

**IDENTIFICAÇÃO E ESTUDO DE EXPRESSÃO DE
VARIANTES DE POLIADENILAÇÃO E DE
MICRORNAS ATRAVÉS DE SEQUENCIAMENTO EM
LARGA ESCALA EM AMOSTRAS DE TUMOR
COLORRETAL**

CAMILA MIRANDA LOPES RAMOS

**Tese apresentada à Fundação Antônio
Prudente para a obtenção do título de Doutor
em Ciências**

Área de Concentração: Oncologia

Orientador: Dr. Raphael Bessa Parmigiani

Co-Orientador: Dr. Pedro A. F. Galante

São Paulo

2013

FICHA CATALOGRÁFICA

Preparada pela Biblioteca da Fundação Antônio Prudente

Ramos, Camila Miranda Lopes

Identificação e estudo de expressão de variantes de poliadenilação e de microRNAs através de sequenciamento em larga escala em amostras de tumor colorretal / Camila Miranda Lopes Ramos – São Paulo, 2013.

206p.

Tese (Doutorado)-Fundação Antônio Prudente.

Curso de Pós-Graduação em Ciências - Área de concentração: Oncologia.

Orientador: Raphael Bessa Parmigiani

Descritores: POLIADENILAÇÃO. 2. NEOPLASIAS COLORRETAIS. 3. SEQUENCIAMENTO EM LARGA ESCALA. 4. MICRORNAS.

AGRADECIMENTOS

Ao meu orientador, Dr. Raphael B. Parmigiani, pelo grande aprendizado, inúmeras oportunidades, incentivo e paciência. É ótimo poder contar com o seu entusiasmo, sinceridade e disponibilidade para discutir sobre tudo.

Ao meu co-orientador, Dr. Pedro A.F. Galante, pela orientação, tempo dedicado, várias discussões e análises de bioinformática.

À Dra. Anamaria A. Camargo, pelas oportunidades, atenção e discussões enriquecedoras.

Ao Dr. Rodrigo Perez, pela colaboração, disponibilidade e vários ensinamentos clínicos.

À Dra Angelita Habr-Gama, pela colaboração fundamental para o desenvolvimento desta tese.

A todos do Centro de Oncologia Molecular - HSL, Ana Paula, Andrei, Bruna Correa, Bruna P. Barros, Daniel, Elisa, Érico, Fábio, Gabriela, Gustavo, Laura, Lilian, Luis Felipe, Luiza, Juliana, Mariana, Nayara, Paola. Agradeço pela ajuda na bancada e análises de bioinformática, pelas ótimas discussões, almoços e churrascos. Vocês tornaram o dia a dia no laboratório mais agradável.

À Paola Carpinetti-Oliveira e à Gabriela Barnabé que estiveram sempre prontas para me ajudar e discutir experimentos, comemorar com aqueles lindos resultados ou me acalmar nas horas de desespero! Sem contar nas longas conversas sobre a vida!

À Nayara Trevisan e à Bruna Paes de Barros que acompanharam a correria final e estiveram dispostas a me ajudar com vários experimentos, principalmente os funcionais.

À Paula Asprino, Fabiana Bettoni e Fernanda Koyama pelas constantes discussões e pelos experimentos de sequenciamento.

À Natália Felício e à Bruna Quevedo pela ajuda com as amostras e pela dedicação ao banco de tumores.

Ao Departamento de Radioterapia do Centro de Oncologia do Hospital Sírio-Libanês, em nome do Dr. Gustavo Marta.

Aos meus grandes amigos "da Ilha" e "da Bio04" presentes em todas as fases da minha vida. Principalmente à Carol e ao Treta que me ouviram e me aconselharam muito durante todo o doutorado. Sinto muito falta de estar com vocês todos os dias!

À minha segunda família Toninho, Branca e Carol pela torcida, atenção e curiosidade pelo meu trabalho, além da preocupação com as "minhas células".

À minha família amada, que sempre compartilha todas as minhas conquistas e dificuldades. Minha mãe, tão carinhosa, calma, meu maior apoio em simplesmente tudo, e ainda é tão atenciosa com a minha vida acadêmica. Meu pai, meu grande exemplo acadêmico de dedicação e realização. Suas opiniões e conselhos são sempre imprescindíveis, sem contar com os ensinamentos de Matlab durante os feriados. Aos meus irmãos Luana e Gabriel que dividem tantas alegrias e risadas e só aumentam minha vontade de ir para casa.

Ao querido Filipe, que com muito amor e paciência conseguiu fazer dessa jornada muito mais fácil. É maravilhoso poder chegar em casa e ter sempre alguém disposto à vibrar com as minhas conquistas ou à me acalmar e aconselhar.

À agência financiadora Fundação de Amparo à Pesquisa do Estado de São Paulo pela concessão da bolsa de doutorado e do auxílio financeiro ao projeto.

RESUMO

Ramos CML. **Identificação e estudo de expressão de variantes de poliadenilação e de microRNAs através de sequenciamento em larga escala em amostras de tumor colorretal.** São Paulo; 2013. [Tese de Doutorado-Fundação Antônio Prudente].

A maioria dos mRNAs de eucariotos adquire uma cauda de poliA não codificadora na extremidade 3' durante a maturação, em um processo chamado de poliadenilação (poliA). Mais da metade dos genes de mamíferos apresentam múltiplos sítios de poliA, que podem levar à formação de transcritos com tamanhos variáveis da 3' UTR. A poliA alternativa (APA) pode influenciar a localização, estabilidade e transporte dos transcritos, de maneira tecido-específica ou até mesmo doença-específica. Isto é resultante de muitos elementos em cis situados nas 3' UTRs estarem envolvidos na regulação pós-transcricional, tais como os sítios de ligação a miRNAs. Neste trabalho foi utilizado o sequenciamento de última geração para identificar e avaliar a expressão de variantes de poliA, e ainda correlacioná-las à expressão de miRNAs presentes nas mesmas amostras. Para tanto, um protocolo original de construção de bibliotecas de cDNAs enriquecidas para a extremidade 3' dos transcritos foi desenvolvido, permitindo o sequenciamento preferencial de 3' UTRs e a identificação de sítios de poliA em duas linhagens derivadas de tumor colorretal (HCT116 e SW480). Variantes de poliA com 3' UTR mais curtas do que o transcrito referência (RefSeq com maior 3' UTR) foram identificadas para mais de 4.000 genes, correspondendo a cerca de 30% dos genes expressos nestas células. Cerca de 60% das variantes curtas perderam pelo menos um sítio alvo de miRNA. Especificamente para estes genes foi observada uma diminuição da razão da expressão da variante longa pela expressão da variante curta (razão VL/VC) em relação aos genes que não possuíam sítios alvo. Além disso, foi encontrada uma diferença significativa da razão VL/VC

entre os genes que possuíam sítios alvo para miRNAs não expressos e genes que possuíam sítios alvo para miRNAs expressos. Essa redução foi ainda maior para os genes nos quais a variante curta perde sítios para miRNAs altamente expressos comparando-se com genes alvo de miRNAs de baixa expressão. Esta abordagem inovadora aumentou a capacidade de identificação e análise de expressão de variantes de poliA e, ainda confirmou a importância da expressão de miRNAs sobre a expressão de variantes de poliA. Paralelamente, foram investigados miRNAs que pudessem ser utilizados como biomarcadores preditivos para pacientes com câncer de reto. O tratamento padrão do câncer de reto consiste na quimiorradioterapia neoadjuvante (QRTn) seguida de cirurgia radical. No entanto, a QRTn pode levar à completa regressão do tumor em até 42% dos pacientes. Para este seletivo grupo de pacientes, com ausência de células tumorais após a QRTn, têm sido considerados procedimentos alternativos à cirurgia radical, para evitar a mortalidade e morbidade associada à mesma. Contudo, a seleção desses pacientes é difícil, principalmente por ser baseada em exames clínicos e radiológicos muitas vezes imprecisos e subjetivos. O objetivo deste trabalho foi identificar biomarcadores capazes de prever a resposta à QRTn. Inicialmente, avaliou-se o perfil global de expressão de miRNAs através do sequenciamento de última geração de 27 pacientes com câncer de reto antes da QRTn. Foram identificados 4 miRNAs (miR-205-5p, miR-21-5p, miR-1246, e miR-1290-3p) diferencialmente expressos entre os pacientes de resposta completa e incompleta. Além disso, o perfil de expressão desses 4 miRNAs foi capaz de agrupar parcialmente os pacientes de acordo com a resposta ao tratamento. A maior acurácia de predição da resposta foi obtida com o miR-21-5p, com uma área sob a curva ROC de 94%. O potencial do miR-21-5p como biomarcador preditivo foi confirmado em um segundo conjunto de amostras de pacientes, para o qual também houve um aumento significativo de sua expressão em pacientes de resposta completa comparados aos de resposta incompleta. Neste grupo de validação, a expressão de miR-21-5p classificou corretamente 57% (4/7) dos pacientes de resposta completa e 80% (4/5) dos pacientes de resposta

incompleta. Buscando-se por genes alvo do miR-21-5p que possivelmente estariam envolvidos na resposta ao tratamento, foi encontrado SATB1, um gene de resistência a múltiplas drogas, que apresentou expressão inversamente correlacionada com a expressão de miR-21-5p. Além disso, a expressão de SATB1 foi significativamente menor nos pacientes de resposta completa em relação aos de resposta incompleta. Deste modo, este estudo apontou o miR-21-5p como um potencial biomarcador preditivo para o tratamento do câncer de reto e, juntamente com exames clínicos e de imagem, poderia ajudar na seleção de pacientes com resposta clínica completa candidatos a estratégias de tratamento alternativas à cirurgia radical.

SUMMARY

Ramos CML. **[Identification and expression study of polyadenylation variants and microRNAs through large-scale sequencing in samples from colorectal tumor]**. São Paulo; 2013. [Tese de Doutorado-Fundação Antônio Prudente].

Most eukaryotic mRNAs acquire an uncoded polyA tail at their 3' end during maturation in a process called polyadenylation (polyA). More than half of mammalian genes have multiple polyA sites, which can lead to the formation of variant transcripts with different 3' UTRs. Alternative polyA (APA) may influence location, stability and transport of transcripts, in tissue or disease-specific manner. This results from many cis-acting elements located within the 3' UTRs being involved in post-transcriptional regulation, such as the miRNA binding sites. Here we used next generation sequencing to identify and evaluate the expression of polyA variants and also correlate them with miRNA expression in the same samples. To do so, an original cDNA library protocol to enrich for the transcripts' 3' ends was developed, enabling the preferential sequencing of the 3' UTRs and identification of polyA sites in two colorectal cancer cell lines (HCT116 and SW480). PolyA variants with 3' UTR shorter than the reference transcript (RefSeq with the longest 3' UTR) were identified for more than 4000 genes, corresponding to 30% of expressed genes in these cells. About 60% of short polyA variants identified have lost at least one conserved miRNA target site. Specifically for these genes, there was a reduction on the ratio of long variants expression by short variants expression (LV/SV ratio) comparing to genes without miRNA target sites. Moreover, there was a significant difference of the LV/LS ratio between genes with target sites for non-expressed miRNAs and genes with target sites for expressed miRNAs. This reduction was even greater for those genes with short variants that lost target sites for highly expressed miRNAs when compared to genes with target sites for miRNAs with low expression. This innovative approach increased the capacity of identification and expression analysis of polyA variants and also confirmed the key role of

miRNA expression over polyA variants expression. In parallel, we investigated miRNAs that could be used as predictive biomarkers for rectal cancer patients. Standard treatment for rectal cancer consists on neoadjuvant chemoradiotherapy (nCRT) followed by radical surgery. However, nCRT can lead to complete tumor regression in up to 42% of patients. For this selective group of patients, with no tumor cells after nCRT, it has been considered alternative procedures other than radical surgery, in order to avoid the postoperative morbidity and mortality associated with it. However, selection of these patients is difficult mainly because it relies on clinical and radiologic exams, often imprecise and subjective. The goal of this study was to identify biomarkers capable of predicting response to nCRT. Initially, global miRNA profile was evaluated by high throughput sequencing in 27 rectal cancer patients before undergoing nCRT. Four miRNAs (miR-205-5p, miR-21-5p, miR-1246, and miR-1290-3p) were found as differently expressed between the patients with complete and incomplete response. Moreover, the expression pattern of these 4 miRNAs was partially capable of clustering patients accordingly to treatment response. The most accurate response prediction was achieved with miR-21-5p, with an area under the ROC curve of 94%. The potential of miR-21-5p as a predictive biomarker was confirmed in second set of patients' samples, where there was also a significant increase of miR-21-5p expression in complete responders compared to incomplete responders. In this validation set, miR-21-5p expression correctly classified 57% (4/7) of complete responders and 80% (4/5) of incomplete responders. Searching for miR-21-5p target genes that could possibly be involved with response to treatment, it was found *SATB1*, a multidrug resistance gene, which showed expression inversely correlated with miR-21-5p expression. Additionally, *SATB1* expression was significantly lower in complete responders compared to incomplete ones. Thus, this study pointed out miR-21-5p as a potential predictive biomarker for rectal cancer and, together with clinical and imaging exams, should aid on selection of patients with complete clinical response that are candidates for alternative treatment strategies rather than radical surgery.

LISTA DE FIGURAS

Figura 1	Os componentes e características da poliadenilação.....	4
Figura 2	Principais características do sequenciamento através do SOLiD.....	14
Figura 3	Biogênese dos miRNAs (via canônica).....	17
Figura 4	Principais etapas para construção da biblioteca de cDNA enriquecida para a extremidade 3' e sequenciada no 454.....	35
Figura 5	Esquema representando o desenho dos primers para validação das variantes mais curtas por RT-PCR.....	41
Figura 6	Esquema representando o desenho dos primers para validação das variantes mais longas.....	42
Figura 7	Principais etapas para construção da biblioteca de cDNA enriquecida para a extremidade 3' e sequenciada no SOLiD.....	49
Figura 8	Esquema do tratamento QRTn de pacientes com câncer de reto avançado.....	66
Figura 9	Ilustração do desenho do oligodT e do que ocorre na extremidade 3' dos transcritos durante as diferentes etapas da construção da biblioteca de cDNA enriquecido.....	82
Figura 10	Comparação do número de genes com APA identificados apenas em uma linhagem celular ou presentes em ambas.....	86

Figura 11	Distância, em pares de bases (pb), entre o final 3' das variantes de poliA identificadas e o final 3' dos genes RefSeq.....	87
Figura 12	Validação por RT-PCR das variantes curtas de poliA identificadas no 454.....	90
Figura 13	Validação por RT-PCR das variantes longas de poliA identificadas no 454.....	91
Figura 14	Perfil do RNA no Bioanalyzer utilizado para construção da biblioteca de cDNA.....	95
Figura 15	Esquema ilustrando as sequências geradas no SOLiD enriquecidas para a extremidade 3' de alguns genes conhecidos.....	97
Figura 16	Esquema ilustrando variantes identificadas no 454 e confirmadas no SOLiD.....	98
Figura 17	Número de genes apresentando variantes de poliA mais curtas do que gene RefSeq com maior 3' UTR identificadas nas linhagens HCT116 e SW480.....	101
Figura 18	Validação por RACE 3' das variantes de poliA identificadas em ambas as linhagens.....	107
Figura 19	Perfil de tamanho dos fragmentos de uma biblioteca de miRNA de linhagem celular avaliado em uma corrida de chip de DNA High Sensitivity no Bioanalyzer.....	110
Figura 20	Expressão avaliada por qPCR de 2 miRNAs identificados como diferencialmente expressos nos dados de sequenciamento.....	113

Figura 21	Expressão de variantes de poliA em função do número de sítios alvo para miRNA.....	117
Figura 22	Expressão de variantes de poliA em função da expressão de miRNAs.....	119
Figura 23	Expressão de variantes de poliA em função do nível de expressão dos miRNAs.....	120
Figura 24	RT-PCR semi quantitativa das variantes de poliA longa e curta que perderam sítios alvo de miRNAs.....	124
Figura 25	Validação da expressão das variantes de poliA do gene <i>TIMP2</i>	126
Figura 26	Validação da expressão das variantes de poliA do gene <i>FLNB</i> ...	128
Figura 27	Validação da expressão das variantes de poliA do gene <i>CALM3</i>	130
Figura 28	Perfil de tamanho dos fragmentos de uma biblioteca de miRNA de amostra de paciente avaliado em uma corrida de chip de DNA High Sensitivity no Bioanalyzer.....	143
Figura 29	Expressão (cpm) dos miRNAs diferencialmente expressos identificados no EdgeR (FDR<0,05) comparando os pacientes de resposta completa (n=7) e incompleta (n=20).....	147
Figura 30	Gráfico de escala multidimensional para visualizar a similaridade das amostras.....	149

Figura 31	Clusterização hierárquica das amostras de pacientes de resposta completa e incompleta com base na expressão dos 4 miRNAs identificados como diferencialmente expressos.....	150
Figura 32	Análise de Bootstrap das amostras de pacientes com resposta completa ("C") e incompleta ("I") com base na expressão dos 4 miRNAs identificados como diferencialmente expressos.....	151
Figura 33	Curva ROC para determinar a acurácia da predição de resposta a QRTn baseada na expressão de miRNAs.....	152
Figura 34	Exemplos de curvas de dissociação obtidas por qPCR.....	154
Figura 35	Expressão relativa por qPCR dos miRNAs identificados como diferencialmente expressos através do sequenciamento de última geração.....	156
Figura 36	Correlação dos dados de expressão de miRNAs obtidos por qPCR e por sequenciamento.....	158
Figura 37	Clusterização hierárquica das amostras do grupo de validação com base na expressão dos 4 miRNAs identificados como diferencialmente expressos no grupo de treinamento.....	161
Figura 38	Análise de Bootstrap do grupo de validação dos pacientes com resposta completa ("C") e incompleta ("I") com base na expressão dos 4 miRNAs identificados como diferencialmente expressos no grupo de treinamento.....	162
Figura 39	Expressão (cpm) do miR-21-5p para todas as amostras sequenciadas.....	166

Figura 40	Expressão de genes alvo de miR-21-5p nos pacientes de resposta incompleta (n=13) e de resposta completa (n=6).....	172
Figura 41	Correlação da expressão de miR-21-5p e de 4 dos seus genes alvo.....	173
Figura 42	Expressão analisada por sequenciamento de SATB1, gene alvo do miR-21-5p.....	175
Figura 43	Curva ROC para determinar a acurácia da predição de resposta a QRTn baseada na expressão de SATB1.....	177
Figura 44	Expressão relativa de SATB1 avaliada por qPCR nos pacientes de resposta completa e incompleta ao tratamento de QRTn.....	178
Figura 45	Proliferação e viabilidade celular após diferentes doses de Qt....	181
Figura 46	Proliferação e viabilidade celular após diferentes doses de Rt....	182
Figura 47	Proliferação e viabilidade celular durante 10 dias após o tratamento de Rt.....	183
Figura 48	Inibição de miR-21-5p na HCT116.....	185
Figura 49	Superexpressão de miR-21-5p na SW480.....	185

LISTA DE TABELAS

Tabela 1	Dados obtidos do sequenciamento no 454 das bibliotecas de cDNA enriquecidas para porção 3'.....	80
Tabela 2	Alinhamento das sequências obtidas em relação à porção do gene correspondente.....	80
Tabela 3	Identificação das sequências geradas no 454 que possuem o sítio de restrição enzimática (FatI ou MboI), trímero de A ou T e pelo menos um sinal canônico de poliA (AATAAA e ATATAA).....	83
Tabela 4	Número de genes apresentando variantes de poliA mais curta ou mais longa do que as sequências RefSeq, dados obtidos no 454.....	85
Tabela 5	Dados de sequências do SOLiD 5500 obtidas a partir das bibliotecas de cDNA enriquecidas para a porção 3' dos transcritos e mapeadas contra o genoma humano referência..	100
Tabela 6	Avaliação no Bioanalyzer das amostras de RNA utilizadas na construção das bibliotecas de miRNA.....	109
Tabela 7	Dados obtidos a partir do sequenciamento das bibliotecas de miRNA no SOLiD.....	111
Tabela 8	MiRNAs diferencialmente expressos entre as linhagens HCT116 e SW480 avaliados no EdgeR.....	112
Tabela 9	Dados clínicos de todos os pacientes incluídos no estudo.....	138

Tabela 10	Avaliação no Bioanalyzer das amostras de RNA do grupo de treinamento utilizadas na construção das bibliotecas de miRNA.....	141
Tabela 11	Avaliação no Bioanalyzer da porcentagem de miRNAs presentes nas bibliotecas de pacientes do grupo de treinamento (faixa de fragmentos de 115-130pb).....	143
Tabela 12	Sequências obtidas no SOLiD4 a partir das bibliotecas de miRNAs dos pacientes do grupo de treinamento.....	145
Tabela 13	miRNAs diferencialmente expressos (FDR<0,05) no grupo de treinamento entre os pacientes de resposta completa e de resposta incompleta.....	147
Tabela 14	Ciclo de amplificação (Ct) do gene normalizador RNU6b.....	155
Tabela 15	Sequências obtidas no SOLiD4 a partir das bibliotecas de miRNAs dos pacientes do grupo de validação.....	160
Tabela 16	Expressão obtida por sequenciamento do miR-21-5p para as 2 linhagens de tumor colorretal HCT116 e SW480 (normalizada por cpm - número de sequências por milhão).....	180

LISTA DE QUADROS

Quadro 1	Primers utilizados na etapa de construção de bibliotecas de cDNA compatíveis com o 454.....	38
Quadro 2	Primers desenhados para validação das variantes de poliadenilação por RT-PCR.....	44
Quadro 3	Primers utilizados para a validação das variantes de poliA por RACE 3'.....	53
Quadro 4	Sequência dos primers utilizados para análise de expressão por PCR em tempo real.....	56
Quadro 5	Primers utilizados para análise de expressão relativa do gene SATB1 através de qPCR.....	73
Quadro 6	Genes com APA selecionados para validação por RT-PCR.....	88

LISTA DE SIGLAS E ABREVIATURAS

5-FU	5-fluorouracila
µg	micrograma
µl	microlitro
µM	micromolar
APA	poliadenilação alternativa
ATCC	<i>American Type Culture Collection</i>
ATP	adenosina trifosfato
AUC	área sob a curva, do inglês <i>area under the curve</i>
BLAT	<i>BLAST-Like Alignment Tool</i>
CCR	câncer colorretal
cDNA	DNA complementar
CDS	região codificadora, do inglês <i>coding sequence</i>
cols	colaboradores
cpm	número de sequências por milhão, do inglês <i>counts per million</i>
CPSF	fator específico de clivagem e poliadenilação, do inglês <i>cleavage and polyadenylation specificity factor</i>
cRC	resposta completa clínica
CstF	fator F de estimulação à clivagem, do inglês <i>cleavage stimulation factor F</i>
Ct	<i>Cycle Treshold</i>
DEPC	di-etil pirocarbonato
DNA	ácido desoxirribonucléico
dNTP	desoxirribonucleotídeo trifosfatado
DNase	desoxirribonuclease
EDTA	ácido etilenodiaminotetracético disódio
ESTs	Etiqueta de sequências expressas, do inglês <i>Expressed Sequence Tags</i>
FDR	taxa de descoberta falsa, do inglês <i>false discovery rate</i>
GAPDH	Gliceraldeído 3-fosfato desidrogenase

HAOC	Hospital Alemão Oswaldo Cruz
iPS	células tronco pluripotentes induzidas, do inglês <i>induced pluripotent stem cells</i>
LV	leucovorin
ml	mililitros
mRNA	RNA mensageiro
miRNA	microRNA
MIS	instabilidade de microssatélite, do inglês <i>microsatellite instability</i>
MLH1	<i>human mut-L homologue 1</i>
ng	nanograma
nm	nanômetro
nt	nucleotídeo
ORF	fase aberta de leitura (do inglês "Open Reading Frame")
PABP	proteína de ligação a poliA, do inglês <i>poliA binding protein</i>
pb	pares de base
PCR	reação em cadeia da polimerase, do inglês <i>Polimerase Chain Reaction</i>
Pol2	polimerase II
PoliA	poliadenilação
Ppi	pirofosfato
pRC	resposta completa patológica
pre-mRNA	mRNA precursor
pre-miRNA	miRNA precursor
pri-mRNA	miRNA primário
qPCR	PCR quantitativa, do inglês <i>quantitative PCR</i>
QRTn	quimiorradioterapia neoadjuvante
Qt	quimioterapia
RACE	<i>Rapid Amplification of cDNA Ends</i>
RIN	Número de integridade do RNA, do inglês <i>RNA integrity number</i>
RNA	ácido ribonucléico

RNAse	ribonuclease
ROC	<i>receiver operating characteristic</i>
rRNA	RNA ribossômico
Rt	radioterapia
RT-PCR	transcrição reversa - PCR, do inglês <i>Reverse Transcriptase – PCR</i>
SATB1	<i>special AT-rich binding protein 1</i>
TMM	<i>trimmed mean of M values</i>
TRG	grau de resposta do tumor, do inglês <i>tumor regression grade</i>
UTR	região não traduzida, do inglês <i>untranslated region</i>
VL/VC	Variante longa / variante curta

ÍNDICE

1	INTRODUÇÃO	1
1.1	Poliadenilação alternativa.....	1
1.2	Sequenciadores de última geração	9
1.3	miRNAs	14
1.4	O câncer colorretal	19
1.5	O câncer de reto e o seu tratamento	21
2	OBJETIVOS	27
2.1	Objetivo geral	27
2.1.1	Objetivos específicos.....	27
3	MATERIAL E MÉTODOS.....	29
3.1	PARTE I: Identificação e estudo de expressão de variantes de poliadenilação e de miRNAs.....	29
3.1.1	Linhagens celulares.....	29
3.1.2	Extração de RNA.....	30
3.1.3	Análise computacional do sequenciamento no 454 e identificação de variantes de poliadenilação	31
3.1.4	Análise computacional do sequenciamento no 454 e identificação de variantes de poliadenilação	38
3.1.5	Validação por RT-PCR das variantes de poliadenilação	40
3.1.6	Construção de bibliotecas enriquecidas para extremidade 3' dos transcritos e sequenciamento no SOLiD.....	45
3.1.7	Análise computacional do sequenciamento no SOLiD e identificação de variantes de poliadenilação	49
3.1.8	Validação por RACE 3' das variantes de poliadenilação	51
3.1.9	Análise de expressão de genes com poliadenilação alternativa por qPCR.....	54

3.1.10	Construção e sequenciamento de bibliotecas de miRNAs de linhagens celulares.....	56
3.1.11	Análise computacional do sequenciamento de miRNA.....	60
3.1.12	Avaliação por qPCR da expressão de miRNAs.....	62
3.1.13	Análise estatística.....	63
3.2	PARTE II: Análise de expressão de miRNAs em pacientes com câncer de reto e sua associação com a resposta ao tratamento.....	64
3.2.1	Dados clínicos.....	64
3.2.2	Coleta das amostras.....	68
3.2.3	Construção e sequenciamento de bibliotecas de miRNA de amostras de pacientes.....	69
3.2.4	Análise computacional da expressão de miRNAs e associação com a resposta à QRTn.....	69
3.2.5	Avaliação por qPCR da expressão de miRNAs.....	71
3.2.6	Análise da expressão de genes alvo de miR-21-5p.....	72
3.2.7	Sensibilidade celular à QRT.....	74
3.2.8	Modulação da expressão de miR-21-5p.....	75
4	RESULTADOS E DISCUSSÃO.....	77
4.1	PARTE I: Identificação e estudo de expressão de variantes de poliadenilação e de miRNAs.....	77
4.1.1	Construção de bibliotecas enriquecidas para extremidade 3' dos transcritos e sequenciamento no 454.....	77
4.1.2	Análise computacional do sequenciamento no 454 e identificação de variantes de poliadenilação.....	84
4.1.3	Validação por RT-PCR das variantes de poliadenilação.....	88
4.1.4	Construção de bibliotecas enriquecidas para a extremidade 3' dos transcritos e sequenciamento no SOLiD.....	93
4.1.5	Análise computacional do sequenciamento no SOLiD e identificação de variantes de poliadenilação.....	99
4.1.6	Validação por RACE 3' das variantes de poliadenilação.....	105

4.1.7	Estudo da expressão de miRNAs através do sequenciamento no SOLiD	108
4.1.8	Correlação de variantes de poliadenilação e de miRNAs.....	114
4.1.9	Validação por qPCR da expressão diferencial de variantes de poliadenilação.....	123
4.2	PARTE II: Análise de expressão de miRNAs em pacientes com câncer de reto e sua associação com a resposta ao tratamento	136
4.2.1	Dados clínicos dos pacientes com câncer de reto.....	136
4.2.2	Sequenciamento de miRNAs em amostras de pacientes com câncer de reto	139
4.2.3	Identificação de miRNAs diferencialmente expressos.....	145
4.2.4	Validação técnica da expressão de miRNAs por qPCR	153
4.2.5	Validação experimental da expressão de miRNAs por sequenciamento	159
4.2.6	Avaliação funcional do miR-21-5p.....	169
5	CONCLUSÕES	187
5.1	PARTE I: Identificação e estudo de expressão de variantes de poliadenilação e de miRNAs.....	187
5.2	PARTE II: Análise de expressão de miRNAs em pacientes com câncer de reto e sua associação com a resposta ao tratamento	189
6	REFERÊNCIAS BIBLIOGRÁFICAS	191

ANEXOS

- Anexo 1** Manuscrito do artigo
- Anexo 2** Aprovação do Comitê de Ética em Pesquisa (CEP) - A.C. Camargo Cancer Center
- Anexo 3** Aprovação do Comitê de Ética em Pesquisa - Hospital Alemão Oswaldo Cruz
- Anexo 4** Termo de Consentimento Livre e Esclarecido

1 INTRODUÇÃO

1.1 POLIADENILAÇÃO ALTERNATIVA

Em eucariotos, com raras exceções, os RNA mensageiros (mRNAs) apresentam estruturas e características muito semelhantes. A maioria dos mRNAs possui regiões 3' e 5' não traduzidas (UTR, do inglês *untranslated region*) e uma região central codificadora (CDS, do inglês *coding sequence*) (LUTZ 2008). Estas regiões são definidas pela presença de sítios específicos: na porção 5' do transcrito há o sítio de início da tradução (*start codon*: AUG) e na porção 3', o sítio de parada da tradução (*stop codon*: UAG, UAA ou UGA).

Os precursores de mRNA (pre-mRNA) são transcritos no núcleo a partir do DNA genômico pela RNA polimerase II (polII). Estas moléculas precursoras passam por um extenso processamento co/pós-transcricional antes de serem transportadas para o citoplasma, onde ocorrerá sua tradução em proteínas. Os pre-mRNAs que não são processados eficientemente são rapidamente degradados no núcleo, de maneira que pequenas variações no processamento podem ser pontos importantes no controle da expressão (EDWALDS-GILBERT e MILCAREK 1997). Os principais eventos envolvidos no processamento dos mRNAs são: *capping*, *splicing* e poliadenilação.

O *capping* consiste na adição de uma guanina modificada (contendo um radical metil no carbono 7) na extremidade 5' da molécula de mRNA. A presença do *cap* é importante para o transporte do mRNA do núcleo para o citoplasma, para promover a tradução, proteger contra a atividade de exonucleases e promover a remoção do intron proximal (MOORE e PROUDFOOT 2009). Já o *splicing* consiste na remoção de introns e união de exons e é conduzido por um complexo ribonucleoproteico, chamado spliceossomo, que reconhece sinais específicos na molécula de pre-mRNA (ex: as bordas intron-exon) (BLAUSTEIN et al. 2007). Durante este processo é possível haver a geração de diferentes transcritos a partir de um mesmo pre-mRNA através da inclusão ou remoção de exons e introns, ou ainda através da utilização de diferentes sítios doadores e aceptores de *splicing* (presentes nas bordas intron/exon). Consequentemente, diferentes proteínas podem ser formadas a partir de um mesmo gene, sendo esse processamento alternativo denominado de *splicing* alternativo (BLENCOWE 2006).

Durante a maturação do mRNA também ocorre a poliadenilação, que consiste na clivagem da porção 3' UTR em um sítio específico, e a posterior adição de uma cauda de adeninas a qual não depende da presença de um molde de múltiplas timinas na sequência do DNA a partir do qual o mRNA está sendo transcrito (LUTZ 2008). Ambas as reações são direcionadas por elementos de sequência situados na região 3' UTR do pre-mRNA, também chamados elementos em *cis*. Cabe ressaltar que o processo de

poliadenilação está intrinsecamente associado a outros processos, como *splicing* e a transcrição (LUTZ 2008).

O processamento da extremidade 3' do pre-mRNA possui uma grande importância funcional em eucariotos. Primeiro, o processamento da extremidade 3' promove o transporte do mRNA do núcleo para o citoplasma (VINCIGUERRA e STUTZ 2004). Prova disso, a substituição do sinal de poliadenilação da 3' UTR de um pre-mRNA por um sítio de clivagem de RNA ribossômico produz um mRNA que é clivado, mas não poliadenilado. Esta substituição diminui a razão de concentração de mRNA citoplasmático/nuclear em 10 vezes, indicando uma diminuição no transporte do mRNA e conseqüentemente uma diminuição da expressão proteica.

A poliadenilação promove também a estabilidade dos mRNAs uma vez que, no citoplasma, os mRNAs são degradados inicialmente a partir da extremidade 3' (WICKENS 1997). A proteção é feita não apenas pela presença da cauda poliA em si, mas também pela ligação da PABP (proteína de ligação a poliA, do inglês *polyA binding protein*) à molécula de mRNA. Além disso, a poliadenilação é importante para a tradução do mRNA, através da interação entre a cauda poliA e a proteína PABP com o *cap* da extremidade 5'. Esta interação permite o recrutamento das subunidades de RNAs ribossomais que iniciarão a tradução (SACHS 1997).

Em sua grande maioria, os pre-mRNAs apresentam uma sequência consenso, denominada sinal de poliadenilação (poliA), situada de 10 a 35 nucleotídeos (nt) upstream ao sítio de clivagem (**Figura 1A**). A maior parte dos sinais de poliA são conservados e representados por 2 sequências

canônicas (AAUAAA e AUUAAA), no entanto 15% dos sinais de poliA podem ser sequências variantes, sendo que 10 delas já foram identificadas (BEAUDOING et al. 2000). Outra sequência importante para o reconhecimento do sítio de clivagem pela maquinaria de poliA está normalmente situada dentre os 30nt downstream ao sítio de clivagem, sequência esta rica em “U” e/ou “GU” (**Figura 1A**). Essas duas sequências são consideradas os elementos centrais da poliA. Elementos ou sequências auxiliares, tanto downstream quanto upstream aos elementos centrais também já foram caracterizados e podem aumentar a eficiência de poliA (ELKON e AGAMI 2013).

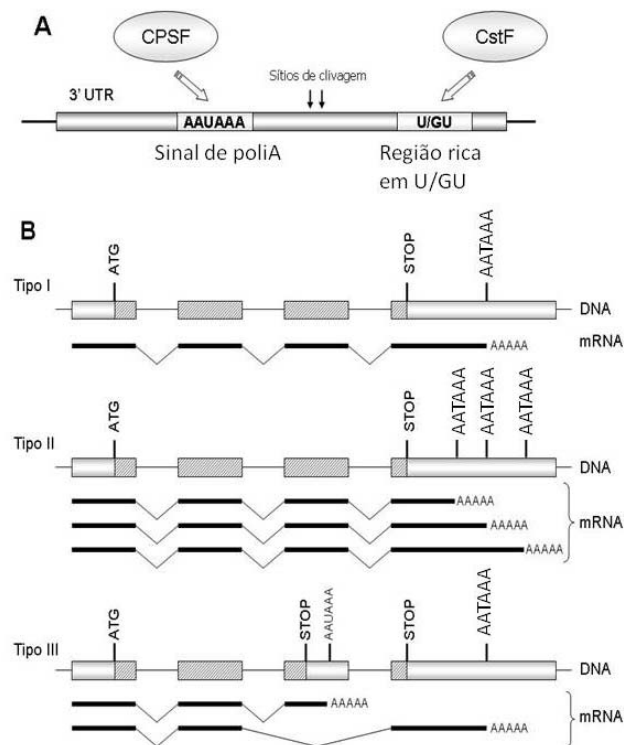


Figura 1 - Os componentes e características da poliadenilação. A) Representação esquemática dos elementos de sequência centrais da poliadenilação e de dois fatores da maquinaria. B) Tipos de poliadenilação alternativa. As caixas representam os exons, sendo a CDS a região hachurada.

Além dos elementos presentes na sequência do mRNA, também são importantes no processo de poliadenilação 5 fatores proteicos de subunidades múltiplas que formam a maquinaria central de clivagem e poliadenilação em mamíferos. Duas proteínas são de especial importância para a clivagem do mRNA, o CPSF (fator específico de clivagem e poliadenilação, do inglês *cleavage and polyadenylation specificity factor*) e o CstF (fator F de estimulação à clivagem, do inglês *cleavage stimulation factor F*) (MANDEL et al. 2008) (**Figura 1A**). O CPSF se associa ao sinal de poliA e determina o sítio de clivagem downstream ao mesmo, preferencialmente à 3' de um dinucleotídeo "CA", embora já tenha sido descrito que a posição de clivagem pode variar em uma distância de 10nt (TIAN et al. 2005; JAN et al. 2011). Já o CstF é responsável pela associação da maquinaria de poliA à sequência rica em "U/GU" do pre-mRNA. Após a clivagem, ocorre a adição de aproximadamente 250nt de adenina à extremidade 3', nesse processo destacam-se a PAP (poliA polimerase), a PABP e o CPSF (COLGAN e MANLEY 1997).

Estudos mostram que mais da metade dos genes humano (54-70%) apresentam mais do que um sinal de poliA e sítio de clivagem (TIAN et al. 2005; DERTI et al. 2012). Esses múltiplos sítios de poliA podem levar à formação de transcritos variantes com diferentes 3' UTR ou mesmo diferentes regiões codificadoras, sendo este fenômeno chamado de poliadenilação alternativa (APA). Considerando-se que durante a maturação do pre-mRNA ocorrem os processos de splicing e poliadenilação, é compreensível que um evento influencie o outro (TIAN e LEE 2007). Dessa

maneira, a poliadenilação pode ser classificada em três tipos de acordo com a influência exercida pelo processo de *splicing* (**Figura 1B**). Tipo I: A poliadenilação é caracterizada pela presença de um único sinal de poliA e não é afetada pelo *splicing*, neste caso não há formação de transcrito alternativo. Tipo II: APA acontece devido à presença de um ou mais sinais de poliA, todos localizados na 3' UTR do mRNA. Tipo III: Um ou mais sinais de poliA estão presentes em diferentes introns ou exons (menos frequentemente) e por isso, APA pode estar associada ao *splicing* alternativo, podendo gerar transcritos que diferem não apenas no tamanho da 3' UTR, mas também na porção codificadora. Este trabalho foi direcionado para o estudo de APA em mRNAs, no entanto este evento também pode ocorrer para RNAs não codificadores, como lncRNAs (RNAs não codificadores longos, do inglês long noncoding RNAs) e miRNAs (microRNAs) (LIN et al. 2012; NI et al. 2013).

Além de poder alterar a proteína codificada, frequentemente esses processamentos alternativos podem influenciar na localização, transporte e estabilidade dos transcritos. Assim, além de um fenômeno de geração de variabilidade, APA também pode ser considerada como uma forma de controle da expressão gênica (LUTZ 2008). Prova disso são os estudos que mostram que APA de mRNAs pode ser diferencialmente regulada em diferentes tecidos (SMIBERT et al. 2012; NI et al. 2013; HAFEZ et al. 2013). Em alguns casos já foi possível fazer a associação entre a presença de promotores específicos com a forma de poliadenilação encontrada em um determinado tecido (COSTESSI et al. 2006; WINTER et al. 2007).

A regulação da escolha do sítio de poliA se dá, principalmente, através da combinação da concentração de fatores da maquinaria de poliA e fatores adicionais de repressão ou indução que são recrutados próximos ao sítio. Os sinais de poliA canônicos são encontrados principalmente próximos aos sítios de poliA distais e possuem maior afinidade pela maquinaria, enquanto os sinais de poliA variantes estão enriquecidos para os sítios proximais e possuem menor afinidade (SHEPARD et al. 2011; LIN et al. 2012). Já foi demonstrado que o uso preferencial de sinais proximais pode estar acompanhado do aumento de expressão de fatores da maquinaria de poliA (DI GIAMMARTINO et al. 2011). Também existem diversas proteínas de ligação ao RNA, como CPEB1 e PABPN1, que podem atuar aumentando ou diminuindo a eficiência da maquinaria de poliA para os sítios proximais, respectivamente (JENAL et al. 2012; BAVA et al. 2013).

APA está intimamente ligada a processos celulares como proliferação e diferenciação. Células com alta taxa de proliferação tendem a apresentar transcritos com 3' UTR mais curtas e menos susceptíveis a mecanismos de controle pós-transcricional (SANDBERG et al. 2008). Com isso, verificou-se um aumento na quantidade da proteína produzida sem haver necessariamente um aumento na expressão do respectivo transcrito. Já para células diferenciadas há uma tendência para o uso de sinais de poliA distais. Por exemplo, durante o desenvolvimento embrionário de camundongos há um aumento global do comprimento da 3' UTR dos transcritos (JI et al. 2009). Essas regiões mais longas permitem um maior controle pós-transcricional da expressão gênica durante o desenvolvimento. Deste modo,

já foram descritos diversos processos nos quais APA tem atuação importante, como: a ativação de células T (SANDBERG et al. 2008), a ativação neuronal (FLAVELL et al. 2008), o desenvolvimento embrionário (JI et al. 2009), diferenciação de células-tronco, geração de células tronco pluripotentes induzidas (iPS) (JI e TIAN 2009) e a ativação de oncogenes (MAYR e BARTEL 2009).

A utilização de sinais de poliA proximais geralmente elimina grande parte da extremidade 3' possibilitando o escape do potencial regulatório presente na mesma (ELKON e AGAMI 2013). Isto se deve ao fato de muitos elementos em cis envolvidos na regulação pós-transcricional estarem localizados nas 3' UTRs, sendo um destes elementos os sítios alvo de microRNA (miRNA). Uma vez que APA pode gerar transcritos com diferentes 3' UTRs, é possível que os transcritos formados sejam seletivamente regulados pelos miRNAs. Corroborando com esta afirmação, no trabalho de Ji e cols (JI et al. 2009), foi verificado que as 3' UTRs mais longas apresentam 30% mais sítios alvo de miRNA, quando comparadas às 3' UTRs mais curtas. Além disso, as 3' UTRs mais longas são formadas por sequências que tendem a ser mais ricas em "AU" e apresentam estruturas menos estáveis de RNA, características que também favorecem a eficiência de ligação dos miRNAs (JI et al. 2009).

Um trabalho utilizou ferramentas computacionais para procurar simultaneamente por sítios de poliA e sítios alvo de miRNA na 3' UTR de genes humanos (MAYR e BARTEL 2009). Os genes que apresentaram o maior número de ambos os sítios foram selecionados para validação

experimental, por meio de experimentos de Northern-blot e RACE (do inglês, *Rapid Amplification of cDNA Ends*). A ocorrência de APA pôde ser confirmada para 6 de 16 genes analisados, sendo que linhagens de células tumorais apresentaram maior número de variantes contendo a 3' UTR menor. Essa alteração trouxe consequências funcionais às células transformadas, já que as variantes com 3' UTR menores apresentaram aumento da estabilidade do mRNA e maior produção de proteína, em parte devido à perda de sequências alvo para miRNA. Foi verificado ainda, que a variante curta do gene *IMP-1* promovia a transformação oncogênica de maneira mais eficiente do que a variante longa, corroborando com a hipótese de que a perda de elementos repressores na 3' UTR através de APA pode promover o fenótipo oncogênico (MAYR e BARTEL 2009).

1.2 SEQUENCIADORES DE ÚLTIMA GERAÇÃO

Inicialmente, os estudos experimentais de variantes geradas por APA foram feitos através de Northern-blot, RACE e *microarray* (SANDBERG et al. 2008; MAYR e BARTEL 2009; JI et al. 2009). Entretanto, todas estas técnicas apresentam significativas limitações. Enquanto as duas primeiras não possibilitam a análise em larga escala, o *microarray* limita-se ao estudo dos genes contidos no *array* e pode haver reação cruzada entre a variante longa e a sonda da variante curta. E ainda, tanto o Northern-blot quanto o *microarray* não possuem sensibilidade suficiente para o estudo de transcritos pouco expressos. Por sua vez, os sequenciadores da nova geração superam

essas limitações. Eles possuem alta sensibilidade e não se restringem ao estudo de transcritos previamente conhecidos e selecionados, permitindo a identificação de novos eventos de APA. Esses sequenciadores são capazes de gerar um enorme número sequências, que podem ser mapeadas no genoma para obter informações sobre o gene correspondente (inclusive sobre variante específica) e estimar sua abundância.

Nos últimos 3 anos, com a utilização da tecnologia de sequenciamento de última geração, nossa compreensão sobre APA melhorou muito. Os estudos foram enriquecedores principalmente para a identificação de um número bem maior de sítios de poliA, bem como a identificação de novos sinais de poliA (OZSOLAK et al. 2010). Um desses trabalhos avaliou 24 tecidos humanos, identificou mais de 280.000 sítios de poliA novos e mostrou que APA ocorre em até 70% do genes humanos DERTI et al. (2012), frequência maior do que o estimado anteriormente (50%) (TIAN et al. 2005). Esses estudos em larga escala também demonstraram o quanto as variantes de poliA podem ser específicas para cada tecido e condição patológica, principalmente em tumores (LIN et al. 2012). Prova disso, um recente trabalho mostrou que é possível diferenciar diferentes tipos tecidos humanos baseando-se na expressão de variantes de poliA (NI et al. 2013). Além disso, pelo menos 10% das variantes de poliA identificadas eram específicas de um único tecido.

Nesses estudos de APA foram desenvolvidos diversos protocolos para sequenciar preferencialmente a extremidade 3' dos transcritos, o que aumentou a capacidade de identificação de sítios de poliA. Vale ressaltar

que quando este trabalho foi iniciado tais protocolos ainda não estavam disponíveis e, portanto, foi estabelecida uma metodologia própria para a identificação e análise de expressão de variantes de poliA através do sequenciamento no 454 (Roche) e no SOLiD (Life Technologies). Posteriormente, a expressão de variantes de poliA foi associada à expressão de miRNAs nas mesmas amostras, uma vez que este estudo em paralelo e em larga escala ainda não havia sido feito.

Existem atualmente no mercado pelo menos três modelos de sequenciadores de última geração, o 454 da Roche, o SOLiD da Life Technologies e o HiSeq da Illumina. O 454 gera em média, por corrida, 1,2 milhões de sequências relativamente mais longas, de 400-500nt. Em contrapartida, os outros dois sequenciadores geram um número bem maior de sequências, de 1 a 3 bilhões, no entanto mais curtas, com 75-200nt.

As plataformas utilizadas nesse trabalho, o 454 e o SOLiD, utilizam como molde para o sequenciamento produtos de PCR originados de uma única molécula de DNA. Durante a construção da biblioteca de DNA ou cDNA a ser sequenciada, adaptadores específicos são ligados aos fragmentos de DNA, os quais permitem sua ligação a um bead e posterior amplificação. A biblioteca e os beads são homogeneizados em fase aquosa e posteriormente adicionados a uma fase oleosa, sob agitação constante. Ao final, forma-se uma emulsão com microbolhas (microrreatores) cujo objetivo é que cada uma contenha um bead e uma molécula de DNA. Nos microrreatores também estão contidos os reagentes necessários para uma PCR convencional, os quais são adicionados durante a preparação da fase

aquosa. A emulsão então é submetida à amplificação em termociclador convencional, de maneira que em cada microrreator ocorre a amplificação de uma única molécula de DNA. Após a amplificação, os beads recobertos por milhares de cópias de uma mesma molécula de DNA são recuperados e submetidos ao sequenciamento.

A tecnologia utilizada pelo sequenciador 454 baseia-se no pirosequenciamento, desenvolvida por Ronaghi e cols (RONAGHI 2001). Para cada etapa da reação apenas um tipo de nucleotídeo não marcado é adicionado e quando um dos dNTP é incorporado durante a reação de extensão, um pirofosfato (Ppi) é liberado. A enzima ATP sulfúrilase converte quantitativamente Ppi em ATP, na presença de adenosina 5' fosfosulfato. Este ATP, por sua vez, possibilita a conversão de luciferina para oxiluciferina pela ação da luciferase que libera luz visível em quantidade proporcional à de ATP. O excesso de nucleotídeos não incorporados é degradado pela apirase, permitindo assim o recomeço do processo através da adição de um novo nucleotídeo. Uma vez que são conhecidos os nucleotídeos adicionados em cada etapa, a sequência pode ser determinada.

O sequenciamento através do SOLiD utiliza química baseada na ligação de oligos marcados com corantes. A primeira etapa consiste na hibridização de um primer ao adaptador, seguida da hibridização por complementaridade de um oligo contendo 8nt, imediatamente adjacente ao primer. Para tanto, uma biblioteca contendo oligos com todas as combinações possíveis de nucleotídeos nas 5 primeiras posições é utilizada em cada reação de ligação ($4^5 = 1024$ oligos diferentes; **Figura 2A**). Já as

últimas três bases são nucleotídeos modificados que fazem o pareamento com qualquer outro nucleotídeo presente na fita molde e são estes os nucleotídeos marcados com os diferentes corantes. Existem quatro corantes possíveis que caracterizam a primeira e segunda base do oligo. Após a ligação do oligo, o sinal do corante é detectado e as três últimas bases removidas enzimaticamente. Os ciclos de hibridização e ligação se repetem, podendo ocorrer até 15 ciclos por primer. Uma vez que o corante utilizado na marcação dos oligos é definido apenas pelos 2 primeiros nucleotídeos de cada oligo (**Figura 2B**), ao final deste primeiro ciclo de ligações, o que se obtém são informações apenas dos nucleotídeos: 1 e 2; 6 e 7; 11 e 12; 16 e 17; e assim sucessivamente até os nucleotídeos 71 e 72. O processo de sequenciamento então se repete utilizando outros primers, que diferem na posição que se anelam ao adaptador, permitindo assim a determinação de outros conjuntos de bases. No total são utilizados 5 primers diferentes com os quais é possível obter as informações sobre todos os nucleotídeos da fita molde. Um programa específico converte a sequência de cores identificada em sequência nucleotídica (**Figura 2D**).

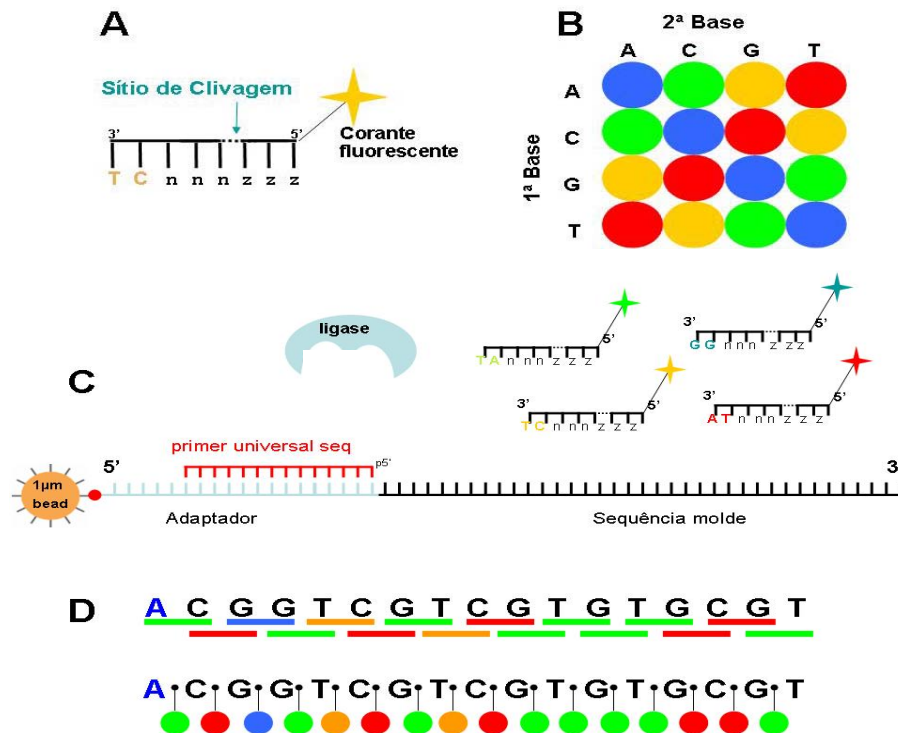


Figura 2 - Principais características do sequenciamento através do SOLiD.

A) Sondas utilizadas no sequenciamento apresentam 8nt, sendo os 3 últimos bases modificadas (z, que pareiam com qualquer nucleotídeo), um sítio de clivagem e um corante fluorescente. B) Os dois primeiros nucleotídeos determinam as cores dos corantes. C) Representação de um *bead* contendo a fita molde, o primer e alguns exemplos de sonda, que será ligada ao molde por complementaridade. D) Cada cor representa 4 possibilidades de dinucleotídeos, a montagem da sequência de nucleotídeos é possível através do conhecimento prévio da primeira base, a qual faz parte da sequência do adaptador (a adenina destacada em azul).

1.3 MIRNAS

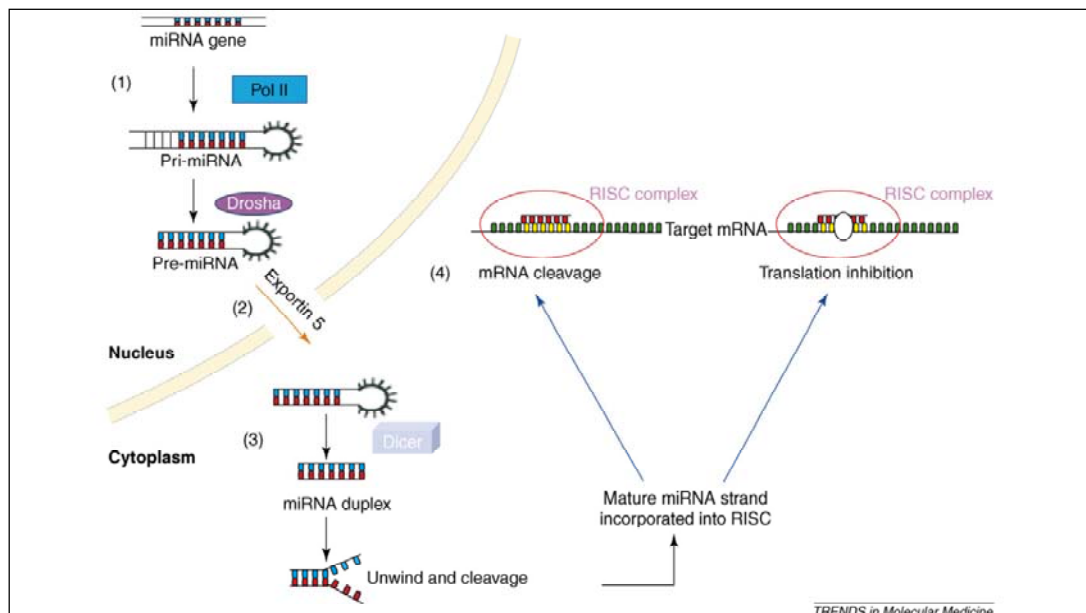
Os miRNAs constituem uma classe de RNAs não codificadores pequenos que agem na repressão de mRNAs em nível pós-transcricional (BARTEL 2004). Esta inibição ocorre através do seu pareamento, por complementaridade, geralmente na 3' UTR do mRNA alvo. Os genes que

codificam miRNAs são transcritos, na sua grande maioria, pela RNA polIII e seu tamanho pode variar de centenas a milhares de nucleotídeos (LEE et al. 2004). A molécula recém transcrita, chamada miRNA primário ou pri-miRNA, assim como mRNAs, possuem cap 5' e são poliadenilados (CAI e CULLEN 2004).

Ainda no núcleo, a molécula de pri-miRNA forma um loop, e conseqüentemente uma região de fita dupla, a qual é reconhecida por um complexo de proteínas que inclui uma RNase III (denominada Drosha) e uma proteína com domínio de ligação a RNA dupla fita (dsRBD, do inglês double-stranded RNA-binding domain). Este complexo cliva o pri-miRNA a partir de suas extremidades 5' e 3', formando uma estrutura em grampo (hairpin) de aproximadamente 70nt, denominado precursor de miRNA ou pre-miRNA (**Figura 3**). Alguns miRNAs estão localizados em introns de genes codificadores, sendo que esses casos não passam pelo processamento da Drosha, os pre-miRNAs são diretamente liberados pela maquinaria de splicing e podem ser clivados por nucleases (RUBY 2007). Ao final, o pre-miRNA é reconhecido e transportado para o citoplasma pela Exportina 5 (LUND et al. 2004).

No citoplasma, ocorre a clivagem do pre-miRNA por um complexo também composto por uma RNase III (denominada Dicer) e proteínas dsRBD (CHENDRIMADA et al. 2005). Neste processo, forma-se uma molécula de fita dupla contendo de 19 a 24 pares de base (pb). Em seguida, as fitas são separadas e a fita guia, ou miRNA maduro, se associa à proteína Argonauta dando origem ao complexo de silenciamento induzido

por RNA (RISC, do inglês *RNA-induced silencing complex*). A outra fita, chamada de passageira, é degradada. A escolha da fita a ser incorporada ao RISC depende da estrutura e propriedades termodinâmicas do duplex de RNA e de interações com proteínas específicas (CZECH e HANNON 2011). Esse processo de escolha pode estar associado ao estágio de desenvolvimento e tecido (GRIFFITHS-JONES et al. 2011) e pode variar, inclusive, entre um tecido normal e tumoral (LI et al. 2012). Sendo assim, a partir de um único pre-miRNA podem ser gerados 2 miRNAs maduros (denominado 5p para a fita referente à porção 5' do pre-miRNA e 3p referente à porção 3'). Por fim, o RISC é direcionado ao mRNA alvo por meio do pareamento de bases. Normalmente, a complementaridade perfeita entre o miRNA e o mRNA leva à clivagem e degradação do mRNA (comum em plantas) (BARTEL 2009). Já o pareamento imperfeito pode levar à repressão da tradução ou desestabilização do mRNA. Em mamíferos, estima-se que o principal mecanismo de ação dos miRNAs se deve à desestabilização do mRNA alvo (para 85% dos casos), sendo o mais comum a deadenilação, que consiste no encurtamento da cauda de poliA, além da remoção do *cap* e degradação do mRNA pela maquinaria de reciclagem (*turnover*) das células (BAEK et al. 2008; GUO et al. 2010).



Fonte: Retirado de GARZON et al. (2006).

Figura 3 - Biogênese dos miRNAs (via canônica). (1) O gene de miRNA é transcrito pela RNA polimerase II formando um transcrito primário longo (pri-miRNA) que é clivado ainda no núcleo pela RNase III Drosha gerando o miRNA precursor (pre-miRNA) com formato de hairpin. (2) O pre-miRNA é exportado do núcleo pela Exportina 5. (3) No citoplasma, o mesmo é processado pela Dicer, produzindo um duplex transitente. Em seguida apenas uma das fitas do miRNA (miRNA maduro) é incorporada ao complexo proteico RISC. (4) Por complementaridade, o miRNA maduro guia o RISC para a degradação do mRNA alvo ou para a repressão da tradução.

A região *seed*, definida pelos nucleotídeos da posição 2 a 7 no miRNA maduro, é determinante para especificar a ligação do miRNA ao seu mRNA alvo (BARTEL 2009). Considerando que a complementaridade não é necessariamente perfeita (com exceção da região *seed*), um mesmo miRNA pode ter centenas de genes alvo e ainda um mesmo transcrito pode ser alvo de diferentes miRNAs (BARTEL 2009). Assim, os miRNAs podem agir de maneira cooperativa, de modo que cada miRNA com sítio alvo em um mesmo mRNA contribui com o efeito de repressão sobre o mRNA (BARTEL 2009).

Em humanos, já foram identificados 2.578 miRNAs maduros, correspondentes a 1.872 pre-miRNAs (de acordo com o banco de dados miRBase versão 20) (GRIFFITHS-JONES et al. 2006; KOZOMARA e GRIFFITHS-JONES 2011). Estima-se que eles sejam capazes de regular a expressão de 30% dos genes humanos (LEWIS et al. 2005), deste modo os miRNAs participam de importantes processos celulares, como a proliferação, diferenciação celular e apoptose. Tais eventos são frequentemente alterados em processos patológicos, em especial na tumorigênese. De fato, diferenças no perfil de expressão de miRNAs em câncer têm sido descritas e sugerem que, assim como os genes codificadores, os miRNAs também podem contribuir para o processo tumorigênico, funcionando como genes supressores de tumor ou oncogenes (CALIN e CROCE 2006; SHENOUDA e ALAHARI 2009). As alterações nos miRNAs podem ser causadas por diversos mecanismos, como deleção, amplificação, mutação, mudanças epigenéticas e transcricionais (CROCE 2009).

O estudo do perfil de expressão de miRNAs em diferentes tipos de tumor é uma ferramenta importante para o diagnóstico, prognóstico e determinação de alvos terapêuticos. Estudos têm mostrado o seu potencial na aplicação clínica, sendo eficientes na distinção entre tecidos normais e tumorais, na determinação da origem do tumor, na predição do comportamento clínico ou resposta terapêutica (LU et al. 2005; NANA-SINKAM et al. 2010).

No câncer colorretal (CCR) os miRNAs estão envolvidos em todo o processo tumorigênico, desde o início, progressão e até a metástase

(SCHETTER et al. 2012). Além disso, o perfil dos miRNAs também está associado à resposta ao tratamento desses pacientes, sendo que a modulação da expressão de miRNAs in vitro pode alterar a sensibilidade das células à quimiorradioterapia (QRT) (SCHETTER et al. 2008; ALLEN e WEISS 2010). Por exemplo, o aumento da expressão de miR-125b e miR-137, resultado da QRT baseada em capecitabina, está associado à pior resposta ao tratamento (SVOBODA et al. 2008). Em outro estudo, pacientes com câncer de cólon que responderam a S-1 (antimetabólitos de 5-fluorouracila) tiveram níveis mais baixos de let-7g e miR-181b do que os pacientes que não responderam (NAKAJIMA et al. 2006). Sendo assim, estudos semelhantes aumentam a expectativa de que os miRNAs possam ser utilizados como marcadores preditivos no câncer (HUMMEL et al. 2010).

1.4 O CÂNCER COLORRETAL

O CCR representa uma das neoplasias mais frequentes e com alta taxa de mortalidade em todo mundo (JEMAL et al. 2011). Esta neoplasia constitui a terceira maior causa de morte por câncer no Brasil e estima-se que cerca de 30.000 novos casos são diagnosticados anualmente (Ministério da Saúde 2011).

O desenvolvimento do CCR consiste em um processo progressivo de acúmulo de alterações genéticas e epigenéticas que transformam o epitélio glandular normal em adenocarcinoma. As etapas envolvidas neste processo foram primeiramente descritas por FEARON e VOGELSTEIN (1990) em um

modelo clássico de progressão a partir da mucosa normal, pólipo, adenoma e adenocarcinoma (FEARON and VOGELSTEIN 1990). Este modelo foi sendo aperfeiçoado com o crescente conhecimento das vias moleculares do CCR, sendo que hoje já foram identificados pelo menos 3 eventos moleculares distintos que podem atuar na progressão pólipo-câncer (AHNEN 2011). Deste modo, dependendo de suas características genéticas e epigenéticas, os tumores colorretais podem ser classificados em 3 tipos moleculares conhecidos como: fenótipo de instabilidade cromossomal (CIN, do inglês *chromosomal instability phenotype*), fenótipo de alta instabilidade de microssatélite (MSI-H, do inglês *high microsatellite instability phenotype*) e fenótipo metilador de ilhas CpG (CIMP, do inglês CpG island methylator phenotype) (AHNEN 2011).

As alterações moleculares presentes nos tumores são específicas para uma dada amostra, mesmo quando tumores de um mesmo tipo histológico ou molecular são comparados. De fato, cada tumor de cólon possui, em média, 90 mutações somáticas em regiões codificadoras e pouquíssimas são compartilhadas entre tumores de diferentes pacientes (SJÖBLOM et al. 2006). Acredita-se que a heterogeneidade molecular do CCR seja responsável pela grande variabilidade na resposta às terapias vista entre os pacientes. Sendo assim, a caracterização de alterações moleculares presentes no CCR são importantes para que novos tratamentos, mais personalizados e efetivos, sejam desenvolvidos ou ainda, que os tratamentos já disponíveis sejam indicados para pacientes que sabidamente têm maior chance de resposta.

1.5 O CÂNCER DE RETO E O SEU TRATAMENTO

Embora o câncer de cólon e de reto normalmente são tratados como um único tipo de câncer, a biologia tumoral e, principalmente, a abordagem terapêutica para cada um é bastante diferente. A maior parte da caracterização molecular do CCR é baseada em amostras provenientes somente do cólon ou os trabalhos não fazem uma clara separação entre cólon e reto. Sendo assim, a caracterização molecular especificamente do câncer de reto ainda é pouco conhecida. Cabe ressaltar que uma taxa significativa dos casos de CCR são decorrentes de câncer de reto, com uma estimativa de 40.000 casos para o ano de 2011 nos Estados Unidos, o que justifica a realização de mais estudos específicos para esta neoplasia (AKLILU e ENG 2011). No Brasil, não existem registros específicos para câncer de reto, mas estima-se que cerca de 20-25% dos casos de CCR sejam originados no reto.

Assim como outros tipos de câncer, o estadiamento anátomo-patológico do câncer de reto tem importância não somente na classificação e determinação do prognóstico da doença, mas também na indicação de tratamento (KOSINSKI et al. 2012). A abordagem do tratamento é influenciada principalmente pela localização do tumor, extensão da invasão na parede do reto, comprometimento de linfonodos e metástase à distância. A excisão local transanal ou microcirurgia endoscópica transanal podem ser recomendados para um pequeno grupo com estágio I com risco mínimo de metástase linfonodal, uma vez que estas modalidades cirúrgicas não

incluem linfadenectomia (PATY et al. 2002). Dessa maneira, podem ser indicados apenas para pacientes de baixo risco, com invasão da parede do reto restrita às camadas da submucosa e muscular própria (T1 e T2) e ausência de metástase em linfonodos regionais (N0). Para os outros casos (T3, T4 e/ou N1), indica-se a excisão total do mesorreto (TME), com ou sem a preservação do esfíncter que depende da extensão e localização do tumor (KOSINSKI et al. 2012).

A ressecção cirúrgica do tumor primário é o tratamento de escolha para o câncer de reto e quando aliada à excisão total do mesorreto é muito eficiente no controle local da doença. No entanto, para pacientes de alto risco, aqueles com tumores invadindo a parede do reto além da camada muscular e/ou linfonodos comprometidos (estádio II e III) o uso de QRT é recomendado, especialmente para reduzir os riscos de recorrência local. A QRT pode ser administrada tanto antes (tratamento neoadjuvante) como após a cirurgia (tratamento adjuvante). No entanto, administra-se principalmente a QRT neoadjuvante (QRTn) por ser menos tóxica e mais eficiente para o controle de recidiva local, com maior impacto na sobrevida global dos pacientes (SAUER et al. 2004). Além disso, o uso de QRT pode reduzir significativamente a massa tumoral permitindo a ressecção total do tumor e aumentando as chances de preservação da função esfínteriana (SAUER et al. 2004).

O esquema de QRTn pode variar entre as diferentes instituições. A radioterapia (Rt) pode ser de baixa duração (25Gy em 5 frações) ou de longa duração (45Gy a 54Gy por um período de 25 a 28 dias) administrada

concomitantemente à quimioterapia (Qt) (KOSINSKI et al. 2012). Já a Qt é baseada, principalmente, em 5-fluorouracila (5-FU) uma vez que a complementação com outras drogas como oxaliplatina e cetuximabe não mostraram benefícios clínicos (GÉRARD et al. 2010; WEISS et al. 2010). O número de ciclos de Qt pode ainda ser estendido durante o intervalo entre a QRTn e a cirurgia definitiva (referida como "período de descanso"), o que tem mostrado aumento das taxas de resposta do tumor (HABR-GAMA et al. 2009; GARCIA-AGUILAR et al. 2011).

De maneira resumida, o tratamento padrão para o câncer de reto avançado consiste na QRTn seguida da cirurgia radical (TME). A resposta ao tratamento neoadjuvante varia significativamente de paciente para paciente, sendo que uma resposta completa patológica (pRC) tem sido relatada para 5 a 42% dos pacientes (SANGHERA et al. 2008). Os casos de pRC estão associados a excelentes resultados oncológicos, especialmente em termos de controle local da doença e aumento da sobrevida (CAPIRCI et al. 2008). Considerando que para estes casos a cirurgia não remove qualquer tipo de câncer residual, estratégias de tratamento alternativo à cirurgia radical têm sido consideradas com o intuito de evitar a morbidade pós-operatória, mortalidade e consequências funcionais associados à TME (KOSINSKI et al. 2012).

O grupo de médicos liderados pela Dra. Angelita Habr-Gama do Hospital Alemão Oswaldo Cruz, São Paulo, propôs uma conduta de observação frequente após a QRTn, sem intervenção cirúrgica imediata, conhecido como "Watch & Wait", para pacientes altamente selecionados,

sem tumor residual clinicamente ou radiologicamente detectáveis (HABR-GAMA et al. 2004). Esta estratégia exige um intenso acompanhamento em que os pacientes são submetidos a frequentes avaliações clínicas, endoscópicas e radiológicas após completar a QRTn, sendo que a cirurgia radical deve ser feita imediatamente em caso de recidiva da doença (HABR-GAMA et al. 2010). Uma conduta semelhante, na qual pacientes com resposta clínica completa após QRTn são acompanhados e não são submetidos imediatamente a cirurgia, também já foi reproduzida por um grupo holandês (MAAS et al. 2011). Uma outra alternativa para os pacientes com pRC seria uma cirurgia menos radical baseada na excisão local ao invés da TME (KIM et al. 2001; SCHELL et al. 2002).

Tendo em vista que a pRC só pode ser confirmada por exames histopatológicos após a cirurgia, os pacientes candidatos a abordagens alternativas só podem ser identificados com base em sua resposta completa clínica (cRC). No entanto, a avaliação clínica da resposta à QRTn ainda é um grande desafio e baseada em exames clínicos, endoscópicos e radiológicos. Pacientes com cRC são aqueles que após o tratamento neoadjuvante apresentam melhora nos sintomas iniciais, resultado negativo no toque retal (ausência de cicatrizes ou úlceras), resultados negativos na tomografia computadorizada ou ultrassonografia endorretal e níveis baixos de antígeno carcinoembrionário (CEA) no sangue (HABR-GAMA 2006). Entretanto, o resultado desses exames são muitas vezes imprecisos e de interpretação subjetiva, além de possuírem baixa capacidade para identificar linfonodos comprometidos. Sendo assim, independente da resposta

observada, a conduta padrão após a QRTn continua sendo a cirurgia radical. Neste cenário, a identificação de marcadores moleculares capazes de prever a resposta tumoral completa permitiriam uma seleção mais rigorosa de pacientes que se beneficiariam de estratégias alternativas e poderiam ser poupados da cirurgia radical imediata após atingir cRC.

A busca por marcadores preditivos tem se intensificado nos últimos anos. A expressão de algumas proteínas pode estar associada à resposta QRTn, como: p53, EGFR, TS, Ki-67, p21 Bcl-2 e Bax (AKIYOSHI et al. 2012). No entanto, os resultados desses trabalhos foram controversos e inconclusivos. Após a introdução das técnicas de análise de expressão gênica em larga escala, os estudos focaram na análise simultânea de um grande número de genes ao invés de focar em genes específicos e, assim, foram obtidos resultados mais animadores. Algumas assinaturas gênicas apresentaram boa acurácia (81-92%) para prever resposta a QRTn, no entanto, existe apenas um gene em comum entre elas, demonstrando que as mesmas possam ser assinaturas específicas das amostras utilizadas em cada estudo (GHADIMI et al. 2005; WATANABE et al. 2006; KIM et al. 2007a; RIMKUS et al. 2008a; DAEMEN et al. 2008a; NISHIOKA et al. 2011). Prova disso é que estes trabalhos foram feitos com um baixo número de amostras (20 a 50) e a maioria das assinaturas não foi validada em um grupo independente de pacientes. Além disso, a classificação da resposta dos pacientes ao tratamento variou muito, sendo que o grupo de "respondedores" pode ter incluído desde pacientes com pRC a pacientes com resposta parcial (com base no grau de regressão do tumor - TRG, do

inglês, *tumor regression grade*). Isso faz com que a comparação entre os estudos seja mais difícil e inconclusiva. É importante ressaltar que o principal interesse desses trabalhos foi identificar biomarcadores que indicassem algum grau de resposta, poucos trabalhos buscaram por marcadores capazes de identificar, exclusivamente, os pacientes com pRC. Em contrapartida, este projeto de doutorado focou no grupo de pacientes que pode se beneficiar de estratégias de tratamento alternativas (ou seja pacientes com pRC ou cRC sustentada) uma vez que aqueles pacientes que não apresentam resposta ou apresentam resposta parcial devem passar pela mesma conduta terapêutica e serem submetidos à cirurgia radical. Por fim, todos os estudos até o momento foram feitos através de microarray ou qPCR e, como descrito anteriormente, o sequenciamento de última geração apresenta algumas vantagens em relação às mesmas e, portanto, foi a metodologia utilizada neste trabalho.

2 OBJETIVOS

2.1 OBJETIVO GERAL

Identificar e estudar a expressão de variantes de poliadenilação e de microRNAs através de sequenciamento em larga escala em amostras de tumor colorretal.

2.2 OBJETIVOS ESPECÍFICOS

- Desenvolver um protocolo de construção de bibliotecas de cDNAs, enriquecidas para a região 3' dos transcritos e que seja compatível com o sequenciamento em larga escala.
- Identificar e avaliar a expressão de variantes de poliadenilação em linhagens celulares de tumor de cólon através de sequenciamento em larga escala.
- Avaliar o perfil de expressão de miRNAs nas mesmas linhagens celulares.
- Correlacionar as variantes de poliadenilação identificadas com a expressão de miRNAs e validar através de PCR em tempo real.
- Avaliar o perfil de expressão de miRNAs em amostras de biópsia de pacientes com câncer de reto.

- Desenvolver um classificador molecular baseado na expressão de miRNAs que seja capaz de prever a resposta ao tratamento de pacientes com câncer de reto.

3 MATERIAL E MÉTODOS

Para facilitar a leitura e apresentação dos dados, a sessão de “material e métodos”, assim como a de “resultados e discussão”, foi dividida em duas partes. A primeira contendo os dados da identificação e estudo de expressão de variantes de poliadenilação e de miRNAs em linhagens celulares e a segunda contendo a análise de expressão de miRNAs em pacientes com câncer de reto e sua associação com a resposta ao tratamento.

3.1 PARTE I: IDENTIFICAÇÃO E ESTUDO DE EXPRESSÃO DE VARIANTES DE POLIADENILAÇÃO E DE MIRNAS

3.1.1 Linhagens celulares

Os experimentos foram realizados com material extraído de duas linhagens derivadas a partir de tumor colorretal, HCT116 e SW480, que diferem quanto à presença de instabilidade de microssatélite (MIS), estando presente na HCT116 e ausente na SW480. As linhagens foram obtidas comercialmente da *American Type Culture Collection (ATCC)* e foram cultivadas de acordo com as recomendações do fornecedor.

3.1.2 Extração de RNA

As linhagens celulares foram cultivadas até atingirem confluência de 70-80% em 6 garrafas de cultura grande, 162cm². O meio de cultura foi aspirado e as células “lavadas” com PBS 1X estéril. Em seguida, adicionou-se 6ml de Trizol (Invitrogen) em cada garrafa. Os lisados celulares de cada garrafa foram transferidos para um tubo Falcon de 15ml e adicionou-se 1200µl de clorofórmio. Os tubos foram agitados vigorosamente durante 30s e, após um intervalo de 5min à temperatura ambiente, foram centrifugados a 3200xg por 30min a 4°C. A fase superior de cada amostra foi transferida para 4 tubos do tipo eppendorf, sendo adicionados 750µl de isopropanol em cada um. As amostras foram homogeneizadas e após 10min à temperatura ambiente, foram centrifugadas a 12000xg por 30min a 4°C. Os sobrenadantes foram descartados e os pellets, lavados duas vezes com 1ml de etanol 70%. Após centrifugar a 12000xg por 5min a 4°C, os sobrenadantes foram descartados e cada um dos pellets foi ressuscitado em 50µl de água DEPC.

A determinação da concentração de RNA foi feita em espectrofotômetro no comprimento de onda de 260 nm (NanoDrop, Thermo Scientific). Também foi analisada a pureza do material através da razão das leituras a 260 e 280 nm, a qual foi utilizada como parâmetro na estimativa do grau de contaminação do RNA por proteínas. Para que fosse considerado puro, o RNA deveria apresentar o valor desta relação igual ou superior a 1,8. A integridade dos RNAs extraídos foi analisada aplicando-se 300ng de RNA em gel de 1% agarose. Antes de ser aplicado, o RNA foi desnaturado a 65°C

por 5min, sendo mantido sob condição desnaturante em tampão de amostra contendo uréia (2x TAE; 30% glicerol; 7M uréia; traços de azul de bromofenol). Foram considerados íntegros os RNAs que apresentaram as bandas correspondentes aos RNAs ribossômicos 28S e 18S bem evidentes e na razão 2:1.

A qualidade das amostras de RNA também foi avaliada com auxílio do RNA 6000 Nano Chip, no Agilent 2100 Bioanalyzer. Utilizando o 2100 expert software, foram obtidas as informações de concentração e o número de integridade do RNA (RIN). Considerou-se RNA íntegro aqueles com RIN acima de 6.

3.1.3 Construção de bibliotecas enriquecidas para extremidade 3' dos transcritos e sequenciamento no 454

Para o estudo de APA, foi desenvolvido um protocolo inédito para o enriquecimento da extremidade 3' dos transcritos e sequenciamento no 454 (Roche).

Inicialmente realizou-se a purificação de RNA poliadenilado (poliA+) utilizando o kit Oligotex mRNA (Qiagen) seguindo as recomendações do fabricante. Foram realizadas três reações independentes a partir de 1mg de RNA total. Ao final, os produtos das três reações foram combinados e tratados para eliminar possível contaminação com DNA genômico. Para tanto, o RNA poliA+ foi incubado com 20U de DNaseI (Ambion) e 200U de RNase out (Invitrogen) durante 40min a 37°C, seguindo com extração fenol/clorofórmio e precipitação com etanol 100% e acetato de sódio 3M. O

precipitado foi ressuspendido em água DEPC e quantificado em espectrofotômetro, NanoDrop, Thermo Scientific.

Foi realizada uma PCR com primers específicos para a porção intrônica do gene *MLH1* (human mut-L homologue 1) para confirmar que a amostra estava livre de contaminação com DNA genômico. A reação de amplificação foi realizada utilizando-se 100ng de RNA poliA+, 1U de Taq Platinum DNA polimerase (Invitrogen), 1x tampão da enzima, 1,5mM MgCl₂, 0,2mM dNTPs e 0,32μM de cada um dos primers para MLH1 (**Quadro 1**). As condições da reação foram: desnaturação a 95°C por 5min; 35 ciclos de 45s a 95°C, 45s a 55°C, 1min a 72°C; e extensão final de 7min a 72°C. Como controle positivo foi utilizada uma amostra de RNA contaminada com DNA genômico. Os produtos das reações foram analisados em gel de 8% poliacrilamida corado com prata e a ausência de produto amplificado de tamanho esperado (250pb) indicou a eficácia da eliminação do DNA genômico contaminante.

Mesmo realizando a purificação para RNA poliA+ ainda resta RNA ribossômico (rRNA) na amostra, portanto o rRNA foi depletado utilizando o kit Ribominus Transcriptome Isolation (Invitrogen). A metodologia se baseia na remoção seletiva de RNA ribossomal grande (18S e 28S) através da hibridização com sondas específicas. A amostra de RNA foi incubada com as sondas a 75°C por 5min, seguindo 15min a 50°C e 25min a 37°C, continuando com as instruções do fabricante. O RNA purificado resultante foi novamente incubado com as sondas, seguindo o mesmo protocolo, para garantir a eliminação do rRNA.

Posteriormente, realizou-se a síntese de cDNA dupla fita utilizando o kit Message Amp™ II aRNA Amplification (Ambion). Foram feitas quatro reações independentes, cada uma a partir de 110ng de RNA poliA+ e ribo-. O oligodT do kit foi substituído por um desenhado especificamente para este projeto. Para garantir a ligação do oligodT no início da cauda de poliA, adicionou-se na extremidade 3' do oligodT uma base diferente de T, portanto foram utilizados 3 oligodTs tendo A, C ou G na extremidade 3'. Já na extremidade 5' foi adicionado o sítio de restrição para a enzima BsmFI e uma biotina conjugada (**Quadro 1**). Para a síntese da primeira fita de cDNA o RNA foi incubado a 42°C durante 2h na presença do oligodT (5µM), 1x first strand buffer, 4µl de dNTPs, 1µl de RNase inhibitor e 1,2µl ArrayScript, em um volume final de 20µl. Em seguida, foi realizada a síntese da segunda fita de cDNA, adicionando à reação inicial de 20µl: 1x second strand buffer, 4µl de dNTPs, 2,5µl de DNA polimerase e 1µl de RNase H, para um volume final de 100µl. A reação foi incubada a 16°C por 2h. No final, cada reação de cDNA foi purificada utilizando 250µl de cDNA binding buffer e centrifugação na coluna proveniente do kit, seguindo de lavagem com wash buffer e a eluição do cDNA em 30µl de água.

A qualidade dos cDNAs sintetizados foi analisada por duas PCRs, sendo uma para a amplificação do gene *GAPDH* e outra para a amplificação da região 5' do gene *NOTCH2*. O cDNA sintetizado foi diluído 10 vezes e 1µl foi utilizado em cada uma das reações de amplificação. Além disso, foram adicionados 0,4µM de cada um dos primers específicos (**Quadro 1**), 1U de Taq Platinum DNA polimerase (Invitrogen), 1x tampão da enzima, 1,5mM

MgCl₂ e 0,2mM dNTPs. As condições da PCR foram: desnaturação a 95°C por 5min; 28 ciclos para GAPDH e 35 para NOTCH2 de 45s a 95°C, 45s a 60°C, 1min a 72°C; e extensão final de 7min a 72°C. Os produtos das reações foram analisados em gel de 1% agarose corado com GelRed (Biotium).

A **Figura 4** resume as principais etapas da construção das bibliotecas de cDNA enriquecidas para extremidade 3' dos transcritos e compatível com o sequenciamento no 454. Metade da massa obtida do cDNA fita dupla foi digerida com a enzima de restrição Mbol (40U, Fermentas) durante 3h a 37°C e a outra metade com a enzima FatI (4U, New England Biolabs) durante 3h a 55°C. Para recuperação da extremidade 3' (que contém biotina incorporada durante a síntese do cDNA), os produtos da digestão foram incubados com *beads* magnéticas conjugadas à estreptavidina. Cada reação de digestão foi incubada com 500µg de *beads* magnéticas (Dynabeads M-280 Streptavidin, Invitrogen), 400µl de bind and wash buffer (10 mM Tris-HCl pH 7.5, 1 mM EDTA, 2 M NaCl) em volume final de 800µl, durante 30min em rotação a temperatura ambiente. Com auxílio de uma estante magnética as *beads* ficaram imobilizadas na estante, foram lavadas com o mesmo tampão (bind and wash buffer), os sobrenadantes foram desprezados e as *beads* ligadas às extremidades 3' dos cDNAs foram reservadas. As mesmas foram utilizadas em uma reação de digestão com a enzima BsmFI (20U, New England Biolabs) durante 3h a 65°C, cujo sítio foi inserido junto ao oligodT. Deste modo, os cDNAs foram liberados das *beads* e, por fim, purificados

segundo as instruções do kit MinElute e eluidos em 20µl de tampão de eluição (Qiagen).

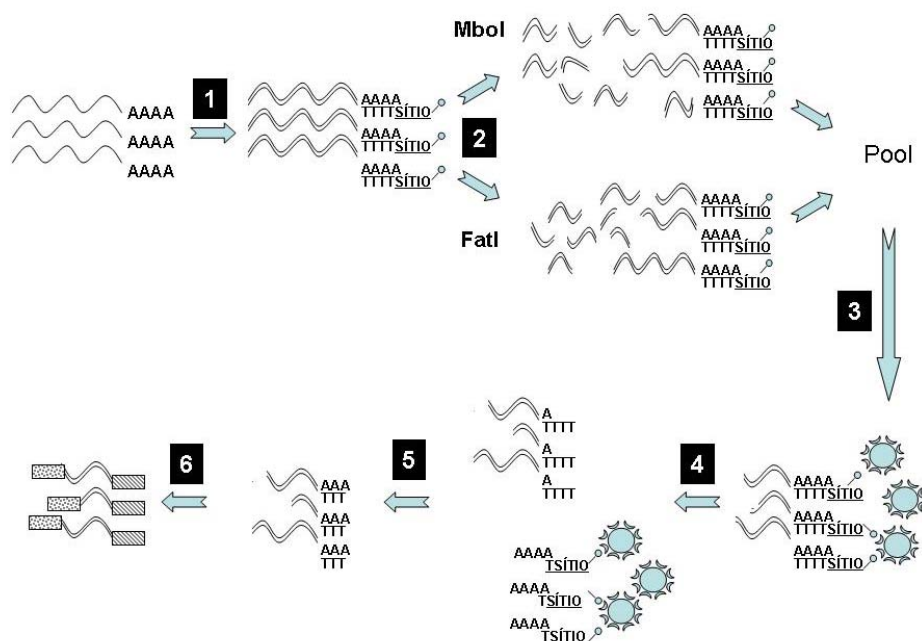


Figura 4 - Principais etapas para construção da biblioteca de cDNA enriquecida para a extremidade 3' e sequenciada no 454. 1. O mRNA foi purificado a partir do RNA total e submetido a síntese de cDNA dupla fita, utilizando o oligodT com sítio de restrição para a enzima BsmFI e biotina conjugada. 2. O cDNA foi digerido com as enzimas de restrição Mbol ou FatI separadamente e, posteriormente, foram combinados. 3. Recuperação da porção 3' dos fragmentos através da incubação com beads magnéticas conjugadas à estreptavidina. 4. Digestão com a enzima BsmFI cujo sítio de restrição foi adicionado junto ao oligodT. 5. Os fragmentos de cDNA livres no sobrenadante foram recuperados e suas extremidades reparadas com T4 polimerase e Taq polimerase. 6. Ligação dos adaptadores (caixas preenchidas) necessários para a amplificação e sequenciamento no 454.

Para confirmar o enriquecimento para a extremidade 3' dos cDNAs e ausência de digestão parcial, foi feita uma reação de RT-PCR utilizando primers para a porção mais 5' do gene *FKBP1B*. A ausência de produto amplificado na porção 3' recuperada mostrou que não houve digestão parcial para o gene *FKBP1B*. Também foram desenhados primers para a porção mais 3' do gene *RPS10*, sendo assim a amplificação do produto demonstrou que o cDNA estava presente na porção recuperada. Cada reação foi feita a partir de 1µl de cDNA, 0,4µM de cada um dos primers específicos (**Quadro 1**), 1U de Taq Platinum DNA polimerase (Invitrogen), 1x tampão da enzima, 1,5mM MgCl₂ e 0,2mM dNTPs. As condições da reação foram: desnaturação a 95°C por 5min; 35 ciclos de 45s a 95°C, 45s a 60°C, 1min a 72°C; e extensão final de 7min a 72°C. Os produtos das reações foram analisados em gel de 1% agarose corado com GelRed (Biotium).

A última etapa foi feita a partir do Library Prep Kit – Rapid Library Rgt/adaptors (Roche), próprio para a construção de bibliotecas para o 454. Inicialmente, as extremidades dos fragmentos foram reparadas utilizando a T4 polimerase e Taq polimerase. Em seguida, foram ligados os adaptadores específicos, compatíveis com o sequenciamento no 454. Ao final, a biblioteca foi purificada utilizando-se *beads* magnéticas (Agencourt Ampure XP for Purification), que se ligam seletivamente a fragmentos maiores do que 100pb e permitem a eliminação de nucleotídeos, excesso de adaptadores, sais e enzimas. Ao final, o cDNA de interesse foi eluído em 60µl de tampão TE.

Devido às inúmeras etapas de digestão, ligação e purificações, a massa obtida foi muito pequena e insuficiente para dar seguimento ao sequenciamento no 454. Assim, as bibliotecas foram amplificadas utilizando-se primers específicos para os adaptadores. A PCR foi feita a partir de 10 μ l da biblioteca, 0,4 μ M de primers (**Quadro 1**), 1U de Taq Platinum DNA polimerase (Invitrogen), 1,5mM de MgCl₂, 0,2mM dNTPs e tampão 1x da enzima. Foram testadas 3 reações com número de ciclos de amplificação diferentes, sendo 15, 20 ou 25. Iniciou-se com a desnaturação a 95°C por 5min, diferente número de ciclos de 45s a 95°C, 45s a 60°C e 1min a 72°C, finalizando com a extensão de 7min a 72°C. Posteriormente, os produtos foram purificados com *beads* magnéticas, Agencourt Ampure XP for Purification.

As bibliotecas foram quantificadas em espectrofotômetro (NanoDrop, Thermo Scientific) e o perfil dos fragmentos obtidos foi analisado com o auxílio do High Sensitivity DNA Chip, Agilent 2100 Bioanalyzer. A massa obtida na reação de amplificação com o menor número de ciclos (15) foi suficiente para prosseguir com PCR em emulsão e sequenciamento no 454 (Roche). Foi feita uma corrida para cada linhagem.

Quadro 1 - Primers utilizados na etapa de construção de bibliotecas de cDNA compatíveis com o 454.

Gene	Primer
Adaptador_Fw	5'- CCA TCT CAT CCC TGC GTG TC -3'
Adaptador_Rv	5'- CCT ATC CCC TGT GTG CCT TG – 3'
OligodTbio	5'- /5Biosg/CGG GGA CTT TTT TTT TTT TTV N -3'
FKBP1B_Fw	5' GCA GCA GCA GGG ACC C 3'
FKBP1B_Rv	5' GTA GTG CAC CAC ACA CGT TTG 3'
GAPDH_Fw	5' CTG CAC CAC CAA CTG CTT A 3'
GAPDH_Rv	5' CAT GAC GGC AGG TCA GGT C 3'
MLH1_Fw	5' TGG TGT CTC TAG TTC TGG 3'
MLH1_Rv	5' CAT TGT TGT AGT AGC TCT GC 3'
NOTCH2_Fw	5' CCA ACC AGT TCT CCT GC 3'
NOTCH2_Rv	5' GTG GTT AGG GCA GTC ATC 3'
RPS10_Fw	5' TCA AGT CCC GAG GCT ACG 3'
RPS10_Rv	5' TCT CGC AGG TCG CTC AC 3'

3.1.4 Análise computacional do sequenciamento no 454 e identificação de variantes de poliadenilação

As sequências obtidas no 454 foram analisadas sob a responsabilidade do Dr. Pedro Galante. As sequências foram alinhadas contra a sequência do genoma humano (GRCh37/hg19) com o auxílio do programa BLAT (KENT 2002). Em seguida, estes resultados foram submetidos a um protocolo de seleção do alinhamento mais confiável para cada sequência (GALANTE et al. 2007). Ao final, as sequências foram anotadas em relação às sequências de genes presentes no banco de dados público RefSeq (<http://www.ncbi.nlm.nih.gov/refseq/>). Deste modo, baseando-se na orientação dos genes RefSeq foi possível avaliar se as sequências geradas poderiam representar novas variantes de poliA, não descritas no RefSeq.

Os critérios para a seleção de possíveis variantes de poliA mais curtas (último exon 3' mais curto que o do RefSeq) foram:

- 1- A coordenada no genoma do começo/final do alinhamento da sequência gerada distar pelo menos 30pb da coordenada correspondente à extremidade 3' do RefSeq. O final da sequência gerada estar localizado dentro da 3' UTR do RefSeq.
- 2- Na extremidade da sequência gerada que corresponde ao 3' do transcrito ser encontrado um trecho de "AAA" ou "TTT" (que indica a posição do sítio de poliA).
- 3- A sequência gerada conter pelo menos um sinal canônico de poliA (AATAAA e ATATAA, como se trata de DNA, foram usadas estas sequências ao invés de AAUAAA e AUUAAA).
- 4- A extremidade 3' da sequência gerada não se alinhar a uma região do genoma rica em "A", a fim de minimizar artefatos experimentais provenientes do anelamento do oligodT a um trecho de "A" no meio do gene. Foram excluídos casos com 10 ou mais "As" em uma janela de 12nt considerando o intervalo de 24nt na extremidade 3' da sequência.
- 5- Pelo menos 2 sequências seguir esses critérios e, portanto, reportar a variante.

Os critérios para a seleção de possíveis variantes de poliA mais longas do que o RefSeq foram:

- 1- A coordenada no genoma do começo/final do alinhamento da sequência gerada distar pelo menos 30pb da coordenada

correspondente à extremidade 3' do RefSeq. O final da sequência gerada estar localizado downstream à 3' UTR do RefSeq.

- 2- A sequência gerada conter pelo menos um sinal de poliA (AATAAA e ATTAAA).
- 3- A extremidade 3' da sequência gerada não se alinhar a uma região do genoma rica em "A" (10 ou mais "As" em uma janela de 12nt considerando o intervalo de 24nt na extremidade 3' da sequência).
- 4- Pelo menos 2 sequências seguir esses critérios e, portanto, reportar a variante.

3.1.5 Validação por RT-PCR das variantes de poliadenilação

Após a identificação de possíveis variantes, foram desenhados primers para a validação experimental por RT-PCR de 8 variantes curtas e 5 variantes longas (Quadro 2). Os primers foram desenhados da maneira descrita abaixo.

Para variantes mais curtas (**Figura 5**):

- 1- Um primer forward "FW1" localizado no segundo exon mais 3' do transcrito, sempre que possível e desde que o mesmo não estivesse descrito como um exon que sofresse *splicing* alternativo.
- 2- Um primer forward "FW2" localizado no exon mais 3' do transcrito quando não foi possível desenhar o primer "FW1" ou o produto amplificado com o "FW1" fosse maior que 2000pb.
- 3- Um primer reverso "R1" para realização de uma PCR como controle positivo da expressão do gene.

- 4- Um primer reverso “R2” cuja sequência consiste em 5nt que se alinham aos últimos nucleotídeos mais 3’ da variante identificada, seguido de um poliT.
- 5- Um primer reverso “R3”, como um segundo controle positivo da presença da variante já descrita (RefSeq).
- 6- Um primer reverso “R_Ctrl”, como controle negativo, para a verificação de que o primer reverso desenhado da maneira sugerida (5nt mais 3’ específicos de cada variante, acrescidos da cauda poliT) não estava inserindo artefato de amplificação para as variantes identificadas.

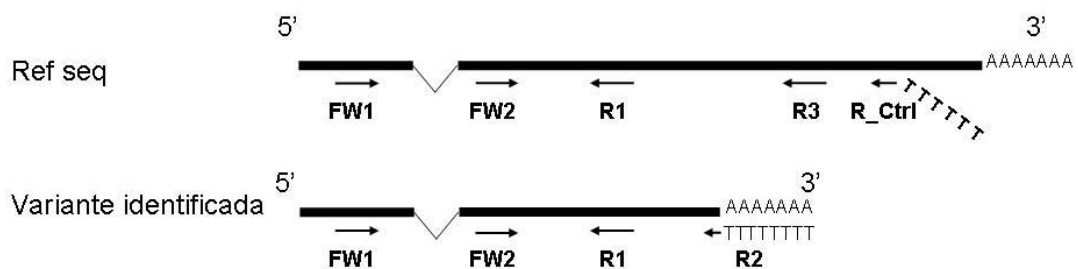


Figura 5 - Esquema representando o desenho dos primers para validação das variantes mais curtas por RT-PCR.

Para variantes longas (**Figura 6**):

- 1- Um primer forward “FW1” localizado no segundo exon mais 3’ do transcrito, sempre que possível e desde que o mesmo não esteja descrito como um exon que sofra *splicing* alternativo.

- 2- Um primer forward “FW2” localizado no exon mais 3’ do transcrito quando não foi possível desenhar o primer “FW1” ou o produto amplificado com o “FW1” fosse maior que 2000pb.
- 3- Um primer reverso “R1”, como controle positivo da presença da variante já descrita (Ref Seq).
- 4- Um primer reverso “R2” cuja sequência consiste em 5 nucleotídeos que se alinham aos últimos nucleotídeos mais 3’ do gene RefSeq, seguido de um poliT, como um segundo controle positivo.
- 5- Um primer reverso “R3”, para confirmação da variante identificada.
- 6- Um primer reverso “R4”, para confirmação da extremidade 3’ da variante identificada. Neste caso, este primer só foi desenhado para as variantes longas que foram confirmadas por sequências que terminaram com o trîmero de A ou T.
- 7- Um primer reverso “R_Ctrl”, como controle negativo, para verificar se o primer reverso desenhado da maneira sugerida não estava inserindo artefato de amplificação para as variantes identificadas.

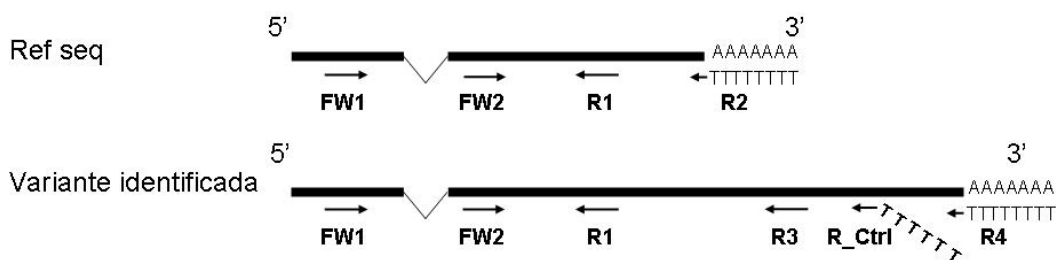


Figura 6 - Esquema representando o desenho dos primers para validação das variantes mais longas.

Para a síntese de cDNA foram utilizados 2µg de RNA total (tratado com DNase) de cada uma das linhagens celulares e ainda: 500ng oligo dT,

0,5mM dNTPs, em 12 μ l de reação. Esta solução foi incubada a 65°C durante 5 minutos e armazenada em gelo. Em seguida, foram adicionados 1X First strand buffer, 0,01M DTT, 40U de RNaseOUT e 200U de SuperScript III RT (Invitrogen), em volume final de 20 μ l. A reação seguiu-se com incubação a 42°C por 1h e depois a 70°C por 15min para inativação da enzima. Como foram feitas várias reações de PCR, foram realizadas 5 reações independentes de síntese de cDNA de cada linhagem e ao final as mesmas foram homogeneizadas e aliquotadas com o intuito de trabalhar com a mesma fonte de material sempre que possível.

Antes de utilizar o cDNA para a PCR de interesse, foi realizado um teste para avaliar a qualidade do mesmo. Para tanto, foi feita uma PCR utilizando-se primers específicos para gene *GAPDH* e para *NOTCH2*, seguindo a mesma reação já apresentada na seção 3.1.3.

Quadro 2 - Primers desenhados para validação das variantes de poliadenilação por RT-PCR.

Gene	Variante identificada	Primers
ABL2	Curta	ABL2_Fw1 - 5' TGC ACA GAG GGT GAG GTT T 3'
		ABL2_R2 - 5' TTT TTT TTT TTT TTT TTC AAG G 3'
		ABL2_R3 - 5' GGC AAA CCT TGT GAG AAG AC 3'
ANKRD28	Curta	ANKRD28_Fw1 - 5' TGA GGG TCT TAC AGG TAT GG 3'
		ANKRD28_Fw2 - 5' GGT GGT TCA GGA ACT TTT GG 3'
		ANKRD28_R2 - 5' TTT TTT TTT TTT TTT TTT CAA CT 3'
ANXA2	Curta	ANKRD28_R3 - 5' GGG ACT CCT GCA TTT GTT AAG 3'
		ANXA2_Fw1 - 5' GGC AAG TCC CTG TAC TAT TAT 3'
		ANXA2_R2 - 5' TTT TTT TTT TTT TTT GTT TCT AA 3'
ASL	Curta	ANXA2_R3 - 5' GAT AGG GGC AAC CAT ACT G 3'
		ASL_Fw1 - 5' TCC GGG AAA GCT GTG TTC 3'
		ASL_R1 - 5' TCT CGC GCC CAC TTT ATT AG 3'
NQO1	Curta	ASL_R2 - 5' TTT TTT TTT TTT TTT TCT CCT 3'
		ASL_R3 - 5' CCT GAC TGT CCC CAC TG 3'
		ASL_R_Ctrl - 5' TTT TTT TTT TTT TTT TCA CGT 3'
POLR2J	Curta	NQO1_Fw1 - 5' ATC CTG CCG AGT CTG TTC 3'
		NQO1_R1 - 5' AAG TGA TGG CCC ACA GAA AG 3'
		NQO1_R2 - 5' TTT TTT TTT TTT TTT TTA CAG T 3'
PSMD8	Curta	NQO1_R3 - 5' TTC AAT GCA CCA CAA GAG G 3'
		POLR2J_Fw1 - 5' TCC TAA AAG ACC CGC AAG TG 3'
		POLR2J_R2 - 5' TTT TTT TTT TTT TTT TGT CTC 3'
ZNF697	Curta	POLR2J_R3 - 5' GGT TTG CAC ACT TCT CTC C 3'
		PSMD8_Fw1 - 5' CCG GAT CCT CTT CTT CAA C 3'
		PSMD8_R1 - 5' TTC TGT GGA GGG AAT GGT G 3'
DUS2L	Longa	PSMD8_R2 - 5' TTT TTT TTT TTT TTT TGT CAG 3'
		PSMD8_R3 - 5' TGG TCC AGC ATT CCT ATG TG 3'
		PSMD8_R_Ctrl - 5' TTT TTT TTT TTT TTT GGC TG 3'
ERO1L	Longa	ZNF697_Fw1 - 5' TAA CCC CTT GAG CCA TGT G 3'
		ZNF697_R1 - 5' ATG CTG TCC AAG GAA CCA TC 3'
		ZNF697_R2 - 5' TTT TTT TTT TTT TTT TGA CTG 3'
MARVELD2	Longa	ZNF697_R3 - 5' TGG CAA AGA AGA ATG ACA TGG 3'
		DUS2L_Fw1 - 5' TCC TCT ATT GTC ACC GTT GC 3'
		DUS2L_R1 - 5' TCT TCA CCA CTT CCC AAG G 3'
RANBP1	Longa	DUS2L_R3 - 5' CTG AAG CAG CCA GCC GT 3'
		ERO1L_Fw1 - 5' TGC CAC CAT ACC CAG TTT C 3'
		ERO1L_Fw2 - 5' TTGGGCACTGCTCTGAAG 3'
ZNF207	Longa	ERO1L_R1 - 5' AGA GGG CAC GAT TTT GAG G 3'
		ERO1L_R3 - 5' GCC ATT TGA ACC CAC ACT G 3'
		ERO1L_R_Ctrl - 5' TTT TTT TTT TTT TTT TCC ATG 3'
RANBP1	Longa	MARVELD2_Fw1 - 5' AGC TGG ATG CAG TGA TGA G 3'
		MARVELD2_R1 - 5' CGC TTG CTT CTC TAC CAC 3'
		MARVELD2_R3 - 5' TCC ACT TGC CCA TCA CAC 3'
ZNF207	Longa	RANBP1_Fw1 - 5' GCG GGC AAA ACT GTT CC 3'
		RANBP1_R1 - 5' CCA CTT TTT CGG CAT GAT CA 3'
		RANBP1_R2 - 5' TTT TTT TTT TTT TTT TTG AGT T 3'
ZNF207	Longa	RANBP1_R3 - 5' AAG TCA TCT TCA ATA GGG C 3'
		RANBP1_R4 - 5' TTT TTT TTT TTT TTT TTC CCA G 3'
		ZNF207_Fw1 - 5' CCA CAG CAA CAA GGA ATG AG 3'
ZNF207	Longa	ZNF207_R1 - 5' CCT TAT TGA CGG AAG CTT GG 3'
		ZNF207_R3 - 5' AAT GGG TCT GGC ACT AGG 3'

3.1.6 Construção de bibliotecas enriquecidas para extremidade 3' dos transcritos e sequenciamento no SOLiD

Foi desenvolvido um novo protocolo de enriquecimento para a extremidade 3' dos transcritos seguido do sequenciamento no SOLiD 5500 (Life Technologies). Utilizou-se a mesma fonte de RNA total das linhagens HCT116 e SW480 e as principais etapas estão resumidas na **Figura 7**.

Após a obtenção do RNA livre de contaminação com DNA genômico foi realizada a depleção do RNA ribossomal utilizando o kit Ribominus Eukaryote for RNA-seq (Invitrogen). Foram realizadas 4 reações independentes, cada uma a partir de 10µg de RNA seguindo as recomendações do fabricante. No final foi feita a purificação utilizando o Ribominus Concentration Module (Invitrogen). O volume das 4 reações foi combinado, adicionado igual volume de binding buffer 1x e etanol para um volume final de 60%. Essa solução foi transferida para a coluna proveniente do kit e centrifugada a 12000xg por 1min. Procedeu-se com uma centrifugação com o tampão de lavagem e eluição do RNA em 32µl de água DEPC. Para a confirmação da remoção dos RNAs ribossomais 18 e 28S, as amostras foram analisadas no Agilent 2100 Bioanalyzer, com o auxílio do RNA 6000 Nano Chip.

Posteriormente, foi realizada a fragmentação do RNA através de uma reação com RNase III, proveniente do kit SOLiD™ Total RNA-Seq. Foram feitas 2 reações separadas, cada uma com 2U de RNase III e 1x tampão da enzima em um volume final de 20µl. As reações foram incubadas a 37°C por 10min e imediatamente adicionado 180µl de água. As 2 reações foram

combinadas e purificadas com Ribominus Concentration Module (Invitrogen), finalizando com a eluição do RNA em 50µl de água DEPC.

Após a fragmentação do RNA a porção 3' dos transcritos foi recuperada utilizando um kit de purificação de RNA poliA+, Oligotex mRNA (Qiagen), que consiste na hibridização da cauda poliA dos transcritos com um oligômero dT conjugado a uma matriz de fase sólida. No final os transcritos poliA+ são liberados abaixando a força iônica e desestabilizando a ligação poliA - oligo dT. O volume total de RNA fragmentado (50µl) foi incubado com 50µl de tampão OBB e 3µl de oligotex por 3min a 70°C, seguido de 10min a temperatura ambiente. Após a incubação, a amostra foi centrifugada para concentrar a matriz e o sobrenadante removido. A matriz foi lavada 2 vezes com tampão OW2 e por fim o RNA foi recuperado com 25µl de tampão de eluição (OEB).

Tendo as amostras enriquecidas para o final 3' dos transcritos, foram construídas as bibliotecas de cDNA utilizando o SOLiD™ Total RNA-Seq Kit (Applied Biosystems). Foi utilizado todo o RNA obtido após a purificação de poliA+ e para tanto o volume foi concentrado através de uma centrifugação a vácuo. Os adaptadores foram ligados através de uma reação contendo 3µl de RNA, 3µl da solução de hibridização e 2µl do mix de adaptadores. As amostras foram incubadas a 65°C por 10min e a 16°C por 5min. Em seguida, foram adicionados 10µl de tampão de ligação 2x, 2µl de Ligation Enzyme Mix e as amostras foram incubadas a 16°C durante 16h. Posteriormente, a essa reação foram adicionados 1x RT buffer, 2µl de dNTPs e 2µl SOLiD RT primer, em um volume final de 19µl, seguido de incubação a 70°C por 5min.

Por último, acrescentou-se 1µl de ArrayScript Reverse Transcriptase e as amostras foram incubadas a 42°C por 30min para a realização da transcrição reversa. Após essa etapa o cDNA foi purificado em colunas MinElute (Qiagen) e eluído em 10µl de tampão de eluição.

O cDNA de interesse foi selecionado em gel de poliacrilamida 6% contendo 7M uréia. Todo o volume de cDNA obtido (10µl) foi diluído em tampão de amostra 2x (Novex TBE-Urea Sample Buffer 2x, Invitrogen) e desnaturado a 95°C por 3min seguido do resfriamento em gelo. As amostras foram aplicadas no gel e submetidas a uma corrida realizada a 180V por 25min. O gel foi corado com SYBR Gold (Invitrogen) diluído em TBE 1x por 8min. Com auxílio de luz UV e baseando-se no marcador de peso molecular 50pb (Invitrogen), a região contendo cDNA de 150 a 250nt foi cortada, a qual corresponde a insertos de 100 a 200nt. Essa região foi dividida em 3 partes iguais longitudinalmente. Para cada porção de gel foi realizada uma reação de amplificação do cDNA contendo 1x tampão da enzima, 0,2mM dNTPs, 1,2µl de AmpliTaq DNA polymerase, 1µl SOLiD 5' PCR primer e 1µl barcoded SOLiD 3' PCR primer. As condições da reação foram: desnaturação a 95°C por 5min; 20 ciclos de 30s a 95°C, 30s a 62°C, 30s a 72°C; e extensão final de 7min a 72°C. O produto das reações foi analisado em gel de 1% agarose corado com GelRed (Biotium). Confirmada a amplificação, o produto da PCR foi purificado utilizando *beads* magnéticas Agencourt AMPure XP. Foram adicionados 1,8µl de *beads* para cada 1µl do volume da PCR seguindo incubação de 5min, na qual ácidos nucleicos de 100pb ou mais se ligam às *beads* magnéticas. O tubo contendo essa

solução foi transferido para uma rack magnética, permitindo a separação entre *beads* ligadas aos fragmentos maiores de 100pb e o sobrenadante, o qual foi descartado. Por fim, as *beads* foram submetidas a duas lavagens com etanol 70% e os fragmentos eluidos com TE.

Finalmente, a biblioteca purificada foi quantificada com o auxílio do fluorômetro Qubit (Invitrogen) e o perfil de tamanho dos fragmentos avaliados no Agilent 2100 Bioanalyzer através de uma corrida com o High Sensitivity DNA Chip.

As bibliotecas das duas linhagens foram misturadas em quantidades iguais (com relação à massa de DNA) e submetidas à PCR em emulsão. O objetivo desta reação é o isolamento e amplificação de moléculas únicas de DNA dentro de microbolhas (microrreatores). Para tanto, o “template” de bibliotecas foi homogeneizado em fase aquosa com *beads* (recobertas com sequências complementares a um dos adaptadores utilizados na construção das bibliotecas) e reagentes necessários para uma PCR convencional. Em seguida, esta solução aquosa foi adicionada a uma fase oleosa e submetida à agitação constante. Ao final, formou-se uma emulsão com microrreatores os quais, idealmente, continham uma *bead* e uma molécula de DNA. A emulsão foi submetida à amplificação em termociclador convencional, de maneira que em cada microrreator ocorreu a amplificação de uma única molécula de DNA. Após a amplificação, as *beads* recobertas por milhares de cópias de uma mesma molécula de DNA foram recuperadas, depositadas em uma lâmina específica e submetidas ao sequenciamento no SOLiD 5500 (Life Technologies).

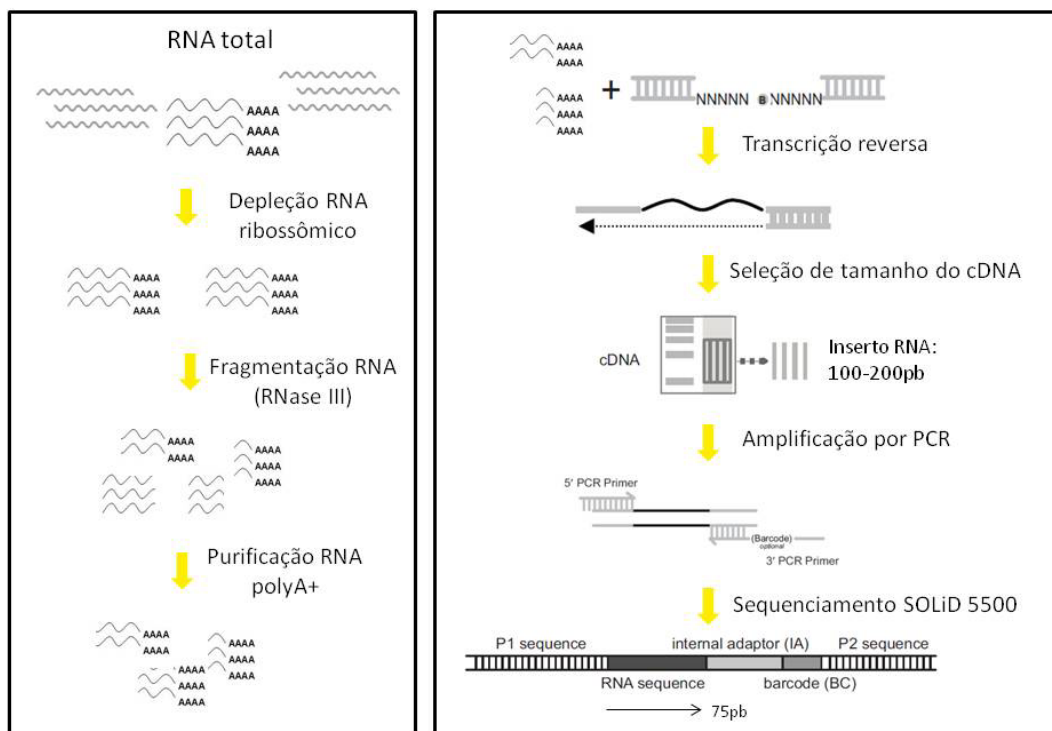


Figura 7 - Principais etapas para construção da biblioteca de cDNA enriquecida para a extremidade 3' e sequenciada no SOLiD.

3.1.7 Análise computacional do sequenciamento no SOLiD e identificação de variantes de poliadenilação

As análises computacionais para a identificação de variantes de poliA foram feitas sob responsabilidade do Dr. Pedro A. F. Galante. As sequências das bibliotecas de cDNA enriquecidas para a porção 3' geradas no SOLiD5500 foram alinhadas contra a sequência do genoma humano (GRCh37/hg19) com o auxílio do programa Bioscope (Life Technologies). Foram mantidas apenas as sequências que apresentaram qualidade de alinhamento ($Q > 20$). Todas estas foram alinhadas em um único local no genoma humano. As coordenadas de alinhamento foram comparadas às coordenadas de alinhamento dos genes Refseq. Deste modo, baseando-se

na orientação dos genes Refseq foi possível avaliar se as sequências geradas apresentavam sítios de poliA diferentes dos sítios de poliA do Refseq. Devido ao curto tamanho das sequências do SOLiD5500 (75pb), foram estudados apenas os sítios de poliA presentes no último exon 3' e, portanto, apenas variantes de poliA que apresentavam a 3'UTR mais curta do que a 3'UTR do gene Refseq foram identificadas.

Primeiramente, foram identificados os sítios de poliA seguindo os critérios abaixo:

- 1- A coordenada no genoma do começo ou final do alinhamento da sequência gerada no SOLiD deve estar dentro da 3'UTR do Refseq.
- 2- Na extremidade da sequência gerada que corresponde ao 3' do transcrito ser encontrado um trecho de no mínimo 3 adeninas ou 3 timinas (correspondentes a uma cauda de poliA) que não se alinham ao genoma.
- 3- A extremidade 3' da sequência gerada não se alinhar a uma região do genoma rica em adeninas (10 ou mais adeninas em uma janela de 12nt considerando o intervalo de 24nt na extremidade 3' da sequência).

Sabe-se que durante o evento de poliA o sítio de clivagem e posterior adição da cauda de poliA pode variar em alguns nucleotídeos (TIAN et al. 2005; JAN et al. 2011). Sendo assim, os sítios de poliA próximos foram agrupados. Começando a partir do sítio de poliA identificado na extremidade mais 5' do Refseq todas as sequências localizadas nos 24nt adjacentes foram agrupadas (TIAN et al. 2005). Para estimar a expressão gênica foi

contabilizado o número total das sequências geradas que apresentavam o seu final alinhado dentro do intervalo de 24nt estabelecido. Foram consideradas como variantes de poliA somente os casos que apresentavam pelo menos 10 sequências com final de alinhamento no intervalo de 24nt, sendo que pelo menos 5 delas deveriam reportar o sítio de poliA, ou seja, possuir 3 adeninas não alinhadas no genoma. Para cada gene foram consideradas apenas a variante definida na extremidade mais 5' e a variante definida na extremidade mais 3' do Refseq, que foram denominadas de variante curta e variante longa, respectivamente. Cabe ressaltar que a variante curta e longa deveriam distar pelo menos 75nt.

Dentro da região que diferenciava a variante de poliA curta da variante longa foi feito uma busca por sítios alvo de miRNAs. Para tanto, foram cruzadas as coordenadas genômicas da região que é perdida na variante curta contra as coordenadas genômicas dos sítios de ligação de miRNAs. Foram considerados todos os sítios alvo conservados que são preditos pelo TargetScan, versão 6.2 (LEWIS et al. 2005).

3.1.8 Validação por RACE 3' das variantes de poliadenilação

Após a identificação computacional das variantes de poliA, foi feita a validação experimental de 10 variantes curtas por RACE 3'. Esta metodologia permite a amplificação da extremidade 3' dos transcritos utilizando um primer *forward* gene-específico e um primer reverso universal complementar ao adaptador inserido durante a transcrição reversa na extremidade 5' do oligodT. Para tanto, a síntese do cDNA foi feita utilizando-

se o kit 3' RACE System for Rapid Amplification of cDNA Ends (Invitrogen), seguindo-se as recomendações do fabricante. Resumidamente, para ambas as linhagens a transcrição reversa foi feita a partir de 5µg de RNA total (tratado com DNase) e um adaptador contendo um trecho de oligodT (500nM). Foram adicionados ainda 20mM Tris-HCl pH 8,4, 50mM KCl, 2,5mM MgCl₂, 10mM DTT, 500µM dNTPs e 1µl SuperScript II RT. As amostras foram incubadas a 42°C por 50min e posteriormente foi feita a inativação da enzima com incubação a 70°C por 15min. Finalmente, foi adicionada 1µl da enzima RNase H às amostras seguido de incubação a 37°C por 20min.

Após a síntese de cDNA, para cada variante candidata foi realizada uma PCR utilizando um primer *forward* gene-específico e um primer reverso complementar ao adaptador, comum para todas as variantes (**Quadro 3**). Para cada variante foi desenhado um primer forward distando cerca de 300pb da extremidade 3' que foi identificada como variante curta pelas sequências geradas no SOLiD. O ideal é que para este mesmo primer, além da variante curta, também seja amplificado o transcrito referência (3' UTR mais longa), o que é importante para confirmar a eficiência da reação. No entanto, uma única reação não é possível para os casos nos quais a extremidade 3' da variante curta se distancia muito da extremidade 3' do transcrito referência. Portanto, para os casos onde a distância entre as extremidades 3' era maior que 1.500pb, foi desenhado um outro primer *forward*, mais próximo da extremidade 3' do gene referência para que, se

necessário, fosse feita uma reação independente para demonstrar a expressão do transcrito referência.

Quadro 3 - Primers utilizados para a validação das variantes de poliA por RACE 3'.

Gene	Primer
RV_adaptador	RV - 5' GGCCACGCGTCTCGACTAGTAC 3'
ACTG1	Fw1 - 5' GAATAGAAATTTGCCCTGG 3'
ANAPC11	Fw - 5' GCATTGCATCCTCAAGTGG 3'
ARF1	Fw1 - 5' CCACCTCTGTGTGTGATGTAG 3'
ATP2A2	Fw1 - 5' GGCCTCCAGTTCATTTTCAG 3'
CALM3	Fw1 - 5' TCGCCTCTGGGACAAGTAAG 3'
FLNB	Fw1 - 5' CTTAGTCCTGAGGAAGTGG 3'
FNDC3B	Fw1 - 5' AACTCATCCCTGGCACAAG 3'
KCMF1	Fw - 5' CACAATCCACAGCAACAACC 3'
TIMP2	Fw1 - 5' GCTGCCGTAATTTAAAGCTC 3'
UQRQ	Fw - 5' TACAGCTTGTCACCGTTTCG 3'

As reações de PCR foram feitas utilizando-se cDNA (50ng de RNA convertido em cDNA) e ainda: 20mM Tris-HCl (pH 8,4), 50mM KCl, 1,5mM MgCl₂, 200nM primer específico, 200nM primer do adaptador, 200µM dNTPs, 1U Taq Platinum DNA polymerase (Invitrogen), em um volume final de 25µL. As condições de ciclagem foram: 95°C por 5 min, seguidos de 5 ciclos de 95°C por 45seg, 65°C por 45seg e 72°C por 1min; 5 ciclos de 95°C por 45seg, 60°C por 45seg e 72°C por 1min; 30 ciclos de 95°C por 45seg, 55°C por 45seg e 72°C por 1min; e extensão final a 72°C por 7min. Os produtos das reações foram analisados em gel de agarose 1% corado com GelRed (Biotium). Para a confirmação da especificidade dos produtos obtidos, os mesmos foram sequenciados pelo método de Sanger. Quando necessário, antes do sequenciamento, os produtos foram purificados a partir de gel de agarose utilizando-se o QIAquick Gel extraction Kit (Qiagen), sob condições recomendadas pelo fabricante. A reação de sequenciamento dos

produtos da PCR foi feita utilizando-se o kit Big Dye Terminator versão 3.1 (Applied Biosystems), de acordo com as instruções do fabricante, e posteriormente analisadas no sequenciador de capilar (ABI3130XL, Applied Biosystems). As sequências obtidas foram estudadas manualmente através de alinhamentos feitos em bancos de dados públicos presentes no site da UCSC (<http://genome.ucsc.edu/>).

3.1.9 Análise de expressão de genes com poliadenilação alternativa por qPCR

Foi realizada a análise por qPCR da expressão das variantes curtas e das variantes longas (que coincidiram com o transcrito RefSeq com maior 3' UTR) para *TIMP2*, *CALM3* e *FLNB*. A mesma fonte de RNA total utilizada na construção das bibliotecas de cDNA e RACE 3' foi utilizada para a transcrição reversa. A síntese de cDNA foi feita a partir de 2µg de RNA, 50µM de oligodT e SuperScript III RT (Invitrogen), seguindo as recomendações do fabricante. A qualidade da síntese de cDNA foi avaliada através de duas PCR para os genes *GAPDH* e *NOTCH2*, como mencionado anteriormente (item 3.1.3).

Para cada gene, foi desenhado um par de primer que amplifica uma região específica da variante longa e um par de primer que amplifica a região comum entre a variante longa e a curta (**Quadro 4**). Inicialmente, foram feitos testes com diferentes concentrações de primer para estabelecer a melhor condição de amplificação, evitando a saturação da reação por falta de primer e também evitando a formação de dímeros de primer ou

amplificação inespecífica por excesso de primer. Após a padronização, as reações seguintes foram realizadas com 400nM de primer, com exceção do CALM3 para amplificação da região comum entre a variante longa e a curta que foi utilizado 100nM. Além disso, também foi feita uma curva com diluição seriada de cDNA (de 250 a 0,4ng) para avaliar a eficiência de cada primer. Todos os pares de primers utilizados apresentaram eficiência maior do que 1,94, sendo que idealmente este valor seria 2. As reações de qPCR foram realizadas, em triplicata, a partir de 25ng de cDNA e GoTaq qPCR Master Mix (Promega), seguindo as condições estabelecidas pelo fabricante. A reação foi realizada no termociclador 7900 Applied Biosystems, sendo as condições de ciclagem: 95°C por 2min, 40 ciclos de 95°C por 15s e 60°C por 60s, finalizando com uma curva de dissociação (95°C por 15s, 60°C por 15s e 95°C por 15s).

A análise de expressão da variante longa entre as duas linhagens celulares foi determinada através da quantificação relativa da expressão, utilizando o método comparativo $\Delta\Delta CT$ (LIVAK e SCHMITTGEN 2001). Para cada amostra, o Ct da variante longa foi normalizado pela expressão total do gene (Ct dado pelo primer que amplifica a região comum entre a variante longa e a curta). Posteriormente, foi feita a comparação dos valores normalizados de uma linhagem em relação à outra, sendo que a amostra com menor Ct foi utilizada como valor de referência. Assim, é possível verificar a diferença de uma linhagem em relação à outra, uma vez que não há uma amostra controle de referência.

Quadro 4 - Sequência dos primers utilizados para análise de expressão por PCR em tempo real.

Gene	Região amplificada	Primer
CALM3	Apenas referência	Fw – 5' CCTGCCCAGCTTTGAGAATC 3'
		Rv– 5' ACTATGCCCCATCCCTCCTT 3'
CALM3	Referência e variante curta	Fw - 5' AGAAGCTGACCGATGAGGAG 3'
		Rv - 5' TACGCAGGGGAGTGTTGAAG 3'
FLNB	Apenas referência	Fw– 5' CTTAGTCCTGAGGAAGTGG 3'
		Rv– 5' TACAGTCGGCGTTTGATTCC 3'
FLNB	Referência e variante curta	Fw– 5' ACCTGCTATAGCGCCATTC 3'
		Rv– 5' GCTTCATGGAGACCTCCTC 3'
TIMP2	Apenas referência	Fw – 5' GCT GCC GTA ATT TAA AGC TC 3'
		Rv– 5' GCA GTC TGC TTG TCA ACT TTC 3'
TIMP2	Referência e variante curta	Fw– 5' TGGATGGACTGGGTCACAGA 3'
		Rv – 5' CTTGATGCAGGCGAAGAACT 3'

3.1.10 Construção e sequenciamento de bibliotecas de miRNAs de linhagens celulares

A análise de expressão de miRNAs foi realizada nas mesmas amostras de RNA da HCT116 e SW480. Foram obtidos os dados de qualidade e concentração das amostras de RNA total com o auxílio do RNA 6000 Nano Chip, Agilent 2100 Bioanalyzer. Em seguida, as amostras também foram analisadas no Small RNA Chip com o auxílio do Bioanalyzer (Agilent) para obtenção da concentração e porcentagem de miRNA na amostra de RNA total. Tendo a concentração do RNA total e a concentração dos miRNAs, foi possível calcular a porcentagem de miRNAs na amostra inicial de RNA total. Como a porcentagem de miRNA era baixa, as amostras foram enriquecidas para RNAs pequenos.

O RNA total (50µg) foi enriquecido para RNAs pequenos utilizando o Invitrogen PureLink™ miRNA Isolation Kit. A amostra de RNA foi diluída em

90µl de água, 300µl de binding buffer (L3) e 210µl de etanol 100%, em seguida a amostra foi centrifugada a 12000xg por 1min na coluna proveniente do kit, de maneira que o RNA de maior peso molecular ficou retido no filtro e os RNAs pequenos permaneceram no eluato. Ao mesmo foram adicionados 700µl de etanol 100% e essa solução foi transferida para uma segunda coluna, de maneira que os RNAs pequenos permaneceram retidos no filtro. Finalmente, o material foi lavado 2 vezes com wash buffer (W5) e eluído em 18µl de água DEPC. Após esse enriquecimento, a amostra foi novamente analisada no Small RNA Chip (Agilent) para obtenção da concentração e porcentagem de miRNA. Uma massa mínima de 1ng de miRNA era necessária para a construção das bibliotecas.

Tendo as amostras enriquecidas para RNAs pequenos, foram construídas as bibliotecas de cDNA utilizando o SOLiD™ Total RNA-Seq Kit (Applied Biosystems). Para tanto, adaptadores foram ligados ao RNA enriquecido (foi utilizada toda a massa de RNA desde que não fosse superior a 1µg) em uma reação contendo 3µl de RNA, 3µl da solução de hibridização e 2µl do mix de adaptadores. As amostras foram incubadas a 65°C por 10min e a 16°C por 5min. Em seguida, foram adicionados 10µl de tampão de ligação 2x, 2µl da enzima de ligação e as amostras incubadas a 16°C durante 16h. Posteriormente, a essa reação foram adicionados 1x tampão de transcriptase reversa, 2µl de dNTPs e 2µl SOLiD RT primer, em um volume final de 19µl, seguido de incubação a 70°C por 5min. Por último, acrescentou-se 1µl de ArrayScript Reverse Transcriptase e as amostras foram incubadas a 42°C por 30min para a realização da transcrição reversa.

Após essa etapa o cDNA foi purificado em colunas MinElute (Qiagen) e eluído em 10 μ l de tampão de eluição.

O cDNA de interesse foi selecionado em gel de poliacrilamida (Novex 10% TBE-Urea Gel, Invitrogen). O volume total de cDNA sintetizado (10 μ l) foi diluído em tampão de amostra (Novex TBE-Urea Sample Buffer 2x, Invitrogen) e desnaturado a 95°C por 3min seguido do resfriamento em gelo. As amostras foram aplicadas no gel e submetidas a uma corrida realizada a 180V por 45min. O gel foi corado com SYBR Gold (Invitrogen) diluído em TBE 1x por 8min. Com auxílio de luz UV e baseando-se no marcador de peso molecular 10pb (Invitrogen), a região contendo cDNA de 60 a 80pb foi excisionada, a qual corresponde a insertos de 18 a 38pb. Essa região foi dividida longitudinalmente em 3 partes iguais, sendo uma delas utilizada como molde para reação de amplificação do cDNA.

A reação de amplificação foi realizada utilizando-se 1x tampão apropriado, 0,2mM dNTPs, 1,2 μ l de AmpliTaq DNA polymerase, 1 μ l SOLiD 5' PCR primer e 1 μ l barcoded SOLiD 3' PCR primer. As condições da reação foram: desnaturação a 95°C por 5min; 15 a 20 ciclos de 30s a 95°C, 30s a 62°C, 30s a 72°C; e extensão final de 7min a 72°C. Cada reação foi primeiramente realizada com 15 ciclos de amplificação, seguida da análise em gel de agarose 1% corado com GelRed (Biotium). Para os casos que a amplificação não foi visível no gel, a reação foi submetida a mais alguns ciclos de amplificação, sendo 20 o máximo. Quando 20 ciclos não foi suficiente para visualizar a amplificação, as outras 2 partes cortadas do gel

contendo cDNA também foram utilizadas como molde para PCR, sendo realizadas 2 reações separadas e ao final todas elas combinadas.

Confirmada a amplificação, o produto de PCR foi purificado utilizando *beads* magnéticas Agencourt AMPure XP. Foram adicionados 1,8µl de *beads* para cada 1µl do volume da PCR seguindo incubação de 5min, na qual ácidos nucleicos de 100pb ou mais se ligam às *beads* magnéticas. O tubo contendo essa solução foi transferido para uma rack magnética, permitindo a separação das *beads* contendo fragmentos maiores de 100pb do sobrenadante, o qual foi descartado. Por fim, foram realizadas duas lavagens com etanol 70% e eluição dos fragmentos com TE. Foi realizada uma segunda purificação com as *beads*, dessa vez, aumentando o volume de *beads* para 2µl para cada 1µl de reação.

Finalmente, as bibliotecas purificadas foram quantificadas com o auxílio do fluorômetro Qubit (Invitrogen) e o perfil de tamanho dos fragmentos avaliados no Agilent 2100 Bioanalyzer através de uma corrida com o High Sensitivity DNA Chip.

As bibliotecas foram misturadas em quantidades iguais (mesma massa) para obter um número de sequências semelhantes entre as mesmas. Foram misturadas 16 bibliotecas (as 2 bibliotecas de linhagens foram misturadas com as bibliotecas de pacientes - item 3.2.3), as quais foram submetidas à PCR em emulsão (seção 3.1.6) e sequenciadas no SOLiD4 (Applied Biosystems). Para gerar entre 25 e 30 milhões de sequências por amostra, foram sequenciadas 16 bibliotecas por lâmina.

3.1.11 Análise computacional do sequenciamento de miRNA

As sequências geradas a partir das bibliotecas de miRNA sequenciadas na plataforma SOLiD4 foram analisadas com o auxílio do *CLC Genomics Workbench 5.1* (CLC Bio). Como as sequências geradas possuem 35nt, espera-se que essas sequências contenham a sequência completa do miRNA (em torno de 22nt) e parte da sequência do adaptador utilizado na construção da biblioteca. O primeiro passo, portanto, consistiu em identificar a sequência do adaptador (CTGCTGTACGGCCAAGGCG) e removê-la do restante da sequência. Em seguida, essas sequências trimadas foram alinhadas contra o banco de miRNAs humano, miRBase (www.mirbase.org, versão 18) (GRIFFITHS-JONES et al. 2006; KOZOMARA e GRIFFITHS-JONES 2011), permitindo no máximo uma base não pareada. Após o alinhamento, as sequências foram anotadas como miRNAs maduros conhecidos. O reconhecimento do mRNA alvo depende principalmente da sequência seed do miRNA localizada na sua porção 5' (2 a 7nt), portanto a anotação das sequências geradas foi feita considerando-se a posição 5' exata do miRNA conhecido comparado com a sequência gerada no SOLiD. Em relação a posição 3' dos miRNAs já foi descrito que eles podem apresentar variação do comprimento na porção 3' e são conhecidos como isomirs (MORIN et al. 2008). Devido a essa variação, a anotação foi feita permitindo que as sequências geradas apresentassem uma diferença de um nucleotídeo a mais ou a menos no final 3' em relação ao miRNA maduro descrito no miRBase. Por último, as sequências anotadas foram contadas e

uma tabela gerada, contendo o número de sequências para cada miRNA maduro conhecido em cada uma das linhagens.

A normalização dos dados foi feita dividindo-se o número de sequências de um determinado miRNA pelo número total de sequências mapeadas em miRNAs maduros para a amostra correspondente e multiplicado por milhão, assim para cada miRNA é gerado um valor correspondente ao número de sequências por milhão (cpm, do inglês counts per million), permitindo comparar a expressão desse miRNA em diferentes amostras. Os miRNAs de baixa expressão foram filtrados da análise após a normalização. Foram mantidos apenas os miRNAs que apresentavam pelo menos 10cpm em uma das linhagens.

A análise de expressão diferencial entre as linhagens foi feita utilizando-se o EdgeR (ROBINSON 2010), um pacote para análise estatística de dados provenientes de sequenciamento em larga escala que faz parte do software Bioconductor 2.10 (GENTLEMAN et al. 2004). Para esta análise foi feita a normalização padrão proposta pelo EdgeR denominada TMM (do inglês trimmed mean of M values). Esta normalização, além de considerar o tamanho da biblioteca também contabiliza a diferença na composição de diferentes RNAs presentes em cada amostra. Foram utilizados os parâmetros padrões do EdgeR para a análise de expressão diferencial. O programa é capaz de inferir a expressão diferencial considerando a distribuição binomial negativa dos dados e modelando a relação média-variância entre os grupos.

3.1.12 Avaliação por qPCR da expressão de miRNAs

Dois miRNAs que apresentaram diferença de expressão utilizando o sequenciamento como metodologia, também tiveram sua expressão avaliada por qPCR, através do sistema miScript PCR (Qiagen). Neste sistema, os miRNAs maduros são seletivamente convertidos em cDNA. Inicialmente, os miRNAs são poliadenilados, seguindo com uma transcrição reversa utilizando primers oligodT que contêm uma sequência universal no final 5'. Essa reação foi feita a partir de 2µg de RNA total, 5x miScript HiSpec Buffer, 10x miScript Nucleics Mix e 2µl miScript Reverse Transcriptase Mix em uma reação de 20µl. A reação foi incubada a 37°C por 1h, seguido de 5min a 95°C. Em seguida, foi feito o qPCR, em triplicata, a partir de 3ng de cDNA, 2x QuantiTect SYBR Green PCR Master Mix, 10x miScript Universal Primer, 10x miScript Primer Assay (primer *forward* específico para o miRNA de interesse) em uma reação de 25µl. A reação foi realizada no 7900 Applied Biosystems, sendo as condições de ciclagem: 95°C por 15 min, 40 ciclos de 94°C por 15s, 55°C por 30s e 70°C por 30s, finalizando com uma curva de dissociação (95°C por 15s, 60°C por 15s e 95°C por 15s).

A expressão diferencial dos miRNAs de interesse foi determinada através da quantificação relativa da expressão, utilizando o método comparativo $\Delta\Delta\text{CT}$ (LIVAK e SCHMITTGEN 2001). Para cada amostra, o Ct do miRNA foi normalizado pelo Ct do gene RNU6b, que tem sido comumente utilizado na literatura como um bom normalizador de dados de expressão de miRNAs (MOTOYAMA et al. 2009; PIEPOLI et al. 2012). Posteriormente, foi feita a comparação dos valores normalizados em cada

linhagem em relação à uma amostra utilizada como referência (RNA extraído de tecido de cólon normal obtido comercialmente da Clontech).

3.1.13 Análise estatística

Os testes estatísticos foram realizados através do programa GraphPad Prism versão 4.03. A diferença da média entre dois grupos de dados que apresentavam distribuição normal foi analisada através de teste t de Student. A diferença da média entre três grupos de dados que apresentavam distribuição normal foi analisada através de ANOVA, seguindo com pós teste de Bonferroni. A diferença da mediana entre dois grupos de dados que não apresentavam distribuição normal foi analisada através de teste de Mann-Whitney. Os testes foram considerados estatisticamente significativo quando $p < 0,05$.

3.2 PARTE II: ANÁLISE DE EXPRESSÃO DE MIRNAS EM PACIENTES COM CÂNCER DE RETO E SUA ASSOCIAÇÃO COM A RESPOSTA AO TRATAMENTO

3.2.1 Dados clínicos

Foram incluídos no estudo 43 pacientes com adenocarcinoma de reto distal, localizado até 7cm da borda anal e estadiamento clínico cT2-4N0M0. Todos os pacientes foram avaliados clinicamente e radiologicamente com ressonância magnética de alta resolução da pelve, tomografias de abdômen e tórax. Pacientes grávidas ou com idade inferior a 18 foram excluídos do estudo. Os pacientes foram avaliados e acompanhados pelo grupo de cirurgiões liderados pela Dra. Angelita Habr-Gama no Hospital Alemão Oswaldo Cruz (HAOC) em São Paulo e no Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo. As amostras foram coletadas somente após o consentimento do paciente e assinatura do termo de consentimento livre e esclarecido do doador do Banco de Tecido e Seroteca do HAOC. Este projeto foi submetido e aprovado pelo comitê de ética em pesquisa do HAOC, nº 02/10.

- **Tratamento dos pacientes e avaliação da resposta**

Após a coleta da biopsia do tumor primário, os pacientes foram submetidos a QRTn, sendo que os pacientes incluídos no estudo não haviam sido tratados previamente. O esquema de tratamento consistiu em doses diárias de Rt por 6 semanas consecutivas e descanso aos finais de

semana, totalizando 54Gy. Os pacientes também receberam 5-FU (425 mg/m²/d) e LV (20 mg/m²/d) durante 3 dias consecutivos ao início e ao final da Rt (HABR-GAMA et al. 2004). Oito semanas (período de "descanso") após o término da QRT os pacientes foram re-estadiados. Este esquema de tratamento é denominado de tratamento padrão (**Figura 8A**). Durante o desenvolvimento deste trabalho, o tratamento QRTn dos pacientes foi aprimorado, uma vez que pacientes tratados com maior número de ciclos de Qt apresentavam maiores taxas de resposta ao tratamento (HABR-GAMA et al. 2009). Este novo tratamento foi denominado de estendido e consistiu na mesma dosagem de Rt, 54Gy durante 6 semanas (**Figura 8B**). Concomitantemente, os pacientes receberam 5-FU (450 mg/m²/d) e LV (dose fixa de 50mg) durante 3 dias consecutivos ao início da Rt e esse ciclo se repetiu a cada 3 semanas. Após o término da Rt, os pacientes ainda receberam mais 3 ciclos de Qt a cada 3 semanas, sendo que ao final das 9 semanas (período de "descanso") o tratamento foi finalizado e os pacientes re-estadiados. Tendo em vista essa mudança de tratamento, foram incluídos nesse estudo pacientes submetidos aos dois tipos de tratamento.

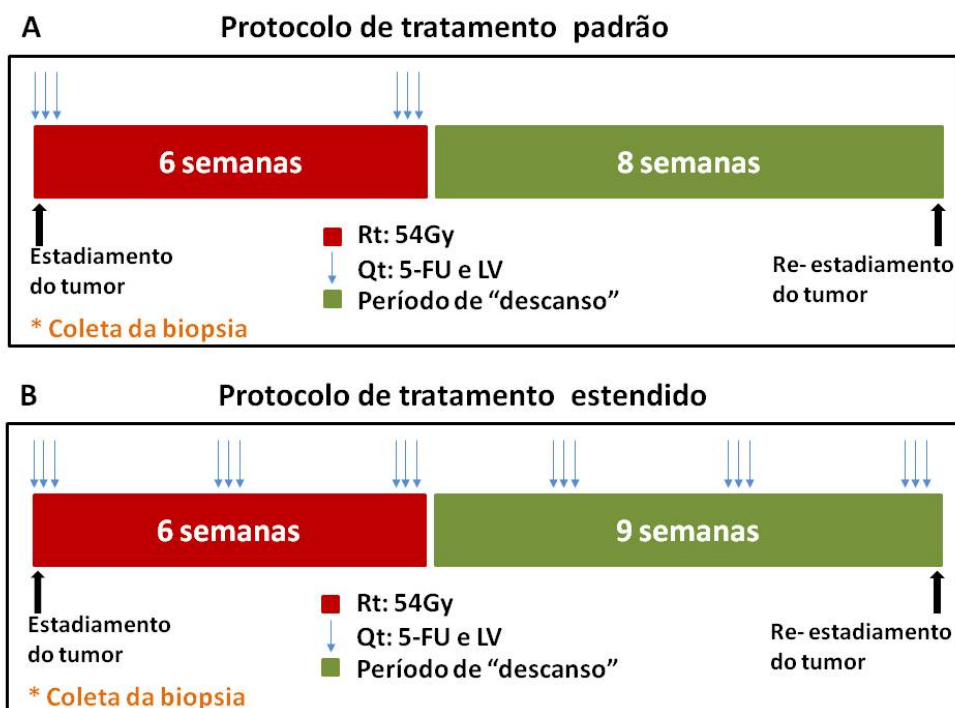


Figura 8 - Esquema do tratamento QRT neoadjuvante de pacientes com câncer de reto avançado. Antes de iniciar o tratamento é feito o estadiamento inicial do tumor e a coleta da biopsia. Durante 6 semanas os pacientes recebem doses diárias de radioterapia (Rt). Três ciclos de quimioterapia (Qt), são administrados no início e no final da Rt no protocolo de tratamento padrão (A). Já para o protocolo de tratamento estendido, são administrados um número maior de ciclos de Qt, inclusive durante o período de "descanso" (B). Ao final do tratamento, o tumor é novamente estadiado.

Após a conclusão do tratamento QRTn, a avaliação da resposta tumoral foi feita por um cirurgião colorretal experiente. A avaliação consistiu de exames clínicos, radiológicos e endoscópicos, como exame de toque retal, níveis de CEA no sangue, proctoscopia rígida, colonoscopia (quando possível), ressonância magnética pélvica, tomografia de abdômen e tórax. Durante o exame clínico, pacientes com evidência de doença residual como qualquer ulceração retal, irregularidade, massa ou estenose foram encaminhados para a cirurgia radical imediata, que consistiu da excisão total

do mesorreto (TME). Da mesma forma, pacientes com evidência radiológica ou endoscópica de doença residual, como a presença de metástase linfonodal ou tumor primário residual também foram encaminhados para a cirurgia radical. Para os pacientes com evidência clínica, endoscópica e radiológica de resposta completa clínica (cRC) não foi recomendado cirurgia radical imediata. Estes pacientes foram incluídos em um rigoroso programa de acompanhamento, sendo a cirurgia radical realizada apenas em caso de recidiva do tumor (HABR-GAMA et al. 2004).

- **Grupos de pacientes analisados**

Após a avaliação da resposta tumoral ao tratamento QRTn, baseando-se em achados clínicos e patológicos, os pacientes foram separados em resposta completa ou resposta incompleta.

O grupo de resposta completa incluiu pacientes com evidência clínica de resposta completa (cRC) sustentada por pelo menos 24 meses de acompanhamento rigoroso e também incluiu pacientes com resposta patológica completa (pRC), que passaram por uma cirurgia radical. É importante notar que os pacientes pRC foram submetidos à cirurgia por suspeita de apresentarem doença residual após o tratamento mas que, após avaliação histopatológica do espécime cirúrgico, não foram encontradas células neoplásicas. Sendo assim, uma vez que os casos de cRC e pRC representam pacientes que responderam completamente ao tratamento QRTn, os mesmos foram incluídos em um único grupo de resposta completa, contendo 9 pacientes de cRC e 5 pacientes de pRC.

O grupo de resposta incompleta incluiu 25 pacientes com doença residual significativa, incluindo ypT2-4 ou ypN+ e >10% de células neoplásicas residuais, correspondendo aos graus de regressão do tumor (TRG) de 0 a 2 de acordo com a classificação de Dworak (DWORAK et al. 1997). Pacientes com resposta incompleta, mas que apresentaram uma regressão do tumor muito acentuada foram excluídos do estudo, são estes casos ypT1N0 ou $\leq 10\%$ de células neoplásicas residuais (TRG3). Desta maneira, foram estudados apenas tumores com respostas extremas ao tratamento (um grupo com regressão total do tumor e outro com baixa ou nenhuma regressão).

O último grupo de pacientes foi denominado "recidiva". Este grupo foi formado por 4 pacientes classificados inicialmente como cRC, no entanto eles apresentaram recidiva local da doença em menos de 18 meses e foram submetidos à cirurgia radical. Desta maneira, como estes pacientes não apresentaram cRC sustentada (pelo menos 24 meses sem doença), os mesmos não foram incluídos no grupo de resposta completa e foram analisados separadamente.

3.2.2 Coleta das amostras

A coleta das amostras foi realizada antes do início do tratamento QRTn através de biopsias endoscópicas do tumor primário. Os tecidos foram congelados imediatamente em nitrogênio líquido e, posteriormente, armazenados a -80°C . A presença de células neoplásicas foi confirmada por um médico patologista através de lâminas histológicas com coloração

padrão de hematoxilina e eosina. Foram utilizados apenas fragmentos compostos por pelo menos 80% de células neoplásicas. Caso necessário, foi feita a macrodissecção para a remoção de tecido normal presente na peça.

3.2.3 Construção e sequenciamento de bibliotecas de miRNA de amostras de pacientes

O tecido, congelado em nitrogênio líquido, foi fragmentado seguindo com a extração de RNA total utilizando o reagente Trizol (Invitrogen) como descrito no item 3.1.2.

As bibliotecas de miRNA foram construídas a partir de 10µg de RNA total seguindo o mesmo protocolo do item 3.1.6.

3.2.4 Análise computacional da expressão de miRNAs e associação com a resposta à QRTn

As análises de expressão diferencial de miRNAs foram realizadas com todas as amostras disponíveis inicialmente, as quais constituíram o grupo de pacientes de treinamento, composto por 7 pacientes de resposta completa e 20 de resposta incompleta. Em um segundo momento, foram analisados novos pacientes, os quais formaram o grupo de validação, composto por 7 pacientes com resposta completa e 5 pacientes com resposta incompleta. As análises de expressão diferencial dos dados de sequenciamento foram feitas da mesma forma nos dois momentos, conforme descrito abaixo.

Inicialmente, os dados de sequenciamento das bibliotecas de miRNA obtidos no SOLiD4 foram analisados com o auxílio do software *CLC Genomics Workbench 5.1* (CLC Bio), conforme descrito no item 3.1.11. Para cada amostras, o número de sequências de cada miRNA foi normalizado pelo número total de sequências mapeadas em miRNA por milhão (cpm). Os miRNAs de baixa expressão foram filtrados da análise, sendo mantidos apenas aqueles com mínimo de 20 cpm em pelo menos 7 amostras. Posteriormente, foi realizada a análise de expressão diferencial entre os grupos de pacientes utilizando-se o EdgeR (item 3.1.11). Foram considerados miRNAs diferencialmente expressos aqueles com $p < 0,01$ e $FDR < 0,05$.

Para avaliar se os miRNAs diferencialmente expressos eram capazes de prever a resposta ao tratamento QRTn, primeiramente foi avaliada a similaridade entre as amostras com base na expressão dos miRNAs identificados como diferencialmente expressos. Foi construído um gráfico de escala multidimensional para avaliar a variação entre as amostras, através da função MDS no EdgeR. Também foi feita a clusterização hierárquica e análise estatística de Bootstrap utilizando-se a função *heatmap* e *pvclust* no RStudio (versão 0.96.122). Para essas análises foram utilizados os dados já normalizados por cpm, e aplicada a distância euclidiana para medir a distância entre os pontos e a média para calcular a distância entre os grupos.

Para cada miRNA identificado como diferencialmente expresso, calculou-se a área sob a curva (AUC) ROC (do inglês *receiver operating*

characteristic) para estimar a acurácia de cada um desses miRNAs em prever a resposta a QRT. As curvas ROC foram geradas no programa GraphPad Prism (versão 4.03) utilizando os valores de expressão obtidos por sequenciamento e normalizados por cpm. A partir da curva ROC foi possível estabelecer o valor de expressão de corte que melhor separava os pacientes de acordo com a resposta ao tratamento, a fim de maximizar o número de amostras classificadas corretamente. Por fim, foram calculados os valores de sensibilidade e especificidade com base no valor de corte estabelecido.

3.2.5 Avaliação por qPCR da expressão de miRNAs

Os miRNAs identificados como diferencialmente expressos através do sequenciamento como metodologia, também tiveram sua expressão avaliada por qPCR, utilizando o sistema miScript PCR (Qiagen), conforme descrito anteriormente (item 3.1.12).

A diferença de expressão dos miRNAs entre os dois grupos de pacientes foi determinada através da quantificação relativa da expressão, utilizando o método comparativo $\Delta\Delta Ct$ (LIVAK e SCHMITTGEN 2001). A reação foi feita em triplicata, sendo que para cada amostra, a média do Ct do miRNA de interesse foi normalizada pela média do Ct do gene normalizador RNU6b. Posteriormente, foi feita a comparação dos valores normalizados para cada amostra em relação à referência (mediana do grupo de menor expressão). Para determinar se havia uma diferença estatística significativa da expressão dos miRNAs entre os dois grupos de pacientes, foi realizado o

teste de Mann-Whitney utilizando o programa GraphPad Prism (versão 4.03).

Os dados de expressão obtidos por sequenciamento foram comparados aos dados obtidos por qPCR. Para tanto, foi feita uma análise de correlação entre os valores negativos de ΔCt (Ct do miRNA - Ct gene normalizador) obtidos por qPCR e os valores de expressão normalizados por cpm e em escala \log_2 obtidos por sequenciamento. Foi calculada a correlação de Spearman para cada um dos miRNAs utilizando o programa GraphPad Prism (versão 4.03).

3.2.6 Análise da expressão de genes alvo de miR-21-5p

Inicialmente, utilizando-se o Target Scan foram identificados todos os alvos conservados de miR-21-5p com alto valor de predição (total context+ score mínimo de -0,3) (LEWIS et al. 2005). O laboratório possui dados de expressão de transcriptoma total para 19 dos pacientes incluídos nesse estudo (6 de resposta completa e 13 de resposta incompleta), os dados foram obtidos por sequenciamento de última geração e ainda não foram publicados. Sendo assim, foi feita uma busca por correlação inversa da expressão de miR-21-5p e a expressão dos seus genes alvo, através da análise de correlação de Spearman (GraphPad Prism versão 4.03). Para os genes alvos que apresentaram correlação significativa, também foi avaliado se havia uma diferença de expressão entre os grupos de resposta completa e incompleta, através do teste de Mann-Whitney (GraphPad Prism, versão 4.03).

A expressão obtida por sequenciamento do gene SATB1, um dos genes alvo de miRNA-21-5p, foi validada por qPCR. Foram analisadas amostras de 8 pacientes de resposta incompleta e 3 de resposta completa. A reação seguiu o protocolo do item 3.5.4, usando 400nM de cada primer (**Quadro 5**). A diferença de expressão do SATB1 entre os pacientes de resposta completa e incompleta foi avaliada através da quantificação relativa da expressão, utilizando o método comparativo $\Delta\Delta Ct$ (LIVAK & SCHMITTGEN 2001). A reação foi feita em triplicata, sendo que para cada amostra, a média do Ct do SATB1 foi normalizada pela média do Ct dos genes normalizadores HMBS e PUM1. Posteriormente, foi feita a comparação dos valores normalizados para cada amostra em relação à uma amostra utilizada como referência (RNA extraído de tecido de cólon normal obtido comercialmente da Clontech). Para determinar se havia uma diferença estatística significativa da expressão de SATB1 entre os pacientes com os dois tipos de resposta, foi realizado o teste de Mann-Whitney utilizando o programa GraphPad Prism (versão 4.03).

Quadro 5 - Primers utilizados para análise de expressão relativa do gene SATB1 através de qPCR.

Gene	Primer
SATB1_Fw	5' GGTACAAACATTTCAAGAAGAC 3'
SATB1_Rv	5' CATGATTGGCGCCTTGCT 3'
HMBS_Fw	5' GGCAATGCGGCTGCAA 3'
HMBS_RV	5' GGGTACCCACGCGAATCAC 3'
PUM1_Fw	5' CCAGGCCAGCGGAAGAT 3'
PUM1_Rv	5' TGTACTTACGAAGAGTTGCGATGTG 3'

3.2.7 Sensibilidade celular à QRT

A sensibilidade das linhagens HCT116 e SW480 à Qt e Rt foi avaliada *in vitro*. Para tanto, foram plaqueadas 5×10^4 células por poço em placas de 24 poços. Após 24h, as células foram lavadas com PBS e adicionou-se 1ml de meio de cultura acrescido do quimioterápico em diferentes concentrações ou adicionou-se 1ml de meio de cultura e as células foram submetidas a radiação. As células foram contadas 24, 48 e 72h após o tratamento com o auxílio do Countess Automated Cell Counter (Invitrogen). Para cada condição foram contadas tanto as células mortas em suspensão quanto as células viáveis aderidas, sendo que através da adição de Trypan Blue foi possível identificar as células mortas (coradas de azul) e as células viáveis (não coradas). A partir da contagem do número total de células viáveis para cada tempo e condição analisada foi possível construir um gráfico para avaliar a taxa de proliferação celular. Também foi possível construir um gráfico de viabilidade celular a partir da porcentagem de células viáveis para cada condição. Todas as análises foram realizadas em triplicata.

Para os ensaios de sensibilidade à Qt foram utilizados 5-FU (Faudfluor, Libbs) e LV (Faudleuco, Libbs) em diferentes concentrações: 5, 50 e 500 μ M de 5-FU concomitante a 0,25; 2,5 e 25 μ M de LV, respectivamente. Para os ensaios de Rt foram utilizadas 3 dosagens de radiação: 2, 4 e 8 Gy através do equipamento GammaCell 3000 Elan do Departamento de Radioterapia do Centro de Oncologia do Hospital Sírio-Libanês.

3.2.8 Modulação da expressão de miR-21-5p

A expressão de miR-21-5p foi modulada *in vitro* através da transfecção das linhagens celulares com oligos miméticos de miR-21-5p ou anti-miR-21-5p. A linhagem SW480 expressa baixos níveis de miR-21-5p e, portanto, foi realizada a transfecção com hsa-miR-21-5p mirVana miRNA mimic ou com a molécula controle, mirVana miRNA Mimic Negative Control (Invitrogen). Foram plaqueadas 5×10^4 células por poço em placas de 24 poços, com um volume final de meio de cultura de 900 μ l. Após 24h, as células foram transfectadas seguindo-se a recomendação do fabricante. A molécula mimética foi diluída em 50 μ l (10nM final) de meio de cultura sem antibiótico e também foi feita a diluição de 3 μ l do reagente de transfecção lipofectamine RNAiMAX (Invitrogen) em 47 μ l de meio de cultura sem antibiótico. Em seguida, as duas soluções foram misturadas e incubadas por 20min. Ao final, 100 μ l dessa solução foram adicionados por poço (com volume final de 1ml). A análise de expressão de miR-21-5p foi realizada 24, 48, 72 e 120h após a transfecção. Cada condição foi realizada em triplicata. Após o período determinado as células foram lavadas com PBS e foram adicionados 300 μ l de Trizol (Invitrogen). As triplicatas correspondentes foram combinadas para a extração de RNA e realização da qPCR (item 3.1.12). No total foram realizados 2 experimentos independentes. Por último, foi avaliada a expressão relativa de miR-21-5p em cada uma das condições (item 3.1.12) tendo como referência as células transfectadas com o molécula controle.

A linhagem HCT116 expressa altos níveis de miR-21-5p e, portanto, foi transfectada com molécula anti-miR-21-5p, mirVana™ miRNA inhibitor, ou com a molécula controle, mirVana miRNA Inhibitor Negative Control

(Invitrogen). Seguiu-se com o mesmo protocolo de transfecção, com exceção do uso de uma maior concentração da molécula de anti-miR-21-5p e do seu controle (30nM final). Da mesma forma, foi avaliada a expressão relativa de miR-21-5p em cada uma das condições.

4 RESULTADOS E DISCUSSÃO

4.1 PARTE I: IDENTIFICAÇÃO E ESTUDO DE EXPRESSÃO DE VARIANTES DE POLIADENILAÇÃO E DE MIRNAS

4.1.1 Construção de bibliotecas enriquecidas para extremidade 3' dos transcritos e sequenciamento no 454

O estudo de APA foi feito em duas linhagens derivadas a partir de tumor colorretal, HCT116 e SW480. Estas duas linhagens foram escolhidas para favorecer a identificação de variantes de um maior número de genes pois elas se diferem quanto à presença de instabilidade de microssatélite (MIS) e um estudo com amostras de tumor de cólon revelou que o perfil de expressão de miRNAs é capaz de discriminar as amostras quanto à positividade de MIS (SCHEPELER et al. 2008; SARVER et al. 2009). Isto nos levou a especular que estas linhagens celulares possuíssem não apenas padrões distintos de expressão de miRNAs, mas também de variantes de poliA.

Além disso, o estudo de APA foi feito através do sequenciamento em larga escala de bibliotecas de cDNA construídas através de protocolos desenvolvidos neste trabalho.

Inicialmente, foi desenvolvido um protocolo inédito de construção de bibliotecas de cDNA que tivessem enriquecimento da extremidade 3' dos transcritos e que fossem compatíveis com o sequenciador em larga escala

454 (Roche). O protocolo de construção das bibliotecas e todos os controles utilizados durante sua construção estão detalhados no material e métodos (**Figura 4**). Resumidamente, foi obtido RNA de alta qualidade, livre de RNA ribossômico e enriquecido para RNA contendo cauda de poliA. Realizou-se a transcrição reversa utilizando um primer oligodT com sítio de restrição para a enzima BsmFI e biotina conjugada. Em seguida, o cDNA fita dupla foi digerido com as enzimas FatI ou MboI, que são frequentemente encontradas no genoma, e a porção 3' dos fragmentos foi recuperada através da incubação com *beads* magnéticas conjugadas à estreptavidina. Por último, o cDNA foi digerido com BsmFI para liberar as moléculas das *beads* e remover grande parte do trecho de poliA/poliT restando apenas 3 As ou Ts devido à maneira como o primer foi desenhado.

Inicialmente foram feitas algumas análises de bioinformática para avaliar se os fragmentos gerados a partir da digestão com as enzimas propostas representariam a maior parte dos transcritos presentes no banco de dados do RefSeq e com tamanho compatível com o sequenciador 454. Considerando todos os genes codificadores humanos definidos por sequências RefSeq, são 18.457 genes e destes 18.285 (98,9%) têm pelo menos um sítio para FatI e 18.014 (97,5%) têm pelo menos um sítio para MboI. Somando os dois conjuntos, existem 18.419 (99,7%) genes com pelo menos um sítio para FatI ou MboI. Portanto, apenas 56 genes (0,3%) não possuem sítios para pelo menos uma das duas enzimas utilizadas neste trabalho e que talvez pudessem não ser sequenciados ou simplesmente não seriam enriquecidos para sua porção 3'.

Também foi avaliada a distância média esperada entre o sítio de restrição mais 3' de cada uma das enzimas e a cauda poliA dos genes do RefSeq. A distância média do sítio da enzima FatI ao início da cauda poliA é 251nt e para a enzima MboI é 341nt. Considerando que o tamanho médio das sequências geradas no 454 é de 400nt e que o tamanho médio entre o sítio de restrição mais 3' e a cauda poliA é menor do que 400nt seria possível analisar a maior parte dos eventos de poliA. Em uma análise mais detalhada, foram encontrados 1.796 (9,7%) genes que apresentam uma distância entre o sítio de restrição (de FatI ou MboI) mais 3' e a cauda poliA maior do que 400nt e que, portanto, suas variantes provavelmente não estariam representadas por sequências únicas cobrindo toda sua extensão. Mas isto não impede que sequências obtidas para os mesmos sejam utilizadas para a identificação de variantes de poliA caso elas apresentem os nucleotídeos AAA ou TTT em uma das extremidades.

O sequenciamento no 454 das duas bibliotecas enriquecidas para a extremidade 3' gerou mais de 1,2 milhões de sequências para cada linhagem (**Tabela 1**). As sequências obtidas foram mapeadas na sequência genômica humana referência (hg19), sendo mapeadas 1.135.000 (89,45%) e 1.161.366 (90,35%) para HCT116 e SW480, respectivamente. Essas sequências corresponderam a mais de 15 mil genes do banco do RefSeq. Além disso, confirmando o enriquecimento para a porção 3' dos transcritos nestas bibliotecas, observou-se sobreposição (de no mínimo 10nt) com o exon mais 3' de um determinado gene para 911.222 (80,28%) das sequências mapeadas de HCT116 e para 920.118 (79,22%) da SW480.

Tabela 1 - Dados obtidos do sequenciamento no 454 das bibliotecas de cDNA enriquecidas para porção 3'.

	HCT116	SW480
Total de sequências geradas	1.268.859	1.285.317
Tamanho médio das sequências (pb)	111	120
Sequências mapeadas	1.135.000 (89,45%)	1.161.366 (90,35%)
Sequências sobrepondo ao exon mais 3'	911.222 (80,28%)	920.118 (79,22%)
Total de genes expressos	15.127	15.823

Dado que muitos genes humanos não possuem introns ou a maior parte de sua sequência encontra-se no exon mais 3', foi feita outra análise para a confirmação do enriquecimento das sequências para a porção 3'. Nesta análise os genes foram divididos em 4 partes iguais de acordo com o comprimento do mRNA maduro (por exemplo, um gene cujo transcrito apresentasse 2000pb foi dividido em 4 partes de 500pb cada) e avaliou-se em qual das partes as sequências geradas foram mapeadas. Confirmando os resultados obtidos anteriormente, foi verificado que 99,8% das sequências correspondiam à porção (um quarto) mais 3' dos genes (**Tabela 2**).

Tabela 2 - Alinhamento das sequências obtidas em relação à porção do gene correspondente. Os genes foram divididos em quatro partes: do início 5' até 25% do seu tamanho total; de 25 a 50%; de 50 a 75% e de 75 a 100%.

Amostra	Porção do transcrito			
	0-25% (extremidade 5')	25-50%	50-75%	75-100% (extremidade 3')
HCT116	0,0%	0,10%	0,10%	99,8%
SW480	0,0%	0,10%	0,10%	99,8%

As sequências geradas a partir das bibliotecas enriquecidas apresentaram um tamanho médio relativamente curto. De fato, o tamanho das sequências obtidas (111pb para HCT116 e 120pb para SW480) foi bem menor do que o esperado (considerando-se a distância média entre os sítios das enzimas de restrição e a extremidade 3' dos genes). As sequências geradas parecem estar truncadas e as evidências estão descritas adiante.

Devido ao protocolo estabelecido para a construção dessas bibliotecas esperava-se encontrar o sítio de restrição de uma das enzimas utilizadas na digestão do cDNA (MboI ou FatI) em uma das extremidades das sequências geradas. Na outra extremidade esperava-se encontrar um trímero de A ou T, pois na extremidade 3' dos transcritos foi preservada a presença de três nucleotídeos A ou T, restantes da cauda poliA e do oligodT (**Figura 9**). Neste caso, a extremidade da sequência que apresentasse “AAA” ou “TTT” representaria a extremidade 3' do transcrito, indicando a posição exata do sítio de poliA. A metodologia de sequenciamento do 454 não é sentido direcionado e, portanto, parte das sequências são geradas a partir do 5' do transcrito (sítio de restrição FatI ou MboI) e a outra parte das sequências são geradas a partir do 3' do transcrito (trímero de A ou T).

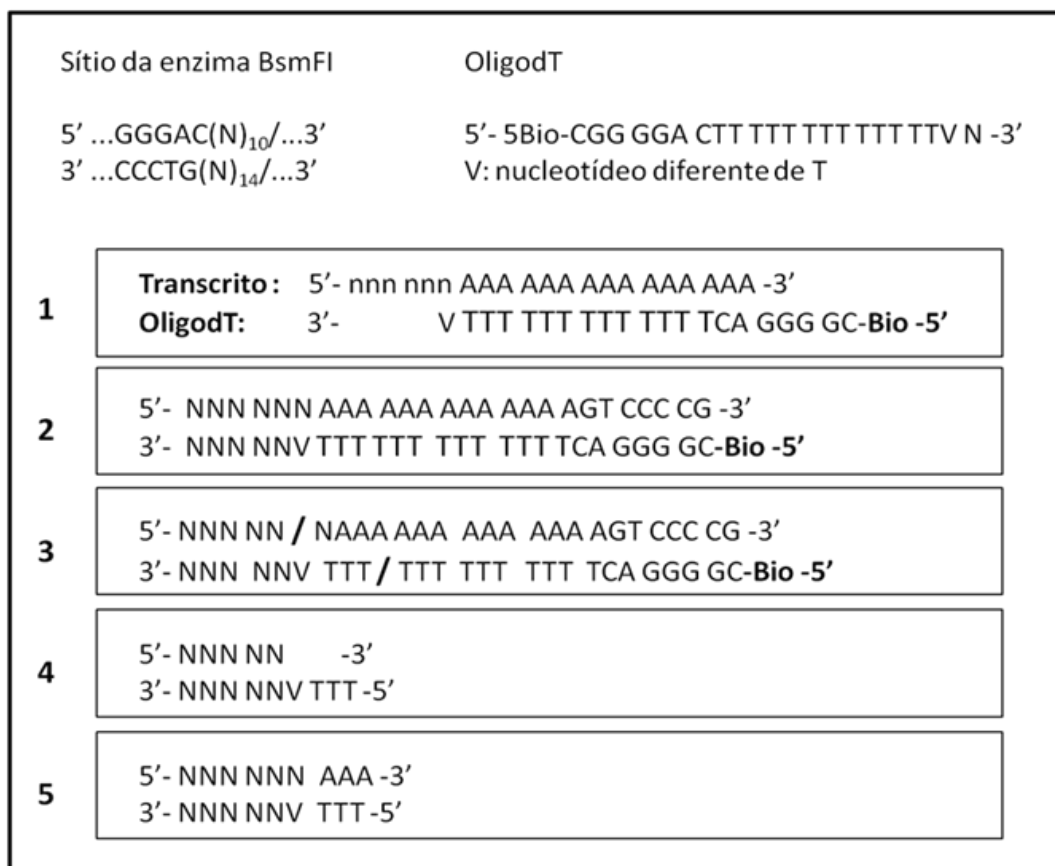


Figura 9 - Ilustração do desenho do oligodT e do que ocorre na extremidade 3' dos transcritos durante as diferentes etapas da construção da biblioteca de cDNA enriquecido. 1- Alinhamento durante a síntese da primeira fita do cDNA. Na sequência do oligo dT também foi inserido o sítio de restrição da enzima BsmFI (CGGGAC) e biotina (Bio). 2- Síntese do cDNA dupla fita; 3- Digestão com BsmFI (posição do corte representada com “ / ”). 4- cDNA de interesse é liberado da *bead*. 5- Reparo da extremidade 3'.

Foi investigada a presença dos sítios de restrição de *FatI* e *MboI* (CATG ou GATC) e dos trimeros de A ou T nas extremidades das sequências geradas (**Tabela 3**). Um dos sítios de restrição foi identificado em 40,18% das sequências geradas para HCT116 e em 36,86% para SW480. Por outro lado, apenas 1,5% das sequências continham o trímico de A ou T, sendo que menos de 1% apresentavam o sítio de restrição em

uma extremidade e o trímero de A ou T na outra. Como descrito anteriormente, esperava-se que a maior parte das sequências geradas apresentassem o sítio de restrição e trímero de A ou T. Isto só não deveria acontecer para os casos em que o fragmento gerado após digestão fosse maior do que o tamanho de sequência que pode ser gerado no 454 (400pb). Assim, embora as sequências geradas correspondessem à porção mais 3' dos transcritos, a maior parte delas estava truncada na terminação 3', não sendo possível identificar o sítio de poliA. Outra observação que confirma a falta da extremidade mais 3' dos transcritos é que apenas 10% das sequências geradas apresentavam os sinais canônicos de poliA (AATAAA e ATATAA – como as sequências vêm de cDNA as uracilas foram substituídas por timinas).

Tabela 3 - Identificação das sequências geradas no 454 que possuem o sítio de restrição enzimática (Fatl ou Mbol), trímero de A ou T e pelo menos um sinal canônico de poliA (AATAAA e ATATAA).

Amostra	Sítio de restrição	Trímero de A ou T	Sítio de restrição e trímero de A ou T	Sinal poliA
HCT116	509.894 (40,18%)	18.288 (1,45%)	8.019 (0,63%)	133.773 (10,54%)
SW480	473.749 (36,86%)	20.748 (1,61%)	8.203 (0,64%)	134.995 (10,50%)

Não foi possível identificar em qual etapa do protocolo de construção das bibliotecas tal artefato poderia ter sido gerado. A amplificação com baixos ciclos que foi feita para o aumento da massa das bibliotecas não deve ter causado tal problema, uma vez que os primers utilizados na reação são específicos para os adaptadores, externos às sequências de interesse.

O fato de não ser identificado os trimeros de A ou T na extremidade das sequências, evidentemente limitou bastante os resultados dessa parte do trabalho, uma vez que a presença dos mesmos seria um forte indício do sítio de poliA de cada um dos transcritos identificados. Ainda assim, conduzimos a busca por variantes de poliA com os dados disponíveis.

4.1.2 Análise computacional do sequenciamento no 454 e identificação de variantes de poliadenilação

Foram consideradas variantes de poliA as sequências cujo final 3' distanciou no mínimo 30nt downstream ou upstream do sítio de poliA de todas as sequências RefSeq do gene correspondente. Para as variantes de poliA mais curtas foram consideradas apenas as sequências que apresentavam o sítio de poliA (identificado pelo trímero de A ou T) e o sinal canônico de poliA (AATAAA ou ATTAAA). Este cuidado foi tomado, pois especialmente a ausência dos trimeros é um forte indício de que o final da sequência gerada não corresponde à extremidade mais 3' do transcrito analisado. Com isso foi minimizada a seleção de candidatos falso positivos. Cabe notar que para aumentar a confiança das variantes identificadas foram considerados apenas os 2 sinais canônicos de poliA, no entanto existem pelo menos mais 10 sinais alternativos e se esses outros sinais tivessem sido incorporados à análise certamente seria identificado um maior número de variantes (BEAUDOING et al. 2000). Por último, foi incluído um filtro para eliminação de falso positivos causados por *internal priming* do oligodT. Neste caso, foram desconsideradas as sequências cuja extremidade 3' alinhou em

uma região genômica rica em adeninas. Para a identificação de variantes mais longas que as sequências representadas pelo RefSeq foram utilizadas sequências que apresentavam pelo menos o sinal de poliA. Nesses casos a ausência do trímero de A ou T não implica necessariamente em um falso positivo. A ausência dos mesmos significa que não foi possível identificar o sítio de poliA, mas a variante não deixa de ser potencialmente verdadeira. Na linhagem HCT116 foram identificadas 994 variantes de poliA mais curtas e 250 variantes mais longas enquanto na SW480 foram identificadas 1.113 variantes mais curtas e 247 variantes mais longas (**Tabela 4**). Para HCT116 22 genes apresentaram tanto variantes mais curtas quanto mais longas, sendo que no total 1.222 genes apresentaram variantes de poliA. Na SW480 23 genes apresentaram variantes curtas e longas, sendo no total 1.337 genes com variantes de poliA. Mais da metade dos genes que apresentaram variantes de poliA foram identificados exclusivamente em apenas uma das linhagens (**Figura 10**).

Tabela 4 - Número de genes apresentando variantes de poliA mais curta ou mais longa do que as sequências RefSeq, dados obtidos no 454.

Amostra	Variante mais curta	Variante mais longa	Total*
HCT116	994	250	1.222
SW480	1.113	247	1.337

*A soma dos genes com variantes não é exata pois alguns genes possuem variantes curtas e longas.

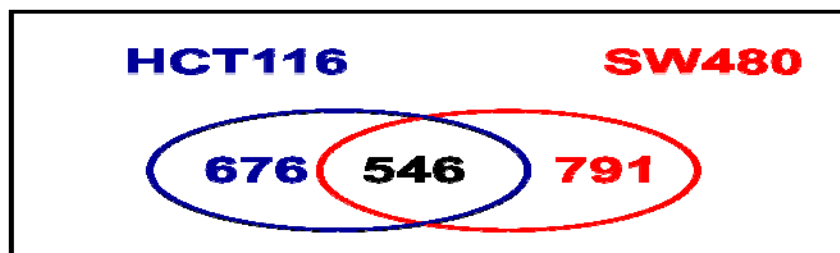


Figura 10 - Comparação do número de genes com APA identificados apenas em uma linhagem celular ou presentes em ambas.

Embora sejam dados um pouco limitados, chama atenção o número bem maior de variantes mais curtas identificadas em ambas as linhagens. Esses achados corroboram com dados da literatura que demonstraram que durante o processo tumorigênico há um aumento da expressão de transcritos com 3'UTRs mais curtas, potencialmente de oncogenes (MAYR e BARTEL 2009). Além disso, já foi descrito que células com maior taxa proliferativa apresentam uma maior expressão de variantes com 3' UTRs mais curtas (SANDBERG 2008). Estes transcritos estariam menos susceptíveis aos efeitos regulatórios de longas 3'UTR e assim a célula poderia aumentar a produção da proteína de interesse sem necessariamente aumentar o nível de expressão gênica.

Como dito anteriormente, um dos critérios adotados para a seleção das variantes foi a extremidade 3' das sequências geradas mapear no genoma humano a no mínimo 30pb da posição de mapeamento da extremidade 3' do gene RefSeq correspondente. No entanto, quanto maior esta distância, maiores as chances das variantes de poliA de um mesmo gene se comportarem de maneira diferente, desde a sua estabilidade, presença de sítios alvo de miRNAs na 3'UTR e até mesmo à sua região

codificadora. Desta maneira, foi calculada a distância em pares de base entre a extremidade 3' das variantes identificadas e a extremidade 3' dos genes RefSeq correspondentes (**Figura 11**). Em ambas as linhagens, a maior parte das sequências que foram consideradas para a identificação das variantes, sejam mais curtas ou mais longas, apresentaram uma distância maior que 200pb entre sua extremidade 3' e a do gene RefSeq, o que aumenta a probabilidade de serem de fato variantes de poliA e de apresentarem diferentes sequências regulatórias na 3' UTR. Outros trabalhos também verificaram uma grande distância entre o final da 3' UTR da referência e da variante. Spies e cols reportou um distância média de 1.209nt e mediana de 775 (SPIES et al. 2013).

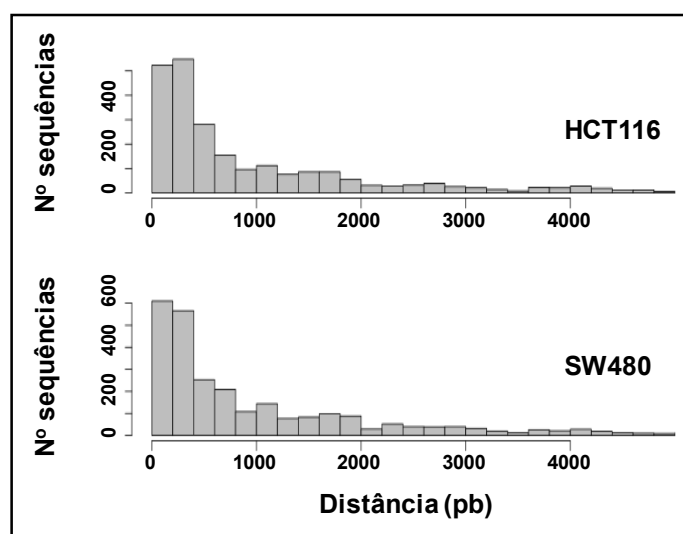


Figura 11 - Distância, em pares de bases (pb), entre o final 3' das variantes de poliA identificadas e o final 3' dos genes RefSeq.

4.1.3 Validação por RT-PCR das variantes de poliadenilação

Treze variantes de poliA identificadas nos dados de sequenciamento foram avaliadas por RT-PCR. Para tanto, foram selecionados genes que apresentassem pelo menos 2 sequências representando a variante identificada (**Quadro**).

Quadro 6 - Genes com APA selecionados para validação por RT-PCR.

Gene	Variante identificada	Linhagem em que foi detectada
ABL2	Curta	HCT116
ANKRD28	Curta	HCT116 e SW480
ANXA2	Curta	HCT116 e SW480
ASL	Curta	HCT116 e SW480
NQO1	Curta	HCT116 e SW480
POLR2J	Curta	HCT116 e SW480
PSMD8	Curta	HCT116 e SW480
ZNF697	Curta	SW480
DUS2L	Longa	HCT116
ERO1L	Longa	HCT116 e SW480
MARVELD2	Longa	HCT116
RANBP1	Longa	SW480
ZNF207	Longa	HCT116 e SW480

Os primers para validação das variantes por RT-PCR foram desenhados contendo 5nt complementares à extremidade 3' da variante a ser validada, seguidos de uma cauda poliT. Além disso, para cada variante foram desenhados primers que funcionaram como controle negativo, os quais continham 5nt complementares a uma região central da 3'UTR seguidos da cauda de poliT. Inicialmente foi realizada a RT-PCR com os primers dos controles negativos para verificar a especificidade dos primers

reversos desenhados da maneira proposta. Para 2 genes, ao contrário do esperado, houve amplificação de um fragmento. Os produtos amplificados foram sequenciados através da metodologia de Sanger. Para um dos genes verificou-se que a amplificação foi inespecífica. No entanto, para o outro foi específica. Sendo assim, com base no resultado desse controle negativo, para as validações das variantes mais curtas não é possível afirmar se o produto amplificado é realmente resultado do anelamento do primer com a extremidade 3' juntamente com a cauda de poliA ou com uma região interna do gene. Ainda assim foram realizadas as reações de RT-PCR e para todos os 8 candidatos de variantes curtas testados foram obtidas amplificação dos fragmentos no tamanho esperado (**Figura 12**). Os produtos amplificados foram sequenciados e todos se referiam ao gene específico em questão. Embora este método desenvolvido para validação não tenha sido capaz de confirmar com segurança a existência das variantes de poliA mais curtas, essas 8 variantes também foram identificadas por Li e cols (LIN et al. 2012). Esses autores realizaram uma extensa caracterização de variantes de poliA em 5 linhagens celulares de tecido normal e tumoral (mama, cólon, pulmão, fígado, rim) e desenvolveram um banco com os dados gerados que pode ser visualizado no UCSC genome browser (*Expression and Polyadenylation Database*).

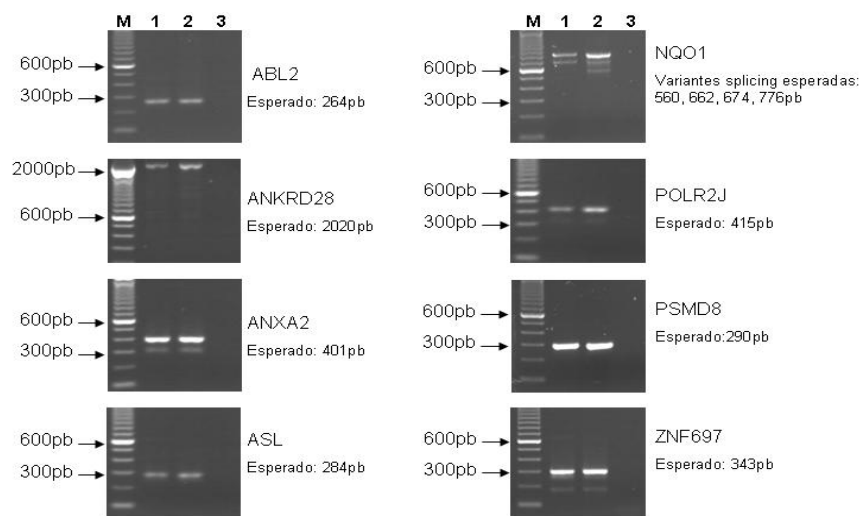


Figura 12 - Validação por RT-PCR das variantes curtas de poliA identificadas no 454. Todas as variantes apresentaram amplificação no tamanho esperado, sendo até mesmo identificadas variantes de splicing alternativo para o gene *NQO1*. M = marcador de peso molecular 100pb; 1= HCT116; 2= SW480; 3= controle negativo (sem cDNA).

Já para as variantes mais longas, foi possível a validação por RT-PCR, pois o primer reverso possui ao redor de 20nt complementares à região que difere a variante longa da variante representada pelo RefSeq. As 5 candidatas mais longas foram validadas por RT-PCR e o sequenciamento confirmou a especificidade dos produtos amplificados (**Figura 13**). Cabe ressaltar que, excetuando-se para o gene *RANBP1*, as variantes longas confirmadas ainda não possuem seu sítio de poliA definido. Por fim, essas variantes também foram comparadas com o banco de poliA mencionado acima (*Expression and Polyadenylation Database*). Para 4 genes avaliados (*DUS2L*, *ERO1L*, *MARVELD2*, *ZNF207*) também já foram identificadas variantes de poliA mais longas. Entretanto, não é possível afirmar com

certeza se são as mesmas variantes identificadas neste trabalho, uma vez que aqui não foram identificados os respectivos sítios de poliA.

Nesse banco de poliA não está presente a variante longa do gene *RANBP1*, a qual pode representar uma variante de poliA ainda não descrita. Como pode ser verificado, apenas uma porcentagem muito pequena das prováveis variantes de poliA aqui identificadas foram avaliadas por outra metodologia para terem sua existência confirmada. Certamente, assim como a variante longa do gene *RANBP1*, outras que não foram testadas por RT-PCR podem ser variantes até então não conhecidas. Cabe notar, que ainda será necessária uma extensa comparação com outros trabalhos já publicados para afirmar quais e quantas variantes realmente foram identificadas pela primeira vez.

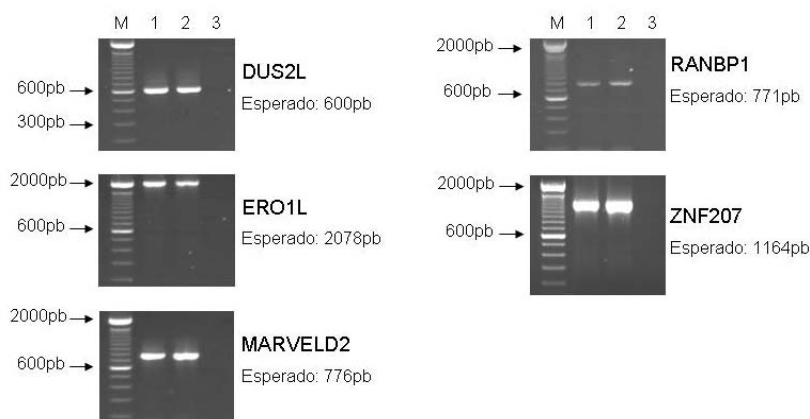


Figura 13 - Validação por RT-PCR das variantes longas de poliA identificadas no 454. Todas as variantes apresentaram amplificação no tamanho esperado. M = marcador de peso molecular 100pb; 1= HCT116; 2= SW480; 3= controle negativo (sem cDNA).

Embora algumas das variantes tenham sido identificadas originalmente em apenas uma das linhagens (de acordo com as sequências do 454), na validação por RT-PCR todas as variantes foram detectadas nas duas linhagens. O relativo pequeno número de sequências utilizadas para a identificação das variantes comprometeu a identificação de um número maior de variantes e, ainda, impossibilitou inferir a especificidade das variantes a qualquer uma das linhagens, bem como estimar níveis de expressão. Além disso, a ideia de identificação de variantes de poliA em larga escala ficou comprometida devido ao problema mencionado anteriormente.

A vantagem de usar o 454 em relação aos outros sequenciadores de larga escala (como SOLiD da Life Technologies e o HiSeq da Illumina) é a capacidade deste equipamento em gerar sequências maiores, em média de 400nt. Considerando o tamanho médio dos fragmentos obtidos durante a construção da biblioteca enriquecida para porção 3', o sequenciamento no 454 geraria a sequência completa para a maioria dos fragmentos, com uma cobertura razoável e tamanho suficiente para identificar o gene correspondente ao transcrito e o sítio de poliA, facilitando principalmente a identificação de variantes longas. Infelizmente o protocolo proposto não foi eficiente e, portanto, foi desenvolvido um outro protocolo, desta vez compatível com o sequenciador SOLiD.

4.1.4 Construção de bibliotecas enriquecidas para a extremidade 3' dos transcritos e sequenciamento no SOLiD

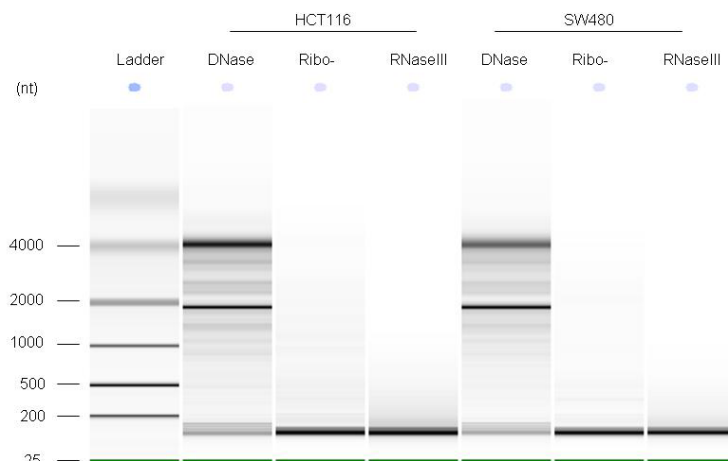
Após a obtenção dos resultados descritos acima, foi desenvolvido um novo protocolo de construção de bibliotecas enriquecidas para a porção 3', agora compatível com outro sequenciador, o SOLiD 5500 (Life Technologies). Embora essa tecnologia gere sequências menores do que o 454, o número de sequências geradas é bem maior, possibilitando a identificação do sítio de poliA pela densidade diferencial de sequências na região, além do sequenciamento de adeninas (proveniente da cauda poliA) ausentes no genoma. O protocolo original de construção de biblioteca de cDNA deste equipamento foi ligeiramente modificado, sem haver a introdução de qualquer etapa de amplificação ou digestão com enzima de restrição, além do que já é recomendado pelo fabricante, afim de evitar novos artefatos.

Inicialmente o RNA total foi tratado com DNase para eliminar qualquer contaminação com DNA genômico, o que poderia interferir na identificação de transcritos expressos na amostra. Em seguida, foi feita a depleção de rRNA e fragmentação com RNase III, capaz de clivar RNA fita dupla formado devido à estrutura secundária. O perfil do tamanho das moléculas de RNA obtidas após cada uma dessas etapas foi avaliado no Bioanalyzer (**Figura 14**). É possível visualizar a ausência das bandas referentes ao rRNA após a depleção (bandas do tamanho do 18S = ~2000nt; e 28S = ~4000nt). Além disso, a **Figura 14b** compara a distribuição de tamanho das moléculas de RNA, antes e após a fragmentação pela RNase

III. Utilizando-se o Bioanalyzer, foi calculada a porcentagem de fragmentos com tamanho dentro da faixa adequada para o sequenciamento (de 100 a 200nt). No RNA total, a porcentagem de fragmentos de RNA nesta faixa é muito baixa, correspondendo a 17% da massa total para HCT116 e 18% para SW480. Após a depleção do RNA ribossômico essa porcentagem passou para 56% na HCT116 e 67% na SW480. Finalmente, após a fragmentação, a porcentagem aumentou para 72% e 79%.

Uma vez que a maior parte dos fragmentos encontravam-se dentro da faixa de tamanho esperada, foram recuperados somente os fragmentos contendo a cauda poliA, enriquecendo assim as amostras para a extremidade 3' dos transcritos. Estes fragmentos poliA+ foram convertidos em cDNA e utilizados na construção das bibliotecas.

A



B

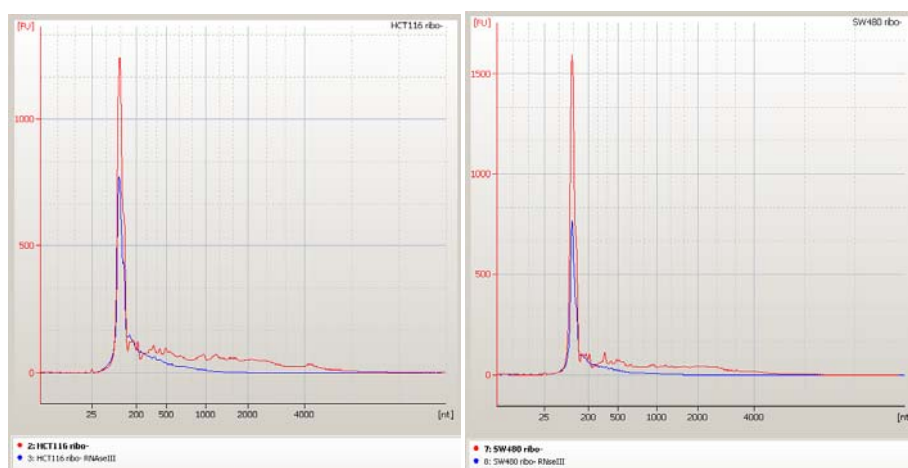


Figura 14 - Perfil do RNA no Bioanalyzer utilizado para construção da biblioteca de cDNA. A) Perfil obtido em uma corrida no chip RNA 6000 Pico do RNA total após tratamento com DNase, do RNA depletado de rRNA e fragmentado por RNase III para as duas linhagens celulares, HCT116 e SW480. B) A mesma corrida está representada em eletroferograma para comparar o perfil de tamanho do RNA depletado para rRNA antes da fragmentação (em vermelho) e após (em azul). À esquerda está o RNA da HCT116 e à direita o da SW480.

Dado o elevado custo do sequenciamento em larga escala e o insucesso do protocolo desenvolvido para o 454, inicialmente foi gerado um baixo número de seqüências e de tamanho reduzido (35pb) no SOLiD para avaliar a eficiência desse novo protocolo na identificação de variantes de

poliA. Foram geradas cerca de 5 milhões de sequências de 35pb para cada linhagem. Este número é bem maior do que o gerado no 454, mas dado o seu tamanho muito curto, a porcentagem de sequências mapeadas no genoma é bem menor.

Foram feitas análises bastante simples dos dados mas que revelaram a eficiência de enriquecimento do novo protocolo desenvolvido. Inicialmente, foi verificado se as sequências realmente mapearam no exon mais 3' dos genes. Na **Figura 15** estão alguns exemplos de sequências geradas e que foram mapeadas preferencialmente na porção 3' de diferentes genes conhecidos. Além disso, também foi analisado se haviam sequências geradas por este protocolo que confirmassem as variantes identificadas com o protocolo anterior (do sequenciador 454). Das 13 variantes identificadas nos dados de 454 e validadas por RT-PCR (8 curtas e 5 longas), as 8 variantes curtas estavam representadas por sequências geradas no SOLiD, assim como 4 das 5 variantes longas. Apenas a variante longa do gene ZNF207 não foi identificada no SOLiD. Na Figura 16 estão 2 exemplos desses casos demonstrando a grande capacidade desse novo protocolo em identificar variantes de poliA.

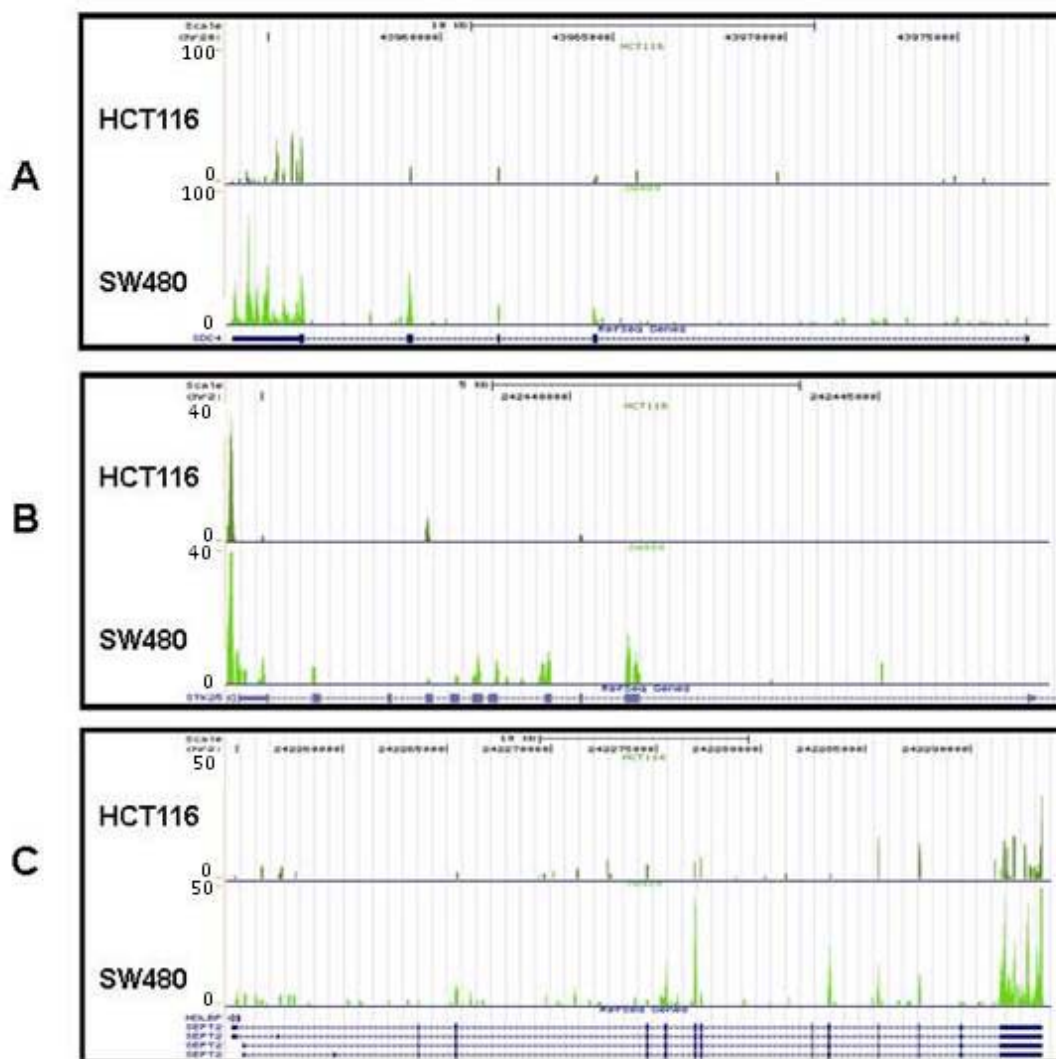


Figura 15 - Esquema ilustrando as sequências geradas no SOLiD enriquecidas para a extremidade 3' de alguns genes conhecidos. Os picos em verde representam as sequências geradas para cada linhagem, sendo sua altura proporcional à quantidade de sequências que mapearam naquela região. Em A e B o gene está na orientação invertida, 3' – 5'.

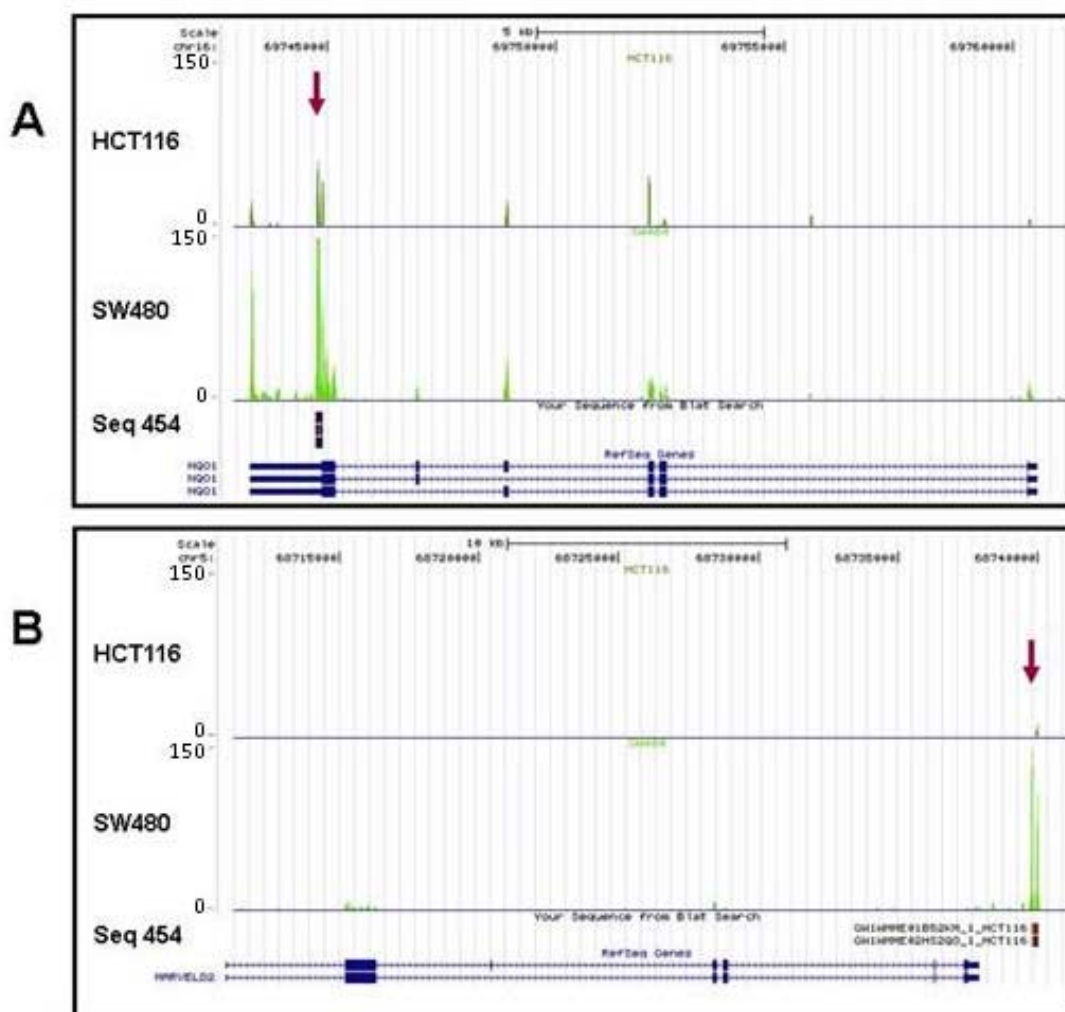


Figura 16 - Esquema ilustrando variantes identificadas no 454 e confirmadas no SOLiD. Estão representadas uma variante mais curtas (A) e uma mais longa (B). Além das seqüências geradas no SOLiD (em verde - pico proporcional à quantidade de seqüências que mapearam na região) também estão representadas as seqüências do 454 que foram utilizadas inicialmente para a identificação das variantes (em preto). As setas vermelhas indicam o sítio de poliA das variantes identificadas. A = NQO1 (na orientação invertida 3' – 5') e B= MARVELD2.

4.1.5 Análise computacional do sequenciamento no SOLiD e identificação de variantes de poliadenilação

As análises preliminares confirmaram o sucesso do novo protocolo no enriquecimento das bibliotecas para a porção 3' dos transcritos. Sendo assim, as bibliotecas enriquecidas das duas linhagens celulares foram sequenciadas no SOLiD 5500 em grande quantidade e com o maior comprimento possível. Foram geradas 88,9 milhões de sequências para HCT116 e 50,5 milhões para a SW480 (**Tabela 5**). As sequências obtidas foram mapeadas na sequência referência do genoma humano sendo que aproximadamente 27% e 57% das sequências de HCT116 e SW480, respectivamente, mapearam unicamente, com qualidade e comprimento médio de 70pb. As sequências mapeadas estavam enriquecidas para o exon mais 3', sendo 44% para HCT116 e 35% para a SW480. Este enriquecimento aumentou a capacidade de identificar a extremidade 3' dos transcritos e de fato, 10% dessas sequências possuíam pelo menos 3 adeninas na extremidade 3'. Não era esperado um enriquecimento perfeito já que os fragmentos recuperados para o sequenciamento tinham de 100-200nt e o fragmento sequenciado era de 70nt, sendo assim alguns fragmentos não eram sequenciados até o início da cauda de poliA. Além disso, uma vez que o enriquecimento dos transcritos foi feito através da incubação com uma matriz de oligodT, fragmentos ricos em adeninas (que não eram provenientes da cauda de poliA) também foram recuperados. Por fim, as sequências mapeadas corresponderam a mais de 15 mil genes do banco do RefSeq.

Tabela 5 - Dados de sequências do SOLiD 5500 obtidas a partir das bibliotecas de cDNA enriquecidas para a porção 3' dos transcritos e mapeadas contra o genoma humano referência.

	HCT116	SW480
Total de sequências geradas	88.901.086	50.471.171
Sequências mapeadas	23.924.681 (27%)	28.628.159 (57%)
Tamanho médio das sequências (nt)	69	68
Sequências mapeadas no exon mais 3'	10.409.970 (44%)	10.106.936 (35%)
Sequências com cauda de poliA	1.494.208 (14%)	989.969 (10%)
Total de genes expressos	15.317	15.723

Com o grande número de sequências geradas foi possível identificarmos um número bem maior de variantes de poliA do que foi identificado com as sequências geradas no 454. Entretanto, devido ao relativo tamanho curto das sequências do SOLiD, as análises focaram-se apenas na identificação de variantes com 3' UTR mais curtas do que as presentes nos genes do RefSeq. Isto porque sequências mapeadas próximas (e downstream) à extremidade 3' de um determinado gene, como se fossem parte de uma variante de poliA com 3' UTR mais longa, poderiam pertencer a um outro gene e serem erroneamente classificadas caso não houvesse uma sobreposição entre as mesmas e o gene do RefSeq.

Sendo assim, foram estabelecidos alguns critérios para a identificação de variantes de poliA mais curtas, conforme detalhado em material e métodos. O principal critério para a determinação de um sítio de poliA foi a presença de pelo menos 5 sequências que apresentassem no mínimo 3 adeninas na extremidade 3' do transcrito e que não fossem mapeadas no genoma. Foram identificados 3.946 e 4.359 genes nas linhagens HCT116 e SW480, respectivamente, que apresentavam mais de um sítio de poliA no último exon 3'. Destes genes 2.533 eram comuns às duas linhagens (**Figura**

17). Comparando-se com dados recentes da literatura, este número é semelhante ao que outros autores têm verificado em diferentes tipos celulares humanos. Os trabalhos têm demonstrado que cerca de 30 a 70% dos genes humanos apresentam variantes de poliA (DERTI et al. 2012; LIN et al. 2012). A análise foi feita apenas para a identificação de variantes mais curtas com sítio de poliA no último exon 3' do gene RefSeq e ainda assim foram identificadas variantes para 26% (3.946 de 15.317) dos genes na linhagem HCT116 e 28% (4.359 de 15.723) na linhagem SW480. Se também tivesse sido trabalhado com variantes mais longas ou mais curtas em introns e exons diferentes do 3' provavelmente teria sido identificado um maior número de genes com variantes expressas.

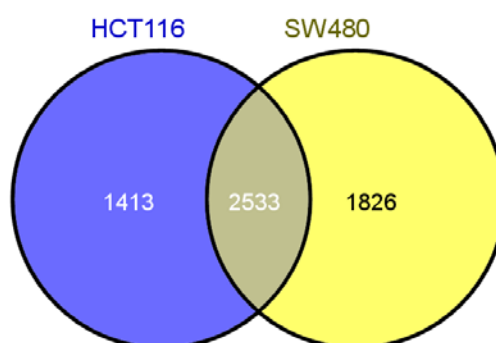


Figura 17 - Número de genes apresentando variantes de poliA mais curtas do que gene RefSeq com maior 3' UTR identificadas nas linhagens HCT116 e SW480.

A tecnologia de sequenciamento de última geração tem contribuído muito para os estudos de APA. Recentemente, dois trabalhos fizeram uma extensa caracterização dos sítios de poliA em humanos após o enriquecimento para o final 3' dos transcritos (os protocolos variaram um

pouco entre si e o que foi aqui desenvolvido). Em um primeiro trabalho, foram estudados 5 tecidos diferentes (testículo, fígado, rim, músculo e cérebro) em 5 mamíferos, o que aumentou a capacidade de encontrar variantes tecido específicas (DERTI et al. 2012). Foi verificado que 69% dos genes expressos nesses tecidos possuíam múltiplos sítios de poliA, sendo que 49% tinham 3 ou mais sítios. E ainda, a expressão de variantes específicas de poliA apresentava maior similaridade entre os tecidos de diferentes espécies do que entre diferentes tecidos da mesma espécie.

Em um outro estudo no qual foram analisadas amostras de tumores derivados de diferentes tecidos (mama, cólon, pulmão, fígado ou rim), o número de genes que apresentavam variantes de poliA foi semelhante entre as diferentes amostras (LIN et al. 2012). Além disso, nas amostras tumorais houve preferência pela maior expressão de variantes curtas quando comparadas às amostras normais. O número de genes apresentando variantes mais curtas foi de aproximadamente 4.000, correspondendo a cerca de 30% dos genes expressos naquelas amostras, bastante semelhante ao que foi encontrado neste trabalho.

Não foi possível comparar as variantes de poliA aqui identificadas com as variantes disponíveis nos bancos de dados públicos e, portanto, inviabilizando estimar quantas variantes novas foram identificadas. A ideia de incluir 2 linhagens celulares no trabalho foi para aumentar o número de variantes identificadas e de fato 36% das variantes identificadas na HCT116 são específicas dessa linhagem enquanto que 42% são específicas da SW480. Entretanto, não sabemos quantas dessas variantes ainda não foram

descritas em outros trabalhos. Outro dado interessante que não foi possível ser gerado foi a avaliação da presença de sinais de poliA nestas variantes, o que daria maior confiabilidade na identificação das mesmas.

Após o início deste trabalho, foram desenvolvidos diversos protocolos para sequenciar preferencialmente a extremidade 3' dos transcritos para o estudo de APA. Um desses protocolos consistiu na transcrição reversa utilizando oligo T₁₀VN (10 timinas seguidas de uma base diferente de T e de uma base randômica) conjugado com sequência adaptadora compatível com o sequenciamento no Illumina (DERTI et al. 2012). Em seguida, foi realizada a síntese da segunda fita de cDNA com primer randômico seguido de sequência 5' adaptadora. Assim, foi possível realizar uma PCR utilizando primers para as sequências adaptadoras inseridas durante a síntese de cDNA e realizar o sequenciamento direcionado para o sítio de poliA, já que o sequenciamento utiliza-se da sequência adaptadora inserida na extremidade 3' do transcrito.

Também já foram desenvolvidos outros protocolos baseados na seleção da extremidade 3' através da transcrição reversa utilizando primer de oligodT modificado (MANGONE et al. 2010; FU et al. 2011; SHEPARD et al. 2011; FOX-WALSH et al. 2011). Todos esses protocolos, inclusive o que foi desenvolvido neste trabalho, podem resultar na identificação de sítios de poliA falsos. A transcrição reversa a partir do oligodT pode introduzir artefatos decorrentes de *internal priming*, o anelamento do oligodT a regiões internas do mRNA, o que pode gerar artefatos e levar à identificação de falsos sítios de poliA. No nosso protocolo, após a fragmentação do RNA, os

fragmentos foram incubados com uma matriz de oligodT para o enriquecimento da extremidade 3' dos transcritos que contém a cauda de poliA, mas também pode ocorrer o enriquecimento de outras regiões dos RNAs, que sejam ricas em adeninas. Ainda assim, esses artefatos podem ser reduzidos através da implementação de filtros computacionais para descartar sítios de poliA candidatos complementares a regiões do genoma ricas de adeninas. No nosso protocolo, a transcrição reversa não foi realizada a partir de oligodT e sim a partir dos adaptadores de sequenciamento ligados ao RNA, o que deve minimizar a ocorrência de *internal priming*. Entretanto, como o enriquecimento da extremidade 3' baseou-se na incubação com uma matriz de oligodT, este enriquecimento não foi 100% e, além disso, o início do sequenciamento não foi direcionado para coincidir com o sítio de poliA. Assim, comparado com os demais protocolos que geraram principalmente sequências contendo o sítio de poliA (já que foram adicionados adaptadores específicos na extremidade 3' para o sequenciamento), nosso protocolo apresentou muitas sequências que não possuíam o sítio de poliA. A vantagem do nosso protocolo, no entanto, foi a facilidade de execução, ausência de oligos customizados e realização apenas de etapas já estabelecidas para o sequenciamento de RNA total no SOLiD, diminuindo a possibilidade de artefatos para identificação de sítios de poliA e para a estimativa de expressão de variantes de poliA.

A tecnologia de sequenciamento direto do RNA, no equipamento da Helicos BioSciences, também já foi utilizada para a identificação de APA em larga escala (OZSOLAK et al. 2010; LIN et al. 2012). Nesses trabalhos, o

RNA poliA+ foi capturado através de hibridização em superfície coberta com oligodT e realizado o sequenciamento direto. Essa metodologia é vantajosa por eliminar os passos de transcrição reversa, ligação e amplificação, que podem produzir artefatos. Embora seja feita a captura dos transcritos com oligodT, foi reportado um número baixo de variantes falsas por *internal priming*. No entanto, as desvantagens desse método são: sequências curtas geradas com 35nt em média, alta taxa de erro de sequenciamento (4%) e o alto custo do equipamento da Helicos.

4.1.6 Validação por RACE 3' das variantes de poliadenilação

A eficiência dos critérios estabelecidos computacionalmente para identificar variantes de poliA foi avaliada através da validação experimental por RACE 3'. Foram selecionadas 10 variantes identificadas nas 2 linhagens e que apresentavam pelo menos 10 sequências reportando a existência das mesmas. Os genes selecionados para validação foram: *ACTG1*, *ANAPC11*, *ARF1*, *ATP2A2*, *CALM3*, *FLNB*, *FNDC3B*, *KCMF1*, *TIMP2*, *UQCRQ*. Primers específicos para cada variante foram desenhados e utilizados na amplificação por RACE 3' (**Figura 18A**). Os produtos das reações foram purificados e sequenciados, e a existência das variantes curtas foi confirmada pela presença de adeninas ao final da extremidade 3' e que não estavam presentes no genoma (**Figura 18B**). Das 10 variantes mais curtas preditas, 8 foram confirmadas após o sequenciamento do produto do RACE 3'. Ao analisar as sequências na interface gráfica do Genome Browser da UCSC (<http://genome.ucsc.edu/index.html>), foi possível visualizar

perfeitamente que o final do alinhamento entre o genoma humano e as sequências geradas no SOLiD coincide com a extremidade 3' do produto amplificado pelo RACE 3' (**Figura 18C**). Apenas 2 candidatos não foram confirmados: ARF1 e ACTG1. Para estes casos, apenas o transcrito referência com maior 3' UTR foi amplificado, confirmando que o gene é expresso nas linhagens e que o primer específico desenhado foi eficiente na amplificação. A região genômica destes genes bem como as sequências obtidas no SOLiD que reportaram a existência da variante curta foram reavaliadas e nada em particular foi verificado.

Por fim, todos estes genes avaliados por RACE 3' foram comparados aos sítios de poliA presentes no banco de dados *Expression and Polyadenylation Database* (LIN et al. 2012). As 2 variantes curtas que não foram validadas por RACE 3' não estão presentes nesse banco de dados, já outras 7 variantes validadas por RACE 3' coincidiram com variantes descritas. Apenas a variante do gene *FNDC3B* não consta nesse banco de dados, e pode representar uma nova variante de poliA, embora não tenha sido feita uma comparação com outros trabalhos já publicados.

fizeram tal validação experimental, não permitindo uma comparação direta com o nosso trabalho. Um dos trabalhos avaliou 7 variantes de poliA por RACE 3' e todas foram confirmadas (FU et al. 2011). Outro trabalho utilizou RT-PCR para confirmar a existência de 20 variantes de poliA identificadas por sequenciamento (NI et al. 2013). No entanto, esta validação foi semelhante à que foi realizada em nosso trabalho para validação dos dados do 454 (item 3.1.5) e, como já discutido, este não foi um método de validação confiável. Por fim, a eficiência de validação de 80% verificada é uma porcentagem bastante razoável para uma metodologia em larga escala como a utilizada neste trabalho.

4.1.7 Estudo da expressão de miRNAs através do sequenciamento no SOLiD

O mesmo RNA total utilizado na construção das bibliotecas enriquecidas para a porção 3' dos transcritos foi utilizado na construção das bibliotecas de miRNA. Para a construção dessas bibliotecas, a avaliação da qualidade da amostra, baseada na integridade do RNA (RIN), é um passo importante para uma análise de expressão de miRNAs confiável e reproduzível. Assim, a qualidade dos RNAs foi avaliada no Agilent 2100 Bioanalyzer, sendo que ambos mostraram-se bastante íntegros com RIN superior a 9,0, sendo 10 o valor máximo possível (**Tabela 6**). Além da qualidade, com o auxílio do software disponível no Bioanalyzer, foi analisada a porcentagem de miRNA presente em cada amostra. O RNA total foi quantificado através do chip RNA 6000 Nano e o miRNA (faixa de 10 a 40nt)

através do chip small RNA, assim, pôde ser estimada a porcentagem de miRNA em cada amostra de RNA total. Inicialmente, as amostras possuíam uma porcentagem baixa de miRNAs, menor do que 1% (**Tabela 6**). Antes de iniciar a construção das bibliotecas, as amostras foram enriquecidas para RNAs pequenos (de até 300nt), conforme recomendado. Após o enriquecimento, a porcentagem de miRNAs foi novamente quantificada verificando-se um significativo aumento da porcentagem de miRNA presente nas amostras para 10% e 8% na HCT116 e SW480, respectivamente (**Tabela 6**).

Tabela 6 - Avaliação no Bioanalyzer das amostras de RNA utilizadas na construção das bibliotecas de miRNA.

Amostra	RIN	% miRNA inicial	% miRNA final (pós enriquecimento)
HCT116	9,6	0,91	10,4
SW480	9,1	0,6	8,5

O RNA enriquecido para fragmentos pequenos foi utilizado na construção das bibliotecas de miRNA. Ao final da construção, as mesmas foram avaliadas no Bioanalyzer quanto ao perfil de tamanho dos fragmentos e quantificação do DNA na faixa de tamanho de interesse. Um exemplo de perfil gerado pelo Bioanalyzer está mostrado na **Figura 19**. Nele observa-se um pico menor referente aos adaptadores sem inserto (90-100pb) e um pico maior referente aos miRNAs ligados a adaptadores, que deve apresentar de 115-130pb considerando o tamanho dos insertos selecionados em gel no início da construção da biblioteca (18-38pb). A partir da faixa de tamanho

esperada para os miRNAs, é possível estimar a porcentagem de miRNAs na biblioteca. As bibliotecas das duas linhagens apresentaram uma porcentagem de miRNA de acordo com o recomendado (acima de 50%), sendo 62,4% para HCT116 e 59,6% para SW480.

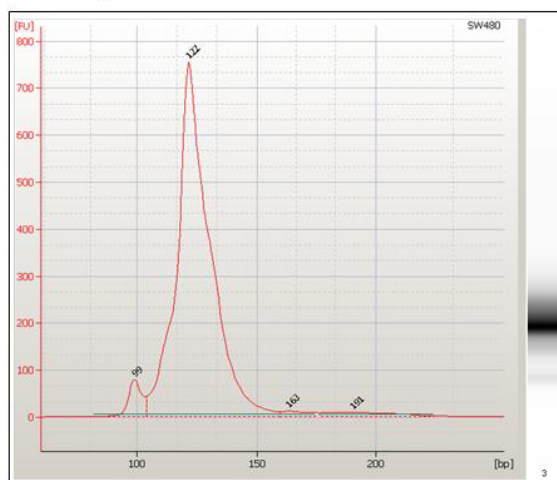


Figura 19 - Perfil de tamanho dos fragmentos de uma biblioteca de miRNA de linhagem celular avaliada em uma corrida de chip de DNA High Sensitivity no Bioanalyzer. O tamanho dos fragmentos, em pares de bases (bp), está representado no eixo x em relação à sua abundância (fluorescência - FU) no eixo y.

As bibliotecas foram sequenciadas no SOLiD4, sendo obtidas aproximadamente 27 e 32 milhões de seqüências para as linhagens HCT116 e SW480, respectivamente. Estas seqüências foram alinhadas contra o banco de dados do miRBase versão 18, com o auxílio do *CLC Genomics Workbench*. Em ambas as linhagens, foi identificada uma grande quantidade de miRNAs maduros conhecidos: 837 na HCT116 e 785 na SW480 (**Tabela 7**). Este número é semelhante ao número de miRNAs encontrados em trabalhos que utilizaram o sequenciamento em larga escala para o estudo da expressão de miRNAs (MEIRI et al. 2010; SCHULTE et al. 2010).

Tabela 7 - Dados obtidos a partir do sequenciamento das bibliotecas de miRNA no SOLiD.

	HCT116	SW480
Sequências geradas	27.085.160	31.777.820
Sequências mapeadas em miRNAs maduros	7.696.425 (28,42%)	4.322.884 (13,60%)
miRNAs maduros expressos	837	785

Após a identificação de miRNAs conhecidos, foi avaliada a expressão de cada um deles. O número de sequências obtidas para um dado miRNA foi normalizado pelo número total de sequências de miRNAs maduros obtidas para cada linhagem, gerando-se um valor correspondente ao número de sequências para um dado miRNA por milhão de sequências mapeadas em miRNAs (*counts per million, cpm*). Para uma análise mais robusta de expressão diferencial, apenas miRNAs que apresentassem pelo menos 10 cpm em uma das linhagens foram considerados nas análises seguintes, o que consistiu em um grupo de 348 miRNAs. A análise de expressão diferencial foi feita utilizando-se o programa EdgeR (ROBINSON 2010). Embora não seja ideal comparar apenas 2 amostras, pois o programa não é capaz de estimar com confiança a variabilidade biológica e dispersão dos dados, é possível estabelecer parâmetros mais rigorosos para melhorar essas estimativas e com isso identificar um grupo de miRNAs com expressão diferencial significativa. Assim, foram encontrados 39 miRNAs com diferença de expressão significativa ($p < 0,05$) entre as duas linhagens, no entanto com $FDR > 0,05$ (**Tabela 8**).

Tabela 8 - MiRNAs diferencialmente expressos entre as linhagens HCT116 e SW480 avaliados no EdgeR. logFC: diferença de expressão da SW480 em relação a HCT116 em escala log2.

miRNA	logFC	p
miR-138-1-3p	14,37	0,001
miR-744-3p	-13,67	0,002
miR-31-5p	-11,35	0,003
miR-26a-1-3p	11,23	0,004
miR-744-5p	-11,22	0,006
miR-100-3p	-10,85	0,007
miR-150-5p	9,93	0,007
miR-373-3p	11,03	0,007
miR-489-3p	-10,54	0,008
miR-138-5p	9,41	0,009
miR-21-5p	-9,23	0,010
miR-449b-5p	-10,04	0,010
miR-376c-3p	-9,81	0,011
miR-3667-3p	-9,65	0,012
miR-100-5p	-8,73	0,013
miR-891a-5p	9,72	0,014
miR-181a-3p	-8,60	0,014
miR-504-5p	9,52	0,015
miR-503-3p	9,49	0,016
miR-1	9,36	0,017
miR-517a-3p//miR-517b-3p	9,26	0,018
miR-125a-3p	9,24	0,018
miR-509-3-5p	9,11	0,019
miR-4301	9,08	0,019
miR-3925-3p	8,91	0,021
miR-499a-5p	7,88	0,022
miR-544b	8,70	0,023
miR-589-3p	8,59	0,025
miR-452-5p	-7,62	0,025
miR-1246	8,51	0,026
miR-767-5p	7,28	0,029
miR-125b-5p	-6,90	0,034
miR-185-5p	-6,90	0,034
miR-145-5p	-6,70	0,038
let-7a-2-3p	-6,59	0,041
miR-31-3p	-6,50	0,042
let-7b-5p	6,46	0,043
miR-130a-3p	-6,40	0,044
miR-376a-3p	-6,55	0,046

Da lista de miRNAs diferencialmente expressos foram selecionados 2 para a validação por uma outra metodologia, o qPCR, sendo eles: mir-100-5p e mir-21-5p. Por sequenciamento, a HCT116 expressa um maior nível desses 2 miRNAs, o que também foi observado por qPCR (**Figura 20**). No entanto, a escala de diferença de expressão é diferente entre as 2 metodologias, provavelmente por serem tecnologias muito diferentes. Características intrínsecas de cada miRNA (como sua própria sequência) podem afetar tanto sua avaliação por qPCR quanto sequenciamento. No caso da PCR, não há muita opção para o desenho dos primers utilizados na detecção de cada miRNA (como pode ser feito para mRNAs) pois o transcrito maduro é muito pequeno. Assim, nem sempre é possível atingir 100% de eficiência de amplificação. De qualquer maneira, foi possível confirmar a diferença de expressão desses 2 miRNAs estudados.

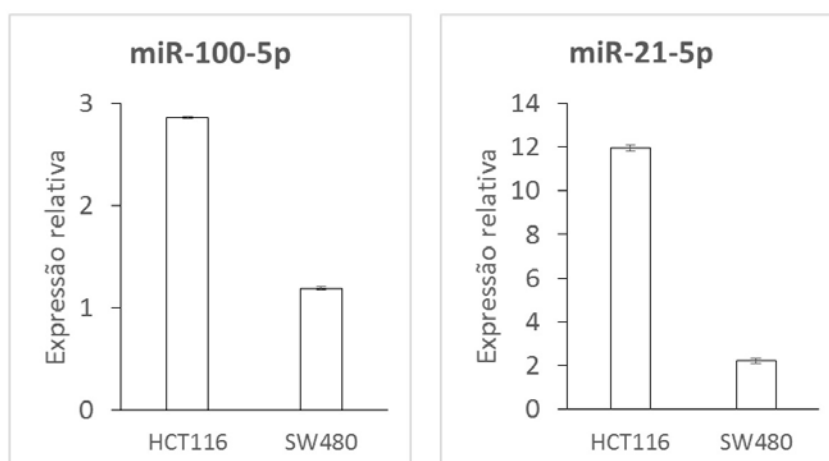


Figura 20 - Expressão avaliada por qPCR de 2 miRNAs identificados como diferencialmente expressos nos dados de sequenciamento. A expressão de cada miRNA foi normalizada pelo RNU6b, o gráfico mostra a expressão relativa ($2^{-\Delta\Delta Ct}$) a uma amostra de tecido de cólon normal.

A expressão diferencial de uma grande quantidade de miRNAs entre a HCT116 e a SW480 confirmou a hipótese de que cada linhagem apresentaria um perfil de expressão associado à instabilidade de microssatélites (MIS). Como descrito anteriormente, essas duas linhagens celulares se diferenciam quanto à MIS e já havia sido descrito que a MIS em pacientes com câncer de cólon pode ser predita com base no perfil de expressão dos miRNAs (SCHEPELER et al. 2008; SARVER et al. 2009). Dentre os miRNAs identificados por Sarver e cols, o miR-31-5p também estava diferencialmente expresso entre a HCT116 e a SW480. Em ambos os trabalhos, o aumento da expressão de miR-31-5-p estava associada à MIS. Entretanto, este foi o único miRNA comum entre os estudos, provavelmente por terem sido estudadas apenas duas linhagens celulares, enquanto os outros estudos foram feitos com mais de 50 amostras de pacientes.

Assim, trabalhando com linhagens contendo diferentes perfis de miRNA também seria esperado encontrar diferenças entre as variantes de poliA presentes em cada uma delas. Além de possibilitar a identificação de um maior número de variantes, esta abordagem também poderia facilitar a identificação de casos de miRNAs e variantes de poliA correlacionados.

4.1.8 Correlação de variantes de poliadenilação e de miRNAs

Além de estabelecer um protocolo para identificar e quantificar a expressão de variantes de poliA em larga escala, dentre nossos objetivos iniciais estava tentar associar o impacto dos miRNAs na expressão de variantes de poliA que poderiam ser reguladas pelos mesmos.

- **Variantes de poliadenilação e sítios alvo de miRNA**

Inicialmente, os sítios alvo de miRNA preditos pelo Target Scan (LEWIS et al. 2005) foram identificados nas regiões 3' UTR de todos os genes humanos. Das 3.946 variantes curtas identificadas na linhagem HCT116, 2.371 (60%) apresentaram a perda de pelo menos um sítio alvo de miRNA em relação ao gene referência (o transcrito com 3' UTR mais longa). Para a linhagem SW480, 2.674 variantes (61%) apresentaram a perda de pelo menos um sítio de miRNA. Assim, foi avaliado o impacto da presença de sítios alvo de miRNA na expressão de variantes de genes que apresentam APA. Para uma análise mais robusta e confiável, os dados de expressão das 2 linhagens celulares foram analisados conjuntamente e descritos a seguir.

Para os genes em que mais de 2 variantes de poliA foram identificadas, foram consideradas apenas 2 delas: a variante cujo sítio de poliA estivesse situado mais 5' no último exon do gene RefSeq (variante curta) e a variante cujo sítio de poliA estivesse situado mais 3' (variante longa).

Comparando-se os genes cuja variante curta não apresenta perda de sítio alvo de miRNA com genes cuja variante curta perde um ou mais sítios de miRNA, foi observada uma redução da razão de expressão da variante longa pela expressão da variante curta (chamada daqui para frente, razão VL/VC) (**Figura 21A**). Este resultado era esperado, uma vez que variantes curtas podem escapar da regulação exercida pelos miRNAs enquanto as

longas são negativamente reguladas pelos miRNAs que se ligam ao(s) sítio(s) alvo presente apenas nas mesmas.

É comum um determinado mRNA ser alvo de mais de um miRNA. No caso das variantes de poliA aqui encontradas, o número de sítios alvo de miRNA perdidos na variante curta variou muito, de 1 a 95 sítios, sendo a média de sítios perdidos igual a 6,8 e mediana igual a 4. Considerando-se que quanto mais sítios alvo de miRNA forem encontrados exclusivamente na variante longa, maior a chance da mesma ter sua expressão diminuída, foi feita uma análise para avaliar o efeito do número de sítios alvo de miRNA na razão VL/VC. Para tanto, subdividimos os casos de variantes de poliA com perda de sítio de miRNA em 1) genes cuja variante curta perdeu de 1 a 5 sítios de miRNA e 2) genes cuja variante curta perdeu mais de 5 sítios de miRNA. Conforme pode ser verificado na **Figura 21B**, para os genes com perda de mais de 5 sítios, a razão VL/VC é estatisticamente menor do que a razão observada nos casos sem perda de sítio e inclusive aqueles com perda de até 5 sítios de miRNA. O maior número de sítios de miRNA aumenta a chance de pelo menos algum destes miRNAs ser expresso na célula e assim regular negativamente a variante longa. Além disso, a atuação de mais de um miRNA em um mesmo mRNA alvo pode aumentar a eficiência de sua degradação (BARTEL 2009).

Em um trabalho feito com uma linhagem de mama normal e uma tumoral também foi verificado a redução da razão VL/VC para genes que apresentavam sítios alvo de miRNA quando comparados a genes sem sítios (LIAW et al. 2013). Legendre e cols identificaram variantes de poliA a partir

de dados de EST e em seguida buscaram por sítios alvo de miRNA. Os autores verificaram que 52% dos sítios alvo de miRNA preditos estão localizados downstream do primeiro sítio de poliA. Sendo assim, as variantes de APA são reguladas diferencialmente por miRNAs de acordo com a presença ou ausência dos respectivos sítios alvo em suas regiões 3' UTR (LEGENDRE et al. 2006). Mais uma vez, corroborando para a importância de sítios de miRNA em genes de APA, no nosso trabalho 60% das variantes curtas de poliA identificadas perderam pelo menos um sítio de miRNA.

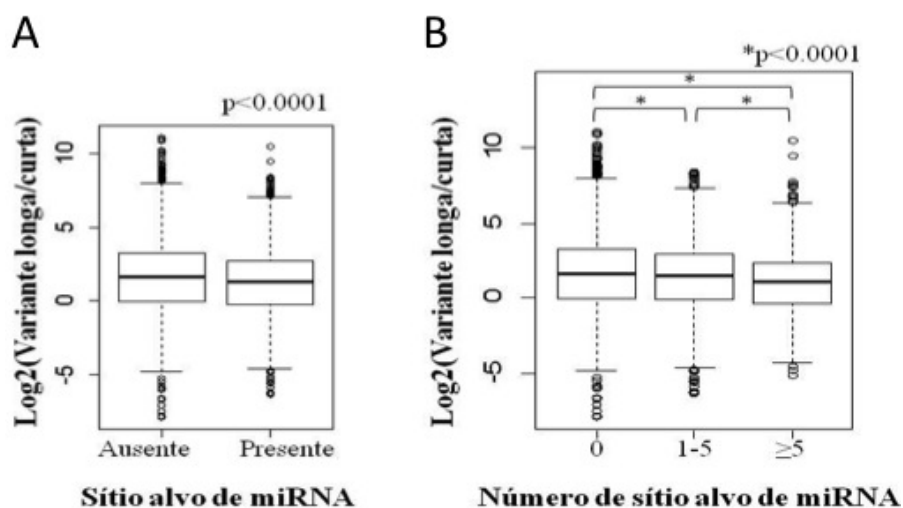


Figura 21 - Expressão de variantes de poliA em função do número de sítios alvo para miRNA. Boxplot representa os valores da razão variante longa pela variante curta (VL/VC), em escala log₂, considerando a ocorrência de sítios alvo para miRNAs na região que diferencia a variante longa da curta. A razão VL/VC dos genes que não possuem sítio alvo para miRNA (n=3.260) foi comparada com a razão VL/VC dos genes que perderam pelo menos um sítio (n=5.045) (A). Os genes que perderam sítio de miRNA foram subcategorizados de acordo com o número de sítios perdidos, de 1 a 5 (n=3.031) ou mais do que 5 (n=2.014) (B). Para os genes que possuem sítio alvo para miRNA observa-se uma redução da razão VL/VC, sendo essa redução ainda maior para genes que perderam mais do que 5 sítios. Os círculos se referem aos valores extremos, aqueles valores 1,5 vezes maior do que a amplitude interquartil. Para análise estatística foi feito teste T ou ANOVA e Bonferroni para o pós-teste.

- **Variantes de poliadenilação e expressão de miRNAs**

O impacto da presença dos sítios de miRNA na razão VL/VC só faria sentido se os miRNAs correspondentes fossem expressos nas linhagens celulares. Para confirmar esta hipótese, foram comparados genes que perderam sítios de miRNA cuja expressão não foi observada nos dados de sequenciamento com genes que perderam sítios de miRNA expressos em cada uma das linhagens. Como esperado, foi verificada uma menor razão VL/VC para os genes cuja variante longa apresenta sítios de miRNA expressos nas células analisadas (**Figura 22**). É interessante notar que não há diferença estatística na razão VL/VC de genes em que a variante longa não apresenta sítio alvo de miRNA e a razão de genes cuja variante longa apresenta sítios de miRNAs não expressos. Esse resultado reforça uma associação entre os miRNAs e as variantes de poliA.

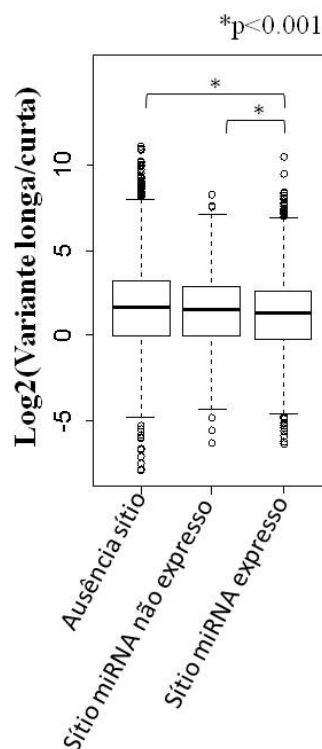


Figura 22 - Expressão de variantes de poliA em função da expressão de miRNAs. Boxplot representa os valores da razão variante longa pela variante curta (VL/VC), em escala log2, considerando a expressão de miRNAs que possuem sítio alvo na região que diferencia a variante longa da curta. Foram comparados genes que não possuem sítios alvo para miRNA (n=3.260), com genes que possuem sítios para miRNAs não expressos na célula (n=634) e genes que possuem sítios para miRNAs expressos (n=4.411). Não há diferença da razão VL/VC para os genes que não possuem sítio ou possuem sítio para miRNAs não expressos, no entanto essa razão diminui quando compara-se com genes que possuem sítios alvo para miRNAs expressos. Os círculos se referem aos valores extremos, aqueles valores 1,5 vezes maior do que a amplitude interquartil. Para análise estatística foi feito ANOVA e Bonferroni para o pós-teste.

Por último, foi avaliado um pouco mais o impacto do nível de expressão dos miRNAs sobre a razão VL/VC dos genes alvo. Para tanto, para cada gene foi feita a soma da expressão de todos os miRNAs que possuíam sítios alvo na região que diferenciava a variante longa da curta e esta soma de expressão (em cpm) foi dividida nos 75% menos expressos (denominados miRNAs de baixa expressão) e 25% mais expressos (miRNAs de alta expressão). Em seguida, para cada linhagem foram comparadas as

razões VL/VC dos genes alvos de miRNAs não expressos, dos genes alvos dos miRNAs de baixa expressão e dos genes alvos dos miRNAs de alta expressão. Não houve diferença estatística entre a razão VL/VC dos genes alvos de miRNAs não expressos e a razão dos genes alvos dos miRNAs de baixa expressão. Interessantemente, foi observada uma redução estatisticamente significativa da razão VL/VC dos genes alvos dos miRNAs de alta expressão quando comparada aos outros dois grupos de genes (Figura 23).

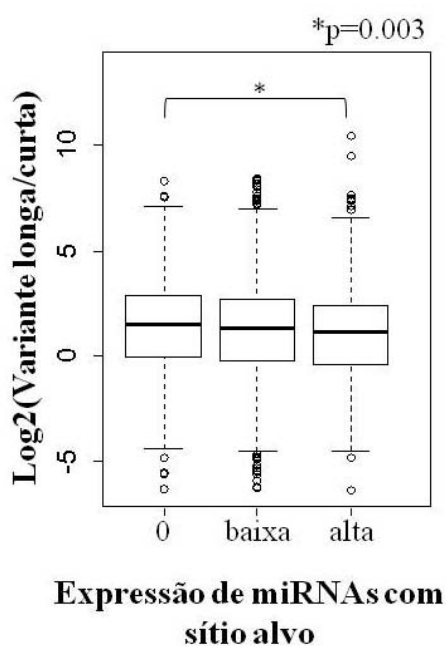


Figura 23 - Expressão de variantes de poliA em função do nível de expressão dos miRNAs. Boxplot representa os valores da razão variante longa pela variante curta (VL/VC), em escala log₂, considerando a expressão de miRNAs que possuem sítio alvo na região que diferencia a variante longa da curta. Para cada gene foi feita a soma da expressão de todos os miRNAs que possuem sítios alvo na variante longa, em seguida, os genes foram separados em função de possuírem sítio para miRNAs não expressos (n=634), de possuírem sítio para miRNAs cuja soma da expressão é baixa (75% inferiores, n=3.308) e de expressão alta (25% superiores, n=1.103). Observa-se uma redução estatisticamente significativa da razão VL/VC quando a variante longa é alvo para miRNAs com maior taxa de expressão. Os círculos se referem aos valores extremos, aqueles valores 1,5 vezes maior do que a amplitude interquartil. Para análise estatística foi feito ANOVA e Bonferroni para o pós-teste.

Essa associação entre a expressão de miRNAs e variantes de poliA já havia sido descrita para dois miRNAs (miR-1 e miR-124) em um estudo de análise global com ESTs, sendo que para os tecidos onde havia expressão dos miRNAs observou-se uma redução da expressão de variantes com sítios alvo para os mesmos (LEGENDRE et al. 2006). Mais recentemente, utilizando dados de sequenciamento em larga escala disponíveis para miRNAs e para variantes de poliA, foi visto que os genes alvo de miRNAs mais expressos em uma linhagem tumoral de mama do que em uma linhagem normal apresentaram uma redução da razão VL/VC na linhagem tumoral em relação à normal (LIAW et al. 2013). No entanto, a redução observada também foi pequena, de maneira semelhante ao que foi aqui encontrado.

- **Comparação da expressão de variantes de poliadenilação e de miRNAs entre as linhagens**

Nas análises acima, foi verificada uma tendência global de redução da razão VL/VC de genes alvos de miRNA expressos, sendo essa redução ainda maior para os genes alvos de miRNAs de alta expressão. Posteriormente, através da comparação entre as duas linhagens, foram investigados casos específicos onde há uma associação entre o aumento de expressão de miRNAs e a redução da razão VL/VC. Para tanto, foram identificados os casos que haviam uma diferença na expressão entre as linhagens tanto para as variantes (razão VL/VC) quanto para a expressão de

miRNAs, sendo que uma correlação inversa seria um indício do papel regulatório exercido pelo miRNA.

Foram analisadas apenas as variantes encontradas em ambas as linhagens e que perderam sítios para miRNAs expressos, constituindo um conjunto de 696 genes. Comparando-se a razão VL/VC destes genes na HCT116 com a razão VL/VC na SW480, foram identificados 125 genes (18%) com razão VL/VC 4 vezes menor na HCT116, ou seja, uma menor expressão da variante longa na HCT116. Dentre esses genes apenas 8 (6%) apresentaram um aumento de pelo menos 4 vezes da soma da expressão de todos os miRNAs que tiveram sítios perdidos pela variante curta quando comparados com a SW480. Já para a SW480, 93 genes (13%) apresentaram redução da razão VL/VC e, dentre esses, 35 genes (38%) apresentaram aumento da soma da expressão de todos os miRNAs que tiveram sítios perdidos pela variante curta. A partir dessas comparações entre as duas linhagens é difícil assumir se a correlação entre o aumento de expressão de miRNAs e a redução da razão VL/VC é realmente rara ou se a maneira como os dados foram analisados não foi apropriada para identificar os casos em que isto de fato acontece. A comparação entre duas linhagens pode não ser suficiente para evidenciar esses casos, principalmente por se tratarem de linhagens tumorais distintas, nas quais os mecanismos importantes na regulação gênica podem ser diferentes entre elas. Uma possível falha na análise pode ter sido considerar a soma da expressão normalizada (cpm) de todos os miRNAs com sítio alvo na variante longa em questão. Sendo assim, também poderia ter sido feita uma busca por

candidatos considerando-se somente os miRNAs com sítio alvo na variante longa que tivessem uma expressão diferencial entre as linhagens, já que a soma da expressão de todos os miRNAs pode reduzir consideravelmente a influência de miRNAs pouco expressos e aumentar a influência de miRNAs altamente expressos. Outra possibilidade seria identificar todos os genes alvos para um determinado miRNA com expressão diferencial entre as linhagens e, posteriormente, avaliar se para este grupo de genes há uma redução da razão VL/VC.

4.1.9 Validação por qPCR da expressão diferencial de variantes de poliadenilação

Dentre os genes que tiveram diferença de pelo menos 4x da razão VL/VC entre a HCT116 e a SW480 foram escolhidos 3 candidatos para a validação experimental (TIMP2, FLNB, CALM3). Inicialmente, foi feita uma análise por RT-PCR semi-quantitativa e não qPCR porque as reações de amplificação sempre apresentavam mais de um produto amplificado, o que impossibilitava a análise confiável pela metodologia de qPCR a partir dos mesmos primers utilizados para RACE 3'. O método de RT-PCR semi-quantitativa não apresentou sensibilidade suficiente para verificar uma diferença de expressão entre as 2 linhagens. A diferença é pequena, mas parece haver uma diminuição da expressão tanto da variante longa quanto da variante curta de TIMP2 e FLNB na HCT116 em relação à SW480 (**Figura 24**). Já para CALM3 parece que a SW480 possui uma expressão menor tanto da variante longa quanto da curta. A especificidade dos

produtos de amplificação apresentados nas figuras foram confirmados por sequenciamento Sanger.

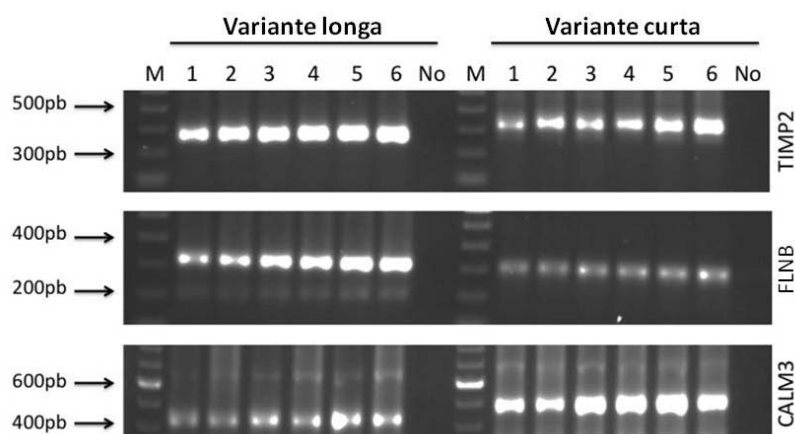


Figura 24 - RT-PCR semi quantitativa das variantes de poliA longa e curta que perderam sítios alvo de miRNAs. HCT116: 1 (22 ciclos), 3 (25 ciclos) e 5 (28 ciclos). SW480: 2 (22 ciclos), 4 (25 ciclos) e 6 (28 ciclos). Controle negativo: No. M= Marcador de peso molecular 100pb.

Embora a metodologia de RT-PCR semi-quantitativa seja interessante por possibilitar a visualização de bandas específicas para as variantes longa e curta, como mencionado acima, ela não apresentou sensibilidade suficiente para confirmar a diferença de expressão observada entre as variantes nos dados de sequenciamento. Evidentemente também não poderia ser descartada a possibilidade da diferença de expressão observada no sequenciamento em larga escala ser decorrente de um artefato inerente desta metodologia. Portanto, foi realizado qPCR para quantificar a expressão da variante longa em relação à expressão total do gene. Foi desenhado um par de primers que amplifica uma região específica da variante longa e um par de primer que amplifica a região comum entre as variantes longa e curta (a localização dos primers está representada na

Figura 25A). A expressão da variante longa foi normalizada pela expressão total do gene (que é dada pelo primer que amplifica a região comum entre a longa e a curta). É necessário normalizar pela expressão do gene para evitar assumir erroneamente que a diminuição de expressão da variante longa não é simplesmente decorrente da menor expressão do gene como um todo em uma linhagem em relação à outra.

As sequências geradas no SOLiD apontam para uma menor expressão da variante longa para o gene *TIMP2* na HCT116 em relação à SW480 (**Figura 25A**). Essa menor expressão da variante longa na HCT116 foi confirmada por qPCR (**Figura 25B**). Assim, procuramos por miRNAs superexpressos em HCT116 que pudessem regular especificamente a variante longa de *TIMP2*. Foram encontrados 27 miRNAs, sendo 22 expressos, que apresentavam sítios alvo na região que diferenciava a variante longa da curta de *TIMP2*. Dentre esses miRNAs, 6 apresentavam uma diferença de pelo menos 4 vezes entre as duas linhagens, 2 estavam mais expressos na HCT116 em relação à SW480 (miR-301B, miR-222-5p), no entanto 4 miRNAs estavam mais expressos na SW480 (miR-425, miR-421, miR-489, miR-93, **Figura 25C**). Esperava-se encontrar miRNAs mais expressos na HCT116 do que na SW480 que pudessem explicar a menor expressão da variante longa de *TIMP2* na HCT116. Contudo, tanto a HCT116 quanto a SW480 apresentavam miRNAs com diferença de expressão entre as mesmas e não foi possível associar a diminuição da variante longa de *TIMP2* na HCT116 ao aumento de miRNAs superexpressos nesta linhagem.

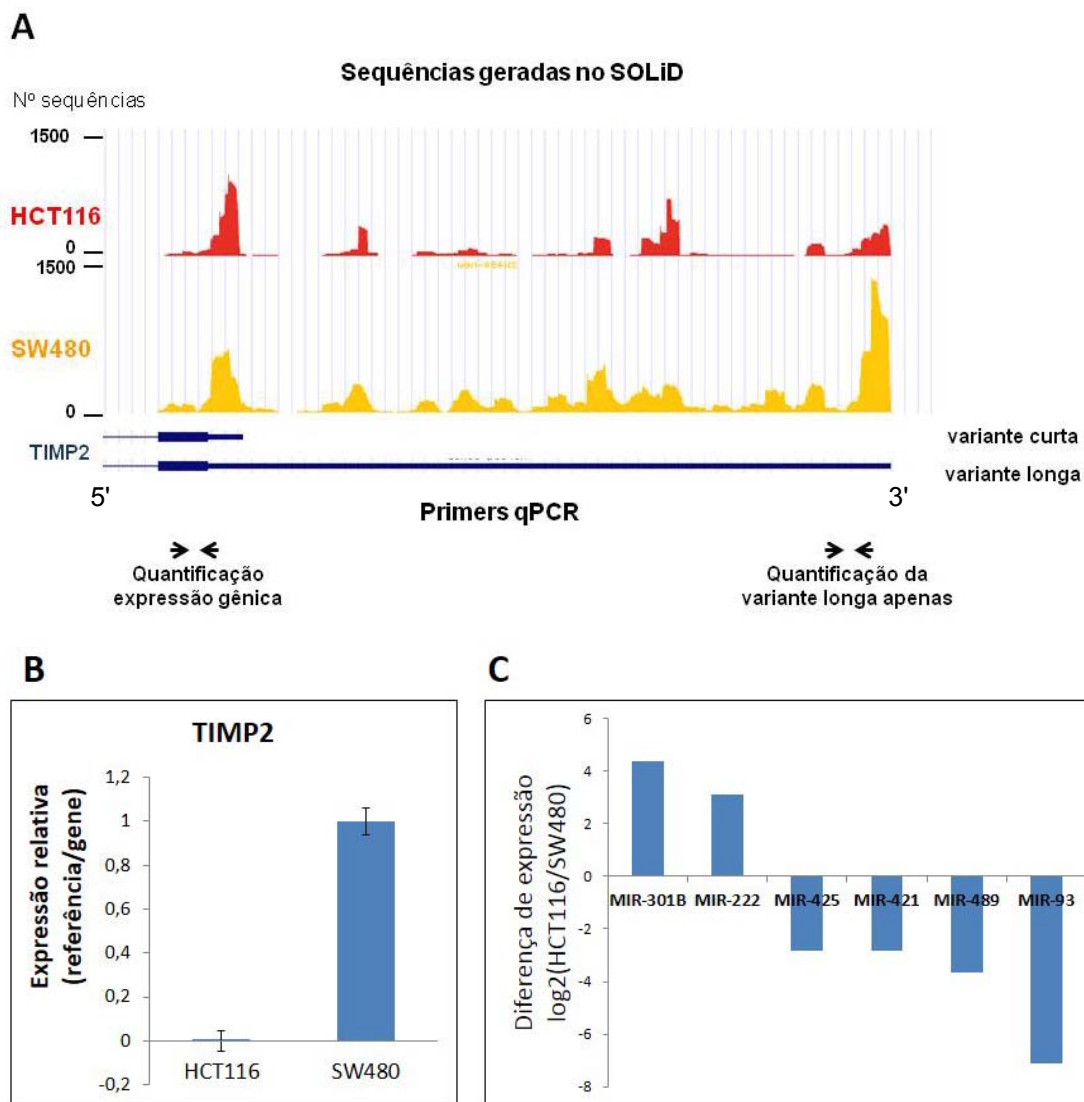


Figura 25 - Validação da expressão das variantes de poliA do gene *TIMP2*.

A) Interface gráfica demonstrando as sequências obtidas no SOLiD para o gene *TIMP2* (HCT116 em vermelho e SW480 em amarelo). Em azul está representado o último exon 3' do gene referência (coincidindo com a variante longa) e da variante curta detectada a partir do sequenciamento. Também está representada a região onde foram desenhados os primers para qPCR, um par está na região comum entre a variante curta e a longa e o outro par está na região específica da longa. B) Expressão avaliada por qPCR da variante longa de *TIMP2* normalizada pela expressão total do gene. O eixo y mostra a expressão relativa à SW480 ($2^{-\Delta\Delta C_t}$). C) Diferença de expressão (\log_2) dos miRNAs na HCT116 em relação à SW480. Estão representados os miRNAs que possuem sítios alvo apenas na variante longa e diferença de expressão de pelo menos 4 vezes entre as linhagens. A expressão dos miRNAs foi obtida por sequenciamento e foi calculada em número de sequências por milhão (cpm).

Essa análise também foi feita para o gene *FLNB*. No entanto, enquanto o sequenciamento revelou uma menor expressão da variante longa para HCT116 em relação à SW480, essa diferença de expressão não foi confirmada por qPCR (**Figura 26**). Por sequenciamento foi observada uma diferença de expressão de 4 vezes entre as duas linhagens e talvez essa diferença não seja grande o suficiente para ser detectada por qPCR já que neste caso não é feita uma razão VL/VC e sim uma razão “variante longa/expressão total do gene”. Assim, não foi investigada a expressão de miRNAs que continham sítios alvo no gene *FLNB* uma vez que mesmo se fossem encontrados miRNAs diferencialmente expressos entre as duas linhagens, eles certamente não estariam desempenhando um papel regulador neste contexto.

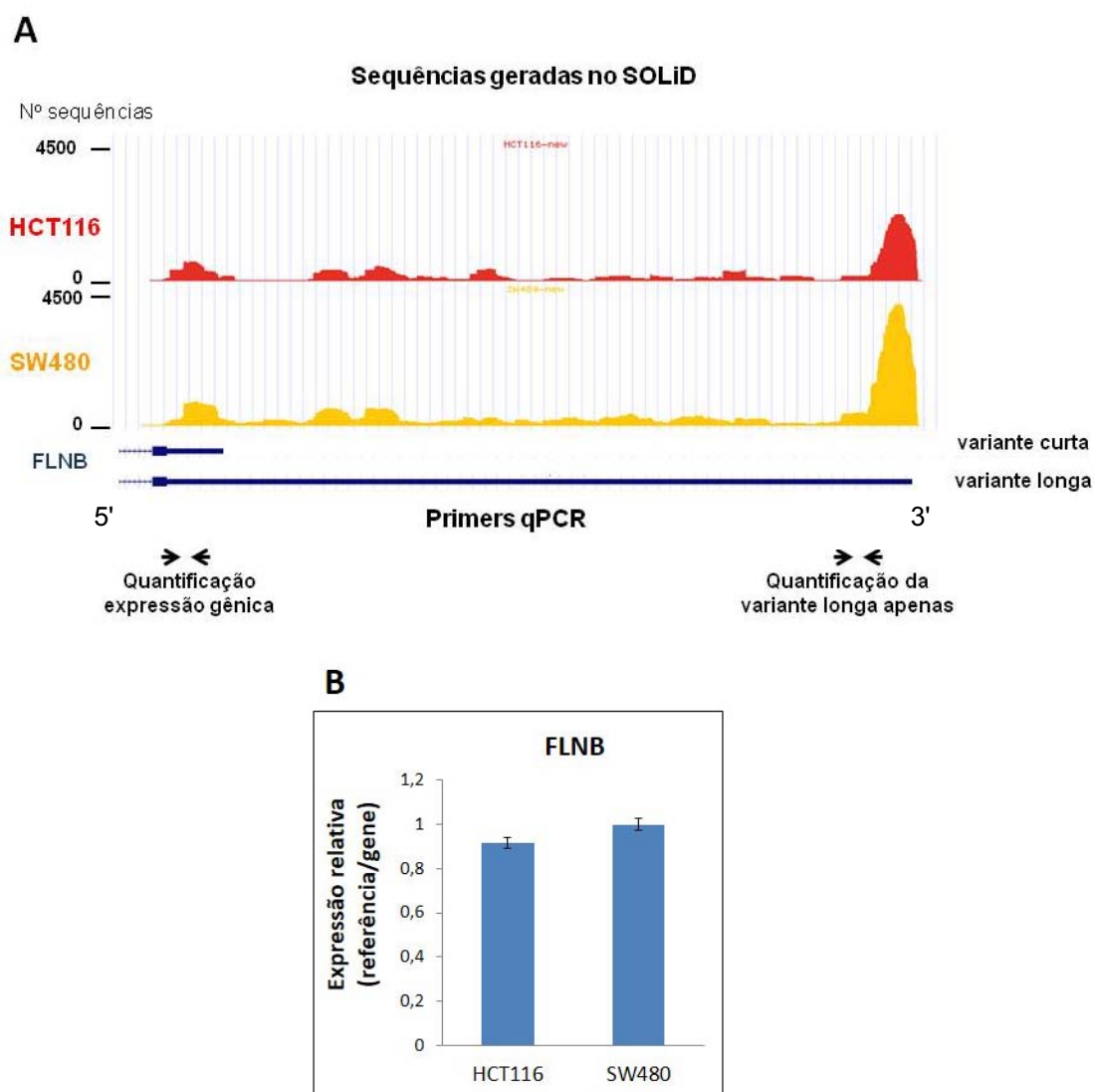


Figura 26 - Validação da expressão das variantes de poliA do gene *FLNB*. A) Interface gráfica demonstrando as sequências obtidas no SOLiD para o gene *FLNB* (HCT116 em vermelho e SW480 em amarelo). Em azul está representado o último exon 3' do gene referência (coincidindo com variante longa) e da variante curta detectada a partir do sequenciamento. Também está representada a região onde os primers para qPCR foram desenhados, um par está na região comum entre as variantes curta e longa e o outro par está na região específica para a longa. B) Expressão avaliada por qPCR da variante longa de *FLNB* normalizada pela expressão total do gene. O eixo y mostra a expressão relativa à SW480 ($2^{-\Delta\Delta C_t}$).

Por fim, também foi avaliado o candidato CALM3, que apresentou redução da expressão da VL/VC na SW480 em relação a HCT116 por sequenciamento (**Figura 27A**). Novamente, a redução da expressão da variante longa na SW480 foi confirmada por qPCR (**Figura 27B**). Além disso, foi verificado que a variante mais curta de CALM3 perde sítio alvo para 25 miRNAs, sendo 10 expressos, dos quais 5 apresentaram uma diferença de expressão de pelo menos 4 vezes entre as duas linhagens. Como esperado, todos os miRNAs com diferença de expressão entre as linhagens estavam mais expressos na SW480 (miR-320C, miR320A, miR-92A-3p, miR-361-5p, miR-27B, **Figura 27C**). Desse modo, uma maior expressão desses miRNAs poderia explicar a menor expressão da variante longa em relação à curta na SW480. Embora essa evidência ainda seja muito indireta.

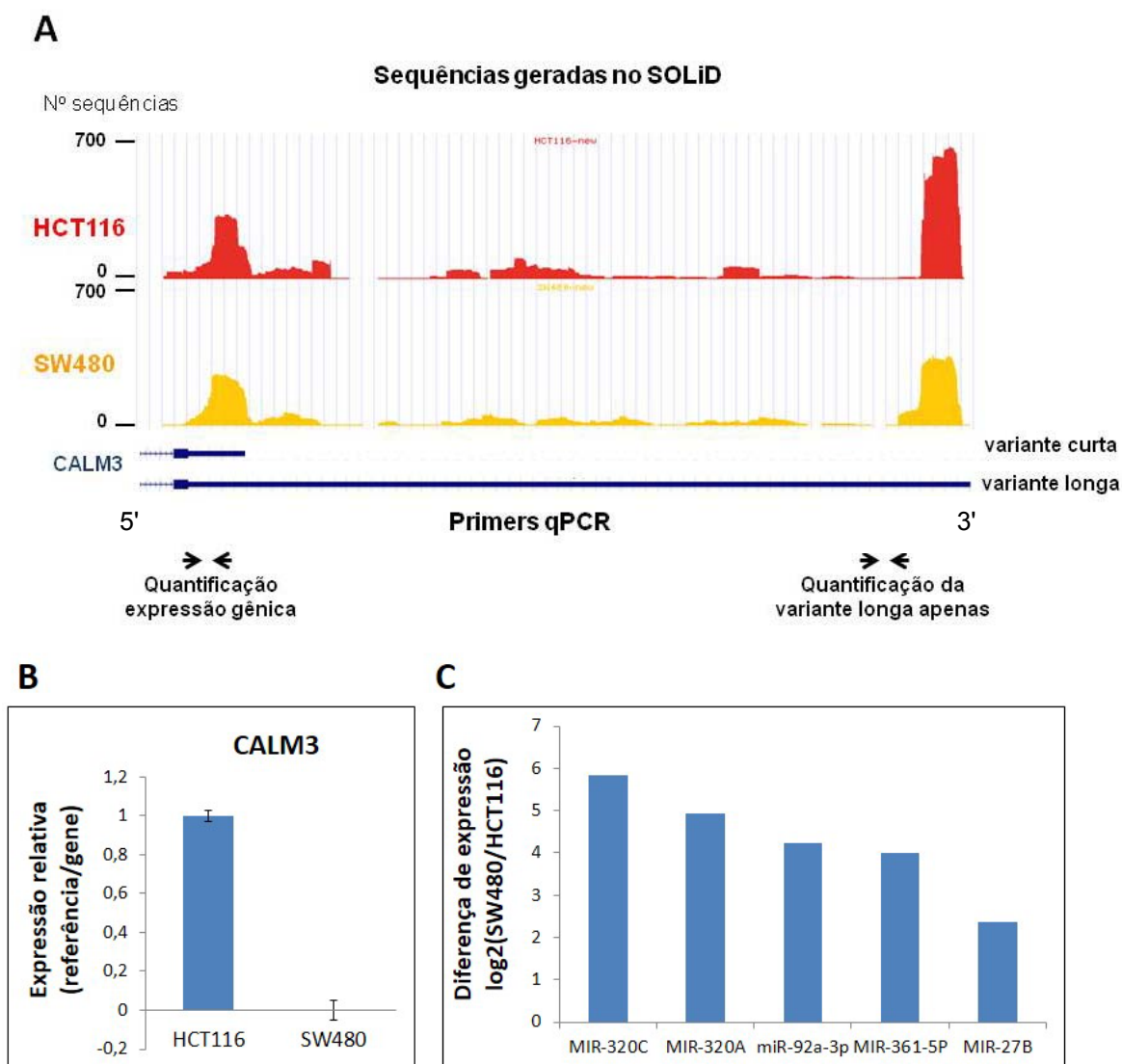


Figura 27 - Validação da expressão das variantes de poliA do gene *CALM3*.

A) Interface gráfica demonstrando as sequências obtidas no SOLiD para o gene *CALM3* (HCT116 em vermelho e SW480 em amarelo). Em azul está representado o último exon 3' do gene referência (coincidindo com variante longa) e da variante curta detectada a partir do sequenciamento. Também está representada a região onde os primers para qPCR foram desenhados, um par está na região comum entre as variantes curta e longa e o outro par está na região específica para a longa. B) Expressão avaliada por qPCR da variante longa de *CALM3* normalizada pela expressão total do gene. O eixo y mostra a expressão relativa à HCT116 ($2^{-\Delta\Delta C_t}$). C) Diferença de expressão (\log_2) dos miRNAs na SW480 em relação à HCT116. Estão representados os miRNAs que possuem sítios alvo apenas na variante longa e diferença de expressão de pelo menos 4 vezes entre as linhagens. A expressão dos miRNAs foi obtida por sequenciamento e foi calculada em número de sequências por milhão (cpm).

Resumidamente, por qPCR foram validados 2 genes para os quais a razão VL/VC tinha uma diferença de pelo menos 4 vezes entre as duas linhagens visto por sequenciamento. Pelo menos para o CALM3 também houve uma correlação inversa entre a expressão da variante longa e de miRNAs que apresentavam sítios alvo na mesma (os quais não estavam presentes nas variantes curtas) comparando a HCT116 e a SW480. Nestes casos, as variantes longas podem ser negativamente reguladas pelos miRNAs ou pode ocorrer uma preferência de expressão das variantes curtas como um mecanismo de escape da regulação exercida pelos miRNAs.

Essa estratégia parece bastante promissora, mas a identificação desta correlação inversa da expressão de miRNAs e variantes de poliA de maneira automática (computacionalmente) ainda não foi otimizada. A validação por qPCR da diferença de expressão da variante longa pela curta de 2 dos 3 genes testados, revela o potencial da nossa estratégia em identificar tais candidatos de APA. No entanto, ainda é necessário refinar a busca por genes com APA que são alvos de miRNAs diferencialmente expressos entre as linhagens. A busca de candidatos foi feita considerando-se a soma da expressão normalizada (cpm) de todos os miRNAs com sítio alvo na variante longa em questão, no entanto seria importante incorporar na análise tanto o número de miRNAs atuando no transcrito quanto o nível de expressão de cada miRNA.

Por fim, para confirmar experimentalmente a regulação exercida por um miRNA sobre um determinado mRNA alvo devem ser realizados experimentos *in vitro* com vetor contendo a 3' UTR da variante longa ou

curta e o gene repórter da luciferase. O vetor deve ser transfectado em linhagem celular isoladamente ou juntamente com o miRNA de interesse. Assim, a quantificação da atividade da luciferase permite avaliar a regulação das variantes pelo miRNA. Esse seria um experimento importante para confirmar a regulação da variante longa pelo miRNA, dando maior confiabilidade às correlações de expressão encontradas em larga escala.

Os trabalhos de identificação de variantes de poliA publicados até o momento têm sido enriquecedores principalmente no sentido de aumentar a porcentagem de transcritos descritos que sofrem tal processamento durante a maturação e por mostrarem a especificidade de APA em tecidos e condições fisiológicas e patológicas. Como já foi visto para o CCR, o padrão global de APA é alterado durante a progressão tumoral (MORRIS et al. 2012). Os genes que apresentam redução do tamanho da 3' UTR estão funcionalmente enriquecidos, principalmente para as categorias associadas a ciclo celular, fatores de ligação a ácido nucleicos e fatores de processamento de splicing (MORRIS et al. 2012). Já os genes que apresentam aumento do tamanho da 3' UTR estão enriquecidos para adesão célula-célula e matriz extracelular. Por último, a expressão diferencial de variantes de APA foi utilizada para a identificação de biomarcadores para a progressão do CCR, mostrando o potencial desse tipo de análise. Sendo assim, a grande quantidade de dados gerados no nosso trabalho também poderia ser utilizada para identificar novos genes que sofrem APA, regulados por miRNAs, e que podem desempenhar papéis importantes na tumorigênese.

Apesar de inúmeros trabalhos atuais corroborar com nossos achados, o verdadeiro impacto de APA na estabilidade de mRNAs foi questionado em um trabalho recente. SPIES et al. (2013) estudaram na linhagem NIH3T3 de fibroblasto de camundongo como a estabilidade e eficiência da tradução dos mRNAs podem estar associadas a diferentes características dos mesmos como: 5' UTR, ORF e 3' UTR (SPIES et al. 2013). Ao contrário do esperado, o impacto de APA foi baixo quando foram comparadas as variantes curtas e longas dos mesmos genes em uma escala global. Ainda assim, foi verificado uma associação negativa entre o tamanho da 3' UTR e a estabilidade do mRNA, sendo que o tamanho da 3' UTR explicou 28-32% da variação da meia-vida dos mRNAs. No modelo estabelecido a principal diferença na estabilidade dos mRNAs estava associada ao número de junções de exons por kilobase das sequências codificadoras. Já a comparação das taxas médias de tradução entre as variantes de poliA longas e curtas variou em menos de 3%, sendo que o recrutamento de fatores trans ou simplesmente o maior tamanho da 3' UTR poderiam aumentar a eficiência da tradução.

Embora tenha sido verificada uma baixa associação entre o tamanho da 3' UTR, a estabilidade dos mRNAs e a tradução das proteínas codificadas pelos mesmos, os elementos da 3' UTR com maior impacto foram os sítios alvo de miRNA. Os autores avaliaram que o discreto efeito dos miRNAs pode ter sido decorrente da escassa presença de sítios alvos para miRNAs altamente expressos. No entanto, os autores não avaliaram a expressão de variantes de poliA e de miRNAs nas mesmas amostras e como foi verificado no nosso trabalho o impacto sobre a razão VL/VC só

ocorre para genes alvo de miRNAs expressos, não ocorrendo uma diminuição da razão VL/VC para genes alvo de miRNAs não expressos na célula. Naquele trabalho também foi feita uma busca por domínios (*motifs*) significativamente associados com a desestabilização dos mRNAs e repressão da tradução das variantes longas. Entre os principais domínios, foram encontrados os sítios alvo de 2 miRNAs (miR-29 e let-7) que são sabidamente expressos em altos níveis nas células NIH3T3. Mais uma vez, reforçando a importância do nível de expressão dos miRNAs na regulação de variantes de poliA. Além disso, nos mRNAs alvo destes miRNAs verificou-se em média uma redução de 4% da tradução da proteína e uma redução de 30% na expressão do mRNA, confirmando trabalhos anteriores que mostram que em mamíferos os miRNAs agem predominantemente desestabilizando os mRNAs alvo (BAEK et al. 2008; GUO et al. 2010). Vale ressaltar que no nosso trabalho não foi possível avaliar o impacto dos miRNAs sobre mRNAs que estão sujeitos à repressão da tradução proteica, uma vez que para estes casos não há alteração do nível de expressão de mRNAs. No entanto, conforme demonstrado na literatura, a menor prevalência deste mecanismo não deve ter comprometido significativamente as análises do impacto global dos miRNAs sobre APA que foi verificado neste trabalho.

Esse último trabalho de Spies e cols mostrou que a correlação global entre a ocorrência de APA e a abundância dos mRNAs é baixa e menor do que o até então esperado. No entanto, em nosso trabalho foi observada uma redução global, estatisticamente significativa, da razão VL/VC para genes

que possuem sítios alvo para miRNAs expressos na região que diferencia a variante longa da curta. Vários trabalhos já demonstraram que a variante curta é mais estável do que a variante longa para genes específicos. Essa diferença na estabilidade dos mRNAs já foi demonstrada principalmente para genes envolvidos com proliferação (SANDBERG et al. 2008), desenvolvimento (JI et al. 2009) e diferenciação (JI e TIAN 2009). É possível que o efeito observado no nosso trabalho se torne mais evidente a partir de uma análise focada nesses grupos de genes para os quais já foi verificado um importante papel da APA. Como descrito anteriormente, um trabalho recente mostrou que para os genes alvo de miRNAs mais expressos em uma linhagem tumoral de mama em relação à normal houve uma redução global da razão VL/VC na linhagem tumoral, sendo assim este estudo também corrobora para a importância da expressão de miRNAs sobre a expressão de variantes de poliA (LIAW et al. 2013).

Além dos miRNAs outros elementos também são importantes e podem atuar na redução da razão VL/VC, por exemplo, as longas 3' UTR apresentam diversas sequências regulatórias, como sequências ricas em AU ou GU e sítios para proteínas de ligação ao RNA. Além disso, pode ocorrer uma preferência para a transcrição de variantes curtas quando há aumento da expressão de genes da maquinaria de poliA como CPSF e CSTF (JI e TIAN 2009). Por último, a atividade transcricional (determinada pela regulação do promotor, do posicionamento de nucleossomos e das marcas de histona) pode influenciar a abundância de variantes curtas, uma vez que já foi demonstrado que genes altamente expressos geram uma maior quantidade de variantes curtas enquanto genes de baixa expressão tendem a produzir variantes mais longas (JI et al. 2011).

4.2 PARTE II: ANÁLISE DE EXPRESSÃO DE MIRNAS EM PACIENTES COM CÂNCER DE RETO E SUA ASSOCIAÇÃO COM A RESPOSTA AO TRATAMENTO

4.2.1 Dados clínicos dos pacientes com câncer de reto

As características clínicas de todos os pacientes incluídos no estudo estão apresentadas na **Tabela 9**. No total, foram estudados 43 pacientes, que foram divididos em 3 grupos de acordo com a etapa do estudo. A primeira etapa foi realizada com 27 pacientes, sendo 7 de resposta completa e 20 de resposta incompleta ao tratamento QRTn. Estes pacientes foram utilizados inicialmente para a identificação de biomarcadores preditivos e fizeram parte do grupo denominado de treinamento. A segunda etapa do estudo foi realizada com 12 pacientes para a validação dos biomarcadores encontrados inicialmente e, portanto, estes pacientes fizeram parte do grupo de validação. A separação desses dois grupos foi feita por ordem cronológica da coleta do material para o estudo. Por último, também foi avaliado o grupo de recidivas. Este grupo é composto por amostras de 4 pacientes que apresentaram inicialmente cRC mas que tiveram recidiva precoce (tempo inferior a 16 meses) e por isso foram submetidos à cirurgia. Sendo assim, estes pacientes foram excluídos do grupo de treinamento pois não satisfaziam o critério de cRC sustentada (tempo maior do que 24 meses). No entanto, ao final do estudo os dados obtidos para o grupo de recidiva foram comparados com os demais.

As amostras de tumor foram coletadas por biópsia antes do início do tratamento QRTn. Inicialmente, foi feita a análise de expressão diferencial de miRNAs no pacientes do grupo de treinamento, comparando-se os pacientes com resposta completa ao tratamento em relação aos pacientes com resposta incompleta. Não houve diferença estatisticamente significativa das características clínicas entre esses dois subgrupos de pacientes, como idade, sexo, distância do tumor até a borda anal, tamanho do tumor e estadiamento inicial (**Tabela 9**).

Tabela 9 - Dados clínicos de todos os pacientes incluídos no estudo.

	Grupo de treinamento			Grupo de validação			Grupo recidiva
	Resposta completa	Resposta incompleta	p	Resposta completa	Resposta incompleta	p	
Número pacientes	7	20		7	5		4
Idade (anos)	58,9 ± 10,9	54,2 ± 13,7	ns	60,4 ± 13,6	59,6 ± 14,5	ns	69,3 ± 15,0
Sexo (M/F)	2/5	9/11	ns	3/4	3/2	ns	3/1
Distância da borda anal (cm)	4,0 ± 2,6	4,3 ± 2,2	ns	4,6 ± 2,4	2,2 ± 2,8	ns	4 ± 0,8
Tamanho do tumor (cm)	4,4 ± 0,8	4,6 ± 1,5	ns	3,6 ± 0,9	5 ± 1,9	ns	5 ± 1,4
Estadiamento inicial			ns			ns	
T2	0	2		4	2		0
T3	7	17		2	2		4
T4	0	1		1	1		0
N0	4	7		2	2		2
N+	3	13		5	3		2
Estadiamento após QRT							
cRC	3	-		6	-		4
pRC	4	-		1	-		-
ypT2	-	6		-	1		-
ypT3	-	13		-	4		-
ypT4	-	1		-	0		-
ypN0	-	12		-	1		-
ypN+	-	8		-	2		-
ypNx	-	0		-	2		-

Os pacientes foram divididos em 3 grupos (treinamento, validação e recidiva). Para as características descritas, não houve diferença estatisticamente significativa (ns) entre os pacientes de resposta completa e de resposta incompleta. Para cada característica está representada a média ± desvio padrão. cRC = resposta completa clínica. pRC = resposta completa patológica. Grupo de recidiva é composto de pacientes com cRC inicial mas que apresentaram recidiva em menos de 16 meses.

4.2.2 Sequenciamento de miRNAs em amostras de pacientes com câncer de reto

A análise da expressão de miRNAs nas amostras de pacientes foi feita de maneira semelhante à análise das duas linhagens apresentadas anteriormente. Como já descrito, é importante que as bibliotecas de miRNAs sejam construídas a partir de amostras de RNA de alta qualidade. Sendo assim, a qualidade das amostras foi avaliada no Bioanalyzer e o resultado está apresentado na **Tabela 10**. Apenas um caso apresentou RIN menor do que o recomendado, que é 6 (sendo 10 o número máximo calculado pelo Bioanalyzer). Entretanto, foi importante incluir este caso para aumentar o grupo dos pacientes com resposta completa, tendo em vista que apenas 30% dos pacientes apresentam resposta completa e conseqüentemente há uma maior dificuldade de obtenção de tais amostras.

Além da qualidade, também foi estimada no Bioanalyzer a porcentagem de miRNAs na amostra de RNA total. Observa-se uma grande variabilidade na porcentagem de miRNAs dessas amostras (1,5 a 24%, **Tabela 10**) e, ainda, essa porcentagem é maior quando comparada com as linhagens (0,6 e 0,9%, **Tabela 6**). Parte dessa diferença provavelmente se deve à diferença na qualidade entre as amostras das linhagens e dos pacientes. Como o cálculo da porcentagem de miRNAs é feito considerando-se os fragmentos presentes em uma determinada faixa de tamanho (10 a 40nt), vários desses fragmentos podem ser resultantes da degradação de outras classes de RNAs (como mensageiro, transportador e ribossômico). Se apenas a integridade das moléculas justificasse tal diferença, esperar-se-

ia uma correlação inversa entre o RIN de uma dada amostra e sua porcentagem de miRNA. Entretanto, tal correlação não é verificada, provavelmente devido a outros fatores também influenciarem a quantidade de miRNAs presente em cada amostra. A expressão aberrante de miRNAs (oncomiRs) resultante da instabilidade genômica tumoral, já foi descrita em inúmeros tipos de câncer, e também poderia explicar esta maior quantidade de miRNAs em algumas amostras (CALIN e CROCE 2006; Esquela-KERSCHER e SLACK 2006).

Antes da construção das bibliotecas de miRNA, as amostras de RNA total foram enriquecidas para RNAs pequenos (de até 300nt) e a porcentagem de miRNAs foi novamente estimada. Pode-se observar que houve um significativo enriquecimento nas amostras, resultando em amostras com porcentagens de miRNAs de 12 a 58% (**Tabela 10**).

Tabela 10 - Avaliação no Bioanalyzer das amostras de RNA do grupo de treinamento utilizadas na construção das bibliotecas de miRNA.

Paciente	Resposta ao tratamento	RIN	% fragmentos 10-40nt inicial	% fragmentos 10-40nt final
1	incompleta	7,0	6,7	40,2
2	incompleta	8,6	2,9	11,8
3	incompleta	7,1	3,8	27,0
4	completa	5,2	12,0	57,7
5	incompleta	6,9	4,8	49,7
6	incompleta	6,8	9,9	35,5
7	completa	6,5	6,0	38,0
8	completa	7,8	10,5	26,5
9	incompleta	6,6	4,23	33,0
10	completa	7,5	12,5	43,5
11	incompleta	7,7	3,2	29,0
12	incompleta	6,6	9,5	52,6
13	incompleta	7,7	17,1	43,0
14	incompleta	6,8	9,0	55,9
15	incompleta	7,4	2,0	39,4
16	incompleta	6,6	7,0	43,7
17	incompleta	7,0	9,9	57,4
18	completa	6,9	11,4	35,0
19	incompleta	6,5	3,48	32,1
20	incompleta	7,9	1,49	23,0
21	completa	6,3	5,7	33,0
22	incompleta	6,9	3,7	27,6
23	completa	7,2	7,0	38,0
24	incompleta	7,2	3,43	21,14
25	incompleta	6,5	24,1	50,2
26	incompleta	7,1	6,6	46,2
27	incompleta	8,0	3,36	35,1
Média		7,33	7,17	33,22

Porcentagem de fragmentos com tamanho esperado para miRNAs na amostra de RNA total (inicial) e após o enriquecimento para RNAs pequenos (final). RIN = integridade do RNA.

Ao final da construção das bibliotecas de miRNA, o perfil das mesmas foi avaliado no Bioanalyzer e a porcentagem de miRNAs na biblioteca foi estimada mais uma vez. Em média, 35% da massa total das bibliotecas era composto por miRNAs (**Tabela 11**). Novamente, observou-se uma grande variabilidade na porcentagem de miRNA entre as bibliotecas dos pacientes (23 a 56,5%) e, ainda, essa porcentagem foi menor do que a encontrada para as duas linhagens já que a porcentagem de fragmentos provenientes de RNAs degradados é maior para as bibliotecas de pacientes. Cabe lembrar que todas estas estimativas de porcentagens de miRNAs nas bibliotecas, são bastante especulativas e que a real porcentagem de miRNAs só foi verificada após o sequenciamento de cada biblioteca.

Como descrito anteriormente, o perfil da biblioteca das linhagens apresentou um pico menor referente aos adaptadores (90-100pb) e um pico maior referente aos miRNAs ligados aos adaptadores (115-130pb, Figura 19). Já as bibliotecas de pacientes, além desses 2 picos, apresentaram um terceiro pico com 135pb em média (**Figura 28**). Este pico se refere à ligação dos adaptadores a moléculas resultantes da degradação de RNAs transportadores ou ribossômicos. Considerando-se o menor RIN das amostras de pacientes, já seria esperado encontrar tal pico nas bibliotecas dos mesmos. Entretanto, não é possível remover tais fragmentos, nem mesmo através de purificação em gel de agarose, devido ao tamanho muito próximo ao dos fragmentos de interesse.

Tabela 11 - Avaliação no Bioanalyzer da porcentagem de miRNAs presentes nas bibliotecas de pacientes do grupo de treinamento (faixa de fragmentos de 115-130pb).

Paciente	Resposta ao tratamento	% miRNA na biblioteca
1	incompleta	28,5
2	incompleta	52,5
3	incompleta	47,5
4	completa	31,6
5	incompleta	28,5
6	incompleta	38,3
7	completa	41,6
8	completa	45,8
9	incompleta	39,2
10	completa	39,1
11	incompleta	34,0
12	incompleta	32,6
13	incompleta	26,0
14	incompleta	42,8
15	incompleta	49,3
16	incompleta	49,0
17	incompleta	23,3
18	completa	40,6
19	incompleta	29,1
20	incompleta	31,0
21	completa	45,5
22	incompleta	42,6
23	completa	56,5
24	incompleta	37,1
25	incompleta	42,4
26	incompleta	25,2
27	incompleta	29,6
Média		35,00

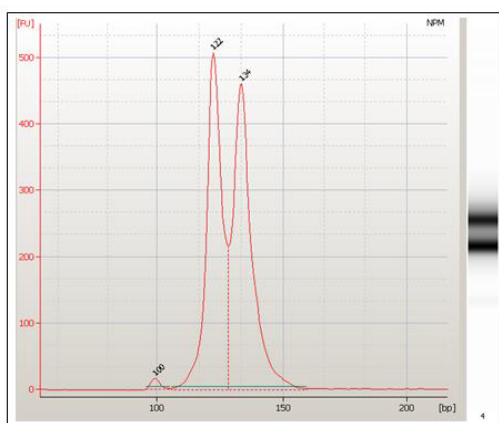


Figura 28 - Perfil de tamanho dos fragmentos de uma biblioteca de miRNA de amostra de paciente avaliado em uma corrida de chip de DNA High Sensitivity no Bioanalyzer. O tamanho dos fragmentos, em pares de bases (bp), está representado no eixo x em relação à sua abundância (fluorescência - FU) no eixo y. O primeiro pico apresenta 100pb, o segundo 122pb e o terceiro 134.

As bibliotecas de miRNA foram sequenciadas no SOLiD4, sendo geradas, em média, 40 milhões de sequências por amostra (variando de 28,8 a 60,8 milhões, **Tabela 12**). Assim como foi feito para as linhagens, foram estudados apenas os miRNAs conhecidos utilizando um programa disponível comercialmente, o CLC Genomics Workbench (CLCBio). Inicialmente as sequências obtidas foram mapeadas contra o miRBase (versão 18, contendo 1921 miRNAs humanos maduros) e, em média, 5,5 milhões de sequências corresponderam a miRNAs. Como esperado, porcentagem de sequências correspondentes a miRNAs conhecidos variou bastante entre as diferentes amostras (7 a 28%, **Tabela 12**). Embora esta porcentagem tenha variado muito, o número de miRNAs identificados em cada amostra variou em uma escala menor (630 a 946) e não foi dependente da quantidade de sequências geradas para cada amostra.

Tabela 12 - Sequências obtidas no SOLiD4 a partir das bibliotecas de miRNAs dos pacientes do grupo de treinamento.

Paciente	Resposta ao tratamento	Sequências geradas	Sequências mapeadas miRNAs	Porcentagem sequências mapeadas miRNAs	Número miRNAs expressos
1	incompleta	50.172.700	5.486.141	10,93	820
2	incompleta	29.444.051	5.222.439	17,74	630
3	incompleta	30.222.945	4.638.707	15,35	908
4	completa	54.158.517	7.451.686	13,76	854
5	incompleta	53.618.746	4.507.681	8,41	885
6	incompleta	31.091.042	3.936.602	12,66	769
7	completa	42.737.245	4.839.742	11,32	886
8	completa	26.184.560	5.237.994	20,00	800
9	incompleta	26.934.474	3.903.521	14,49	836
10	completa	31.528.426	4.381.148	13,90	676
11	incompleta	37.032.364	5.546.587	14,98	894
12	incompleta	33.610.904	3.926.311	11,68	774
13	incompleta	54.550.819	3.614.647	6,63	942
14	incompleta	32.084.705	5.926.807	18,47	770
15	incompleta	31.452.650	7.465.445	23,74	901
16	incompleta	26.986.387	6.149.544	22,79	910
17	incompleta	45.907.246	2.815.616	6,13	651
18	completa	30.073.433	4.623.558	15,37	797
19	incompleta	44.521.292	2.362.828	5,31	797
20	incompleta	41.177.074	3.416.683	8,30	946
21	completa	26.339.749	4.405.923	16,73	821
22	incompleta	30.194.181	6.470.087	21,43	838
23	completa	42.780.410	2.449.337	5,73	849
24	incompleta	33.446.117	3.045.721	9,11	666
25	incompleta	29.357.756	2.699.764	9,20	829
26	incompleta	51.372.267	3.257.262	6,34	894
27	incompleta	44.310.006	2.297.596	5,19	662
Média		37.455.188	4.447.384	12,80	815

4.2.3 Identificação de miRNAs diferencialmente expressos

Os miRNAs diferencialmente expressos entre os pacientes de resposta completa e incompleta foram obtidos através do EdgeR, programa desenvolvido para os dados de contagem de sequências obtidos nos equipamentos de última geração (ROBINSON 2010). A partir dos dados brutos de número de sequências obtidas para cada miRNA, o programa normaliza, modela a dispersão e a variância dos dados e por fim calcula a probabilidade da expressão de cada miRNA ser estatisticamente diferente

entre os grupos de amostras que estão sendo comparadas. Para uma análise estatística mais robusta, os miRNAs pouco expressos foram excluídos e apenas os miRNAs com expressão maior que 20cpm em pelo menos 7 amostras foram utilizados, correspondendo a um total de 331 miRNAs. Foram considerados miRNAs diferencialmente expressos aqueles que apresentaram $p < 0,01$ e $FDR < 0,05$. Como são realizados múltiplos testes espera-se por chance que alguns candidatos que tenham diferença de expressão significativa ($p < 0,05$) sejam falso positivos. Portanto, o valor de p foi corrigido pelo método de BENJAMINI e HOCHBERG (1995) que calcula a taxa de descoberta falsa (FDR) dentre os candidatos com p significativo (BENJAMINI e HOCHBERG 1995). Assim, considerando-se um $FDR < 0,05$ espera-se que no máximo 5% dos miRNAs reportados como estatisticamente significativos sejam falso positivos. Foram encontrados 4 miRNAs diferencialmente expressos entre os grupos, sendo 3 mais expressos nos pacientes de resposta completa (miR-21-5p, miR-1246 e miR1290-3p) com uma diferença média de expressão variando de 3,35 a 3,83 vezes maior nestes pacientes em relação aos de resposta incompleta (**Tabela 13**). Apenas um miRNA foi mais expresso nos pacientes com resposta incompleta (miR-205-5p), com uma diferença média de expressão de 0,002 vezes nos pacientes de resposta completa em relação aos de resposta incompleta (**Tabela 13**). A dispersão dos valores de expressão (cpm) dentro de cada grupo pode ser visualizada na **Figura 29**.

Tabela 13 - miRNAs diferencialmente expressos (FDR<0,05) no grupo de treinamento entre os pacientes de resposta completa e de resposta incompleta.

miRNA	Média expressão do grupo de resposta completa (cpm)	Média expressão do grupo de resposta incompleta (cpm)	Razão (completa/incompleta)	FDR
miR-205-5p	3,67	2.099,32	0,002	0,002
miR-21-5p	106.020,55	31.438,53	3,37	0,028
miR-1246	65,58	19,60	3,35	0,028
miR-1290-3p	115,14	30,02	3,83	0,028

Dados obtidos no SOLiD4 e analisados no EdgeR. cpm = número de sequências por milhão; FDR = taxa de descoberta falsa.

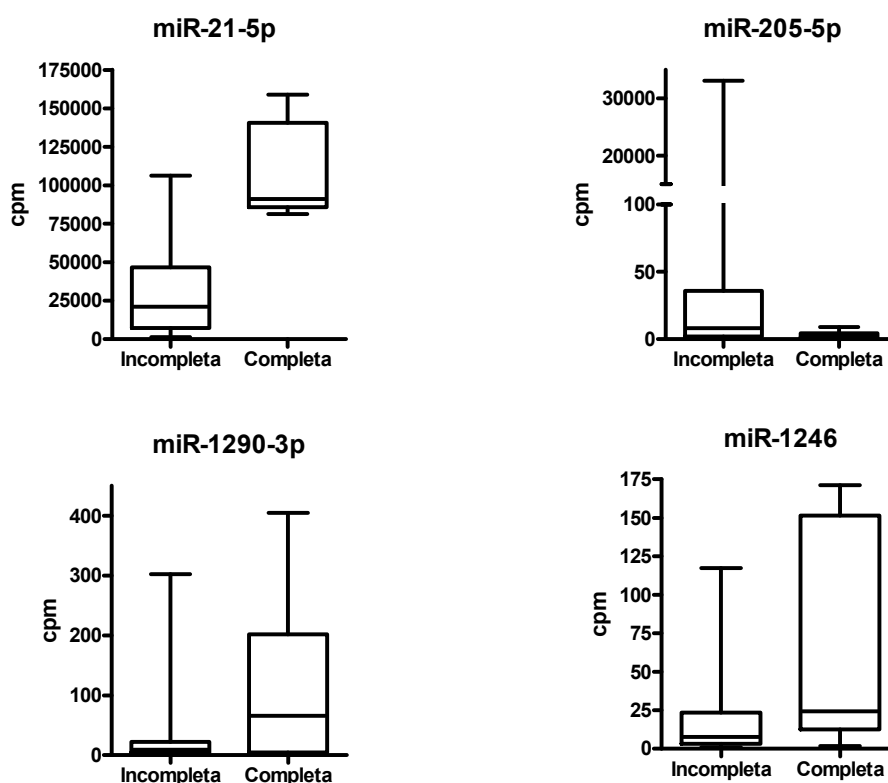


Figura 29 - Expressão (cpm) dos miRNAs diferencialmente expressos identificados no EdgeR (FDR<0,05) comparando os pacientes de resposta completa (n=7) e incompleta (n=20). cpm = número de sequências por milhão.

Após a identificação dos miRNAs diferencialmente expressos entre os pacientes de resposta completa e incompleta, foi avaliada a capacidade desses miRNAs em prever a resposta ao tratamento QRTn. No pacote de análises do EdgeR é possível construir um gráfico para avaliar a similaridade das amostras considerando-se um menor número de dimensões. A maior parte da informação contida em 4 variáveis (número de miRNAs diferencialmente expressos) pode ser representada em um gráfico de 2 dimensões, demonstrando a variabilidade das amostras em relação à expressão desses miRNAs. A primeira dimensão (eixo x) é a direção que melhor explica a variação presente nos dados, a segunda dimensão (eixo y) é a direção seguinte, não correlacionada com a primeira. A **Figura 30** mostra que existe uma boa separação entre as amostras de resposta completa e incompleta. No entanto, há 3 amostras de resposta incompleta que se aproximam mais daqueles de resposta completa, são elas as amostras 15, 19 e 26.

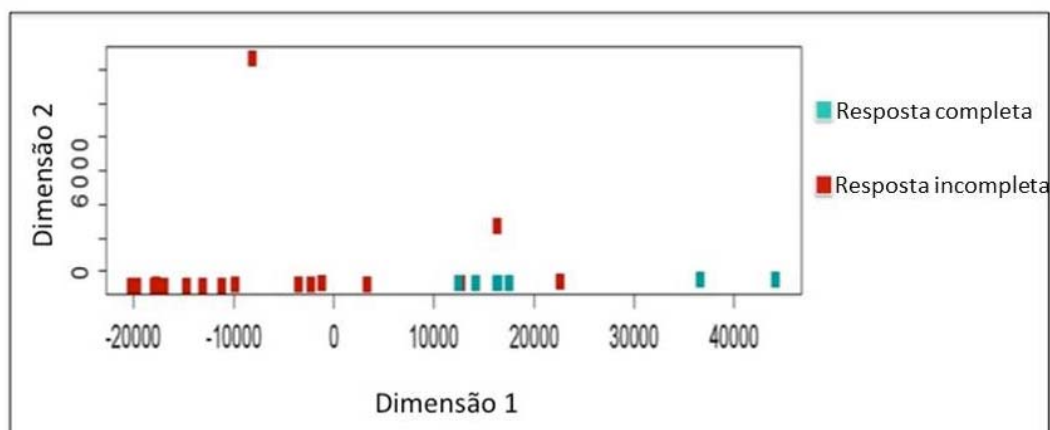


Figura 30 - Gráfico de escala multidimensional para visualizar a similaridade das amostras. Figura gerada no EdgeR considerando os 4 miRNAs diferencialmente expressos ($FDR < 0,05$). Dimensão 1 (eixo x) é a direção que melhor explica a variação presente nos dados e a dimensão 2 (eixo y) é a segunda melhor, não correlacionada com a primeira. Em vermelho estão representadas as amostras de resposta incompleta e em azul as de resposta completa. As 3 amostras em vermelho que ficaram em meio às amostras azuis correspondem aos pacientes 15, 19 e 26.

Foi realizada uma segunda análise para explorar a similaridade entre as amostras. A clusterização hierárquica agrupa amostras similares em *clusters* (grupos). O dendograma gerado mostra a proximidade das amostras baseando-se na expressão dos 4 miRNAs que apresentaram diferença de expressão significativa. Essa clusterização separou as amostras em dois grandes grupos, tendo de um lado apenas amostras de pacientes com resposta incompleta e do outro lado todas as amostras dos pacientes com resposta completa, juntamente com 3 amostras de resposta incompleta (**Figura 31**). Estas são as mesmas 3 amostras de pacientes com resposta incompleta (15, 19 e 26) que aparecem em meio às amostras de resposta completa no gráfico de escala multidimensional (**Figura 30**). Também foi feita uma análise estatística de Bootstrap para avaliar a confiabilidade do

agrupamento das amostras. Nesta análise, o agrupamento é repetido 1000 vezes (randomizando as amostras selecionadas para cada agrupamento) e em vermelho está mostrada a porcentagem de vezes em que os ramos se agruparam da maneira mostrada na **Figura 32**, o que resultou em uma confiabilidade de 97%. Assim, todas estas análises revelaram que a expressão dos 4 miRNAs permite o agrupamento das amostras em 2 conjuntos: um deles contendo apenas amostras de pacientes com resposta incompleta, e o outro contendo todos os pacientes com resposta completa e ainda 3 amostras de pacientes com resposta incompleta.

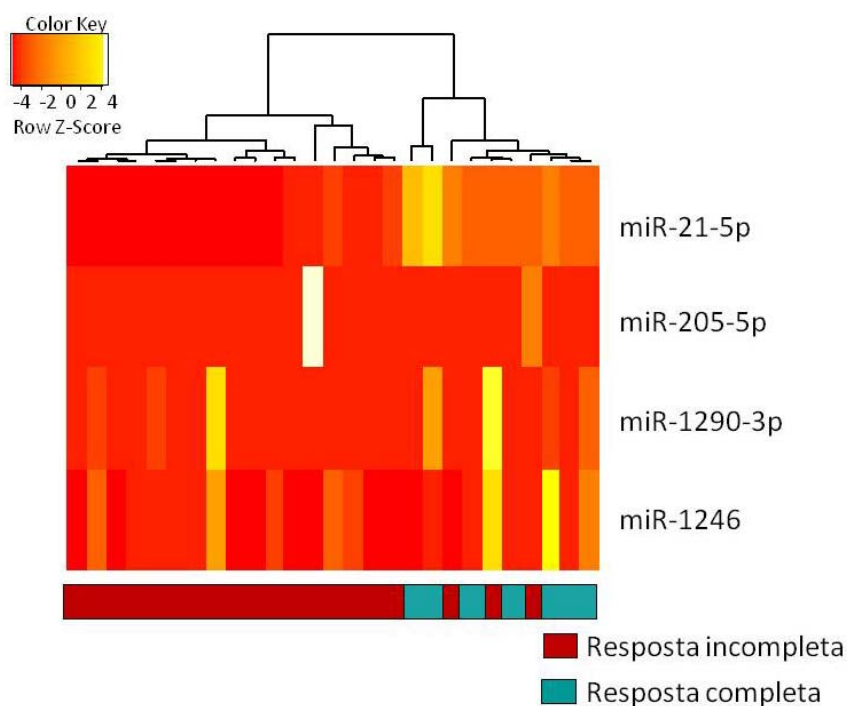


Figura 31 - Clusterização hierárquica das amostras de pacientes de resposta completa e incompleta com base na expressão dos 4 miRNAs identificados como diferencialmente expressos. As 3 amostras de resposta incompleta (em vermelho) que foram agrupadas com as amostras dos completos (em azul) correspondem aos pacientes 15, 19 e 26.

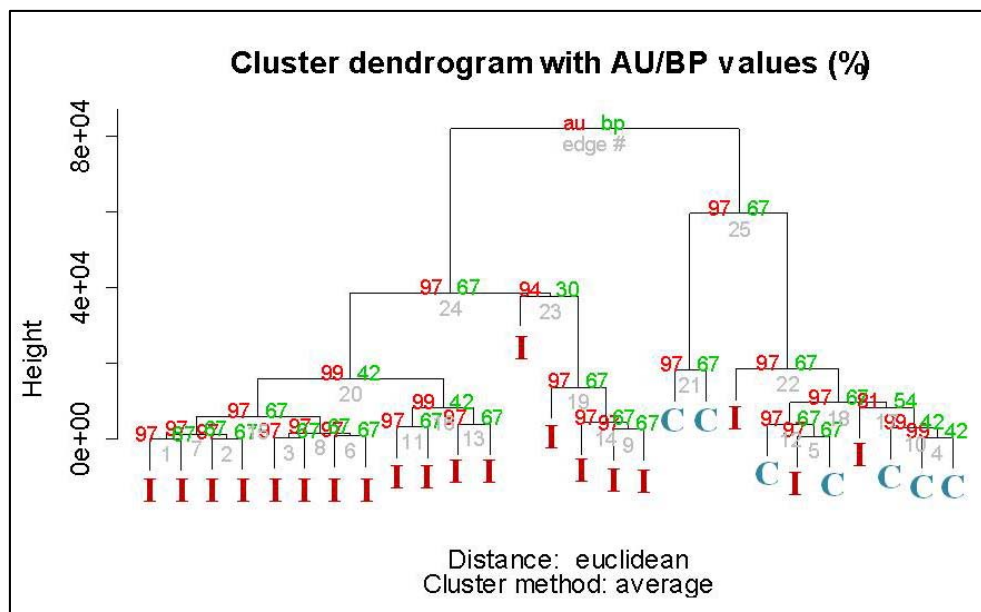


Figura 32 - Análise de Bootstrap das amostras de pacientes com resposta completa ("C") e incompleta ("I") com base na expressão dos 4 miRNAs identificados como diferencialmente expressos. A confiabilidade de cada agrupamento está mostrada em vermelho.

Por último, também foi avaliado o potencial de cada miRNA ser utilizado individualmente como um biomarcador preditivo. Calculou-se a área sob a curva (AUC, do inglês *area under the curve*) ROC (do inglês *receiver operating characteristic*) para estimar a eficiência geral em classificar corretamente cada paciente de acordo com a resposta ao tratamento (**Figura 33**). O valor da AUC foi 94%, 63%, 70% e 74% para miR-21-5p, miR-205-5p, miR-1290-3p e miR-1246, respectivamente. A partir da curva ROC foi possível estabelecer o ponto de corte que melhor separava os pacientes de resposta completa e incompleta. Para o miR-21-5p, o valor de corte de expressão estabelecido foi de 70.000 cpm, alcançando 100% de sensibilidade e 85% de especificidade para este grupo de treinamento. O miR-205-5p apresentou 86% de sensibilidade e 55% de especificidade com

um valor de corte de 6 cpm. Já o miR-1290-3p apresentou 71% de sensibilidade e 75% de especificidade com um valor de corte de 22 cpm. Por último, o miR-1246 teve 86% de sensibilidade e 65% de especificidade com um valor de corte de 11 cpm. Visto a boa acurácia para prever a resposta à QRTn, em especial do miR-21-5p, esses 4 miRNAs foram considerados como potenciais candidatos a biomarcadores preditivos.

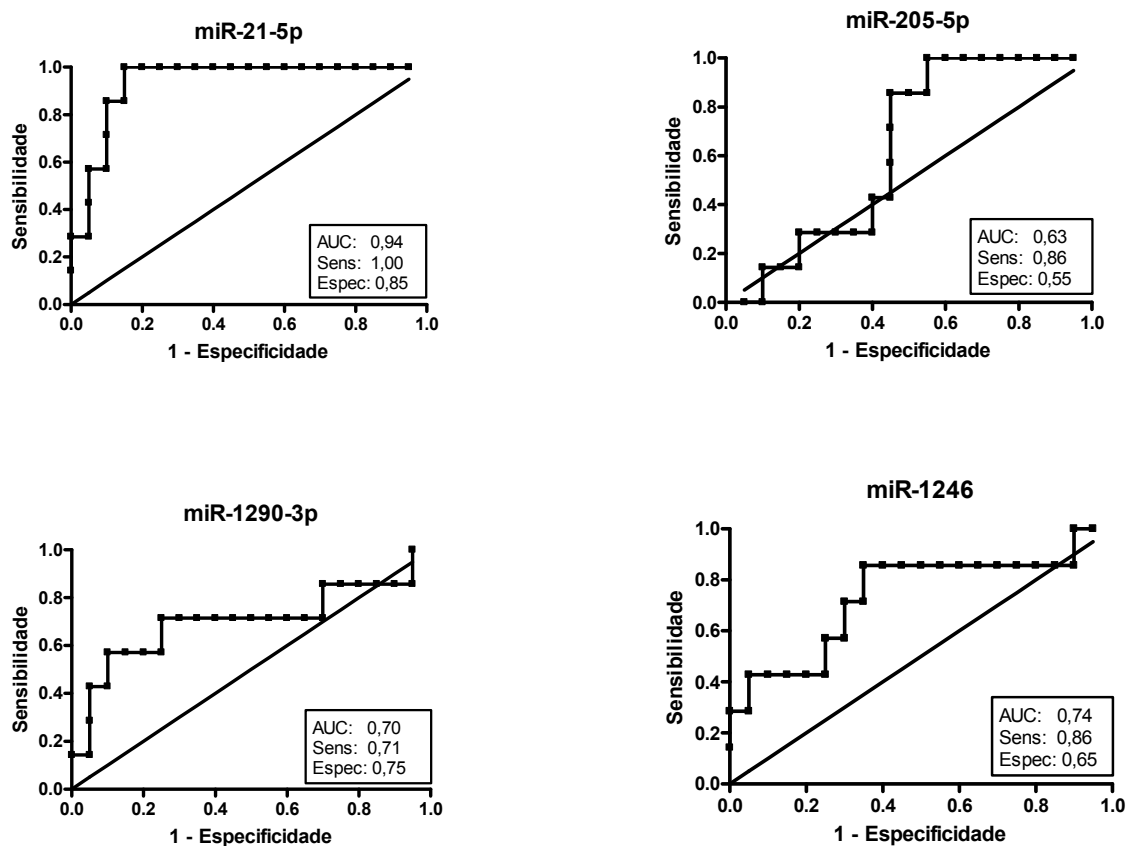


Figura 33 - Curva ROC para determinar a acurácia da predição de resposta a QRTn baseada na expressão de miRNAs. Para cada miRNA foram calculados a área sob a curva (AUC), sensibilidade e especificidade após identificar o melhor valor de corte para a separação dos grupos.

4.2.4 Validação técnica da expressão de miRNAs por qPCR

Para aumentar a confiabilidade dos resultados obtidos por sequenciamento, o primeiro passo realizado foi uma validação da técnica de sequenciamento. Para tanto, a expressão dos 4 miRNAs identificados como diferencialmente expressos foi avaliada por uma segunda metodologia, o qPCR. A análise foi feita para todas as amostras do grupo de treinamento sequenciadas e que ainda dispunham de RNA, o que totalizou em 4 pacientes de resposta completa e 9 de resposta incompleta. De modo geral, os experimentos de qPCR apresentaram baixa eficiência e especificidade. Observou-se uma baixa especificidade dos primers principalmente para as amostras que foram detectadas com baixa expressão no sequenciamento. Para estes casos a curva de dissociação após a reação de amplificação apresentou mais de um pico, revelando a amplificação de produtos inespecíficos ou a formação de dímeros de primer. Mesmo após diferentes tentativas e modificações nas condições de amplificação, as reações continuaram apresentando tal inespecificidade (Figura 34). Como os miRNAs são moléculas pequenas não é possível testar diferentes combinações de primers com a finalidade de melhorar as condições da reação e, por isso, a eficiência de amplificação e confiabilidade dos resultados pode variar muito para cada miRNA.

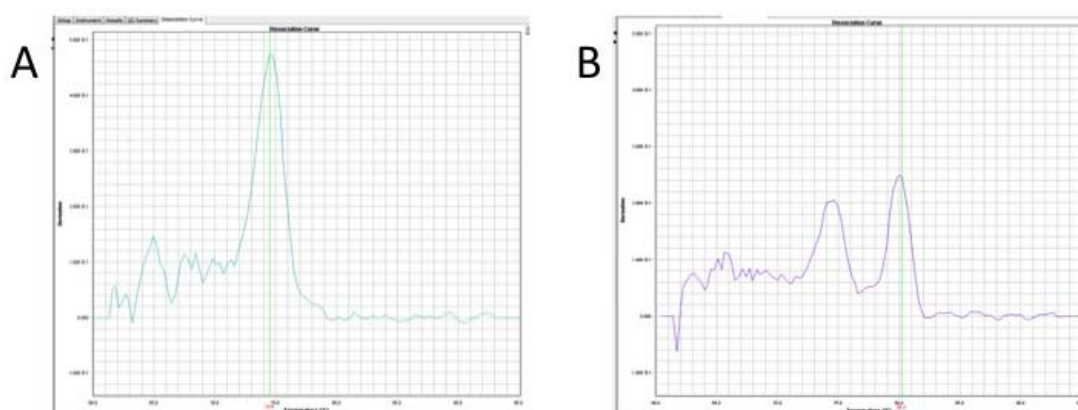


Figura 34 - Exemplos de curvas de dissociação obtidas por qPCR. A) Curva de dissociação esperada, com pico único, representando amplificação específica. B) Curva de dissociação com pico inespecífico e que não deve ser utilizada para a avaliação da expressão do transcrito.

Para comparar a expressão de um determinado miRNA entre as diferentes amostras, a técnica de qPCR depende da normalização por um gene endógeno com expressão constante na condição estudada. A escolha do gene normalizador foi feita com base na literatura, e embora não haja um consenso de qual normalizador é o mais constante, foi escolhido o RNU6b (um RNA nuclear pequeno), o qual já foi utilizado em diversos trabalhos de câncer colorretal (PIEPOLI et al. 2012; NG et al. 2009; SCHETTER et al. 2008). Ainda assim, a expressão do RNU6b variou muito entre as amostras analisadas, o que pode prejudicar a análise de expressão diferencial. Foi testado também um outro normalizador bastante utilizado na literatura (SNORD68, dados não apresentados), mas cuja expressão variou ainda mais, e portanto foram mantidos os resultados obtidos com o RNU6b. Vale ressaltar que todas as quantificações e reações foram realizadas simultaneamente e partindo da mesma quantidade de RNA, e portanto, não

sendo esperada uma alta variação da expressão de RNU6b entre as amostras (**Tabela 14**). Dado este resultado, seria necessário um estudo maior para identificar um gene normalizador com expressão constante neste grupo de amostras.

Tabela 14 - Ciclo de amplificação (Ct) do gene normalizador RNU6b.

Paciente	Ct RNU6b
1	23,1
5	22,7
7	24,2
8	21,4
10	22,8
13	23,9
15	22,9
16	22,3
18	23,5
19	22,8
22	23,2
25	22,6
26	23,1

Os resultados do qPCR foram analisados mesmo sabendo de todas as limitações apresentadas. Conforme pode ser verificado na **Figura 35**, comparando a mediana da expressão relativa entre os pacientes de resposta completa e os de resposta incompleta, 3 miRNAs (miR-21-5p, miR-1290-3p e miR-1246) apresentaram a mesma tendência de expressão do que foi observado nos dados de sequenciamento. Entretanto, não foi verificada diferença estatística significativa entre os grupos de pacientes para nenhum dos 4 miRNAs. Os resultados obtidos por qPCR são difíceis de serem interpretados devido às limitações técnicas já descritas. Além disso, a análise foi feita com um baixo número de amostras e observou-se uma alta

variabilidade entre as mesmas. Cabe notar que não foram utilizadas todas as 27 amostras nesta verificação por qPCR devido à falta da maioria delas, uma vez que por se tratar de amostras de biópsias, a massa de RNA normalmente obtida foi muito baixa e a maioria delas foi integralmente utilizada na construção das bibliotecas de miRNAs para o sequenciamento em larga escala.

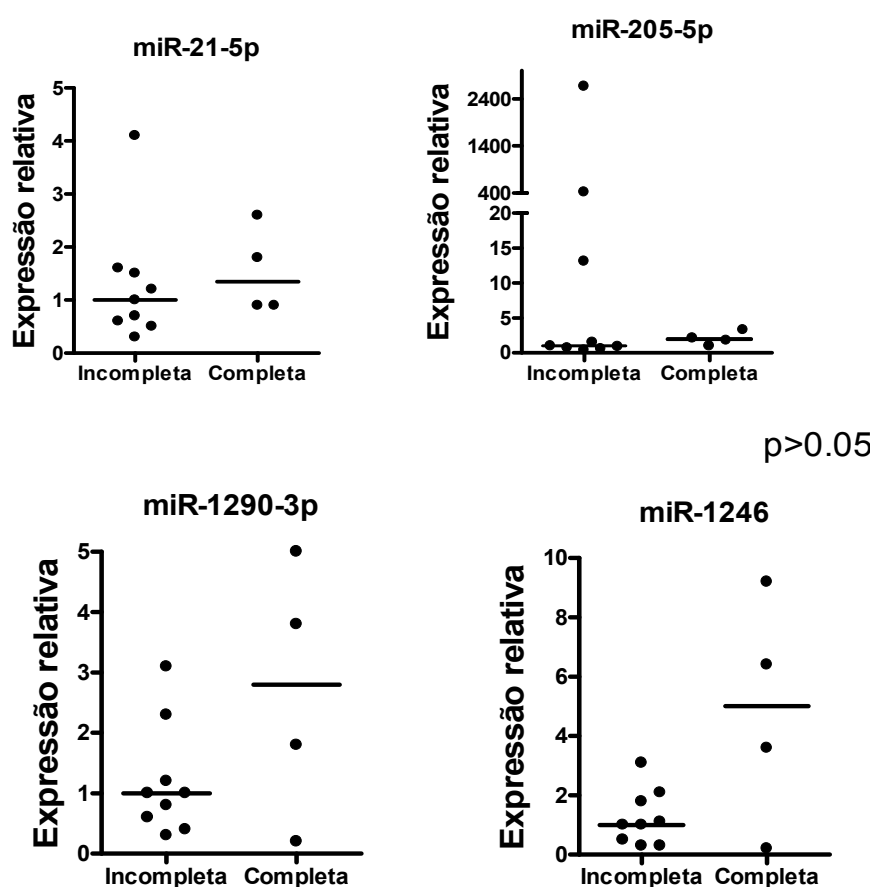


Figura 35 - Expressão relativa por qPCR dos miRNAs identificados como diferencialmente expressos através do sequenciamento de última geração. Não há diferença estatística significativa entre os pacientes de resposta incompleta e completa (teste Mann-Whitney), mas observa-se a mesma tendência obtida por sequenciamento, com exceção do miR-205-5p. O eixo y representa o valor de expressão do miRNA normalizado pelo RNU6b em relação à mediana do grupo de resposta incompleta ($2^{-\Delta\Delta CT}$). Cada paciente está indicado por um ponto e a barra horizontal representa a mediana do grupo.

Por fim, para ilustrar como estas limitações técnicas poderiam impactar os resultados de validação da expressão diferencial, foi feita uma análise de correlação entre os resultados do sequenciamento e do qPCR. Dados da literatura mostram que há uma boa correlação geral entre os dados de expressão obtidos por essas duas metodologias, no entanto, a correlação não é perfeita e varia muito para cada miRNA estudado (HAMFJORD et al. 2012; SCHULTE et al. 2010). A partir do qPCR pode ser obtido o valor de ΔCt , que é a diferença do Ct do miRNA de interesse em relação ao normalizador, sendo que o Ct, do inglês threshold cycle, é o ciclo em que cada curva de amplificação atravessa o limiar de detecção – threshold - o qual é definido arbitrariamente durante a fase exponencial da amplificação. Como a PCR amplifica as moléculas exponencialmente, Rahmann e cols demonstraram que os valores negativos de ΔCt se correlacionam linearmente com os valores de expressão obtidos por sequenciamento em escala log2 (RAHMANN et al. 2013). Neste mesmo trabalho foi feita a análise de correlação da expressão para 202 miRNAs em 10 amostras e observou-se que para 75% desses miRNAs havia uma correlação significativa, que variou de 0,655 a 0,732. Não foi verificado um padrão que pudesse explicar os casos de baixa correlação, sendo que não havia uma característica restrita às sequências desses miRNAs e nem mesmo uma dependência do seu nível de expressão.

Assim, o mesmo exercício foi aplicado para os dados aqui obtidos. Conforme pode ser verificado na Figura 36, para 3 miRNAs houve uma correlação positiva e estatisticamente significativa entre os dados de

sequenciamento e de qPCR, sendo 0,62 para o miR-205-5p, 0,60 para miR-1290-3p e 0,59 para miR-1246. Já para o miR-21-5p não foi observada uma correlação estatisticamente significativa. Sendo assim, através de qPCR não foi possível validar os dados obtidos por sequenciamento, em grande parte devido às limitações técnicas apresentadas acima.

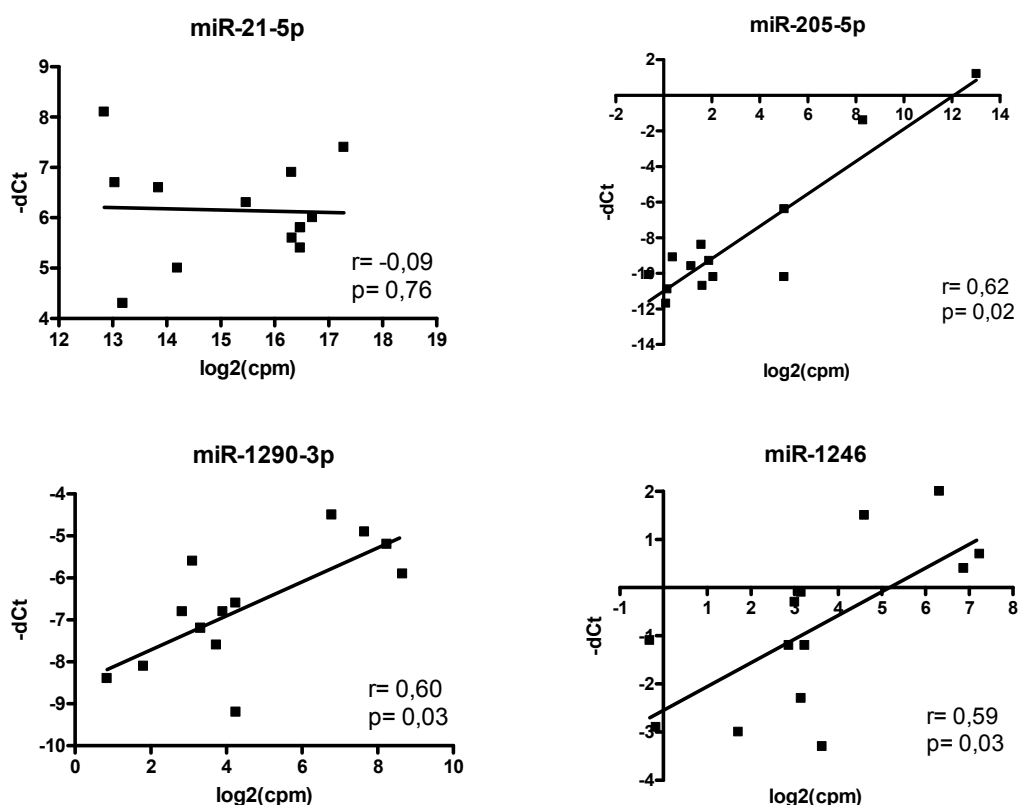


Figura 36 - Correlação dos dados de expressão de miRNAs obtidos por qPCR e por sequenciamento. O gráfico apresenta, no eixo x, os dados de expressão obtidos por sequenciamento (normalizados por cpm e em escala log₂) e no eixo y os valores negativos de Δ Ct (Ct do miRNA normalizado pelo Ct do RNU6b) obtidos por qPCR. A correlação de Spearman (r) e valor de p estão mostrados para cada miRNA. Cada ponto representa um paciente e a linha traça a tendência linear dos dados.

4.2.5 Validação experimental da expressão de miRNAs por sequenciamento

O verdadeiro potencial desses 4 miRNAs como biomarcadores só poderia ser confirmado através da avaliação da expressão dos mesmos em um grupo independente de amostras. Sendo assim, novas amostras foram obtidas e formaram o grupo de validação, o qual continha 7 amostras de pacientes de resposta completa e 5 de resposta incompleta. Devido à dificuldade de interpretar os resultados dos experimentos de qPCR, que tiveram baixa eficiência e especificidade para alguns primers e alta variabilidade do gene normalizador, essas novas amostras também foram utilizadas para construção de bibliotecas de miRNAs e sequenciadas. O número de sequências geradas foi semelhante ao sequenciamento das amostras do grupo de treinamento. Em média, foram geradas 52 milhões de sequências para cada paciente, sendo que 11,6% corresponderam a miRNAs conhecidos (**Tabela 15**).

Tabela 15 - Sequências obtidas no SOLiD4 a partir das bibliotecas de miRNAs dos pacientes do grupo de validação.

Paciente	Resposta ao tratamento	Sequências geradas	Sequências mapeadas miRNAs	Porcentagem de sequências mapeadas miRNAs	Número miRNAs expressos
28	completa	50.715.095	4.841.085	9,55	818
29	completa	19.061.826	2.268.017	11,90	649
30	incompleta	59.729.745	2.360.510	3,95	714
31	completa	45.731.316	2.289.984	5,01	809
32	completa	33.644.141	4.348.408	12,92	889
33	incompleta	59.101.716	4.430.414	7,50	936
34	incompleta	93.193.034	5.929.896	6,36	912
35	completa	41.123.068	3.264.029	7,94	840
36	incompleta	23.367.580	2.565.471	10,98	742
37	completa	34.092.250	4.760.712	13,96	843
38	incompleta	133.185.797	5.155.644	3,87	837
39	completa	32.079.436	6.146.284	19,16	912
40	recidiva	30.518.047	5.638.091	18,47	870
41	recidiva	30.606.777	3.436.187	11,23	788
42	recidiva	32.985.246	4.478.279	13,58	956
43	recidiva	31.271.842	4.339.892	13,88	897
Média		46.900.432	4.140.806	10,64	838

Inicialmente, foi avaliado se o perfil de expressão dos 4 miRNAs identificados como diferencialmente expressos no grupo de treinamento continuavam apresentando a capacidade de agrupar as amostras do grupo de validação com base na resposta ao tratamento. Para tanto, a similaridade entre essas novas amostras foi avaliada através da clusterização hierárquica e análise de Bootstrap. Conforme pode ser verificado, os pacientes do grupo de validação não foram agrupados corretamente de acordo com sua resposta ao tratamento (**Figura 37**, **Figura 38**). Isto sugere que a expressão

diferencial desses miRNAs era característico apenas das amostras do grupo de treinamento.

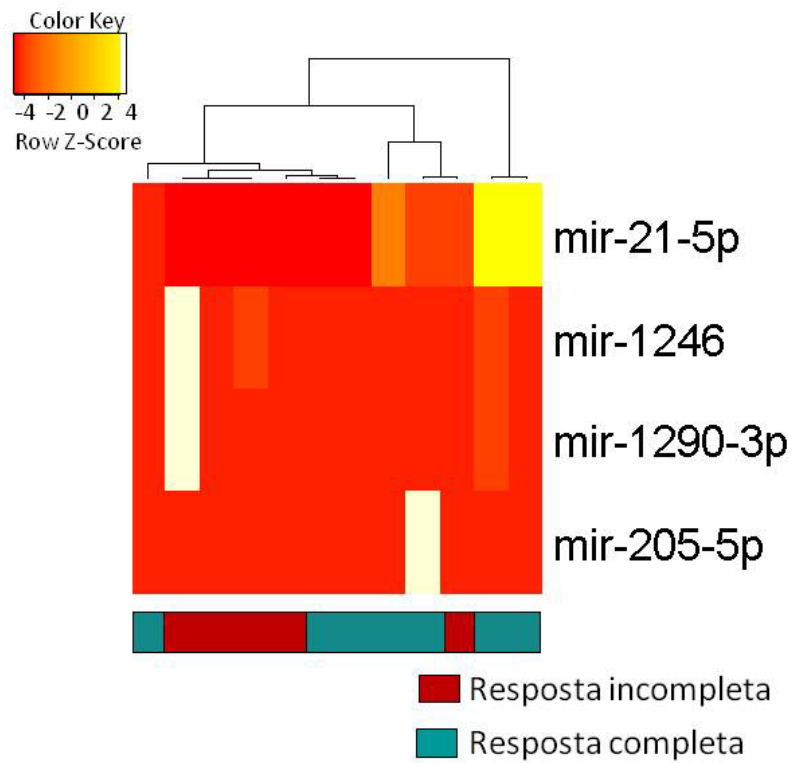


Figura 37 - Clusterização hierárquica das amostras do grupo de validação com base na expressão dos 4 miRNAs identificados como diferencialmente expressos no grupo de treinamento. Amostras de pacientes com resposta completa em vermelho e incompleta em azul.

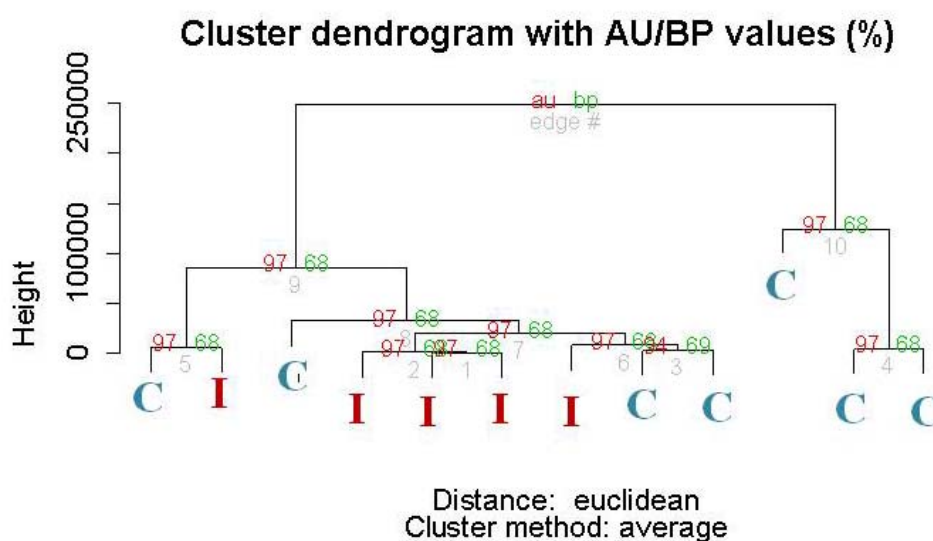


Figura 38 - Análise de Bootstrap do grupo de validação dos pacientes com resposta completa ("C") e incompleta ("I") com base na expressão dos 4 miRNAs identificados como diferencialmente expressos no grupo de treinamento. A confiabilidade de cada agrupamento está mostrada em vermelho.

Embora os dados de expressão destes miRNAs não tenham agrupado as amostras do grupo de validação da maneira esperada, ainda poderia haver diferença de expressão de algum dos 4 miRNAs comparando-se pacientes com resposta completa e incompleta. Para responder essa pergunta, foi realizada a análise de expressão diferencial no EdgeR somente considerando-se os dados de sequenciamento dos pacientes do grupo de validação. Assim, como observado no grupo de pacientes de treinamento, a expressão de miR-21-5p foi significativamente maior nos pacientes de resposta completa comparando-se aos pacientes de resposta incompleta ($p < 0,01$). Embora tenha sido observado uma diferença estatística significativa ($p < 0,01$), o FDR calculado foi maior do que 0,05. Os demais miRNAs não apresentaram diferença de expressão entre os pacientes com

resposta completa e incompleta e portanto, foram considerados como não validados. Este resultado reforça o potencial do miR-21-5p como biomarcador preditivo em pacientes com câncer de reto, o qual já tinha apresentado a maior acurácia para a predição de resposta ao tratamento no grupo de treinamento.

Como descrito anteriormente, os dados de expressão dos 4 miRNAs diferencialmente expressos no grupo de treinamento foram utilizados para estabelecer os valores de corte para separação dos pacientes de resposta completa e incompleta com base nas curvas ROC construídas. Posteriormente, com base nos valores de corte de cada miRNA, os 12 pacientes do grupo de validação foram classificados como resposta completa ou incompleta, permitindo assim uma estimativa mais fiel da acurácia de cada miRNA como biomarcador preditivo.

Para os 3 miRNAs que não apresentaram expressão diferencial no grupo de validação, já seria esperado encontrar valores baixos de sensibilidade e especificidade. De fato, para o miR-205-5p o valor de sensibilidade foi de 86% (6/7) e o valor de especificidade foi 0 (0/5). Já a miR-1290-3p apresentou 80% (4/5) de especificidade, o miR-1246 60% (3/5) de especificidade e ambos tiveram 14% (1/7) de sensibilidade. Estes resultados de validação revelaram que os 3 miRNAs (miR-205-5p, miR-1290-3p, miR-1246) não são bons biomarcadores de resposta ao tratamento QRTn em pacientes com câncer de reto.

Já o miR-21-5p apresentou um resultado interessante na validação. Utilizando-se o valor de corte de 70.000 cpm estabelecido pela curva ROC

(Figura 33), 57% (4/7) dos pacientes com resposta completa foram corretamente classificados e 80% (4/5) dos pacientes com resposta incompleta foram corretamente classificados. Embora um valor de sensibilidade de 57% ainda pareça baixo vale ressaltar que atualmente não há teste clínico ou de imagem eficiente para garantir a ausência de células tumorais após a QRTn e, por isso, na maioria das instituições independente da resposta ao tratamento todos os pacientes com tumor de reto passam por uma cirurgia radical. Este resultado mostra que mais da metade dos pacientes de resposta completa poderiam ser identificados e oferecidos tratamentos alternativos à cirurgia radical. Também é importante notar que a especificidade se manteve alta (80%) no grupo de validação. Embora, seja mais fácil identificar um paciente de resposta incompleta com tumor residual, estes não podem ser erroneamente classificados em um teste molecular e não serem submetidos à cirurgia. Ainda assim, cabe ressaltar que os pacientes com resposta incompleta possuem doença evidente após a QRTn e portanto, mesmo que o marcador preditivo indique uma resposta completa, o cirurgião certamente não deixaria de submeter estes pacientes à cirurgia.

O valor preditivo positivo também foi alto, 80% (4/5), ou seja, houve uma alta porcentagem de resultado verdadeiro entre todos os pacientes que tiveram exame positivo (miR-21-5p acima de 70.000cpm). Já o valor preditivo negativo foi mais baixo, 57% (4/7), mostrando que de todos os resultados negativos obtidos (miR-21-5p menor do que 70.000cpm) 57% eram realmente pacientes de resposta incompleta. Ainda é necessário estabelecer o real valor preditivo do miR-21-5p em um grupo maior de

pacientes. Se confirmado, esse teste molecular poderia ser utilizado para identificar mais da metade dos pacientes que apresentam resposta completa à QRTn e para os quais poderiam ser indicadas condutas alternativas à cirurgia radical.

Para melhor visualizar a diferença de expressão de miR-21-5p em relação à resposta ao tratamento QRTn, na **Figura 39** estão representadas todas as amostras sequenciadas neste estudo: as 27 amostras do grupo de treinamento e as 12 do grupo de validação. Neste total incluem 14 pacientes de resposta completa e 25 de resposta incompleta. Além disso, também foi avaliada a expressão do miR-21-5p em 4 amostras de pacientes que haviam sido excluídos do estudo após terem sido sequenciados por não apresentarem cRC sustentada. Estes 4 pacientes apresentaram inicialmente uma cRC e não foram operados logo após o término do tratamento de QRTn, mas posteriormente eles apresentaram uma recidiva local precoce (em tempo menor do que 16 meses). Este grupo de pacientes foi identificado como grupo de "recidiva".

Como esperado, observa-se uma diferença de expressão significativa entre todos os pacientes de resposta completa comparando com todos os pacientes de resposta incompleta (análise estatística no EdgeR com $FDR < 0,05$, **Figura 39**). Interessantemente, a expressão de miR-21-5p nas amostras dos pacientes que apresentaram recidiva precoce é semelhante à do grupo de pacientes de resposta incompleta, sendo inclusive estatisticamente diferente da expressão dos pacientes com resposta completa (**Figura 39**). Novamente, este resultado ilustra o potencial de uso

do miR-21-5p como marcador molecular no auxílio à conduta terapêutica oferecida aos pacientes com câncer de reto submetidos ao tratamento QRTn. Apesar dos pacientes com recidiva precoce terem apresentado inicialmente uma cRC, a detecção da baixa expressão do miR-21-5p na biópsia poderia ter auxiliado na correta identificação desses pacientes como de resposta incompleta, indicando que na verdade eles deveriam ter sido submetidos à cirurgia imediatamente após o período de descanso do tratamento.

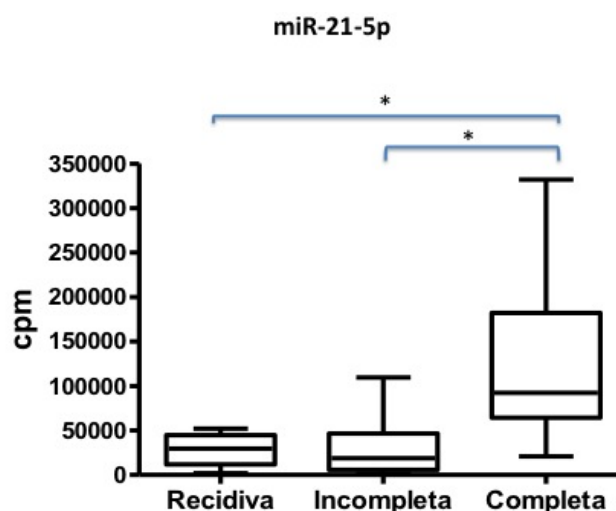


Figura 39 - Expressão (cpm) do miR-21-5p para todas as amostras sequenciadas. Os pacientes foram divididos em 3 grupos: pacientes com boa resposta clínica que recidivaram precocemente (n=4), com resposta incompleta (n=25) e completa (n=14) ao tratamento QRTn. Há uma diferença estatisticamente significativa dos grupos recidiva e incompleta quando comparados ao grupo de resposta completa (EdgeR, FDR<0,05). O valor de expressão está normalizado por cpm (número de sequências por milhão). *FDR<0,05

Diversos trabalhos têm demonstrado que o perfil de expressão de miRNAs pode ser utilizado na identificação de biomarcadores para o diagnóstico de diferentes tipos de câncer. Piepoli e cols identificaram uma

assinatura câncer-específica com 42 miRNAs, entre eles o miR-1290-3p, miR-1246 e miR-21-5p estavam super expressos no tumor colorretal em comparação ao tecido normal pareado (PIEPOLI et al. 2012). No nosso trabalho, as amostras de tumor de reto apresentaram expressão diferencial de miR-21-5p de acordo com a resposta ao tratamento QRTn. Sendo assim, esse miRNA pode representar não apenas um biomarcador para o diagnóstico do câncer de reto mas também um biomarcador preditivo.

Recentemente, foram publicados três trabalhos que identificaram assinaturas de miRNAs capazes de predizer a resposta ao tratamento neoadjuvante do câncer de reto (KHEIRELSEID et al. 2013; DELLA VITTORIA SCARPATI et al. 2012; SVOBODA et al. 2012). Entretanto, não há sobreposição entre o conjunto de miRNAs que compõe cada uma dessas 3 assinaturas e também com os 4 miRNAs identificados nesse trabalho. Além disso, diferentemente do nosso, dois trabalhos realizaram o estudo de expressão dos miRNAs através de microarray e um através de qPCR. Kheirelseid e cols obtiveram uma assinatura com 3 miRNAs (miR-16, miR-153, miR-590-5p) com acurácia de 100% na predição de resposta ao tratamento (KHEIRELSEID et al. 2013). No entanto, o estudo foi feito com um número pequeno de amostras, composto por apenas 8 pacientes de resposta incompleta e 4 de resposta completa ao tratamento. Além disso, não foram utilizadas amostras independentes para a validação do resultado obtido. No outro trabalho, foram estudados 373 miRNAs em 38 pacientes (sendo 9 de resposta completa e 29 incompleta) e uma assinatura de 13 miRNAs foi identificada (Della Vittoria Scarpati et al. 2012). A especificidade

foi calculada para cada miRNA, variando de 62 a 100% e a sensibilidade variou de 78 a 100%. Apesar dos resultados de sensibilidade e especificidade serem interessantes, a expressão diferencial dos miRNAs não foi validada em um grupo independente de amostras. O terceiro trabalho separou os pacientes com base no TRG, sendo 10 amostras de bons respondedores (TRG 1 e 2, de acordo com classificação de Mandard) e 10 amostras de maus respondedores (TRG 3 e 4) (SVOBODA et al. 2012). Este tipo de separação das amostras dificulta a comparação entre os estudos. De qualquer modo, foram estudados 667 miRNAs por qPCR e 8 deles apresentaram expressão diferencial e uma acurácia de 90% para predizer a resposta ao tratamento. No entanto, também não foi feita a validação dos resultados em um grupo independente.

As diferentes assinaturas obtidas em cada trabalho podem ser decorrentes das diferentes metodologias experimentais empregadas, do pequeno número de amostras estudadas nos trabalhos, bem como das diferentes definições de resposta ao tratamento QRTn para a separação dos pacientes de resposta completa e incompleta. Apesar do número de pacientes incluídos em nosso estudo também ter sido pequeno, foi o único entre os quatro a realizar a validação da expressão diferencial dos miRNAs candidatos em um grupo independente de amostras. Tal validação foi bastante importante para a eliminação de 3 miRNAs candidatos a marcadores preditivos e ainda para reforçar o potencial do miR-21-5p como um bom biomarcador preditivo de resposta à QRTn nos pacientes com

câncer de reto. Vale ressaltar que o real valor preditivo do miR-21-5p só pode ser confirmado após a avaliação em um maior grupo de pacientes.

4.2.6 Avaliação funcional do miR-21-5p

- **Genes alvo de miR-21-5p**

O miR-21-5p já foi descrito como superexpresso em diferentes tipos de tumores, incluindo o câncer colorretal (SCHMITZ et al. 2009; YAMAMICHI et al. 2009) e foi caracterizado como um importante onco-mir, podendo atuar na proliferação celular, apoptose e invasão (CHAN et al. 2005; REN et al. 2010). Estudos mais recentes também relacionaram a expressão de miR-21-5p com a resposta a diversos quimioterápicos. Especificamente para 5-FU, experimentos in vitro demonstraram uma correlação direta entre a superexpressão de miR-21-5p e a resistência a droga em linhagens derivadas de tumores colorretal e hepático, contrariamente ao que foi verificados nesse estudo (VALERI et al. 2010; TOMIMARU et al. 2010; FENG et al. 2012). Nestes trabalhos, a transfecção das células com oligo mimético de miR-21-5p reduziu a taxa de apoptose induzida por 5-FU. Por outro lado, a transfecção com anti-miR-21-5p aumentou a sensibilidade ao 5-FU, o que pôde ser parcialmente revertido com a transfecção de siRNAs que silenciavam os alvos de miR-21-5p. Como um miRNA possui diversos alvos, cada trabalho mostrou que esse efeito no aumento da resistência ao tratamento com 5-FU pode ser decorrente da repressão de um determinado alvo do miR-21-5p, tais como SPRY2, PTEN, PDCD4 e MSH2 (VALERI et al. 2010; TOMIMARU et al. 2010; FENG et al. 2012). Poucos estudos

correlacionaram a expressão de miR-21-5p à resposta ao tratamento em amostras de tumor de pacientes. Já foi demonstrada uma associação entre a alta expressão de miR-21-5p em adenocarcinoma de cólon avançado e pior taxa de sobrevida (SCHETTER et al. 2008; OUE et al. 2013). Além disso, em pacientes tratados com Qt adjuvante, a alta expressão de miR-21-5p estava associada a pior sobrevida global, sugerindo uma associação entre a alta expressão de miR-21-5p e pior resposta terapêutica. No entanto, a comparação entre esses trabalhos e os resultados obtidos nos pacientes com câncer de reto não é simples, uma vez que os mesmos investigaram a expressão de miR-21-5p no contexto de adenocarcinoma do cólon e Qt adjuvante.

Para tentar entender o conflito entre os dados obtidos nesse trabalho e os da literatura, primeiro foi avaliado se havia uma correlação inversa entre a expressão do miR-21-5p e a desses genes alvos já descritos. Para tanto, foram utilizados dados de sequenciamento do transcriptoma total disponíveis no laboratório (dados ainda não publicados) de algumas das mesmas amostras de pacientes incluídas nesse trabalho (6 de resposta completa e 13 de resposta incompleta). Interessantemente, para estes 4 alvos de miR-21-5p não foi observada uma diferença de expressão significativa entre os pacientes com resposta completa e incompleta (Figura 40). Corroborando esta informação, também não foi observada uma correlação inversa entre a expressão de miR-21-5p e a expressão dos genes *SPRY2*, *PTEN*, *PDCD4* e *MSH2* (**Figura 41**), indicando que talvez miR-21-5p não esteja regulando esses genes neste grupo de amostras.

Cabe ressaltar que os trabalhos que confirmaram estes 4 genes como alvos do miR-21-5p, o fizeram através de experimentos com vetores contendo a região 3' UTR dos mesmos associada ao gene da luciferase. Ao medir a atividade da luciferase nas células transfectadas com tais vetores, esta metodologia permite avaliar o efeito inibitório do miR-21-5p sobre o seu alvo, porém não é possível distinguir se o efeito inibitório se deve à degradação do mRNA ou da inibição da tradução da proteína. Deste modo, as análises de transcriptoma das amostras dos pacientes revelaram apenas que o miR-21-5p não está induzindo a degradação dos mRNAs destes 4 genes mas não é possível descartar a possibilidade de haver a repressão da tradução de suas proteínas (mesmo que este mecanismo seja o menos frequentemente utilizado pelos miRNAs em mamíferos). Esta resposta só poderia ser obtida através de experimentos de western-blot ou imunohistoquímica.

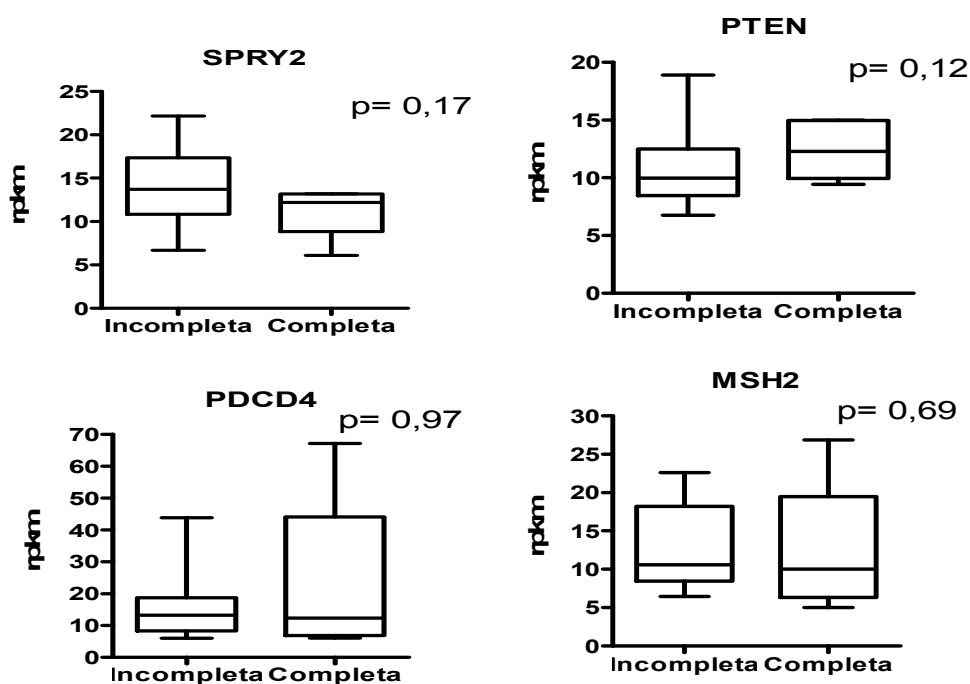


Figura 40 - Expressão de genes alvo de miR-21-5p nos pacientes de resposta incompleta (n=13) e de resposta completa (n=6). Dados obtidos por sequenciamento no SOLiD4 mostrando que não há diferença de expressão entre os 2 grupos (teste de Mann-Whitney). Expressão normalizada pelo número de sequências por milhão e pelo tamanho em kilobase do gene (rpkkm). Na literatura, a inibição desses 4 genes alvo de miR-21-5p está associada à resistência ao 5-FU *in vitro*.

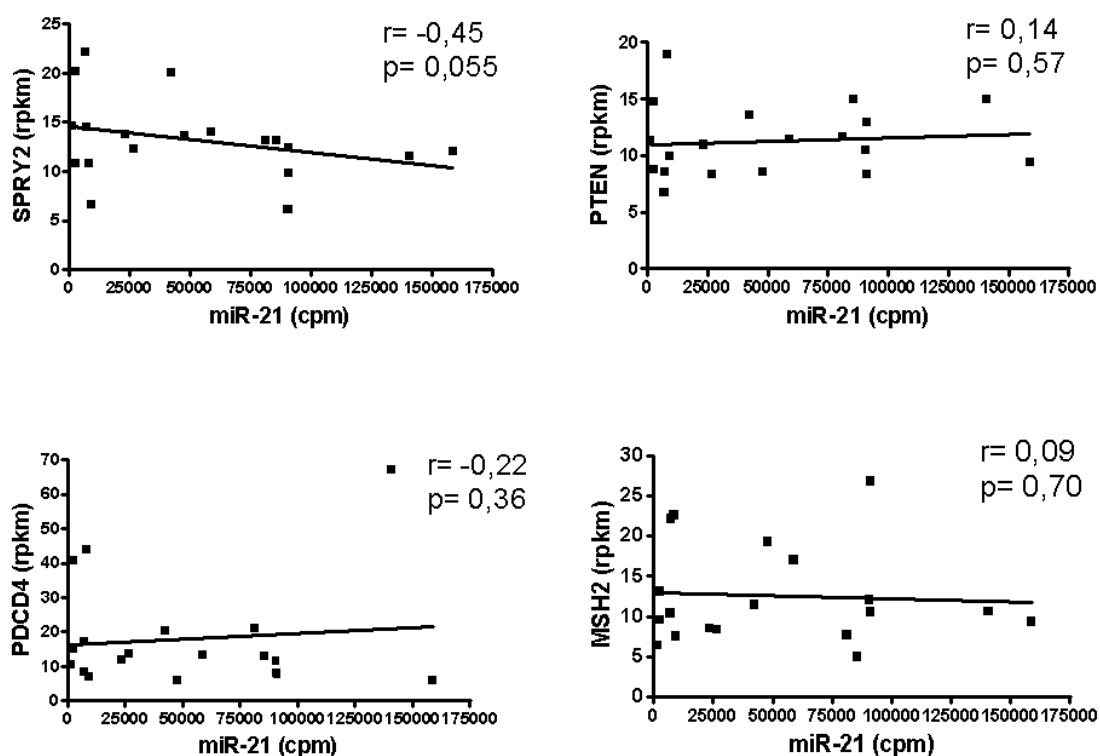


Figura 41 - Correlação da expressão de miR-21-5p e de 4 dos seus genes alvo. Não há correlação de Spearman estatisticamente significativa entre a expressão de miR-21-5p e os 4 genes alvo, para os quais, na literatura, a redução de expressão foi associada à resistência ao 5-FU *in vitro*. Expressão normalizada pelo número de sequências por milhão e pelo tamanho em kilobase do gene (rpkm).

Assim como outros miRNAs, o miR-21-5p é capaz de regular vários mRNAs, sendo que a sensibilidade ao tratamento QRTn poderia estar associada a alvos diferentes desses descritos na literatura. Assim, foi feita uma análise de correlação da expressão do miR-21-5p e a expressão dos seus alvos, preditos pelo programa Target Scan (LEWIS et al. 2005). Foi feita uma busca por alvos que apresentassem correlação inversa entre sua expressão e a do miR-21-5p e ainda apresentassem diferença de expressão estatisticamente significativa entre os grupos de resposta completa e incompleta. Vale lembrar que esta abordagem ainda é uma evidência

indireta da regulação dos alvos por miR-21-5p e que somente com ensaios como o de gene repórter é possível afirmar se um determinado gene realmente é alvo deste miRNA.

Foram avaliados 56 genes alvos de miR-21-5p conservados em diferentes espécies e que apresentavam um alto valor de predição pelo Target Scan, *total context+ score* de pelo menos -0,3 (LEWIS et al. 2005). O gene *SATB1* foi o único que apresentou uma correlação inversa entre as duas expressões ($r = -0,5$ e $p = 0,03$) (**Figura 42A**). *SATB1* é um alvo do miR-21-5p bastante interessante por estar associado ao fenótipo de resistência a múltiplas drogas. Além disso, a regulação de *SATB1* por miR-21-5p já foi confirmada experimentalmente (BEYER et al. 2011). Sendo assim, foi avaliada a expressão de *SATB1* comparando os pacientes de resposta incompleta ($n=19$) com os pacientes de resposta completa ($n=9$). Para essa análise foram utilizados todos os dados de sequenciamento de transcriptoma disponíveis no laboratório, independente deles terem sido incluídos para o estudo de expressão de miRNAs. Como esperado, foi verificada uma maior expressão de *SATB1* nos pacientes de resposta incompleta em relação aos pacientes de resposta completa, sendo essa diferença de expressão estatisticamente significativa ($p=0,02$).

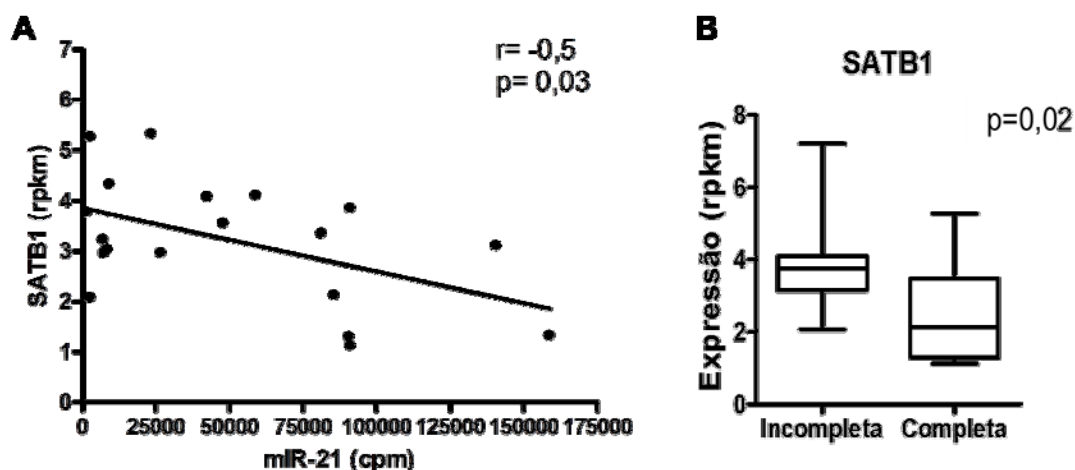


Figura 42 - Expressão, analisada por sequenciamento, de *SATB1*, gene alvo do miR-21-5p. A) Correlação inversa da expressão de miR-21-5p e do seu gene alvo *SATB1* (r : fator de correlação de Pearson). B) Expressão de *SATB1* comparada entre os pacientes de resposta incompleta ($n=19$) e de resposta completa ($n=9$), normalizada pelo número de sequências por milhão e pelo tamanho em kilobase do gene (rpkm). Teste estatístico Mann-Whitney.

SATB1 (proteína especial de ligação em região rica em AT, do inglês *special AT-rich binding protein 1*) atua na regulação da expressão gênica, podendo ter uma atuação direta através de sua ligação a promotores ou indireta por atuar na organização da cromatina (MIR et al. 2012). Os genes regulados por *SATB1* podem variar de acordo com o contexto e o tipo celular. No câncer de mama, *SATB1* promove a progressão e metástase através da regulação de genes envolvidos na adesão celular, sinalização celular e ciclo celular (HAN et al. 2008). Além disso, *SATB1* pode ter papel importante na resposta à Qt uma vez que linhagens de câncer de mama que desenvolveram resistência a múltiplas drogas *in vitro* apresentaram maior expressão de *SATB1* em relação às linhagens celulares parentais (LI et al. 2010). É importante ressaltar também que a modulação da expressão de

SATB1 *in vitro* e *in vivo* mostrou que o aumento da expressão de SATB1 promovia a multirresistência enquanto a sua depleção levava a sensibilidade a várias drogas, incluindo-se o 5-FU. Este foi o primeiro estudo a mostrar que SATB1 pode promover resistência à Qt.

Esse efeito de SATB1 também foi demonstrado, posteriormente, em linhagens derivadas de câncer gástrico e de bexiga (SUN et al. 2012; HAN et al. 2013). Um mecanismo semelhante poderia explicar, pelo menos em parte, a maior resistência ao tratamento QRTn dos pacientes com câncer de reto que apresentaram resposta incompleta. Nestes pacientes a menor expressão de miR-21-5p e consequente maior expressão de SATB1 levaria ao aumento da expressão de genes associados ao fenótipo de resistência a múltiplas drogas e consequentemente à redução das taxas de apoptose induzida pela QRTn.

Como foi observada uma diferença de expressão de SATB1 entre os pacientes de resposta completa e incompleta, também foi calculada a AUC-ROC para avaliar a acurácia da expressão de SATB1 em prever a resposta ao tratamento QRTn considerando todas as amostras disponíveis (n=28), sendo a acurácia igual a 78% (**Figura 43**). Além disso, utilizando o valor de corte de 2,5 rpk, foi possível prever corretamente a resposta de 56% (5/9) dos pacientes de resposta completa e 92% (18/19) dos pacientes de resposta incompleta. Portanto, além do miR-21-5p, o seu alvo SATB1 também é um bom candidato a biomarcador para predição de resposta à QRTn em pacientes com câncer de reto.

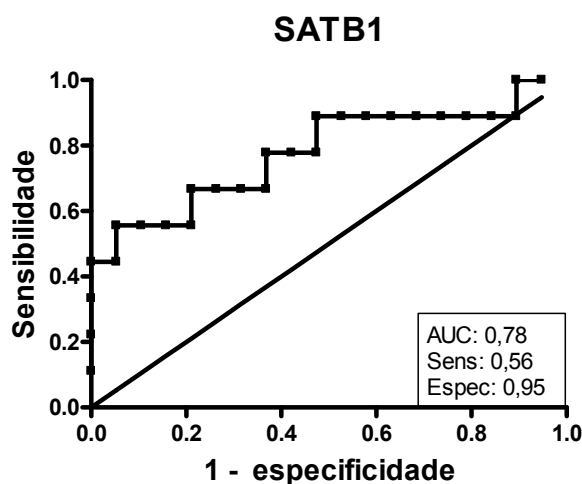


Figura 43 - Curva ROC para determinar a acurácia da predição de resposta a QRTn baseada na expressão de SATB1. Foram calculados a área sob a curva (AUC), sensibilidade (Sens) e especificidade (Espec) após identificar o melhor valor de corte (2,5rpkm) para a separação dos pacientes de resposta completa (n=9) e incompleta (n=19).

Os resultados obtidos por sequenciamento foram avaliados por uma segunda metodologia, sendo ela o qPCR. Infelizmente, poucas amostras ainda dispunham de RNA total para essa análise que foi feita com amostras de 8 pacientes de resposta incompleta e 3 de resposta completa. Observou-se a mesma tendência entre os resultados obtidos por sequenciamento e por qPCR, sendo a expressão de SATB1 maior nos pacientes de resposta incompleta em relação aos pacientes de resposta completa (**Figura 44**). No entanto, não houve uma diferença estatisticamente significativa entre os dois tipos de resposta, o que pode ser decorrente do baixo número de amostras disponíveis para essa análise.

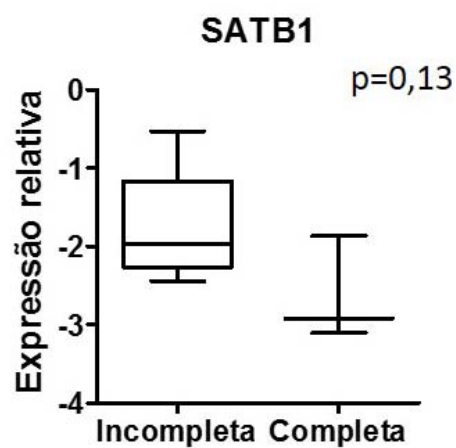


Figura 44 - Expressão relativa de SATB1 avaliada por qPCR nos pacientes de resposta completa e incompleta ao tratamento de QRTn. O eixo y representa o valor de expressão relativa ($2^{-\Delta\Delta Ct}$) do SATB1 normalizado por PUM1 e HMBS em relação a uma referência (tecido de cólon normal obtido comercialmente da Clontech).

- **Sensibilidade à QRT *in vitro***

Embora parcialmente conflitante com os dados da literatura, os resultados de associação da expressão do miR-21-5p com a resposta ao tratamento QRTn tornaram-se bastante consistentes após a identificação do SATB1. Portanto, foram realizados ensaios funcionais de manipulação da expressão do miR-21-5p em linhagens celulares cultivadas *in vitro* e submetidas ao um tratamento semelhante a QRT dos pacientes. Como dito anteriormente, a literatura tem mostrado a associação da superexpressão de miR-21-5p e resistência ao tratamento isolado com 5-FU. No entanto, o tratamento padrão para os pacientes com câncer de reto é a combinação de Rt e Qt (KOSINSKI et al. 2012). Na verdade, a Rt parece ter papel fundamental no tratamento, uma vez que a Qt (5-FU e LV) age principalmente sensibilizando as células tumorais à Rt (ZHU e WILLETT 2003; GRADE et al. 2012). Assim, a avaliação *in vitro* foi feita não apenas para a resposta ao tratamento com 5-FU e LV, mas também à Rt.

Foram escolhidas as mesmas duas linhagens celulares utilizadas no estudo de APA: HCT116 e SW480. A linhagem derivada de tumor colorretal HCT116 superexpressa miR-21-5p em relação à SW480, uma diferença de 277 vezes foi verificada por sequenciamento (**Tabela 16**). Corroborando com os dados obtidos nos pacientes, já foi demonstrado na literatura que a HCT116 é mais sensível ao tratamento isolado com 5-FU do que a SW480 (MARIADASON et al. 2003).

Tabela 16 - Expressão obtida por sequenciamento do miR-21-5p para as 2 linhagens de tumor colorretal HCT116 e SW480 (normalizada por cpm - número de sequências por milhão).

miR-21-5p (cpm)		diferença expressão
HCT116	SW480	HCT116/SW480
341.199	1.232	277

Para a padronização dos experimentos, em um primeiro momento os tratamentos de Rt e Qt *in vitro* foram feitos de maneira isolada, utilizando diferentes doses de cada. Durante o tratamento com 5-FU e LV, observou-se uma inibição do crescimento celular em ambas as linhagens, mas a linhagem HCT116 apresentou uma diminuição da viabilidade celular de maneira dose-dependente e bem mais acentuada do que a linhagem SW480 (**Figura 45**). Para a dose mais concentrada de quimioterápicos (500 μ M 5-FU e 25 μ M LV), após 3 dias de tratamento a HCT116 apresentava 55% de células viáveis enquanto a SW480 apresentava 70% (**Figura 45**).

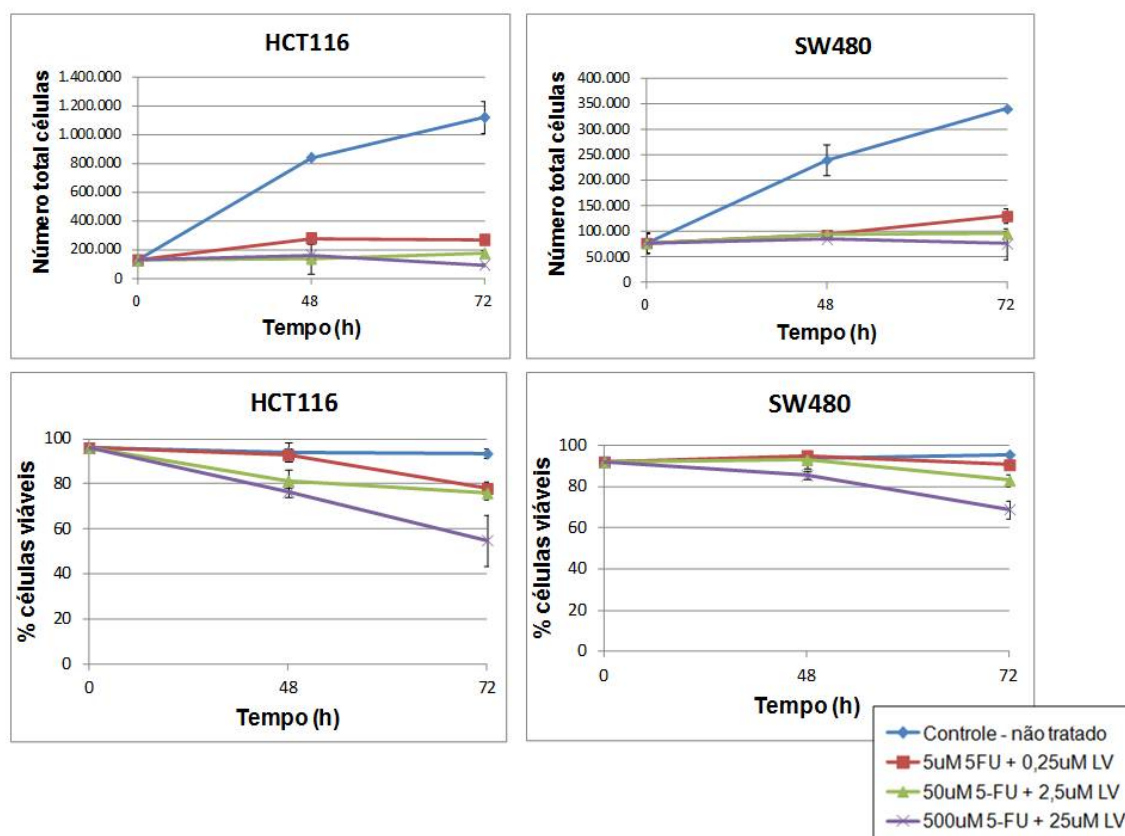


Figura 45 - Proliferação e viabilidade celular após diferentes doses de Qt. As linhagens HCT116 e SW480 foram tratadas com as doses descritas e o número total de células viáveis e mortas foi obtido através de contagem utilizando Trypan Blue. A barra representa o desvio padrão de 3 replicatas.

Após a padronização da Qt, foram feitos os experimentos de simulação da Rt, utilizando-se 3 doses diferentes de radiação. Novamente, observou-se uma redução na taxa de proliferação celular dose-dependente e semelhante entre as 2 linhagens. A diminuição da viabilidade celular foi pequena e ocorreu apenas para a HCT116 após 72h com a dosagem mais alta (8Gy, **Figura 46**). Devido a este discreto efeito da radiação na viabilidade celular, as células foram novamente submetidas ao tratamento com a maior dose, mas mantidas em cultura por um maior período de tempo. Isto porque acredita-se que o principal efeito de dano ao DNA causado pela

radiação é indireto, devido ao aumento de radicais livres na célula, o que portanto pode requerer um maior tempo para desencadear a morte celular (GRADE et al. 2012). Assim, estendendo o tempo de análise após a Rt, a HCT116 apresentou-se mais sensível, sendo que após 6 dias da radiação apenas 58% das células estavam viáveis enquanto a viabilidade da SW480 era de 83% (Figura 47).

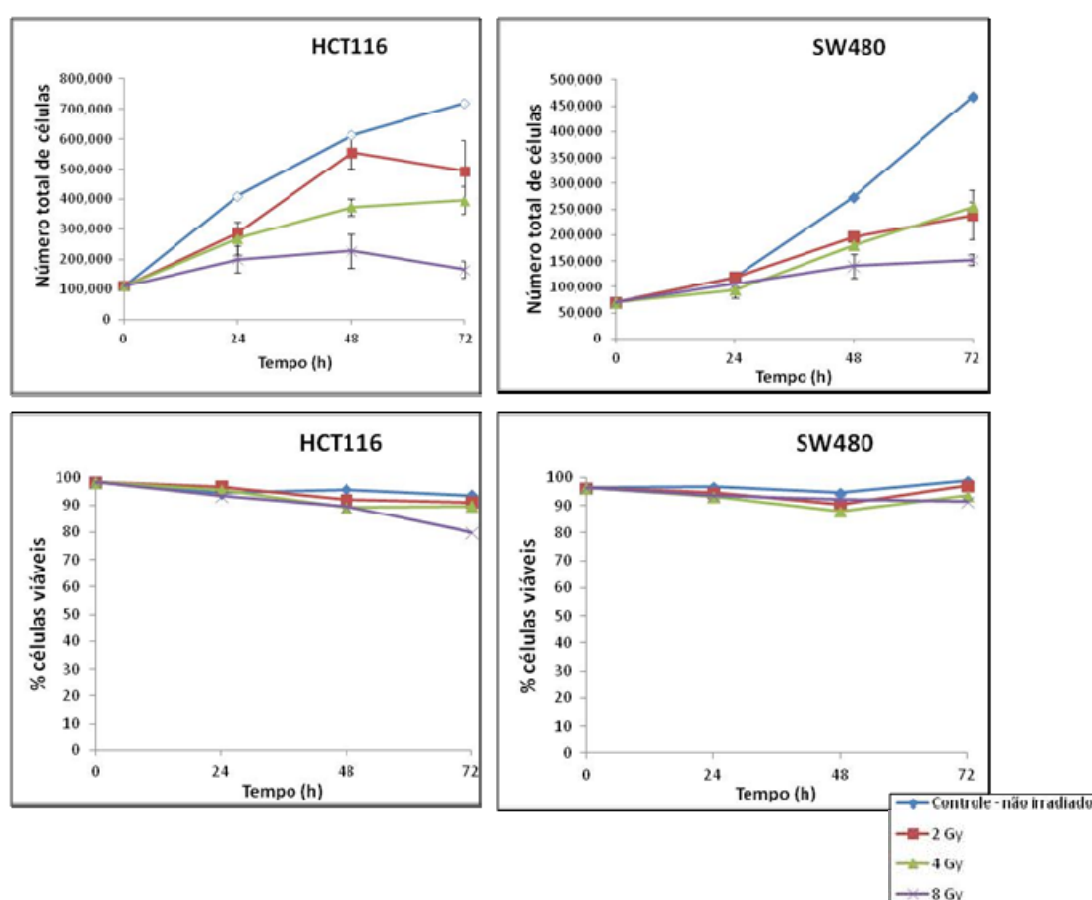


Figura 46 - Proliferação e viabilidade celular após diferentes doses de Rt. As linhagens HCT116 e SW480 foram irradiadas com as doses descritas e o número total de células viáveis e mortas foi obtido através de contagem utilizando Trypan Blue. A barra representa o desvio padrão de 3 replicatas.

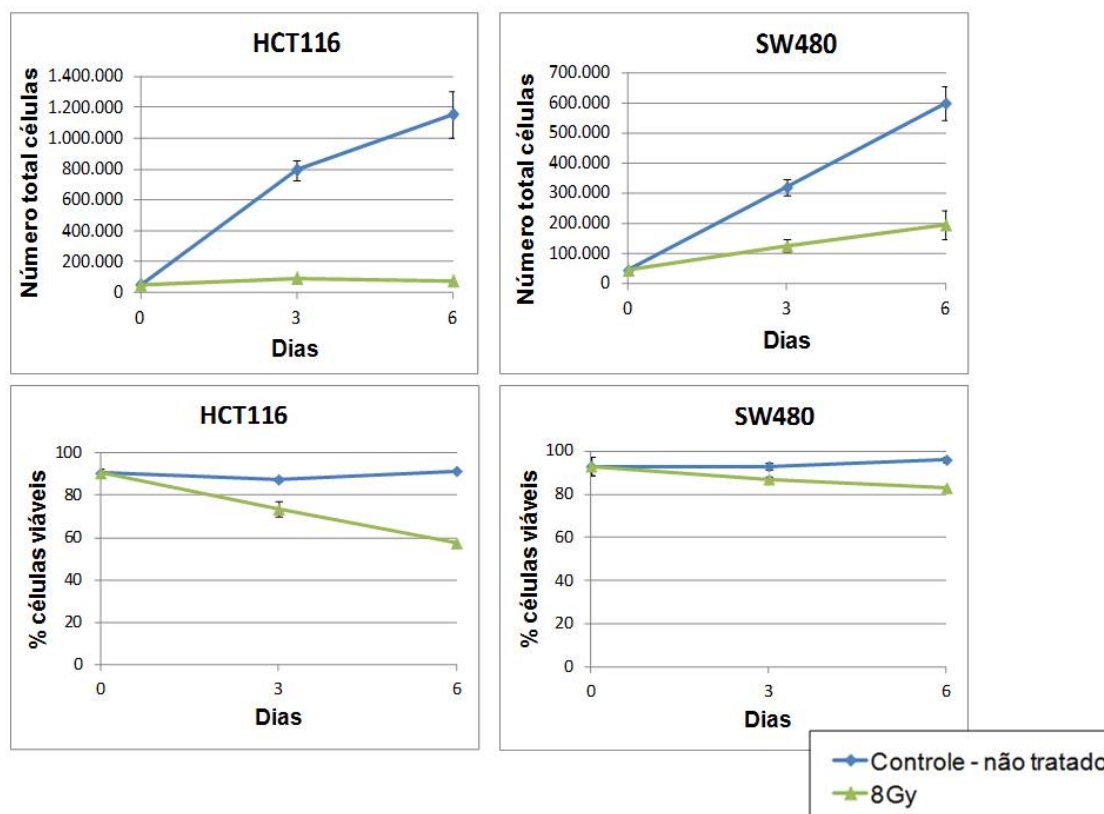


Figura 47 - Proliferação e viabilidade celular durante 10 dias após o tratamento Rt. As linhagens HCT116 e SW480 foram irradiadas com uma única dose (8Gy) e avaliadas durante 10 dias. O número total de células viáveis e mortas foi obtido através de contagem utilizando Trypan Blue. A barra representa o desvio padrão de 3 replicatas.

Esses resultados preliminares estão de acordo com a nossa hipótese sobre o papel do miR-21-5p na resposta ao tratamento QRT: a linhagem HCT116 expressa mais miR-21-5p do que a linhagem SW480 e é a mais sensível a ambos os tratamentos de Qt e Rt isolados. Entretanto, como as 2 linhagens estudadas são derivadas de tumores diferentes e geneticamente distintas não é possível assumir que a diferença de sensibilidade entre as duas seja necessariamente decorrente da expressão diferencial de miR-21-5p. Assim, fez-se necessária a modulação da expressão de miR-21-5p em

cada uma delas e o posterior tratamento de Qt e Rt para verificar se há alteração na resposta ao mesmo. Paralelamente às padronizações de dose e tempo de tratamento com Qt e Rt, também foi padronizada a superexpressão ou inibição transiente de miR-21-5p através da transfecção com oligos miméticos de miR-21-5p ou anti-miR-21-5p, respectivamente.

A linhagem HCT116, que apresenta alta expressão constitutiva de miR-21-5p, foi transfectada com o anti-miR-21. Após 24h já pôde ser observada uma redução de 70% da expressão de miR-21-5p nesta linhagem, o que se manteve até 72h após a transfecção. No entanto, a expressão de miR-21-5p é logo reestabelecida, após 120h (**Figura 48**). Estes resultados revelaram que a diminuição da expressão do miR-21-5p é bastante transiente. De maneira semelhante, a superexpressão de miR-21-5p foi feita na linhagem SW480 (que apresenta baixos níveis de expressão constitutiva do mesmo). Vinte e quatro horas após a transfecção, foi observado um aumento de 5 vezes de expressão e já após 72h os níveis de expressão voltaram ao normal (**Figura 49**).

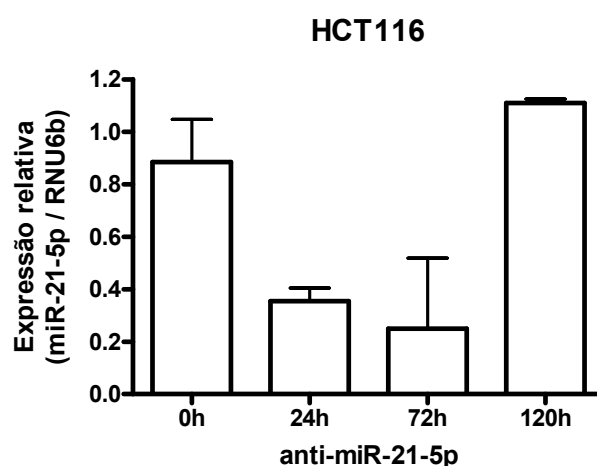


Figura 48 - Inibição de miR-21-5p na HCT116. Análise por qPCR da expressão do miR-21-5p após transfecção com o anti-miR-21-5p. Expressão analisada para condição transfectada com molécula controle (0h) e 24h, 72h e 120h após transfecção com molécula anti-miR-21-5p. O eixo y representa a expressão de miR-21-5p normalizada pelo RNU6b e em relação a condição controle - 0h ($2^{-\Delta\Delta Ct}$). A barra representa o desvio padrão de 2 experimentos em triplicata.

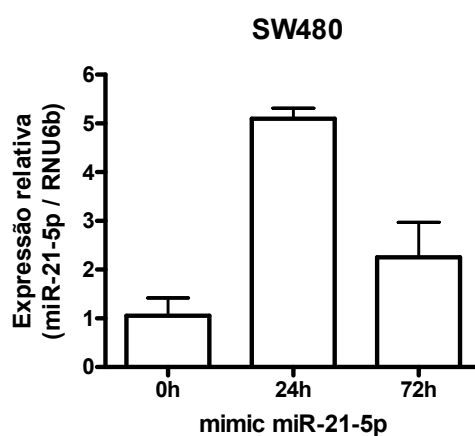


Figura 49 - Superexpressão de miR-21-5p na SW480. Análise por qPCR da expressão do miR-21-5p após transfecção com uma molécula mimética de miR-21-5p. A expressão foi analisada para a condição transfectada com molécula controle (0h) e 24h e 72h após transfecção com molécula mimética de miR-21-5p. O eixo y representa a expressão de miR-21-5p normalizada pelo RNU6b e em relação a condição controle - 0h ($2^{-\Delta\Delta Ct}$). A barra representa o desvio padrão de 2 experimentos em triplicata.

Visto que a inibição e superexpressão transientes de miR-21-5p se mantiveram por um tempo relativamente curto (72h), não seria informativo prosseguir com os experimentos funcionais uma vez que os efeitos da Rt (provavelmente o principal componente do tratamento dado aos pacientes) só puderam ficar evidentes após 6 dias da exposição. Cabe lembrar que manipular a expressão de miRNAs e observar seus efeitos funcionais nas células exige que as mesmas mantenham a expressão alterada por um período mínimo que garanta que os alvos do miRNA em estudo também tenham sua expressão alterada. Portanto, é necessário aprimorar o modelo de estudo, sendo necessário um vetor de expressão ou de inibição estável de miR-21-5p, assim como a seleção de clones celulares que apresentam a expressão ou inibição de miR-21-5p confirmadas por qPCR. Após estabelecer este modelo, também seria interessante avaliar o efeito combinado de Qt e Rt e, ainda, seria possível avaliar a expressão dos genes alvo que têm sua expressão alterada após a manipulação da expressão do miR-21-5p e que poderiam estar relacionados à resposta ao tratamento. Certamente, avaliar a expressão de SATB1 nestas condições poderia revelar se de fato o mesmo está envolvido na determinação da sensibilidade à QRT. De qualquer maneira, estes resultados preliminares de experimentos funcionais indicam que a maior expressão do miR-21-5p pode estar diretamente envolvida com uma maior sensibilidade ao tratamento QRTn e corroborar com os achados de expressão nas amostras dos pacientes com câncer de reto.

5 CONCLUSÕES

5.1 PARTE I: IDENTIFICAÇÃO E ESTUDO DE EXPRESSÃO DE VARIANTES DE POLIADENILAÇÃO E DE MIRNAS

Para o estudo de APA, foram desenvolvidos dois protocolos inéditos de construção de bibliotecas de cDNA enriquecidas para a extremidade 3' dos transcritos, um compatível com o 454 (Roche) e outro com o SOLiD5500 (Life Technologies). Embora 80% das sequências geradas no 454 correspondessem ao último exon 3', estas sequências estavam truncadas e apenas 1,5% apresentaram o sítio de poliA, limitando os resultados obtidos com este protocolo. Ainda assim, foram identificadas 994 variantes de poliA mais curtas e 250 variantes mais longas na linhagem HCT116, enquanto que para a SW480 foram identificadas 1.113 variantes mais curtas e 247 variantes mais longas.

As bibliotecas sequenciadas no SOLiD foram eficientes para identificar um maior número de variantes de poliA, bem como estimar a expressão das mesmas. Quase 30% dos genes expressos em cada linhagem apresentaram variantes curtas de poliA. Foram identificados 3.946 e 4.359 genes nas linhagens HCT116 e SW480, respectivamente, com mais de um sítio de poliA no último exon 3', sendo 1.413 variantes exclusivas da HCT116 e 1.826 variantes exclusivas da SW480. Finalmente, essa metodologia em larga escala foi adequada para identificar variantes curtas

de poliA, com uma taxa de sucesso de 80% (8/10) na validação experimental por RACE 3'.

Cerca de 60% das variantes curtas de poliA identificadas perderam pelo menos um sítio alvo para miRNA. Houve uma redução global da razão da expressão da variante longa pela expressão da variante curta (razão VL/VC) para os genes cuja variante curta perdeu sítio alvo de miRNA (na região que diferenciava a variante curta da longa). Sendo assim, a abundância de variantes longa e curta para um mesmo gene pode estar associada à regulação por miRNAs. De fato, enquanto não houve diferença da razão VL/VC entre os genes que não possuíam sítios alvo ou possuíam sítios para miRNAs não expressos, essa razão foi significativamente menor nos genes que possuíam sítios para miRNAs expressos. Esses resultados mostraram a importância da expressão de miRNAs sobre a expressão de variantes de poliA.

5.2 PARTE II: ANÁLISE DE EXPRESSÃO DE MIRNAS EM PACIENTES COM CÂNCER DE RETO E SUA ASSOCIAÇÃO COM A RESPOSTA AO TRATAMENTO

A análise de expressão de miRNAs que pudessem ser utilizados como biomarcadores de resposta ao tratamento QRTn em pacientes com câncer de reto identificou 4 miRNAs diferencialmente expressos entre os pacientes de resposta completa e de resposta incompleta (miR-21-5p, miR-1246, miR1290-3p e miR-205-5p).

O perfil de expressão desses 4 miRNAs foi capaz de agrupar parcialmente os pacientes de acordo com a resposta à QRTn, sendo que apenas 3 dos 20 pacientes de resposta incompleta clusterizaram incorretamente.

O poder preditivo de cada miRNA foi estimado por meio do cálculo da AUC-ROC e variou de 63% a 94%, sendo que o miR-21-5p apresentou a melhor acurácia para a classificação dos pacientes em resposta completa e incompleta.

A expressão desses 4 miRNAs foi avaliada em um grupo independente de amostras sendo observado um aumento estatisticamente significativo da expressão do miR-21-5p nos pacientes de resposta completa em relação aos de resposta incompleta. Sendo assim, somente o miR-21-5p teve sua expressão diferencial confirmada em um grupo independente de amostras. Utilizando-se a expressão de miR-21-5p, foram corretamente classificados 57% (4/7) dos pacientes de resposta completa e 80% (4/5) dos pacientes de resposta incompleta.

A expressão de miR-21-5p em pacientes que apresentaram recidiva local precoce foi semelhante à de pacientes de resposta incompleta e significativamente menor do que nos pacientes de resposta completa sustentada.

Avaliando-se a expressão de genes alvo de miR-21-5p nas amostras de pacientes, foi encontrado o gene *SATB1* com correlação inversa entre sua expressão e a do miR-21-5p. *SATB1* apresentou uma diferença de expressão estatisticamente significativa entre os pacientes de resposta completa e incompleta e ainda apresentou boa acurácia para predição de resposta à QRTn, com AUC-ROC de 78%.

Em experimentos *in vitro*, foi observado que a linhagem HCT116 expressa mais miR-21-5p do que a linhagem SW480 e é a mais sensível a ambos os tratamentos de Qt e Rt isolados. Ainda é necessário modular a expressão *in vitro* de miR-21-5p para melhor compreensão do papel desse miRNA, de qualquer maneira, estes resultados preliminares indicam que a maior expressão do miR-21-5p pode estar envolvida com uma maior sensibilidade ao tratamento QRTn.

Além de evidências preliminares da associação direta de miR-21-5p à sensibilidade à QRT, este estudo apontou o miR-21-5p como um potencial biomarcador no auxílio à conduta terapêutica oferecida aos pacientes com câncer de reto. Juntamente com exames clínicos e de imagem, a avaliação da expressão de miR-21-5p em biópsias de tumor de reto poderia ajudar na seleção de pacientes com resposta clínica completa potencialmente candidatos para estratégias de tratamento alternativas à cirurgia radical.

6 REFERÊNCIAS BIBLIOGRÁFICAS

Ahnen DJ. The American College of Gastroenterology Emily Couric Lecture--the adenoma-carcinoma sequence revisited: has the era of genetic tailoring finally arrived?. **Am J Gastroenterol** 2011; 106:190-8.

Akiyoshi T, Kobunai T, Watanabe T. Predicting the response to preoperative radiation or chemoradiation by a microarray analysis of the gene expression profiles in rectal cancer. **Surg Today** 2012; 42:713-9.

Aklilu M, Eng C. The current landscape of locally advanced rectal cancer. **Nat Rev Clin Oncol** 2011; 8:649–59.

Allen KE, Weiss GJ. Resistance may not be futile: microRNA biomarkers for chemoresistance and potential therapeutics. **Mol Cancer Ther** 2010; 9:3126-36.

Baek D, Villén J, Shin C, Camargo FD, Gygi SP, Bartel DP. The impact of microRNAs on protein output. **Nature** 2008; 455:64-71.

Bartel DP. MicroRNAs: target recognition and regulatory functions. **Cell** 2009; 136:215-33

Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. **Cell** 2004; 116:281-97.

Bava FA, Eliscovich C, Ferreira PG, et al. CPEB1 coordinates alternative 3'-UTR formation with translational regulation. **Nature** 2013; 495:121-5.

Beaudoing E, Freier S, Wyatt JR, Claverie JM, Gautheret D. Patterns of variant polyadenylation signal usage in human genes. **Genome Res** 2000; 10:1001-10.

Benjamini Y, Hochberg Y, Controlling the false discovery rate: a practical and powerful approach to multiple testing. **J Royal Statistical Society Series B (Methodological)** 1995; 57:289-300.

Beyer M, Thabet Y, Müller RU, et al. Repression of the genome organizer SATB1 in regulatory T cells is required for suppressive function and inhibition of effector differentiation. **Nat Immunol** 2011; 12:898-907.

Blaustein M, Pelisch F, Srebrow A. Signals, pathways and splicing regulation. **Int J Biochem Cell Biol** 2007; 39:2031-48.

Blencowe BJ. Alternative splicing: new insights from global analyses. **Cell** 2006; 126:37-47.

Cai X, Hagedorn CH, Cullen BR. Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. **RNA** 2004; 10:1957-66.

Calin GA, Croce CM. MicroRNA signatures in human cancers. **Nat Rev Cancer** 2006; 6:857-66.

Capirci C, Valentini V, Cionini L, et al. Prognostic value of pathologic complete response after neoadjuvant therapy in locally advanced rectal cancer: long-term analysis of 566 ypCR patients. **Int J Radiat Oncol Biol Phys** 2008; 72:99-107.

Chan JA, Krichevsky AM, Kosik KS. MicroRNA-21 is an antiapoptotic factor in human glioblastoma cells. **Cancer Res** 2005; 65:6029-33.

Chendrimada TP, Gregory RI, Kumaraswamy E, et al. TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing. **Nature** 2005; 436:740-4.

Colgan DF, Manley JL. Mechanism and regulation of mRNA polyadenylation. **Genes Dev** 1997; 11:2755-66.

Costessi L, Devescovi G, Baralle FE, Muro AF. Brain-specific promoter and polyadenylation sites of the beta-adducin pre-mRNA generate an unusually long 3'-UTR. **Nucleic Acids Res** 2006; 34:243-53.

Croce CM. Causes and consequences of microRNA dysregulation in cancer. **Nat Rev Genet** 2009; 10:704-14.

Czech B, Hannon GJ. Small RNA sorting: matchmaking for Argonautes. **Nat Rev Genet** 2011; 12:19-31.

Daemen A, Gevaert O, De Bie T, et al. Integrating microarray and proteomics data to predict the response on cetuximab in patients with rectal cancer. **Pac Symp Biocomput** 2008:166-77.

Della Vittoria Scarpati G, Falchetta F, Carlomagno C, et al. A specific miRNA signature correlates with complete pathological response to neoadjuvant chemoradiotherapy in locally advanced rectal cancer. **Int J Radiat Oncol Biol Phys** 2012; 83:1113-9.

Derti A, Garrett-Engle P, Macisaac KD, et al. A quantitative atlas of polyadenylation in five mammals. **Genome Res** 2012; 22:1173-83.

Di Giammartino DC, Nishida K, Manley JL. Mechanisms and consequences of alternative polyadenylation. **Mol Cell** 2011; 43:853-66.

Dworak O, Keilholz L, Hoffmann A. Pathological features of rectal cancer after preoperative radiochemotherapy. **Int J Colorectal Dis** 1997; 12:19-23.

Edwards-Gilbert G, Veraldi KL, Milcarek C. Alternative poly(A) site selection in complex transcription units: means to an end?. **Nucleic Acids Res** 1997; 25:2547-61.

Elkon R, Ugalde AP, Agami R. Alternative cleavage and polyadenylation: extent, regulation and function. **Nat Rev Genet** 2013; 14:496-506.

Esquela-Kerscher A, Slack FJ. Oncomirs - microRNAs with a role in cancer. **Nat Rev Cancer** 2006; 6:259-69.

Fearon ER, Vogelstein B. A genetic model for colorectal tumorigenesis. **Cell** 1990; 61:759-67.

Feng YH, Wu CL, Shiau AL, et al. MicroRNA-21-mediated regulation of Sprouty2 protein expression enhances the cytotoxic effect of 5-fluorouracil and metformin in colon cancer cells. **Int J Mol Med** 2012; 29:920-6.

Flavell SW, Kim TK, Gray JM, et al. Genome-wide analysis of MEF2 transcriptional program reveals synaptic target genes and neuronal activity-dependent polyadenylation site selection. **Neuron** 2008; 60:1022-38.

Fox-Walsh K, Davis-Turak J, Zhou Y, Li H, Fu XD. A multiplex RNA-seq strategy to profile poly(A+) RNA: application to analysis of transcription response and 3' end formation. **Genomics** 2011; 98:266-71.

Fu Y, Sun Y, Li Y, et al. Differential genome-wide profiling of tandem 3' UTRs among human breast cancer and normal cells by high-throughput sequencing. **Genome Res** 2011; 21:741-7.

Galante PA, Vidal DO, de Souza JE, Camargo AA, de Souza SJ. Sense-antisense pairs in mammals: functional and evolutionary considerations. **Genome Biol** 2007; 8:R40.

Garcia-Aguilar J, Smith DD, Avila K, Bergsland EK, Chu P, Krieg RM, Timing of Rectal Cancer Response to Chemoradiation Consortium. Optimal timing of surgery after chemoradiation for advanced rectal cancer: preliminary results of a multicenter, nonrandomized phase II prospective trial. **Ann Surg** 2011; 254:97-102.

Garzon R, Fabbri M, Cimmino A, Calin GA, Croce CM. MicroRNA expression and function in cancer. **Trends Mol Med** 2006; 12:580-7.

Gentleman RC, Carey VJ, Bates DM, et al. Bioconductor: open software development for computational biology and bioinformatics. **Genome Biol** 2004; 5:R80.

Gérard JP, Azria D, Gourgou-Bourgade S, et al. Comparison of two neoadjuvant chemoradiotherapy regimens for locally advanced rectal cancer: results of the phase III trial ACCORD 12/0405-Prodige 2. **J Clin Oncol** 2010; 28:1638-44.

Ghadimi BM, Grade M, Difilippantonio MJ, et al. Effectiveness of gene expression profiling for response prediction of rectal adenocarcinomas to preoperative chemoradiotherapy. **J Clin Oncol** 2005; 23:1826-38.

Grade M, Wolff HA, Gaedcke J, Ghadimi BM. The molecular basis of chemoradiosensitivity in rectal cancer: implications for personalized therapies. **Langenbecks Arch Surg** 2012; 397:543-55.

Griffiths-Jones S, Hui JH, Marco A, Ronshaugen M. MicroRNA evolution by arm switching. **EMBO Rep** 2011; 12:172-7.

Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ. miRBase: microRNA sequences, targets and gene nomenclature. **Nucleic Acids Res** 2006; 34(Database issue):D140-4.

Guo H, Ingolia NT, Weissman JS, Bartel DP. Mammalian microRNAs predominantly act to decrease target mRNA levels. **Nature** 2010; 466:835-40.

Habr-Gama A. Assessment and management of the complete clinical response of rectal cancer to chemoradiotherapy. **Colorectal Dis** 2006; 8 (Suppl 3):21-4.

Habr-Gama A, Perez RO, Wynn G, Marks J, Kessler H, Gama-Rodrigues J. Complete clinical response after neoadjuvant chemoradiation therapy for distal rectal cancer: characterization of clinical and endoscopic findings for standardization. **Dis Colon Rectum** 2010; 53:1692-8.

Habr-Gama A, Perez RO, Sabbaga J, Nadalin W, São Julião GP, Gama-Rodrigues J. Increasing the rates of complete response to neoadjuvant chemoradiotherapy for distal rectal cancer: results of a prospective study using additional chemotherapy during the resting period. **Dis Colon Rectum** 2009; 52:1927-34.

Habr-Gama A, Perez RO, Nadalin W, et al. Operative versus nonoperative treatment for stage 0 distal rectal cancer following chemoradiation therapy: long-term results. **Ann Surg** 2004; 240:711-7.

Hafez D, Ni T, Mukherjee S, Zhu J, Ohler U. Genome-wide identification and predictive modeling of tissue-specific alternative polyadenylation. **Bioinformatics** 2013; 29:i108-16.

Hamfjord J, Stangeland AM, Hughes T, et al. Differential expression of miRNAs in colorectal cancer: comparison of paired tumor tissue and adjacent normal mucosa using high-throughput sequencing. **PLoS One** 2012; 7:e34150.

Han B, Luan L, Xu Z, Wu B. Expression and biological roles of SATB1 in human bladder cancer. **Tumour Biol** 2013; 34:2943-9.

Han HJ, Russo J, Kohwi Y, Kohwi-Shigematsu T. SATB1 reprogrammes gene expression to promote breast tumour growth and metastasis. **Nature** 2008; 452:187-93.

Hummel R, Hussey DJ, Haier J. MicroRNAs: predictors and modifiers of chemo- and radiotherapy in different tumour types. **Eur J Cancer** 2010; 46:298-311.

Jan CH, Friedman RC, Ruby JG, Bartel DP. Formation, regulation and evolution of *Caenorhabditis elegans* 3'UTRs. **Nature** 2011; 469:97-101.

Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. **CA Cancer J Clin** 2011; 61:69-90.

Jenal M, Elkon R, Loayza-Puch et al. The poly(A)-binding protein nuclear 1 suppresses alternative cleavage and polyadenylation sites. **Cell** 2012; 149:538-53.

Ji Z, Lee JY, Pan Z, Jiang B, Tian B. Progressive lengthening of 3' untranslated regions of mRNAs by alternative polyadenylation during mouse embryonic development. **Proc Natl Acad Sci U S A** 2009; 106:7028-33.

Ji Z, Tian B. Reprogramming of 3' untranslated regions of mRNAs by alternative polyadenylation in generation of pluripotent stem cells from different cell types. **PLoS One** 2009; 4:e8419.

Ji Z, Luo W, Li W, et al. Transcriptional activity regulates alternative cleavage and polyadenylation. **Mol Syst Biol** 2011; 7:534.

Kent WJ. BLAT--the BLAST-like alignment tool. **Genome Res** 2002; 12:656-64.

Kheirelseid EA, Miller N, Chang KH, et al. miRNA expressions in rectal cancer as predictors of response to neoadjuvant chemoradiation therapy. **Int J Colorectal Dis** 2013; 28:247-60.

Kim CJ, Yeatman TJ, Coppola D, et al. Local excision of T2 and T3 rectal cancers after downstaging chemoradiation. **Ann Surg** 2001; 234:352-8.

Kim IJ, Lim SB, Kang HC, et al. Microarray gene expression profiling for predicting complete response to preoperative chemoradiotherapy in patients with advanced rectal cancer. **Dis Colon Rectum** 2007; 50:1342-53.

Kosinski L, Habr-Gama A, Ludwig K, Perez R. Shifting concepts in rectal cancer management: a review of contemporary primary rectal cancer treatment strategies. **CA Cancer J Clin** 2012; 62:173-202.

Kozomara A, Griffiths-Jones S. miRBase: integrating microRNA annotation and deep-sequencing data. **Nucleic Acids Res** 2011; 39:D152-7.

Lee Y, Kim M, Han J, et al. MicroRNA genes are transcribed by RNA polymerase II. **EMBO J** 2004; 23:4051-60.

Legendre M, Ritchie W, Lopez F, Gautheret D. Differential repression of alternative transcripts: a screen for miRNA targets. **PLoS Comput Biol** 2006; 2:e43.

Lewis BP, Burge CB, Bartel DP. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. **Cell** 2005; 120:15-20.

Li QQ, Chen ZQ, Xu JD, et al. Over expression and involvement of special AT-rich sequence binding protein 1 in multidrug resistance in human breast carcinoma cells. **Cancer Sci** 2010; 101:80-6.

Li SC, Liao YL, Ho MR, Tsai KW, Lai CH, Lin WC. miRNA arm selection and isomiR distribution in gastric cancer. **BMC Genomics** 2012; 13 (Suppl 1):S13.

Liaw HH, Lin CC, Juan HF, Huang HC. Differential microRNA regulation correlates with alternative polyadenylation pattern between breast cancer and normal cells. **PLoS One** 2013; 8:e56958

Lin Y, Li Z, Oszolak F, et al. An in-depth map of polyadenylation sites in cancer. **Nucleic Acids Res** 2012; 40:8460-71.

Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. **Methods** 2001; 25:402-8.

Lu J, Getz G, Miska EA, et al. MicroRNA expression profiles classify human cancers. **Nature** 2005; 435:834-8.

Lund E, Güttinger S, Calado A, Dahlberg JE, Kutay U. Nuclear export of microRNA precursors. **Science** 2004; 303:95-8.

Lutz CS. Alternative polyadenylation: a twist on mRNA 3' end formation. **ACS Chem Biol** 2008; 3:609-17.

Maas M, Beets-Tan RG, Lambregts DM, et al. Wait-and-see policy for clinical complete responders after chemoradiation for rectal cancer. **J Clin Oncol** 2011; 29:4633-40.

Mandel CR, Bai Y, Tong L. Protein factors in pre-mRNA 3'-end processing. **Cell Mol Life Sci** 2008; 65:1099-122.

Mangone M, Manoharan AP, Thierry-Mieg D, et al. The landscape of *C. elegans* 3'UTRs. **Science** 2010; 329:432-5.

Mariadason JM, Arango D, Shi Q, et al. Gene expression profiling-based prediction of response of colon carcinoma cells to 5-fluorouracil and camptothecin. **Cancer Res** 2003; 63:8791-812.

Mayr C, Bartel DP. Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. **Cell** 2009; 138:673-84.

Meiri E, Levy A, Benjamin H, et al. Discovery of microRNAs and other small RNAs in solid tumors. **Nucleic Acids Res** 2010; 38:6234-46.

Ministério da Saúde. Instituto Nacional do Câncer. **Estimativa/2012 Incidência de câncer no Brasil**. Rio de Janeiro: INCA; 2011.

Mir R, Pradhan SJ, Galande S. Chromatin organizer SATB1 as a novel molecular target for cancer therapy. **Curr Drug Targets** 2012; 13:1603-15.

Moore MJ, Proudfoot NJ. Pre-mRNA processing reaches back to transcription and ahead to translation. **Cell** 2009; 136:688-700.

Morin RD, O'Connor MD, Griffith M, et al. Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. **Genome Res** 2008; 18:610-21.

Morris AR, Bos A, Diosdado B, et al. Alternative cleavage and polyadenylation during colorectal cancer development. **Clin Cancer Res** 2012; 18:5256-66.

Motoyama K, Inoue H, Takatsuno Y, et al. Over- and under-expressed microRNAs in human colorectal cancer. **Int J Oncol** 2009; 34:1069-75.

Nakajima G, Hayashi K, Xi Y, et al. Non-coding MicroRNAs hsa-let-7g and hsa-miR-181b are Associated with Chemoresponse to S-1 in Colon Cancer. **Cancer Genomics Proteomics** 2006; 3:317-24.

Nana-Sinkam SP, Fabbri M, Croce CM. MicroRNAs in cancer: personalizing diagnosis and therapy. **Ann N Y Acad Sci** 2010; 1210:25-33.

Ng EK, Chong WW, Jin H, et al. Differential expression of microRNAs in plasma of patients with colorectal cancer: a potential marker for colorectal cancer screening. **Gut** 2009; 58:1375-81.

Ni T, Yang Y, Hafez D, et al. Distinct polyadenylation landscapes of diverse human tissues revealed by a modified PA-seq strategy. **BMC Genomics** 2013; 14:615.

Nishioka M, Shimada M, Kurita N, et al. Gene expression profile can predict pathological response to preoperative chemoradiotherapy in rectal cancer. **Cancer Genomics Proteomics** 2011; 8:87-92.

Oue N, Anami K, Schetter AJ, et al. High miR-21 expression from FFPE tissues is associated with poor survival and response to adjuvant chemotherapy in colon cancer. **Int J Cancer** 2013.

Ozsolak F, Kapranov P, Foissac S, et al. Comprehensive polyadenylation site maps in yeast and human reveal pervasive alternative polyadenylation. **Cell** 2010; 143:1018-29.

Paty PB, Nash GM, Baron P, et al. Long-term results of local excision for rectal cancer. **Ann Surg** 2002; 236:522-29; discussion 529-30.

Piepoli A, Tavano F, Copetti M, et al. Mirna expression profiles identify drivers in colorectal and pancreatic cancers. **PLoS One** 2012; 7:e33663.

Rahmann S, Martin M, Schulte JH, Köster J, Marschall T, Schramm A. Identifying transcriptional miRNA biomarkers by integrating high-throughput sequencing and real-time PCR data. **Methods** 2013; 59:154-63.

Ren J, Zhu D, Liu M, Sun Y, Tian L. Downregulation of miR-21 modulates Ras expression to promote apoptosis and suppress invasion of Laryngeal squamous cell carcinoma. **Eur J Cancer** 2010; 46:3409-16.

Rimkus C, Friederichs J, Boulesteix AL, et al. Microarray-based prediction of tumor response to neoadjuvant radiochemotherapy of patients with locally advanced rectal cancer. **Clin Gastroenterol Hepatol** 2008; 6:53-61.

Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. **Bioinformatics** 2010; 26:139-40.

Ronaghi M. Pyrosequencing sheds light on DNA sequencing. **Genome Res** 2001; 11:3-11.

Ruby JG, Jan CH, Bartel DP. Intronic microRNA precursors that bypass Drosha processing. **Nature** 2007; 448:83-6.

Sachs AB, Sarnow P, Hentze MW. Starting at the beginning, middle, and end: translation initiation in eukaryotes. **Cell** 1997; 89:831-8.

Sandberg R, Neilson JR, Sarma A, Sharp PA, Burge CB. Proliferating cells express mRNAs with shortened 3' untranslated regions and fewer microRNA target sites. **Science** 2008; 320:1643-7.

Sanghera P, Wong DW, McConkey CC, Geh JI, Hartley A. Chemoradiotherapy for rectal cancer: an updated analysis of factors affecting pathological response. **Clin Oncol (R Coll Radiol)** 2008; 20:176-83.

Sarver AL, French AJ, Borralho PM et al. "Human Colon Cancer Profiles Show Differential microRNA Expression Depending on Mismatch Repair Status and Are Characteristic of Undifferentiated Proliferative States." **BMC Cancer** 2009; 9: 401.

Sauer R, Becker H, Hohenberger W, et al. German Rectal Cancer Study Group. Preoperative versus postoperative chemoradiotherapy for rectal cancer. **N Engl J Med** 2004; 351:1731-40.

Schell SR, Zlotecki RA, Mendenhall WM, Marsh RW, Vauthey JN, Copeland EM 3rd. Transanal excision of locally advanced rectal cancers downstaged using neoadjuvant chemoradiotherapy. **J Am Coll Surg** 2002; 194:584-90; discussion 590-1.

Schepeler T, Reinert JT, Ostensfeld MS, et al. Diagnostic and prognostic microRNAs in stage II colon cancer. **Cancer Res** 2008; 68:6416-24.

Schetter AJ, Leung SY, Sohn JJ, et al. MicroRNA expression profiles associated with prognosis and therapeutic outcome in colon adenocarcinoma. **JAMA** 2008; 299:425-36.

Schetter AJ, Okayama H, Harris CC. The role of microRNAs in colorectal cancer. **Cancer J** 2012; 18:244-52.

Schmitz KJ, Hey S, Schinwald A, et al. Differential expression of microRNA 181b and microRNA 21 in hyperplastic polyps and sessile serrated adenomas of the colon. **Virchows Arch** 2009; 455:49-54.

Schulte JH, Marschall T, Martin M, et al. Deep sequencing reveals differential expression of microRNAs in favorable versus unfavorable neuroblastoma. **Nucleic Acids Res** 2010; 38:5919-28.

Shenouda SK, Alahari SK. MicroRNA function in cancer: oncogene or a tumor suppressor?. **Cancer Metastasis Rev** 2009; 28:369-78.

Shepard PJ, Choi EA, Lu J, Flanagan LA, Hertel KJ, Shi Y. Complex and dynamic landscape of RNA polyadenylation revealed by PAS-Seq. **RNA** 2011; 17:761-72.

Sjöblom T, Jones S, Wood LD, et al. The consensus coding sequences of human breast and colorectal cancers. **Science** 2006; 314:268-74.

Smibert P, Miura P, Westholm JO, et al. Global patterns of tissue-specific alternative polyadenylation in *Drosophila*. **Cell Rep** 2012; 1:277-89.

Spies N, Burge CB, Bartel DP. 3' UTR-isoform choice has limited influence on the stability and translational efficiency of most mRNAs in mouse fibroblasts. **Genome Res** 2013

Sun F, Lu X, Li H, et al. Special AT-rich sequence binding protein 1 regulates the multidrug resistance and invasion of human gastric cancer cells. **Oncol Lett** 2012; 4:156-62.

Svoboda M, Izakovicova Holla L, Sefr R, et al. Micro-RNAs miR125b and miR137 are frequently upregulated in response to capecitabine chemoradiotherapy of rectal cancer. **Int J Oncol** 2008; 33:541-7.

Svoboda M, Sana J, Fabian P, et al. MicroRNA expression profile associated with response to neoadjuvant chemoradiotherapy in locally advanced rectal cancer patients. **Radiat Oncol** 2012; 20:7-195.

Tian B, Hu J, Zhang H, Lutz CS. A large-scale analysis of mRNA polyadenylation of human and mouse genes. **Nucleic Acids Res** 2005; 33:201-12.

Tian B, Pan Z, Lee JY. Widespread mRNA polyadenylation events in introns indicate dynamic interplay between polyadenylation and splicing. **Genome Res** 2007; 17:156-65.

Tomimaru Y, Eguchi H, Nagano H, et al. MicroRNA-21 induces resistance to the anti-tumour effect of interferon- α /5-fluorouracil in hepatocellular carcinoma cells. **Br J Cancer** 2010; 103:1617-26.

Valeri N, Gasparini P, Braconi C, et al. MicroRNA-21 induces resistance to 5-fluorouracil by down-regulating human DNA MutS homolog 2 (hMSH2). **Proc Natl Acad Sci U S A** 2010; 107:21098-103.

Vinciguerra P, Stutz F. mRNA export: an assembly line from genes to nuclear pores. **Curr Opin Cell Biol** 2004; 16:285-92.

Watanabe T, Komuro Y, Kiyomatsu T, et al. Prediction of sensitivity of rectal cancer cells in response to preoperative radiotherapy by DNA microarray analysis of gene expression profiles. **Cancer Res** 2006; 66:3370-4.

Weiss C, Arnold D, Dellas K, et al. Preoperative radiotherapy of advanced rectal cancer with capecitabine and oxaliplatin with or without cetuximab: A pooled analysis of three prospective phase I-II trials. **Int J Radiat Oncol Biol Phys** 2010; 78:472-8.

Wickens M, Anderson P, Jackson RJ. Life and death in the cytoplasm: messages from the 3' end. **Curr Opin Genet Dev** 1997; 7:220-32.

Winter J, Kunath M, Roepcke S, Krause S, Schneider R, Schweiger S. Alternative polyadenylation signals and promoters act in concert to control tissue-specific expression of the Opitz Syndrome gene MID1. **BMC Mol Biol** 2007; 8:105.

Yamamichi N, Shimomura R, Inada K, et al. Locked nucleic acid in situ hybridization analysis of miR-21 expression during colorectal cancer development. **Clin Cancer Res** 2009; 15:4009-16.

Zhu AX, Willett CG. Chemotherapeutic and biologic agents as radiosensitizers in rectal cancer. **Semin Radiat Oncol** 2003; 13:454-68.

Anexo 1 - Manuscrito do Artigo

Overexpression of miR-21 as a predictive marker for complete tumor regression to neoadjuvant chemoradiotherapy in rectal cancer patients

Camila Miranda Lopes-Ramos^{1,2}, Angelita Habr-Gama³, Bruna de Souza Quevedo, Natália Mariana Felício³, Fabiana Bettoni¹, Fernanda Christtanini Koyama¹, Paula Fontes Asprino¹, Pedro AF Galante¹, Joaquim Gama-Rodrigues³, Anamaria Aranha Camargo^{1,4}, Rodrigo Oliva Perez^{3,4,5}, Raphael Bessa Parmigiani¹

[1] Centro de Oncologia Molecular, Hospital Sírio-Libanês, São Paulo, Brazil

[2] Fundação Antônio Prudente, São Paulo, Brazil

[3] Angelita & Joaquim Gama Institute, São Paulo, Brazil

[4] Ludwig Institute for Cancer Research, São Paulo, Brazil

[5] University of Sao Paulo, São Paulo, Brazil

Financial support: Two grants (2009/54349-9 and 2009/17785-5) from São Paulo Research Foundation (FAPESP).

ABSTRACT

Purpose: Neoadjuvant chemoradiotherapy (nCRT) followed by radical surgery is the preferred treatment strategy for locally advanced rectal cancer. However, complete tumor regression is observed in a significant proportion of patients after nCRT, making them ideal candidates for alternative treatment strategies to radical surgery. Still, identification of such patients based on clinical findings (complete clinical response - cCR) is difficult mainly because

it relies on subjective clinical and imaging studies. Our goal was to identify biomarkers capable of predicting complete response to nCRT.

Experimental Design: We analyzed miRNA expression profile using deep sequencing in rectal tumor biopsies prior to nCRT.

Results: We found 4 miRNAs differently expressed between complete and incomplete responders to nCRT. In addition, validation was performed using an independent group of patients and miR-21-5p was confirmed as being overexpressed in complete responders. Overall sensitivity and specificity of miR-21-5p expression in predicting complete response to nCRT was 78% and 86% respectively. Interestingly, in a subset of patients with cCR followed by early local recurrence the expression level of miR-21-5p was considerably low, similarly to incomplete responders. Finally, we analyzed the expression of miR-21-5p target genes in primary rectal tumors and found SATB1 expression being inversely correlated with miR-21-5p expression. SATB1 is a known multidrug resistance gene and may be involved with poor response to nCRT in rectal cancer patients.

Conclusions: This study suggests miR-21-5p as a promising predictive biomarker, which should aid in the selection of patients with cCR to nCRT that potentially could be spared from radical surgery.

INTRODUCTION

Treatment of locally advanced rectal cancer includes in most cases a combination of surgery and radiotherapy with or without chemotherapy (CRT and RT, respectively). In recent years, preoperative delivery of RT or CRT has become one of the preferred initial treatment for the management of rectal cancer due to improved local disease control and significant tumor regression (including tumor downstaging and downsizing) (1–3).

Response to neoadjuvant chemoradiation (nCRT) varies substantially among patients, and ultimately, complete pathological response (pCR) may develop in up to 42% of them (4). These patients with pCR are associated with excellent oncological outcomes, particularly in terms of local disease control. Considering that radical surgery does not remove any residual

cancer in these patients, alternative treatment strategies to this procedure have been considered in an effort to avoid unnecessary postoperative morbidity, mortality and functional consequences associated with total mesorectal excision (TME). A strategy of close observation without immediate surgery, known as "Watch and Wait", has been proposed for highly selected patients with no clinically or radiologically detectable residual tumor (complete clinical response – cCR) (5). This strategy requires an intensive follow-up in which patients are submitted to frequent clinical, endoscopic and radiological assessments after completing nCRT (6).

However, clinical and radiological assessment of tumor response remains a significant challenge due to their subjectivity and inherent limitations of currently available studies. In this setting, identification of molecular markers capable of predicting complete response to nCRT would not only allow accurate selection of patients that benefit the most from nCRT but also identify ideal candidates to alternative treatment strategies without immediate radical surgery after achieving cCR.

miRNAs are small non-coding RNAs (18-25nt) capable of regulating mRNAs post-transcriptionally by inducing their destabilization or translational repression (7). miRNAs play important regulatory roles in several cellular processes such as cell proliferation, differentiation, and apoptosis. Changes in miRNA expression profile have been reported for several types of cancer, due to different mechanisms, such as deletion, amplification, mutation or epigenetic changes (8). In fact, miRNAs may contribute to the tumorigenic process functioning as tumor suppressor genes or oncogenes depending on the genes they regulate (9,10).

Importantly, different studies have also investigated how miRNA profiling may be associated with treatment response in different tumor types and how *in vitro* modulation of miRNA expression can change tumor sensitivity to chemo and radiotherapy (11–13). Besides the increasing evidences of association between miRNA expression and treatment response in cancer patients, the high stability of miRNAs motivate the search for their role as predictive molecular markers in cancer. The goal of the

present study was to compare miRNA expression profile of treatment-naive tumor biopsies derived from rectal cancer patients with complete or incomplete response to nCRT in order to identify specific miRNAs as predictive biomarkers.

MATERIAL AND METHODS

Patients and Biological samples

Patients with biopsy-proven rectal adenocarcinoma, cT2-4N0-2M0 located no more than 7cm from the anal verge measured by rigid proctoscopy were eligible for the study. All patients were radiologically staged using high-resolution pelvic MRI, abdominal and chest CT scans and CEA levels. Patients with metastatic disease at baseline staging, pregnancy or under the age of 18 were excluded from the study.

Endoscopic biopsies were taken from primary tumors, properly identified and snap frozen using liquid nitrogen prior to storage at -80° C. Tumor fragments were verified for the presence of ≥80% viable cancer cells prior to RNA extraction using standard hematoxylin-eosin staining. Specimens with <80% cancer cells were dissected to result in ≥80% of cancer cells. Fragments with <80% of cancer cells after dissection were excluded from the study.

This study was approved by local IRB (Hospital Alemão Oswaldo Cruz, São Paulo, Brazil). Informed consent was obtained from all study participants prior to sample collection.

Treatment and Assessment of Response

All patients underwent nCRT as described elsewhere (5). Briefly, nCRT consisted of 50.4-54Gy of radiation and concomitant 5FU-based chemotherapy. Patients were clinically reassessed for tumor response at least 8 weeks from nCRT completion using digital rectal examination, rigid proctoscopy, pelvic MRI and CEA levels (6). Patients with evidence of residual disease (incomplete clinical response) such as residual ulcers,

irregularity, mass or stenosis at clinical examination were referred to immediate radical surgery. Likewise, patients with radiological evidence of residual disease including the presence of nodal metastases or residual primary cancer were also referred to radical surgery. Patients with clinical, endoscopic or radiological evidence of complete clinical response were recommended no immediate radical surgery and were enrolled in a strict follow-up program (Watch & Wait strategy) as described elsewhere (5).

Groups for comparison

Patients were grouped accordingly to response to nCRT based on clinical and pathological findings. The complete responders group included patients with clinical evidence of complete response (cCR) sustained for at least 24 months of follow-up and patients with complete pathological response (pCR) who underwent radical surgery due to inability to rule out residual disease. The incomplete responders group included patients with significant residual disease detected after pathological examination of the resected specimen, including ypT2-4 or ypN+ and >10% residual cancer cells (tumor regression grades - TRG0-2 according to Dworak's classification (14). To increase the possibility of identifying predictive miRNAs, we concentrated our study on patients with extreme responses to nCRT. Therefore, patients with incomplete but with "near-complete" response including ypT1N0 or ≤10% residual cancer cells (TRG3) were excluded from the study. Differential miRNA expression analyses were performed comparing the complete responders to incomplete responders.

miRNA sequencing and analysis

Total RNA was extracted with Trizol reagent (Invitrogen), and RNA quality was evaluated on 2100 Bioanalyzer (Agilent). All samples had RNA integrity number (RIN) above 6. Total RNA (10µg) was enriched for small RNAs (up to 250bp) using PureLink™ miRNA Isolation Kit (Invitrogen). miRNA libraries were prepared using SOLiD™ Total RNA-Seq Kit (Life

Technologies), accordingly to the manufacturer's recommendations and were sequenced on SOLiD4 (Life Technologies) in 35 bp single read runs.

Sequencing data was analyzed using the CLC Genomics Workbench 5.1 software (CLC Bio). Initially, sequencing adapters were trimmed from sequencing reads. The remaining sequences were mapped against the human miRNA database, miRBase (www.mirbase.org, release 18) (15,16), allowing one mismatch. At last, known mature miRNAs were annotated according to miRBase and the number of sequences for each miRNA was used for differential expression analysis. Normalization was done by dividing the number of sequences of a certain miRNA by the total number of miRNA mapped sequences for a given sample and multiplying by one million. This resulted in a value corresponding to the number of sequences (counts) per million (cpm) for each miRNA. After normalization, low expressed miRNAs were filtered from further analysis to increase detection power of the statistic tests. Only miRNAs with a minimum of 20 cpm in at least seven samples were kept in the analyses.

Differential expression between the two groups of patients was performed with EdgeR (version 2.6.7) (17) available on Bioconductor 2.10 (18). P-values were adjusted for multiple testing using the Benjamini and Hochberg method to control the false discovery rate (FDR) (19). Only miRNAs with p-value under 0.001 and FDR under 0.05 were considered as differently expressed.

Clustering and statistical analysis

After identifying differentially expressed miRNAs, hierarchical clustering and bootstrap statistical analysis were performed using the functions 'heatmap' and 'pvclust' available in the R statistical software (version 3.0.1). These analyzes were based on Euclidean distance and average linkage clustering using cpm normalized data. For the same list of differentially expressed miRNAs we performed a Principal Component Analysis (PCA) using the Matlab language function 'princomp' (version R2009b). PCA was carried out by eigenvalue decomposition of the covariance matrix after

mean centering the data matrix for each attribute. The first principal component (*i.e.*, the direction along which miRNAs expression show the largest variance) and the second principal component (*i.e.*, the next best direction uncorrelated with first one) were retrieved and plotted to illustrate similarities between samples.

Receiver operating characteristic (ROC) curve

ROC curve for each miRNA was generated using cpm values obtained from RNA-Seq data, using GraphPad Prism software (version 4.03). To determine accuracy to predict nCRT, the area under the ROC curve (AUC) was calculated. The cut off expression values (cpm) were defined in order to maximize the number of correctly classified samples. Finally, sensitivity and specificity based on this cut off were calculated.

miR-21-5p putative target genes

We initially used Targetscan software (www.targetscan.org) to search for predicted miR-21-5p targets (20). We considered only conserved sites in conserved target genes. Using whole transcriptome data from nineteen of the primary rectal tumors used in this study (unpublished data), we searched for the expression of predicted targets that were inversely correlated (Pearson correlation) to miR-21-5p expression.

SATB1 gene expression was also investigated on all in-house whole transcriptome data available from rectal tumors, including 9 complete responders and 19 incomplete responders. Finally, SATB1 gene expression was evaluated through qPCR in 11 samples (3 complete responders and 8 incomplete responders). Reactions were performed in triplicates using the following primers: SATB1Fw: GGTACAAACATTTCAAGAAGAC and SATB1Rev: CATGATTGGCGCCTTGCT. PUM1 and HMBS genes were used for normalization (PUM1Fw: TGTACTIONTACGAAGAGTTGCGATGTG PUM1Rev: CCAGGCCAGCGGAAGAT; HMBSFw: GGCAATGCGGCTGCAA HMBSRev: GGGTACCCACGCGAATCAC).

RESULTS

Training and validation sets of patients

Clinical and demographics features of all patients included in this study are summarized in Table 1. Overall, 43 patients were included in our study and had biopsies collected prior to nCRT. Initially, 27 patients were analyzed in the training set, including 7 complete responders and 20 incomplete responders (as described in Material and Methods).

In addition, 12 patients were analyzed as a validation set, being 7 complete responders, 5 incomplete responders. Finally, 4 patients that were initially classified as complete clinical responders but developed early local recurrence (within the first 16 months of follow-up after nCRT) were also included in this validation set.

miRNA differential expression to predict nCRT

Primary tumor biopsies were collected before nCRT and used for RNA extraction followed by small RNA sequencing as described in material and methods. On average, 37 million sequences were generated for each sample (Table S1). After mapping against miRBase and annotation, the expression of 711 mature miRNAs was detected per sample (Table S1). The number of sequences for each specific miRNA was used to estimate its expression. After removing low expressed miRNAs (details in Material and Methods), differential expression analysis was performed with 331 unique mature miRNAs (Table S1).

The first approach to identify molecular markers capable of predicting tumor response to nCRT was to perform differential miRNA expression analysis using the 27 samples from the training set. Four miRNAs were identified as differentially expressed (p -value <0.001 and $FDR < 0.05$) when comparing complete responders to incomplete responders: miR-205-5p, miR-21-5p, miR-1246, and miR-1290-3p (table 2). miRNAs overexpressed in complete responders (miR-21-5p, miR-1246, and miR-1290-3p) showed a 3.3 to 3.8 fold change of expression relative to incomplete responders. The

only downregulated miRNA, miR-205-5p, showed a 0.002 fold change comparing complete responders to incomplete responders.

After determining differentially expressed miRNAs, we evaluated their efficiency in discriminating complete responders from incomplete responders' patients. First, we performed a hierarchical clustering, which is capable of grouping samples based on the similarity of their expression profile. Clustering analysis resulted in two major groups, one containing only incomplete responders and the other one with all complete responders plus 3 incomplete responders (Figure 1a). Second, the cpm values of the 4 differentially expressed miRNAs were used to perform principle component analysis (PCA) in order to reduce data dimension and to provide informative visualization of clusters. PCA revealed good separation between complete and incomplete responders (Figure 1b). Consistent with the hierarchical clustering analysis, the same 3 incomplete responders showed greater similarity with patients from the complete responders group.

miRNAs as potential predictive markers

We have also investigated the potential of each of the 4 differently expressed miRNAs to be individually used as a predictive biomarker. In order to estimate the efficiency on correctly classifying patients accordingly to treatment response, we calculated the area under the ROC curve (AUC) for each miRNA. We found an AUC of 0.63, 0.94, 0.74 and 0.70, for miR-205-5p, miR-21-5p, miR-1246, and miR-1290-3p, respectively (Figure 2). Also, the expression cut off point was determined for each miRNA to maximize sensitivity and specificity (Figure 2). Considering the normalized values (in cpm) obtained by RNA-seq, the established cut off for each miRNA resulted in a sensitivity and specificity of 86% and 55% for miR-205-5p, 100% and 85% for miR-21-5p, 86% and 65% for miR-1246 and 71% and 75% for miR-1290-3p, respectively. These results suggested miR-21-5p as the best putative predictive marker.

Validation of miRNA differential expression in independent samples

In order to verify whether the differential expression of the 4 miRNAs between complete and incomplete responders was not limited to the training set of samples, we have also analyzed a validation set of samples, including 7 complete responders and 5 incomplete responders. miRNAs libraries from these primary tumors were sequenced and global differential expression analysis between the two groups of patients was performed. Thus, we found miR-21-5p as a commonly differently expressed miRNA (p -value < 0.01). The other 3 miRNAs didn't show different expression in the validation set of samples and therefore, were not used in the following analyses.

To further validate miR-21-5p as a predictive marker for nCRT, we also analyzed its expression in samples from patients that initially presented cCR to nCRT and were spared from immediate surgery but during follow up visits, presented early local recurrence requiring salvage resection (named here as recurrence group). These patients most definitely had clinically undetectable residual disease after nCRT. Noteworthy, miR-21-5p expression pattern in these patients was very similar to incomplete responders (Figure 3). More importantly, miR-21-5p expression in these two groups of patients (incomplete responders and recurrence group) was significantly lower than the complete responders group (p -value < 0.01).

After this independent validation, we determined the predictive power of miR-21-5p expression by using the cutoff established on the ROC curve and all the samples included in the study (27 from training set, 12 from the validation set and 4 from the recurrence group). For this estimation we considered the recurrence group as incomplete responders. Overall, miR-21-5p expression presented 78.5% (11/14) sensitivity and 86% (25/29) specificity to identify patients with complete response to nCRT.

Potential role of miR-21 on treatment response

In order to investigate whether miR-21-5p expression is directly involved in treatment response, we analyzed the expression pattern of its

target genes in primary rectal tumors using whole transcriptome RNA-seq data from a subset of the patients' samples (6 complete responders and 13 incomplete responders – unpublished data). Using Target Scan software, we found 307 miR-21-5p targets genes whose context+ score varied from -0.76 to >-0.01. Considering the negative regulation of miRNAs over their target genes expression, we searched for predicted targets for which the expression pattern was inversely correlated to miR-21-5p expression. Using this approach, we found a strong negative correlation ($r = -0.5$ and p -value = 0.03; Pearson correlation) for SATB1 (context+ score = -0.31). This is a global gene regulator that has been reported to confer malignant behavior and multidrug resistance (MDR), as well as being associated with poor prognosis in several cancer types, including rectal cancer (21–23). It has already been shown that miR-21-5p can negatively regulate SATB1 expression (24,25). Indeed, we have also experimentally confirmed this regulation *in vitro* by knocking down miR-21 expression in HCT116 colon cancer cell line and observing an increased expression of SATB1 (data not shown).

Further exploring all in-house whole transcriptome data we had available from rectal cancer patients, including 19 that were also used in the miRNA expression analysis plus 8 independent patients, we compared SATB1 expression between complete and incomplete responders. Accordingly to our hypothesis, we found a significantly higher SATB1 expression in incomplete responders (p -value = 0.02) (Figure 4b). In addition, we have also evaluated SATB1 expression by qPCR in a few primary tumors that still had biological tissue available. A higher SATB1 expression in incomplete response patients was once again observed. However, possibly due to the small number of samples available (3 complete responders and 8 incomplete responders), this difference was not statistically significant (p -value = 0.13) (supplement Figure 1). These results suggest SATB1 as a potential candidate for functional studies, which could elucidate the possible direct involvement of miR-21-5p on response to nCRT in rectal cancer patients.

DISCUSSION

Standard treatment of locally advanced rectal cancer may include nCRT followed by radical surgery in most patients. However, the observation of pCR in up to 42% of patients undergoing nCRT has challenged the role of radical surgery in this setting and alternative treatment strategies for these highly selected patients have been suggested (4,5,26). Identification of these patients by clinical, endoscopic and radiological assessments may be very difficult and remains restricted to specific dedicated centers. Therefore, most of these patients still undergo potentially unnecessary radical surgery including total mesorectal excision with or without a permanent stoma. Therefore, molecular markers capable of accurately predicting tumor response to nCRT would be of great clinical significance since these patients could be spared from radical surgery and from morbidities associated with such aggressive procedure.

Previous studies searching for predictive molecular markers in rectal cancer patients have used microarray and qPCR to evaluate mRNA or miRNA profiling (27–34). However, patients' response classification varied considerably among these studies. The “Responders” group frequently included complete pathological responders and partial responders (based on T-level downstaging or tumor regression grades). In addition, although it has been shown that pCR rates may considerably increase up to 12 weeks after nCRT, most of these studies have assessed tumor response at considerably short intervals from nCRT (6 weeks on average) (27,29,30,35). Thus, comparison between studies may be quite difficult and rather inconclusive. Regarding miRNA profiling, three recent studies have identified signatures with good accuracy to predict nCRT response (32–34). However, one of them used only 12 samples and all of them didn't validate their signatures in an independent group of patients. Importantly, comparing the molecular signatures, there is no overlapping miRNA among them, suggesting that these signatures might be characteristic of the samples used in each study.

Here we performed global miRNA sequencing of primary rectal tumor samples from patients with complete and incomplete response to nCRT

followed by differential expression analysis. This is the first study to use, as complete response group, samples from patients with pCR as well as with sustained cCR to nCRT. In addition, the exclusion of patients with “near-complete” responses rather than including them in complete or incomplete responder groups allowed us to compare only patients/tumors with significant clinical discrepancies in their outcomes (response to nCRT), which potentially favor the identification of molecular differences. Using this approach we identified 4 differently expressed miRNAs (miR-205-5p, miR-21-5p, miR-1246, and miR-1290-3p) between complete and incomplete responders. Based on their expression, we performed hierarchical clustering and PCA to group patients with similar expression profiles. Only 3 out of 20 (85%) incomplete responders were incorrectly clustered with the complete responders. More importantly, all 7 complete responders were clustered together. The predictive power of each miRNA was also estimated through ROC curves and miR-21-5p showed the best accuracy with an AUC of 0.94.

miRNA differential expression was further validated in 12 additional samples. Among the 4 miRNAs, only miR-21-5p maintained a significant difference in expression levels, being overexpressed in patients with complete response to nCRT. Interestingly, in all patients with initial cCR followed by early local recurrence, miR-21-5p showed an incomplete response expression pattern. As mentioned, these patients are currently very difficult to identify and molecular markers useful for their identification are truly needed. Thus, if miR-21-5p expression was used to help physicians to decide whether to operate patients or not, they would have been submitted to surgery immediately after the resting period post-nCRT.

When used as a predictive marker on all samples included in the study, miR-21-5p expression showed high sensitivity (78.5%, 11/14) and specificity (86%, 25/29) on the identification of complete responders. Similarly, miR-21-5p expression also showed good positive (73%, 11/15) and negative predictive values (89%, 25/28). Importantly, considering only the complete pathological response patients, 4 out of 5 (80%) were correctly predicted based on miR-21-5p expression, suggesting that a significantly

high percentage of patients could benefit from such biomarker and be spared from radical surgery. Although we have studied a limited number of samples, this is one of the few studies that have used a validation set of samples to confirm the differential expression observed in a training set. Thus, miR-21-5p emerges as a promising predictive to nCRT that still needs to be further validated in a larger cohort.

miR-21-5p is overexpressed in different tumor types, including colorectal cancer (CRC), and also acts as an oncogenic miRNA (oncomiR) with important roles in cell proliferation, apoptosis and invasion (36–38). Therefore, miR-21-5p has been considered a promising diagnostic/prognostic biomarker (39–41). However, the literature is controversial regarding miR-21-5p expression and treatment response in CRC. Although several *in vitro* studies using CRC cell lines have shown that overexpression of miR-21 induces resistance to chemotherapy (42–44), few studies evaluated miR-21-5p expression in CRC primary tumors. In one of these studies, high miR-21-5p expression was associated with colon adenocarcinoma advanced stages and poor survival (45). Moreover, for patients treated with adjuvant chemotherapy, high miR-21-5p expression predicted worse overall survival, suggesting an association between high miR-21-5p expression and poor therapeutic outcome. However, comparison between these results and ours is not straightforward, since they investigated miR-21-5p expression in the context of colon adenocarcinoma and adjuvant chemotherapy. First, it is not clear whether colon and rectal adenocarcinomas have similar biological behavior. Also, neoadjuvant treatment with radiation and 5FU used as a sensitizing agent may be quite different from adjuvant systemic therapy using variable combinations of chemotherapy/biological agents. Finally, response to nCRT may provide a useful and short-term end-point that is quite different from long-term survival outcomes in patients undergoing surgery for colon cancer followed by adjuvant chemotherapy. All of these discrepancies lead to inherent limitations in comparisons between studies and possibly yielding controversial findings.

miRNAs' functions are usually inferred through the identification of their target genes. Considering the negative regulation imposed by miRNAs it can be assumed that they perform an opposite role of their targets. Searching for the expression of miR-21-5p targets in our samples, we found an interesting negative correlation for SATB1 gene and miR-21-5p expression. As expected, when we compared SATB1 expression among complete and incomplete responders, a significantly higher expression was observed in incomplete responders. SATB1 is a chromatin modifier and transcription factor and it has also been shown to be a multidrug resistance gene in several cancer types (22,46,47). Although very interesting, our expression analysis results are still only suggestive and future functional studies must be performed to confirm whether miR-21-5p/SATB1 plays a direct role in the response to nCRT.

In conclusion, we have identified miR-21-5p as a promising predictive marker for response to nCRT in rectal cancer patients. Patients with cCR based on very stringent criteria and high levels of miR-21-5p expression may be ideal candidates for alternative treatment strategies to radical surgery including a watch and wait approach. Further validation of these findings in larger cohorts may lead to the inclusion of this miRNA-based test in standard workup of patients prior to nCRT and help in clinical decision-making for patients with apparent cCR.

Table 1. Clinical and demographics data of patients included in the study.

	Training set			Validation set			
	Complete responders	Incomplete responders	p	Complete responders	Incomplete responders	Recurrence	p
N	7	20		7	5	4	
Age (years)	58.9 ± 10.9	54.2 ± 13.7	ns	60.4 ± 13.6	59.6 ± 14.5	69.3 ± 15.0	ns
Gender (M/F)	2/5	9/11	ns	3/4	3/2	3/1	ns
Distance anal verge (cm)	4.0 ± 2.6	4.3 ± 2.2	ns	4.6 ± 2.4	2.2 ± 2.8	4 ± 0.8	ns
Tumor size (cm)	4.4 ± 0.8	4.6 ± 1.5	ns	3.6 ± 0.9	5 ± 1.9	5 ± 1.4	ns
Initial staging			ns				ns
T2	0	2		4	2	0	
T3	7	17		2	2	4	
T4	0	1		1	1	0	
N0	4	7		2	2	2	
N+	3	13		5	3	2	
Staging after CRT							
cCR	3	-		6	-	4	
pCR	4	-		1	-	-	
ypT2	-	6		-	1	-	
ypT3	-	13		-	4	-	
ypT4	-	1		-	0	-	
ypN0	-	12		-	1	-	
ypN+	-	8		-	2	-	
ypNx	-	0		-	2	-	

Table 2. Differentially expressed miRNAs. miRNAs mean expression is shown for each response group to nCRT and fold change of complete responders relative to incomplete responders. Statistical significance can be verified by p-values and false discovery rate (FDR), using Benjamini and Hochberg's method to adjust for multiple testing.

miRNA	Complete response group mean expression (cpm)	Incomplete response group mean expression (cpm)	Fold change	P-value	FDR
miR-205-5p	3.67	2,099.32	0.002	6.21E-06	0.002
miR-21-5p	106,020.55	31,438.53	3.372	1.99E-04	0.028
miR-1290-3p	115.14	30.02	3.835	2.65E-04	0.028
miR-1246	65.58	19.60	3.346	3.36E-04	0.028

FIGURES

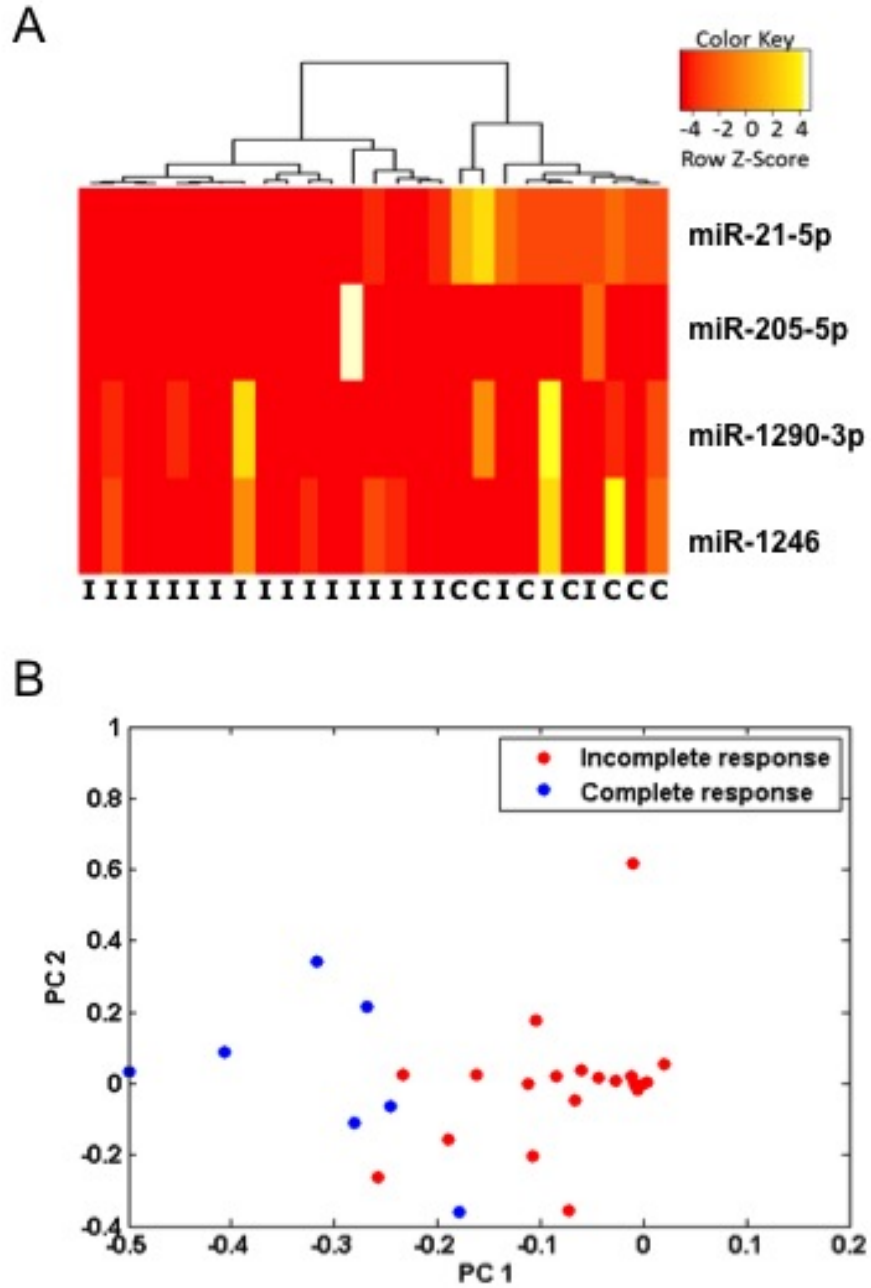


Figure 1. Patients' clustering based on the expression of 4 differently expressed miRNAs. A) Hierarchical clustering heatmap based on Euclidean distance and average-linkage clustering for incomplete responders (I) and complete responders (C). B) Principal component analysis (PCA) was conducted for 27 samples (patients) and 4 variables (differently expressed miRNAs). Blue dots represent complete responders and red dots represent incomplete responders.

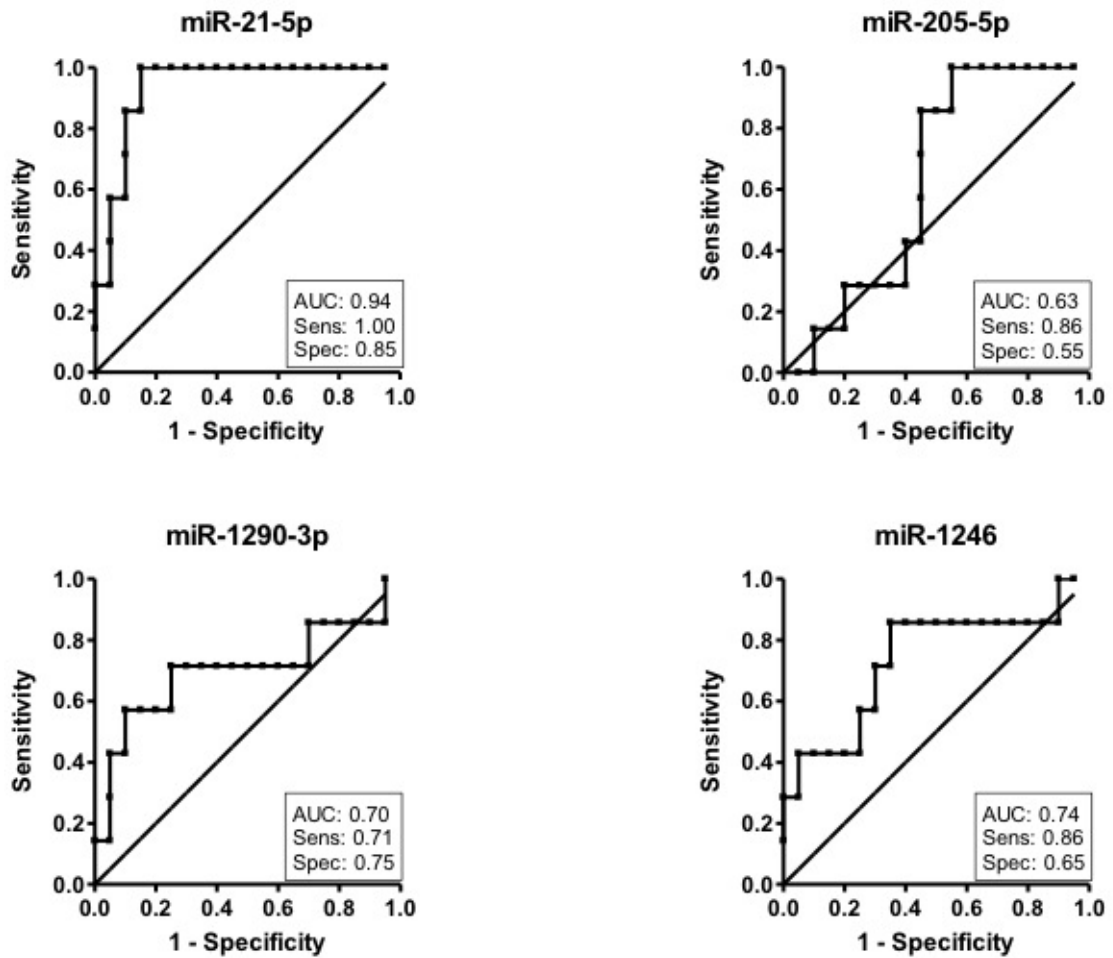


Figure 2. Accuracy to predict nCRT based on the expression of differently expressed miRNAs. Receiver operating characteristic (ROC) curve based on miRNA expression determined by RNA-Seq. Area under the ROC curve (AUC), sensitivity (Sens) and specificity (Spec) were calculated and displayed for each miRNA. miR-21-5p showed the best accuracy to predict complete response to nCRT.

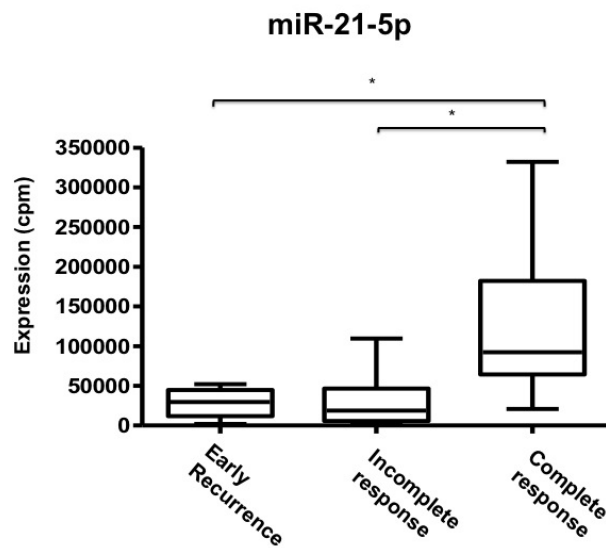


Figure 3. miR-21-5p expression in all samples included in the study. There is a significant difference between complete and incomplete responders (p value < 0.05). Interestingly, patients with initial complete clinical response followed by early local recurrence showed similar miR-21-5p expression to incomplete responders. miR-21-5p expression is giving as normalized values (cpm) from RNA-Seq data.

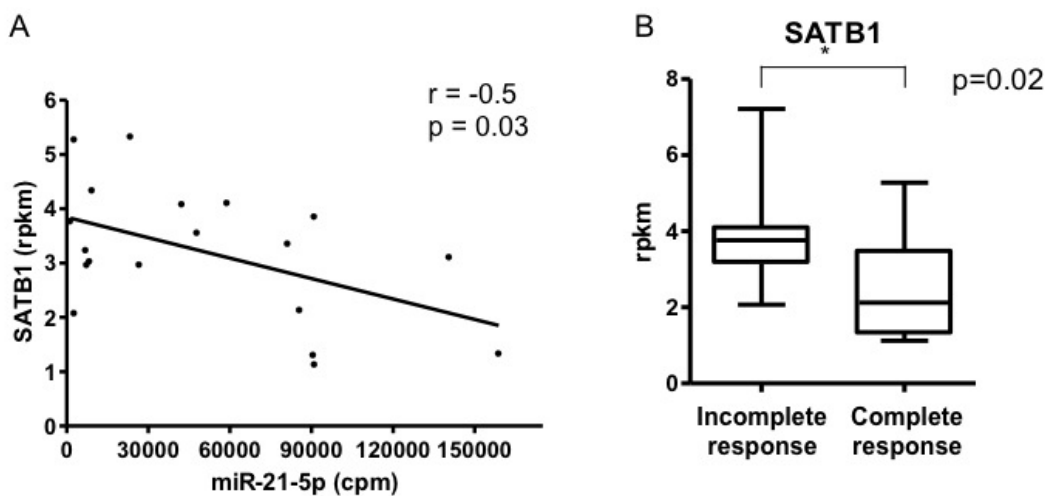


Figure 4. Expression analysis of SATB1 gene in patients' samples. A) Correlation between SATB1 and miR-21-5p expression in the same primary rectal tumor samples. SATB1 and miR-21-5p expression was evaluated through RNA-Seq in 13 incomplete response and 6 complete response patients. r : Pearson correlation factor. B) SATB1 gene expression accordingly to treatment response. All whole transcriptome data from rectal tumors available was used, including data from 9 complete responders and 19 incomplete responders to nCRT. Mann-Whitney test was applied to compare SATB1 median expression between the two groups of patients.

Supplemental material

Table S1. Summary of miRNA sequencing and mapping data for all patients. miRNA sequences were obtained on SOLiD4 sequencer and mapped to miRBase v18 database (details in material and methods).

Patient	Total reads	Reads mapped to mature miRNAs	Percentage of reads mapped to mature miRNAs	Number of miRNAs detected*	Number of miRNAs used for differential expression**
1	50,172,700	5,486,141	10.93	719	256
2	29,444,051	5,222,439	17.74	583	258
3	30,222,945	4,638,707	15.35	798	288
4	54,158,517	7,451,686	13.76	770	288
5	53,618,746	4,507,681	8.41	805	304
6	31,091,042	3,936,602	12.66	711	314
7	42,737,245	4,839,742	11.32	760	288
8	26,184,560	5,237,994	20.00	736	309
9	26,934,474	3,903,521	14.49	695	264
10	31,528,426	4,381,148	13.90	605	278
11	37,032,364	5,546,587	14.98	857	332
12	33,610,904	3,926,311	11.68	718	281
13	54,550,819	3,614,647	6.63	809	302
14	32,084,705	5,926,807	18.47	734	264
15	31,452,650	7,465,445	23.74	537	282
16	26,986,387	6,149,544	22.79	824	295
17	45,907,246	2,815,616	6.13	603	295
18	30,073,433	4,623,558	15.37	705	272
19	44,521,292	2,362,828	5.31	683	299
20	41,177,074	3,416,683	8.30	588	325
21	26,339,749	4,405,923	16.73	758	281
22	30,194,181	6,470,087	21.43	759	271
23	42,780,410	2,449,337	5.73	674	264
24	33,446,117	3,045,721	9.11	616	262
25	29,357,756	2,699,764	9.20	782	316
26	51,372,267	3,257,262	6.34	769	294
27	44,310,006	2,297,596	5.19	587	281
28	50,715,095	4,841,085	9.55	818	268
29	19,061,826	2,268,017	11.90	649	260
30	59,729,745	2,360,510	3.95	714	280
31	45,731,316	2,289,984	5.01	809	302
32	33,644,141	4,348,408	12.92	889	287

33	59,101,716	4,430,414	7.50	936	306
34	93,193,034	5,929,896	6.36	912	262
35	41,123,068	3,264,029	7.94	840	286
36	23,367,580	2,565,471	10.98	742	288
37	34,092,250	4,760,712	13.96	843	299
38	133,185,797	5,155,644	3.87	837	278
39	32,079,436	6,146,284	19.16	912	256
40	30,518,047	5,638,091	18.47	870	270
41	30,606,777	3,436,187	11.23	788	276
42	32,985,246	4,478,279	13.58	956	293
43	31,271,842	4,339,892	13.88	897	279
Average	40,969,697	4,333,309	12.00	758	285

* At least 1 read **More than 20 cpm

Supplemental Figure

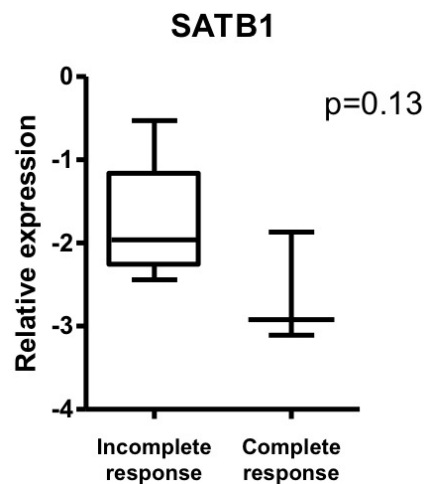


Figure S1: Evaluation of SATB1 gene expression by qPCR. Only few patients' samples were available, including 3 complete and 8 incomplete responders. A CRC cell line (HCT116) was used as reference sample and the relative expression was calculated based on $\Delta\Delta CT$ method (48). Mann-Whitney test was applied to compare SATB1 median expression between the two groups of patients.

REFERENCES

1. Kapiteijn E, Marijnen CA, Nagtegaal ID, Putter H, Steup WH, Wiggers T, et al. Preoperative radiotherapy combined with total mesorectal excision for resectable rectal cancer. *N Engl J Med* 2001;345:638–46.
2. Minsky BD, Cohen AM, Kemeny N, Enker WE, Kelsen DP, Reichman B, et al. Combined modality therapy of rectal cancer: decreased acute toxicity with the preoperative approach. *J Clin Oncol* 1992;10:1218–24.
3. Habr-Gama A, Perez RO, Kiss DR, Rawet V, Scanavini A, Santinho PM, et al. Preoperative chemoradiation therapy for low rectal cancer. Impact on downstaging and sphincter-saving operations. *Hepatogastroenterology* 2004;51:1703–7.
4. Sanghera P, Wong DWY, McConkey CC, Geh JI, Hartley A. Chemoradiotherapy for rectal cancer: an updated analysis of factors affecting pathological response. *Clin Oncol (R Coll Radiol)* 2008;20:176–83.
5. Habr-Gama A, Perez RO, Nadalin W, Sabbaga J, Ribeiro U, Silva e Sousa AH, et al. Operative versus nonoperative treatment for stage 0 distal rectal cancer following chemoradiation therapy: long-term results. *Ann Surg* 2004;240:711–7.
6. Habr-Gama A, Perez RO, Wynn G, Marks J, Kessler H, Gama-Rodrigues J. Complete clinical response after neoadjuvant chemoradiation therapy for distal rectal cancer: characterization of clinical and endoscopic findings for standardization. *Dis Colon Rectum* 2010;53:1692–8.
7. Bartel DP, Lee R, Feinbaum R. MicroRNAs: Genomics, Biogenesis, Mechanism, and Function. *Cell* 2004;116:281–297.
8. Croce CM. Causes and consequences of microRNA dysregulation in cancer. *Nat Rev Genet* 2009;10:704–14.
9. Calin GA, Croce CM. MicroRNA signatures in human cancers. *Nat Rev Cancer* 2006;6:857–66.
10. Shenouda SK, Alahari SK. MicroRNA function in cancer: oncogene or a tumor suppressor? *Cancer Metastasis Rev* 2009;28:369–78.
11. Nakajima G, Hayashi K, Xi Y, Kudo K, Uchida K, Takasaki K, et al. Non-coding MicroRNAs hsa-let-7g and hsa-miR-181b are Associated with Chemoresponse to S-1 in Colon Cancer. *Cancer Genomics Proteomics* 2006;3:317–24.
12. Svoboda M, Izakovicova Holla L, Sefr R, Vrtkova I, Kocakova I, Tichy B, et al. Micro-RNAs miR125b and miR137 are frequently upregulated in response to capecitabine chemoradiotherapy of rectal cancer. *Int J Oncol* 2008;33:541–7.
13. Hummel R, Hussey DJ, Haier J. MicroRNAs: predictors and modifiers of chemo- and radiotherapy in different tumour types. *Eur J Cancer* 2010;46:298–311.
14. Dworak O, Keilholz L, Hoffmann A. Pathological features of rectal cancer after preoperative radiochemotherapy. *Int J Colorectal Dis* 1997;12:19–23.

15. Kozomara A, Griffiths-Jones S. miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res* 2011;39:D152–7.
16. Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ. miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res* 2006;34:D140–4.
17. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 2010;26:139–40.
18. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, et al. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* 2004;5:R80.
19. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J R Stat Soc B* 1995;57:289–300.
20. Lewis BP, Burge CB, Bartel DP. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* 2005;120:15–20.
21. Han H-J, Russo J, Kohwi Y, Kohwi-Shigematsu T. SATB1 reprogrammes gene expression to promote breast tumour growth and metastasis. *Nature* 2008;452:187–93.
22. Li Q-Q, Chen Z-Q, Xu J-D, Cao X-X, Chen Q, Liu X-P, et al. Overexpression and involvement of special AT-rich sequence binding protein 1 in multidrug resistance in human breast carcinoma cells. *Cancer Sci* 2010;101:80–6.
23. Meng W-J, Yan H, Zhou B, Zhang W, Kong X-H, Wang R, et al. Correlation of SATB1 overexpression with the progression of human rectal cancer. *Int J Colorectal Dis* 2012;27:143–50.
24. Beyer M, Thabet Y, Müller R-U, Sadlon T, Classen S, Lahl K, et al. Repression of the genome organizer SATB1 in regulatory T cells is required for suppressive function and inhibition of effector differentiation. *Nat Immunol* 2011;12:898–907.
25. Yang S, Banerjee S, Freitas A de, Cui H, Xie N, Abraham E, et al. miR-21 regulates chronic hypoxia-induced pulmonary vascular remodeling. *Am J Physiol Lung Cell Mol Physiol* 2012;302:L521–9.
26. Maas M, Beets-Tan RGH, Lambregts DMJ, Lammering G, Nelemans PJ, Engelen SME, et al. Wait-and-see policy for clinical complete responders after chemoradiation for rectal cancer. *J Clin Oncol* 2011;29:4633–40.
27. Ghadimi BM, Grade M, Difilippantonio MJ, Varma S, Simon R, Montagna C, et al. Effectiveness of gene expression profiling for response prediction of rectal adenocarcinomas to preoperative chemoradiotherapy. *J Clin Oncol* 2005;23:1826–38.
28. Watanabe T, Komuro Y, Kiyomatsu T, Kanazawa T, Kazama Y, Tanaka J, et al. Prediction of sensitivity of rectal cancer cells in response to preoperative radiotherapy by DNA microarray analysis of gene expression profiles. *Cancer Res* 2006;66:3370–4.

29. Kim I-J, Lim S-B, Kang HC, Chang HJ, Ahn S-A, Park H-W, et al. Microarray gene expression profiling for predicting complete response to preoperative chemoradiotherapy in patients with advanced rectal cancer. *Dis Colon Rectum* 2007;50:1342–53.
30. Rimkus C, Friederichs J, Boulesteix A-L, Theisen J, Mages J, Becker K, et al. Microarray-based prediction of tumor response to neoadjuvant radiochemotherapy of patients with locally advanced rectal cancer. *Clin Gastroenterol Hepatol* 2008;6:53–61.
31. Nishioka M, Shimada M, Kurita N, Iwata T, Morimoto S, Yoshikawa K, et al. Gene expression profile can predict pathological response to preoperative chemoradiotherapy in rectal cancer. *Cancer Genomics Proteomics* 2011;8:87–92.
32. Kheirlesei E a H, Miller N, Chang KH, Curran C, Hennessey E, Sheehan M, et al. miRNA expressions in rectal cancer as predictors of response to neoadjuvant chemoradiation therapy. *Int J Colorectal Dis* 2013;28:247-60.
33. Della Vittoria Scarpati G, Falcetta F, Carlomagno C, Ubezio P, Marchini S, De Stefano A, et al. A specific miRNA signature correlates with complete pathological response to neoadjuvant chemoradiotherapy in locally advanced rectal cancer. *Int J Radiat Oncol Biol Phys* 2012;83:1113–9.
34. Svoboda M, Sana J, Fabian P, Kocakova I, Gombosova J, Nekvindova J, et al. MicroRNA expression profile associated with response to neoadjuvant chemoradiotherapy in locally advanced rectal cancer patients. *Radiat Oncol* 2012;7:195.
35. Kalady MF, de Campos-Lobato LF, Stocchi L, Geisler DP, Dietz D, Lavery IC, et al. Predictive factors of pathologic complete response after neoadjuvant chemoradiation for rectal cancer. *Ann Surg* 2009;250:582–9.
36. Yamamichi N, Shimomura R, Inada K, Sakurai K, Haraguchi T, Ozaki Y, et al. Locked nucleic acid in situ hybridization analysis of miR-21 expression during colorectal cancer development. *Clin Cancer Res* 2009;15:4009–16.
37. Chan JA, Krichevsky AM, Kosik KS. MicroRNA-21 is an antiapoptotic factor in human glioblastoma cells. *Cancer Res* 2005;65:6029–33.
38. Ren J, Zhu D, Liu M, Sun Y, Tian L. Downregulation of miR-21 modulates Ras expression to promote apoptosis and suppress invasion of Laryngeal squamous cell carcinoma. *Eur. J. Cancer* 2010;46:3409–16.
39. Wu CW, Ng SSM, Dong YJ, Ng SC, Leung WW, Lee CW, et al. Detection of miR-92a and miR-21 in stool samples as potential screening biomarkers for colorectal cancer and polyps. *Gut* 2012;61:739–45.
40. Toiyama Y, Takahashi M, Hur K, Nagasaka T, Tanaka K, Inoue Y, et al. Serum miR-21 as a Diagnostic and Prognostic Biomarker in Colorectal Cancer. *J Natl Cancer Inst* 2013;105:849–859.

41. Xia X, Yang B, Zhai X, Liu X, Shen K, Wu Z, et al. Prognostic Role of microRNA-21 in Colorectal Cancer: a Meta-Analysis. Cho WC, editor. *PloS One* 2013;8:e80426.
42. Valeri N, Gasparini P, Braconi C, Paone A, Lovat F, Fabbri M, et al. MicroRNA-21 induces resistance to 5-fluorouracil by down-regulating human DNA MutS homolog 2 (hMSH2). *Proc Natl Acad Sci USA* 2010;107:21098–103.
43. Feng Y-H, Wu C-L, Shiau A-L, Lee J-C, Chang J-G, Lu P-J, et al. MicroRNA-21-mediated regulation of Sprouty2 protein expression enhances the cytotoxic effect of 5-fluorouracil and metformin in colon cancer cells. *Int J Mol Med* 2012;29:920–6.
44. Yu Y, Sarkar FH, Majumdar APN. Down-regulation of miR-21 Induces Differentiation of Chemoresistant Colon Cancer Cells and Enhances Susceptibility to Therapeutic Regimens. *Transl Oncol* 2013;6:180–6.
45. Schetter AJ, Leung SY, Sohn JJ, Zanetti KA, Bowman ED, Yanaihara N, et al. MicroRNA expression profiles associated with prognosis and therapeutic outcome in colon adenocarcinoma. *JAMA* 2008;299:425–36.
46. Sun F, Lu X, Li H, Peng Z, Wu K, Wang G, et al. Special AT-rich sequence binding protein 1 regulates the multidrug resistance and invasion of human gastric cancer cells. *Oncol Lett* 2012;4:156–62.
47. Han B, Luan L, Xu Z, Wu B. Expression and biological roles of SATB1 in human bladder cancer. *Tumour Biol* 2013;34:2943–9.
48. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* 2001;25:402–8.

Anexo 2 - Aprovação do Comitê de Ética em Pesquisa (CEP) - A.C. Camargo Cancer Center

São Paulo, 15 de Abril de 2010.

Ao
Dr. Raphael Bessa Parmigiani

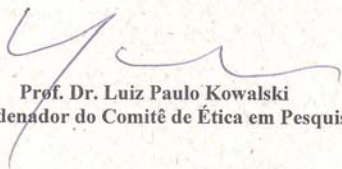
Ref.: Projeto de Pesquisa n°. 1387/10
“Identificação e estudo de expressão de variantes de poliadenilação e de microRNAs através de sequenciamento em larga escala em amostras de tumor colorretal”.

Os membros do Comitê de Ética em Pesquisa em Seres Humanos da Fundação Antonio Prudente – Hospital do Câncer - A.C. Camargo/SP, em sua última reunião de 13/04/2010, **aprovaram** a realização do projeto em referência, datado de 29 de Março de 2010 e tomaram conhecimento dos seguintes documentos:

- Folha de Rosto para Pesquisa Envolvendo Seres Humanos;
- Termo de Compromisso do Pesquisador com as Resoluções do Conselho Nacional de Saúde;
- Termo de Dispensa do Consentimento Livre e Esclarecido;
- Orçamento Financeiro Detalhado;
- Declaração de Ciência do Laboratório Biologia Molecular e Genômica do Instituto Ludwig de Pesquisa sobre o Câncer, Hospital Alemão Oswaldo Cruz, SP.

Informações a respeito do andamento do referido projeto deverão ser encaminhadas à assistente do CEP dentro de 12 meses.

Atenciosamente,


Prof. Dr. Luiz Paulo Kowalski
Coordenador do Comitê de Ética em Pesquisa

PARECER CONSUBSTANCIADO

Projeto: 1387/10

Título: "Identificação e estudo de expressão de variantes de poliadenilação e de microRNAs através de sequenciamento em larga escala em amostras de tumor colorretal"

Pesquisador Responsável: Raphael Bessa Parmigiani

Aluna: Camila Miranda Lopes (Doutorado)

Patrocinador: FAPESP

Comentários gerais:

Trata-se de um projeto de Mestrado. Os pesquisadores propõem avaliar a expressão de variantes de poliadenilação em amostras de tumores colorretais pela metodologia de seqüenciamento em larga escala. Para isso desenvolverão um protocolo de construção de bibliotecas de transcritos enriquecidas para a porção 3' compatíveis com o seqüenciador 454 (Roche) e de microRNAs usando o sistema SOLID (Applied Biosystems). Utilizarão inicialmente duas linhagens celulares de carcinoma colorretal. As variantes de poliadenilação e os miRNAs detectados serão validados por outras estratégias em amostras a fresco (cinco casos).

Embora relevante, os autores informam que todo o protocolo deverá ser implantado. Esta não é uma estratégia trivial e depende de equipe coordenada em métodos extremamente laboriosos e de bioinformática.

Adequação do título: Adequado.

Adequação da revisão de literatura: Adequada.

Adequação dos objetivos frente à literatura: Adequados.

O objetivo geral é "identificar e estudar a expressão de variantes de poliadenilação e de miRNA utilizando seqüenciamento em larga escala em amostras de tumor colorretal.

Desenho e metodologia do projeto (grupos experimentais, procedimentos, indicadores de resultado, tipo de estudo, fase da pesquisa):

Pretende-se utilizar duas linhagens celulares para os experimentos iniciais usando os seqüenciadores de alta performance. Após as análises de bioinformática, serão utilizadas 5 amostras a fresco (Banco de Tumores do Hospital Oswaldo Cruz) para avaliar o padrão de expressão dos variantes de poliadenilação e dos miRNA.

O cronograma de execução não inclui a análise de bioinformática que é crucial neste estudo.

Critérios de inclusão:

Serão utilizadas duas linhagens celulares (SW480 e HCT116) de carcinomas colorretais.

Os autores propõem analisar amostras normais e tumorais de 5 pacientes com câncer colorretal já coletadas e aprovadas pelo Comitê de Ética do Hospital Alemão Oswaldo Cruz, SP. Não é informado qual o critério de seleção destes casos.

Critérios de exclusão:

Não informados.

Riscos e benefícios aos sujeitos: Não se aplica

Termo de Consentimento Livre e Esclarecido:

Foi solicitada a dispensa do Termo de Consentimento por se tratar de pesquisa retrospectiva.

O projeto foi aprovado pelo CEP do Hospital Alemão Oswaldo Cruz, SP, onde estão incluídas 20 casos como número amostral.

Análise crítica de riscos: N/A

Duração total da pesquisa: 36 meses

Explicitação das responsabilidades:

- do pesquisador: adequado
- do patrocinador: adequado

Explicitação dos critérios para suspender ou encerrar a pesquisa: N/A

Local de pesquisa:

- Laboratório de Biologia Molecular e Genômica do Instituto Ludwig de Pesquisa sobre o Câncer, Hospital Alemão Oswaldo Cruz, SP.

Fonte financeira:

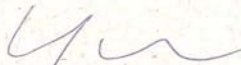
O presente projeto foi submetido para análise na FAPESP (Orçamento estimado em R\$92500,00).

Análise da Comissão de Pesquisa: Aprovado

A proposta do estudo é avaliar a expressão de variantes de poliadenilação em amostras de tumores colorretais pela metodologia de seqüenciamento em larga escala. Os autores propõem o desenvolvimento de um protocolo de construção de bibliotecas de transcritos enriquecidas para a porção 3' compatíveis com o seqüenciador 454 (Roche). Inicialmente devem utilizar duas linhagens celulares de carcinoma colorretal e depois amostras a fresco. As variantes de poliadenilação e os miRNAs detectados serão validados por outras estratégias. Trata-se de uma abordagem muito interessante e que poderá contribuir com a pesquisa básica da área. Entretanto, todo o protocolo, como apresentado pelos autores, deverá ser implantado e não se trata de estudo de fácil execução e análise. Este seria um projeto adequado para um pós-doutorado, por exemplo. Os autores informam que um especialista em bioinformática participará do estudo.

Parecer Final:

Projeto Aprovado.



FUNDAÇÃO ANTONIO PRUDENTE
Hospital do Câncer A. C. Camargo
Comitê de Ética em Pesquisa
Prof. Dr. Luiz Paulo Kowalski

Anexo 3 - Aprovação do Comitê de Ética em Pesquisa - Hospital Alemão Oswaldo Cruz



PROTÓCOLO DE PESQUISA N.º 02/10	Data de entrada: 05/01/10	Data da resposta:
PARECER do CEP N.º 02/10	Data do parecer: 11/01/10	Data do parecer final:

Parecer Consubstanciado

Título da Pesquisa: "Identificação e estudo de expressão de variantes de poliadenilação e de microRNAs em amostras de tumor colorretal".

Pesquisador Principal: Raphael Bessa Parmigiani

Instituição: Hospital Alemão Oswaldo Cruz

Sumário Geral do Protocolo: Trata-se de pesquisa experimental retrospectiva na área de genética humana. Objetivo Geral: Identificar e estudar a expressão de variantes de poliadenilação e de microRNAs em amostras de tumor colorretal através de sequenciamento em larga escala, utilizando se sequenciadores da nova geração. Avaliar o envolvimento dos microRNAs no desenvolvimento de câncer do intestino grosso. Avaliar as alterações moleculares presentes nestes tumores, que servirão de base para desenvolvimento de novas terapias e avaliação para diagnóstico e prognóstico. Identificar os miRNAs que podem regular diferentes variantes de poliadenilação. Objetivos Específicos: 1-desenvolver protocolo inédito que visa a construção de bibliotecas de cDNAs, enriquecidas para a região 3 dos transcritos e que seja compatível com o sequenciamento em larga escala; 2-identificar e avaliar a expressão de variantes de poliadenilação em linhagens celulares de tumor de cólon através de sequenciamento em larga escala; 3- avaliar o perfil de expressão de microRNAs nas mesmas linhagens celulares e em amostras de tecido normal e tumoral de pacientes com câncer colorretal; correlacionar as variantes de poliadenilação identificadas com a expressão de miRNAs e validar através de PCR em tempo real; 4-avaliar a expressão de microRNAs e variantes de poliadenilação que apresentarem uma maior correlação inversa de expressão em amostras pareadas de pacientes com tumor colorretal.

Considerações/Comentários:

1. Não existem implicações éticas e não haverá nenhuma intervenção no paciente. Serão utilizadas 20 amostras do Banco de Tecidos do HAOC;
2. Já se dispõe de termo de consentimento (TCLE) autorizando o uso dos espécimes em pesquisa, armazenados no Banco de Tecidos; será utilizado também um TCLE específico para este projeto.
3. Trata se de um projeto de Área Temática Especial (Grupo I.1: Genética Humana), porém, não está inserido em nenhuma das situações de genéticas especificadas para este grupo, portanto deverá ser incluído no Grupo II, conforme resolução CNS 340, de 08 de julho de 2004.

Diante dos pontos analisados o projeto foi considerado "APROVADO"

- APROVADO
 APROVADO COM RECOMENDAÇÕES
 NÃO APROVADO
 COM PENDÊNCIAS

(A ausência de resposta em 30 dias acarretará em arquivamento do processo por falta de interesse do pesquisador)


Projeto nº 02/10 Data: 11/01/2010

TEMÁTICA ESPECIAL				SIM	<input checked="" type="checkbox"/>	NÃO	<input type="checkbox"/>	
CONEP	SIM	<input checked="" type="checkbox"/>	NÃO	<input type="checkbox"/>	Completo	<input type="checkbox"/>	Folha de rosto+ Parecer Consubiastanciado	<input checked="" type="checkbox"/>

- Lembramos que os pesquisadores têm como dever enviar relatórios anuais e o relatório final ao CEP-HAOC.
- O CEP também deve ser informado sobre qualquer "alteração" ou "emenda" ao referido projeto, e sobre quaisquer efeitos adversos.

Data do 1º Relatório Parcial: **11/07/2010**

Data do 2º Relatório Parcial: **11/01/2011**


Dr. José Marcio Neves Jorge
Coordenador do Comitê de Ética em Pesquisa
Hospital Alemão Oswaldo Cruz

Anexo 4 - Termo de Consentimento Livre e Esclarecido

 HOSPITAL ALEMÃO
OSWALDO CRUZ
Anexo
Estudo de microRNAs
Câncer Colorretal

TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

(Instruções para preenchimento no verso)

I - DADOS DE IDENTIFICAÇÃO DO SUJEITO DA PESQUISA OU RESPONSÁVEL
LEGAL

1. NOME DO PACIENTE

DOCUMENTO DE IDENTIDADE Nº : SEXO : M F

DATA NASCIMENTO:/...../.....

ENDEREÇO

BAIRRO.....CIDADE.....

CEP:..... TELEFONE: DDD (.....).....

2.RESPONSÁVEL LEGAL.....

NATUREZA (grau de parentesco, tutor, curador etc.).....

DOCUMENTO DE IDENTIDADE :SEXO: M F

DATA NASCIMENTO:/...../.....

ENDEREÇO:

BAIRRO: CIDADE:.....

CEP: TELEFONE: DDD (.....).....



HOSPITAL ALEMÃO
OSWALDO CRUZ

Anexo
**Estudo de microRNAs
Câncer Colorretal**

II - DADOS SOBRE A PESQUISA CIENTÍFICA

1. TÍTULO DO PROJETO DE PESQUISA: **Identificação e estudo de expressão de variantes de poliadenilação e de microRNAs em amostras de tumor colorretal.**

2. INVESTIGADOR PRINCIPAL: Dr Raphael Bessa Parmigiani
CARGO/FUNÇÃO: Pesquisador do Instituto Ludwig de Pesquisa sobre o Câncer.

PESQUISADORES ASSOCIADOS: Dr Angelita Habr-Gama (CRM 8345)¹, Dr Rodrigo Oliva Perez (CRM 93950)¹, Dra Anamaria Aranha Camargo², Dr Sandro de Souza³ e Dr Pedro Galante³.

CARGO/FUNÇÃO: 1- Médicos Cirurgiões e Pesquisadores do Departamento de Gastroenterologia FMUSP e do Hospital Alemão Oswaldo Cruz (HAOC); 2- Diretora Associada do Instituto Ludwig de Pesquisa sobre o Câncer. 3- Pesquisadores do grupo de Bioinformática do Instituto Ludwig de Pesquisa sobre o Câncer.

3. AVALIAÇÃO DO RISCO DA PESQUISA:

SEM RISCO RISCO MÍNIMO RISCO MÉDIO
RISCO BAIXO RISCO MAIOR

(probabilidade de que o indivíduo sofra algum dano como consequência imediata ou tardia do estudo)

A qualquer momento você pode entrar em contato com o Dr Raphael Bessa Parmigiani.
Endereço: Rua João Julião, 245 Tel: 33883248, ou Celular 11-81368656.

O estudo foi analisado e aprovado pelo Comitê de Ética em pesquisa do Hospital Alemão Oswaldo Cruz. Se você tiver quaisquer dúvidas sobre a ética da pesquisa, seus direitos como um paciente de pesquisa ou caso tenha uma lesão relacionada ao estudo, você pode contatar o Comitê de Ética através do telefone (11) 3549-0000 para mais informações.

III - Informações ao Paciente

TÍTULO DO PROJETO DE PESQUISA: Identificação e estudo de expressão de variantes de poliadenilação e de microRNAs em amostras de tumor colorretal.

Você foi convidado a participar de um estudo de pesquisa. Você precisará decidir se quer participar ou não. Por favor, fique a vontade para tomar sua decisão.

Este termo de consentimento pode conter palavras que você não entenda. Leia cuidadosamente o que segue e peça ao médico do estudo ou a alguém da equipe do estudo que explique as palavras ou informações que você não entenda completamente.

Introdução

O câncer de intestino grosso ainda representa uma das doenças malignas, mais frequentes e mortais em todo o mundo. Atualmente, o tratamento convencional é a quimioterapia (algumas vezes associado à radioterapia), seguido de cirurgia. Este tipo de tratamento apresenta uma série de complicações e efeitos colaterais e ainda pode não trazer benefícios evidentes ao paciente. Assim, são necessários estudos que procurem identificar novos alvos terapêuticos para um tratamento mais específico e eficiente da doença. Diferentes tipos de moléculas podem funcionar como alvo terapêutico, dentre elas, os microRNAs têm merecido especial destaque e pouco se conhece sobre a presença dos mesmos nos tumores colorretais. Estas moléculas regulam a atividade dos genes que produzem as proteínas e são, portanto, importante para o funcionamento normal de nossas células. Já foi demonstrada a alteração de microRNA em células neoplásicas de diferentes tipos de tumor.

Objetivo

O objetivo desta pesquisa é avaliar como moléculas específicas denominadas microRNAs estão representadas nas células malignas dos pacientes com câncer do intestino grosso.

Quem pode participar deste estudo?

Todos os pacientes portadores de câncer de intestino grosso (cólon ou reto), maiores de 18 anos, ambos os sexos.

Procedimentos

Caso você aceite participar desta pesquisa, após ter sido plenamente informado sobre o estudo que se realizará, você assinará o termo de consentimento livre e esclarecido e receberá uma cópia do mesmo. Nenhum paciente incluído neste estudo terá sua rotina alterada ou tratamento modificado. Todo o estudo será feito apenas com o material já coletado durante a cirurgia e que foi armazenado no banco de tecidos, conforme consentimento do próprio paciente.

Duração da Pesquisa

O período de inclusão dos pacientes no estudo será de dois anos. Serão necessários também 2 anos para processar as amostras e avaliar os dados. Portanto, a duração do estudo esta prevista para 4 anos.

Possíveis riscos e desconfortos

Não há riscos para os pacientes.



HOSPITAL ALEMÃO
OSWALDO CRUZ

Benefícios:

Você entende que a sua participação neste estudo pode não trazer benefícios diretos a você. Entre os benefícios indiretos está o possível avanço do conhecimento científico, tornando possível aos cientistas descobrir tratamentos mais eficazes e mais seguros para indivíduos com condições iguais às suas.

Procedimentos alternativos que possam ser vantajosos para o indivíduo

Não há alternativa que possa ser vantajosa para o indivíduo.

Confidencialidade

Todos os dados obtidos no decorrer deste protocolo de pesquisa são confidenciais, o sigilo destes e a privacidade de todos os pacientes serão mantidos. Os pacientes serão registrados pelas iniciais e número de registro. Se os resultados do ensaio forem publicados, sua identidade permanecerá confidencial, conforme exigido pelas leis e normas aplicáveis.

Participação voluntária

Sua decisão em participar deste estudo é totalmente voluntária e não trará prejuízos nem a você nem ao seu tratamento. Não há penalidades nem perdas de benefícios (a que teria direito de outra forma) se optar por não participar. Você terá liberdade de retirar seu consentimento a qualquer momento e de deixar de participar do estudo, sem que isto traga prejuízo à continuidade da assistência.

Acesso à informação adicional

Você poderá ter acesso a qualquer tempo, diretamente com o médico do estudo, às informações sobre procedimentos, riscos e benefícios relacionados à pesquisa, inclusive para esclarecer eventuais dúvidas.

Pagamento pela participação

Você não receberá qualquer tipo de compensação financeira decorrente da sua participação no estudo.

Indenização por eventual dano causado à saúde decorrentes da pesquisa

Não haverá indenização de forma alguma por eventuais danos à saúde decorrentes da pesquisa dado que esta não representa risco algum aos pacientes.

Dúvidas sobre a pesquisa

Em caso de dúvidas, entre em contato com o Dr. Raphael Bessa Parmigiani ou Dr Rodrigo Oliva Perez. Não tenha pressa para entender este informativo escrito e as explicações e informações orais dadas pelo médico que assina este formulário. Você também pode falar com seu médico particular.

XX



HOSPITAL ALEMÃO
OSWALDO CRUZ

CONSENTIMENTO PÓS-ESCLARECIDO DO PACIENTE

- Eu li ou foi lido para mim em uma língua que eu entendi este Termo de Consentimento e as Informações ao Paciente.
- Eu tive oportunidade para questionar e estou satisfeito com as respostas que eu recebi.
- Eu livremente concordei em doar material coletado durante minha cirurgia.
- Eu entendi que esta pesquisa pode levar muitos anos para ser concluída e qualquer informação obtida não beneficiará a mim ou minha família pessoalmente ou financeiramente.
- Declaro que, após convenientemente esclarecido pelo pesquisador e ter entendido o que me foi explicado, consinto em participar do presente Protocolo de Pesquisa.

Nome do Participante: _____ Data: / /

Assinatura: _____

Nº Registro: _____

Nome da testemunha: _____ Data: / /

DECLARAÇÃO DA PESSOA PROVEDORA DA INFORMAÇÃO:

Eu dei uma explicação verbal dos objetivos do Banco de Tecido, seus possíveis procedimentos e riscos e acredito que o participante entendeu minhas explicações.

Nome e

Título: _____

Assinatura: _____ Data / /

(Nota: Todas as partes que assinam o Termo de Consentimento devem datar sua própria assinatura)
