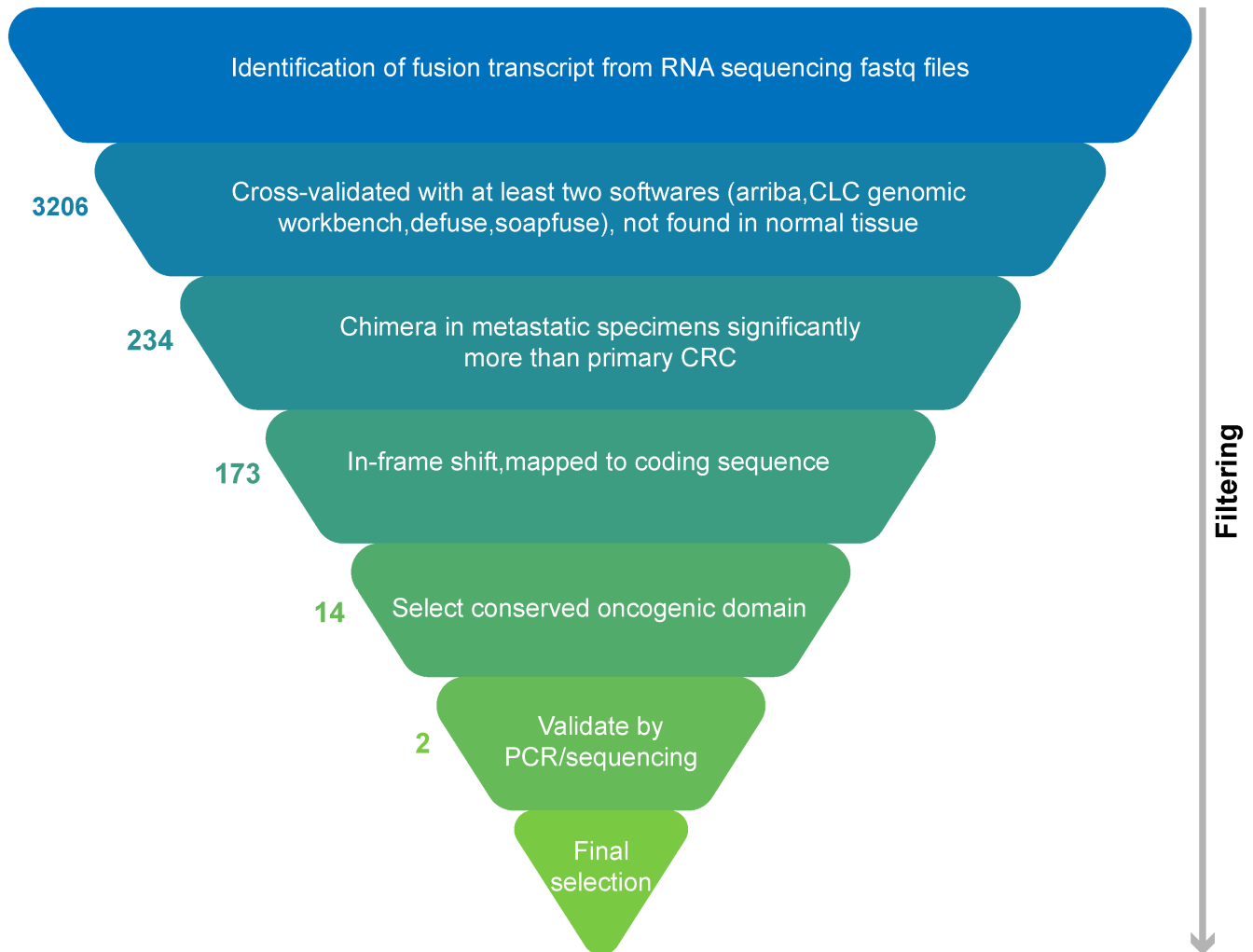


FIGURE 1 An overview representing the fusion transcript identification filtering pipeline from paired-end RNA sequencing fastq data in metastatic tissue samples.

5'-TCAGCTGAGCGTGAAGTGC-3' from cDNAs isolated from SW48 cells, the sequence was cloned into pcDNA3.1/V5-His B vector by one-step ligation method.

2.5 | Functional validation

For in vitro functional assessment of ADAP1-NOC4L multiple assays including clonogenic cell survival assay, Wound-healing assay, Transwell invasion assay, PI Annexin V apoptosis assay, and fusion expression analysis were performed. The detailed description of assays used are explained in file S1.

2.6 | Chimera expression comparison analysis

The relative expression level of candidate chimera of sequentially validated positive samples in metastatic tissue

compared to primary CRC were investigated by qPCR in 17 fresh tissues (11 liver metastasis and 6 primary tumors) and 23 formalin-fixed paraffin-embedded (FFPE) tissues (12 liver metastasis and 11 primary tumors).

2.7 | Survival analysis for candidate genes

We examined the effect of the selected chimera expression levels on the OS rate of patients with primary and metastatic colorectal cancer.

2.8 | Statistical analysis

Gene expression data are presented as the mean \pm standard error of the mean. Data distribution was evaluated using the D'Agostino test. Comparison between two groups was conducted using the Student's *t* test and Mann-Whitney

test. Kaplan–Meier curves and the log-rank test were generated to assess the survival data. p value <0.05 was considered to indicate a statistically significant difference. Statistical analysis was performed using SPSS software version 22 (IBM Corp.)³⁴ and GraphPad Prism software version 8 (GraphPad software, Inc.).³⁵

3 | RESULTS

3.1 | Prediction of oncogenic chimeras in metastasis

Oncofuse software was applied to investigate preserved oncogenic domains in chimeras with the continuous open reading frame (ORF), which uses a Bayesian machine learning algorithm. Fusion transcripts with high driver scores that included exonic or coding regions were considered. In other cases, candidate driver fusion selection included cases in which at least one of the two fusion partners had a conserved oncogenic domain based on domain analysis with InterPro or Uniprot online tools, or the results of previous fusion transcript studies.

By using filtering pipeline criteria, we discovered a total of 3206 cross-validated fusion transcripts with at least two out of four fusion detection tools. Among these fusion transcripts, 1617 were exclusively in metastatic tissues with 1245 mapped to the coding sequences, and 136 located in continuous ORFs (Table 1). Our results revealed that in some paired specimens (primary & metastasis CRC) the fusion was exclusively expressed in metastasis. Fourteen fusion transcripts (51% of mCRC) differentially upregulated in metastatic tissue were found to contain at least one preserved oncogenic domain or driver score >0.8 , analyzed by Oncofuse (Table 2).

Finally, five candidates fusion transcripts in 12 metastatic tissue samples (39% of mCRC samples) including two and previously known fusion transcripts *ADAP1-NOC4L* found in seven (23%) mCRC specimens and also *RNF43-SUPT4H1* discovered in eight (26%) of mCRC specimens and three chimeric transcript with highest driver score ($>80\%$) found exclusively in metastatic samples were selected for validation with RT-PCR and Sanger sequencing (Table 2, Figure 2). The two recurrent fusions (*ADAP1-NOC4L* and *RNF43-SUPT4H1*) out of five chimeras were validated in SW48 and HT29 colon cancer cell lines as well as patient tissue samples (Figures 3 and 4A). These samples include 21 *ADAP1-NOC4L* positive samples with 13(56%) mCRC specimens [7 (58%) FFPE and 6 (54%) fresh tissues], eight (47%) pCRC specimens [5 (45%) FFPE and 3 (50%) fresh tissues], 25

RNF43-SUPT4H1 positive samples with 15 (60%) mCRC specimens [8 (66%) FFPE and 7 (63%) fresh tissues] and 10 (40%) pCRC specimens [6 (54%) FFPE and 4 (66%) fresh tissues].

The *ADAP1-NOC4L* fusion transcripts as a result of joining exon 4 *ADAP1* (ENST00000265846) to exon 10 *NOC4L* (ENST00000330579) had a continuous open reading frame (ORF) (Figure 3A). The DNA binding domains in *NOC4L* and the GTPase domain in *ADAP* were also preserved (Figure 3B). *RNF43-SUPT4H1* fusion transcript was identified as the result of joining the 5'UTR region of the *RNF43* (ENST00000407977) to exon 2 of the *SUPT4H1* (ENST00000225504) (Figure 4A). Domain analysis with Uniport, and InterPro online database tools indicated an existence of a protected transcription elongation factor SPT4 domain in *SUPT4H1* transcript (Figure 4B).

The other fusion transcripts identified in this study, have not been previously reported. They include *EVI5* (Ecotropic Viral Integration Site 5) -*GF11* (Growth Factor Independent 1 Transcriptional Repressor) characterized by 5' transcript Rab-GTPase-TBC domain and 3' transcript Zinc finger C2H2 type preserved functional domain; *DCAF11* (DDB1 And CUL4 Associated Factor 11)-*PSME1* (Proteasome Activator Subunit 1) contained 5'transcript WD domain repeats and 3'transcript proteasome activator pa28 alpha and beta subunit functional domain and *APLF* (Aprataxin And PNKP Like Factor)- *SPTLC1* (Serine Palmitoyltransferase Long Chain Base Subunit 1) marked with aminotransferase class I and II functional domain in 3' region (Table 2, Figure S1).

3.2 | Functional analysis of the ADAP1-NOC4L chimera

For in vitro functional validation, the full coding sequence of *ADAP1-NOC4L* (ENST00000265846:ENST00000330579 ADAP1-001: NOC4L-001) fusion transcript presented in SW48 cells was amplified. It was confirmed with RT-PCR followed by Sanger sequencing, whereas the *ADAP1-NOC4L* expression detected by RT-PCR was significantly low in SW48 (Figure S2C) and absent in HT29 cell lines.

3.3 | Clonogenic cell survival assay

A clonogenic assay was performed to determine the effect of the *ADAP1-NOC4L* on the proliferation of cancer cell line SW48. The proliferation of over-expressed SW48 and HT29 cell lines with *ADAP1-NOC4L* increased at day 7 of transfection, compared to the negative control ($p < 0.001$) (Figure 5A,B).

TABLE 1 Summary of final in-frame driver fusion transcript discovered by filtering pipeline in mCRC

| 5' Gene | 5' Gene_transcript_id | Breakpoint 1 Location | 3' Gene | 3' Gene_transcript_id | Breakpoint 2 Location | Strand |
|---------------|-----------------------|--------------------------|----------------|-----------------------|--------------------------|--------|
| EVI5 | ENST00000370331 | chr1:93029199 | GFI1 | ENST00000427103 | chr1:92944310 | -/- |
| LNK1 | ENST00000306888 | chr4:54342920 | FIP1L1 | ENST00000358575 | chr4:54324820 | -/+ |
| PRKAA1 | ENST00000397128 | chr5:40771821 | PTPN22 | ENST00000528414 | chr1:114372329 | -/- |
| MAP3K5 | ENST00000359015 | chr6:136901439 | MAP7 | ENST00000354570 | chr6:136742937 | -/- |
| DCAF11 | ENST00000559115 | chr14:24592286 | PSME1 | ENST00000382708 | chr14:24606194 | +/+ |
| CLRN3 | ENST00000368671 | chr10:129690820 | DOCK1 | ENST00000280333 | chr10:128768966 | -/+ |
| THBS1 | ENST00000260356 | chr15:39885003 | NFIB | ENST00000380959 | chr9:14179779 | +/- |
| LIME1 | ENST00000309546 | chr20:62369852 | SLC2A4RG | ENST00000266077 | chr20:62371730 | +/+ |
| LARS2 | ENST00000265537 | chr3:45565600 | LIMD1 | ENST00000273317 | chr3:45677642 | +/+ |
| ADAP1 | ENST00000265846 | chr7:959605 | NOC4L | ENST00000330579 | chr12:132635526 | -/+ |
| SPPL2A | ENST00000261854 | chr15:51012137 | TRPM7 | ENST00000560955 | chr15:50955243 | -/- |
| APLF | ENST00000303795 | chr2:68753374 | SPTLC1 | ENST00000262554 | chr9:94871116 | +/- |
| COL4A1 | ENST00000375820 | chr13:110817209 | CAMK2D | ENST00000296402 | chr4:114582928 | -/- |
| RNF43 | ENST00000407977 | chr17:56494378 | SUPT4H1 | ENST00000225504 | chr17:56428869 | -/- |

Note: Bold letters in 5'gene and 3'gene columns indicate finally selected chimeras. (-) and (+) in strand column shows sense and antisense strand and ordered from 5' to 3' gene.

Abbreviation: LM, liver metastasis.

3.4 | Wound-healing assay

The results of a wound-healing assay revealed that the 2D migration of the SW48 and HT29 cells were increased following transfection with the overexpressing *ADAP1-NOC4L* vector (Figure 5C,D).

3.5 | Transwell cell invasion assay

To investigate the effects of fusion transcript on cancer progression, we conducted a transwell migration assay for SW48 cancer cell lines. Seventy-two hours after transfection, SW48 cells harboring *ADAP1-NOC4L* overexpression

| In-frame | Functional domain | Type | Sample ID (type) | Read count |
|------------------|---|------------------|------------------|------------|
| 1 > 0 | Zinc finger C2H2-type | Intrachromosomal | SRR975592 (LM) | 7 |
| | | | SRR3827632 (LM) | 6 |
| | | | SRR10160732 (LM) | 6 |
| 0 > 1 | C3HC4 RING-type | Intrachromosomal | SRR975587 (LM) | 5 |
| | | | SRR10160731 (LM) | 5 |
| 2 > 2 | Serine/threonine-protein kinase | Interchromosomal | SRR975589 (LM) | 5 |
| | | | SRR3827633 (LM) | 5 |
| 2 > 2 | Serine/threonine-protein kinase | Intrachromosomal | SRR975587 (LM) | 6 |
| | | | SRR10160731 (LM) | 5 |
| 2 > 0 | Proteasome activator pa28 | Intrachromosomal | SRR975593 (LM) | 7 |
| | | | SRR975594 (LM) | 5 |
| | | | SRR10160731 (LM) | 6 |
| 2 > 2 | DNA-binding domain | Intrachromosomal | SRR975587 (LM) | 8 |
| | | | SRR3827632 (LM) | 5 |
| 0 > 2 | Zinc finger, C2H2-like | Interchromosomal | SRR10160731 (LM) | 6 |
| | | | SRR975588 (LM) | 5 |
| 2 > 0 | Zinc finger, C2H2-like | Intrachromosomal | SRR975588 (LM) | 9 |
| | | | SRR975595 (LM) | 6 |
| 0 > 2 | Zinc finger, LIM-type | Intrachromosomal | SRR3827632 | 5 |
| | | | SRR3827633 | 5 |
| 2 > 2 | Arf GTPase activating protein / CCAATbinding factor | Interchromosomal | SRR5131510 (LM) | 6 |
| | | | SRR10160731 (LM) | 8 |
| | | | SRR987590 (LM) | 6 |
| | | | SRR975592 (LM) | 5 |
| | | | SRR975594 (LM) | 7 |
| | | | SRR10160732 (LM) | 6 |
| 1 > 0 | MHCK/EF2 kinase | Intrachromosomal | SRR975593 (LM) | 5 |
| | | | SRR10160731 (LM) | 5 |
| 2 > 0 | SMAD/FHA domain | Interchromosomal | SRR4457128 (LM) | 5 |
| | | | SRR3827632 (LM) | 5 |
| | | | SRR10160732 (LM) | 5 |
| 2 > 2 | calmodulin-dependent protein kinase II | Interchromosomal | SRR975587 (LM) | 6 |
| | | | SRR975594 (LM) | 5 |
| - | Transcription elongation factor (<i>Spt4</i>) | Intrachromosomal | SRR975588 (LM) | 6 |
| | | | SRR987590 (LM) | 6 |
| | | | SRR975592 (LM) | 6 |
| | | | SRR975597 (LM) | 5 |
| | | | SRR4457128 (LM) | 8 |
| | | | SRR3827632 (LM) | 6 |
| | | | SRR975593 (LM) | 5 |
| SRR10160731 (LM) | 5 | | | |

exhibited 6.5 fold increased invasiveness ($p < 0.001$) (Figure 5E,F).

3.6 | PI Annexin V apoptosis assay

To further investigate the role of *ADAPI-NOC4L* in the progression of CRC cells, we used the Annexin V propidium

iodide staining. The percentage of apoptotic cells was determined by flow cytometric analysis. With the *ADAPI-NOC4L* upregulation in SW48 cells, the population of early apoptotic cells (Annexin V-positive, PI-negative) decreased to 3.57% compared to negative control. The percentage of late apoptotic cells (positive for both Annexin V and PI) increased from 56.3% in negative control cells to 57.2% in SW48 cells upregulating *ADAPI-NOC4L*.

| | Liver metastasis (%) | Primary CRC (%) | Common*(%) |
|------------------|----------------------|-----------------|------------|
| Intrachromosomal | 614 (37.9) | 838 (56.1) | 74 (79.5) |
| Interchromosomal | 1003 (62.1) | 658 (43.9) | 19 (20.5) |
| Coding | 1245 (76.9) | 1137 (76) | 73 (78.5) |
| Non-coding | 372 (23.1) | 359 (24) | 20 (21.5) |
| In-frame | 136 (8.5) | 47(3.1) | 5 (5.3) |
| Out of frame | 1481 (91.5) | 1449 (96.9) | 88 (94.7) |

TABLE 2 Detail representing types of fusion transcript regarding unique tissue type-specific and also (*) Shows shared chimeras among liver metastasis and primary colorectal cancer

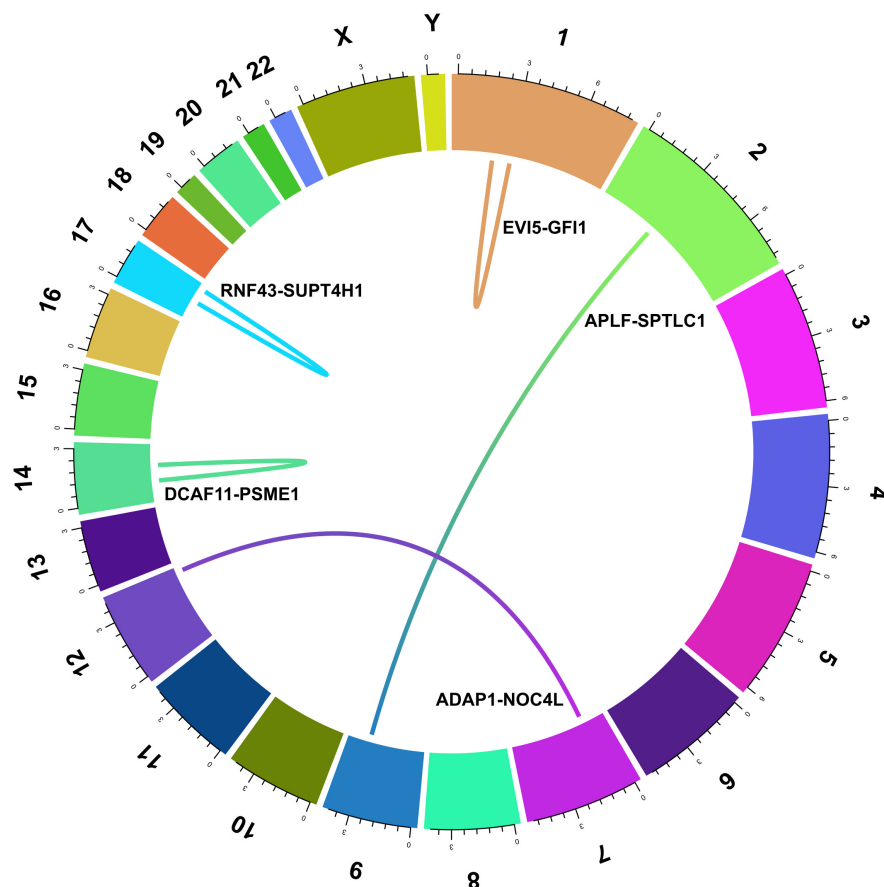


FIGURE 2 Circular view of the final 5 candidate chimeras with higher driver score or frequency representing the chromosomal regions involved, associated with each gene. ADAP1-NOC4L and APLF-SPTLC1 fusions are illustrated with interchromosomal mechanisms and other fusions in this image are characterized by intrachromosomal mechanisms.

The population of viable cells (Annexin V-negative and PI-negative) decreased to 13.2%, and there was an increase in the percentage of necrotic cells (Annexin V-negative, PI-positive), from 10.9% for the control group to 26% in the cells that overexpressed the fusion transcript (Figure S2A, B).

3.7 | Expression analysis of metastasis biomarkers

The expression of MMP9 associated with metastasis was found by RT-qPCR to be increased by 2.4 fold in SW48 ($p = 0.001$) and 2.8 fold in HT29 ($p = 0.171$) overexpressed cells. The results of qPCR also demonstrated higher expression of *FNI*, vimentin, and N-cadherin in SW48

($p = 0.001$, $p = 0.001$, $p = 0.001$) and HT29 cells containing fusion transcript ($p = 0.836$, $p = 0.001$, $p = 0.336$). *BCL2* anti-apoptotic and *BAX* proapoptotic markers showed higher expression and downregulation in SW48 ($p = 0.001$, $p = 0.001$, $p = 0.001$) and HT29 ($p = 0.001$, $p = 0.171$) cells overexpressing *ADAP1-NOC4L* chimeras compared to cells harboring empty vectors (mock), respectively (Figure 5G,H).

3.8 | Chimera expression level comparison

To further investigate the expression level difference of sequentially validated fusion transcripts between metastatic and primary CRC tissues, we analyzed the

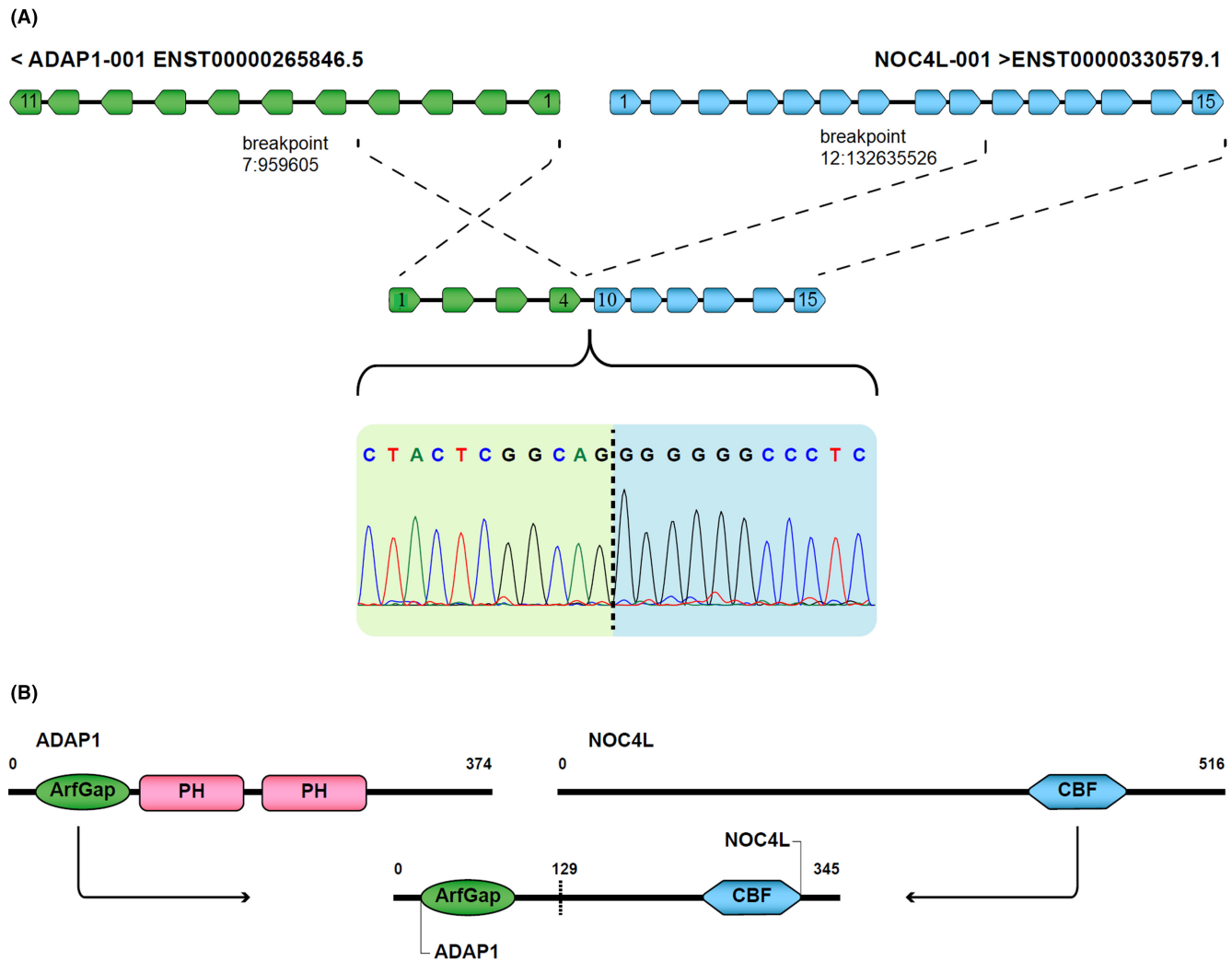


FIGURE 3 Schematic view of ADAP1-NOC4L fusion and the domains A. ADAP1-NOC4L fusion resulting from the binding of exon 4 of the ADAP1 gene and exon 10 of the NOC4L gene and electropherogram of the Sanger sequencing validation in the cell line expressing this fusion. B. Indicates the preservation of the ArfGap domain with GTPase function in the 5' region and CBF domain in the 3' end of the putative fusion protein.

21 *ADAP1-NOC4L* positive samples [13(56%) mCRC and 8(47%) pCRC] and 25 *RNF43-SUPT4H1* positive samples [15(60%) mCRC and 10(40%) pCRC], and found that higher relative expression of both candidate fusion transcripts ($p = 0.041$, $p = 0.037$) in liver metastasis samples compared to primary CRC tissues (Figure 6A,B).

3.9 | Survival analysis

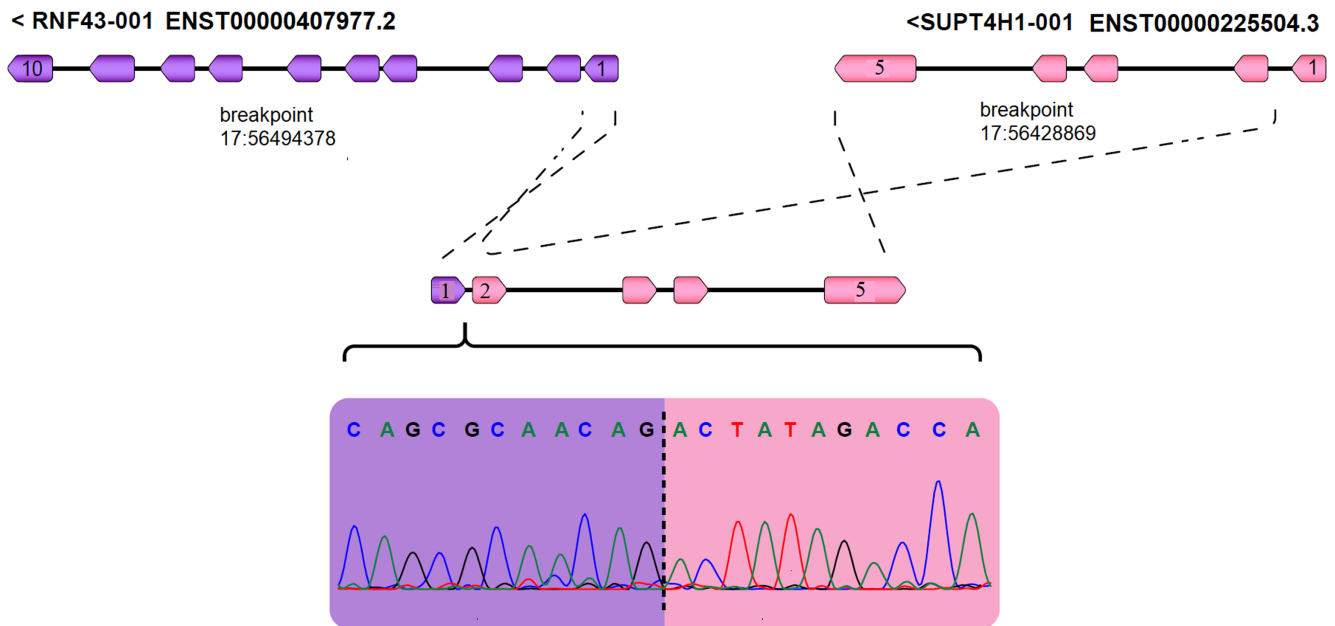
The association between chimera expression and OS time was analyzed using the Kaplan–Meier method to determine the prognostic value of *ADAP1-NOC4L* and *RNF43-SUPT4H1* in patients with pCRC and mCRC. Only cases with survival information were included, which resulted in a total of 23. The results demonstrated that patients

with positive expression of *ADAP1-NOC4L* had a significantly shorter OS time compared with those with the absence of *ADAP1-NOC4L* expression in mCRC ($p = 0.02$) (Figure 6C). Patients with pCRC, with positive expression of *RNF43-SUPT4H1*, exhibited a significantly shorter OS time than those with negative chimera expression ($p = 0.02$). (Figure 6F). In other cases, the fusions identified did not demonstrate a significant correlation with OS of patients (Figure 6D,E).

4 | DISCUSSION

In the present study, we aimed to identify important recurrent driver fusion transcripts involved in mCRC pathogenesis. We discovered 14 fusions exclusively in mCRC patients through RNA sequencing raw data analysis of

(A)



(B)

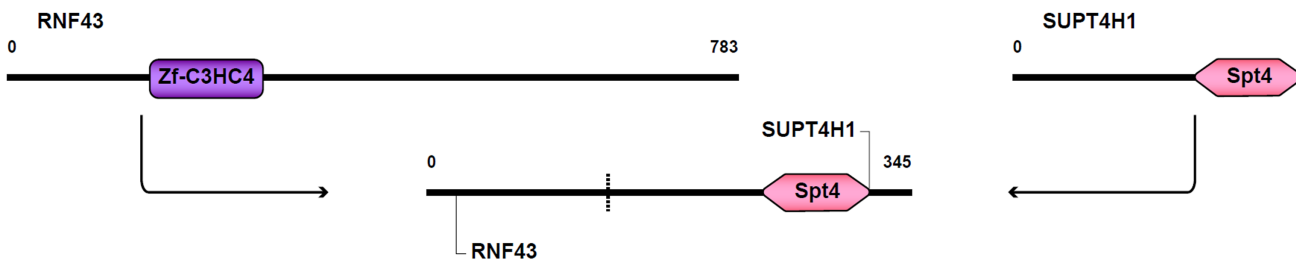


FIGURE 4 Schematic representation of RNF43-SUPT4H1 fusion and its domains (A). The RNF43-SUPT4H1 fusion is the result of the binding of the 5'UTR region of the RNF43 transcript to exon 2 of the SUPTH1 transcript, and the electropherogram shows Sanger sequencing validation in the expressing cell line. (B) Indicates the preservation of the SPT4 domain in the 3' fusion region. This domain is involved in transcription regulation and is predicted to be involved in the oncogenic activity of this fusion.

four different public databases including 12 novel fusions that were not previously reported and two previously known recurrent fusion transcripts. Finally, we validated in vitro the driver function of *ADAP1-NOC4L* fusion transcripts in CRC cell lines.

Multiple studies have discovered several fusion genes in mCRC, including *NTRK*, *ALK*, *ROS*, *RET*, *BRAF*, and *FGFR2* fusions genes, harboring a prognostic and/or predictive value or representing potentially candidates of targeted therapy.^{15–19,36} By applying fusion gene databases such as TCGA, FusionGDB, ChimerDB 4.0, Mitelman databases, our filtering strategy discovered novel fusion transcripts in mCRC.

In the current study based on the presence of putative oncogenic domains in fusion partners, we found 14 potential driver fusion transcripts in metastatic CRC by using Oncofuse, listed in Table 2. The expression of functional proteins resulting from the driver chimeric

transcripts have been shown to promote cancer development and invasive migration.^{37,38} Thus, we suggest that these cancer-type specific transcriptomic alterations may influence metastatic cell growth, development, and differentiation of primary CRC cells. In addition, it could potentially be utilized to contribute to the diagnosis of colorectal metastatic cancer at the histological examination or early detection in plasma using circulating tumor RNA specimens.^{39–41} Furthermore, studies to focus on tissue specificity could leverage these fusion transcripts as diagnostic biomarkers to identify the tissue of origin of the metastatic lesions in cases of diagnostic ambiguity.

In the present study, we compared the expression of *ADAP1-NOC4L* and *RNF43-SUPT4H1* between mCRC and pCRC patients and we found that the expression of these fusions was not limited to metastatic tissue and was expressed in varying quantities in primary CRC tissue.

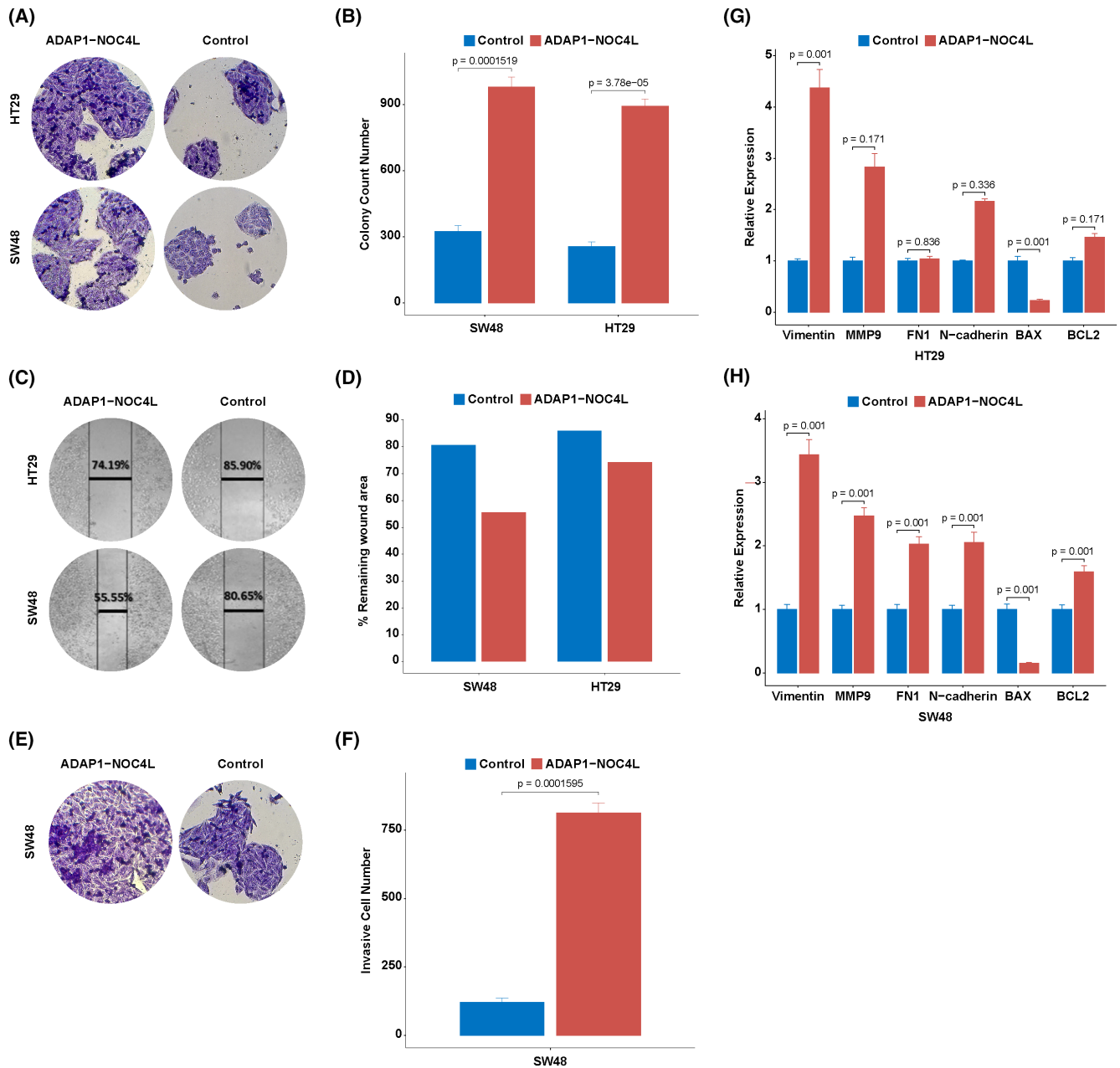


FIGURE 5 The effect of increasing ADAP1-NOC4L over-expression on functional and molecular characteristics of primary colon cancer cell lines. (A) and (B) Increased expression of the chimeric transcript has significantly increased the formation of cancer cell colonies ($p < 0.001$). C and D. shows 2D cell migration elevation in primary colon cancer. E and F. Increased expression of chimeric fusion caused a significant increase in cell invasion in SW48 cell line ($p < 0.001$). G. Increased ADAP1-NOC4L expression significantly enhanced Vimentin expression ($p = 0.001$) and decreased BAX (apoptosis-promoting biomarker) expression ($p = 0.001$) in HT29 cell line. H. Increased expression of ADAP1-NOC4L significantly upregulated the expression of metastasis (EMT) and BCL2 markers ($p = 0.001$, $p = 0.001$) and decreased the expression of BAX (apoptosis-promoting biomarker) in SW48 cell line ($p = 0.001$). EMT: Epithelial mesenchymal transition; MMP9: Matrix Metalloproteinase 9; FN1: Fibronectin 1.

Big Bang model of tumor evolution suggests that tumor metastatic potential is specified ab initio during early stages of tumor development and the metastatic tumors are in fact as a result of few subclones from previous molecularly determined cells.^{4,7} A recent study by Simeonov et al. has supported this idea.⁴² From this perspective, it can be expected that some driver alteration

including fusion transcript formation involved in the development of metastatic clones, previously developed in primary CRC, might be upregulated. In this aspect, therapeutic management similar to primary tumor tissue could be also effective on metastatic tumors.

Moreover, Kaplan–Meier survival analysis revealed that the presence of *ADAP1-NOC4L* in mCRC patients and

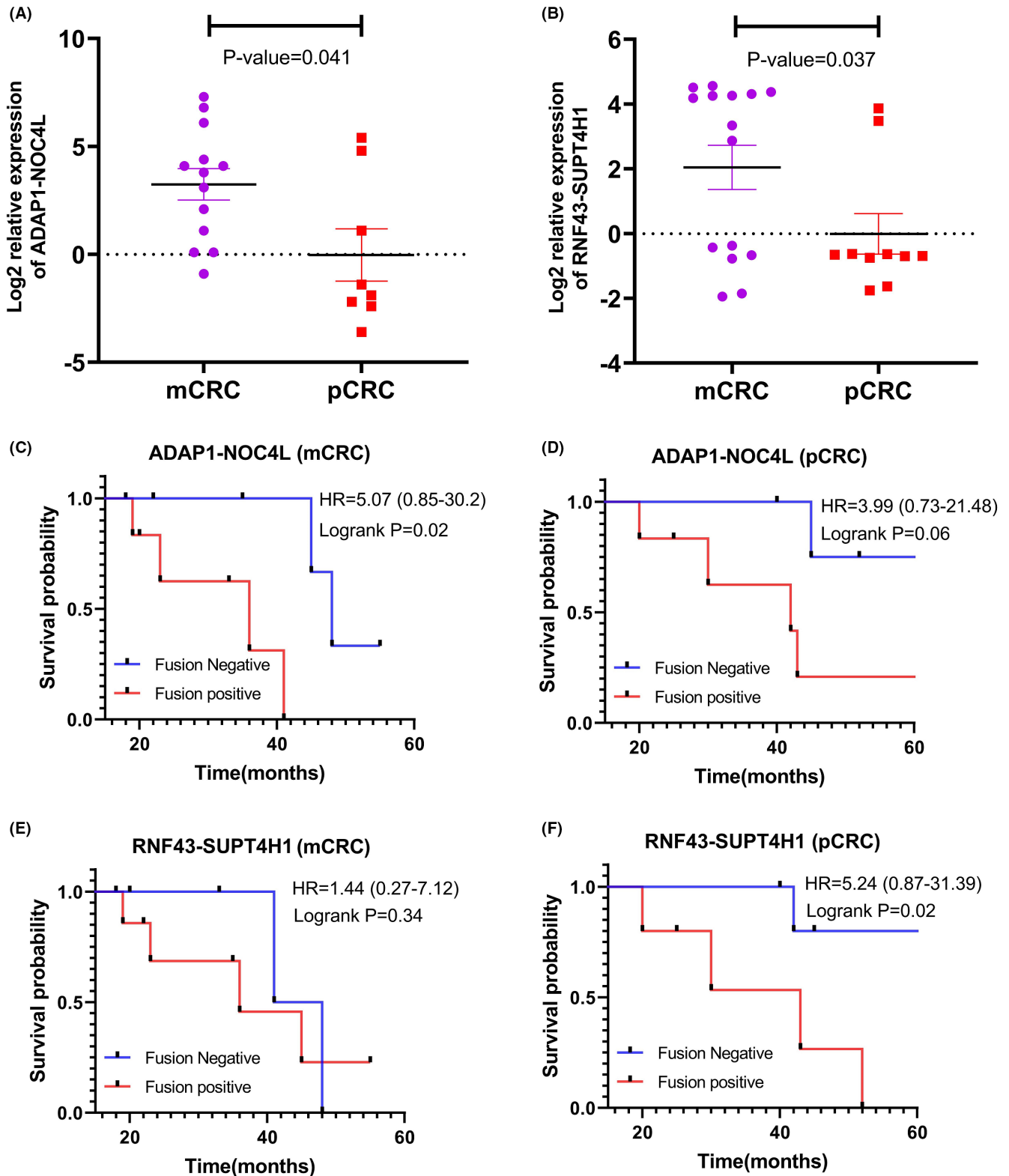


FIGURE 6 Relative expression comparison and overall survival analysis of chimeric transcripts candidates in fusion-positive liver metastatic and primary CRC tumor. (A) Shows an increase in the relative expression of ADAP1-NOC4L in metastatic colon cancer compared to the primary tumor ($p = 0.041$). (B) The relative expression of RNF43-SUPT4H1 is also shown elevation in metastatic colon cancer compared to the primary tumor ($p = 0.037$). (C) Patients with positive expression of ADAP1-NOC4L had a significantly shorter survival time compared with those with the absence of ADAP1-NOC4L expression in mCRC. ($p = 0.02$) and (D) pCRC ($p = 0.04$). (E) Positive expression of RNF43-SUPT4H1 in mCRC patients had no effect ($p = 0.34$) and (F) in pCRC patients had a direct effect on shorter overall survival rate ($p = 0.02$). mCRC: Metastatic colorectal cancer, pCRC: Primary colorectal cancer.

RNF43-SUPT4H1 fusion transcripts in pCRC patients were significantly associated with a shorter OS time. The findings indicate that *ADAP1-NOC4L* and *RNF43-SUPT4H1* expression may be a potential molecular marker for predicting the development and prognosis of patients with CRC. However, due to the limited number of samples examined, the current results should be interpreted with caution.

In this study, we validated two final candidate fusion transcripts by RT-PCR followed by Sanger sequencing (Figure 3A, 4A). *ADAP1-NOC4L* transcript involves exon 10 of the *ADAP1* gene linked with exon 9 of the *NOC4L*. *ADAP1* (Arf-GAP with dual PH domain-containing protein 1) is a protein containing two main domains including Arf-GAP which is a putative zinc finger with GTPase activating proteins (GAPs) and a Pleckstrin homology (PH) domain (Figure 3B). The Arf-GAP domain is critical in endocytic recycling and cytoskeleton remodeling. PH domain, plays a role in recruiting proteins to different membranes, thus targeting them to appropriate cellular compartments or enabling them to interact with other components of the signal transduction pathways.^{43,44} However, the association between *ADAP1* and tumorigenesis has only been rarely investigated. By using high throughput approaches, numerous fusion transcripts, including *ADAP1* in the 3' regions, have been described in different tumors.^{30,45} Only a handful of them have been functionally proven, and there is only one known study on *ADAP1-NOC4L*. In their research, Oga et al. introduced multiple *ADAP1*-based in-frame fusion transcripts. *ADAP1-NOC4L* was reported out of frame, contrary to our study in which the ORF of the *ADAP1-NOC4L* fusion was preserved. CBF is also critical in the function of *NOC4L*-based fusion genes. As a result, we hypothesized that the excessive *ADAP1-NOC4L* expression could contribute to the metastatic dissemination of a subset of primary colon cancer cells.

We further examined how *ADAP1-NOC4L* is involved in CRC metastatic progression, not previously investigated. In order to validate the biological functions of the *ADAP1-NOC4L* fusion transcript; first, we performed a functional analysis of *ADAP1-NOC4L* fusion transcript in vitro. Overexpression of *ADAP1-NOC4L* significantly increased cell growth and migration, compared to cells transfected with an empty vector. Then, we found that *ADAP1-NOC4L* enhanced EMT in SW48 and HT29 cell lines. It has been shown that EMT is associated with invasion and metastasis in numerous carcinomas.^{46–48} In the last decade, a growing number of studies have demonstrated the critical involvement of EMT in the dissemination of various carcinomas including CRC.^{49,50} Our results revealed that the mRNA level of EMT markers increased in SW48 and HT29 cells, suggesting that *ADAP1-NOC4L*

could promote EMT in CRC. It has been widely proposed that Bcl-2 expression in cancer patient samples can promote cell migration, invasion, and metastasis by inducing MMP9 protein expression in various tumors.⁵¹ In contrast, cell invasion has been shown to be impeded by BAX and other cell death inducers and negative regulators of apoptosis such as BAK.⁵² In our study, *ADAP1-NOC4L* overexpression increased cell motility and invasiveness, which was in the same way as *MMP9*, *BCL2* expression, and *BAX* downregulation and thus could be inferred to be effective in colon carcinoma cell metastasis.

RNF43-SUPT4H1 contains the joining of the 5'UTR region of *RNF43* RNA and exon 2 of *SUPT4H1* RNA which is a read-through transcript and is likely not related to structural variation (Figure 4A). The *RNF43-SUPT4H1* predicted protein lacks all *RNF43* domains but has preserved SPT4 domain of *SUPT4H1* (Figure 4B). It is likely to have nuclear localization and acts in a similar way to the wild-type *SUPT4H1* transcript. A study by Lee et al. showed that this fusion transcript commonly occurs in primary CRC samples.³⁷ They discovered that *RNF43-SUPT4H1* is prevalent in a variety of cell lines, including DLD-1, HT29, HCT116, and HCT15. They also functionally validated the driver potential of this chimera in primary CRC cell lines; but it was not identified in metastatic tissues in an early bioinformatics analysis with SOAPfuse. However, in our study, bioinformatics analysis with CLC Genomics Workbench software showed differential expression metastatic samples, which was cross-validated by Arriba and SOAPfuse. In addition, we elucidated the presence of this chimera in the SW48 cell line among different CRC cell lines (Figure 4A). In agreement with the previous similar study,³⁷ we also speculated that this fusion event is a frequent molecular alteration and can serve as a potential diagnostic biomarker in metastatic CRC.

In the current study, we confirmed the expression of *ADAP1-NOC4L* in mCRC patients' tissue samples and discovered discrepancies regarding the frequency of these fusion transcripts in mCRC between public data analysis and our patient clinical samples. In our specimens, *ADAP1-NOC4L* fusion-positive samples were identified in 56% of our mCRC tissues. In comparison, public data analysis revealed that only 23% of mCRC samples expressed fusion genes. These discrepancies have been also observed among other similar studies performed on fusion discovery in solid tumors^{11,21,38} and may be due to differences in sample collection, preparation, quality control, and bioinformatics analysis pipelines to select fusion transcripts.⁵³

Here, we used several computational fusion detection tools such as Arriba 2.1.0,²³ SOAPfuse-v1.27,²⁵ CLC Genomics Workbench 20,²⁴ and defuse-0.6.2²⁶ to identify fusion events at the transcriptome level. Although

RNA sequencing is a highly sensitive method for identifying RNA fusions, it is error prone at various stages such as pre-sequencing and sequencing.^{54,55} In addition, different fusion detection algorithms differ significantly in sensitivity and specificity, and therefore the use of different pipelines and experimental confirmation of results seems crucial to reduce false positive and negative results,^{14,53,56} which could be a major strength of the current study.

Another strength of this study is the investigation of recurrent fusions. As previously noted, fusion transcripts in solid tumors are uncommon and infrequently occur among individuals.²¹ Surprisingly, none of the fusion transcripts identified in Choi et al. study³⁸ were detected in our analysis, even when using the identical (defuse) algorithms in metastatic or primary cancer specimens. In this regard, chimeras with higher frequency are clinically more important. The exploitation of public samples allows the possibility of studying a more comprehensive range of specimens. As a result, the possibility of discovering clinically significant genetic alterations rises.

There are some limitations in this study that could be addressed in future research. First, the limited number of samples available and second, the unavailability of patients' clinical data, which has restricted more comprehensive examination of the current discoveries. Another limitation is that the algorithms we used were unable to discriminate fusions happening at the transcriptome level from those occurring at the genome level. Future research is needed to gain complete knowledge of the selected fusions. Furthermore, additional investigations are recommended by using animal models of metastasis to investigate the role of fusion proteins in conferring the complex multistep process of metastasis as well as further wet lab research to assess the specific pathways involved in metastasis.

5 | CONCLUSION

In the present study, we identified multiple novel chimeras as well as 2 recurrent driver protein-coding fusion RNAs (*ADAP1-NOC4L* and *RNF43-SUPT4H1*) in mCRC. *In-silico* analysis elucidated that the protein-coding fusions account for a considerable amount of RNA-based fusion in cancer. *In vitro*, functional evaluation of the *ADAP1-NOC4L* chimeric transcript revealed a potential contributory effect in EMT and metastasizing the primary CRC. Survival analysis showed a prognostic biomarker potentiality of *ADAP1-NOC4L* and *RNF43-SUPT4H1* in metastatic and primary CRC, respectively.

AUTHOR CONTRIBUTIONS

Mohammad Amin Kerachian and Soodabeh Shahidsales designed the experiments. Amin Talebi performed the experiments and analyzed the results. Mohsen Aliakbarian and Masoud Pezeshki Rad provided samples. All authors contributed to drafting the article and revising it.

ACKNOWLEDGMENTS

Our special thanks goes to Mashhad University of Medical Sciences for supporting financially this study (Grant# 981199). This study was part of a Ph.D. dissertation (AT).

CONFLICT OF INTEREST

Authors have no financial conflicts of interests.

DATA AVAILABILITY STATEMENT

The datasets generated and/or analysed during the current study are available in the The Jackson Laboratory, Tumor Fusion Gene Data Portal [<https://tumorfusions.org/>], UTHHealth hosts 2021 Commencement Ceremonies. fusion gene annotation DataBase [<https://ccsm.uth.edu/FusionGDB/>], ChimerDB 4.0 [<http://www.kobic.re.kr/chimerdb/>], Mitelman F JB, Mertens F Mitelman Database of Chromosome Aberrations and Gene Fusions in Cancer [<https://mitelmandatabase.isb-cgc.org/>].

ETHICS STATEMENT

Ethical approval for this study was given by the ethics committee of the Mashhad University of Medical Sciences, Mashhad, Iran (Approval ID: IR.MUMS.MEDICAL.REC.1399.106).

AVAILABILITY OF DATA AND MATERIALS

The datasets generated and/or analyzed during the current study are available in the The Jackson Laboratory, Tumor Fusion Gene Data Portal [<https://tumorfusions.org/>], UTHHealth hosts 2021 Commencement Ceremonies. fusion gene annotation DataBase [<https://ccsm.uth.edu/FusionGDB/>], ChimerDB 4.0 [<http://www.kobic.re.kr/chimerdb/>], Mitelman F JB, Mertens F Mitelman Database of Chromosome Aberrations and Gene Fusions in Cancer [<https://mitelmandatabase.isb-cgc.org/>].

INFORMED CONSENT

The patients/participants provided their written informed consent to participate in this study.

ORCID

Mohammad Amin Kerachian  <https://orcid.org/0000-0001-5016-2882>

REFERENCES

- Sung H, Ferlay J, Siegel RL, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2021;71:209-249.
- Siegel RL, Miller KD, Goding Sauer A, et al. Colorectal cancer statistics, 2020. *CA Cancer J Clin.* 2020;70:145-164.
- Oki E, Ando K, Nakanishi R, et al. Recent advances in treatment for colorectal liver metastasis. *Annals of Gastroenterological Surgery.* 2018;2:167-175.
- Wang J, Li S, Liu Y, Zhang C, Li H, Lai B. Metastatic patterns and survival outcomes in patients with stage IV colon cancer: a population-based analysis. *Cancer Med.* 2020;9:361-373.
- Heinemann V, von Weikersthal LF, Decker T, et al. FOLFIRI plus cetuximab or bevacizumab for advanced colorectal cancer: final survival and per-protocol analysis of FIRE-3, a randomised clinical trial. *Br J Cancer.* 2021;124:587-594.
- Osterlund P, Salminen T, Soveri L-M, et al. Repeated centralized multidisciplinary team assessment of resectability, clinical behavior, and outcomes in 1086 Finnish metastatic colorectal cancer patients (RAXO): a nationwide prospective intervention study. *Lancet Regional Health-Europe.* 2021;3:100049.
- Hu Z, Ding J, Ma Z, et al. Quantitative evidence for early metastatic seeding in colorectal cancer. *Nat Genet.* 2019;51:1113-1122.
- Nappi A, Berretta M, Romano C, et al. Metastatic colorectal cancer: role of target therapies and future perspectives. *Curr Cancer Drug Targets.* 2018;18:421-429.
- Haan JC, Labots M, Rausch C, et al. Genomic landscape of metastatic colorectal cancer. *Nat Commun.* 2014;5:1-12.
- Yaeger R, Chatila WK, Lipsyc MD, Hechtman JF, Cercek A, Sanchez-Vega F, Jayakumaran G, Middha S, Zehir A, Donoghue MTA, You D, Viale A, Kemeny N, Segal NH, Stadler ZK, Varghese AM, Kundra R, Gao J, Syed A, Hyman DM, Vakiani E, Rosen N, Taylor BS, Ladanyi M, Berger MF, Solit DB, Shia J, Saltz L, Schultz N. Clinical sequencing defines the genomic landscape of metastatic colorectal cancer. *Cancer Cell* 2018; 33: 125-136. e123, 125, 136.e3.
- Pagani F, Randon G, Guarini V, et al. The landscape of actionable gene fusions in colorectal cancer. *Int J Mol Sci.* 2019;20:5319.
- Malone ER, Oliva M, Sabatini PJ, Stockley TL, Siu LL. Molecular profiling for precision cancer therapies. *Genome Med.* 2020;12:1-19.
- Ogunwobi OO, Mahmood F, Akingboye A. Biomarkers in colorectal cancer: current research and future prospects. *Int J Mol Sci.* 2020;21:5311.
- Taniue K, Akimitsu N. Fusion genes and RNAs in cancer development. *Non-coding RNA.* 2021;7:10.
- Cocco E, Scaltriti M, Drilon A. NTRK fusion-positive cancers and TRK inhibitor therapy. *Nat Rev Clin Oncol.* 2018;15:731-747.
- Hsiao SJ, Zehir A, Sireci AN, Aisner DL. Detection of tumor NTRK gene fusions to identify patients who may benefit from tyrosine kinase (TRK) inhibitor therapy. *J Mol Diagn.* 2019;21:553-571.
- Alese OB, El-Rayes BF, Sica G, et al. Anaplastic lymphoma kinase (ALK) gene alteration in signet ring cell carcinoma of the gastrointestinal tract. *Ther Adv Med Oncol.* 2015;7:56-62.
- Le Rolle A-F, Klempner SJ, Garrett CR, et al. Identification and characterization of RET fusions in advanced colorectal cancer. *Oncotarget.* 2015;6:28929-28937.
- Peeters M, Oliner KS, Parker A, et al. Massively parallel tumor multigene sequencing to evaluate response to panitumumab in a randomized phase III study of metastatic colorectal cancer. *Clin Cancer Res.* 2013;19:1902-1912.
- Valeri N. Streamlining detection of fusion genes in colorectal cancer: having "faith" in precision oncology in the (tissue)"agnostic" era. *Cancer Res.* 2019;79:1041-1043.
- Kloosterman WP, van den Braak RRC, Pieterse M, et al. A systematic analysis of oncogenic gene fusions in primary colon cancer. *Cancer Res.* 2017;77:3814-3822.
- Pietrantonio F, Di Nicolantonio F, Schrock A, et al. RET fusions in a small subset of advanced colorectal cancers at risk of being neglected. *Ann Oncol.* 2018;29:1394-1401.
- Uhrig S, Ellermann J, Walther T, et al. Accurate and efficient detection of gene fusions from RNA sequencing data. *Genome Res.* 2021;31:448-460.
- QIAGEN. QIAGEN CLC Genomics.
- Jia W, Qiu K, He M, et al. SOAPfuse: an algorithm for identifying fusion transcripts from paired-end RNA-seq data. *Genome Biol.* 2013;14:1-15.
- McPherson A, Hormozdiari F, Zayed A, et al. deFuse: an algorithm for gene fusion discovery in tumor RNA-seq data. *PLoS Comput Biol.* 2011;7:e1001138.
- Murphy C, Elemento O. AGFusion: annotate and visualize gene fusions. *bioRxiv.* 2016;080903.
- The Jackson Laboratory. Tumor Fusion Gene Data Portal.
- UTHealth hosts 2021 Commencement Ceremonies. fusion gene annotation DataBase.
- ChimerDB 4.0
- Mitelman FJB, Mertens F. Mitelman Database of Chromosome Aberrations and Gene Fusions in Cancer
- Shugay M, Ortiz de Mendibil I, Vizmanos JL, Novo FJ. Oncofuse: a computational framework for the prediction of the oncogenic potential of gene fusions. *Bioinformatics.* 2013;29:2539-2546.
- Huggett JF. The digital MIQE guidelines update: minimum information for publication of quantitative digital PCR experiments for 2020. *Clin Chem.* 2020;66:1012-1029.
- IBM SPSS Statistics for Windows. Version 22.0 edn.: IBM Corp, Released 2013.
- GraphPad prism version 8.0.0 for windows. GraphPad Software. Accessed November 20, 2021. www.graphpad.com
- Cocco E, Benhamida J, Middha S, et al. Colorectal carcinomas containing hypermethylated MLH1 promoter and wild-type BRAF/KRAS are enriched for targetable kinase fusions. *Cancer Res.* 2019;79:1047-1053.
- Lee J-R, Kwon CH, Choi Y, et al. Transcriptome analysis of paired primary colorectal carcinoma and liver metastases reveals fusion transcripts and similar gene expression profiles in primary carcinoma and liver metastases. *BMC Cancer.* 2016;16:1-11.
- Choi Y, Kwon CH, Lee SJ, Park J, Shin J-Y. Integrative analysis of oncogenic fusion genes and their functional impact in colorectal cancer. *Br J Cancer.* 2018;119:230-240.
- Streubel A, Mairinger T, Kollmeier J, et al. Detection of clinically relevant gene fusions in plasma samples of lung cancer patients in routine molecular diagnostics. American society of. *Clin Oncol.* 2018;36:e24191.

40. Przybyl J, van de Rijn M, Rutkowski P. Detection of SS18-SSX1/2 fusion transcripts in circulating tumor cells of patients with synovial sarcoma. *Diagn Pathol*. 2019;14:1-2.
41. illumina. Get more from your sarcoma analysis 2017.
42. Simeonov KP, Byrns CN, Clark ML, Norgard RJ, Martin B, Stanger BZ, Shendure J, McKenna A, Lengner CJ Single-cell lineage tracing of metastatic cancer reveals selection of hybrid EMT states. *Cancer Cell* 2021; 39: 1150–1162. e1159, 1150, 1162.e9.
43. Wang D-S, Shaw R, Winkelmann JC, Shaw G. Binding of PH domains of β -adrenergic-receptor kinase and β -spectrin to WD40/ β -transducin repeat containing regions of the β -subunit of trimeric G-proteins. *Biochem Biophys Res Commun*. 1994;203:29-35.
44. Yao L, Kawakami Y, Kawakami T. The pleckstrin homology domain of Bruton tyrosine kinase interacts with protein kinase C. *Proc Natl Acad Sci*. 1994;91:9175-9179.
45. Oga T, Yamashita Y, Soda M, et al. Genomic profiles of colorectal carcinoma with liver metastases and newly identified fusion genes. *Cancer Sci*. 2019;110:2973-2981.
46. Cao H, Xu E, Liu H, Wan L, Lai M. Epithelial–mesenchymal transition in colorectal cancer metastasis: a system review. *Pathology-Research and Practice*. 2015;211:557-569.
47. Vu T, Datta PK. Regulation of EMT in colorectal cancer: a culprit in metastasis. *Cancer*. 2017;9:171.
48. Gooding AJ, Schiemann WP. Epithelial–mesenchymal transition programs and cancer stem cell phenotypes: mediators of breast cancer therapy resistance. *Mol Cancer Res*. 2020;18:1257-1270.
49. Niknami Z, Eslamifar A, Emamirazavi A, Ebrahimi A, Shirkoohi R. The association of vimentin and fibronectin gene expression with epithelial-mesenchymal transition and tumor malignancy in colorectal carcinoma. *EXCLI J*. 2017;16:1009.
50. Zeisberg M, Neilson EG. Biomarkers for epithelial-mesenchymal transitions. *J Clin Invest*. 2009;119:1429-1437.
51. Dewson G, Kluck RM. Bcl-2 family-regulated apoptosis in health and disease. *Cell Health and Cytoskeleton*. 2010;2:9-22.
52. Wojcik I, Szybka M, Golanska E, et al. Abnormalities of the P53, MDM2, BCL2 and BAX genes in acute leukemias. *Neoplasma*. 2005;52:318-324.
53. Talebi A, Thiery JP, Kerachian MA. Fusion transcript discovery using RNA sequencing in formalin-fixed paraffin-embedded specimen. *Crit Rev Oncol Hematol*. 2021;103303:103303.
54. Lou DI, Hussmann JA, McBee RM, et al. High-throughput DNA sequencing errors are reduced by orders of magnitude using circle sequencing. *Proc Natl Acad Sci*. 2013;110:19872-19877.
55. Sleep JA, Schreiber AW, Baumann U. Sequencing error correction without a reference genome. *BMC Bioinformatics*. 2013;14:1-9.
56. Carrara M, Beccuti M, Lazzarato F, Cavallo F, Cordero F, Donatelli S, Calogero RA State-of-the-art fusion-finder algorithms sensitivity and specificity. *Biomed Res Int* 2013; 2013, 1, 6.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Talebi A, Shahidsales S, Aliakbarian M, Pezeshki Rad M, Kerachian MA. Oncogenic fusion transcript analysis identified ADAP1-NOC4L, potentially associated with metastatic colorectal cancer. *Cancer Med*. 2023;12:525-540. doi: [10.1002/cam4.4943](https://doi.org/10.1002/cam4.4943)