



Università di Foggia

UNIVERSITÀ DEGLI STUDI DI FOGGIA

Dipartimento di Medicina Clinica e Sperimentale

PhD Course

“ Medicina Sperimentale e Rigenerativa ”

XXX Cicle

PhD Thesis

Title

**“IDENTIFICATION OF FUSION GENES
ASSOCIATED WITH SPORADIC COLORECTAL
CARCINOGENESIS”**

Tutor

Prof. Vescovi Angelo

PhD Student

Dott.ssa Panza Anna

Supervisor

Dott. Palmieri Orazio

A. A. 2016-2017

INDEX

SUMMARY	pag.3
1 INTRODUCTION	pag.5
1.1 Sporadic colorectal cancer	pag.5
1.2 Fusion genes	pag.9
1.3 Fusion genes and CRC	pag.10
1.4 Aim of the study	pag.13
2 MATERIALS AND METHODS	pag.15
2.1 Clinical Samples	pag.15
2.2 RNA extraction from fresh frozen tissue	pag.15
2.3 Library preparation for RNA Sequencing	pag.17
2.4 Sample pooling and sequencing	pag.19
2.5 Fusion Search	pag.20
2.6 Fusion gene validation by Reverse Transcription-PCR	pag.20
2.7 DNA Sequencing	pag.22
2.8 FISH Assay	pag.23
3 RESULTS	pag.24
3.1 Identification of fusion transcripts by EricScript and ChimeraScan algoritms	pag.24
3.2 Fusion junction validation by RT-PCR and Sanger sequencing	pag.27
3.3 MRPS31 gene rearrangement validation by FISH assay	pag.31
4 DISCUSSION	pag.34
APPENDIX A	pag.39
REFERENCES	pag.41

SUMMARY

Sporadic colorectal cancer (CRC) is a somatic genetic disease that progresses as a multistep process arising from adenomatous polyps and developing into locally invasive and metastatic cancer. The understanding of the mechanisms behind adenoma-to-carcinoma progression is crucial to identify biomarkers that characterize polyps at high risk of cancer development. New advances in sequencing have made possible to confirm prior identified genetic aberrations, expand new data, revealing several diagnostic and prognostic biomarkers including fusion genes. The latter are a class of genomic rearrangements in which two genes are fused, leading to the production of chimeric transcripts that may have an aberrant activity.

In the present study, the RNA-Sequencing (RNA-Seq) technology has been used in order to interrogate the transcriptome and to detect novel fusion genes involved in sporadic CRC progression in a cohort of 11 CRC patients. For this purpose, adenomatous and cancerous synchronous lesions, as well as normal mucosa, were collected. The potential fusion events were analyzed by means bioinformatics analysis using 2 different fusion-search algorithms: EricScript and ChimeraScan. Moreover, using specific filtering criteria, a dataset of 12 tumor-specific fusion events has been selected for further validation analyses. Many of the identified genes, have a known involvement in biological pathways linked in mitosis, cell cycle control, cell proliferation, apoptosis, invasion and in cytoskeletal remodelling. The subsequent validation by means RT-PCR and Sanger sequencing on whole CRC

cohort, confirmed that one of these 12 fusions, named MRPS31-SUGT1, was a tumor specific recurrent event. In particular, this newly fusion was present in adenoma and adenocarcinoma tissues of 3 patients and, as expected, no band was detect in normal tissues. Furthermore, the MRPS31 rearrangement was confirmed using FISH assay (Break Apart probe) in the cells of polyps and in the tumor cells of two CRC patients positives to MRPS31-SUGT1 fusion. This fusion gene represents a newly, unreported and recurrent event in promoting colorectal tumorigenesis and may be used a potential targeted therapeutics.

1.INTRODUCTION

1.1 Sporadic colorectal cancer

Sporadic colorectal cancer (CRC) is the third most common human malignancy and the fourth leading cause of cancer related death worldwide [Parker BC, et al. 2013; Shike M, et al. 1990]. Each year over one million people are diagnosed with CRC worldwide with a mortality rate of ~30% [Torre LA, et al. 2015]. It onset late in life, being detected at an average age of 70 years and affect men and women with nearly equal frequency.

In morphological terms, sporadic CRC progresses as a multistep process arising from adenomatous polyps and developing into locally invasive and metastatic cancer, without known contribution from germline causes or significant family history of cancer or inflammatory bowel disease. According to the multistep process, an initiating genetic event occurs in a normal colonic stem cell involving overactivation of Wnt signaling and the following grow of a observable polyp. This step, called “tumor initiation”, may take starting from 30 to 60 years. Progressively, a polyp will acquired the makeup to transform into malignancy. This is called “tumor progression” step may be as 1-5 years after tumor initiation [Carethers JM, et al. 2015].

Early detection of premalignant lesions such as adenomatous polyps has decreased the risk of CRCs [Winawer SJ, et al. 1993], however, cases which are initially undetected and advanced in distant metastasis have a poor prognosis [O’Connell

JB, et al. 2004]. To date, surgery of the primary cancer (or the defined metastasis) and the chemoradiation improving outcome are the best approaches for attempted cure [Carethers JM, et al. 2015]. Recently, the development of next generation sequencing (NGS) approaches has confirmed prior identified genetic aberrations, expanded new data and allowed to classify sporadic CRCs into hypermutated (16% of sporadic CRCs) and non-hypermutated (84% of sporadic CRCs) based on the molecular features [Carethers JM, et al. 2015; Lee HS, et al. 2017].

The low group of *hypermutated* CRC is characterized by microsatellite instability (MSI) resulting from defects in the DNA mismatch-repair (MMR) system. Hypermutated tumors revealed somatic mutation of the DNA MMR genes such as hMLH1, hMSH6, hMSH2, hMSH3, and mutations in the genes POLE and POLD1.

Non-hypermutated sporadic CRCs are characterized by chromosomal instability (CIN), with multiple somatic copy number alterations and aneuploidy.

The model of molecular pathogenesis of non-hypermutated adenomas and CIN CRCs (showed in figure 1) is complex and heterogeneous and involve several somatic mutations of oncogenes (for example, KRAS, PIK3CA and BRAF) and tumor suppressor genes (for example, APC, TP53 and FBXW7) that are required for cancer initiation and progression [Fearon ER, et al. 1990; Wong SH, et al. 2013].

The accumulation of somatic genetic events is clonal and cause deregulation of different intracellular signals, resulting in uncontrolled cell proliferation, resistance to cell death, cellular invasiveness and metastasis, thereby contributing to colorectal tumorigenesis [Wu WK, et al. 2013]. The earliest and most prevalent genetic event yet identified colorectal tumorigenesis is genetic disruption of the APC and the following over-activation of Wnt signaling pathway [Kinzler KW, et al. 1991]. This pathway is documented to be altered in approximately 95% of colorectal tumors (hypermuted and non hypermutated) [Powell SM, et al. 1992]. Inactivation of APC leads to increased β -catenin/Tcf-mediated transcription growth-promoting genes [He TC, et al. 1998]. Loss of APC function or gain of β -catenin function leads to clonal expansion of the mutated epithelial cell, giving rise to a small adenoma [Su LK, et al. 1992]. Mutations in the RAS/RAF pathway, the p53 pathway, and several other genes/pathways drive tumor progression towards malignancy and metastasis (figure 1) [Vogelstein B, et al. 2001]. It is notorious that oncogenic activation of KRAS increase proliferation and growth in size of the tumor. KRAS belongs to the ERBB/KRAS/BRAF/MAPK signaling pathways in which each member can mimic the effects of mutant KRAS protein. Furthermore KRAS was present in 13% of non-hypermuted CRCs. Moreover, a low part of non-hypermuted CRCs (27%) shows disruption of tumor suppressor TGF β signaling pathways during tumor progression [Carethers JM, et al. 2015].

In the 60% of non-hypermuted CRCs was found the down-regulation of TP53. The latter is a tumor suppressor gene involved in the regulation of the cell cycle

and in the and in repair of DNA. TP53 mutations, coinciding with conversion from benign to malignancy, confer poor prognosis for patients with CRC [Carethers JM, et al. 2015].

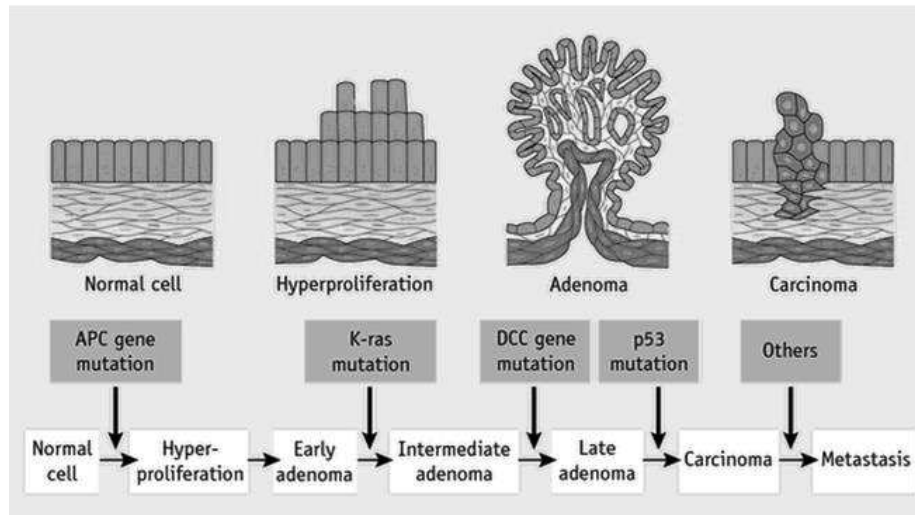


Figure 1. The genetic model of colorectal tumorigenesis. Colorectal cancers develop over the course of 20-40 years due to genetic disruption of the APC, RAS and p53 pathways.

Genomic instability is a frequent hallmark of non-hypermethylated sporadic CRCs. Previous studies of genomic alterations have demonstrated that somatic changes, including point mutations, DNA rearrangements and copy number variations, may drive the development of CRC [Markowitz SD, et al. 2009]. Studies on independent genome-wide expression data sets of CRC have identified several biologically relevant pathways [Abatangelo L, et al. 2009; Maglietta R, et al. 2010], and alternative splicing variants by measuring the differential expression of exons [Consiglio A, et al. 2012]. New advances in sequencing have made it possible to reveal a diagnostic and prognostic aspects in the genetic variants.

Among these, the fusion genes represent an important and emerging class of oncogenes [Medves S, et al. 2012].

1.2 Fusion genes

Fusion genes originate from a chromosomal breakage (translocation) and junction of two noncontiguous genetic loci, thereby juxtaposing regulatory and/or protein-coding elements from the two loci. The resulting fusion genes express transcript and protein products with altered regulation and/or structure, play an important role in the initiation of tumorigenesis and have been strongly associated with distinct cancer subtypes [Barr FG, et al. 2016]. The biological significance of fusion genes, together with their specificity to cancer cells, has made them into excellent biomarkers that might be utilized to predict prognosis, staging, and treatment approaches for a more personalized form of medicine [Carethers JM, et al. 2015].

The first translocation $t(9:22)(q34;q11)$ identified in chronic myeloid leukemia, was resulted in discovery of Philadelphia chromosome [Rowley JD, et al. 1973]. This translocation juxtaposes the 5' portion of the BCR gene with the 3' portion of the tyrosine kinase ABL1, generating the BCR-ABL1 fusion gene with constitutive kinase activity. Recent studies revealed a number of genes involved in translocations and in tumorigenesis. However, the study of translocations in solid tumors has been less appreciated compared with hematologic disorders due

principally to technical and analytical problems. The chromosome morphology is often poor to that in hematologic cancer. Moreover, the karyotypes are complex and make it difficult the identification of the abnormality. Despite this, using technologies based on fluorescence in situ hybridization (FISH) and microarray approach, several fusion genes have been identified, and the application of NGS has led to the confirmation of their existence in solid tumors [Barr FG, et al. 2016]. Fusion genes, such as TMPRSS2-ERG is present in about 50% of prostate cancer [Tomlins SA, et al. 2005], EML4-ALK was identified in about 8.6% of non-small cell lung cancers [Soda M, et al. 2007; Lin E, et al. 2009]. In breast cancer were detected different fusion genes such as ETV6-NTRK3, MYB-NFIB and more recently EEFDIP3-FRY and PPIRIB-STAR3 [Kim J, et al. 2015]. The FGFR3-TACC3 fusion gene, has been reported in lung cancers, bladder and recently in a subset of glioblastoma patients [Singh D, et al. 2012; Williams SV, et al. 2013; Wu YM, et al. 2013].

1.3 Fusion genes and CRC

Thus far, important progress has been made in the characterization of genetic alterations in CRC [Kim TM, et al. 2013]. New advances in sequencing have made possible to identify many genetic variants in CRC samples, by revealing several diagnostic and prognostic biomarkers including fusion genes. In 2011 was reported VTI1A-TCF7L2 as the first fusion discovered in CRC. The latter was detected in three of 97 CRCs (3%) as well as the colon cancer cell line NCI-H508

[Bass AJ, et al. 2011; Nome T, et al. 2014] (figure 2). TCF7L2 encodes a transcription factor, TCF4, that dimerizes with β -catenin involved in regulation of the Wnt signaling pathway. RNA-interference vectors targeting the sequence spanning the fusion reduce the expression of the fusion mRNA leading to a dramatic reduction in the anchorage-independent growth of cells from NCI-H508 that harbor the fusion gene.

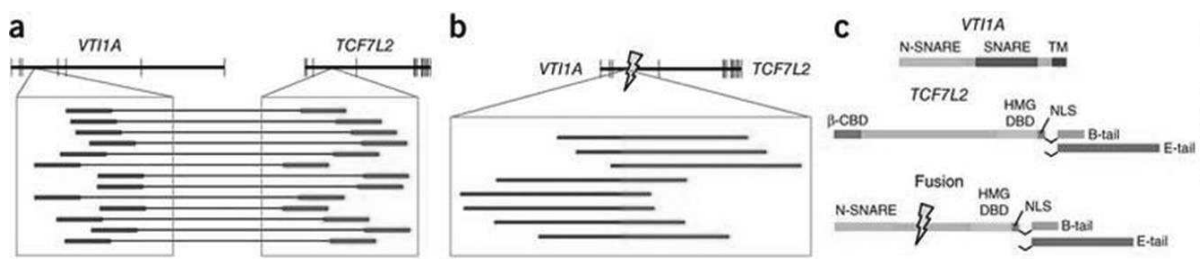


Figure 2: VTI1A-TCF7L2 fusion gene. a) The upper schematic depicts the positions of exons (vertical lines) within VTI1A and TCF7L2, which reside adjacent to each other on chromosome 10. The blowup displays the locations of discordant paired-end reads found in tumor CRC-9 for which one read (labeled in blue) is in an intron of VTI1A and the other read (labeled in red) is in an intron of TCF7L2. b) The upper schematic depicts the structure of the predicted fusion transcript generated by the fusion. The presence of the exact reads spanning the fusion of the two introns (marked by lightning bolt) is depicted in the inset with regions of the reads corresponding to original VTI1A intron in blue and those of TCF7L2 in red. c) The protein domain structure of native VTI1A and TCF4-TCF7L2 are shown.

One of the most frequent fusion genes in CRC is represented by R-spondin family members found in approximately 10% of colon tumor [Seshagiri S, et al. 2012] (figure 3). R-spondin family proteins (RSPO) are involved in cellular proliferation, differentiation, and maintenance of stem cells by modulating the Wnt/ β -catenin pathway.

The PTPRK-RSPO3 and EIF3E-RSPO2 chimeric proteins activated the Wnt signaling pathway in vitro, and fusion-positive tumors can carry alterations in the Wnt pathway.

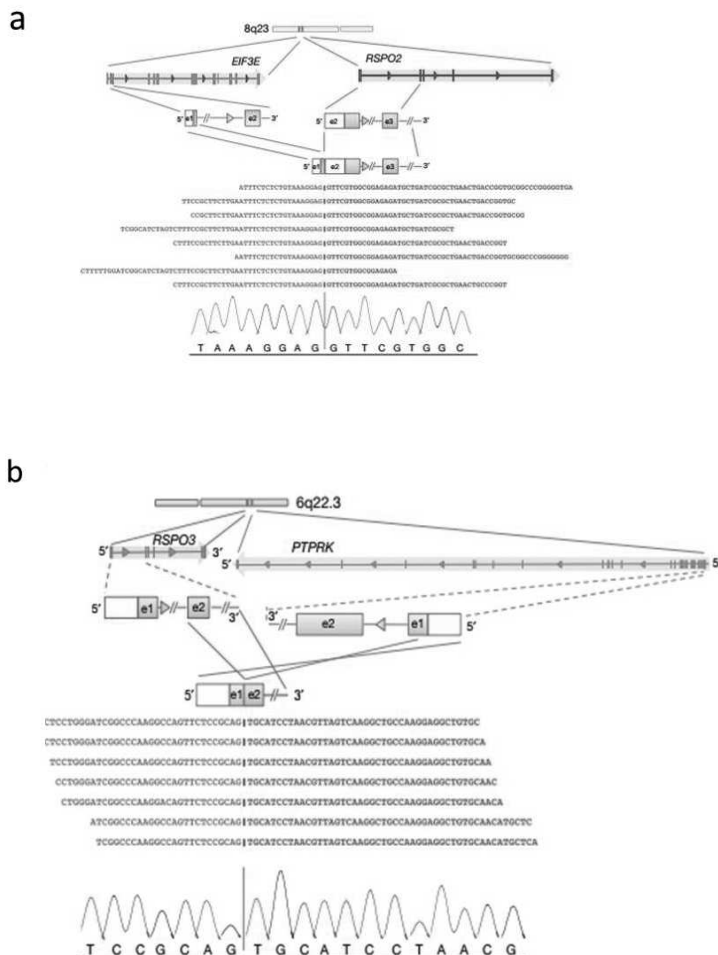


Figure 3: EIF3E-RSPO2 and PTPRK-RSPO3 fusion transcripts. a) EIF3E–RSPO2 gene fusion. The read evidence for the EIF3E(e1)–RSPO2(e2) fusion identified using RNA-Seq data is shown; b) PTPRK–RSPO3 gene fusion. The read evidence for PTPRK(e1)–RSPO3(e2) fusion identified using RNA-Seq data is shown.

Recently, Storm et al., have showed that inhibition of PTPRK-RSPO3 fusion prevents tumor growth and promotes differentiation [Storm EE, et al. 2016]. Until now, other fusion transcripts of various prevalence have been documented in

CRC, such as VWA2-TCF7L2, DHX35-BPIFA2, CASZ1-MASP2 [Hoff AM, et al. 2015].

Another fusion gene, recurrent in CRC and derived by intra-chromosomal translocation (chromosome 8), is a fusion involving the 1st exon of LACTB2 and the 22nd exon of NCOA2. The fusion protein lacks major functional domains of respective genes, indicative of a loss-of-function rearrangement and inactivation of the negative growth regulatory gene NCOA2. Therefore, the fusion with LACTB2 disrupted the tumor suppressing function of NCOA2, thereby promoting colorectal tumorigenesis [Yu J, et al. 2016].

1.4 Aim of the study

Colorectal carcinogenesis is a multistep process involving the gradual accumulation of genetic changes in colon epithelial cells. These alterations promote the malignant transformation of pre-malignant lesions of the colorectal mucosa into carcinoma. The understanding of the genomics and post-transcriptional mechanisms behind adenoma-to-carcinoma progression is crucial to identify biomarkers that characterize polyps at high risk of cancer development.

The aim of the study was to identify a molecular signature peculiar of the different steps of sporadic CRC development in synchronous lesions (polyps and cancer) and to detect novel fusion genes involved in sporadic CRC progression by RNA-

Sequencing (RNA-Seq) analysis by using an Illumina NextSeq 500 sequencing system.

The potential fusion events were analyzed by means bioinformatics analysis. As for the sample-specific fusion gene pattern, a combination of 2 different tools named Chimerascan and EricScript were used.

The identified fusion transcripts were further validated by RT-PCR, Sanger sequencing and FISH assay.

2. MATERIALS AND METHODS

2.1 Clinical Samples

To characterize the molecular architecture of the colonic tumorigenesis and to better describe the deregulated molecular signatures in each phase of CRC progression, eleven patients (8 males, 3 female; mean age: 67 ± 12 years) affected by CRC who underwent colorectal endoscopy at IRCCS-Casa Sollievo della Sofferenza Hospital, were prospectively recruited. Adenomatous and cancerous synchronous lesions, as well as normal mucosa, of CRC patients were collected. The samples obtained were immediately frozen in liquid nitrogen and then stored at -80°C . All the specimens were reviewed by the same experienced pathologist to confirm the histological diagnosis. Informed consent to take part in this study was obtained from all the patients. The study was approved by the Hospital's Ethics Committee.

2.2 RNA extraction from fresh frozen tissue

For the total RNA extractions, the fresh frozen tissue samples were transferred into M tubes with 1000 μl of TRIzol reagent (Thermo Fisher Scientific, Somerset, NJ 08873, USA) and were homogenized by gentleMACS™ Dissociator (Miltenyi Biotec, Bergisch-Gladbach, Germany) with “RNA_02_01” programme. Total RNA purification was performed with RNeasy Mini Kit (Qiagen, Valencia, CA)

according to manufacturers' instructions and treated with DNase I RNase free kit (Qiagen) to remove genomic DNA followed by precipitation.

Total RNA concentration was measured using the NanoDrop 1000 spectrophotometer (Thermo Fisher Scientific) and RNA quality was evaluated with RNA 6000 Nano Chip kit by using BioAnalyzer 2100 microcapillary electrophoresis system (Agilent Technologies, Santa Clara, CA, USA). Total RNAs with an RNA integrity number (RIN) ≥ 7.0 were accepted for the following RNA-Seq analysis.

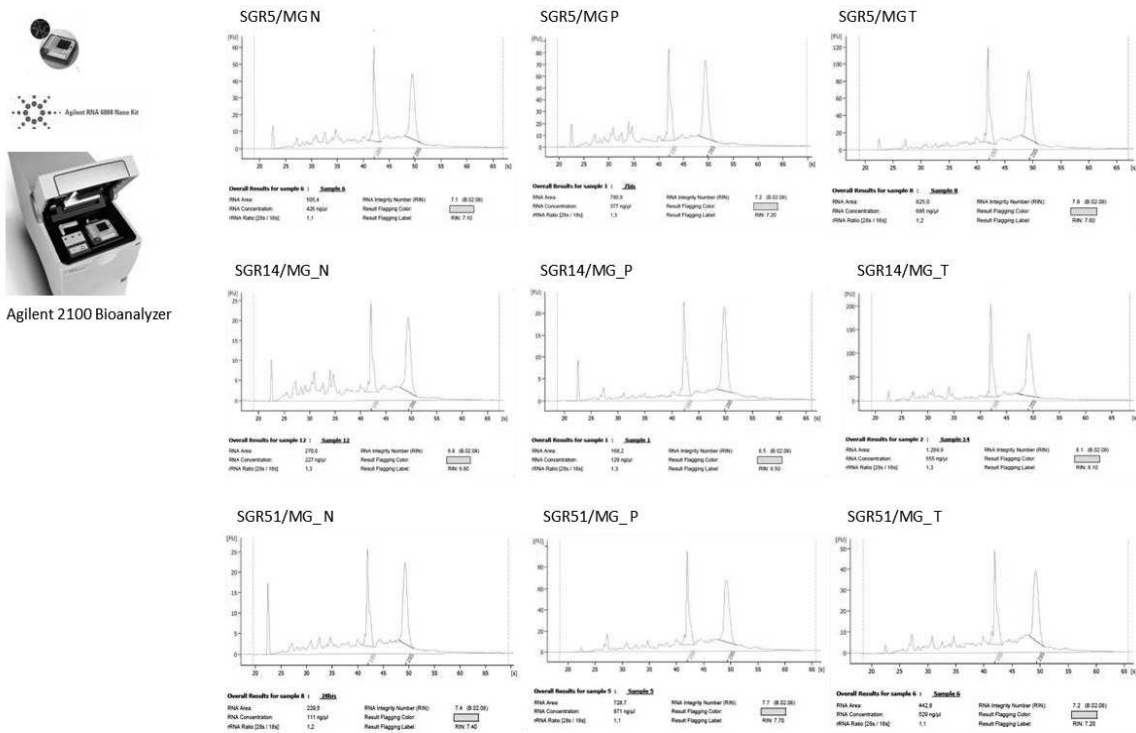


Figure 4. Chromatograms of microcapillary electrophoresis from RNA samples. A typical electropherogram of high quality RNA (RIN ≥ 7) includes clearly visible 28S/18S rRNA peak ratio and small 5S RNA.

2.3 Library preparation for RNA Sequencing

An aliquot (Ci=100ng/ μ l) of total RNA was used to construct cDNA libraries according to the TruSeq® Stranded Total RNA Sample Preparation kit as provided by the manufacturer (Illumina, San Diego, CA, USA) (figure 5).

In the first step, cytoplasmic and mitochondrial ribosomal RNAs (rRNAs) were removed using biotinylated, target-specific oligos combined with Ribo-Zero rRNA removal beads (Human Ribo-Zero Gold kit). Subsequently, the RNAs were purified and fragmented into small pieces by divalent cations under elevated temperature. The cleaved RNA fragments were copied into first strand cDNA using reverse transcriptase and random primers. The addition of Actinomycin D to the mix improved the strand specificity. A second strand cDNA was synthesized using DNA Polymerase I (Illumina).

After that, a single 'A' base was added to the 3' ends of these cDNA fragments (to prevent them from ligating to one another) and then was performed the ligation of the multiple indexing adapters, preparing them for hybridization onto a flow cell. The products were purified and enriched with PCR to create the final cDNA library.

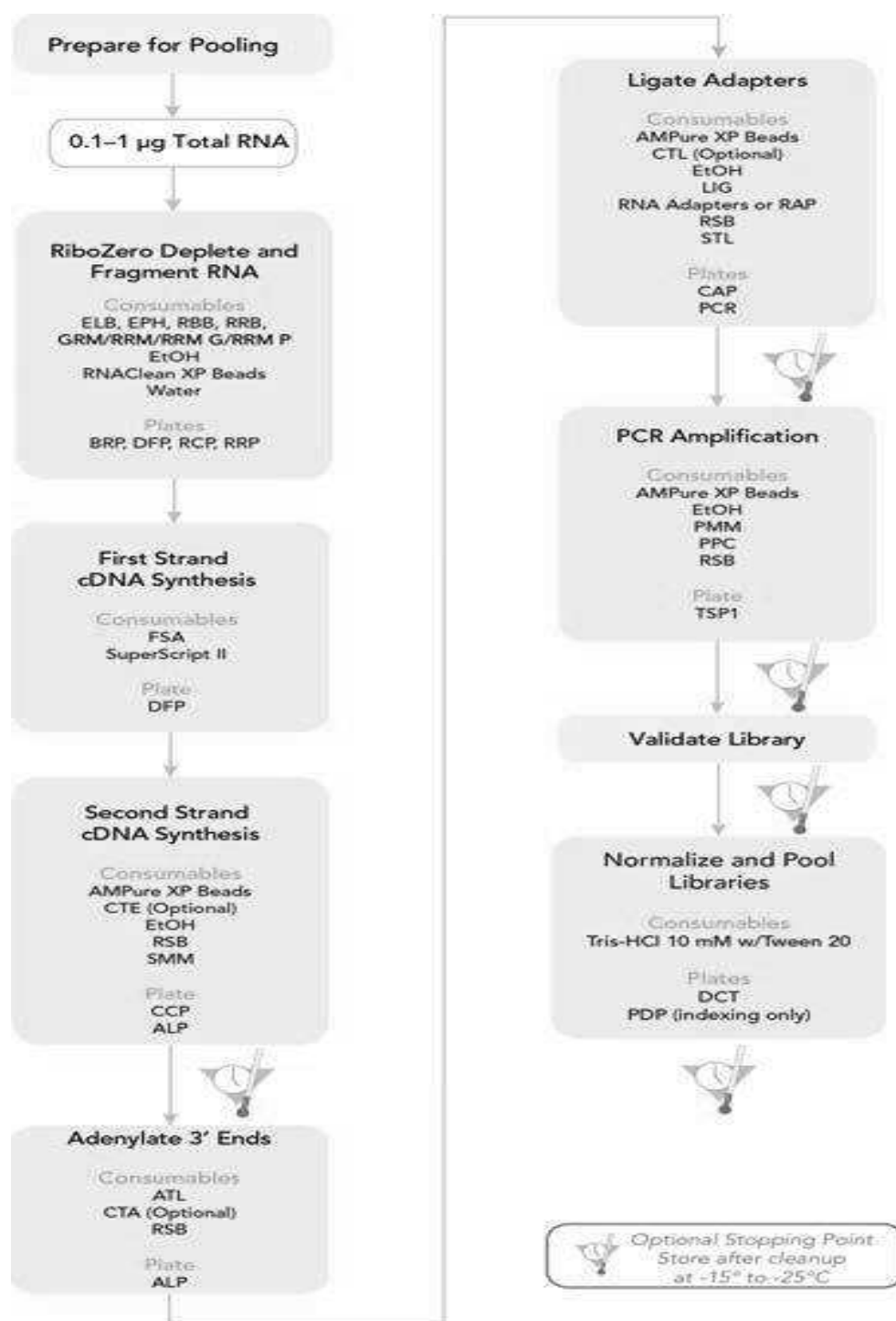


Figure 5: TruSeq Stranded Total RNA Sample Preparation LS Workflow

To validate the quality and to assess the size distribution of cDNA library, an aliquot was loaded on an Agilent High Sensitivity DNA chip and was running on

Agilent Technologies 2100 Bioanalyzer (figure 6). The final products showed a band of about 300 bp.

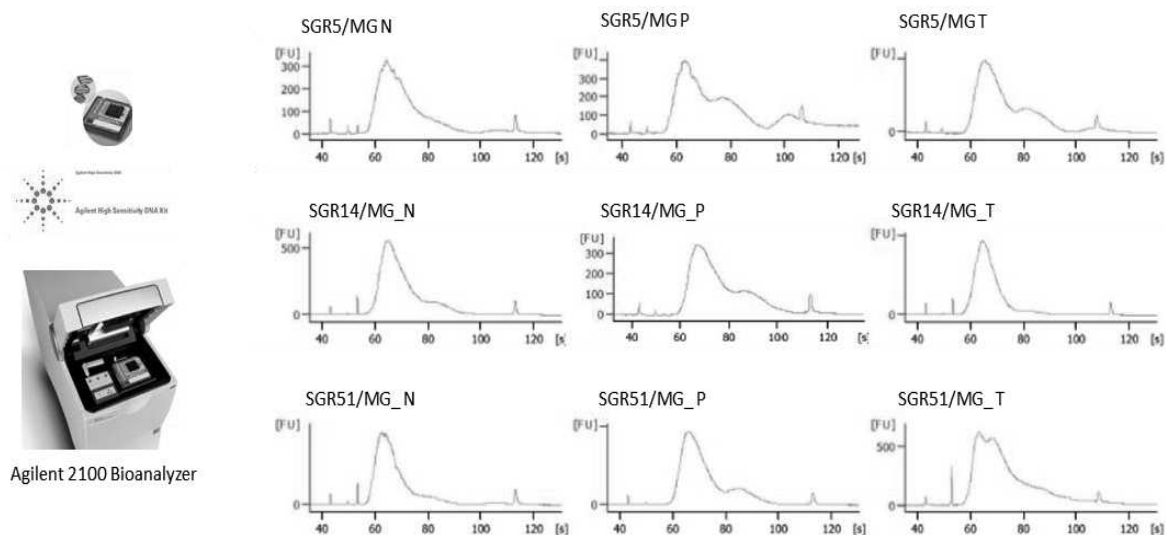


Figure 6: Example of TruSeq Stranded Total RNA Sample Preparation Library Size Distribution

For accurate quantitation of DNA library, the samples were analyzed by using a fluorometric based system (Qubit dsDNA HS Assay System; Thermo Fisher Scientific).

2.4 Sample pooling and sequencing

The DNA libraries were pooled and an aliquot (1,4 pM) was loaded into a High Output flow cell and was sequenced by means NextSeq500 Systems (Illumina). Considering the potential of RNA-Seq and its versatility, was used a paired-end approach (2X75bp), with around 80 million reads per sample.

2.5 Fusion Search

RNA-Seq data were analyzed in order to discovery the potential fusion genes in the different tumor stages.

Raw data (.fastq files) were quality-controlled using the FastQC v0.11.5 software package (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Reads were discarded if the average per-base phred values resulted less than 20, or trimmed by Trimmomatic [Bolger AM, et al. 2014] if the phred values of more than 5% of nucleotides at the extremities of the reads were lower than 20. Residual adapter sequences were removed by cutadapt [Martin M, 2011].

Around 50 million reads per sample were analyzed with a pool of tools: Chimerascan [Iyer MK, et al. 2011], EricScript [Benelli M, et al. 2012], pyPrada [Torres-García W, et al. 2014] and FusionCatcher [Daniel N, et al. 2014]. Each software package was run with standard parameters, yielding a list of putative fusion genes, annotated with the coordinates of the portions of the partner genes, together with the estimated breakpoints, the type of fusions (e.g., Inter-chromosomal, Read Through) and a reliability score. Their results were merged and considered together for further validation.

2.6 Fusion gene validation by Reverse Transcription-PCR

Expression of the candidate fusion genes was validated by Reverse Transcription-PCR (RT-PCR). A mixture containing 0.1µg of total RNA from each sample was

reverse transcribed for 10 min at 25°C and 2 h at 37°C using the High Capacity cDNA Reverse Transcription Kit (Thermo Fisher Scientific). Junction PCR specific primers were designed based on the RNA-Seq chimeric junction reads using primer3 software (https://primer3plus.com/primer3web/primer3web_input.htm) and the amplicon sequences were checked by BLAST against the human genome to ensure specificity.

PCR was performed in a final volume of 25 µL containing 2.5 µl 10× PCR Buffer (Thermo Fisher Scientific), 2.5 mM dNTPs, 25 mM MgCl₂, 15 pM junction specific PCR primers, 0.15 µL AmpliTaq Gold polymerase (Thermo Fisher Scientific), and 1 µL cDNA. Cycling PCR conditions consisted of an initial 10 minutes denaturation step at 94°C, followed by 35 cycles of 94°C for 1 minute, 58°C for 1 minute and 72°C for 1 minute, with a final extension at 72°C for 10 minutes. PCR products were visualized by ethidium bromide staining on 3% agarose gels.

cDNA samples from matching normal tissue were used as controls to be able to confirm that fusion genes were tumor-specific.

The junction sequences of potential fusion genes are reported in Appendix A while, PCR primers sequence, size of PCR products are listed in Table 1.

GENES	AMPLICON SIZE (bp)	PRIMERS SEQUENCES 5'-3'
RNF123 EX2 STAT3 EX24	137	<i>F:GCATGTCTTTCTCCCGCAAG</i> <i>R:TGGTGCAATTTTGGCTCACT</i>
PLK1 EX10 ERN2 EX22	150	<i>F:TTTGTACATGTTCTGGGTGTGG</i> <i>R:GTCTCGAACTCCTGAGCTCA</i>
PDLIM2 EX10 CCAR2 EX2	118	<i>F:CCTGCGCAGAGATTGGCT</i> <i>R:CTGGCGCTTAAACTGGGAC</i>
MRPS31 EX6 SUGT1 EX3	131	<i>F:TGAACAACCCCTTCAGAATGG</i> <i>R:CTGTGCATCATCTGGTTTCTGT</i>
LPHN1 EX1 SUZ12 EX2-3	167	<i>F:CACCTCCTACCCGCTTCC</i> <i>R:TGGTGCTATGAGATTCCGAGT</i>
HPSE2 EX1 HSD11B2 EX3	173	<i>F:AGCACCTCATTCAATCCCT</i> <i>R:CATGCAGCTACGGAAAGTGG</i>
HDAC1 EX14 MARCKSL1 EX2	152	<i>F:GAGGTTGCTAGTCTAGTTTCCTT</i> <i>R:AACCATTCTTCCACCTGG</i>
ARSA EX1 TNS4 EX13	259	<i>F:GTAAGTCACTTGGCGCTGAC</i> <i>R:CGCCAAAGTCTGCAGAATGT</i>
EIF5AL1 EX1 MSH3 EX24	155	<i>F:AGACTAGTCATGTGGCAGGA</i> <i>R:GGTGATCTGCCTACCTTGG</i>
ERBB2 EX25 MIEN1 EX3	187	<i>F:GCAGGGCTTCTTCTGTCCA</i> <i>R:CCACTCTTCGCCTCGGTAAT</i>
GUCY2C EX27 PLBD1 EX 1	186	<i>F:TGAATCAGATGTGTGTTCTCAGT</i> <i>R:CCATTTATCTCCAGCAGCGG</i>
HSPE1 EX3 MOB4 EX2	131	<i>F:ACCAGTTAGCGTGAAAGTTGG</i> <i>R:CAGCTAGTGTACTGTCCATTCA</i>

Table 1. Junction PCR specific primers sequence, size of PCR products

2.7 DNA Sequencing

The amplicons were sequenced from both ends using an aliquot (3.2 pM) of the PCR reaction primers in presence of BigDye Terminator Cycle Sequencing Kit v. 1.1 (Thermo Fisher Scientific). After purification by using centrisep columns (Thermo Fisher Scientific), sequencing reactions were loaded on 3500 DX Genetic Analyzer capillaries (Applied Biosystems) and analyzed using the Sequencing Analysis software v5.4.

2.8 FISH Assay

The FISH assay was performed on 4-5 μm thick sections obtained from formalin-fixed, paraffin-embedded (FFPE) pathologic tissue. The sections mounted on coated slides were deparaffinized overnight at 60°C and in 3 change of Bioclear (Bio-Optica, Milan, Italy) at 40°C, 40 minutes each. The slides were incubated in pretreatment solution (Pretreatment KIT II) (Abbott Molecular, Abbott Park, Illinois, USA) at 80°C for 10 minutes, in protease solution at 37°C for 16 minutes, and in H₂O at room temperature (RT) for 3 minutes. The slides were dehydrated by immersing in 70%, 85% and 100% ethanol solution at RT. The slides were held dry at RT for several hours. An aliquot of denatured probe (Vysis FOXO1 Break Apart FISH Probe Kit; Abbott Molecular) was applied and the samples were covered with a cover glass and sealed carefully with rubber cement. The slides were incubated at 80°C for 5 minutes and overnight (18-24 hs) at 37°C. The cover glasses were removed and the sections were washed with 0.4xSSC, 0.1% NP-40 at 72°C for 2 minutes, and 2xSSC, 0.1% NP-40 at RM for 2 minutes, and air-dried in darkness. An aliquot of DAPI solution (Abbott Molecular) was applied and a cover glass was put on.

The slides were examined on a ZEISS AXIO M1 (Carl Zeiss Meditec, Jena, Germany) fluorescence microscope with an 120W halogen lamp and spectrum filters (Green, Orange and DAPI) provided by Abbott Molecular. The images were analyzed by means the ASIs (Applied Spectral Imaging) FISHView software.

3.RESULTS

3.1 Identification of fusion transcripts by EricScript and ChimeraScan algorithms

Fusion genes were investigated in normal, polyps and tumor tissues of eleven individuals affected by CRC by means of four different fusion-search algorithms widely used in scientific literature [Carrara M, et al. 2013]. Each of such methods made use of different strategies for finding fusion candidates, both in the evaluation of the reads spanning the potential fusion junctions and in the manner of scoring the inferred fusion gene pairs.

Given the different outputs of the 4 software packages, we filtered our fusions according to several criteria, including the fact that junctions were spanned by a sufficient number of short-reads, and whether the candidate fusions were present in known datasets. In particular, we focused on the results of EricScript and ChimeraScan, since the requirements of both FusionCatcher and pyPrada were only partially compliance with the specifics of our experiments.

While the ability to identify fusion genes by means a RNA-Seq is very powerful, verification and validation of this data is labor intensive, and was performed by means of the following workflow.

A filter has been applied to select fusion genes that:

- i. were in common between adenoma and adenocarcinoma samples, while, the same had to be absent in the normal mucosa;
- ii. were composed of a genes with a low number of isoforms;
- iii. were consisted of a genes with a function associated with cancer such as invasion, cell movement, apoptosis, cell death, tumorigenesis and differentiation.

Using these specific filtering criteria, a dataset of 12 tumor-specific fusion events has been selected for further validation analyses.

In Table 2 are indicated the selected fusion genes, the estimated breakpoints, the type of fusions and a reliability score.

fusion (EricScript)	GeneName 5p	GeneName 3p	chr 5p	Breakpoint1 (end 5p)	strand 5p	chr 3p	Breakpoint2 (start 3p)	strand 3p	fusiontype	JunctionSequence	EricScore
RNF123-STAT3	RNF123	STAT3	3	49728680	+	17		-	inter-chr	ccgcaagagctataggctgacctcagatgctgagaaatccag ggtcagagCTACTCGGGAAGCTGAGGCAAGGAG AATCGCTTGAACCTGAGAGGCGGAGG	0.91586235
PLK1-ERN2	PLK1	ERN2	16	23701614	+	16	23702074	-	Cis	gtgggttctacagcctgtcccccctcccctcaacccaccat gaattGCTGGGTGCACTGGCTCACACCTGTAA TCCCAAGCATTTTGGGAGGCTGAG	0.690112004
MRPS31-SUGT1	MRPS31	SUGT1	13	41323274	-	13	53231667	+	intra-chr	gtggacaaaaggggaaactatgggagtcccaat gaagcagGAGCTGACTAAAGCTTTGGAAACAG AAACCAAGATGATGCACAGATATTATG	0.734629203
LPHN1-SUZ12	LPHN1	SUZ12	19	14316797	-	17	30267305	+	inter-chr	cgagccgaggagagacacgctggccgaccagagaggg cgtgtgacagAGCCAAACACAGATCTATAGAT TTCTTGAACCTCGGAAATCTCATAGCACA	0.855475528
EIF5AL1-MSH3	EIF5AL1	MSH3	10	81274508	+	5		+	inter-chr	aagactgtaaaatgaatccagaggtgaccaagcattgaatt aacaatGGTGGCTCATGCCTGTAAATCCCAAGCA CTTTGGGAGGCCAAAGGTAGGCAAG	0.532811583
GUCY2C-PLBD1	GUCY2C	PLBD1	12	14765813	-	12	14721126	-	Read-Through	acctccaactctggaaccttaccagcagttgtccagggagct tctacCTGTGGAGCCCTCTCCAAGAACAGCA GAGGATCCGAGCTGCGTGTAGGCA	0.896360711
HSPE1-MOB4	HSPE1	MOB4	2	198367852	+	2	198388348	+	Read-Through	aagtcttctccagaatgaggcacaagatggttctagat gacaagGATTTCTATAATTGGCCTGATGAAT CCTTTGATGAAATGGACAGTACT	0.821493951
fusion (ChimeraScan)	GeneName 5p	GeneName 3p	chr 5p	Breakpoint1 (end 5p)	strand 5p	chr 3p	Breakpoint2 (start 3p)	strand 3p	fusiontype	JunctionSequence	Chimera Score
PDLIM2-CCAR2	PDLIM2	CCAR2	8	22455537	+	8	22463248	+	intra-chr	agagatggctgtgggectcagttcccatttataaagttaa aatctGCCTTTTCCCAACGACTCTGAAAAGAGG ACAGCGTCCCAATGICCCAGTIT	5
HPSE2-HSD11B2	HPSE2	HSD11B2	10	100995631	-	16	67469859	+	inter-chr	tctcttactggtctcgtactgactaattgtcttatctaaag tgtgGGCCTGTGGGCTCTGCAACAACGCA GGCCAAATGAAGTAGTGTCTGAT	2
HDAC1-MARCKSL1	HDAC1	MARCKSL1	1	32799223	+	1	32799429	-	Adjacent_C onverging	agatactatttctttttgagcctcttgaataaafgtacat ttcTAAAGCACCACTAAAGGACGACATTT ATTCTTTTCCAAATGITACAGTA	2
ARSA-TNS4	ARSA	TNS4	22	51066600	-	17	38632079	-	inter-chr	gccggtaccggctgccggcctcccgctcgccccccc cgtgacctgtCTACTGTTTGCAAAAGACAAA CATTTTATTTTTCATGATAGGAGCTGTAG	4
ERBB2-MIEN1	ERBB2	MIEN1	17	37883255	+	17	37885408	-	Adjacent_C onverging	ccccggcctggggcaggtccacacaggcaccgagct catctaccagATTAGTGTGAGGCCACTTT ACTGCCAATAGCTGACATTCCTGGGT	4

Table2: List of putative fusion genes selected by EricScript and ChimeraScan tools, with the name of the partner genes, the estimated breakpoints, the type of fusions (e.g., Inter-chromosomal, Read Through) and a reliability score.

Our analysis revealed that the largest subset of these genes (involved in the fusion events) were cancer related (SUZ12, HSD11B2, ERBB2, HSPE1), some were colon cancer related (SUGT1, HPSE2, TNS4, MSH3, GUCY2C), some were related to the cytoskeleton, cell adhesion and migration (MARCKSL1, MIEN1)

followed by other functional categories such as cellular growth, proliferation and apoptosis (STAT3, PLK1, CCAR2, HDAC1).

3.2 Fusion junction validation by RT-PCR and Sanger sequencing

The candidate fusion gene MRPS31-SUGT1 was successfully validated in the pathologic specimens (polyp and tumor) of 3 patients (27.3%) (BKCH26810, BKCH29410, SGR32) and was absent in the corresponding control tissue.

MRPS31-SUGT1 is a fusion between exon 6 of MRPS31 (mitochondrial ribosomal protein S31) and exon 3 of SUGT1 (SGT1 homolog, MIS12 kinetochore complex assembly cochaperone). Both genes are located on the long arm of chromosome 13.

MRPS31 encodes for a mitochondrial ribosomal protein (395 aa) involved in protein synthesis within the mitochondrion and associated with type 1 diabetes. This gene is encoded by 7 exons and has only one known transcript (NM_005830) in RefSeq database.

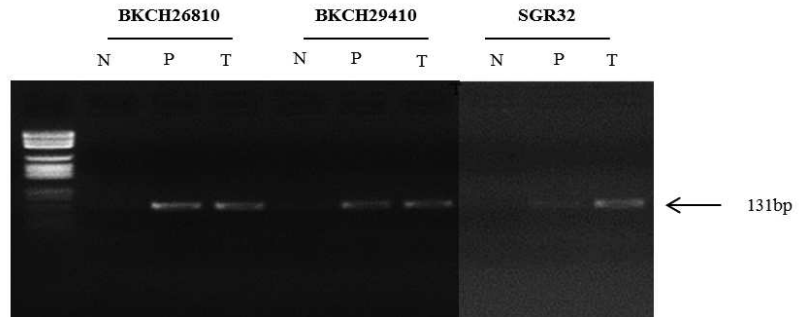
SUGT1 encodes for a highly conserved nuclear protein (365 aa) that interacts with heat shock protein 90 (Hsp90). SUGT1 is encoded by 14 exons and has two known transcripts (NM_001130912; NM_006704) in RefSeq database.

The fusion of these two adjacent genes resulted in the early truncation, loss of domains of the SUGT1 gene.

The validation results of the fusion between MRPS31 and SUGT1 genes were shown in figure 7. RT-PCR products, visualized on a agarose gel, presented the 131 bp band (figure 7a). Sequence of the band (figure 7b) confirmed the fusion between exon 6 of MRPS31 and exon 3 of SUGT1.

a

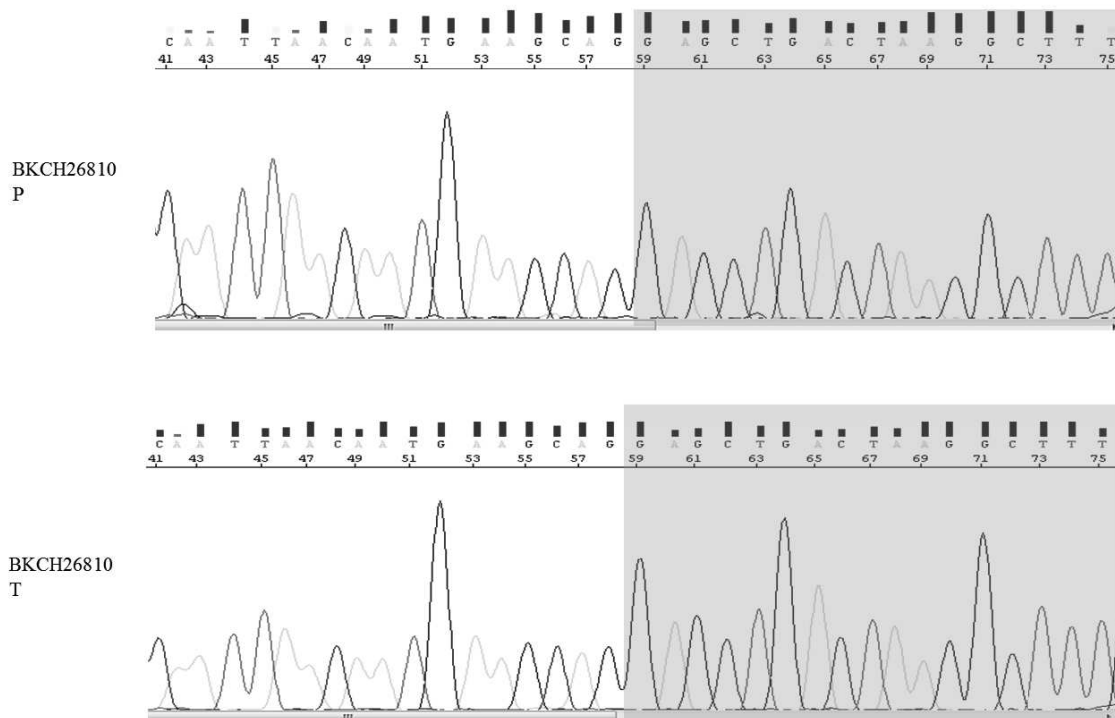
MRPS31-SUGT1



b

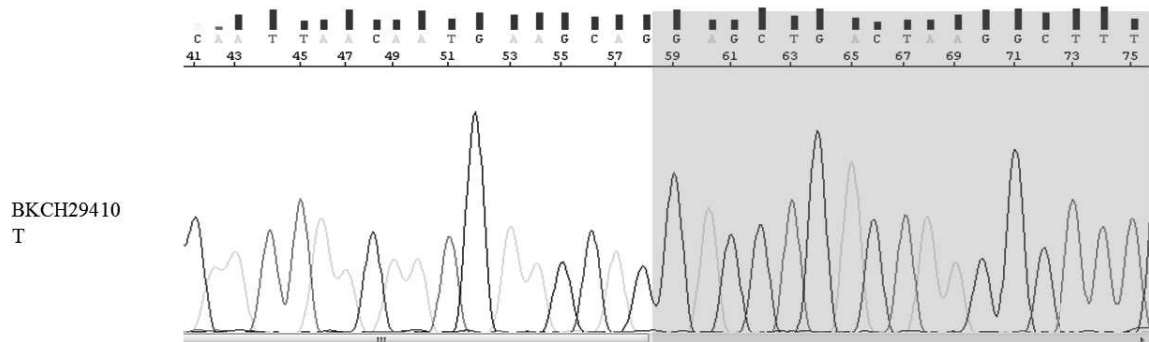
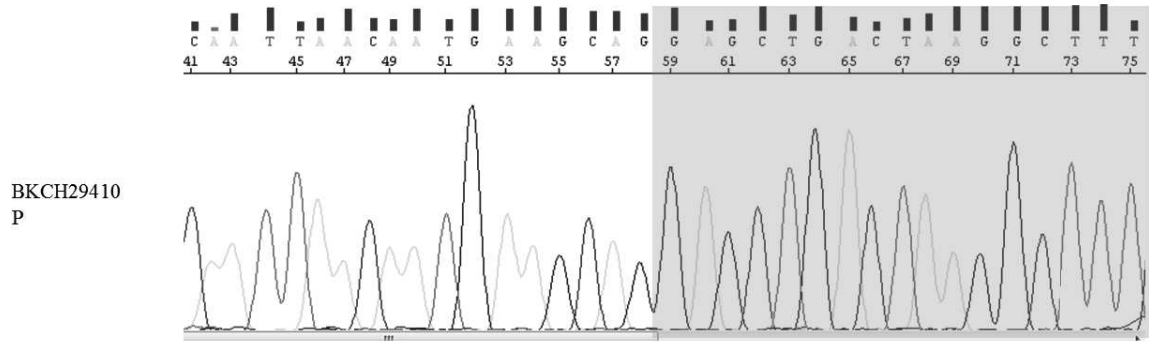
MRPS31 exon6

SUGT1 exon3



MRPS31 exon6

SUGT1 exon3



MRPS31 exon6

SUGT1 exon3

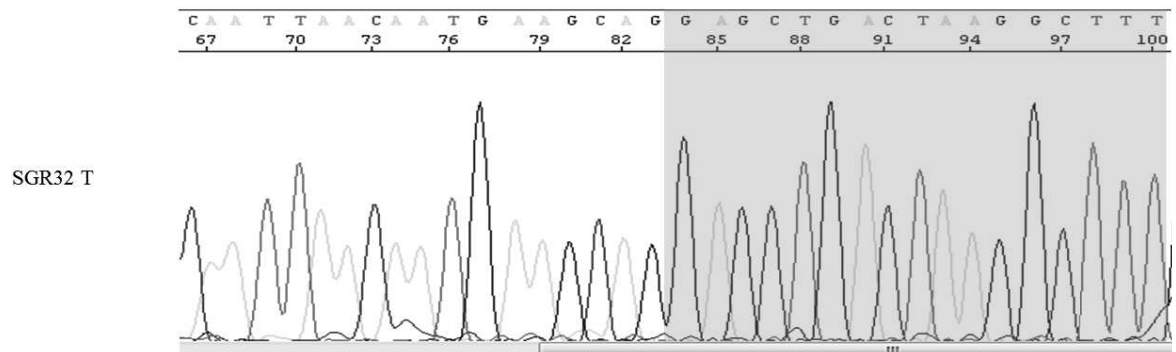
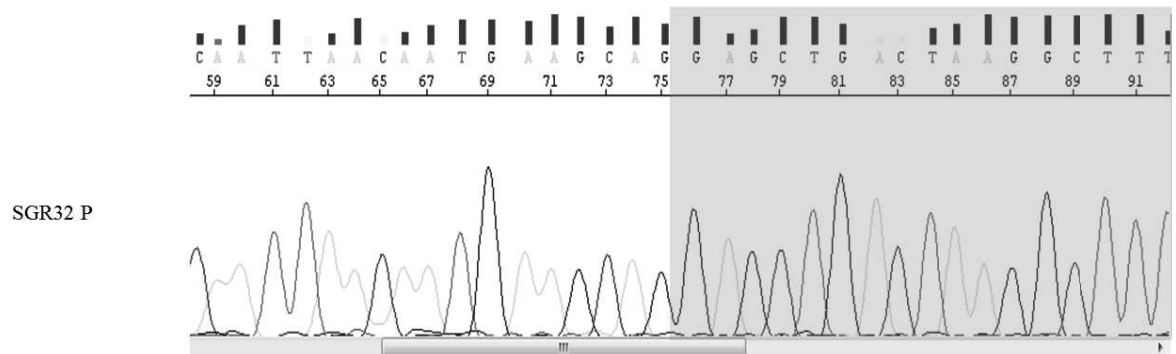


Figure 7: Detection of novel fusion gene, MRPS31-SUGT1. a) RT-PCR analysis of cDNA derived from pathologic tissues (P: polyp; T: tumor) and adjacent normal mucosa (N) in three CRC patients (BKCH26810, BKCH29410, SGR32). RT-PCR products were visualized on a agarose gel; b) Sequencing analysis of the MRPS31-SUGT1 fusion transcript in patients: BKCH26810, BKCH29410, SGR32. The sequencing electropherograms revealed the fusion occurring between exon 6 of MRPS31 and exon 3 of SUGT1.

For 8 of predicted fusion genes named RNF123-STAT3, PDLIM2-CCAR2, LPHN1-SUZ12, HPSE2-HSD11B2, ARSA-TNS4, EIF5AL1-MSH3, ERBB2-MIEN1, HSPE1-MOB4 we did not observed an RT-PCR product, indicating that these are either false positive calls or that the fusion gene RT-PCR detection assay was suboptimal. In addition, the PLK1-ERN2, HDAC1-MARCKSL1 and GUCY2C-PLBD1 fusions showed a weak RT-PCR product in the corresponding control tissue resulting not specific for the tumor sample.

3.3 MRPS31 gene rearrangement validation by FISH assay

To evaluate the MRPS31 gene region rearrangement in 2 FFPE patients (BKCH26810, BKCH29410), positive to the MRPS31-SUGT1 fusion, a FISH assay by means Vysis FOXO1 Break Apart FISH Probe Kit was performed. One case (SGR32) was excluded due to insufficient tissue.

The Vysis FOXO1 Break Apart FISH Probe consists of two DNA probes, targeting the 13q14 region :

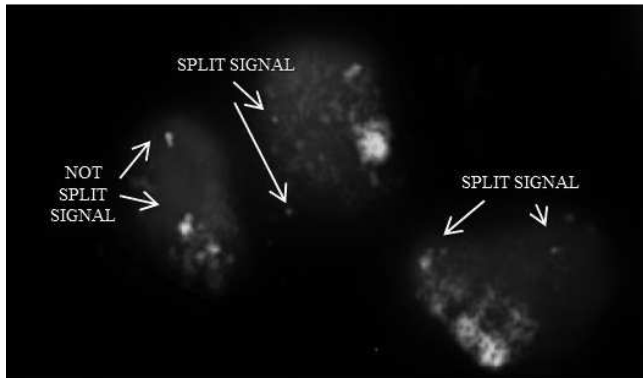
- a 720kb probe, labeled in Spectrum Green, spreads proximally from the FOXO1 gene;

- a 650kb probe, labeled in Spectrum Orange, extends distally from the FOXO1 gene. In particular, this probe covers the entire part of MRPS31 gene (13q14.1) which is involved in MRPS31-SUGT1 fusion.

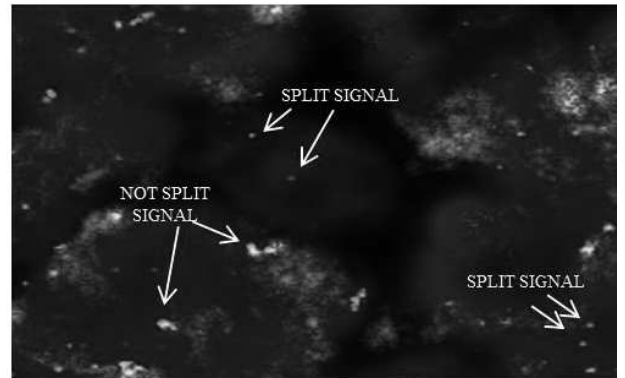
As shown in figure 8a, in the cells of polyps were observed both single signals, indicative of MRPS31 gene-breaking, and fusion signals usually detected in normal cells. Whereas, the cells of tumor tissues showed single signals, suggesting rearrangement of the MRPS31 gene region (figure 8b).

a

BKCH26810P



BKCH29410P



b

BKCH26810T



BKCH29410T

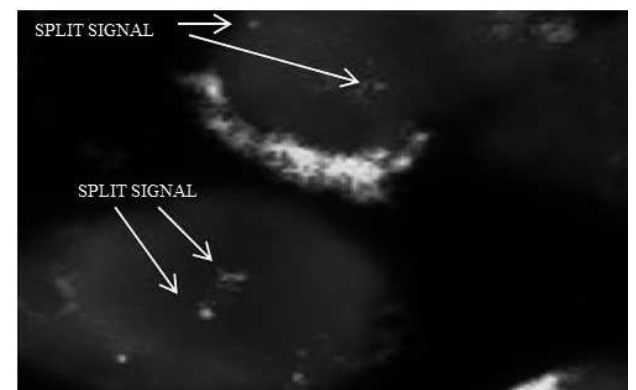


Figure 8: FISH detection of MRPS31 gene region rearrangement in CRC patients positives to the MRPS31-SUGT1 fusion. Results showed an aberrant hybridization of the Vysis FOXO1 Break Apart probe in the pathologic tissues. a) in the cells of polyps were observed both single orange and green signals (split signals), and fusion signals (not split signal); b) in the tumor cells were observed single signals (split signals) suggesting rearrangement of the MRPS31 gene region.

4. DISCUSSION

Fusion genes are a class of genomic rearrangements in which two genes are fused, leading to the production of chimeric transcripts that may have an aberrant activity. In recent years, the development of high-throughput sequencing has led to an increasing number of fusion genes identified in CRC, although the results are not quite conclusive. Therefore, further studies are required to unveil the effects of these chromosomal aberrations on changes in early phases of sporadic CRC tumorigenesis and to identify clinically confident targets for the therapeutic intervention of CRC.

In the present study, the RNA-Seq technology has been used in order to interrogate the transcriptome and to identify structural rearrangements peculiar of each phase of the tumorigenic process in a cohort of 11 CRC patients. For this purpose, adenomatous and cancerous synchronous lesions, as well as normal mucosa, were collected. The extracted total RNAs were used for DNA library preparation and further sequencing by paired-end RNA-Seq.

Fusions were detected in RNA-Seq data by means of two different fusion-search algorithms: EricScript and ChimeraScan.

EricScript represents a computational methods that uses a combination of four alignment processes to identify reads spanning the potential fusion junctions and a single score to distinguish genuine fusions from false positive events [Benelli M, et al. 2012].

ChimeraScan is a computational tool that allows to process long (>75 bp) paired-end reads and to detect of reads spanning a fusion junction [Iyer MK, et al. 2011].

By means of these two approaches several fusion events were found. Following a restrictive set of filtering criteria (only tumor specific fusions; only fusions composed of a genes with a low number of isoforms and with a function associated with cancer), 12 of these were selected for validation (Table 2). Many of the identified genes, have a known involvement in biological pathways linked in mitosis, cell cycle control, cell proliferation, apoptosis, invasion and in cytoskeletal remodelling.

The subsequent validation by means RT-PCR and Sanger sequencing on whole CRC cohort, confirmed that one of these 12 fusions, named MRPS31-SUGT1, was a tumor specific recurrent event. In particular, this newly fusion was present in adenoma and adenocarcinoma tissues of 3 patients (3/11 patients, 27.3%), and as expected no band was detect in normal tissues. The fusion, between exon 6 of MRPS31 (codon 319) and exon 3 of SUGT1 (codon 33), was generated through an intra-chromosomal translocation, clustered on chromosome 13. The protein produced from this fusion transcript was 352 amino acids in length and the rearrangement caused the complete loss of SUGT1 domains.

SUGT1 is devoid of catalytic domains, but contains three distinct protein interaction motifs: the amino-terminal region containing a tetratricopeptide repeat

(TPR), a central CS domain (shared by CHORD-containing protein and SGT1), and a carboxyl-terminal SGS domain (SGT1-specific).

A previous study has demonstrated that the CS and SGS domains are required for association with nucleotide oligomerization domain (NOD)1. The wild type SUGT1 plays an essential role in NOD1 activation and it is involved in immune response [da Silva Correia J, et al. 2007], in bacterial recognition and host defence [Franchi L, et al. 2009]. In a recent pediatric study, based on whole-exome sequencing, both SUGT1 and NOD2 were found significantly associated with pediatric inflammatory bowel disease (pIBD) in a cohort of 136 patients [Andreoletti G, et al.2017]. IBD is a chronic inflammatory disorder and represents a major risk factor for CRC. The pathogenesis of CRC in IBD is influenced by environmental and genetic factors and many of the genetic alterations associated with development of sporadic CRC also play roles in colitis-associated CRC [Kim ER, et al. 2014].

SUGT1 is a cochaperone of Hsp90, and the TPR domain of SUGT1 binds Hsp90 while the CS domain of SGT1 interacts with the ATPase domain of Hsp90 [Takahashi A, et al. 2003].

SUGT1, as a cochaperone of Hsp90, is involved in multiple biological processes through interaction with different protein complexes. For example SUGT1 mediates the kinetochore assembly through its interaction with Skp1; SUGT1 and Hsp90 collaborates in kinetocore-microtubule attachment by stabilizing the MIS12

complex at kinetochores; SUGT1-HSP90 complex is required for CENP-A deposition at centromeres [Niikura Y, et al. 2017].

RNA interference-mediated depletion of SUGT1 results in a marked alteration of kinetochores, problems in chromosome segregation, aneuploidy and thus cancer [Steensgaard P, et al. 2004].

MRPS31 (or Imogen 38) encodes for a mitochondrial ribosomal protein that helps in protein synthesis within the mitochondrion. It is more abundant in the mitochondria of pancreatic β cell, but is also distributed in other tissues. This protein has been associated with type 1 diabetes (insulin dependent diabetes mellitus) and plays a pathological role in autoimmune attack functioning as an auto antigen recognized by T cells. Its relation to the etiology of this disease remains to be clarified [Arden SD, et al.1996; Cavdar Koc E, et al.2001]. Furthermore, in a recent work based on transcriptome array data, MRPS31 was found related with thyroid cancer progression [Xu Y, et al.2014].

In the present study, was further confirmed the MRPS31 rearrangement using FISH assay. The results from Vysis FOXO1 Break Apart FISH Probe kit were in 100% concordant to the previous validation tests in two CRC patients positives to MRPS31-SUGT1 fusion. Results showed an aberrant hybridization of the Vysis FOXO1 Break Apart probe and evident MRPS31 gene breaking signals, both in the cells of polyps and in the tumor cells. A future work should involve *in vitro*

assays to elucidate the biological significance of MRPS31-SUGT1 fusion gene, the oncogenic capacity and response to antitumor drugs.

In conclusion, we performed whole-transcriptome sequencing of 11 CRC patients to identify novel and recurrent fusions already present in early stage of this disease. A MRPS31-SUGT1 fusion was detected in synchronous lesions (polyps and tumours) of three subjects and further validated using RT-PCR, Sanger sequencing and FISH. This fusion gene represents a newly, unreported and recurrent event in promoting colorectal tumorigenesis and may be used a potential targeted therapeutics. The development and use of such therapeutics approaches will allow us to practice personalized medicine and improve health care.

APPENDIX A

**(RNF123).....ccgcaagagctataggctgacctcagatgctgagaaatccagggtcacagCTACTCGGGAGGCT
GAGGCAGGAGAATCGCTTGAACCTGAGAGGCGGAGG...(STAT3)**

**(PLK1)...gtgggttctacagccttgccccctccccctcaaccccaccatatgaattGCTGGGTGCAGTGGCTCA
CACCTGTAATCCCAGCATTTTGGGAGGCTGAG...(ERN2)**

**(PDLIM2)...agagattggctgtgggcctcagttccccattttataaagttttaaatactGCCTTTTCCCCACGACT
CTGAAAGAGGACAGCGTTCCAATGTCCCAGTTT...(CCAR2)**

**(MRPS31)...gtggacaaaagaggggaaactatgggagtccaattaacaatgaagcagGAGCTGACTAAGGC
TTTGGAACAGAAACCAGATGATGCACAGTATTATTG...(SUGT1)**

**(LPHN1)...cgagccgcaggagagacacgctgggcccagaccagagagcgtggacagAGCCAACACAGATC
TATAGATTTCTTCGAACTCGGAATCTCATAGCACCA...(SUZ12)**

**(HPSE2)...tctcttctactgggtctcgctagtactaattgtccttatctaaagtgGGCCTGTGGGGCCTCGTC
AACAAACGCAGGCCACAATGAAGTAGTTGCTGAT...(HSD11B2)**

**(HDAC1)...agatactatthtcattttgtgagcctctttgtaataaaaatggtacatttcTAAAGCACCACTAAAGG
GACGACATTTATTCTTTTCCAAATGTTACAGTA...(MARCKSL1)**

**(ARSA)...gccgtaccgggctgcccgccttcgcctcgcccgccttgacctgtCTTACTGTTTTGCAAAG
ACAAACATTTTATTTTTCATGATAGGAGCTGTAG...(TNS4)**

**(EIF5AL1)...aagactgtgaaaatgaatccagaggtgaccaagcattgaatttaacaatGGTGGCTCATGCCTG
TAATCCCAGCACTTTGGGAGGCCAAGGTAGGCAGA...(MSH3)**

**(ERBB2)...cccggcgctggggcatggtccaccacaggcaccgcagctcatctaccagATTAGTGTTTGTAGC
GCCACTTTACTGCCAATAGCTGACATTGCCCTGGGT...(MIEN1)**

**(GUCY2C)...accttccactctggaaccttattccagcagttgtccagggagcttctacCTGTGGAGGCCTCTCC
AGAAACAGCAGAGGATCCGAGCTGCGTGTAGGCA...(PLBD1)**

**(HSPE1)...aagttcttctcccagaatattggaggcaccaaagtagttctagatgacaagGATTTCTATAATTGGCC
TGATGAATCCTTTGATGAAATGGACAGTACACT...(MOB4)**

REFERENCES

- Abatangelo L, et al. Comparative study of gene set enrichment methods. *BMC Bioinformatics*. 2009 Sep 2;10:275.
- Andreoletti G, et al. Exome Analysis of Rare and Common Variants within the NOD Signaling Pathway. *Sci Rep*. 2017 Apr 19;7:46454.
- Arden SD, et al. Imogen 38: a novel 38-kD islet mitochondrial autoantigen recognized by T cells from a newly diagnosed type 1 diabetic patient. *J Clin Invest*. 1996;97:551-61.
- Barr FG. Fusion genes in solid tumors: the possibilities and the pitfalls. *Expert Rev Mol Diagn*. 2016 Sep;16(9):921-3.
- Bass AJ, et al. Genomic sequencing of colorectal adenocarcinomas identifies a recurrent VTI1A TCF7L2 fusion. *Nat Genet*. 2011; 43: 964-968.
- Benelli M, et al. Discovering chimeric transcripts in paired-end RNA-seq data by using EricScript. *Bioinformatics*. 2012 Dec 15;28:3232-9.
- Bolger AM, et al. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014 Aug 1;30:2114-20.
- Carethers JM, et al. Genetics and Genetic Biomarkers in Sporadic Colorectal Cancer. *Gastroenterology*. 2015 Oct;149:1177-1190.e3.
- Carrara M, et al. State-of-the-art fusion-finder algorithms sensitivity and specificity. *Biomed Res Int*. 2013;2013:340620.

- Cavdar Koc E, et al. The small subunit of the mammalian mitochondrial ribosome. Identification of the full complement of ribosomal proteins present. *J Biol Chem*. 2001 Jun 1;276:19363-74.
- Consiglio A, et al. BEAT: Bioinformatics Exon Array Tool to store, analyze and visualize Affymetrix GeneChip Human Exon Array data from disease experiments. *BMC Bioinformatics*. 2012 Mar 28;13 Suppl 4:S21.
- da Silva Correia J, et al. SGT1 is essential for Nod1 activation. *Proc Natl Acad Sci U S A*. 2007 Apr 17;104:6764-9.
- Daniel N, et al. FusionCatcher - a tool for finding somatic fusion genes in paired-end RNA-sequencing data. *BioRxiv: the preprint server for biology*. November 19, 2014.
- Fearon ER, et al. A genetic model for colorectal carcinogenesis. *Cell*. 1990;61:759–67.
- Franchi L, et al. Function of Nod-like receptors in microbial recognition and host defense. *Immunol Rev*. 2009 Jan;227:106-28.
- He TC, et al. Identification of c-MYC as a target of the APC pathway. *Science*. 1998;281:1509–12.
- Hoff AM, et al. Novel RNA variants in colorectal cancers. *Oncotarget*. 2015 Nov 3;6:36587-602.
- Iyer MK, et al. ChimeraScan: a tool for identifying chimeric transcription in sequencing data. *Bioinformatics*. 2011 Oct 15;27:2903-4.

- Kim ER, et al. Colorectal cancer in inflammatory bowel disease: the risk, pathogenesis, prevention and diagnosis. *World J Gastroenterol*. 2014 Aug 7;20:9872-81.
- Kim J, Recurrent fusion transcripts detected by whole-transcriptome sequencing of 120 primary breast cancer samples. *Genes Chromosomes Cancer*. 2015 Nov;54:681-91.
- Kim TM, et al. Clinical applications of generation sequencing in colorectal cancers. *World J Gastroenterol*. 2013; 19: 6784-6793.
- Kinzler KW, et al. Identification of a gene located at chromosome 5q21 that is mutated in colorectal cancers. *Science*. 1991;251:1366-70.
- Lee HS, et al. Molecular Testing for Gastrointestinal Cancer. *J Pathol Transl Med*. 2017 Mar;51:103-121.
- Lin E, et al. Exon array profiling detects EML4-ALK fusion in breast, colorectal, and non-small cell lung cancers. *Mol Cancer Res*. 2009 Sep;7:1466-76.
- Maglietta R, et al. On the reproducibility of results of pathway analysis in genome-wide expression studies of colorectal cancers. *J Biomed Inform*. 2010 Jun;43:397-406.
- Markowitz SD, et al. Molecular origins of cancer: Molecular basis of colorectal cancer. *N Engl J Med*. 2009. 361: 2449-2460.
- Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, [S.l.], v. 17, n. 1, p. pp. 10-12, may. 2011.

- Niikura Y, et al. SGT1-HSP90 complex is required for CENP-A deposition at centromeres. *Cell Cycle*. 2017 Sep 17;16:1683-1694.
- Nome T, et al. High frequency of fusion transcripts involving TCF7L2 in colorectal cancer: novel fusion partner and splice variants. *PLoS One*. 2014 Mar 7;9:e91264.
- O'Connell JB, et al. Colon cancer survival rates with the new American Joint Committee on Cancer sixth edition staging. *J Natl Cancer Inst*. 2004; 96: 1420-1425.
- Parker BC, et al. Fusion genes in solid tumors: an emerging target for cancer diagnosis and treatment. *Chin J Cancer*. 2013 Nov;32:594-603.
- Powell SM, et al. APC mutations occur early during colorectal tumorigenesis. *Nature*. 1992;359:235-7.
- Rowley JD. A new consistent chromosomal abnormality in chronic myelogenous leukemia identified by quinacrine fluorescence and Giemsa staining. *Nature*. 1973;243:290-3.
- Seshagiri S, et al. Recurrent R-spondin fusions in colon cancer. *Nature*. 2012 Aug 30;488:660-4.
- Shike M, et al. Primary prevention of colorectal cancer. The WHO Collaborating Centre for the Prevention of Colorectal Cancer. *Bull World Health Organ*. 1990;68:377-85.
- Singh D, et al. Transforming fusions of FGFR and TACC genes in human glioblastoma. *Science*. 2012;337:1231-1235.

- Soda M, et al. Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer. *Nature*. 2007;448:561-6.
- Steensgaard P, et al. Sgt1 is required for human kinetochore assembly. *EMBO Rep*. 2004 Jun;5:626-31.
- Storm EE, et al. Targeting PTPRK-RSPO3 colon tumours promotes differentiation and loss of stem-cell function. *Nature*. 2016 Jan 7;529:97-100.
- Su LK, et al. Multiple intestinal neoplasia caused by a mutation in the murine homolog of the APC gene. *Science*. 1992;256:668-70.
- Takahashi A, et al. HSP90 interacts with RAR1 and SGT1 and is essential for RPS2-mediated disease resistance in Arabidopsis. *Proc Natl Acad Sci U S A*. 2003 Sep 30;100:11777-82.
- Tomlins SA, et al. Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science*. 2005;310:644-648.
- Torre LA, et al. Global cancer statistics, 2012. *CA Cancer J Clin*. 2015 Mar;65:87-108.
- Torres-García W, et al.. PRADA: pipeline for RNA sequencing data analysis. *Bioinformatics*. 2014 Aug 1;30:2224-6.
- Vogelstein B, et al. The genetic basis for human cancer. 2nd ed. Toronto: McGraw-Hill. 2001.
- Williams SV, et al. Oncogenic FGFR3 gene fusions in bladder cancer. *Hum Mol Genet*. 2013;22:795-803.

- Winawer SJ, et al. Prevention of colorectal cancer by colonoscopic polypectomy. *N Engl J Med.* 1993; 329: 1977-1981.
- Wong SH, et al. Genome-wide association and sequencing studies on colorectal cancer. *Semin Cancer Biol.* 2013; 23: 502-511.
- Wu WK, et al. Dysregulation and crosstalk of cellular signaling pathways in colon carcinogenesis. *Crit Rev Oncol Hematol.* 2013; 86: 251-277.
- Wu YM, et al. Identification of targetable FGFR gene fusions in diverse cancers. *Cancer Discov.* 2013;3:636-647.
- Xu Y, et al. Identification of thyroid carcinoma related genes with mRMR and shortest path approaches. *PLoS One.* 2014 Apr 9;9:e94022.
- Yu J, et al. Disruption of NCOA2 by recurrent fusion with LACTB2 in colorectal cancer. *Oncogene.* 2016 Jan 14;35:187-95.