 doi:10.20944/preprints202106.0102.v2

Article

Characterization of alternatively spliced isoforms of MUC4 and ADAM12 genes in a metastatic colorectal cancer cell line model

Saleh Althenayyan[#1,2], Mohammed H AlMuhanna[1,2] , Abdulkareem Abdulrahman[2,4], Bander AlGhanem[2,3], Sulaiman AlSaqaby[5] and Mohammad Azhar Aziz[1,2*]

[1] King Abdullah International Medical Research Center, Colorectal Cancer Research Program, Department of Cellular therapy and cancer research, Riyadh, 11481, Saudi Arabia

[2]King Saud Bin Abdulaziz University for Health Sciences, Riyadh, 11481, Saudi Arabia

[3]King Abdullah International Medical Research Center, Department of Core facilities and platforms, Riyadh, 11481, Saudi Arabia

[4] King Abdullah International Medical Research Center, Department of Medical genomics, Riyadh, 11481, Saudi Arabia

[5]Department of Medical Laboratory Sciences, College of Applied Medical Sciences, Majmaah University, Majmaah 11932, Saudi Arabia

[*] Corresponding Author:

Mohammad A. Aziz, PhD

King Abdullah International Medical Research Center (KAIMRC)

King Saud Bin Abdulaziz University for Health Sciences

King Abdulaziz Medical City

Ministry of National Guard Health Affairs

PO box 3660 Riyadh 11481, Saudi Arabia

Mail Code 1515

Tel. No.: +966 11 4294582

Email: azizmo@ngha.med.sa, maa32@case.edu

Simple Summary: Biomarker driven precision and personalized medicine is pertinent for improving therapeutic outcome for colorectal cancer especially for metastatic stage. In this study we studied two genes which were previously reported by our group to be differently expressed and spliced in human colorectal cancer samples. Gene splicing is a way to generate multiple forms of the same gene. We characterized these multiple forms known as 'isoforms' by studying their genomic organization, expression pattern as well as response to drugs. This evidence will be helpful in establishing the role of the two genes as a biomarker in colorectal cancer metastasis.

Abstract: Colorectal cancer prognosis get worse with advancement of disease into metastatic stage. There is a pertinent need to develop prognostic biomarkers that can be used for personalized and precision medicine. Alternative splicing provides an insight into understanding of changes at isoform expression level which may not be evident at gene level. In this direction, we utilized our prior knowledge about significant alternatively spliced genes and chose *ADAM12* and *MUC4* for further characterization in a metastatic cell line model. These genes were found to be good prognostic indicators in The Cancer Genome Atlas database. We studied the gene organization and designed primers to specifically amplify a group of isoforms. Differential expression of these group of isoforms was observed in normal, primary and metastatic colorectal cancer cell lines. We further validated the results using sanger sequencing. Isoform expression was found to respond to the 5-fluorouracil treatment. RNAseq analysis of the cell lines further validated the differential expression of gene isoforms. Successful detection of ADAM12 and MUC4 in cell lysates varied according to the antibody used which may reflect differential expression of isoforms. This comprehensive study underscores the importance of studying alternatively spliced isoforms and their probable used as prognostic or predictive biomarkers.

Keywords: colorectal cancer, alternative splicing, mucins, biomarkers, precision medicine

1. Introduction

Colorectal cancer (CRC) is one of the commonest malignancies in the world, including Saudi Arabia [1] with very low 5-year survival rate (10%) in advanced metastatic stages [2,3]. However, there is hope generated by the advent of precision/personalized medicine especially in colorectal cancer [4,5]. Personalized medicine has come of age and is complementing therapeutic decision making process in different clinical settings with the aid of biomarkers.

Biomarker is a statistically significant measurable indicator of a patient. Prognostic biomarker is defined as a biomarker that is used to predict the progression of a disease. Predictive biomarker can be used to predict the efficacy of a particular therapeutic regimen. The microsatellite instability status, location of tumor (right or left side) has helped in making guided decision about administering available therapies. However, there is still lack of enough biomarkers that can be used as prognostic and predictive biomarkers for colorectal cancer metastasis [5,6]. There is an unfulfilled need to develop molecular biomarkers that can help in predicting the course of disease progression after administering standard of care regimens for metastatic colorectal cancer.

Isoforms of a gene may serve as precise biomarkers. Gene isoform is defined as an alternatively spliced form of messenger RNA. Alternative splicing (AS) based biomarkers have not been studied in detail due to technical complexity but carry huge potential in complex diseases like cancer and diabetes. More than 90% genes are alternatively spliced and any disturbance in this complex machinery could result into initiation and progression of cancer [7,8]. The changes in AS mechanism resulting into altered ratio of splice variants causing increased progression and metastasis have been observed in prostate cancer [9]. In general, alternatively spliced isoforms have been less studied to understand their role in initiation and progression of cancer.

Based on prior evidence from our study on patient exon profiling [10], we studied two genes (MUC4 and ADAM12) with highest statistical significance in terms of exon level expression changes. These two genes were found to be highly significant in cytogenetic, gene expression and splicing studies on the Saudi colorectal cancer patients [10,11]. In this study, we attempt to characterize the isoforms of MUC4 and ADAM12 genes in SW480 and SW620 cell lines. We tried to amplify specific alternatively spliced isoforms and sequence them to confirm their identity. We compared the relative abundance of the gene transcripts and protein in the two cell lines which represent primary (SW480) and metastatic (SW620) cells. We also employed global gene expression profiling (RNAseq) of the two cell lines to interrogate the expression of MUC4 and ADAM12 gene isoforms. RNASeq data also provided avenues to study the expression changes in pathways and molecules associated with AS. Finally we checked the prognostic significance of the two genes in colorectal cancer using The Cancer genome Atlas (TCGA) database. This study provides comprehensive analy-

sis of MUC4 and ADAM12 gene isoforms in colorectal cancer cell lines which could be useful in their development as prognostic biomarkers.

2. Materials and Methods

Cell Culture

CCD841, SW480, and SW620 cell lines (ATCC® CCL-247TM) were obtained from American Type Culture Collection (USA). CCD841 represents normal colon mucosa whereas SW480 and SW620 cells are from primary and metastatic colorectal cancer cells, respectively. Cells were cultured in 5%CO_2 at 37°C. Cells were grown in advanced Dulbecco's modified Eagle's medium (DMEM) containing 10% FBS, 100 IU/mL of penicillin, 100 µg/mL of streptomycin and 2 mmol/L L- glutamine (Gibco, Thermo Fisher Scientific, USA).

RNA extraction and cDNA synthesis

Total RNA was extracted from CCD841, SW480 and SW620 cell lines using PureLink RNA mini kit (Cat#12183025, Thermo Fisher Scientific, USA). 2µg of RNA was reverse transcribed for cDNA synthesis using High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems, Thermo Fisher Scientific; Cat# 4374967).

Primer design

Alternatively spliced isoforms of ADAM12 and MUC4 were analysed by SnapGene software (GSL Biotech, USA). We used global alignment (Needleman-Wunsch) algorithm for aligning two sequences. For more than two sequence alignment we used Clustal Omega option in Snapgene. ADAM12 and MUC4 isoform sequences were obtained from genome database NCBI as well as ENSEMBL. Five well known ADAM12 isoforms were aligned to identify the difference among the isoforms. The possible difference between the available isoform sequence was used to design primers. Three set of primers were designed to amplify ADAM12-All isoforms, – isoform 1&3, -isoform 2,4 &5. Similar design strategy was employed to design four set of primers to detect four well described isoforms of MUC4. Both GAPDH and ACTNB genes were used as house keeping control.

Semi-quantitative endpoint PCR

Primers designed as described above were used to amplify the targets of alternatively spliced isoforms of ADAM12 and MUC4. We used the Master Mix DreamTaq™ Hot Start Green (Cat# K9022, Thermo Fisher Scientific, USA).

PCR amplification was carried out using thermal cycler Veriti™ 96-Well (Applied Biosystems, Thermo Fisher Scientific; Cat# 4375786,). PCR cycling conditions were: 95ºC for 5 minutes initially, followed by 40 repeats of 95ºC for 1 minutes, 60ºC for 30 seconds and 72ºC for 30 seconds. Then one cycle of 72 °C for 7 minutes, and lastly 4°C indefinitely. All PCR products were analysed using 2% agarose on horizontal gel electrophoresis. 50bp ladder was used to estimate the band size.

Gel extraction and purification of amplified PCR products

We excised the gel slices that contain the DNA fragment using Thermo Scientific GeneJET Gel Extraction Kit, (cat#K0692). The gel slices were then placed into pre-weighed 1.5 mL tubes to weigh to the gel slice. Then we add 2:1 volume of the binding buffer because, we used 2% agarose. The following step was to incubate the eppendorf tubes in 50-60 °C for 10 min or until the gel slice is completely dissolved. We used 1 gel volume of 100% isopropanol because our bands sizes are ≤500 bp. Then we followed the steps of the protocol using GeneJET purification column to extract the DNA fragments. We checked the purity and quantity using nanodrop (Thermofisher scientific)

Sanger Sequencing

The PCR products using different set of primers were sequenced using DNA Analyzer (Applied Biosystems™; Cat# 3730XL). For clean-up of endpoint PCR product, ExoSAP-IT™ kit was used (Thermo Fisher Scientific; Cat# 78200.200.UL). Cycling sequencing was performed with BigDye™ Terminator v3.1 Kit (Thermo Fisher Scientific; Cat#4337455). Clean up for the cycling sequencing was with BigDye XTerminator™ Purification Kit (Thermo Fisher Scientific; Cat# 4376486).

Quantitive Real Time PCR

Quantitaive real time PCR (qRT-PCR) was performed by using both SYBR green and TaqMan universal PCR master mix (Applied Biosystems; Cat# 4304437) and the QuantStudio 6 Flex Real-Time PCR System (Thermo Fisher Scientific, Inc) according to the manufacturer's instructions. The gene expression was

analyzed using TaqMan probe for ADAM12 and MUC4 genes (Thermo Fisher Scientific, USA). Eukaryotic 18S rRNA was used as endogenous control gene. All reactions were performed in triplicate and the qRT-PCR data was analyzed by using the relative quantitative (RQ) method ($2^{-\Delta\Delta Ct}$) through Expression Suite software version 1.1 (Thermo Fisher Scientific, USA). For SYBR green PCR reactions same primers as used in endpoint PCR were used and SYBR™ Green PCR Master Mix; Cat# 4309155, from Applied Biosystems™. All experiments were done in triplicate and repeated at least twice.

RNA Sequencing

Total RNA was extracted from the cell line SW480 and SW620 using the kit (Ambion, PureLink RNA mini; Cat# 12183025). RNA quality and concentration was evaluated using ND-1000 UV-Vis Spectrophotometer; highly qualified samples were selected for RNA seq.

RNA sequencing was performed according to manufacturer's protocol as follows, poly (A) messenger RNA was captured using RiboMinus™ Eukaryote System v2 (Cat# 4481370), and the purified samples were cleaned from any residuals by RiboMinus™ Magnetic Bead Cleanup Module (Cat# 4481370). RNA was then measured by (Qubit™ RNA HS Assay kit; Cat# 032852).

Library was then constructed by Ion Total RNA-Seq Kit v2 (Cat# 4475936). This library was then converted to cDNA using Ion Total RNA-Seq Primer Set v2 (Cat# 4474810). After that cDNA was indexed by Ion Express™ RNA-Seq Barcode 01-16 Kit (Cat# 4475485). Next, we measured cDNA using NanoDrop1000.

High qualified library was then injected into the Ion PI™ Hi-Q™ Chip Kit v3 (Cat# A26770) and then inserted in the Ion Chef™ Instrument (4484177) to start emulsion and enrichment for the library. Next, sample sequencing was performed using the Proton Semiconductor Sequencer.

Western Blot

Whole cell extract was prepared from each cell line CCD841, SW480, and SW620 with NP40 buffer (Invitrogen, Paisley, UK) containing protease inhibitor cocktail (Sigma-Aldrich, Taufkirchen, Germany). The protein lysate was incubated on ice for 30 minutes, centrifuged at 13,000 rpm for 10 minutes and the supernatant was collected. Protein concentrations were determined by Qubit protein assay kit (Thermo Fisher Scientific). 50µg of whole cell lysate was mixed with 4X Laemmli buffer (Bio-Rad) and loaded onto

ready-made gel (4–20% Mini-PROTEAN® TGX™ Precast Protein Gels ;Bio-Rad). The gel was subjected to electrophoresis at 100V for approximately 90 minutes and transferred overnight at 5V to polyvinyl PVDF membranes using semidry transfer cell (Bio-Rad). The membranes were blocked using 5% Bovine Serum Albumin (MILLIPORE, Germany) in TBST (Tris Buffer Saline Tween) for 1 hour with shaking at room temperature followed by the incubation of the membrane with 1:1000 of respective antibodies raised in rabbit (MUC4 : Cat# orb399150 & Cat# orb306041, Biorbyt; ADAM12 :Cat# orb155592 & Cat# orb373867; GAPDH as loading control (Cat# orb234217) in 5% BSA with gentle shaking overnight (O/N) at 4°C. GAPDH was diluted at a ratio of 1:10000. Secondary antibodies were goat anti-rabbit (Cat# orb43514) for MUC4 and ADAM12, and goat anti-mouse (Cat# orb500708) for GAPDH. The secondary antibodies were diluted at a ratio of 1:10000, and incubated at room temperature for 1 hour. The signals were detected with Chemiluminiscent HRP Substrate (Bio-Rad). Images were captured and analysed using Chemidoc gel documentation system (Bio- Rad).

Statistical analysis

Prism software (GraphPad) and Microsoft Excel were used to analyze data and calculate *p* values determined by statistical tests indicated within the figure legends. Differences between groups were considered significant for p values ≤ 0.05.

Survival analysis

Two transcriptomics data sets that were generated from colorectal adenocarcinoma patients and published by The Cancer Genomic Atlas (TCGA) research network[12] were used. The TCGA term of the data sets was as follows: “colorectal adenocarcinoma TCGA Firehose Legacy” and “colorectal adenocarcinoma TCGA PanCancer Atlas”. The prognostic importance of *ADAM12* and *MUC4* was examined using these two data sets through **“**cBioPortal” coupled with Onco Query Language (OQL) tools [13]. The selected genomic profile for the analysis in both data sets was: mRNA expression (RNA sequencing) z-scores relative to all samples (log RNA Seq V2 RSEM).

3. Results

Expression of *ADAM12* and *MUC4* in CRC patients

We had previously studied exon level expression pattern of genes significantly altered in tumor samples of our patient cohort [10]. From this study, we further analyzed the expression pattern of *ADAM12* and *MUC4* which were among the genes with highest significance in terms of AS in CRC. We analyzed the expression level of the two genes in patient tumor and normal samples. ADAM12 was upregulated while MUC4 gene was downregulated significantly in tumor samples (Figure 1 A&B).

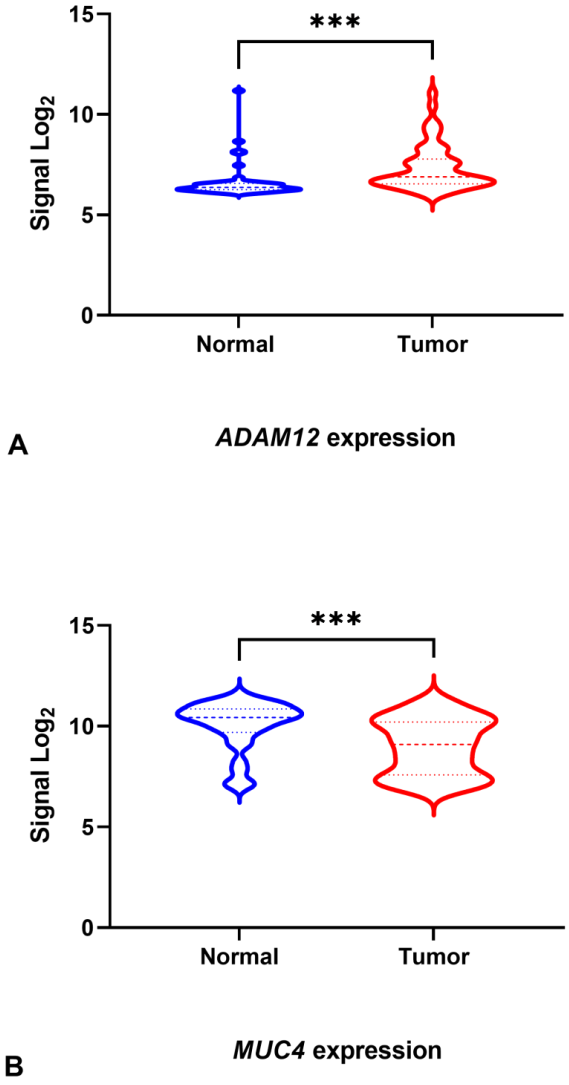


Figure 1: Analysis of microarray data from tumour and normal CRC patient samples

A: Expression patter of *ADAM12* as analysed from a previous microarray study. *ADAM12* was found to be significantly higher in tumor samples. **B:** Expression patter of *MUC4* gene as analysed from a previous microarray study. *MUC4* was found to be significantly lower in tumour samples. Graphs represent data from 48 patients. Welch's t-test was used to determine statistical significance.

Potential of *ADAM12* and *MUC4* as biomarkers in CRC

The prognostic potential of *ADAM12* and *MUC4* was assessed using two transcriptomics data sets from TCGA. *ADAM12* significantly predicted OS in patients with colorectal adenocarcinoma. In the data sets termed "colorectal adenocarcinoma TCGA Firehose Legacy", patients with high expression of *ADAM12* (z-score>0.2) had short OS (median OS = 63 months), whereas those with low expression (z-score<0.2) showed long survival (median OS = undefined; Figure 2A; p= 0.006; HR= 1.84). Similarly, in the data set "colorectal adenocarcinoma TCGA PanCancer Atlas", *ADAM12* appeared to predict the disease-specific survival (DSS); on the month number 75 the proportion of DSS event was 16% in high expression group (z-score>0.3) compared to 11% in the low expression group (z-score<0.03; Figure 2B, p= 0.05, HR= 1.61). In contrast to *ADAM12*, *MUC4* was found to predict good prognosis of colorectal adenocarcinoma. Increased expression of *MUC4* (z-score>-0.83) indicated long survival (median OS= 100 months), whereas decreased expression (z-score<-0.83) identified patients with shorter survival (median OS = 54 months; Figure 2C, p<0.05, HR= 0.59). Supporting this finding, the analysis conducted on data set termed "colorectal adenocarcinoma TCGA PanCancer Atlas" found *MUC4* to be informative of DSS. Figure 2D shows that the proportion of DSS event post month number 75 was 8% on the high expression group (z-score>0.62) compared with 15% in low expression group (z-score<0.62; p= 0.02, HR= 0.59).

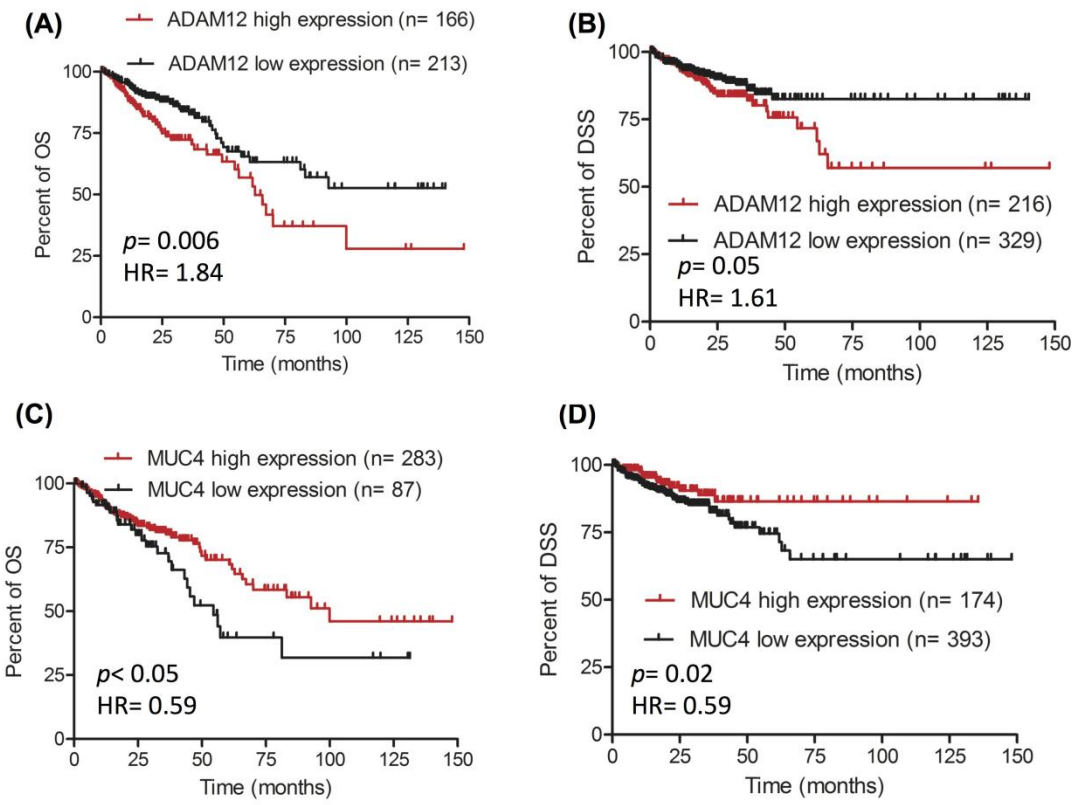


Figure 2: Survival curves depicting prognostic potential of ADAM12 and MUC4 The transcript expression of *ADAM12* and *MUC4* predict the prognosis of colorectal carcinoma. Two TCGA transcriptomics data sets that were generated from patients with colorectal adenocarcinoma were used to evaluate the prognostic potential of *ADAM12* and *MUC4*. Increased expression of *ADAM12* predicted short OS (A; data set: colorectal adenocarcinoma TCGA Firehose Legacy) and short DSS (B; data set: colorectal adenocarcinoma TCGA PanCancer Atlas). Decreased expression of *MUC4* predicted short OS (C; data set: colorectal

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 June 2021 doi:10.20944/preprints202106.0102.v2

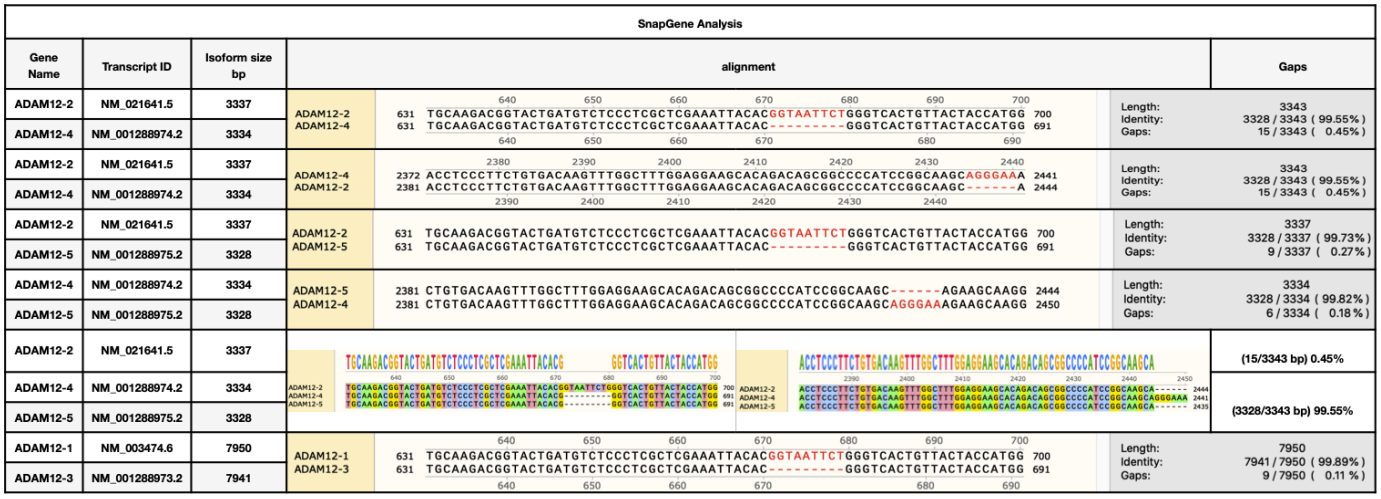
adenocarcinoma TCGA Firehose Legacy) and short DSS (D data set: colorectal adenocarcinoma TCGA PanCancer Atlas). OS: overall survival; DSS: disease-specific survival.

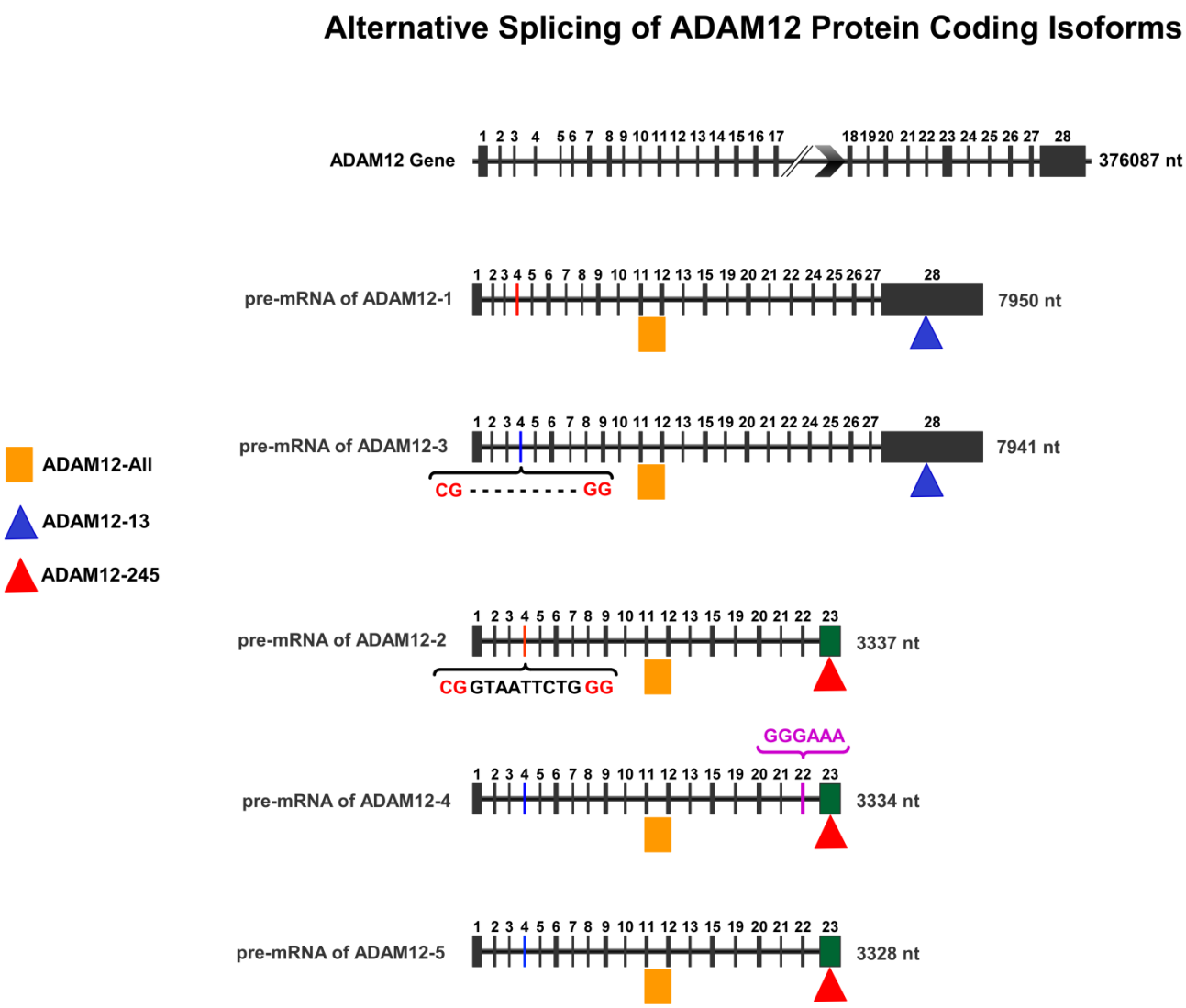
In silico identification of *ADAM12* and *MUC4* isoforms

We catalogued the reported isoforms of ADAM12 and MUC4 genes reported in public databases. We used the sequence information to design primers for amplification of isoforms as discussed below.

ADAM12 isoforms

In order to characterize the presence of *ADAM12* isoforms in the cell lines, we consulted the annotations from NCBI database. There are five listed isoforms of *ADAM12*. Table 1 provides details of the transcript along with annotation from ENSEMBL database and pimer sequence used to amplify group of transcripts. Primer designed for amplifying all the five protein coding isoforms of ADAM12, is annotated as (ADAM12-All). We considered isoforms 2,4&5 (ADAM12-245) as one isoform for primer design, due to their sequence similarity that exceeds 99%. Similar approach was employed for isoforms 1&3 (ADAM12-13), see Figure 3A. For ADAM12-All, ADAM12-245, and ADAM12-13, we designed three sets of primers. The location of primers along with exon organization for ADAM12 is illustrated in Figure 3B. The isoforms 1 and 3 show skipping of exon 23. ADAM12-1&2 have a sequence of 9 nucleotides (GTAATTCTG) that does not exist in ADAM12-3, 4, & 5. ADAM12-2,4 & 5 are unique with the presence of exon 23. ADAM12-4 has a stretch of 6 base pairs that are not unique among the isoforms but unique at this location when all the isoforms are aligned.





246

Figure 3: Sequence similarity analyses for *ADAM12* 247

A: Sequence similarity of *ADAM12* isoforms. **B:** Illustration of exon organization of ADAM12 that was used for designing primers 248
for specific amplification of alternatively spliced isoforms. There are five alternatively spliced isoforms of ADAM12. The gene 249
ADAM12 and the isoforms are presented here as horizontal lines. Each vertical bar represent the exon separated by intronic regions. 250
The isoforms depicts the phenomenon of alternative splicing via skipping of exon 14 in all the isoforms, and pretermination in the 251
isoforms 2,4, and 5. 252

MUC4 Isoforms

We used the annotation of NCBI database to identify MUC4 isoforms.. Table 2 provides details of the transcripts along with annotation from ENSEMBL database. According to NCBI database, MUC4 isoform 1 size is 16756 bp, MUC4 isoform 4 size is 4048 bp, MUC4 isoform 5 size is 3895 bp, and MUC4 isoform 6 size is 22824 bp.

We designed primers that amplify the four MUC4 protein coding isoforms and we called them MUC4-All. There was high similarity (99%) between MUC4 isoform 1 (MUC4-1) and MUC4 isoform 6 (MUC4-6) without their corresponding tandem repeat sequences. Therefore, we designed primers that amplify these two, and we call them MUC4-16 . There were two gap regions and 7 mismatches (only 2 are represented in the figure) among the two sequences (Figure 4A). Skipping of exon 2 and 25 is common in all the isoforms. Additionally, MUC4-4 shows skipping of exon 2&3, and MUC4-5 shows skipping of exons 2,3&4. The location of primers along with exon structure is illustrated in Figure 4B. AS at exon junction creates unique isoforms. We employed junction at exon 1&5 to design primers for MUC4-4.

Gene Name	Transcript ID	Isoform size bp	alignment	Gaps
MUC4-6	NM_001322468.1	22824	MUC4-6 1 CTCTTTTGTCCTCTTCCCAGGTTCCCTGGCCCCTTCGGAGAAACGCACTTGGTTCGGGCCAGCCGCCTGAGGGGACGGGCTCACGTCTGCTCCTCACACTGCAGCTGCTG 110	Length: 6868 Identity: 6751 / 6868 (98.30%) Gaps: 109 / 6868 (1.59%)
MUC4-1	NM_018406.7	16756	MUC4-1 1 --GCACTTGGTTCGGGCCAGCCGCCTGAGGGGACGGGCTCACGTCTGCTCCTCACACTGCAGCTGCTG 66	
MUC4-6	NM_001322468.1	22824	MUC4-6 441 GATGATGACATCAACTCTTTTTTCTTCCCCAAGTGTACACAATGTGATGGAGACTGTTACGCAGGAGACAGCTCCTCCAGATGAAATGACCACATCATTTCCCTCCAGTG 550	Length: 6868 Identity: 6751 / 6868 (98.30%) Gaps: 109 / 6868 (1.59%)
MUC4-1	NM_018406.7	16756	MUC4-1 397 GATGATGACATCAACTCTTTTTTCTTCCCCAAGTGTACACAATGTGAT---------------GGAGACAGCTCCTCCAGATGAAATGACCACATCATTTCCCTCCAGTG 491	
MUC4-6	NM_001322468.1	22824	MUC4-6 991 CTTGGAAACCCAGGGGAGACATCATCAGTACCTGTTACTGGAAGTCTTATGCCAGTCACCTCAGCAGCCTTAGTAACAGTTGATCCAGAAGGACAATCACCAGTAACTTT 1100	Length: 6868 Identity: 6751 / 6868 (98.30%) Gaps: 109 / 6868 (1.59%)
MUC4-1	NM_018406.7	16756	MUC4-1 932 CTTGGAAACCCAGGGGAGACATCATCAGTACCTGTTACTGGAAGTCTTATGCCAGTCACCTCAGCAGCCTTAGTAACATTTGATCCAGAAGGACAATCACCAGCAACTTT 1041	

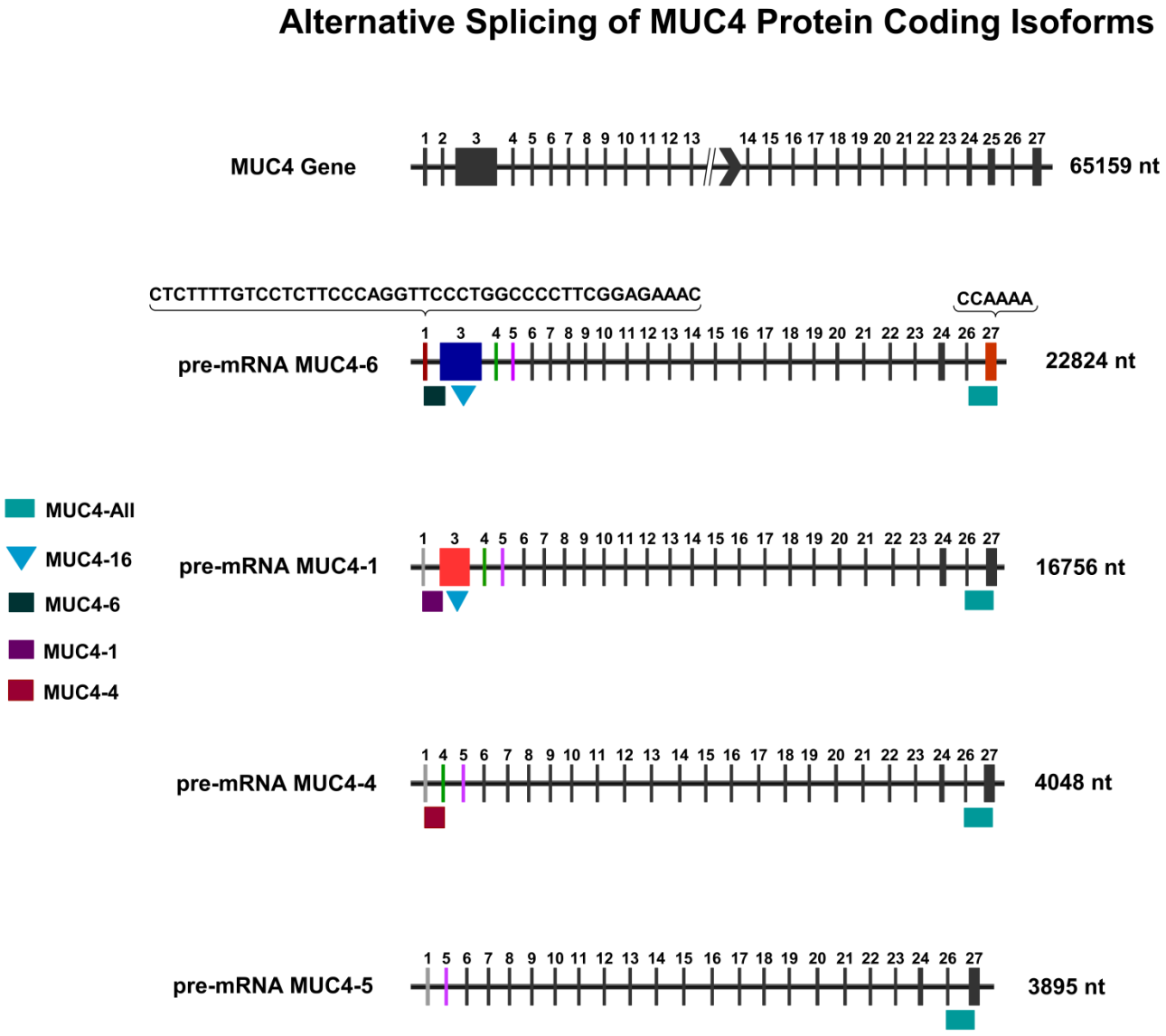


Figure 4: Sequence similarity analyses for *MUC4*

A: Sequence similarity of *MUC4* isoforms. B: Illustration of exon organization of MUC4 that was used for designing primers for specific amplification of isoforms Alternative splicing of MUC4 protein coding isoforms. The gene MUC4 and the isoforms are presented here as horizontal lines. Each vertical bar represent the exon separated by intronic regions. With MUC4-6, we used the first 44 bp that is unique to it in exon 1, with the slight difference in exon 3 as basis for the primer design.

Amplification of *ADAM12* and *MUC4* isoforms

Endpoint PCR results show specific and expected bands sizes for the primers ADAM12-All, ADAM12-245, and ADAM12-13 (Figure 5). ADAM12-All primers show lowest expression levels among the three set of primers we studied. ADAM12-All shows differential expression among the three cell lines. CCD841 shows highest expression level, followed by SW480, and SW620 exhibits least expression (Figure 5A). ADAM12-245 as well as ADAM12-13, do not show any apparent change in expression level among the three cell lines (Figure 5B&C).

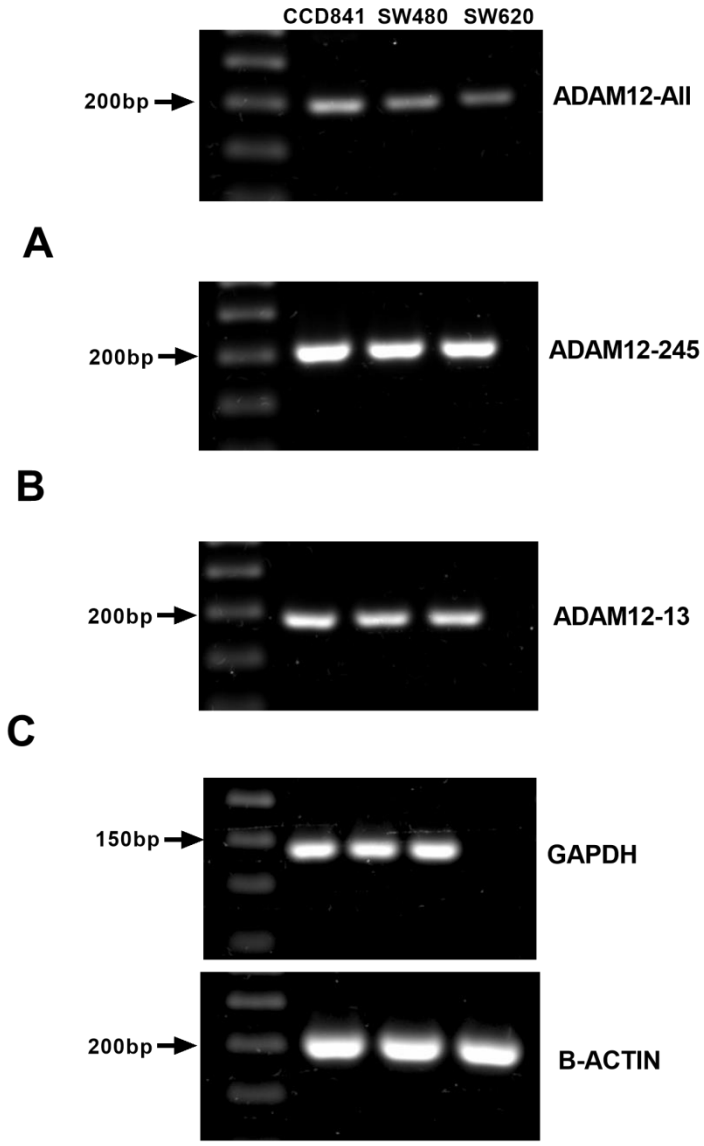


Figure 5: PCR Amplification of specific isoforms of ADAM12

Agarose gel analyses of PCR amplified products using specific set of primers. There was a difference in expression levels when ADAM12-All primers were used **(A)** whereas ADAM12-245 **(B)** and ADAM12-13 **(C)** showed similar expression levels in all three cell lines. Both *GAPDH* and *ACTB* were used as housekeeping endogenous controls.

Further, we confirmed the sequence of the amplicons of ADAM12-All, ADAM12-245, and ADAM12-13 from the cell lines CCD841, SW480 and SW620. Results from Sanger sequencing show the NCBI blasting alignment score of confidence between 99 and 100% (Supplementary Figure 1).

Semi-quantitative estimation from endpoint PCR with the primers for MUC4-All show that CCD841 is least expressed but similar expression pattern was observed in SW480 and SW620 (Figure 6A). Primers for MUC4-1&6, MUC4-4 were able to amplify specifically and show single bands corresponding to expected amplicon size (Figure 6 B&C).

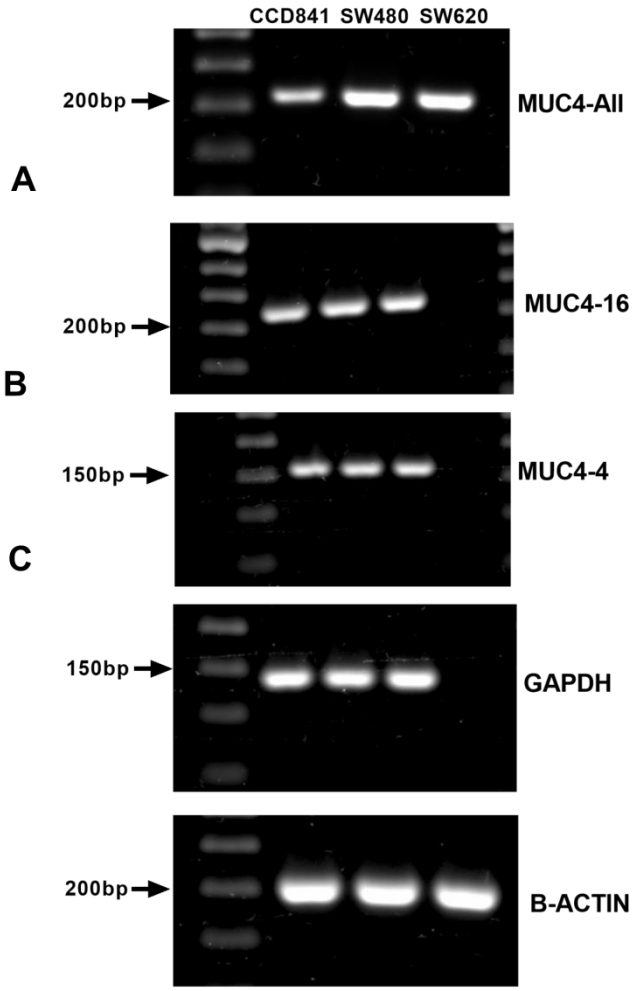


Figure 6: PCR Amplification of specific isoforms of MUC4

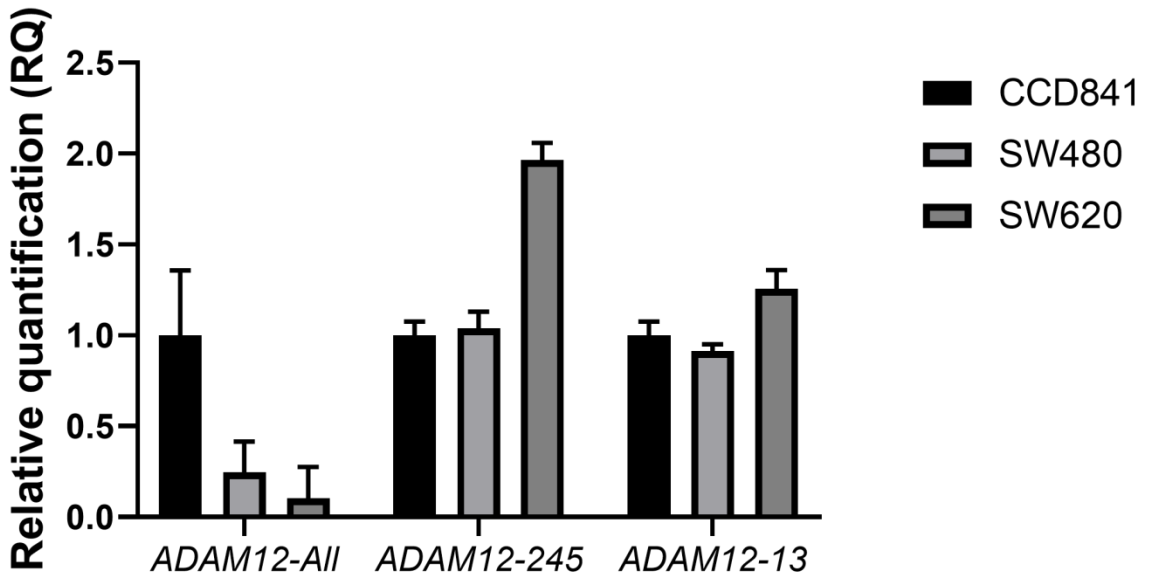
Agarose gel analyses of PCR amplified products using specific set of primers. There was a difference in expression levels when ADAM12-All primers were used **(A)** whereas MUC4-16 **(B)** and MUC4-4 **(C)** showed similar expression levels in all 3 cell lines. Both *GAPDH* and *ACTB* were used as housekeeping endogenous controls.

We were unable to detect specific amplicons for MUC4-6 and MUC4-1. All primers sets used for MUC4-6 and MUC4-1 resulted in different amplicons showed as multiple bands (supplementary Figure 2). Amplicons from these primers were subjected to Sanger sequencing . We then aligned the sequences of the amplicons from MUC4-All, MUC4-16 and MUC4-4 with the reference isoform using SnapGene, and blast the sequences via NCBI. The scores of confidence were between 95 and 100% (Supplementary Figure 3). Amplification of isoforms MUC4-1 and MUC4-6 resulted in multiple bands. We cut the two specific bands of expected size for isoform 1&6 along with two other non-specific bands to sequence them. The Sanger Sequencing results of these four bands show that they correspond to MUC4-1 and MUC4-6 with the same scoring confidence across all the three cell lines CCD841, SW480, and SW620. The scoring confidence is between 98 and 93%, apart from one band (320 kb; unexpected band size) that was amplified by MUC4-6 primers in SW480 cells (Supplementary Figure 4).

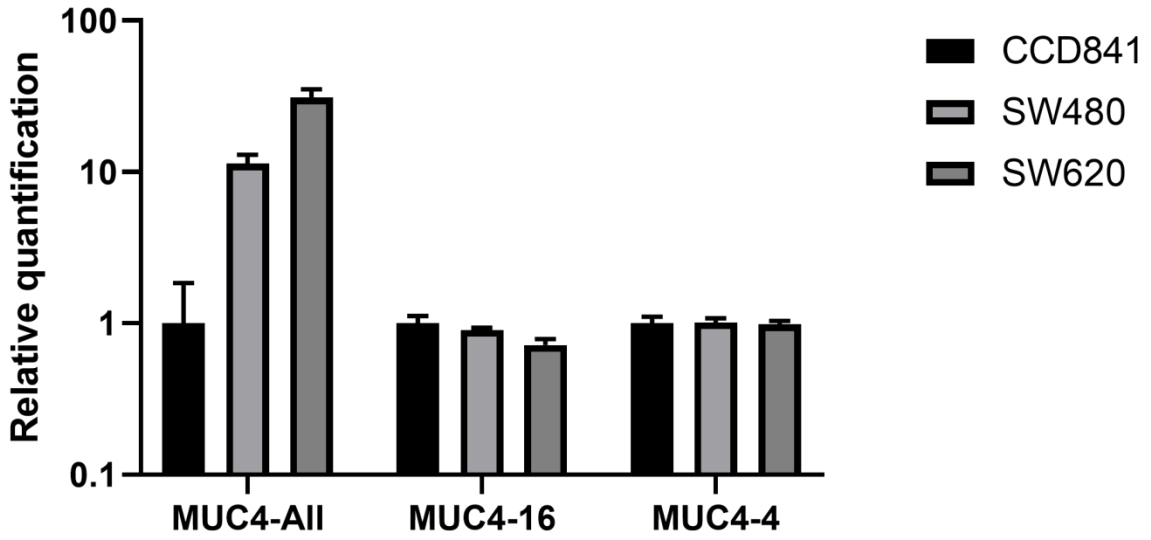
Quantitative assessment of expression of *ADAM12* and *MUC4* isoforms

Expression of ADAM12 as measured by ADAM12-All primers shows downregulation of ADAM12 in SW480 (RQ=0.25), and SW620 expression was approximately half of SW480 with RQ=0.1. Isoforms detected by primers ADAM12-245 show similar expression between CCD841 and SW480 whereas SW620 shows almost double the expression with RQ=2. ADAM12-13 primers exhibited similar expression pattern (Figure 7A).

MUC-4-All primers detected very high expression increase of MUC4 in SW480 cells. Isoforms detected by MUC4-5 primers also exhibited similar pattern. However, MUC4-16 primers detected increasing levels of transcripts in CCD841<SW480<SW620. MUC4-4 primers also resulted in similar expression pattern as MUC4-16, but with higher differential expression values between the three cell lines (Figure 7B).



A



B

Figure 7: Quantification of ADAM12 and MUC4 isoforms in cell lines

Quantitative RT-PCR analyses suggest differential expression of isoforms in normal colon (CCD841) primary (SW480) and metastatic (SW620) colorectal cancer. The pattern of expression is different among the isoforms as well. **A:** ADAM12-All primers showed a decreasing trend in CCD841>SW480>SW620 whereas ADAM12-245 exhibits CCD841=SW480>SW620 pattern. Similar trend was observed with ADAM12-13 primers. B: MUC4-All shows an increasing pattern from normal to metastatic cells whereas MUC4-16 and MUC4-4 primers had no difference in expression levels of amplified isoforms.

Effect of 5-Fluorouracil treatment on expression of ADAM12 and MUC4 isoforms

We studied the effect of 5-fluorouracil on the expression of the isoforms. ADAM12-All primers detected isoforms with maximum expression in SW620 cells treated with 100uM drug for 24h while the least expression was observed in the same cells treated with 10uM in same time frame (Figure 8A). With ADAM12-245 primers SW480 cells treated with 5-FU for 96h showed highest expression while treatment of same cells with 1uM 5-FU for 24h caused highest downregulation (Figure 8B).

For MUC4-All, highest expression was in response to treatment of SW620 cells with 5-FU (100uM) for 24h whereas lowest expression was observed in SW480 cells treated with 100uM 5-FU for 24h (Figure 8C). Isoforms detected by MUC4-4 were found to be at maximum level in SW480 cells treated with 10uM 5-FU for 96h while lowest levels were observed in SW620 cells treated with 1uM 5-FU for 72h (Figure 8D).

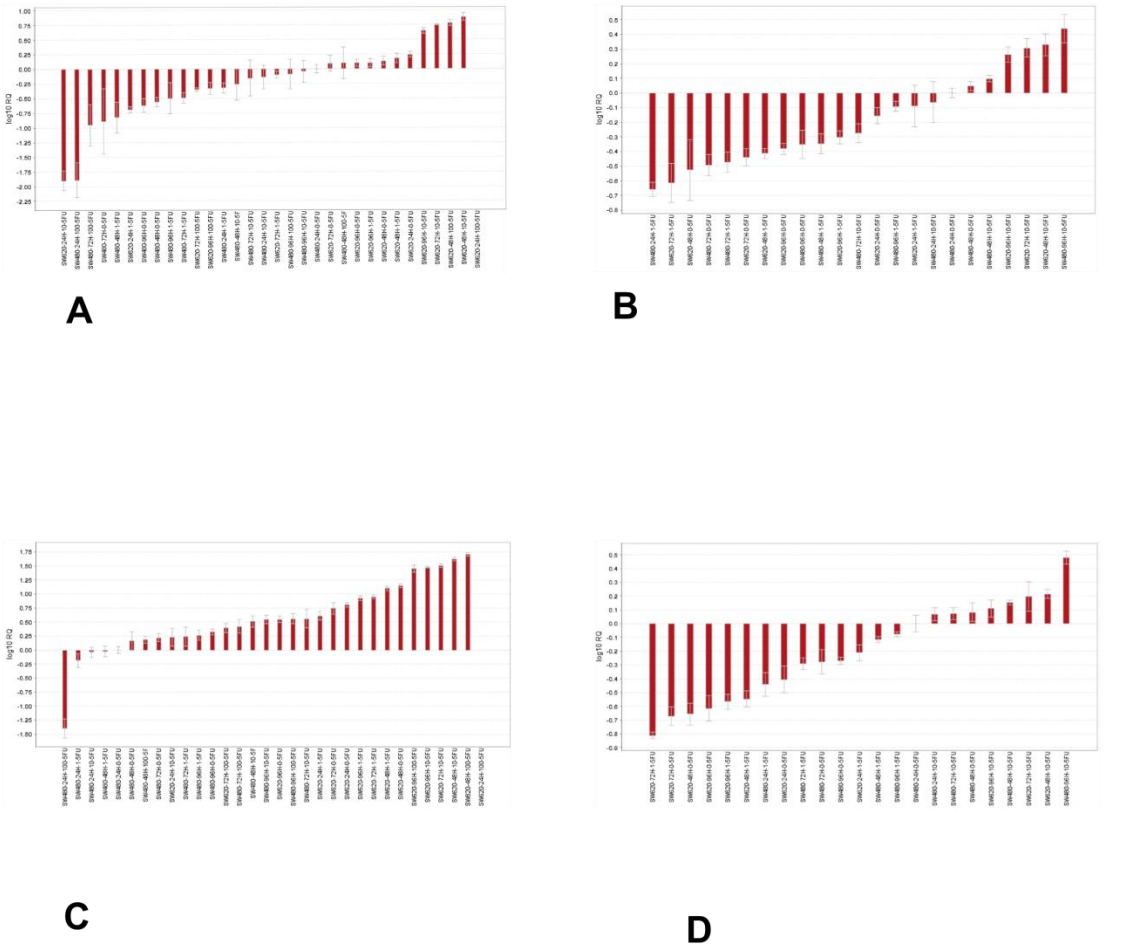


Figure 8: Change in expression level of ADAM12 and MUC4 in response to 5-FU treatments

Expression level of *ADAM12* and *MUC4* isoforms was studied in response to different levels of 5-FU for duration up to 96h. Waterfall plots were generated for 0,1,10 and 100uM of 5-FU treatments for 24,48,72 and 96h in SW480 and SW620 cells. Expression pattern of isoforms detected by **A:** ADAM12-All, **B:** ADAM12-245, **C:**MUC4-All, **D:** MUC4-4 set of primers.

Isoform analysis using RNA sequencing

We used RNA sequencing (RNA-Seq) to quantify the expression of the five isoforms of ADAM12 and the four isoforms of MUC4, according to NCBI RefSeq database. However, the results from the RNA Seq were annotated according to Ensembl GRCh37. RNA-Seq results show expression level for each isoform by Fragments Per Kilobase of transcript per Million mapped reads (FPKM). Table 3 shows the expression values for ADAM12 and MUC4 isoforms.

In order to understand the possible role of splicing genes in colorectal cancer, we analysed the RNA-Seq data. Among the six well known genes associated with AS, FUBP1 was found to be upregulated with >1.5 fold change in SW620 cells whereas U2AF1L4 gene was significantly downregulated (fold change 0.47) (Table 4).

Detection of ADAM12 and MUC4 isoforms at protein level

Probable protein isoforms of ADAM12 and MUC4 that were reported in Uniprot were used as reference for western bot results (Table 5) . We attempted to detect ADAM12 protein from the whole cell extract of three cell lines. Using antibody (cat# orb373867), we identified a ~100KDa isoform of ADAM 12 which matched the expected size of Isoform 1 from Uniprot database. The protein levels were detected at highest level in SW480 followed by SW620 and least was in CCD841 (Figure 9A). Another antibody (cat# orb155592) that could detect any of the remaining three isoforms showed similar expression level of ADAM12 in SW480 and SW620 cells but no protein was found in CCD841 cells (Figure 9B).

Two different MUC4 antibodies detected isoform 1 with an expected size of 230KDa. One of these antibodies (Cat# orb306041) could not detect any band in CCD841 (Figure 9C) whereas other one (Cat# orb399150) did react with a protein band of ~232KDa (Figure 9D).

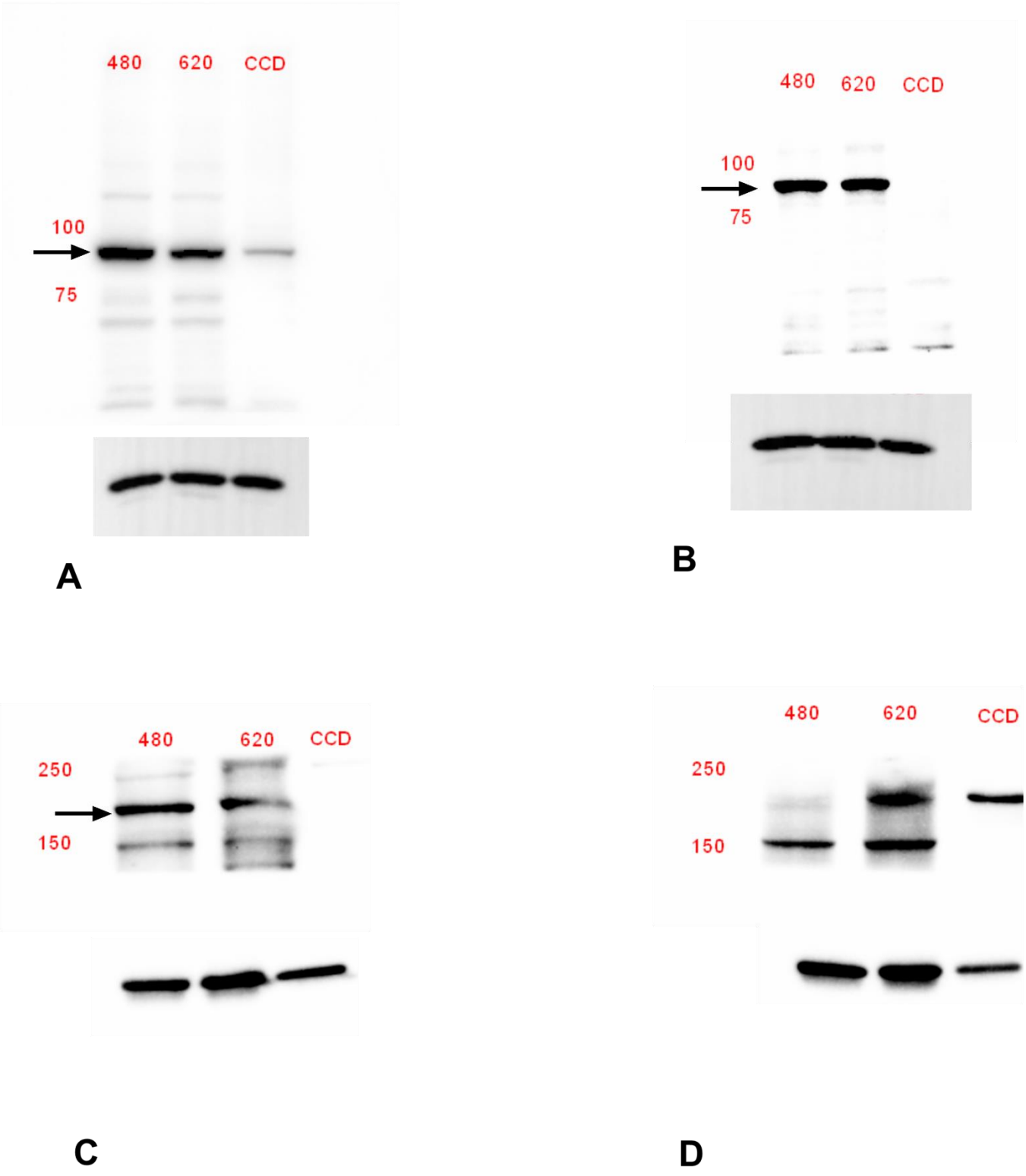


Figure 9: Western blot analyses of ADAM12 and MUC4 isoforms in cell lines

Immunoblot analyses of ADAM12 resulted in successful detection of ADAM12 using two different antibodies with an expected size of ~100KDa (Figure 9A&B). MUC4 was not detected by one of the antibodies in CCD481 cells (Figure 9C) but using another antibody showed a band of ~232KDa (Figure 9D).

We attempted to identify secreted isoforms but could not detect either ADAM12 or MUC4 in conditioned media. All Protein blots were repeated and the original blots are given as supplementary figure 5.

3. Discussion

From our previous study on colorectal cancer patients in Saudi Arabia, we had identified novel genes that were differentially expressed as well as alternatively spliced in a highly significant manner. In the present study, we further characterized the isoforms of these two genes in an unstudied cell line model.

AS is a complex mechanism of generating multiple functional isoforms from the same gene. This allows the genome to generate a diversified transcriptome as well as proteome[14]. Almost all genes with more than one exons undergo AS [15,16]. The changes in alternative splicing mechanism resulting into altered ratio of splice variants causing increased progression and metastasis has been reported [17]. Hence, alternative splicing has been suggested as another hall mark of cancer [18]. A recent review has nicely pointed out the therapeutic potential unleashed by the study on alternative splicing in context of cancer [19]. Alternative splicing allows certain exons to be included or excluded from the final mature mRNA resulting into different protein isoforms that generates functional diversity. It has been demonstrated that approximately 95% of human genes with multiple exons undergo alternative splicing during pre-mRNA maturation [20]. The association of different splice isoforms with particular stage of disease has also been evidenced. CXCL12 and IG20/MADD are good examples of genes whose isoform expression patterns change with disease progression [21]. A comprehensive study of AS events in colorectal cancer suggests its role in determining prognosis [22] which has been provided in our study as a confirmation.

Apart from understanding the functional role of these alternatively spliced isoforms in causing disease, there is an important aspect of their use as biomarkers. Specific splice variants have been reported for several types of cancer [23-25]. For CRC, 5 genes were reported to be alternatively spliced [26]. Recent study suggests the use of AS in characterizing left and right sided CRC which is an established prognostic indicator [27].

ADAM12 has been reported to be involved in a variety of diseases specially in cancer. ADAM12 is from family of disintegrin-containing metalloproteases. It has been shown to be very diverse in function in both normal and pathological states like remodeling of cell surface, shedding of ectodomain, regulation of availability of growth factors and facilitating interactions between cell and matrix [28] [29,30].

Several studies suggest that ADAM 12 is a key player in remodeling of extracellular matrix, an important sign of neoplastic disease.

Mucins are well known to be involved in CRC initiation and progression. Family of mucins was found to be downregulated in the same patient cohort that we report in this study [31]. *MUC4* has [29]been reported to be alternatively spliced to generate secreted as well as membrane associated proteins. The genomic organization of *MUC4* is not well understood especially with a context. RT-PCR based studies analysed *MUC4* exon organization and expression in different types of tissues [32]. This result support our study where we take it further and study its expression in different cell lines using qRT-PCR and RNASeq studies. We detected the membrane bound protein but could not detect it in secreted form. There has been no prior study about detecting MUC4 in cell lines. More sophisticated techniques are needed to detect this heavily glycosylated protein. Our results showing exon skipping events are supported by prior studies where *MUC4* was one of the 6 genes that showed cancer tissue specific differential exon skipping [33]. The product amplicon of MUC4-16 is (223bp). This amplicon has one nucleotide difference in the sequence between MUC4-1 and MUC4-6. MUC4-1 has a C nucleotide, and MUC4-6 has an Adenine (A) nucleotide instead. We were able to detect the MUC4-6 (with 'A' nucleotide) using Sanger sequencing. There is no annotation for MUC4-6 in ENSEMBL. However, MUC4-1 is annotated by ENSEMBL, and our results from the RNA sequencing depend on ENSEMBL and show MUC4-1 expression. The expression of MUC4-1 in comparison to other isoforms is low, and this may explain why we were not able to detect MUC4-1 by Sanger sequencing. It might be that the expression of MUC4-6 is higher than the expression of MUC4-1, and so after the amplification, Sanger sequencing was only able to detect MUC4-6. MUC4 isoform 5 (MUC4-5) is quite interesting; its uniqueness comes from a C nucleotide. Also what makes MUC4-5 different from MUC4-6 is a G nucleotide in MUC4-5, whereas MUC4-6 has an A. Other than these two differences, the whole sequence of MUC4-5 matches to all the four isoforms. These two differences are not part of the amplicon of the primers of MUC4-5. Therefore, MUC4-5 primers are amplifying all the four isoforms, however, they do not show the same expression profile as MUC4-All.

This study provides evidence with higher probability of establishing the role of alternatively spliced isoforms as biomarkers. Comprehensive analyses of gene organization advances the field of largely un-

studied mechanism of AS. With the help of prior evidence, we selected genes that were most probable candidates to serve as biomarkers for CRC. From this study we can conclude that *ADAM12* and *MUC4* isoforms can be potential candidates for CRC metastasis. However, more work is needed to detect ADAM12 and MUC4 in secreted form. Further design of primers that can exclusively amplify a single isoform would be undertaken to study their role as prognostic biomarkers.

Limitations: It was not possible to correlate the gene isoforms with identified protein.

5. Conclusions

This study provides characterization of two genes –*ADAM12* and *MUC4* in terms of their isoforms. We found that the expression pattern of isoforms is different from the gene expression in different cell lines. The isoform expression pattern changed in response to drug treatments as well. Detection of these isoforms at protein level was observed with variation among the cell lines and antibodies used. This evidence supports the notion of using alternatively spliced isoforms at prognostic and predictive biomarkers

Supplementary Materials: Figure S1: title, Table S1: title, Video S1: title.

Author Contributions: Conceptualization, MAA; methodology, SA, MHA, AA, BA and SS .; software, SA and MAA; validation, SA, MHA, AA; formal analysis, SA, BA, SS and MAA.; resources, MAA ; data curation, MAA; writing—original draft preparation, SA, SS and MAA; writing—review and editing, All authors; visualization, MAA; supervision, MAA; project administration, MAA; funding acquisition, MAA. All authors have read and agreed to the published version of the manuscript.

Funding: Please add: This research was funded by King Abdullah International Medical Research Center, grant number RC18/185" and "The APC was funded by King Abdullah International Medical Research Center".

Institutional Review Board Statement The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Institutional Review Board of King Abdullah International Medical Research Center (protocol code RC18/185 initially approved on 15 Oct2018)."

Acknowledgments: We would like to acknowledge the support received from KAIMRC in the form of research grant RC18/185 awarded to MAA. We acknowledge the help of Dr. Ensaf Hujaily in initial designing of the primers. Help and support of Dr. Saleh AlGhamdi in the submission of the project and availability of department resources is much appreciated. The help and support of Dr. Yazeid AlHaidan, Dr. Umair, Mohammed AlIdrees and Hamad from department of medical genomics, KAIMRC is duly acknowledged.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Alsanea, N.; Abduljabbar, A.S.; Alhomoud, S.; Ashari, L.H.; Hibbert, D.; Bazarbashi, S. Colorectal cancer in Saudi Arabia: incidence, survival, demographics and implications for national policies. *Annals of Saudi medicine* **2015**, *35*, 196-202, doi:10.5144/0256-4947.2015.196.
2. Haggar, F.A.; Boushey, R.P. Colorectal cancer epidemiology: incidence, mortality, survival, and risk factors. *Clin Colon Rectal Surg* **2009**, *22*, 191-197, doi:10.1055/s-0029-1242458.
3. Strum, W.B. Colorectal Adenomas. *N Engl J Med* **2016**, *375*, 389-390, doi:10.1056/NEJMc1604867.
4. Salari, K.; Watkins, H.; Ashley, E.A. Personalized medicine: hope or hype? *Eur Heart J* **2012**, *33*, 1564-1570, doi:10.1093/eurheartj/ehs112.

5. Aziz, M.A.; Yousef, Z.; Saleh, A.M.; Mohammad, S.; Al Knawy, B. Towards personalized medicine of colorectal cancer. *Crit Rev Oncol Hematol* **2017**, *118*, 70-78, doi:10.1016/j.critrevonc.2017.08.007.
6. Alhumaid, A.; AlYousef, Z.; Bakhsh, H.A.; AlGhamdi, S.; Aziz, M.A. Emerging paradigms in the treatment of liver metastases in colorectal cancer. *Crit Rev Oncol Hematol* **2018**, *132*, 39-50, doi:10.1016/j.critrevonc.2018.09.011.
7. Lander, E.S.; Linton, L.M.; Birren, B.; Nusbaum, C.; Zody, M.C.; Baldwin, J.; Devon, K.; Dewar, K.; Doyle, M.; FitzHugh, W., et al. Initial sequencing and analysis of the human genome. *Nature* **2001**, *409*, 860-921, doi:10.1038/35057062.
8. Tazi, J.; Bakkour, N.; Stamm, S. Alternative splicing and disease. *Biochim Biophys Acta* **2009**, *1792*, 14-26, doi:10.1016/j.bbadis.2008.09.017.
9. Hagen, R.M.; Ladomery, M.R. Role of splice variants in the metastatic progression of prostate cancer. *Biochem Soc Trans* **2012**, *40*, 870-874, doi:10.1042/BST20120026.
10. Aziz, M.A.; Periyasamy, S.; Al Yousef, Z.; AlAbdulkarim, I.; Al Otaibi, M.; Alfahed, A.; Alasiri, G. Integrated exon level expression analysis of driver genes explain their role in colorectal cancer. *PLoS One* **2014**, *9*, e110134, doi:10.1371/journal.pone.0110134.
11. Eldai, H.; Periyasamy, S.; Al Qarni, S.; Al Rodayyan, M.; Mustafa, S.M.; Deeb, A.; Al Sheikh, E.; Khan, M.A.; Johani, M.; Yousef, Z. Novel genes associated with colorectal cancer are revealed by high resolution cytogenetic analysis in a patient specific manner. *PLoS One* **2013**, *8*, e76251.
12. National Cancer Institute. The Cancer Genome Atlas. NIH. [[Accessed 2021 May 25]]. Available from: https://cancergenome.nih.gov/.
13. Gao, J.; Aksoy, B.A.; Dogrusoz, U.; Dresdner, G.; Gross, B.; Sumer, S.O.; Sun, Y.; Jacobsen, A.; Sinha, R.; Larsson, E., et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal* **2013**, *6*, pl1, doi:10.1126/scisignal.2004088.
14. Nilsen, T.W.; Graveley, B.R. Expansion of the eukaryotic proteome by alternative splicing. *Nature* **2010**, *463*, 457-463, doi:10.1038/nature08909.
15. Pan, Q.; Shai, O.; Lee, L.J.; Frey, B.J.; Blencowe, B.J. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet* **2008**, *40*, 1413-1415, doi:10.1038/ng.259.
16. Wang, E.T.; Sandberg, R.; Luo, S.; Khrebtukova, I.; Zhang, L.; Mayr, C.; Kingsmore, S.F.; Schroth, G.P.; Burge, C.B. Alternative isoform regulation in human tissue transcriptomes. *Nature* **2008**, *456*, 470-476, doi:10.1038/nature07509.
17. Munkley, J.; Livermore, K.; Rajan, P.; Elliott, D.J. RNA splicing and splicing regulator changes in prostate cancer pathology. *Hum Genet* **2017**, *136*, 1143-1154, doi:10.1007/s00439-017-1792-9.
18. Oltean, S.; Bates, D.O. Hallmarks of alternative splicing in cancer. *Oncogene* **2014**, *33*, 5311-5318, doi:10.1038/onc.2013.533.
19. Martinez-Montiel, N.; Rosas-Murrieta, N.H.; Anaya Ruiz, M.; Monjaraz-Guzman, E.; Martinez-Contreras, R. Alternative Splicing as a Target for Cancer Treatment. *Int J Mol Sci* **2018**, *19*, doi:10.3390/ijms19020545.
20. Le, K.Q.; Prabhakar, B.S.; Hong, W.J.; Li, L.C. Alternative splicing as a biomarker and potential target for drug discovery. *Acta Pharmacol Sin* **2015**, *36*, 1212-1218, doi:10.1038/aps.2015.43.
21. Lefave, C.V.; Squatrito, M.; Vorlova, S.; Rocco, G.L.; Brennan, C.W.; Holland, E.C.; Pan, Y.X.; Cartegni, L. Splicing factor hnRNPH drives an oncogenic splicing switch in gliomas. *EMBO J* **2011**, *30*, 4084-4097, doi:10.1038/emboj.2011.259.

22. Xiong, Y.; Deng, Y.; Wang, K.; Zhou, H.; Zheng, X.; Si, L.; Fu, Z. Profiles of alternative splicing in colorectal cancer and their clinical significance: A study based on large-scale sequencing data. *EBioMedicine* **2018**, *36*, 183-195, doi:10.1016/j.ebiom.2018.09.021.
23. Venables, J.P.; Klinck, R.; Bramard, A.; Inkel, L.; Dufresne-Martin, G.; Koh, C.; Gervais-Bird, J.; Lapointe, E.; Froehlich, U.; Durand, M., et al. Identification of alternative splicing markers for breast cancer. *Cancer Res* **2008**, *68*, 9525-9531, doi:10.1158/0008-5472.CAN-08-1769.
24. Klinck, R.; Bramard, A.; Inkel, L.; Dufresne-Martin, G.; Gervais-Bird, J.; Madden, R.; Paquet, E.R.; Koh, C.; Venables, J.P.; Prinos, P., et al. Multiple alternative splicing markers for ovarian cancer. *Cancer Res* **2008**, *68*, 657-663, doi:10.1158/0008-5472.CAN-07-2580.
25. Thorsen, K.; Sorensen, K.D.; Brems-Eskildsen, A.S.; Modin, C.; Gaustadnes, M.; Hein, A.M.; Kruhoffer, M.; Laurberg, S.; Borre, M.; Wang, K., et al. Alternative splicing in colon, bladder, and prostate cancer identified by exon array analysis. *Mol Cell Proteomics* **2008**, *7*, 1214-1224, doi:10.1074/mcp.M700590-MCP200.
26. Yi, Q.; Tang, L. Alternative spliced variants as biomarkers of colorectal cancer. *Curr Drug Metab* **2011**, *12*, 966-974, doi:10.2174/138920011798062355.
27. Huang, X.; Liu, J.; Mo, X.; Liu, H.; Wei, C.; Huang, L.; Chen, J.; Tian, C.; Meng, Y.; Wu, G., et al. Systematic profiling of alternative splicing events and splicing factors in left- and right-sided colon cancer. *Aging (Albany NY)* **2019**, *11*, 8270-8293, doi:10.18632/aging.102319.
28. Buxbaum, J.D.; Liu, K.N.; Luo, Y.; Slack, J.L.; Stocking, K.L.; Peschon, J.J.; Johnson, R.S.; Castner, B.J.; Cerretti, D.P.; Black, R.A. Evidence that tumor necrosis factor alpha converting enzyme is involved in regulated alpha-secretase cleavage of the Alzheimer amyloid protein precursor. *J Biol Chem* **1998**, *273*, 27765-27767, doi:10.1074/jbc.273.43.27765.
29. Chubinskaya, S.; Mikhail, R.; Deutsch, A.; Tindal, M.H. ADAM-10 protein is present in human articular cartilage primarily in the membrane-bound form and is upregulated in osteoarthritis and in response to IL-1alpha in bovine nasal cartilage. *J Histochem Cytochem* **2001**, *49*, 1165-1176, doi:10.1177/002215540104900910.
30. Seals, D.F.; Courtneidge, S.A. The ADAMs family of metalloproteases: multidomain proteins with multiple functions. *Genes Dev* **2003**, *17*, 7-30, doi:10.1101/gad.1039703.
31. Mohammad Azhar Aziz, M.A., Abdulkareem AlAbdulrahman, Mohammed AlDrees and Ibrahim AlAbdulkarim. Mucin Family Genes are Downregulated in Colorectal Cancer Patients. *Journal of Carcinogenesis & Mutagenesis* **2014**, 10.4172/2157-2518.S10-009, doi:10.4172/2157-2518.S10-009.
32. Moniaux, N.; Escande, F.; Batra, S.K.; Porchet, N.; Laine, A.; Aubert, J.P. Alternative splicing generates a family of putative secreted and membrane-associated MUC4 mucins. *Eur J Biochem* **2000**, *267*, 4536-4544, doi:10.1046/j.1432-1327.2000.01504.x.
33. Wu, Y.; Wang, X.; Wu, F.; Huang, R.; Xue, F.; Liang, G.; Tao, M.; Cai, P.; Huang, Y. Transcriptome profiling of the cancer, adjacent non-tumor and distant normal tissues from a colorectal cancer patient by deep sequencing. *PLoS One* **2012**, *7*, e41001, doi:10.1371/journal.pone.0041001.

Table1: Primer design for protein coding alternatively spliced isoforms of ADAM12

Gene isoform	Transcript ID	Primer Sequence	Amplicon Size in base pair (bp)
ADAM12-All		F-GTAGCTGTCAAATGGCG R-CCACAAATCTGTTCCCAC	198bp
ADAM12-1	ENST00000368679.4* NM_003474.6**	F-CCAACTCGTATAGCATGCATC	197bp
ADAM12-3	ENST00000448723.1* NM_001288973.2**	R-CAATGCCCACGTAATGCAC	
ADAM12-2	ENST00000368676.4* NM_021641.5**	F-CTGCTCCTGAGAGAGTAGC	205bp
ADAM12-4	NM_001288974.2**	R-CAGAGCATTAAGTTGCAGCC	
ADAM12-5	NM_001288975.2**		
ADAM12-4	NM_001288974.2**	F-GAAAGAAGCAAGGCAGGA R-CGTTTCCATGACAACAGAC	231bp
GAPDH	ENSG00000111640* NG_007073.2**	F-ACCCAGAAGACTGTG R-CAGTGAGCTTCCCGTTCAG	139bp
ACTNB	ENSG00000075624* NG_007992.1**	F-TGACGTGGACATCCGCAAAG R-CTGGAAGGTGGACAGCGAGG	205bp

*ENSEMBL database ** NCBI database

Table 2. Primer sequence to amplify protein coding alternatively spliced isoforms of MUC4

Gene Isoform	Transcript ID	Primer Sequence	Amplicon Size in base pair (bp)
MUC4-All Isoforms		F-GCTTACTTCAGATGCGATG	210bp
		R-GTCGAGTTTCATGCTCAGG	
MUC4-1	NM_018406.7**	F-GAGAGGCTCTTCTCAGCAG	223bp
	ENST00000463781.3*		
MUC4-6	NM_001322468.1**	R-GATTCTTGTGTGGTCTGCG	
MUC4-6	NM_001322468.1**	F-AGGCCACCCTTCTTCCT	Band A (180bp)
		R-CTGAGGAAGCGTCCGTG	Band B (320bp)
MUC4-5	ENST00000349607.4*	F-CACTTGGTTCGGGCCAGC	186bp
	NM_138297.5**	R-CACATGCGGAAGGAGGCAG	
MUC4-4	ENST00000346145.4*	F-GTCTGCTCCTCACACTGC	158bp
	NM_004532.6**	R-ATTCCTGGGACCACATGCG	
MUC4-1	ENST00000463781.3*	F-CCATCCCTTCCTCAGTATCC	Band A (282bp)
	NM_018406.7**	R-GAAGAGGGGTGGCATGTC	Band B (240bp)

*ENSMBL database **NCBI database

Table 3: FPKM values for ADAM12 and MUC transcripts obtained from RNA Seq data.

ADAM12_ENSG00000148848.10	0.0245429	0.0241492	0.983958701
ADAM12_ENST00000368679.4	0.00916641	0.000339521	0.037039692
ADAM12_ENST00000368676.4	0.0273946	0.0424578	1.549860191
ADAM12_ENST00000467145.1	0	5.45E-05	
ADAM12_ENST00000482291.1	0.0861337	9.29E-05	0.001079091
ADAM12_ENST00000485388.2	0.010143	0	0
ADAM12_ENST00000448723.1	0.0269748	0.00363765	0.134853641
ADAM12_ENST00000494661.1	0	1.44E-18	
MUC4_**ENST00000415455.1**	4.40E-06	4.54E-10	1.03E-04
MUC4_ENST00000308466.8	7.57E-09	1.71E-12	2.26E-04
MUC4_ENST00000392407.2	0.112782	0.0177344	1.57E-01
MUC4_ENST00000339251.5	8.73E-06	0.0596096	6.83E+03
MUC4_ENST00000448861.1	0.00120339	0.00127307	1.06E+00
MUC4_ENST00000349607.4	2.30E-07	4.93E-13	2.14E-06
MUC4_ENST00000346145.4	6.92E-12	2.68E-07	3.87E+04
MUC4_ENST00000478156.1	1.28E-91	1.11E-25	8.67E+65
MUC4_ENSG00000145113.17	0.0510692	0.0332947	6.52E-01
MUC4_ENST00000463781.3	1.02E-66	5.30E-23	5.20E+43
MUC4_ENST00000479406.1	2.42E-93	6.33E-30	2.62E+63
MUC4_ENST00000462323.1	3.35E-94	2.96E-32	8.84E+61
MUC4_ENST00000475231.1	9.90E-95	1.02E-32	1.03E+62
MUC4_ENST00000470451.1	5.26E-94	4.40E-32	8.37E+61
MUC4_ENST00000480843.1	1.55E-94	1.51E-32	9.74E+61
MUC4_ENST00000477086.1	3.49E-90	2.11E-30	6.05E+59
MUC4_ENST00000466475.1	8.35E-91	6.02E-31	7.21E+59
MUC4_ENST00000477756.1	1.63E-90	1.08E-30	6.63E+59
MUC4_ENST00000464234.1	1.63E-56	1.69E-12	1.04E+44
MUC4_ENST00000467235.1	0.0118175	0.0397415	3.36E+00
MUC4_ENST00000469992.1	2.94E-76	0.153917	5.24E+74
MUC4_ENST00000486425.1	2.02E-234	6.08E-94	3.01E+140
MUC4_ENST00000478685.1	2.08E-76	0.11691	5.62E+74

Table 4: RNA-Seq analysis of splicing factor genes in terms of fold change differences between metastatic (SW620) and primary (SW480) colorectal cancer cell lines

Gene	SW620	SW480	Fold Change (SW480/SW620)
SF3B1	287.3	242.6	1.18
U2Af1	91.08	109.47	0.83
U2AF1L4	1.228	2.64	0.47
SRSF2	225.8	200.5	1.12
RBM10	124.05	89.05	1.39
FUBP1	157.3	104.5	1.5
SRPK1	123.48	118.4	1.04

Values represent normalized FPKM values from RNA Seq analyses

Table 5: Details of protein isoforms obtained from Uniprot knowledgebase (UniprotKB). 568

ADAM12	**Mass (Da)**
Isoform 1 *also known as:* 12L	99,542
Isoform 2 *Also known as:* 12S *The sequence of this isoform differs from the canonical sequence as follows:* DNQGLTIGILVTILCLLAAGFVVYLKRKTLIRLL→ EARQEAAESNRERGQGQEPVGSQEHA STASLTLI	80,403
Isoform 3 *The sequence of this isoform differs from the canonical sequence as follows:* 114-116: Missing. 705-738: DNQGLTIGILVTILCLLAAGFVVYLKRKTLIRLL → EARQEAAESNRERGQGQEPVGSQEHASTASLTLI 739-909: Missing.	80,078
Isoform 4 *The sequence of this isoform differs from the canonical sequence as follows:* 114-116: Missing. 705-740: DNQGLTIGILVTILCLLAAGFVVYLKRKTLIRLLFT → GKEARQEAAESNRERGQGQEPVGSQEHASTASLTLI 741-909: Missing.	80,263
MUC4	
Isoform 1 *This isoform has been chosen as the* canonical *sequence*	231,518
Isoform 2 *Also known as:* Sv12, Sv13 *The sequence of this isoform differs from the canonical sequence as follows:* 1117-1157: FTDNGQIIFP...PVALVAPFWD → ASQAGTLWPW...TRRSMVNTAC 1158-2169: Missing.	118,842
Isoform 3 *Also known as:* Sv20 *The sequence of this isoform differs from the canonical sequence as follows:* 1814-1827: VAGCKCDGGTFGRY → LWGALSCVRTSPAL 1828-2169: Missing.	194,147
Isoform 5 *Also known as:* Sv18, Sv19 *The sequence of this isoform differs from the canonical sequence as follows:* 1172-1223: Missing. 1258-1267: GNPVLMGFSS → VEMAFSKTAH 1268-2169: Missing.	1,215

 doi:10.20944/preprints202106.0102.v2

Source: www.uniprot.org. This information was used to analyse ADAM12 and MUC western blotting results from whole cell lysates of CCD841, SW480 and SW620 cells.