

Universidade Federal do Rio Grande do Sul
Instituto de Biociências
Programa de Pós-Graduação em Genética e Biologia Molecular

**CARACTERIZAÇÃO DAS SEQUÊNCIAS REGULATÓRIAS DE
POLIADENILAÇÃO E MODULAÇÃO DA EXPRESSÃO POR MICRORNAS
EM UM CONJUNTO ABRANGENTE DE GENES DE PREDISPOSIÇÃO AO
CÂNCER**

Igor Araujo Vieira

Dissertação submetida ao Programa de Pós-Graduação em Genética e Biologia Molecular da Universidade Federal do Rio Grande do Sul como requisito parcial para a obtenção do grau de Mestre em Genética e Biologia Molecular.

Orientadora: Profa. Dra. Patricia Ashton Prolla

Porto Alegre, Abril de 2017.

Este trabalho foi desenvolvido no Laboratório de Medicina Genômica do Centro de Pesquisa Experimental do Hospital de Clínicas de Porto Alegre (HCPA) em parceria com a bioinformata Profa. Dra. Mariana Recamonde Mendoza, pertencente ao Instituto de Informática da Universidade Federal do Rio Grande do Sul (UFRGS), e com o grupo de pesquisa coordenado pelo Prof. Dr. Sandro José de Souza, pertencente ao Instituto Internacional de Física da Universidade Federal do Rio Grande do Norte (UFRN). O estudo foi apoiado pelo Programa de Pós-Graduação em Genética e Biologia Molecular (PPGBM-UFRGS) e pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) através da concessão da bolsa de mestrado e bolsa de produtividade em pesquisa.

DEDICATÓRIA

A todos os pacientes que lutam contra o câncer, bem como suas famílias, e aos profissionais da área da saúde que contribuem para a pesquisa e atendimento clínico em Oncologia.

AGRADECIMENTOS

À Professora Patricia Ashton-Prolla, agradeço pela confiança e orientação conduzida sempre de maneira respeitosa e agradável. Você é um exemplo de pesquisadora e pessoa no qual me espelho. Obrigado pela grande compreensão nos momentos difíceis da minha vida pessoal.

Aos colaboradores que participaram da execução do projeto: Mariana Mendoza, Isabel, Delva, Marina e Clévia. Agradeço pela amizade, pelo auxílio fundamental nas análises de Bioinformática e pela incansável ajuda na revisão desse trabalho. A todos os colegas e amigos do Laboratório de Medicina Genômica (LMG): Pati Silva, Liliane Todeschini, Gustavo, Ivaine, Tiago, Larissa, Fernanda Vianna, Cleandra, Bárbara, Camila e Gabriel.

Agradeço aos grandes “amigos” que a vida me deu: Ronize, Ethi, Marcus, Su Dalmolin, Su Basgalupp e Raíssa. De modo especial, agradeço a minha amiga-irmã Ronize pelo apoio incondicional durante o período do mestrado. Tua amizade foi muito importante para que eu pudesse finalizar esse ciclo com tranquilidade e positividade. A vocês, deixo registrada minha gratidão pela força, pelas inúmeras horas de terapia gratuita e, principalmente, por me fazerem descobrir a amizade verdadeira.

Aos funcionários do Centro de Pesquisa Experimental do HCPA, especialmente o Everaldo Almeida pelo suporte e atenção. Ao Elmo, pela ajuda em todas as questões burocráticas e pela paciência e atenção de sempre.

Por fim, agradeço ao suporte, carinho e amor que a minha família me forneceu durante os vários anos de estudo que antecederam o meu mestrado: mãe Lia e pai Rubinho, que acreditaram na minha capacidade e apoiaram minhas escolhas profissionais; e ao meu irmão falecido William, que certamente só me transmite energias positivas da “morada de paz” na qual se encontra. “Mano”, eu dedico essa conquista a ti!

SUMÁRIO

CAPÍTULO I – INTRODUÇÃO	13
I.1. Genes de predisposição ao câncer	14
I.2. Síndromes de predisposição ao câncer.....	15
I.3. Poliadenilação e sequências regulatórias.....	18
I.4. Sequências regulatórias de poliadenilação em genes humanos e variantes patogênicas nessas sequências	21
I.5. microRNAs (miRNAs)	26
I.5.1. miRNAs em câncer	27
I.5.2. Predição de sítios-alvo para ligação de miRNAs e bancos de dados com validação experimental.....	29
I.6. Poliadenilação alternativa e regulação diferencial da expressão por miRNAs.....	30
CAPÍTULO II – JUSTIFICATIVA	33
CAPÍTULO III – OBJETIVOS	35
CAPÍTULO IV – MANUSCRITO	37
<i>A comprehensive analysis of core polyadenylation sequences and regulation by microRNAs in a set of cancer predisposition genes</i>	37
CAPÍTULO V– DISCUSSÃO	80
CAPÍTULO VI – CONCLUSÕES	86
CAPÍTULO VII – PERSPECTIVAS	90
REFERÊNCIAS BIBLIOGRÁFICAS	93

LISTA DE ABREVIATURAS

- APC** - *Adenomatous Polyposis Coli Gene*
- BRCA1** - *Breast Cancer Susceptibility Gene 1*
- BRCA2** - *Breast Cancer Susceptibility Gene 2*
- DNA** - Ácido Desoxirribonucleico (*Deoxyribonucleic Acid*)
- FAP** - *Familial Adenomatous Polyposis*
- GPC** - Genes de Predisposição ao Câncer
- HBCC** - *Hereditary Breast and Colon Cancer*
- HBOC** - *Hereditary Breast and Ovarian Cancer*
- HNPCC** - *Hereditary Nonpolyposis Colorectal Cancer*
- IPEX** - *Immune dysregulation, Polyendocrinopathy, Enteropathy, X-linked*
- KRAS** - *v-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog*
- LFL** - Síndrome de Li-Fraumeni-like (*Li-Fraumeni-like Syndrome*)
- miRNA** - **microRNA**
- miRISC** - Complexo de Silenciamento Induzido por RNA (**miRNA-Induced Silencing Complex**)
- mRNA** - RNA mensageiro (*messenger RNA*)
- PAP** – Poli-A Polimerase
- PTEN** - *Phosphatase and tensin homolog gene*
- RNA** - Ácido Ribonucleico (*Ribonucleic acid*)
- SC** - Sítio de Clivagem (ou sítio de poliadenilação)
- SLF** - Síndrome de Li-Fraumeni (*Li-Fraumeni Syndrome*)
- SNP** - Polimorfismo de nucleotídeo único (*Single Nucleotide Polymorphism*)
- SP** - Sinal de Poliadenilação
- SPHC** - Síndromes de Predisposição Hereditária ao Câncer
- SRP** - Sequências Regulatórias de Poliadenilação
- TP53** - *Tumor Protein p53 Gene*
- 3'UTR** - Região 3' não-traduzida (**3' Untranslated Region**)

LISTA DE FIGURAS

- Figura 1.** Processamento da extremidade 3' do pré-mRNA em eucariotos e as sequências regulatórias envolvidas20
- Figura 2.** Comparação entre as sequências dos 13 sinais de poliadenilação funcionais descritos em genes humanos até o momento21
- Figura 3.** microRNAs como supressores de tumor ou oncogenes28
- Figura 4.** A poliadenilação alternativa de um transcrito primário pode interferir na regulação da sua expressão por miRNAs31

LISTA DE TABELAS

Tabela 1. Principais síndromes de câncer hereditário e seus genes associados	17
Tabela 2. Variantes funcionais em sequências regulatórias de poliadenilação associadas com patologias em humanos	24

RESUMO

A poliadenilação consiste na etapa de processamento da extremidade 3' dos pré-mRNAs em eucariotos. Duas sequências regulatórias localizadas na região 3' não-traduzida (3'UTR) desses transcritos primários apresentam um papel fundamental nesse processo: o sinal de poliadenilação (SP) AAUAAA e suas variantes funcionais; e o sítio de clivagem (SC), preferencialmente um dinucleotídeo "CA". Outros elementos de sequência também situados na região 3'UTR são os sítios de ligação para microRNAs (miRNAs), os quais representam uma classe de pequenos RNAs não-codificantes que regulam negativamente a expressão gênica e podem atuar como supressores de tumor e/ou oncogenes. Os principais objetivos desse estudo foram: (1) caracterizar os SP e SC em um conjunto de genes de predisposição ao câncer (GPC); (2) avaliar a frequência de poliadenilação alternativa entre os GPC estudados; e (3) identificar miRNAs e/ou famílias de miRNAs potencialmente oncogênicos e supressores de tumor considerando o mesmo grupo de genes. Para caracterizar essas sequências nos GPC selecionados (n=117) foram utilizados os bancos de dados do NCBI (referência) e os bancos de dados de poliadenilação alternativa APADB e APASdb. A busca por miRNAs reguladores de genes supressores de tumor (n=81) e oncogenes (n=17) foi realizada utilizando fontes de dados preditos computacionalmente e com validação experimental. Em relação ao objetivo (1), 21 dos 117 GPC (~18%) não apresentaram SP indicados no NCBI, sendo que um método computacional foi empregado para detectar SP putativos nesses genes. A maioria dos SP descritos para GPC nesse banco (74,4%) possuía a sequência canônica AAUAAA, enquanto 24,1% continha a variante funcional AUUAAA. Curiosamente, o dinucleotídeo "AA" representou o SC mais frequente dentre os GPC analisados. Além disso, análises relacionadas ao objetivo (2) identificaram evidências de poliadenilação alternativa (mais de um SP funcional) em 105 GPC (~90%), em comparação com uma estimativa anterior de que isso ocorreria em cerca de 50% dos genes humanos, sugerindo uma maior complexidade na regulação da poliadenilação nesses transcritos. Quanto ao objetivo (3), a família de miRNA miR-192 foi significativamente super-representada entre os genes supressores de tumor ($P < 0,01$), o que sugere

uma função potencialmente oncogênica. Em contraste, um número maior de famílias de miRNAs foi fortemente associado com a regulação de oncogenes: miR-128, miR-1471, miR-483, miR-3170 e miR-218. Nossos dados fornecem um mapeamento das sequências regulatórias de poliadenilação (SRP) em GPC, o qual pode ser utilizado no desenvolvimento de análises moleculares incluindo esses elementos da região 3'UTR frequentemente negligenciados na rotina de diagnóstico molecular. Ademais, estudos funcionais devem ser realizados para validar as funções potencialmente atribuídas às famílias de miRNAs aqui mencionadas. Em suma, esse é o primeiro estudo focado nesse tipo de caracterização abrangente (SRP e miRNAs) em GPC.

ABSTRACT

Polyadenylation is the processing step of 3' end in eukaryotic pre-mRNAs. Two regulatory sequences located in the 3' untranslated region (3'UTR) of these primary transcripts play an essential role in this process: the polyadenylation signal (PAS) AAUAAA and its functional variants; and the cleavage site (CS), preferentially a "CA" dinucleotide. Other 3'UTR sequence elements are the binding sites for microRNAs (miRNAs), which represent a class of small non-coding RNAs that negatively regulate gene expression, and can act as tumor suppressors and/or oncogenes. The main goals of this study were: (1) to characterize PAS and CS in a set of cancer predisposition genes (CPGs); (2) assess the frequency of alternative polyadenylation among the studied genes; and (3) identify potentially oncogenic and tumor-suppressive miRNAs or miRNA families considering the same group of genes. In order to characterize these sequences in the selected CPGs (n=117) the NCBI database (reference source) and the alternative polyadenylation databases APADB and APASdb were used. The search for miRNAs that regulate tumor suppressor genes (n=81) and oncogenes (n=17) was performed using computationally predicted and experimentally validated data sources. In relation to the first aim, 21 of the 117 CPGs (~18%) did not exhibit a PAS sequence indicated in NCBI, and a computational method was employed to detect putative PAS in these genes. In this database, most described PAS (74.4%) for CPGs had the canonical sequence AAUAAA, while 24.1% contained the functional variant PAS AUUAAA. Interestingly, "AA" dinucleotide was the most frequent CS associated with the CPGs. In addition, second aim related-analyses identified alternative polyadenylation evidences (more than one functional PAS) for 105 CPGs (~90%), while a previous estimate indicated that it occurs in about 50% of human genes, suggesting a greater complexity in the regulation of polyadenylation in these transcripts. Regarding the third aim, miR-192 family was significantly overrepresented among tumor suppressor genes ($P < 0.01$), which suggests a potentially oncogenic function. In contrast, a larger number of miRNA families was strongly associated with regulation of oncogenes: miR-128, miR-1471, miR-483, miR-3170 and miR-218. Our data provide a mapping of polyadenylation regulatory sequences (PRS) in CPGs, which can be used in the

development of molecular analyses including these 3'UTR elements often neglected during the routine molecular diagnosis. Furthermore, functional studies should be performed to validate the potential functions assigned to the miRNA families mentioned herein. In summary, this is the first study focused on this type of comprehensive characterization (PRS and miRNAs) in CPGs.

CAPÍTULO I – INTRODUÇÃO

I.1. Genes de predisposição ao câncer

Nas últimas décadas, avanços significativos foram feitos no conhecimento dos mecanismos moleculares que originam o câncer. Diversos genes foram identificados e agrupados em duas classes mais abrangentes segundo o risco de o câncer decorrer de uma atividade aumentada ou diminuída do produto do gene (de la Chapelle e Peltomäki, 1998). Os genes nos quais uma mutação que cause aumento de função contribui para o processo de carcinogênese são denominados proto-oncogenes, e os seus mutantes, as formas hiperativas, são chamados de oncogenes. Os proto-oncogenes são responsáveis por estimular a proliferação celular em condições normais e podem ser convertidos de um gene celular normal a um oncogene através de mutações pontuais, pequenas deleções/inserções, translocações cromossômicas e amplificação gênica. Estas alterações em uma única cópia de um proto-oncogene têm um efeito dominante em promover a proliferação celular (Osborne *et al.*, 2004; Croce, 2008; Weinstein e Joe, 2008).

Em contrapartida, genes nos quais as mutações que levam à perda de função contribuem para o desenvolvimento de neoplasias são denominados genes supressores de tumor. Estes genes são geralmente reguladores negativos da proliferação celular ou de outras funções que afetam o potencial invasivo e metastático das células tumorais, sendo que a sua perda de função geralmente decorre da inativação de ambas as cópias do gene normal na célula diplóide somática, através de mutações deletérias ou outros mecanismos que interfiram na sua expressão ou função (efeito recessivo) (Osborne *et al.*, 2004). Essa “hipótese de dois eventos” foi inicialmente postulada por Alfred Knudson em referência ao retinoblastoma, no qual ambas as cópias do gene RB1 selvagem devem ser removidas ou inativadas para desencadear o fenótipo maligno (Knudson, 1971).

Uma terceira classe incluiria os genes de reparo do DNA, os quais estão envolvidos nos processos constitutivos de reparo a danos no DNA, sejam eles alterações espontâneas ou induzidas por fatores ambientais. A inativação desses genes leva ao acúmulo de mutações somáticas e pode resultar no favorecimento de uma instabilidade genômica (Murga e Fernández-Capetillo,

2007; Bauer *et al.*, 2015). Frequentemente o câncer aparece quando o ambiente adverso contribui para esta instabilidade por um período suficientemente longo para permitir o acúmulo de um número de crítico de alterações genéticas (Hanahan e Weinberg, 2011).

I.2. Síndromes de predisposição ao câncer

Embora a maioria das neoplasias seja resultado de interações complexas entre o componente genético do indivíduo e os fatores ambientais, o que denominamos de herança multifatorial, um pequeno percentual de casos decorre principalmente de alterações herdadas que conferem uma maior predisposição ao desenvolvimento de tumores (Kinzler e Vogelstein, 1998). Estima-se que cerca de 5 a 10% de todos os tumores estejam associados à predisposição hereditária e diretamente associados a mutações germinativas em genes de alta penetrância para o câncer (de la Chapelle e Peltomäki, 1998; Rocha, 2010). Quando o câncer ocorre como parte de uma síndrome de predisposição hereditária ao câncer (SPHC), a mutação inicial causadora da doença é herdada através da linhagem germinativa e, portanto, já está presente em cada célula do corpo, desde a formação do embrião (Garber e Offit, 2005).

Os avanços na biologia molecular do câncer propiciaram a identificação de genes associados a determinadas SPHC, mencionados na **Tabela 1**. Dentre estes, destacam-se os genes supressores de tumor *BRCA1* e *BRCA2* para câncer de mama e ovário hereditários (Miki *et al.*, 1994; Wooster *et al.*, 1994); o gene *TP53* para a Síndrome de Li-Fraumeni (SLF) (Malkin *et al.*, 1990); o gene *APC* para a polipose adenomatosa familiar (Grodin *et al.*, 1991); os genes do sistema de reparo de pareamento incorreto do DNA (sistema MMR, *Mismatch Repair*) *MLH1*, *MSH2*, *MSH6*, *PMS1* e *PMS2* para câncer colorretal hereditário não-polipomatoso ou Síndrome de Lynch (Leach *et al.*, 1993; Liu *et al.*, 1996), entre outros. Estas descobertas deram origem a testes moleculares de diagnóstico de predisposição hereditária em diferentes síndromes (Ponder, 1997; Sidransky, 1997) e estimularam o desenvolvimento de programas de

avaliação clínica e aconselhamento genético de famílias em risco (Biesecker e Garber, 1995). A **Tabela 1** aborda os principais genes supressores de tumor, oncogenes e genes de reparo do DNA envolvidos com SPHC e o espectro de tumores característico de cada síndrome. Todos esses genes foram incluídos nas análises do presente estudo, além de outros genes associados com suscetibilidade a determinado(s) tipo(s) de tumor(es) e genes associados com síndromes gênicas nas quais um dos fenótipos é a predisposição hereditária ao câncer.

Tabela 1. Principais síndromes de câncer hereditário e seus genes associados.

Síndromes*	Principais neoplasias associadas	Genes	Função
Síndrome de Câncer de Mama e Ovário Hereditários (HBOC)	Câncer de mama, ovário, próstata e pâncreas.	<i>BRCA1</i> e <i>BRCA2</i>	Supressor tumoral/ Reparo de DNA.
Síndrome de Câncer de Mama e Cólon Hereditários (HBOC)	Câncer de mama e colorretal.	<i>CHEK2</i>	Supressor tumoral
Síndrome de Lynch ou Síndrome de Câncer Colorretal Hereditário não-Polipomatoso (HNPCC)	Câncer colorretal, endométrio, ovário, pelve renal e ureter, gástrico e de intestino delgado.	<i>MLH1</i> , <i>MSH2</i> , <i>MSH6</i> , <i>PMS1</i> e <i>PMS2</i>	Reparo de DNA
Polipose Adenomatosa Familiar (FAP)	Câncer colorretal, câncer gástrico e tumores desmoides.	<i>APC</i>	Supressor tumoral
Síndrome de Li-Fraumeni	Sarcoma de tecidos moles, osteossarcomas, câncer de mama, tumores do sistema nervoso central, leucemia e tumores adrenocorticais.	<i>TP53</i>	Supressor tumoral
Síndrome de Cowden e Bannayan-Riley-Ruvalcaba	Câncer de mama, tireoide, endométrio, renal e meningioma.	<i>PTEN</i>	Supressor tumoral
Neoplasia Endócrina Múltipla tipo 1	Tumores de ilhotas pancreáticas, adenomas de hipófise e paratireoide.	<i>MEN1</i>	Supressor tumoral
Neoplasia Endócrina Múltipla tipo 2	Carcinoma medular de tireoide, feocromocitoma e hiperplasia de tireoide.	<i>RET</i>	Oncogene
Neurofibromatose-1	Neurofibroma, neurofibrossarcoma, glioma óptico, feocromocitoma e meningioma.	<i>NF1</i>	Supressor tumoral
Neurofibromatose-2	Schwannoma vestibular bilateral e de outros nervos cranianos e periféricos, meningiomas e ependimomas.	<i>NF2</i>	Supressor tumoral
Esclerose Tuberosa	Angiofibromas faciais, fibroma subungual, rabdomiomas cardíacos, angiomiolipomas e cistos renais.	<i>TSC1</i> e <i>TSC2</i>	Supressor tumoral
Doença de Von Hippel-Lindau	Hemangioblastomas de retina e sistema nervoso central, câncer renal e feocromocitoma.	<i>VHL</i>	Supressor tumoral
Retinoblastoma Hereditário	Retinoblastoma e osteosarcoma	<i>RB1</i>	Supressor tumoral
Melanoma Familiar	Melanoma maligno e câncer de pâncreas	<i>CDKN2</i> <i>CDK4</i>	Supressor tumoral Oncogene
Tumor de Wilms Hereditário	Tumor de Wilms (tumor renal maligno infantil)	<i>WT1</i>	Supressor tumoral
Síndrome da Polipose Juvenil	Câncer colorretal e outros tumores do trato gastrointestinal	<i>SMAD4</i> e <i>BMPR1A</i>	Supressores de tumor
Câncer Gástrico Difuso Hereditário	Câncer gástrico	<i>CDH1</i>	Supressor tumoral
Síndrome de Gorlin ou Síndrome do Nevo/Carcinoma Basocelular	Carcinoma basocelular e tumores cerebrais (meduloblastoma)	<i>PTCH1</i>	Supressor tumoral

*O padrão de herança de todas as síndromes mencionadas é autossômico dominante. HBOC, *Hereditary Breast and Ovarian Cancer*; HBOC, *Hereditary Breast and Colon Cancer*; HNPCC, *Hereditary Nonpolyposis Colorectal Cancer*; FAP, *Familial Adenomatous Polyposis*. Modificado de Garber e Offit, 2005.

I.3. Poliadenilação e sequências regulatórias

A poliadenilação consiste em uma etapa nuclear de processamento da extremidade 3' dos transcritos primários (pré-mRNAs), sendo dividida em duas etapas: a clivagem da extremidade 3' do pré-mRNA e a subsequente adição da cauda poli-A. Ambas etapas são cruciais para uma terminação apropriada dos transcritos em eucariotos, bem como para sua estabilidade, localização e transporte ao citoplasma, a fim de que a cadeia polipeptídica seja sintetizada nos ribossomos (Sachs, 1990; Proudfoot, 2011). Evidências experimentais indicaram que a sequência sinal de poliadenilação (SP) AAUAAA, geralmente situada na região 3' não-traduzida (3'UTR) do pré-mRNA, apresenta um papel fundamental nesse processo (Zarkower *et al.*, 1986; Zhao *et al.*, 1999; Ryan *et al.*, 2004).

O hexanucleotídeo/hexâmero AAUAAA foi a primeira sequência regulatória de mRNA eucariótico descoberta, mostrando-se altamente conservada nesse grupo de organismos (**Figura 1A**) (Proudfoot e Brownlee, 1976; Proudfoot, 1991). O desenvolvimento de sistemas *in vitro* para avaliar o processo de poliadenilação em mamíferos mostrou a importância do SP AAUAAA na terminação 3' apropriada do pré-mRNA, embora outras sequências contendo mutações pontuais ou polimorfismos de nucleotídeo único (SNPs) nesse hexâmero canônico permitam a poliadenilação em níveis menores, mas mensuráveis *in vivo* e *in vitro* (Moore e Sharp, 1984, 1985; Wickens e Stephenson, 1984; Manley *et al.*, 1985; Conway e Wickens, 1987; Sheets *et al.*, 1990)

Conforme indica a **Figura 1B**, o SP é reconhecido por um grande complexo proteico associado ao domínio carboxi-terminal da RNA polimerase II, o qual cliva a fita simples do pré-mRNA em um sítio localizado 10-30 nucleotídeos a jusante dessa sequência, denominado sítio de clivagem (SC) ou sítio poli-A. A sequência desse SC é descrita como sendo preferencialmente um dinucleotídeo CA (Danckwardt *et al.*, 2008). Após essa clivagem específica, ocorre a adição da cauda poli-A pela enzima poli-A polimerase (PAP), uma sequência com cerca de duas centenas de adenilatos não transcrita pela RNA

polimerase II e que aumenta a estabilidade e tradução do mRNA, já que previne a sua degradação por 3'-5' exonucleases (Zhao *et al.*, 1999; Tian *et al.*, 2005). Uma região rica em U/GU, pouco conservada e possivelmente redundante, situada 14-70 nucleotídeos a jusante de AAUAAA, também estaria envolvida na poliadenilação (Sadofsky e Alwine, 1984; Ryan *et al.*, 2004; Lemay *et al.*, 2010). Essa região e o motivo AAUAAA, juntamente com o SC, constituem os chamados elementos centrais da poliadenilação (Lutz, 2008; Hollerer *et al.*, 2014), destacados na **Figura 1C**. Mais especificamente, estudos demonstraram que a eficiência de um sítio de clivagem no pré-mRNA é definida pela estabilidade de sua interação com o complexo proteico de clivagem, particularmente do fator CPSF (*Cleavage and Polyadenylation Specificity Factor*) com o motivo AAUAAA, bem como do fator CstF (*Cleavage Stimulation Factor*) com a sequência rica em U/GU a jusante desse motivo (Wahle e Rügsegger, 1999; Danckwardt *et al.*, 2008).

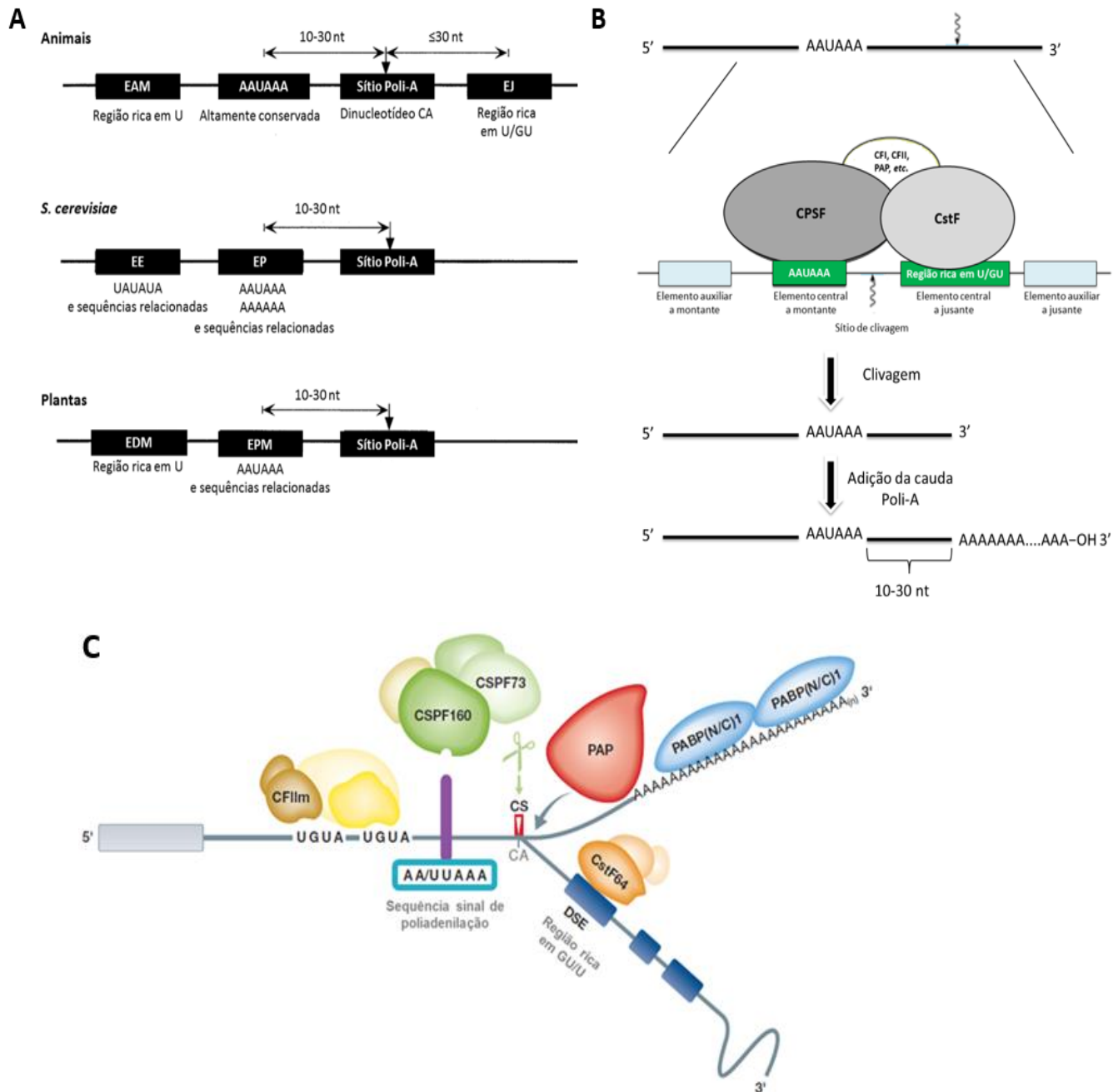


Figura 1. Processamento da extremidade 3' do pré-mRNA em eucariotos e as seqüências regulatórias envolvidas. (A) Comparação das seqüências regulatórias desse processo entre diferentes eucariotos, destacando a alta conservação do sinal de poliadenilação AAUAAA. nt, nucleotídeos; proteínas que fazem parte da maquinaria central de clivagem e poliadenilação em mamíferos: CPSF, *Cleavage and Polyadenylation Specificity Factor*; CstF, *Cleavage Stimulation Factor*; CFI e CFII, *Cleavage Factors* (fatores de clivagem I e II); PAP, *Poly(A) Polymerase* (enzima poli-A polimerase). (B) Seqüência de etapas durante a poliadenilação. EAM, *enhancer* auxiliar a montante; EJ, elemento a jusante; EE, elemento de eficiência; EP, elemento de posicionamento; EDM, elemento distante a montante; EPM, elemento próximo a montante. (C) Reconhecimento das seqüências regulatórias pelo complexo proteico de clivagem do pré-mRNA. CFIIm, fator de clivagem II; CSPF= *Cleavage and Polyadenylation Specificity Factor*; CS, *Cleavage Site* (sítio de clivagem); PABP, *Poly(A)-Binding Protein*; DSE, *Downstream Element* (elemento a jusante). **Modificado de Zhao et al., 1999 (A); Lutz, 2008 (B); e Hollerer et al., 2014 (C).**

I.4. Sequências regulatórias de poliadenilação em genes humanos e variantes patogênicas nessas sequências

Diversos estudos sugerem que a maioria dos SP em transcritos derivados de genes humanos contém o hexâmero AAUAAA (cerca de 60%), porém uma parcela menor apresenta variabilidade nessa sequência (Graber *et al.*, 1999; Beaudoin *et al.*, 2000; MacDonald e Redondo, 2002). Além da sequência canônica AAUAAA, doze SP variantes são conhecidos em genes/transcritos humanos até o momento (Tian *et al.*, 2005; Retelska *et al.*, 2006), os quais estão citados na **Figura 2**.

A	A	U	A	A	A
A	U	U	A	A	A
A	G	U	A	A	A
U	A	U	A	A	A
C	A	U	A	A	A
G	A	U	A	A	A
A	A	U	A	U	A
A	A	U	A	C	A
A	A	U	A	G	A
A	C	U	A	A	A
A	A	G	A	A	A
A	A	U	G	A	A
U	U	U	A	A	A
N	N	U	A	N	A

Figura 2. Comparação entre as sequências dos 13 sinais de poliadenilação funcionais descritos em genes humanos até o momento. Em humanos, as posições 3, 4 e 6 desse hexanucleotídeo mostram-se altamente conservadas (destacadas em negrito), enquanto as posições 1, 2 e 5 parecem ser tolerantes a mutações de ponto (N= qualquer nucleotídeo). Modificado de Beaudoin *et al.*, 2000 e Tian *et al.*, 2005.

Conforme mostra a **Tabela 2**, variantes germinativas patogênicas no SP em genes humanos primeiramente foram identificadas no sinal de poliadenilação do gene *HBA2* (AAUAAA>AAUAAG), que codifica a alfa-2-globina, causando a α -talassemia (Higgs *et al.*, 1983). Posteriormente, foram descritas alterações patogênicas no SP em outros genes, resultando em várias patologias humanas conhecidas (Chen *et al.*, 2006; Michalova *et al.*, 2013), incluindo a β -talassemia (Orkin *et al.*, 1985; Rund *et al.*, 1992; Giordano *et al.*,

2005), doença de Fabry (Brady *et al.*, 1967; Yasuda *et al.*, 2003), imunodeficiência combinada severa ligada ao X (Bonapace *et al.*, 2003), síndrome IPEX (*Immune dysregulation, polyendocrinopathy, enteropathy, X-linked*) (Bennett *et al.*, 2001), entre outras. De modo geral, sabe-se que a presença de alterações na sequência do SP AAUAAA (SNPs, mutações pontuais, inserções/deleções) pode causar instabilidade na sua interação com o complexo proteico de clivagem, ocasionando uma abolição/diminuição das taxas de poliadenilação dos transcritos contendo o alelo variante/mutante e, por sua vez, degradação mais rápida destes (Michalova *et al.*, 2013). Em última análise, o impacto funcional dessas variantes gênicas geralmente consiste na redução dos níveis de mRNA e de expressão proteica (Chen *et al.*, 2006; Danckwardt *et al.*, 2008).

Em relação ao SC, também denominado sítio de poliadenilação, tem sido descrito que a maioria dos mRNAs em eucariotos são preferencialmente clivados imediatamente 3' de um dinucleotídeo CA, localizado 10-30 nucleotídeos a jusante do SP (Danckwardt *et al.*, 2008). Ensaios funcionais mostraram que esse dinucleotídeo CA apresenta maior reconhecimento pela maquinaria de clivagem e, portanto, seria fisiologicamente mais eficiente (Chen *et al.*, 1995; Gehring *et al.*, 2001; Danckwardt *et al.*, 2004). Além disso, o modelo proposto por Chen e colaboradores (1995) indica que a posição exata de clivagem na extremidade 3' do pré-mRNA é determinada pela sequência nucleotídica local na ordem de preferência A > U > C >> G.

Variantes clinicamente relevantes na sequência do SC já foram identificadas no gene *F2*, que codifica o fator de coagulação protrombina, apresentando um efeito de ganho de função, isto é, aumento da eficiência de poliadenilação. Tais variantes ocasionam aumento na concentração plasmática de protrombina, resultando em um risco aumentado para desenvolver trombose (Gehring *et al.*, 2001; Danckwardt *et al.*, 2004). O SNP *F2* 20210*A promove a troca de sequência CG>CA no SC e representa uma causa comum de trombofilia (Poort *et al.*, 1996). Fisiologicamente, a clivagem do pré-mRNA *F2* ocorre na 3' de um dinucleotídeo CG, sendo essa a sequência menos eficiente em promover a reação de clivagem *in vitro* (Chen *et al.*, 1995). Assim, foi observado que a variante *F2* 20210*A aumentava a eficiência de poliadenilação do mRNA *F2* (Gehring *et al.*, 2001; Danckwardt *et al.*, 2004; Kuwahara *et al.*,

2004), sugerindo um novo mecanismo molecular que causa expressão gênica patológica. Subsequentemente, duas mutações raras com o mesmo efeito de ganho de função foram descritas: a mutação *F2* 20209*T, uma troca CG>TG no SC (Danckwardt *et al.*, 2006); e a mutação *F2* 20221*T, que introduz um resíduo adicional de uridina (U) dentro do sítio de ligação putativo para o fator CstF (Balim *et al.*, 2003). Considerando que a região rica em U/GU, um dos elementos em *cis* que regulam a poliadenilação, exibe uma baixa densidade de resíduos de U no pré-mRNA *F2* quando comparada com a maioria dos mRNAs em eucariotos, a inserção de resíduos adicionais de U aumentaria a eficiência de poliadenilação, possivelmente por facilitar a interação de CstF com a extremidade 3' desse transcrito primário (Danckwardt *et al.*, 2004).

Recentemente, variantes germinativas que afetam a poliadenilação de transcritos derivados de genes de predisposição ao câncer apresentaram evidências de associação com síndromes de câncer hereditário: o SNP rs78378222, localizado no SP constitutivo do gene supressor tumoral *TP53*, foi associado com o fenótipo da Síndrome de Li-Fraumeni-like (LFL) (Macedo *et al.*, 2016); e uma variante funcional (duplicação de 20pb) a montante do SP do gene *MSH6* foi detectada em dois pacientes que preenchiam critérios clínicos para a Síndrome de Lynch (Decorsière *et al.*, 2012) (**Tabela 2**). Em um estudo publicado pelo nosso grupo, a variante *TP53* rs78378222 (AAUAAA>AAUACA) não foi identificada em um grupo de pacientes SLF/LFL com mutação germinativa patogênica em *TP53*, mas foi detectada em uma alta frequência em pacientes com fenótipo clínico da síndrome e sem mutações patogênicas identificadas em regiões codificadoras do gene (Macedo *et al.*, 2016). Adicionalmente, foi observada uma redução na expressão da proteína p53 em células normais (fibroblastos) de um dos pacientes LFL portadores da variante em relação a um indivíduo controle (Macedo *et al.*, 2016). Anteriormente, esse SNP havia sido associado com diversos tipos de tumores, incluindo alguns frequentemente descritos em famílias SLF/LFL, verificando-se níveis menores de mRNA *TP53* em indivíduos heterozigotos, em comparação aos homozigotos selvagens (Stacey *et al.*, 2011).

Tabela 2. Variantes funcionais em sequências regulatórias de poliadenilação associadas com patologias em humanos.

Patologia	Gene	Varição no SP ou SC	Principais referências
α-talassemia	HBA2	AATAAA > AATAAG	Higgs <i>et al.</i> , 1983
		AATAAA > AATGAA	Yuregir <i>et al.</i> , 1992
		AATAAA > AATA	Harteveld <i>et al.</i> , 1994
		Deleção de 16 pb incluindo a primeira base do SP AATAAA	Tamary <i>et al.</i> , 1997
β-talassemia	HBB	AATAAA > AACAAA	Orkin <i>et al.</i> , 1985
		AATAAA > AATGAA, AATAAA > AATAGA	Jankovic <i>et al.</i> , 1990
		AATAAA > AATAAG, AATAAA > A	Rund <i>et al.</i> , 1992
		AATAAA > AAAA	Kimberland <i>et al.</i> , 1995
		AATAAA > GATAAA	Waye <i>et al.</i> , 2001
		AATAAA > AAAAAA	Jacquette <i>et al.</i> , 2004
Leucodistrofia Metacromática (pseudodeficiência de arilsulfatase A)	ARSA	AATAAA > AATATA	Giordano <i>et al.</i> , 2005
		AATAAC > AGTAAC	Gieselmann <i>et al.</i> , 1989
Deficiência do fator de crescimento insulina-like I	IL2RG	AATAAA > AATAAG	Hsu <i>et al.</i> , 2000
Modulação do risco ao câncer colorretal	NAT1	AATAAA > inserção AATAAA AAA (alelo NAT1*16)	de Léon <i>et al.</i> , 2000
Síndrome IPEX*	FOXP3	AATAAA > AATGAA	Bennett <i>et al.</i> , 2001
Trombofilia hereditária	F2	SNP G>A no SC (CG>CA)	Gehring <i>et al.</i> , 2001
		Inserção 20221*T; introduz um resíduo adicional de uridina (U) no sítio de ligação putativo para o fator CstF	Balim <i>et al.</i> , 2003
		Mutação C>T no SC (CG>TG)	Danckwardt <i>et al.</i> , 2006
Doença de Fabry	GLA	1277delAA, deleção da última base do SP AATAAA	Yasuda <i>et al.</i> , 2003
Imunodeficiência combinada severa ligada ao X	IGF1	AATATA > AAAATA	Bonapace <i>et al.</i> , 2003
Síndrome de Wiskott-Aldrich	WASP	Mutação complexa (pequena deleção de 9pb, inversão de 151pb e uma grande deleção de 4.3Kb) cria novo SP (AAUUAA) e resulta na produção de dois transcritos alternativos não-funcionais	Andreu <i>et al.</i> , 2006

Patologia	Gene	Varição no SP ou SC	Principais referências
Predisposição a certos tipos de tumores** e possível associação com o fenótipo da Síndrome de Li-Fraumeni-like	<i>TP53</i>	AATAAA > AATACA	Stacey <i>et al.</i> , 2011; Zhou <i>et al.</i> , 2012; Diskin <i>et al.</i> , 2014; Macedo <i>et al.</i> , 2016
Síndrome de Lynch ou Câncer Colorretal Hereditário não-Polipomatoso	<i>MSH6</i>	Duplicação de 20bp (TTCAGACAACATTATGATCT, denominada S20) a montante do SP AATAAA	Decorsière <i>et al.</i> , 2012

SP, sinal de poliadenilação; SC, sítio de clivagem; pb, pares de base. **Immune dysregulation, polyendocrinopathy, enteropathy, X-linked* (IPEX). **Carcinoma basocelular, glioma, câncer de próstata, adenoma colorretal, carcinoma esofágico de células escamosas, neuroblastoma. **Modificado de Chen *et al.*, 2006.**

I.5. microRNAs (miRNAs)

Os microRNAs (também chamados de miRNAs ou miRs) consistem em uma classe de pequenos RNAs não-codificantes de fita simples (18 a 25 nucleotídeos de tamanho) que, de modo geral, regulam negativamente a expressão gênica em todos os eucariotos multicelulares de uma maneira pós-transcricional e sequência-específica (Ambros, 2004; Bartel, 2004). Estima-se que os miRNAs regulam aproximadamente 30% de todos os transcritos gênicos (Rachagani *et al.*, 2010) e, atualmente, a principal base de dados de miRNAs contabiliza 2588 diferentes miRNAs maduros em humanos (miRBase versão 21.0; 2014).

Em geral, a transcrição dos genes codificadores de miRNAs é mediada pela RNA polimerase II. O transcrito primário, ou pri-miRNA, que possui 5' *cap* e cauda poli-A, é formado por algumas centenas a milhares de nucleotídeos. O pri-miRNA é clivado no núcleo pelo complexo enzimático formado pela RNase III *Drosha* e proteínas associadas do Microprocessador, produzindo um pre-miRNA de aproximadamente 70 nucleotídeos. Este pre-miRNA é posteriormente clivado no citoplasma pelo complexo enzimático RNase III *Dicer/Argonata*, a fim de gerar o miRNA maduro, o qual é então incorporado ao complexo de silenciamento induzido por RNA (RISC), que atua no reconhecimento da região 3'UTR dos mRNAs-alvo através de pareamento de bases imperfeito com o miRNA (Bartel, 2004; Friedman *et al.*, 2009). O complexo miRNA e RISC (miRISC) regula negativamente a expressão gênica através de dois mecanismos pós-transcricionais: a degradação do mRNA, quando a complementaridade entre as bases do mRNA e do miRNA for total (mais comum em plantas); ou a inibição da tradução do mRNA, quando essa complementaridade for parcial (Bartel, 2004; Esquela-Kerscher e Slack, 2006). Neste último caso, evidências sugerem que a repressão pode ocorrer durante o início da tradução ou mesmo após este evento. Exemplos de mecanismos de repressão inicial são a deadenilação da cauda poli-A (Beilharz *et al.*, 2009) e o bloqueio da associação entre as subunidades ribossomais 60S e 40S promovidos pelo complexo miRISC (Bartel, 2009; Fabian *et al.*, 2010). miRISC

pode competir com eIF4 e impedir sua ligação à extremidade 5' do mRNA, evitando assim a adesão da subunidade ribossomal 60S, ou impedindo o recrutamento da subunidade 40S (Mathonnet *et al.*, 2007). Os mecanismos posteriores ao início da tradução incluem: dissociação prematura dos ribossomos, diminuição da taxa ou parada no alongamento proteico e degradação do polipeptídeo nascente (Nilsen, 2007).

As descrições mais recentes dos mecanismos de ação dos miRNAs indicam que eles podem ser ativadores e repressores da expressão gênica ao se ligar a sequências reguladoras no DNA. A ligação complementar de miRNAs a regiões promotoras pode reprimir a transcrição através de uma impossibilidade física do alongamento do transcrito, bem como ativar a transcrição quando a sua ligação ocorre em regiões que recrutam modificadores de histonas, conferindo a abertura da cromatina (Portnoy *et al.*, 2011; Schwarzenbach *et al.*, 2014).

I.5.1. miRNAs em câncer

Cerca de 50% dos genes de miRNAs descritos em humanos estão localizados em regiões genômicas associadas ao câncer, pontos de quebra cromossômica e regiões de frequente perda de heterozigosidade, sugerindo que os miRNAs estão relacionados com a iniciação e progressão de diversos tipos de tumores (Calin *et al.*, 2004; Rachagani *et al.*, 2010).

Adicionalmente, sabe-se que os miRNAs estão envolvidos na regulação de genes com papel crucial em todas as etapas da tumorigênese, como proliferação celular, apoptose, angiogênese, invasão, metástase, entre outras (Ruan *et al.*, 2009). Análises de microarranjo indicaram que os perfis de expressão de miRNAs refletem a classificação das malignidades humanas, como o grau de diferenciação tumoral e o tipo de neoplasia (Lu *et al.*, 2005). Nesse contexto, os miRNAs podem atuar como supressores de tumor ou oncogenes, sendo estes últimos chamados de “*oncomirs*”. Conforme mostra a **Figura 3**, em tecidos normais, o funcionamento adequado da transcrição, processamento e ligação dos miRNAs às sequências complementares resulta

em taxas balanceadas de crescimento, proliferação, diferenciação e morte celular. Já em tecidos tumorais, duas alterações podem ser observadas em relação ao nível de expressão de miRNAs, ambas relacionadas com falhas em algum ponto da biogênese desses miRNAs: (a) a redução ou eliminação da expressão de um miRNA que atua como um supressor de tumor, processo este que favoreceria a formação de tumores devido à não-repressão da expressão de oncoproteínas; e (b) o aumento nas taxas de expressão de miRNAs que atuam como oncogenes (Esquela-Kerscher e Slack, 2006).

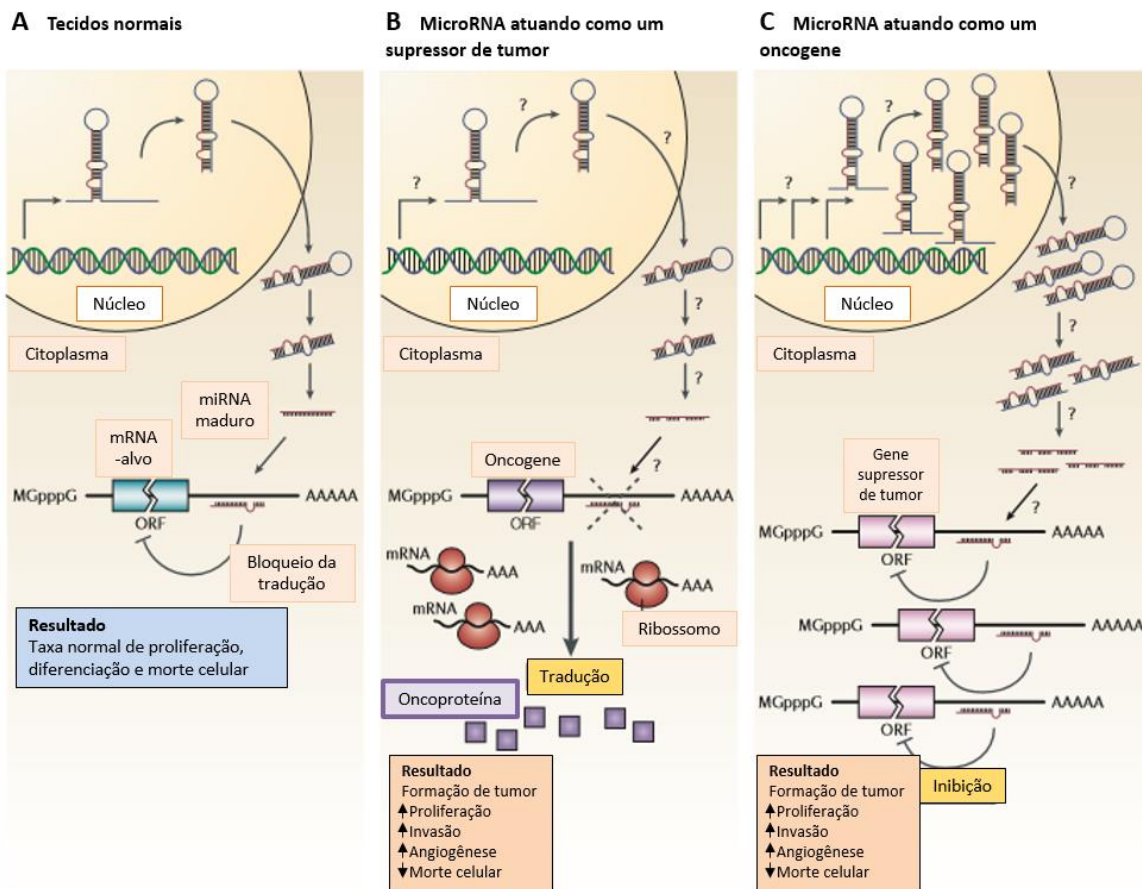


Figura 3. microRNAs podem atuar como supressores de tumor ou oncogenes. (A) Em tecidos normais, níveis de expressão adequados de miRNAs e sua ligação às sequências-alvo na 3'UTR dos mRNAs promove repressão da expressão gênica e resulta, de modo global, em taxas normais de proliferação e morte celular. (B) A diminuição da expressão de um miRNA supressor de tumor pode desencadear um aumento na expressão de oncoproteínas, favorecendo a formação de tumores. (C) A superexpressão de um miRNA oncogênico pode reduzir a expressão de genes supressores de tumor, também constituindo um mecanismo de carcinogênese. A modificação dos níveis de expressão, seja de miRNAs supressores de tumor ou oncogênicos, pode ser decorrente de falhas nas etapas da sua biogênese (indicadas na figura pelos pontos de interrogação). MGpppG, modificação da extremidade 5' do pré-mRNA em eucariotos, denominada 5'cap; ORF, *open reading frame* (fase aberta de leitura). **Modificado de Esquela-Kerscher e Slack, 2006.**

I.5.2. Predição de sítios-alvo para ligação de miRNAs e bancos de dados com validação experimental

Considerando que comumente os miRNAs ligam-se na região 3'UTR dos mRNAs através de complementaridade imperfeita de bases, a identificação de mRNAs-alvo tem sido cada vez mais complexa (Dweep *et al.*, 2013). No entanto, existe uma região correspondente aos nucleotídeos 2-8 da porção 5' do miRNA maduro, denominada sequência *seed*, na qual geralmente ocorre pareamento perfeito com a 3'UTR do mRNA-alvo, determinando em grande parte a especificidade do reconhecimento desses alvos e sendo crucial para a estabilidade e o carregamento da molécula de miRNA no complexo RISC (Friedman *et al.*, 2009). Por esse motivo, ferramentas de bioinformática para predição de alvos levam em consideração essencialmente essa sequência (Witkos *et al.*, 2011). Estudos resultantes dessas análises mostram que um único miRNA pode regular negativamente múltiplos alvos (até 200 mRNAs), com funções completamente distintas (Esquela-Kerscher e Slack, 2006; Ruan *et al.*, 2009). Também tem sido descritos múltiplos sítios de ligação para um mesmo miRNA em um determinado mRNA, o que pode aumentar o nível de regulação da expressão (Grimson *et al.*, 2007).

Além das sequências regulatórias de poliadenilação, a classe mais abundante de elementos na 3'UTR são os sítios de ligação para miRNAs (Friedman *et al.*, 2009). Sendo esses sítios predominantemente localizados na região 3'UTR dos mRNAs, espera-se que variantes somáticas ou germinativas nessas sequências-alvo possam alterar a expressão gênica por abolir, enfraquecer ou criar complementaridade com miRNAs, possivelmente levando à variação fenotípica (Winter *et al.*, 2009; Endale Ahanda *et al.*, 2012). Alterações na expressão gênica representam um dos mecanismos-chave a partir dos quais células normais progridem a um estado tumoral, portanto, variantes nessas sequências-alvo podem influenciar no risco individual para o desenvolvimento de câncer (López de Silanes *et al.*, 2007; Diederichs *et al.*, 2016). Um exemplo clássico é o caso do SNP rs61764370 na região 3'UTR do proto-oncogene *KRAS*, cuja presença rompe a ligação do miRNA *let-7*,

resultando em uma superexpressão de *KRAS* e risco aumentado para o desenvolvimento de câncer de pulmão (Chin *et al.*, 2008).

Um número crescente de estratégias de bioinformática tem sido proposto para avaliar as interações entre miRNAs e regiões 3'UTR de diferentes transcritos (Bentwich, 2005; Witkos *et al.*, 2011; Dweep *et al.*, 2013), dentre as quais destacam-se as ferramentas *TargetScan* (Grimson *et al.*, 2007; Agarwal *et al.*, 2015) e *DIANA-microT-CDS* (Maragkakis *et al.*, 2009; Paraskevopoulou *et al.*, 2013). A maioria das ferramentas descritas exige uma complementaridade de bases perfeita entre a sequência *seed* do miRNA e o seu sítio de ligação na região 3'UTR (Dweep *et al.*, 2013). Algumas utilizam critérios adicionais, como a conservação filogenética desses sítios-alvo (Friedman *et al.*, 2009; Gaidatzis *et al.*, 2007), a estabilidade termodinâmica do pareamento miRNA:mRNA (Rehmsmeier *et al.*, 2004) e características estruturais da região 3'UTR, como a acessibilidade ao sítio-alvo ou contexto local (Long *et al.*, 2007; Marín e Vaníček, 2011).

Adicionalmente, existem fontes de dados de interações miRNA-mRNA em humanos com validação experimental, compilando resultados de diversas metodologias e ensaios funcionais (revisado em Elton e Yalowich 2015), dentre as quais destacam-se *miRTarBase* (Chou *et al.*, 2016), *starBase* (Li *et al.*, 2014), *miRecords* (Xiao *et al.*, 2009) e *TarBase* (Sethupathy *et al.*, 2006). Esses bancos de dados apresentam interfaces que permitem a avaliação crítica dos níveis de evidência experimental sugerindo a interação funcional entre determinado miRNA e mRNA-alvo (Elton e Yalowich 2015).

I.6. Poliadenilação alternativa e regulação diferencial da expressão por miRNAs

Cerca de 50% dos genes humanos possuem dois ou mais SP funcionais, denominados SP alternativos (Tian *et al.*, 2005; Elkon *et al.*, 2013). Nesse grupo de genes, a utilização de um SP específico determina a extensão em pares de bases da região 3'UTR do mRNA e, assim, o mesmo gene pode originar mRNAs com regiões 3'UTR de diferentes tamanhos (isoformas mais

longas e curtas de mRNA). Como os sítios de ligação para miRNAs estão predominantemente localizados na região 3'UTR dos mRNAs-alvo, transcritos primários com regiões 3'UTR de tamanhos diferentes exibem perfis de modulação da expressão por miRNAs diferenciados, sendo que transcritos curtos tendem a apresentar expressão aumentada (revisado em An *et al.*, 2013). Por outro lado, variações de sequência podem criar ou romper SP alternativos, podendo afetar a regulação global da expressão gênica por abolir, enfraquecer ou introduzir novos sítios-alvo para ligação de miRNAs em comparação com os transcritos selvagens mais abundantes derivados do mesmo gene (Legendre *et al.*, 2006; Sandberg *et al.*, 2008; Mayr e Bartel, 2009). Além disso, um padrão alterado de poliadenilação alternativa tem sido observado em diversos tipos de câncer, incluindo sarcomas, tumores de mama, colorretal e de pulmão (Mayr e Bartel, 2009; Morris *et al.*, 2012).

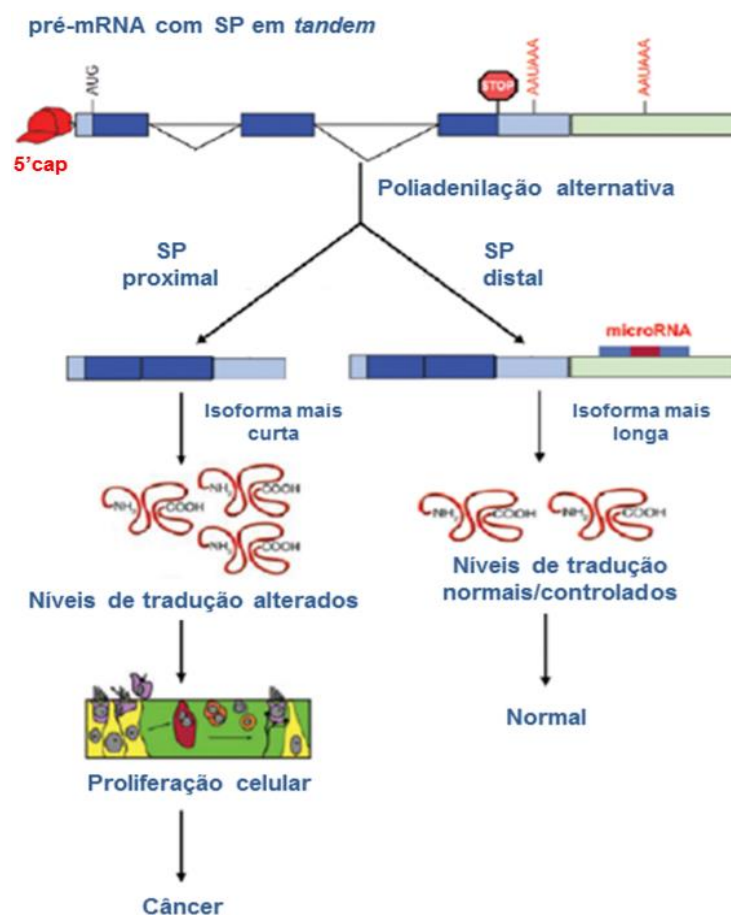


Figura 4. A poliadenilação alternativa de um transcrito primário pode interferir na regulação da sua expressão por miRNAs. Poliadenilação alternativa e modulação por miRNAs são mecanismos de regulação pós-transcricional da expressão gênica. A presença de sinais de poliadenilação alternativos na 3'UTR de determinado mRNA frequentemente resulta

em isoformas com a mesma sequência codificante, mas com diferentes tamanhos de 3'UTR (em pares de bases), enquanto os miRNAs regulam a expressão por se ligar em sítios específicos na 3'UTR desse mRNA. Sendo assim, miRNAs regulam diferencialmente isoformas de mRNA mais longas e curtas, levando a mudanças na expressão gênica que podem induzir proliferação celular (superexpressão de oncogenes, por exemplo) e favorecer mecanismos de carcinogênese. SP, sinal de poliadenilação. **Modificado de An et al., 2013.**

CAPÍTULO II – JUSTIFICATIVA

A análise da região 3'UTR pode ser negligenciada na rotina de diagnóstico molecular das síndromes de predisposição ao câncer, uma vez que os protocolos de sequenciamento abrangem principalmente as regiões codificantes e junções éxon-íntron. Essa região regulatória apresenta elementos de sequência que regulam a poliadenilação dos transcritos, especialmente os sinais de poliadenilação (SP) e sítios de clivagem (SC), cuja importância funcional é corroborada pela descrição de variantes deletérias localizadas nesses elementos em diferentes genes humanos, associadas com um grupo relativamente heterogêneo de doenças. Nesse contexto, e considerando-se os achados recentes do nosso grupo acerca da variante funcional situada no SP constitutivo do gene *TP53* (rs78378222) e sua possível associação com o fenótipo da Síndrome Li-Fraumeni-like, torna-se relevante caracterizar as sequências regulatórias de poliadenilação (SRP) em genes de predisposição ao câncer (GPC). Ademais, existem poucos estudos que avaliaram a variabilidade genética nas SRP e regiões adjacentes em GPC específicos. Portanto, a caracterização das SRP proposta nesse estudo poderá servir como base para um futuro mapeamento de variantes potencialmente patogênicas localizadas nessa região do grupo de genes estudado.

Por outro lado, observa-se uma escassez de dados na literatura sobre a regulação da expressão por miRNAs compartilhada entre genes supressores de tumor e oncogenes (a maioria das análises publicadas envolvendo GPC foca em vias de sinalização celular específicas), bem como poucas estratégias de bioinformática baseadas na combinação de dados derivados de predição computacional e validação experimental que permitam a busca por miRNAs reguladores desses genes. Sendo assim, mostra-se pertinente utilizar essas fontes de dados de maneira conjunta visando a identificação de miRNAs com papel potencialmente oncogênico ou supressor de tumor, considerando-se um conjunto abrangente de GPC implicados em diferentes vias de sinalização celular e processos fisiológicos.

CAPÍTULO III – OBJETIVOS

III. 1. Objetivos gerais

Caracterizar as sequências correspondentes aos sinais de poliadenilação (SP) e sítios de clivagem (SC) em um conjunto abrangente de genes de predisposição ao câncer (GPC), bem como identificar miRNAs e/ou famílias de miRNAs potencialmente oncogênicos e supressores de tumor considerando o mesmo grupo de genes.

III.2. Objetivos específicos

1. Caracterizar a sequência dos SP e SC entre os GPC estudados;
2. Comparar a frequência do tipo de hexâmero constituindo o SP dos GPC estudados com os dados em escala genômica publicados anteriormente;
3. Avaliar a frequência de sítios de poliadenilação alternativos entre os GPC estudados;
4. Determinar a frequência das sequências de dinucleotídeo caracterizando o SC nos GPC analisados;
5. Identificar, através de fontes de dados de predição computacional e com validação experimental, miRNAs e/ou famílias de miRNAs potencialmente oncogênicos e supressores de tumor considerando o conjunto de GPC analisado.

CAPITULO IV – MANUSCRITO

A comprehensive analysis of core polyadenylation sequences and regulation by microRNAs in a set of cancer predisposition genes

(Formatado para submissão ao periódico PLOS ONE)

A comprehensive analysis of core polyadenylation sequences and regulation by microRNAs in a set of cancer predisposition genes

Short title: Polyadenylation and microRNA regulation in cancer predisposition genes

Igor Araujo Vieira^{1,2*}; Mariana Recamonde-Mendoza³; Vandeclecio Lira da Silva^{4,5}; Delva Pereira Leão^{1,6}; Marina Roberta Scheid^{2,7}; Sandro José de Souza^{4,5}; Patricia Ashton-Prolla^{1,2,6,8}.

¹ Programa de Pós-graduação em Genética e Biologia Molecular, Universidade Federal do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, Brazil.

² Laboratório de Medicina Genômica, Serviço de Pesquisa Experimental, Hospital de Clínicas de Porto Alegre, Porto Alegre, Rio Grande do Sul, Brazil.

³ Instituto de Informática, Universidade Federal do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, Brazil.

⁴ Programa de Pós-Graduação em Bioinformática, Universidade Federal do Rio Grande do Norte, Natal, Rio Grande do Norte, Brazil.

⁵ Instituto do Cérebro, Universidade Federal do Rio Grande do Norte, Natal, Rio Grande do Norte, Brazil.

⁶ Serviço de Genética Médica, Hospital de Clínicas de Porto Alegre, Porto Alegre, Rio Grande do Sul, Brazil.

⁷ Programa de Pós-graduação em Ciências Médicas, Universidade Federal do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, Brazil.

⁸ Departamento de Genética, Universidade Federal do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, Brazil.

* Corresponding author

E-mail: igoraraujovieira@gmail.com (IAV)

Abstract

Introduction: Almost all eukaryotic mRNAs acquire a poly(A) tail at their 3' ends in a process termed polyadenylation. Two core polyadenylation elements (CPE) located in the 3' untranslated region (3'UTR) of pre-mRNAs play an essential role in this process: the polyadenylation signal (PAS) AAUAAA or its close variants; and the cleavage site (CS), preferentially a CA dinucleotide. Herein, we characterized PAS and CS sequences in a comprehensive set of cancer predisposition genes (CPGs) and performed an *in silico* investigation of the regulation by microRNAs (miRNAs) in the same gene set, in order to identify potentially tumor-suppressive and oncogenic miRNAs.

Methods: NCBI and alternative polyadenylation databases were queried to characterize CPE sequences in the selected CPGs (n=117), including 81 tumor suppressor genes and 17 oncogenes. microRNA-mediated regulation analysis was performed using predicted and validated data sources.

Results: Based on NCBI analysis, we did not find an established PAS in 21 CPGs, and most PAS already described (74.4%) had the canonical sequence AAUAAA, while 24.1% contained the variant AUUAAA. Interestingly, "AA" dinucleotide was the most common CS associated with this set of CPGs. The majority of CPGs (~90%) exhibited evidences of alternative polyadenylation (more than one functional PAS). Regarding the analysis of another class of 3'UTR-related elements (miRNAs), miR-192 family was significantly overrepresented among tumor suppressor genes ($P<0.01$), suggesting a potentially oncogenic function. In contrast, a larger number of miRNA families was strongly associated with regulation of oncogenes: miR-128, miR-1471, miR-483, miR-3170 and miR-218.

Conclusions: Overall, our study generated a landscape of CPE in CPGs and identified some miRNA families acting as potential tumor suppressors and oncogenes. This is the first study focused on the comprehensive characterization of 3'UTR-related elements in CPGs, which may be useful in the development of molecular analyses covering these frequently neglected regulatory sequences.

Keywords: Polyadenylation regulatory sequences, alternative polyadenylation, cancer predisposition genes, microRNAs.

Introduction

Polyadenylation including the cleavage of pre-mRNA followed by addition of a stretch of adenosine residues (called poly(A) tail) to the 3' end is an essential nuclear step of processing in almost all eukaryotic pre-mRNAs [1, 2]. Three *cis*-acting RNA sequence elements often termed the “core” polyadenylation elements (CPE) precisely determine the site of 3' end cleavage/polyadenylation in mammalian pre-mRNAs: (1) the poly(A) signal (PAS), a highly conserved hexamer AAUAAA or its close variants; (2) the actual polyadenylation site or cleavage site (CS), preferentially a “CA” dinucleotide 10-30 nucleotides (nt) downstream of the signal; and (3) a less well-conserved G-U-rich sequence 14-70 nt downstream of the PAS [3-7]. Mainly five *trans*-acting protein factors make up the machinery involved in this process: cleavage and polyadenylation specificity factor (CPSF), cleavage stimulatory factor (CstF), cleavage factors I and II (CFI and CFII), and poly(A) polymerase (PAP) [1, 2, 8]. The CPSF recognizes the AAUAAA sequence and the CstF binds to the G-U-rich sequence, while CFI and CFII cleave the RNA in CS and generate a 3' end for PAP to perform polyadenylation [9-11]. Moreover, alternative polyadenylation (APA), defined as use of more than one functional PAS/CS, occurs in at least 50% of human genes, allowing a single gene to encode multiple mRNA transcripts with variable 3' untranslated regions (3'UTR) [12, 13].

The known regulatory role of polyadenylation in mRNA localization, stability, and translation and the emerging link between deleterious variants located in CPE and human diseases underscore the need to fully characterize them [1, 14]. Alterations in CPE and neighboring sequences in the 3'UTR disrupt the cleavage and polyadenylation steps resulting in various human pathologies [14-16]. Recently, functional CPE-related germline variants were reported in cancer predisposition genes (CPGs) and were associated with hereditary cancer syndromes (HCS). The single nucleotide polymorphism (SNP) rs78378222 changes the constitutive PAS of the tumor suppressor gene *TP53* [17] and was associated with the Li-Fraumeni-like syndrome phenotype [18], while a functional variant (duplication of 20 base pairs) upstream of the PAS of *MSH6* DNA-repair gene was identified in two patients fulfilling clinical criteria for Lynch syndrome [19].

On the other hand, microRNAs (miRNAs) are endogenous, single-stranded, small non-coding RNAs (18-25 nt in length) that post-transcriptionally regulate gene expression through mRNA degradation and/or translation repression. In addition to CPE sequences, miRNA binding sites represent another class of elements predominantly located in the 3'UTR of mRNAs [20, 21]. Remarkable studies showed that miRNA mutations or altered expression levels correlate with various human cancers and indicate that miRNAs can function as tumor suppressors and/or oncogenes [22-25]. The deletion or downregulation of a miRNA that has a tumor suppressor role, modulating expression of oncogenes, leads to inappropriate amounts of the miRNA-target oncoproteins. The overall outcome might involve increased cellular proliferation, ultimately contributing to carcinogenesis. In addition, the amplification or overexpression of an oncogenic miRNA can reduce or eliminate the expression of a miRNA-target tumor-suppressor gene also resulting in excess cellular proliferation and potentially, tumor formation [25].

In the present study, we aimed to characterize CPE (specifically PAS and CS) in a comprehensive set of CPGs and performed an *in silico* investigation of the regulation by miRNAs in the same group of genes, in order to identify miRNAs with potentially oncogenic or tumor suppressor function. These analyses generated a landscape of the CPE in CPGs that can be used in the development of more comprehensive sequencing analyses including these regulatory sequences often neglected during the routine molecular diagnosis. Furthermore, computationally predicted and experimentally validated data allowed us to suggest a potential oncogenic function of the miR-192 family.

Materials and Methods

Selection criteria of the cancer predisposition genes (CPGs) included in this study

The CPGs included in this study (n=117 genes) were selected from literature search using the following criteria: (1) genes associated with well described genetic syndromes in which the sole or main clinical phenotype is hereditary predisposition to cancer (classic HCS and certain types of familial cancer, n=71); (2) genes associated with syndromes/conditions in which one of

the phenotypes is increased tumor predisposition, including some congenital malformation syndromes (n=36); and (3) susceptibility genes to certain sporadic tumor(s), including genes encoding translesion DNA polymerases implicated in cancer either by germline/somatic mutations or by altered expression in tumors (n=10). The minimum requirement considered for linking genes and predisposition to cancer was the consistent association between sequence variants (or significantly altered expression levels) in such genes and increased risk for developing certain tumor type(s) through independent case-control studies. Most of the CPGs (n=76) were selected from the review of Garber and Offit (2005) [26], and the other CPGs from specific studies mentioned in S1 Table. Based on experimental evidence regarding the effect of cancer-associated mutations in each gene, the CPGs were classified into: oncogenes (n=17), genes whose alterations cause gain-of-function effects to promote carcinogenesis; tumor suppressor genes (n=81), genes in which loss-of-function mutations contribute to the malignant phenotype [27]; and “undetermined function” genes (n=19), which did not fit into these categories and/or for which there is insufficient evidence about the mechanism of association with tumor development (main references related to this classification are shown in S1 Table). In addition, 2 genes previously reported as putative oncogenes (*SHOC2* e *PALLD*), as well as 8 putative tumor suppressor genes (*EPHB2*, *EXT1*, *EXT2*, *GPC3*, *ELAC2*, *MSR1*, *POT1* and *PRKAR1A*), were included in the respective groups.

Analysis of polyadenylation databases

Briefly, “Nucleotide database” within NCBI (reference source) and a computational method were employed to characterize PAS and CS sequences of the selected CPGs. The recently released alternative polyadenylation (APA) databases APADB [28] and APASdb [29] were used to assess the occurrence of APA in these genes.

Regarding the NCBI analyses, we searched for the transcript variant with the largest extension in base pairs (bp) derived from each gene, wherein mRNA sequences predicted by computational analysis were not considered. This specific criterion was used to avoid loss of data on possible PAS/CS sequences related to a certain gene, which may occur, for example, when the transcript

variant has a shorter 3'UTR. After identifying the longest variant transcript, PAS and CS were located in the sequence selecting the features "regulatory" and "polyA_site", respectively. The distance (in bp) between the PAS and CS was calculated according to their positions indicated in the NCBI interface.

Regarding CPGs with no PAS described in the NCBI database, we developed a computational method to identify putative PAS in the full sequence of corresponding transcripts. First, RefSeq mRNA sequences were obtained from UCSC Genome Browser (human genome version hg19). Screening of the canonical AAUAAA hexamer and its functional variants previously reported in human genome [30] was the second step, restricting the definition of putative PAS to the 3'-most hexamers identified in the transcript sequence. This step was conducted using in-house Perl scripts (further information upon request).

For the APA analyses, APADB data were filtered for CS identified only in human peripheral blood samples and supporting read alignments for each CS were visualized in a genome browser. This database was built by 3' end sequencing using massive analysis of complementary DNA ends, a high-throughput next-generation sequencing-based technique [28]. Additionally, APASdb allowed the counting of CS mapped in all transcript variants for each gene from samples of 22 normal human tissues (searching dataset: hg19 human-all22-tissues). This database, based on the sequencing APA sites (SAPAS) method reported previously [31], also provided both the position and usage quantification of a given alternative CS among transcripts derived from the same gene by computing their corresponding normalized-reads [29].

MicroRNA-mediated regulation analysis

To investigate the miRNAs that regulate expression of selected tumor suppressor genes (n=81) and oncogenes (n=17), first we collected experimentally validated data of miRNA-target gene interactions in humans from the miRTarBase release 6.0 [32], starBase v2.0 [33], TarBase v5.0 [34] and miRecords v4.0 [35] databases. Data derived from miRTarBase was restricted to interactions classified as functional, including those with weak experimental evidence, while starBase data was filtered to keep only interactions predicted by two or more software and with experimental support by at least one experiment ("low stringency" option). For other databases, we

included all available interactions in the analysis. To complement experimental data, we performed computational target prediction using TargetScan v7.2 [36], Diana MicroT-CDS [37] and miRanda-mirSVR (August 2010 Release) [38]. To control the false positive rate and restrain the large list of predicted miRNAs, we adopted the following filtering criteria: (a) for TargetScan, we considered only interactions involving 8mer, 7mer and 6mer conserved miRNA sites and with Context++ score lower than -0.2; (b) for Diana MicroT-CDS, we kept only predictions with a score higher than 0.9; (c) for miRanda-mirSVR, we selected interactions with mirSVR score lower than -0.1 (regarded as “good mirSVR score”) involving conserved miRNAs; (d) finally, we defined as the final set of predicted miRNA-target gene interactions all relations suggested by at least two computational tools.

The set of validated and predicted miRNA-target gene interactions collected in the described sources was filtered in order to identify the interactions whose target gene was a tumor suppressor gene or oncogene included in this study. Moreover, whenever possible, the miRNAs found in this step were clustered into families according to the miRBase 21 annotations (some miRNAs do not have a family classification). To better visualize the overlap of validated and predicted data, we created regulatory networks. Cytoscape software version 3.1.0 was used for graphical network visualization [39]. All miRNAs and genes identifiers were mapped to miRBase 21 and HUGO Gene Nomenclature Committee (HGNC) standards, respectively, to maintain an uniformity and enable consistent data integration across different sources.

Statistical analyses

SPSS version 18.0 (IBM) was used for data handling and statistical analyses. Fisher’s exact test was used to compare the frequency of different PAS hexamer types between the groups of oncogenes and tumor suppressor genes. The same test was also applied to identify miRNAs and miRNA families over-represented in the regulation of the groups of oncogenes and tumor suppressors studied, i.e., they have a greater number of target genes among the genes of interest included in this study than would be expected by chance. This analysis compared the proportion of genes in the list of interest that are targets of a specific miRNA/miRNA family against the proportion of targets of

the same miRNA/miRNA family considering a background set. The complete list of validated and predicted interactions collected from the aforementioned sources, without filtering it for interactions involving the CPGs included in this study, was adopted as the background set in this “over-representation analysis”. More details about the composition of the background set are detailed in S2 Table. Statistical tests were two-sided and *P*-values of <0.01 were considered statistically significant.

Results

The number and sequence of constitutive PAS and CS in the longest variant transcript derived from each CPG are shown in Table 1. Distance (in bp) between these specific CPE and analysis of the presence of APA sites are also indicated. Importantly, we did not find description of a PAS sequence in 21 of the 117 studied CPGs (~18%) using NCBI database (reference source). Among these 21 genes were genes associated with highly prevalent HCS or phenotypes such as *BRCA2* (Hereditary Breast and Ovarian Cancer Syndrome), *BRIP1* (Familial Breast and Ovarian Cancer), *CDK4* and *TERT* (Familial Malignant Melanoma), *MEN1* (Multiple Endocrine Neoplasia Type 1), and *PMS2* (Lynch Syndrome).

Table 1. Characterization of polyadenylation regulatory sequences in cancer predisposition genes included in this study.

Gene (ONC, TSG or UND)	Syndrome/Tumor predisposition	NCBI database (Constitutive PAS and CS sequences) ^a			Alternative polyadenylation (Number of poly(A) sites)	
		Number of PAS and sequence	Number of CS and sequence	Distance between the PAS and CS (bp) ^b	APADB ^c	APASdb ^d
<i>APC</i> (TSG)	Familial Adenomatous Polyposis, including attenuated phenotype; Gardner Syndrome and Turcot's Syndrome (mainly colorectal cancer)	1 (CATAAA)	1 (CA)	16	6	18
<i>ATM</i> (TSG)	Ataxia-telangiectasia (breast cancer, leukemia and lymphoma)	1 (AATAAA)	1 (GC)	12	4	31
<i>BAP1</i> (TSG)	Predisposition to malignant mesothelioma, cutaneous and uveal melanoma, and renal cell carcinoma	1 (AATAAA)	1 (AA)	28	2	6
<i>BARD1</i> (TSG)	Hereditary Breast and Ovarian Cancer Syndrome	1 (AATAAA)	1 (TG)	16	2	7
<i>BLM</i> (TSG)	Bloom Syndrome (leukemia, lymphoma and colorectal cancer)	1 (AATAAA)	2 (Both AA)	13-16	3	7
<i>BMPR1A</i> (TSG)	Juvenile Polyposis Syndrome (colorectal cancer and other tumors of the gastrointestinal tract)	1 (AATAAA)	1 (GA)	18	2	10
<i>BRAF</i> (ONC)	LEOPARD, Cardiofaciocutaneous and Noonan syndromes (predisposition to various tumor types); predisposition to melanoma	3 (All hexamers are AATAAA)	3 (All CS are AA)	18-26	2	18
<i>BRCA1</i> (TSG)	Hereditary Breast and Ovarian Cancer Syndrome	1 (AATAAA)	1 (CA)	18	1	5
<i>BRCA2</i> (TSG)	Hereditary Breast and Ovarian Cancer Syndrome and Fanconi Anemia (predisposition to leukemia and certain solid tumors)	NI	NI	NI	1	11
<i>BRIP1</i> (TSG)	Fanconi Anemia; familial breast and ovarian cancer	NI	1 (AA)	NI	NI	7
<i>CDC73</i> (TSG)	Familial Parathyroid Cancer (often called Hyperparathyroidism-Jaw Tumor Syndrome)	4 (2 ATATAA and 2 AATAAA)	4 (2 AA, TA and TC)	8-26	10	31
<i>CDH1</i> (TSG)	Hereditary Gastric Cancer (predisposition to gastric and colorectal tumors)	1 (ATATAA)	1 (AT)	16	1	10

Gene (ONC, TSG or UND)	Syndrome/Tumor predisposition	NCBI database (Constitutive PAS and CS sequences) ^a			Alternative polyadenylation (Number of poly(A) sites)	
		Number of PAS and sequence	Number of CS and sequence	Distance between the PAS and CS (bp) ^b	APADB ^c	APASdb ^d
<i>CDKN1C</i> (TSG)	Beckwith-Wiedemann Syndrome (predisposition to embryonal tumors and childhood tumors)	1 (AATAAA)	1 (GC)	17	NI	3
<i>CDKN2A</i> (TSG)	Melanoma-Pancreatic Cancer Syndrome	2 (Both AATAAA)	1 (CA)	21-27	2	4
<i>CDK4</i> (ONC)	Familial Malignant Melanoma	NI	2 (AA and GC)	NI	1	5
<i>CHEK2</i> (TSG)	Hereditary Breast and Colorectal Cancer	1 (ATTAAA)	1 (GG)	16	NI	2
<i>DDB2</i> (TSG)	Xeroderma pigmentosum (predisposition to skin cancer, melanoma, leukemia and neoplasms in internal organs)	1 (AATAAA)	1 (AG)	14	1	9
<i>ELAC2</i> (TSG)	Hereditary prostate cancer	2 (Both AATAAA)	2 (GA and TA)	17-19	3	7
<i>EPCAM</i> (UND)	Lynch Syndrome (often called Hereditary Nonpolyposis Colorectal Cancer)	2 (ATTAAA and AATAAA)	2 (GA and AA)	22-30	2	7
<i>EPHB2</i> (TSG)	Predisposition to prostate cancer	2 (Both AATAAA)	2 (AA and TA)	13-17	2	7
<i>ERCC2</i> (TSG)	Xeroderma pigmentosum	NI (transcript variant 1) and AATAAA (transcript variant 2)	GA (transcript variant 1) and AA (transcript variant 2)	42 (transcript variant 2)	2	2
<i>ERCC3</i> (TSG)	Xeroderma pigmentosum	1 (AATAAA)	1 (GT)	21	1	3
<i>ERCC4</i> (TSG)	Xeroderma pigmentosum	NI	NI	NI	3	17
<i>ERCC5</i> (TSG)	Xeroderma pigmentosum	1 (AATAAA)	2 (TC and TA)	18-24	1	10
<i>EXT1</i> (TSG)	Hereditary Multiple Osteochondromas and Bone Cancer Predisposition Syndrome (predisposition to chondrosarcoma)	2 (Both ATTAAA)	2 (GA and AA)	17-25	2	19
<i>EXT2</i> (TSG)	Hereditary Multiple Osteochondromas and Bone Cancer Predisposition Syndrome	2 (Both ATTAAA)	2 (TT and CA)	18-22	2	16
<i>FANCA</i> (TSG)	Fanconi Anemia (predisposition to leukemia and certain solid tumors)	NI	NI	NI	4	5
<i>FANCB</i> (TSG)	Fanconi Anemia	NI	NI	NI	NI	NI
<i>FANCC</i> (TSG)	Fanconi Anemia	2 (ATTAAA and AATAAA)	2 (Both AA)	13-17	1	11
<i>FANCD2</i> (TSG)	Fanconi Anemia	1 (AATAAA)	1 (TC)	14	2	5
<i>FANCE</i> (TSG)	Fanconi Anemia	1 (AATAAA)	1 (TG)	15	1	1

Gene (ONC, TSG or UND)	Syndrome/Tumor predisposition	NCBI database (Constitutive PAS and CS sequences) ^a			Alternative polyadenylation (Number of poly(A) sites)	
		Number of PAS and sequence	Number of CS and sequence	Distance between the PAS and CS (bp) ^b	APADB ^c	APASdb ^d
<i>FANCF</i> (TSG)	Fanconi Anemia	1 (AATAAA)	1 (CA)	16	6	6
<i>FANCG</i> (TSG)	Fanconi Anemia	1 (AGTAAA)	1 (AT)	12	1	1
<i>FANCI</i> (TSG)	Fanconi Anemia	NI	NI	NI	2	11
<i>FANCL</i> (TSG)	Fanconi Anemia	1 (AATAAA)	1 (AA)	16	1	6
<i>FAS</i> (UND)	Autoimmune Lymphoproliferative Syndrome (predisposition to lymphoma)	2 (Both ATAAA)	2 (Both TA)	13-18	3	5
<i>FASLG</i> (UND)	Autoimmune Lymphoproliferative Syndrome	1 (AATAAA)	1 (AA)	20	4	1
<i>FH</i> (TSG)	Hereditary Leiomyomatosis and Renal Cell Carcinoma Syndrome	NI	NI	NI	1	5
<i>FLCN</i> (TSG)	Birt-Hogg-Dube Syndrome (predisposition to cutaneous hamartomas and renal tumors)	1 (AATAAA)	1 (AA)	15	3	3
<i>GPC3</i> (TSG)	Simpson-Golabi-Behmel Syndrome (predisposition to embryonal tumors)	1 (AATAAA)	1 (AA)	14	NI	NI
<i>GREM1</i> (ONC)	Juvenile Polyposis Syndrome and Hereditary Mixed Polyposis Syndrome (colorectal cancer)	1 (AATAAA)	1 (TG)	19	NI	5
<i>HRAS</i> (ONC)	Costello and Noonan Syndromes (predisposition to various tumor types)	NI	1 (GA)	NI	1	1
<i>IL2RG</i> (UND)	X-Linked Severe Combined Immunodeficiency (predisposition to lymphoma)	1 (AATAAA)	1 (AA)	23	1	NI
<i>KIT</i> (ONC)	Familial Gastrointestinal Stromal Tumors	1 (AATAAA)	1 (AA)	18	1	3
<i>KRAS</i> (ONC)	Cardiofaciocutaneous, Costello and Noonan Syndromes (predisposition to various tumor types)	NI	NI	NI	6	17
<i>LIG4</i> (TSG)	DNA Ligase IV Syndrome (predisposition to leukemia and lymphoma)	1 (AATAAA)	1 (AA)	20	1	8
<i>MAP2K1</i> (ONC)	Cardiofaciocutaneous and Noonan Syndromes	1 (AATAAA)	1 (AT)	12	1	4
<i>MAP2K2</i> (ONC)	Cardiofaciocutaneous and Noonan Syndromes	NI	1 (AA)	NI	3	3
<i>MAX</i> (TSG)	Hereditary Pheochromocytoma and Paraganglioma	1 (AATAAA)	2 (Both TT)	18 ^e	12	18

Gene (ONC, TSG or UND)	Syndrome/Tumor predisposition	NCBI database (Constitutive PAS and CS sequences) ^a			Alternative polyadenylation (Number of poly(A) sites)	
		Number of PAS and sequence	Number of CS and sequence	Distance between the PAS and CS (bp) ^b	APADB ^c	APASdb ^d
<i>MC1R</i> (UND)	Familial Malignant Melanoma	1 (AATAAA)	1 (TA)	23	1	3
<i>MDM2</i> (ONC)	Related to Li-Fraumeni Syndrome (genetic modifiers; further information in <i>TP53</i> gene)	1 (AATAAA)	1 (CA)	14	7	26
<i>MDM4</i> (ONC)		3 (2 ATTA AAA and AATAAA)	3 (2 CA and TA)	16-22	7	30
<i>MEN1</i> (TSG)	Multiple Endocrine Neoplasia Type 1	NI	1 (TA)	NI	NI	1
<i>MET</i> (ONC)	Hereditary Papillary Renal Cell Carcinoma	1 (ATTAAA)	1 (TA)	22	1	1
<i>MITF</i> (ONC)	Predisposition to malignant melanoma and renal cell carcinoma	1 (ATTAAA)	2 (CA and TA)	14-18	1	18
<i>MLH1</i> (TSG)	Lynch Syndrome	1 (AATAAA)	1 (AA)	13	2	12
<i>MRE11A</i> (TSG)	Ataxia-telangiectasia-like disorder	2 (ATTAAA and AATAAA); overlapping PAS	4 (2 TA, CA and AA)	16-34	3	11
<i>MSH2</i> (TSG)	Lynch Syndrome	1 (AATAAA)	1 (GA)	22	1	15
<i>MSH6</i> (TSG)	Lynch Syndrome	1 (AATAAA)	1 (GA)	18	1	8
<i>MSR1</i> (TSG)	Predisposition to prostate cancer	1 (ATTAAA)	1 (GA)	17	2	16
<i>MUTYH</i> (TSG)	Familial Adenomatous Polyposis 2 (Polyposis associated with <i>MUTYH</i>)	NI	1 (TA)	NI	1	2
<i>NBN</i> (TSG)	Nijmegen Breakage Syndrome (predisposition to leukemia, lymphoma and certain solid tumors)	2 (ATTAAA and AATAAA)	3 (All CS are AA)	17-36	3	17
<i>NF1</i> (TSG)	Neurofibromatosis type 1 (predisposition to neurofibrosarcomas, optic gliomas, meningiomas and other malignant tumors)	1 (AATAAA)	2 (AA and GA)	23-1806	9	32
<i>NF2</i> (TSG)	Neurofibromatosis type 2 (predisposition to vestibular schwannomas, meningiomas and other malignant tumors)	1 (AATAAA)	1 (TA)	18	2	8
<i>NSD1</i> (UND)	Beckwith-Wiedemann Syndrome	1 (AATAAA)	3 (TA, AA, GA)	13-4469	3	23
<i>PALB2</i> (TSG)	Fanconi Anemia; predisposition to breast and pancreas cancer	1 (AATAAA)	1 (TC)	18	1	2
<i>PALLD</i> (ONC)	Familial Pancreatic Adenocarcinoma	2 (Both AATAAA)	2 (AA and TA)	11-13	1	46

Gene (ONC, TSG or UND)	Syndrome/Tumor predisposition	NCBI database (Constitutive PAS and CS sequences) ^a			Alternative polyadenylation (Number of poly(A) sites)	
		Number of PAS and sequence	Number of CS and sequence	Distance between the PAS and CS (bp) ^b	APADB ^c	APASdb ^d
<i>PMS1</i> (TSG)	Lynch Syndrome	2 (AATAAA and ATTAAA)	2 (CA and GA)	13-16	3	17
<i>PMS2</i> (TSG)	Lynch Syndrome	NI	1 (GA)	NI	1	3
<i>POLB</i> (UND)	Alternative DNA polymerase; overexpression in certain solid tumors	1 (AATAAA)	1 (GA)	20	1	3
<i>POLD1</i> (UND)	Alternative DNA polymerase; Polymerase Proofreading-associated polyposis; predisposition to colorectal, breast and endometrial cancer	1 (AATAAA)	1 (TG)	20	1	3
<i>POLE</i> (UND)	Alternative DNA polymerase; Polymerase Proofreading-associated polyposis; predisposition to colorectal, pancreas, ovarian and small intestine cancer	NI	1 (TC)	NI	1	4
<i>POLH</i> (UND)	Translesion DNA Polymerase; Xeroderma pigmentosum Variant Type (predisposition to skin cancer)	2 (Both AATAAA)	2 (CA and TA)	19-20	4	1
<i>POLK</i> (UND)	Translesion DNA Polymerase; overexpression in rectum and lung tumors	NI	NI	NI	2	16
<i>POLQ</i> (UND)	Translesion DNA Polymerase; overexpression in certain solid tumors	1 (AATAAA)	1 (TA)	20	NI	1
<i>POT1</i> (TSG)	Familial Glioma and Melanoma	2 (Both AATAAA)	2 (AA and GA)	16-17	3	13
<i>PRCC</i> (UND)	Papillary Renal Cancer Syndrome	1 (AATAAA)	1 (CA)	18	2	3
<i>PRKAR1A</i> (TSG)	Carney Complex Disorder (predisposition to myxoid subcutaneous, testicular sertoli cell and thyroid cancer, and other solid tumors)	4 (3 AATAAA and 1 ATTAAA)	4 (TT, GA and 2 AA)	16-31	21	25
<i>PRSS1</i> (UND)	Familial pancreatic cancer	1 (AATAAA)	1 (TC)	19	NI	1
<i>PTCH1</i> (TSG)	Gorlin Syndrome, often called Basal Cell Nevus Syndrome (predisposition to basal cell carcinoma and medulloblastoma)	NI	NI	NI	1	10
<i>PTEN</i> (TSG)	Cowden Syndrome (predisposition to breast, thyroid, endometrial and other cancers)	5 (3 ATTAAA and 2 AATAAA)	6 (2 AA, 2 TA, GA and CA)	14-19	21	61
<i>PTPN11</i> (UND)	Noonan and LEOPARD Syndromes	3 (2 AATAAA and 1 ATTAAA)	3 (2 GA and AA)	16-32	3	14

Gene (ONC, TSG or UND)	Syndrome/Tumor predisposition	NCBI database (Constitutive PAS and CS sequences) ^a			Alternative polyadenylation (Number of poly(A) sites)	
		Number of PAS and sequence	Number of CS and sequence	Distance between the PAS and CS (bp) ^b	APADB ^c	APASdb ^d
<i>RAD50</i> (TSG)	Nijmegen Breakage-like Syndrome	1 (AATAAA)	2 (AA and CA)	16-741	7	31
<i>RAD51</i> (TSG)	Fanconi Anemia and Familial Breast and Ovarian Cancer	2 (Both AATAAA)	4 (3 AA and GT)	20-80	NI	6
<i>RAD51C</i> (TSG)	Hereditary Breast and Ovarian Cancer	1 (AATAAA)	2 (CA and TA)	11-30	2	6
<i>RAD51D</i> (TSG)	Hereditary Breast and Ovarian Cancer	1 (AATAAA)	1 (AA)	15	2	4
<i>RAF1</i> (ONC)	Noonan and LEOPARD 2 Syndromes	2 (Both AATAAA); overlapping PAS	1 (AA)	15-20	3	7
<i>RB1</i> (TSG)	Hereditary Retinoblastoma	1 (AATAAA)	1 (TC)	15	6	21
<i>RECQL4</i> (TSG)	Rothmund-Thomson Syndrome (predisposition to basal cell carcinoma, squamous cell carcinoma and osteosarcoma)	1 (AATAAA)	1 (GG)	23	1	1
<i>RET</i> (ONC)	Multiple Endocrine Neoplasia Type 2	1 (AATAAA or ATTAAA; isoform-specific PAS)	1 (TT or AA, isoform- specific CS)	11-15	NI	5
<i>RNASEL</i> (TSG)	Hereditary prostate cancer	1 (AATAAA)	1 (TA)	15	3	NI
<i>SBDS</i> (UND)	Shwachman-Diamond Syndrome (predisposition to myelodysplasia and acute myeloid leukemia)	1 (AATAAA)	1 (TT)	17	NI	1
<i>SDHA</i> (TSG)	Hereditary Paraganglioma-Pheochromocytoma	1 (AATAAA)	2 (AA and CA)	32-415	1	7
<i>SDHAF2</i> (TSG)	Hereditary Paraganglioma-Pheochromocytoma	1 (ATTAAA)	3 (2 AA and TA)	18-42	2	5
<i>SDHB</i> (TSG)	Hereditary Paraganglioma-Pheochromocytoma	2 (Both AATAAA)	2 (Both AA)	15-29	1	1
<i>SDHC</i> (TSG)	Hereditary Paraganglioma-Pheochromocytoma	1 (ATTAAA)	2 (GA and AA)	17-1555	1	5
<i>SDHD</i> (TSG)	Hereditary Paraganglioma-Pheochromocytoma	2 (Both AATAAA)	2 (AA and TA)	0-18	1	1
<i>SH2D1A</i> (UND)	X-Linked Lymphoproliferative Disease	1 (AATAAA)	1 (AA)	24	2	NI
<i>SHOC2</i> (ONC)	Cardiofaciocutaneous and Noonan Syndromes	NI	2 (TA and GA)	NI	3	19
<i>SLX4</i> (TSG)	Fanconi Anemia	1 (AATAAA)	1 (AA)	19	1	2
<i>SMAD4</i> (TSG)	Juvenile Polyposis Syndrome	3 (2 AATAAA and ATTAAA)	3 (2 AA and TT)	11-25	6	17
<i>SMARCA4</i> (TSG)	Rhabdoid Tumor Predisposition Syndrome	2 (AATAAA and ATTAAA); overlapping PAS	2 (GA and CC)	20-25	4	10

Gene (ONC, TSG or UND)	Syndrome/Tumor predisposition	NCBI database (Constitutive PAS and CS sequences) ^a			Alternative polyadenylation (Number of poly(A) sites)	
		Number of PAS and sequence	Number of CS and sequence	Distance between the PAS and CS (bp) ^b	APADB ^c	APASdb ^d
<i>SMARCB1</i> (TSG)	Rhabdoid Tumor Predisposition Syndrome	1 (AATAAA)	1 (CA)	30	1	1
<i>SOS1</i> (ONC)	Noonan Syndrome	1 (AATAAA)	1 (AA)	24	5	47
<i>STK11</i> (TSG)	Peutz-Jeghers Syndrome (tumors of the gastrointestinal tract, breast and ovarian cancer)	1 (AATAAA)	1 (TG)	19	3	4
<i>TERT</i> (UND)	Familial Malignant Melanoma	NI	NI	NI	NI	1
<i>TMEM127</i> (TSG)	Familial Pheochromocytoma and predisposition to renal cell carcinoma	NI	1 (AA)	NI	9	10
<i>TP53</i> (TSG)	Li-Fraumeni Syndrome (sarcoma, breast, brain and adrenocortical tumors, and other cancers)	1 (AATAAA)	1 (CA)	11	3	6
<i>TSC1</i> (TSG)	Tuberous sclerosis (facial angiofibroma, renal angiomyolipoma, renal cell carcinoma, cardiac rhabdomyoma, and other tumors)	1 (AATAAA)	1 (AC)	19	4	19
<i>TSC2</i> (TSG)	Tuberous sclerosis	2 (Both AATAAA); overlapping PAS	4 (2 GA and 2 CA, isoform-specific CS)	20-43	1	4
<i>VHL</i> (TSG)	Von Hippel-Lindau Syndrome (retina and central nervous system hemangioblastomas, and renal cell carcinoma) and Familial Pheochromocytoma	1 (ATTAATA)	1 (AA)	18	6	8
<i>WAS</i> (UND)	Wiskott-Aldrich Syndrome (predisposition to hematopoietic malignancies)	1 (AATAAA)	1 (AA)	26	3	NI
<i>WRN</i> (TSG)	Werner Syndrome (soft tissue sarcomas, osteosarcoma, meningioma, and other tumors)	2 (ATTAATA and AATAATA)	4 (2 AA and 2 TA)	16-24	2	13
<i>WT1</i> (TSG)	Hereditary Wilms tumor (childhood renal cancer)	1 (ATTAATA)	1 (AA)	16	NI	6
<i>XPA</i> (TSG)	Xeroderma pigmentosum	1 (AATAATA)	1 (GA)	20	1	10
<i>XPC</i> (TSG)	Xeroderma pigmentosum	1 (AATAATA)	1 (AA)	12	3	12
<i>XRCC3</i> (TSG)	Familial Malignant Melanoma	1 (AATAATA)	1 (CA)	14	1	2

ONC, oncogene; TSG, tumor suppressor gene; UND, "undetermined function" gene; PAS, polyadenylation signals; CS, cleavage sites; bp, base pairs; NI, not identified.

^a Constitutive PAS and CS indicated in NCBI database for the longest variant transcript (in bp) derived from each cancer predisposition gene; ^b Distance (in bp) between the PAS and CS calculated according to their positions indicated in the NCBI interface; ^c Analysis using this database included only cleavage/polyadenylation sites identified from human peripheral blood samples (Müller *et al.*, 2014, reference 28); ^d Analysis considering alternative CS mapped in all transcript variants (for each gene) from samples of 22 normal human tissues (You *et al.*, 2015, reference 29); ^e One of the polyadenylation sites in the longest variant transcript derived from *MAX* gene (NM_145113.2) does not have a close PAS (upstream region).

Next, we determined the frequency of different PAS hexamers and compared this information with data from two previous genomic-scale studies [12, 30] (Table 2). About 74% of the PAS described for the set of CPGs exhibited the canonical hexamer AAUAAA, while approximately 24% had the functional variant PAS AUUAAA. These frequencies did not differ significantly between the two main groups of genes (oncogenes and tumor suppressor genes) ($p=0.78$). Moreover, we identified two CPGs with less frequently occurring hexanucleotides representing their constitutive PAS: *APC* (CAUAAA) and *FANCG* (AGUAAA) (Table 1). Curiously, overlapping PAS were described for the *MRE11A*, *RAF1*, *SMARCA4* and *TSC2* genes, and the distance calculated between PAS and CS sequences for thirteen CPGs (~11%) was greater than that established by long-standing experimental studies (10-30 bp) [3-5], suggesting that this theoretical convention does not apply to all human genes (Table 1).

Table 2. Comparison between the frequencies of human polyadenylation signal hexamers identified by previous genomic studies and in cancer predisposition genes analyzed in the current study.

Hexamer	Frequency, % (ranking)				
	Genomic data; reported by Tian <i>et al.</i> , 2005 ^a	Genomic data; reported by Beadoing <i>et al.</i> , 2000 ^b	Cancer predisposition genes; N=96 ^c	Oncogenes; N=12 ^d	Tumor suppressor genes; N=69 ^e
AAUAAA	53.18 (1)	58.20 (1)	74.45 (1)	73.68 (1) ^f	76.81 (1) ^f
AUUAAA	16.78 (2)	14.90 (2)	24.09 (2)	26.32 (2) ^f	20.29 (2) ^f
UAUAAA	4.37 (3)	3.20 (3)	NI	NI	NI
AGUAAA	3.72 (4)	2.70 (4)	0.73 (3)	NI	1.45 (3)
AAGAAA	2.99 (5)	1.10 (10)	NI	NI	NI
AAUAUA	2.13 (6)	1.70 (5)	NI	NI	NI
AAUACA	2.03 (7)	1.20 (8)	NI	NI	NI
CAUAAA	1.92 (8)	1.30 (6)	0.73 (3)	NI	1.45 (3)
GAUAAA	1.75 (9)	1.30 (7)	NI	NI	NI
AAUGAA	1.56 (10)	0.80 (11)	NI	NI	NI
UUUAAA	1.20 (11)	1.20 (9)	NI	NI	NI
ACUAAA	0.93 (12)	0.60 (13)	NI	NI	NI
AAUAGA	0.60 (13)	0.70 (12)	NI	NI	NI

NI, not identified.

^a Frequency of hexamer/hexanucleotide types identified as polyadenylation signals (PAS) in the human genome as previously reported by Tian *et al.*, 2005 [12]. The method developed by Beadoing *et al.*, 2000 [30] was applied to the human genome sequences located 1 to 40 nucleotides upstream of polyadenylation sites to detect hexamers that may function as PAS.

^b Frequency of hexamer types identified as PAS in the human genome using a method described by Beadoing *et al.*, 2000 [30].

^c Frequency considering PAS indicated in NCBI for 96 of 117 genes included in this study (the remaining genes have no description of PAS in this database). Each hexamer was counted as a unit, including the case of transcripts with more than one PAS sequence and transcripts derived from *RET* gene that have different hexamers (AAUAAA and AUUAAA) in their specific isoforms.

^d Frequency considering PAS indicated in NCBI for 12 of 17 oncogenes included in this study. Two putative oncogenes were included in this group (*PALLD* and *SHOC2*).

^e Frequency considering PAS indicated in NCBI for 69 of 81 tumor suppressor genes (TSG) included in this study. Eight putative TSG were included in this group (see Materials and Methods).

^f $p=0.78$ (Fisher's exact test); comparison of the frequency of most common hexamers functioning as PAS (AAUAAA and AUUAAA) between oncogenes and tumor suppressor genes.

An integrative analysis of the data obtained through the NCBI, APADB and APASdb databases allowed to identify 105 CPGs (~90%) with evidences of alternative polyadenylation among their transcript variants, using the criterion that more than one functional PAS/CS had been indicated by at least one of these databases (based on the results presented in Table 1). However, we need to emphasize that some CPGs included in this study had little or no data reported regarding their CPE sequences (e.g., *FANCB*, *HRAS*, *MEN1* and *TERT*), which made it impossible to evaluate the occurrence of APA sites in such genes. The strongest evidence of APA arose from the *PTEN* transcript which has 61 APA sites differentially used in its processing among different normal human tissues according to APASdb analysis (Fig 1). Although there is experimental evidence for a large number of APA sites in the *PTEN* gene, two CS located in distinct exonic regions (chr10:89726131 and chr10:89726870) showed higher usage quantification (29.4% and 12.3%, respectively). Both CS are associated with PAS containing the functional AUUAAA hexamer (Fig 1) and are expected to be preferentially used by the cleavage/polyadenylation protein complex.

Regarding the 21 CPGs without PAS sequences described in the NCBI database, our computational strategy was able to detect 3'-most hexamers that may function as PAS for 17 of these genes (Table 3). Although we used the same method to screen the full sequence of corresponding transcripts, we were unable to identify putative PAS for the *ERCC4*, *FH*, *MUTYH* and *SHOC2* genes.

Table 3. Putative polyadenylation signals for cancer predisposition genes without identification of this sequence in NCBI database.

Gene	Chr number	RefSeq	Transcript start	Transcript end	PAS sequence ^a	PAS start	PAS end
<i>BRCA2</i>	chr13	NM_000059	32889616	32973809	AATAAA	32973352	32973357
<i>BRCA2</i>	chr13	NM_000059	32889616	32973809	ATTAAA	32973676	32973681
<i>BRIP1</i>	chr17	NM_032043	59756546	59940920	AATAAA	59938913	59938918
<i>BRIP1</i>	chr17	NM_032043	59756546	59940920	ATTAAA	59939856	59939861
<i>CDK4</i>	chr12	NM_000075	58141509	58146230	AGTAAA	58145246	58145251
<i>CDK4</i>	chr12	NM_000075	58141509	58146230	CATAAA	58144382	58144387
<i>ERCC2</i>	chr19	NM_000400	45854648	45873845	ATTAAA	45871095	45871100
<i>ERCC2</i>	chr19	NM_000400	45854648	45873845	AATACA	45871113	45871118
<i>FANCA</i>	chr16	NM_000135	89803958	89883065	AATAAA	89878113	89878118
<i>FANCA</i>	chr16	NM_000135	89803958	89883065	ATTAAA	89882343	89882348
<i>FANCB</i>	chrX	NM_001018113	14861528	14891191	AATAAA	14889174	14889179
<i>FANCB</i>	chrX	NM_001018113	14861528	14891191	TATAAA	14890834	14890839
<i>FANCI</i>	chr15	NM_001113378	89787193	89860362	AATAAA	89860342	89860347
<i>FANCI</i>	chr15	NM_001113378	89787193	89860362	TATAAA	89857626	89857631
<i>HRAS</i>	chr11	NM_001130442	532241	535567	AGTAAA	532259	532264
<i>HRAS</i>	chr11	NM_001130442	532241	535567	AATATA	534310	534315
<i>KRAS</i>	chr12	NM_004985	25357722	25403865	AATAAA	25401738	25401743
<i>KRAS</i>	chr12	NM_004985	25357722	25403865	TATAAA	25403182	25403187
<i>MAP2K2</i>	chr19	NM_030662	4090319	4124126	AATAAA	4118828	4118833
<i>MAP2K2</i>	chr19	NM_030662	4090319	4124126	ATTAAA	4114316	4114321
<i>MEN1</i>	chr11	NM_130800	64570985	64578035	TATAAA	64572816	64572821
<i>MEN1</i>	chr11	NM_130804	64570985	64578766	AGTAAA	64575874	64575879
<i>PMS2</i>	chr7	NM_000535	6010555	6048737	AATAAA	6047840	6047845
<i>PMS2</i>	chr7	NM_000535	6010555	6048737	ATTAAA	6047310	6047315
<i>POLE</i>	chr12	NM_006231	133200347	133264110	AATAAA	133257497	133257502
<i>POLE</i>	chr12	NM_006231	133200347	133264110	TATAAA	133256791	133256796
<i>POLK</i>	chr5	NM_016218	74807656	74895646	AATAAA	74895618	74895623
<i>POLK</i>	chr5	NM_016218	74807656	74895646	ATTAAA	74894636	74894641
<i>PTCH1</i>	chr9	NM_001083607	98205263	98269481	ATTAAA	98268733	98268738
<i>PTCH1</i>	chr9	NM_001083607	98205263	98269481	TATAAA	98265099	98265104
<i>TERT</i>	chr5	NM_001193376	1253286	1295162	AATAAA	1269949	1269954
<i>TERT</i>	chr5	NM_001193376	1253286	1295162	AGTAAA	1274131	1274136
<i>TMEM127</i>	chr2	NM_001193304	96915945	96931751	ATTAAA	96929713	96929718
<i>TMEM127</i>	chr2	NM_001193304	96915945	96931751	TATAAA	96929395	96929400

Chr Number, chromosome number; PAS, polyadenylation signal.

^a A computational method was used to screen 3'-most canonical AAUAAA hexamers and its human functional variants in the full sequence of corresponding transcripts (RefSeq sequences).

In addition to investigating the distribution of PAS hexamer types, we also assessed the frequency of different dinucleotide sequences identified as cleavage sites (CS) in CPGs (Table 4). It is well known that most pre-mRNAs in eukaryotes are preferentially cleaved immediately 3' from a “CA” dinucleotide [5, 40], but “AA” was the most common sequence used by this specific set of genes. Surprisingly, the “CA” dinucleotide was only the fourth most frequent CS in our analysis.

Table 4. Frequency of dinucleotide sequences identified as cleavage sites in cancer predisposition genes analyzed in the current study.

Dinucleotide sequence	Frequency in transcripts analyzed in this study (%) ^a
AA	37.50
TA	17.05
GA	14.20
CA	13.06
TC	3.98
TT	3.98
TG	2.85
AT	1.71
GC	1.71
GG	1.14
GT	1.14
AC	0.56
AG	0.56
CC	0.56

^a Frequency calculated considering the cleavage sites (CS) indicated in NCBI for 107 of 117 genes included in this study (the remaining genes have no description of CS in this database). Each dinucleotide was counted as a unit, including the case of genes with more than one CS sequence and transcripts derived from *ERCC2*, *RET* and *TSC2* genes that have different CS sequences in their specific isoforms.

Finally, using experimentally validated and computationally predicted data sources, we investigated the miRNAs/miRNA families that modulate expression of a subset of CPGs included in this study: 81 tumor suppressor genes and 17 oncogenes. First, we identified a set of 1396 experimentally validated miRNAs and a set of 184 miRNAs predicted to be regulators of both gene groups, and explored the number of intersections among these sets of miRNAs (S1 Fig). Considering the filtering criteria employed in these *in silico* analyses, only the *FANCB* gene had no reported regulating miRNAs. We then performed an “over-representation analysis” in order to identify miRNAs with potentially oncogenic or tumor suppressor function. Fig 2 shows the regulatory networks obtained by combined analyses of validated and predicted data indicating miRNAs and miRNA families significantly overrepresented ($P < 0.01$) as regulators of expression between the groups of tumor suppressors and oncogenes. As seen in these interactomes, most statistically significant data were derived from experimental sources (Figs 2A and 2B), and the experimental data represented the majority of collected miRNA-target gene interactions compared to the data only predicted (Figs 2C and 2D).

Lastly, we observed that the miR-192 family was the most significantly overrepresented among tumor suppressor genes ($P = 0.002$), which may suggest an oncogenic function. This miRNA family has experimentally validated binding sites to the 3'UTR of 20 tumor suppressor genes included in the current study (Fig 2B). Among them, there are central genes in human DNA repair pathways such as *BRCA1*, *BRCA2*, *MSH6*, *RAD51* and *XPA*. Although the oncogenes group was smaller, a larger number of miRNA families were strongly associated with these genes: miR-128, miR-1471, miR-483, miR-3170 and miR-218. An additional result, important to validate our analysis, relates to miR-34a-5p, a specific miRNA already known to have tumor suppressor function [41, 42]. This miRNA family was significantly overrepresented as regulator of oncogenes ($P = 0.005$), and in our analysis we observed validated and predicted interactions with 8 of the 17 oncogenes studied (Fig 2A).

Discussion

Recent advances have shown that the eukaryotic cleavage and polyadenylation mechanism is regulated via a network of *cis*-acting RNA sequence elements (herein termed CPE) located in the pre-mRNA 3'UTR and *trans*-acting proteins, contributing to the qualitative/quantitative adjustment of gene expression. The arrangement of these CPE determines the efficiency of a given polyadenylation site [2, 16]. The most prominent CPE is the PAS, a hexameric sequence motif located 10–30 nt upstream of the CS that was first described by Proudfoot and Brownlee (1976) [3]. The PAS serves as the binding site for CPSF, the endonuclease responsible for the pre-mRNA cleavage [8]. Previous studies indicated that in ~55% of human mRNAs, the PAS hexamer consists of the consensus sequence AAUAAA [12, 30], recognized as one of the most highly conserved sequence elements known [3, 11, 43, 44]. Most of the remaining ~45% of mRNAs that do not have an exact match to the consensus differ by only a single substitution. An A → U conversion at the second position is the most common variant (AUUAAA) [12, 45].

In this regard, recent studies identified functional germline variants located within or in the vicinity of CPE sequences in two CPGs [17,19], raising the need to characterize these elements. In the present study, we evaluated the frequency and sequence of PAS hexamers and CS dinucleotides in a set of CPGs, as well as the distance (in bp) between them. In agreement with few previous genomic-scale data [12, 30], our analysis revealed that the majority of PAS contained the canonical hexamer AAUAAA, and the AUUAAA variant was the second most frequent. However, we found the “AA” dinucleotide in most CS sequences associated with CPGs. Although the nucleotide sequence of the exact CS is not highly conserved [46], most pre-mRNAs are cleaved 3' of an adenosine residue (in agreement with our data) and “CA” was defined as the optimal CS [5, 47]. Interestingly, “CA” dinucleotide was only the fourth most frequent CS in our gene set. Since more than 50% of human protein coding genes harbor multiple mRNA CS [12], it would appear that a “CA” dinucleotide cannot be an absolute requirement for correct cleavage, a situation analogous to the use of both the canonical AAUAAA PAS and its variants by human

genes. A functional study supporting this hypothesis demonstrated that the “CA” dinucleotide is preferred but not required for cleavage machinery recognition, and CS usage at position –1 was found to be in the order of preference A > U > C >> G [5]. This same study and other previous analyses [3, 4] indicated that the CS is located no closer than 10 bases, but no further than 30 bases from the AAUAAA element, while about 11% of CPGs included in our study exhibited a PAS-CS distance greater than 30bp. Thus, we can speculate that certain estimates provided by long-standing polyadenylation studies do not apply to all human transcripts. Nonetheless, the results of the current study must be viewed in the context of two main limitations: a characterization targeting a specific group of genes, and a relatively small number of genes analyzed.

Strikingly, we did not find reports of a PAS hexamer in about 18% of the CPGs using a reference database (NCBI). For these genes, a computational analysis was developed in this study to identify 3'-most hexamers (putative PAS) in the full corresponding mRNA sequences, because the predominant form of an mRNA sequence is usually the longest one, generated by the 3'-most poly(A) site [12]. Although our analysis has not identified putative PAS for all genes, these findings reinforce the relevance of establishing updated methods and/or databases to detect this regulatory element of 3' end processing in human genes. The strategy used here could be easily applied to similar situations with additional genes outside the CPG context.

In addition, we also explored the frequency of functional APA sites among the CPGs studied. Indeed, APA has only recently appeared as a major player in gene regulation [13], and a widespread APA-mediated 3' UTR shortening has been identified across the cancer genome [48-50]. Since shorter 3' UTR isoforms have higher translational efficiency than their respective long isoforms due to loss of miRNA regulation, APA events can activate some proto-oncogenes in cancer cells [48, 51]. Our analysis using recently released databases indicated that approximately 90% of the selected CPGs have two or more APA sites. In contrast, a previous analysis performed by Tian and colleagues (2005) estimated that about 54% of the human genes are alternatively polyadenylated [12], and a recent genome-wide map of APA sites in major cancers and tumor cell lines identified around 30% of mRNAs containing APA sites, independent of the cell type [49]. Taken together, these

findings suggest a greater complexity in the regulation of polyadenylation in transcripts specifically derived from CPGs. Besides that, APA patterns in mammals seem to be evolutionarily conserved [52] and tissue-specifically modulated [53, 54]. In this sense, here we reported the strong evidence of APA modulation to the *PTEN* tumor suppressor gene: it contains 61 APA sites differentially used in 22 distinct non-tumoral human tissues analyzed by APASdb [29]. For instance, two of these APA sites are preferentially used in *PTEN* mRNA processing, but their usage quantification was lower in specific tissues, such as kidney and spleen (data not shown). Therefore, a closer understanding of APA as a mechanism for transcript diversity and gene regulation in cancer development is still warranted.

Finally, given the scarcity of studies about regulation by miRNAs shared between comprehensive sets of tumor suppressor genes and oncogenes acting on different cell signaling pathways, we investigated whether validated and predicted data sources could identify miRNAs/miRNA families with potential tumor suppressor and oncogenic functions. The miR-192 family was significantly overrepresented as regulator of tumor suppressor genes, having experimentally validated interactions with genes implicated in essential DNA repair pathways, including double-strand break repair by homologous recombination (*BRCA1*, *BRCA2* and *RAD51*), mismatch repair (*MSH6*) and nucleotide excision repair induced by ultraviolet light (*XPA*). An upregulation of this miRNA family was reported in multiple tumor types including gastric cancer, hepatocellular carcinoma, neuroblastoma, pancreatic ductal adenocarcinoma and esophageal squamous cell carcinoma [55-59]. The biological effects of miR-192 in these cancers have been partially elucidated. Initial evidences suggest that it could enhance cell proliferation and migration, reduce apoptosis and promote cell cycle progression [55, 56, 59]. These findings support our hypothesis that miRNAs within this family may function as tumor oncogenes. However, miR-192 was found downregulated in colon, colorectal and lung tumors [60-62]. We also identified some miRNA families that might act as effectors in tumor suppression, including miR-128, miR-1471, miR-483, miR-3170 and miR-218. Up to now, little is known about the role of miR-1471 and miR-3170 in cancer, but recent reports corroborate the function assigned to other miRNA families: miR-128 exerts pro-apoptotic effects and was found as a

tumor suppressor in certain tumor types [63-66]; miR-483 suppresses the proliferation of glioma and squamous cell carcinoma cells [67-68]; and emerging tumor-suppressing roles has been described for miR-218 in prostate, breast and lung cancers [69-71]. Moreover, miR-34a is known tumor-suppressive miRNA and was found to be a regulator of several of the oncogenes included here, validating our *in silico* analysis. Importantly, miR-34a has a positive effect on p53 transcriptional activity and protein stability (encoding by tumor suppressor gene *TP53*), by targeting multiple p53 inhibitor genes (e.g., the *MDM4* oncogene). In turn, p53 up-regulates this miRNA, contributing to apoptosis and senescence [41, 42]. In fact, *TP53* and *MDM4* genes appeared as direct targets of miR-34a in our analysis. Overall, our results using this computational analysis are in agreement with the data already available in the literature, suggesting that this is a suitable strategy to identify potentially tumor-suppressive and oncogenic miRNAs.

In summary, the present study is the first to focus on the comprehensive characterization of 3'UTR-related elements (CPE sequences and regulation by miRNAs) in CPGs. This approach generated a landscape of CPE in CPGs which may be useful in the development of molecular analyses covering these frequently neglected 3'UTR regulatory elements. Nevertheless, further studies including additional cancer-related genes should be conducted in order to confirm these findings. Additionally, functional validation of the potential oncogenic and tumor suppressor functions assigned to these miRNA families from our *in silico* analysis is warranted.

Acknowledgments

We are grateful to Clévia Rosset, Isabel Bandeira da Silva and Bárbara Alemar for their valuable contributions and support. Igor Araujo Vieira and Delva Pereira Leão are recipients of postgraduate fellowships from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). Patricia Ashton-Prolla and Sandro José de Souza are investigators of CNPq.

Author Contributions

Conceived and designed the analyses: IAV MRM VLL DPL SJS PAP.
Performed the analyses: IAV MRM VLL MRS. Analyzed the data: IAV MRM VLL
DPL MRS SJS PAP. Wrote the paper: IAV MRM VLL SJS PAP.

References

1. Zhao J, Hyman L, Moore C. Formation of mRNA 3' ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis. *Microbiol Mol Biol Rev.* 1999;63: 405-445.
2. Millevoi S, Vagner S. Molecular mechanisms of eukaryotic pre-mRNA 3' end processing regulation. *Nucleic Acids Res.* 2010;38: 2757-2774.
3. Proudfoot NJ, Brownlee GG. 3' non-coding region sequences in eukaryotic messenger RNA. *Nature* 1976;263: 211-214.
4. Gil A, Proudfoot NJ. Position-dependent sequence elements downstream of AAUAAA are required for efficient rabbit beta-globin mRNA 3' end formation. *Cell* 1987;49: 399-406.
5. Chen F, MacDonald CC, Wilusz J. Cleavage site determinants in the mammalian polyadenylation signal. *Nucleic Acids Res.* 1995;23: 2614-2620.
6. Legendre M, Gautheret D. Sequence determinants in human polyadenylation site selection. *BMC Genomics* 2003;4: 7.
7. Matoulkova E, Michalova E, Vojtesek B, Hrstka R. The role of the 3' untranslated region in post-transcriptional regulation of protein expression in mammalian cells. *RNA Biol.* 2012;9: 563-576.
8. Ryan K, Calvo O, Manley JL. Evidence that polyadenylation factor CPSF-73 is the mRNA 3' processing endonuclease. *RNA* 2004;10: 565-573.

9. Keller W, Bienroth S, Lang KM, Christofori G. Cleavage and polyadenylation factor CPF specifically interacts with the pre-mRNA 3' processing signal AAUAAA. *EMBO J.* 1991;10: 4241-4249.
10. MacDonald CC, Wilusz J, Shenk T. The 64-kilodalton subunit of the CstF polyadenylation factor binds to pre-mRNAs downstream of the cleavage site and influences cleavage site location. *Mol Cell Biol.* 1994;14: 6647-6654.
11. Proudfoot NJ. Ending the message: poly(A) signals then and now. *Genes Dev.* 2011;25: 1770-1782.
12. 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: 201-212.
13. Lutz CS. Alternative polyadenylation: a twist on mRNA 3' end formation. *ACS Chem Biol.* 2008;3: 609-617.
14. Chen JM, Férec C, Cooper DN. A systematic analysis of disease-associated variants in the 3' regulatory regions of human protein-coding genes I: general principles and overview. *Hum Genet.* 2006;120: 1-21.
15. Michalova E, Vojtesek B, Hrstka R. Impaired pre-mRNA processing and altered architecture of 3' untranslated regions contribute to the development of human disorders. *Int J Mol Sci.* 2013;14: 15681-15694.
16. Hollerer I, Grund K, Hentze MW, Kulozik AE. mRNA 3'end processing: A tale of the tail reaches the clinic. *EMBO Mol Med.* 2014;6: 16-26.
17. Stacey SN, Sulem P, Jonasdottir A, Masson G, Gudmundsson J, Gudbjartsson DF, et al. A germline variant in the TP53 polyadenylation signal confers cancer susceptibility. *Nat Genet.* 2011;43: 1098-10103.

18. Macedo GS, Araujo Vieira I, Brandalize AP, Giacomazzi J, Inez Palmero E, Volc S, et al. Rare germline variant (rs78378222) in the TP53 3' UTR: Evidence for a new mechanism of cancer predisposition in Li-Fraumeni syndrome. *Cancer Genet.* 2016;209: 97-106.
19. Decorsière A, Toulas C, Fouque F, Tilkin-Mariamé AF, Selves J, Guimbaud R, et al. Decreased efficiency of MSH6 mRNA polyadenylation linked to a 20-base-pair duplication in Lynch syndrome families. *Cell Cycle* 2012;11: 2578-2580.
20. Ambros V. The functions of animal microRNAs. *Nature* 2004;431: 350-355.
21. Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 2004;116: 281-297.
22. Calin GA, Dumitru CD, Shimizu M, Bichi R, Zupo S, Noch E, et al. Frequent deletions and down-regulation of micro- RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. *Proc Natl Acad Sci U S A* 2002;99: 15524-15529.
23. Takamizawa J, Konishi H, Yanagisawa K, Tomida S, Osada H, Endoh H, et al. Reduced expression of the let-7 microRNAs in human lung cancers in association with shortened postoperative survival. *Cancer Res.* 2004;64: 3753-3756.
24. Cimmino A, Calin GA, Fabbri M, Iorio MV, Ferracin M, Shimizu M, et al. miR-15 and miR-16 induce apoptosis by targeting BCL2. *Proc Natl Acad Sci U S A* 2005;102: 13944-13949.
25. Esquela-Kerscher A, Slack FJ. Oncomirs - microRNAs with a role in cancer. *Nat Rev Cancer* 2006;6: 259-269.
26. Garber JE, Offit K. Hereditary cancer predisposition syndromes. *J Clin Oncol.* 2005;23: 276-292.

27. Hanahan D, Weinberg RA. The hallmarks of cancer. *Cell* 2000;100: 57-70.
28. Müller S, Rycak L, Afonso-Grunz F, Winter P, Zawada AM, Damrath E, et al. APADB: a database for alternative polyadenylation and microRNA regulation events. *Database* 2014;pii: bau076.
29. You L, Wu J, Feng Y, Fu Y, Guo Y, Long L, et al. APASdb: a database describing alternative poly(A) sites and selection of heterogeneous cleavage sites downstream of poly(A) signals. *Nucleic Acids Res.* 2015;43: D59-67.
30. Beaulieu E, Freier S, Wyatt JR, Claverie JM, Gautheret D. Patterns of variant polyadenylation signal usage in human genes. *Genome Res.* 2000;10: 1001-1010.
31. Sun Y, Fu Y, Li Y, Xu A. Genome-wide alternative polyadenylation in animals: insights from high-throughput technologies. *J Mol Cell Biol.* 2012;4: 352-361.
32. Chou CH, Chang NW, Shrestha S, Hsu SD, Lin YL, Lee WH, et al. miRTarBase 2016: updates to the experimentally validated miRNA-target interactions database. *Nucleic Acids Res.* 2016;44: D239-247.
33. Li JH, Liu S, Zhou H, Qu LH, Yang JH. starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Res.* 2014;42: D92-97.
34. Sethupathy P, Corda B, Hatzigeorgiou AG. TarBase: A comprehensive database of experimentally supported animal microRNA targets. *RNA* 2016;12: 192-197.
35. Xiao F, Zuo Z, Cai G, Kang S, Gao X, Li T. miRecords: an integrated resource for microRNA-target interactions. *Nucleic Acids Res.* 2009;37: D105-110.

36. Agarwal V, Bell GW, Nam J, Bartel DP. Predicting effective microRNA target sites in mammalian mRNAs. *eLife* 2015;4: e05005.
37. Paraskevopoulou MD, Georgakilas G, Kostoulas N, Vlachos IS, Vergoulis T, Reczko M, et al. DIANAmicroT web server v5.0: service integration into miRNA functional analysis workflows. *Nucleic Acids Res.* 2013;41: W169-173.
38. Betel D, Koppal A, Agius P, Sander C, Leslie C. Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites. *Genome Biol.* 2010;11: R90.
39. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 2003;13: 2498-2504.
40. Danckwardt S, Hentze MW, Kulozik AE. 3' end mRNA processing: molecular mechanisms and implications for health and disease. *EMBO J.* 2008;27: 482-498.
41. Raver-Shapira N, Marciano E, Meiri E, Spector Y, Rosenfeld N, Moskovits N, et al. Transcriptional activation of miR-34a contributes to p53-mediated apoptosis. *Mol Cell.* 2007;26: 731-743.
42. Navarro F, Lieberman J. miR-34 and p53: New Insights into a Complex Functional Relationship. *PLoS One* 2015;10: e0132767.
43. Proudfoot N. Poly(A) signals. *Cell* 1991;64: 671-674.
44. Wickens M, Stephenson P. Role of the conserved AAUAAA sequence: four AAUAAA point mutants prevent messenger RNA 3' end formation. *Science* 1984;226: 1045-1051.
45. MacDonald CC, Redondo JL. Reexamining the polyadenylation signal: were we wrong about AAUAAA? *Mol Cell Endocrinol.* 2002;190: 1-8.

46. Sheets MD, Ogg SC, Wickens MP. Point mutations in AAUAAA and the poly (A) addition site: effects on the accuracy and efficiency of cleavage and polyadenylation in vitro. *Nucleic Acids Res.* 1990;18: 5799-5805.
47. Gehring NH, Frede U, Neu-Yilik G, Hundsdorfer P, Vetter B, Hentze MW, et al. Increased efficiency of mRNA 3' end formation: a new genetic mechanism contributing to hereditary thrombophilia. *Nat Genet.* 2001;28: 389-392.
48. Mayr C, Bartel DP. Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. *Cell* 2009;138: 673-684.
49. Lin Y, Li Z, Ozsolak F, Kim SW, Arango-Argoty G, Liu TT, et al. An in-depth map of polyadenylation sites in cancer. *Nucleic Acids Res.* 2012;40: 8460-8471.
50. Lai DP, Tan S, Kang YN, Wu J, Ooi HS, Chen J, et al. Genome-wide profiling of polyadenylation sites reveals a link between selective polyadenylation and cancer metastasis. *Hum Mol Genet.* 2015;24: 3410-3417.
51. An J, Zhu X, Wang H, Jin X. A dynamic interplay between alternative polyadenylation and microRNA regulation: implications for cancer (Review). *Int J Oncol.* 2013;43: 995-1001.
52. Ara T, Lopez F, Ritchie W, Benech P, Gautheret D. Conservation of alternative polyadenylation patterns in mammalian genes. *BMC Genomics* 2006;7: 189.
53. Beaudoin E, Gautheret D. Identification of alternate polyadenylation sites and analysis of their tissue distribution using EST data. *Genome Res.* 2001;11: 1520–1526.
54. Zhang H, Lee JY, Tian B. Biased alternative polyadenylation in human tissues. *Genome Biol.* 2005;6: R100.

55. Feinberg-Gorenshtein G, Guedj A, Shichrur K, Jeison M, Luria D, Kodman Y, et al. MiR-192 directly binds and regulates Dicer1 expression in neuroblastoma. *PLoS One* 2013;8: e78713.
56. Zhao C, Zhang J, Zhang S, Yu D, Chen Y, Liu Q, et al. Diagnostic and biological significance of microRNA-192 in pancreatic ductal adenocarcinoma. *Oncol Rep.* 2013;30: 276-284.
57. Chen Q, Ge X, Zhang Y, Xia H, Yuan D, Tang Q, et al. Plasma miR-122 and miR-192 as potential novel biomarkers for the early detection of distant metastasis of gastric cancer. *Oncol Rep.* 2014;31: 1863-1870.
58. Tan Y, Ge G, Pan T, Wen D, Chen L, Yu X, et al. A serum microRNA panel as potential biomarkers for hepatocellular carcinoma related with hepatitis B virus. *PLoS One* 2014;9: e107986.
59. Li S, Li F, Niu R, Zhang H, Cui A, An W, et al. Mir-192 suppresses apoptosis and promotes proliferation in esophageal squamous cell carcinoma by targeting Bim. *Int J Clin Exp Pathol.* 2015;8: 8048-8056.
60. Feng S, Cong S, Zhang X, Bao X, Wang W, Li H, et al. MicroRNA-192 targeting retinoblastoma 1 inhibits cell proliferation and induces cell apoptosis in lung cancer cells. *Nucleic Acids Res.* 2011;39: 6669-6678.
61. Karaayvaz M, Pal T, Song B, Zhang C, Georgakopoulos P, Mehmood S, et al. Prognostic significance of miR-215 in colon cancer. *Clin Colorectal Cancer* 2011;10: 340-347.
62. Chiang Y, Song Y, Wang Z, Liu Z, Gao P, Liang J, et al. microRNA-192, -194 and -215 are frequently downregulated in colorectal cancer. *Exp Ther Med.* 2012;3: 560-566.

63. Adlakha YK, Saini N. miR-128 exerts pro-apoptotic effect in a p53 transcription-dependent and -independent manner via PUMA-Bak axis. *Cell Death Dis.* 2013;4: e542.
64. Hauser B, Zhao Y, Pang X, Ling Z, Myers E, Wang P, et al. Functions of MiRNA-128 on the regulation of head and neck squamous cell carcinoma growth and apoptosis. *PLoS One* 2015;10: e0116321.
65. Wu L, Shi B, Huang K, Fan G. MicroRNA-128 suppresses cell growth and metastasis in colorectal carcinoma by targeting IRS1. *Oncol Rep.* 2015;34: 2797-2805.
66. Shan ZN, Tian R, Zhang M, Gui ZH, Wu J, Ding M, et al. miR128-1 inhibits the growth of glioblastoma multiforme and glioma stem-like cells via targeting BMI1 and E2F3. *Oncotarget* 2016;7: 78813-78826.
67. Wang L, Shi M, Hou S, Ding B, Liu L, Ji X, et al. MiR-483-5p suppresses the proliferation of glioma cells via directly targeting ERK1. *FEBS Lett.* 2012;586: 1312-1317.
68. Bertero T, Bourget-Ponzio I, Puissant A, Loubat A, Mari B, Meneguzzi G, et al. Tumor suppressor function of miR-483-3p on squamous cell carcinomas due to its pro-apoptotic properties. *Cell Cycle* 2013;12: 2183-2193.
69. Liu B, Tian Y, Li F, Zhao Z, Jiang X, Zhai C, et al. Tumor-suppressing roles of miR-214 and miR-218 in breast cancer. *Oncol Rep.* 2016;35: 3178-3184.
70. Song L, Li D, Zhao Y, Gu Y, Zhao D, Li X, et al. miR-218 suppressed the growth of lung carcinoma by reducing MEF2D expression. *Tumour Biol.* 2016;37: 2891-2900.
71. Guan B, Wu K, Zeng J, Xu S, Mu L, Gao Y, et al. Tumor-suppressive microRNA-218 inhibits tumor angiogenesis via targeting the mTOR component RICTOR in prostate cancer. *Oncotarget* 2017;8: 8162-8172.

Figures

Fig 1. APASdb data showing alternative polyadenylation/cleavage sites (CS) and polyadenylation signals (PAS) mapped to the transcript derived from *PTEN* gene, containing the 5' and 3' flanking region of 1kb. This database identified 61 differentially used alternative CS in the *PTEN* mRNA processing among 22 normal human tissues (Mean reads: 75, searching dataset: hg19 human-all22-tissues). Chromosome position, PAS sequences and usage quantification (%) for each CS are indicated.

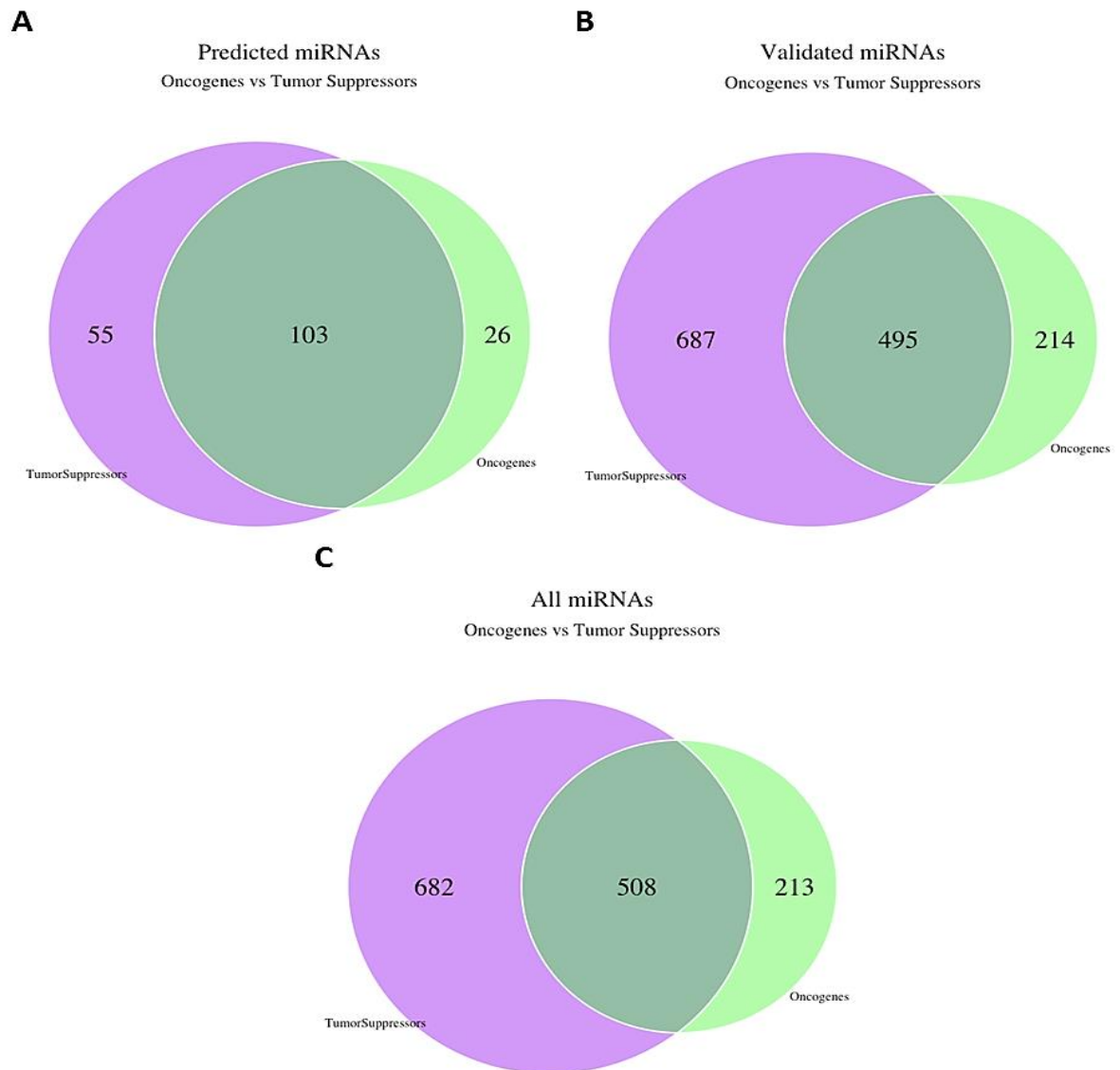
> uc021pww.1; *PTEN*, Homo sapiens phosphatase and tensin homolog (*PTEN*), mRNA.



Fig 2. Regulatory networks and Venn diagrams obtained by combining validated and predicted interactions between miRNAs and tumor suppressor genes/oncogenes. miRNAs (A) and miRNA families (B) significantly overrepresented (P -value <0.01) as regulators (diamond nodes) of expression between the groups of tumor suppressor genes (green circle nodes) and oncogenes studied (red circle nodes) are shown in these interactomes. Red-outlined diamonds indicate miRNAs with potential tumor suppressor function, whereas the green-outlined ones represent potentially oncogenic miRNAs. Solid lines denote experimentally validated interactions and dashed lines are high confidence computationally predicted interactions. Interactome (A) was built considering only miRNAs with at least three target genes among tumor suppressor genes and oncogenes. Intersections among the sets of validated and predicted miRNAs found to modulate expression of tumor suppressor genes (C) and oncogenes (D) are indicated in Venn diagrams.

Supporting Information

S1 Fig. Venn diagrams showing the number of intersections among the sets of predicted (A) and validated miRNAs (B) involved in the regulation of oncogenes and tumor suppressor genes included in this study, as well as for the union of validated and predicted data (C).



S1 Table. Selection criteria and classification of cancer predisposition genes included in the current study.

Gene	Selection criteria (number) ^a	Main reference	Classification (ONC, TSG or UND) ^b	Classification-related reference ^c
<i>APC</i>	1	Garber and Offit, 2005 (PMID: 15637391)	TSG	Vogelstein <i>et al.</i> , 2013 (PMID: 23539594)
<i>ATM</i>	2	Garber and Offit, 2005	TSG	Vogelstein <i>et al.</i> , 2013
<i>BAP1</i>	3	Testa <i>et al.</i> , 2011 (PMID: 21874000)	TSG	Vogelstein <i>et al.</i> , 2013
<i>BARD1</i>	1	Castéra <i>et al.</i> , 2014 (PMID: 24549055)	TSG	Karppinen <i>et al.</i> , 2004 (PMID: 15342711)
<i>BLM</i>	2	Garber and Offit, 2005	TSG	de Voer <i>et al.</i> , 2015 (PMID: 26358404)
<i>BMPR1A</i>	1	Garber and Offit, 2005	TSG	Delnatte <i>et al.</i> , 2006 (PMID: 16685657)
<i>BRAF</i>	2	Niihori <i>et al.</i> , 2006 (PMID: 16474404)	ONC	Vogelstein <i>et al.</i> , 2013
<i>BRCA1</i>	1	Garber and Offit, 2005	TSG	Vogelstein <i>et al.</i> , 2013
<i>BRCA2</i>	1	Garber and Offit, 2005	TSG	Vogelstein <i>et al.</i> , 2013
<i>BRIP1</i>	2	Levitus <i>et al.</i> , 2005 (PMID: 16116423)	TSG	Ren <i>et al.</i> , 2013 (PMID: 24301948)
<i>CDC73</i>	1	Zhang <i>et al.</i> , 2010 (PMID: 20339377)	TSG	Vogelstein <i>et al.</i> , 2013
<i>CDH1</i>	1	Garber and Offit, 2005	TSG	Vogelstein <i>et al.</i> , 2013
<i>CDKN1C</i>	2	Garber and Offit, 2005	TSG	Matsuoka <i>et al.</i> , 1995 (PMID: 7729684)
<i>CDKN2A</i>	1	Garber and Offit, 2005	TSG	Vogelstein <i>et al.</i> , 2013
<i>CDK4</i>	1	Garber and Offit, 2005	ONC	Zuo <i>et al.</i> , 1996 (PMID: 8528263)
<i>CHEK2</i>	1	Garber and Offit, 2005	TSG	Falck <i>et al.</i> , 2001 (PMID: 11298456)
<i>DDB2</i>	1	Garber and Offit, 2005	TSG	Kraemer and DiGiovanna, 2003 (PMID: 20301571)
<i>ELAC2</i>	1	Garber and Offit, 2005	TSG	Korver <i>et al.</i> , 2003 (PMID: 12569551)
<i>EPCAM</i>	1	Tuttlewska <i>et al.</i> , 2013 (PMID: 23938213)	UND	-
<i>EPHB2</i>	3	Huusko <i>et al.</i> , 2004 (PMID: 15300251)	TSG	Huusko <i>et al.</i> , 2004 (PMID: 15300251)
<i>ERCC2</i>	1	Garber and Offit, 2005	TSG	Kraemer and DiGiovanna, 2003 (PMID: 20301571)
<i>ERCC3</i>	1	Garber and Offit, 2005	TSG	Kraemer and DiGiovanna, 2003 (PMID: 20301571)
<i>ERCC4</i>	1	Garber and Offit, 2005	TSG	Kraemer and DiGiovanna, 2003 (PMID: 20301571)
<i>ERCC5</i>	1	Garber and Offit, 2005	TSG	Kraemer and DiGiovanna, 2003 (PMID: 20301571)
<i>EXT1</i>	1	Garber and Offit, 2005	TSG	Sarrión <i>et al.</i> , 2013 (PMID: 23439489)
<i>EXT2</i>	1	Garber and Offit, 2005	TSG	Sarrión <i>et al.</i> , 2013 (PMID: 23439489)
<i>FANCA</i>	2	Garber and Offit, 2005	TSG	Mehta and Tolar, 2002 (PMID: 20301575)
<i>FANCB</i>	2	Garber and Offit, 2005	TSG	Mehta and Tolar, 2002 (PMID: 20301575)
<i>FANCC</i>	2	Garber and Offit, 2005	TSG	Mehta and Tolar, 2002 (PMID: 20301575)
<i>FANCD2</i>	2	Garber and Offit, 2005	TSG	Mehta and Tolar, 2002 (PMID: 20301575)

Gene	Selection criteria (number) ^a	Main reference	Classification (ONC, TSG or UND) ^b	Classification-related reference ^c
<i>FANCE</i>	2	Garber and Offit, 2005	TSG	Mehta and Tolar, 2002 (PMID: 20301575)
<i>FANCF</i>	2	Garber and Offit, 2005	TSG	Mehta and Tolar, 2002 (PMID: 20301575)
<i>FANCG</i>	2	Garber and Offit, 2005	TSG	Mehta and Tolar, 2002 (PMID: 20301575)
<i>FANCI</i>	2	Garber and Offit, 2005	TSG	Mehta and Tolar, 2002 (PMID: 20301575)
<i>FANCL</i>	2	Garber and Offit, 2005	TSG	Mehta and Tolar, 2002 (PMID: 20301575)
<i>FAS</i>	1	Garber and Offit, 2005	UND	-
<i>FASLG</i>	1	Garber and Offit, 2005	UND	-
<i>FH</i>	1	Garber and Offit, 2005	TSG	Toro <i>et al.</i> , 2003 (PMID: 12772087)
<i>FLCN</i>	2	Garber and Offit, 2005	TSG	Schmidt and Linehan, 2015 (PMID: 26334087)
<i>GPC3</i>	2	Garber and Offit, 2005	TSG	Veugelers <i>et al.</i> , 2000 (PMID: 10814714)
<i>GREM1</i>	1	Davis <i>et al.</i> , 2015 (PMID: 25419707)	ONC	Davis <i>et al.</i> , 2015 (PMID: 25419707)
<i>HRAS</i>	2	Denayer <i>et al.</i> , 2008 (PMID: 17979197)	ONC	Vogelstein <i>et al.</i> , 2013
<i>IL2RG</i>	2	Garber and Offit, 2005	UND	-
<i>KIT</i>	1	Garber and Offit, 2005	ONC	Vogelstein <i>et al.</i> , 2013
<i>KRAS</i>	2	Niihori <i>et al.</i> , 2006 (PMID: 16474404)	ONC	Vogelstein <i>et al.</i> , 2013
<i>LIG4</i>	2	O'Driscoll <i>et al.</i> , 2001 (PMID: 11779494)	TSG	O'Driscoll <i>et al.</i> , 2001 (PMID: 11779494)
<i>MAP2K1</i>	2	Schulz <i>et al.</i> , 2008 (PMID: 18042262)	ONC	Vogelstein <i>et al.</i> , 2013
<i>MAP2K2</i>	2	Schulz <i>et al.</i> , 2008 (PMID: 18042262)	ONC	Vogelstein <i>et al.</i> , 2013
<i>MAX</i>	1	Burnichon <i>et al.</i> , 2012 (PMID: 22452945)	TSG	Burnichon <i>et al.</i> , 2012 (PMID: 22452945)
<i>MC1R</i>	1	Perez Oliva <i>et al.</i> , 2009 (PMID: 19338054)	UND	-
<i>MDM2</i>	1	Basu and Murphy, 2016 (PMID: 27037420)	ONC	Momand <i>et al.</i> , 1992 (PMID: 1535557)
<i>MDM4</i>	1	Basu and Murphy, 2016 (PMID: 27037420)	ONC	Paranti <i>et al.</i> , 2001 (PMID: 11528400)
<i>MEN1</i>	1	Garber and Offit, 2005	TSG	Vogelstein <i>et al.</i> , 2013
<i>MET</i>	1	Garber and Offit, 2005	ONC	Vogelstein <i>et al.</i> , 2013
<i>MITF</i>	3	Yokoyama <i>et al.</i> , 2011 (PMID: 22080950)	ONC	Yokoyama <i>et al.</i> , 2011 (PMID: 22080950)
<i>MLH1</i>	1	Garber and Offit, 2005	TSG	Vogelstein <i>et al.</i> , 2013
<i>MRE11A</i>	2	Stewart <i>et al.</i> , 1999 (PMID: 10612394)	TSG	Lavin, 2007 (PMID: 18066087)
<i>MSH2</i>	1	Garber and Offit, 2005	TSG	Vogelstein <i>et al.</i> , 2013
<i>MSH6</i>	1	Garber and Offit, 2005	TSG	Vogelstein <i>et al.</i> , 2013
<i>MSR1</i>	3	Maier <i>et al.</i> , 2006 (PMID: 16287155)	TSG	Maier <i>et al.</i> , 2006 (PMID: 16287155)
<i>MUTYH</i>	1	Al-Tassan <i>et al.</i> , 2002 (PMID: 11818965)	TSG	Al-Tassan <i>et al.</i> , 2002 (PMID: 11818965)
<i>NBN</i>	2	Garber and Offit, 2005	TSG	Varon <i>et al.</i> , 1998 (PMID: 9590180)

Gene	Selection criteria (number) ^a	Main reference	Classification (ONC, TSG or UND) ^b	Classification-related reference ^c
<i>NF1</i>	1	Garber and Offit, 2005	TSG	Vogelstein <i>et al.</i> , 2013
<i>NF2</i>	1	Garber and Offit, 2005	TSG	Vogelstein <i>et al.</i> , 2013
<i>NSD1</i>	2	Garber and Offit, 2005	UND	-
<i>PALB2</i>	2	Xia <i>et al.</i> , 2007 (PMID: 17200672)	TSG	Xia <i>et al.</i> , 2007 (PMID: 17200672)
<i>PALLD</i>	1	Pogue-Geile <i>et al.</i> , 2006 (PMID: 17194196)	ONC	Pogue-Geile <i>et al.</i> , 2006 (PMID: 17194196)
<i>PMS1</i>	1	Garber and Offit, 2005	TSG	Liu <i>et al.</i> , 1996 (PMID: 8574961)
<i>PMS2</i>	1	Garber and Offit, 2005	TSG	Liu <i>et al.</i> , 1996 (PMID: 8574961)
<i>POLB</i>	3	Sobol, 2012 (PMID: 23144636)	UND	-
<i>POLD1</i>	3	Hoffmann and Cazaux, 2010 (PMID: 20934518)	UND	-
<i>POLE</i>	3	Hoffmann and Cazaux, 2010 (PMID: 20934518)	UND	-
<i>POLH</i>	3	Kraemer and DiGiovanna, 2003 (PMID: 20301571)	UND	-
<i>POLK</i>	3	Hoffmann and Cazaux, 2010 (PMID: 20934518)	UND	-
<i>POLQ</i>	3	Hoffmann and Cazaux, 2010 (PMID: 20934518)	UND	-
<i>POT1</i>	1	Robles-Espinoza <i>et al.</i> , 2014 (PMID: 24686849)	TSG	Robles-Espinoza <i>et al.</i> , 2014 (PMID: 24686849)
<i>PRCC</i>	1	Garber and Offit, 2005	UND	-
<i>PRKAR1A</i>	1	Garber and Offit, 2005	TSG	Groussin <i>et al.</i> , 2002 (PMID: 12424709)
<i>PRSS1</i>	1	Garber and Offit, 2005	UND	-
<i>PTCH1</i>	1	Garber and Offit, 2005	TSG	Vogelstein <i>et al.</i> , 2013
<i>PTEN</i>	1	Garber and Offit, 2005	TSG	Vogelstein <i>et al.</i> , 2013
<i>PTPN11</i>	2	Komatsuzaki <i>et al.</i> , 2010 (PMID: 20882035)	UND	-
<i>RAD50</i>	2	Waltes <i>et al.</i> , 2009 (PMID: 19409520)	TSG	Waltes <i>et al.</i> , 2009 (PMID: 19409520)
<i>RAD51</i>	2	Garber and Offit, 2005	TSG	Mehta and Tolar, 2002 (PMID: 20301575)
<i>RAD51C</i>	1	Meindl <i>et al.</i> , 2010 (PMID: 20400964)	TSG	Meindl <i>et al.</i> , 2010 (PMID: 20400964)
<i>RAD51D</i>	1	Loveday <i>et al.</i> , 2011 (PMID: 21822267)	TSG	Loveday <i>et al.</i> , 2011 (PMID: 21822267)
<i>RAF1</i>	2	Komatsuzaki <i>et al.</i> , 2010 (PMID: 20882035)	ONC	Komatsuzaki <i>et al.</i> , 2010 (PMID: 20882035)
<i>RB1</i>	1	Garber and Offit, 2005	TSG	Vogelstein <i>et al.</i> , 2013
<i>RECQL4</i>	2	Garber and Offit, 2005	TSG	Kitao <i>et al.</i> , 1999 (PMID: 10552928)
<i>RET</i>	1	Garber and Offit, 2005	ONC	Vogelstein <i>et al.</i> , 2013
<i>RNASEL</i>	1	Garber and Offit, 2005	TSG	Chen <i>et al.</i> , 2003 (PMID: 12624150)
<i>SBDS</i>	2	Garber and Offit, 2005	UND	-
<i>SDHA</i>	1	Garber and Offit, 2005	TSG	Bardella <i>et al.</i> , 2011 (PMID: 21771581)
<i>SDHAF2</i>	1	Garber and Offit, 2005	TSG	Bardella <i>et al.</i> , 2011 (PMID: 21771581)

Gene	Selection criteria (number) ^a	Main reference	Classification (ONC, TSG or UND) ^b	Classification-related reference ^c
<i>SDHB</i>	1	Garber and Offit, 2005	TSG	Bardella <i>et al.</i> , 2011 (PMID: 21771581)
<i>SDHC</i>	1	Garber and Offit, 2005	TSG	Bardella <i>et al.</i> , 2011 (PMID: 21771581)
<i>SDHD</i>	1	Garber and Offit, 2005	TSG	Bardella <i>et al.</i> , 2011 (PMID: 21771581)
<i>SH2D1A</i>	1	Garber and Offit, 2005	UND	-
<i>SHOC2</i>	2	Komatsuzaki <i>et al.</i> , 2010 (PMID: 20882035)	ONC	Komatsuzaki <i>et al.</i> , 2010 (PMID: 20882035)
<i>SLX4</i>	2	Mehta and Tolar, 2002 (PMID: 20301575)	TSG	Mehta and Tolar, 2002 (PMID: 20301575)
<i>SMAD4</i>	1	Garber and Offit, 2005	TSG	Vogelstein <i>et al.</i> , 2013
<i>SMARCA4</i>	1	Garber and Offit, 2005	TSG	Vogelstein <i>et al.</i> , 2013
<i>SMARCB1</i>	1	Garber and Offit, 2005	TSG	Vogelstein <i>et al.</i> , 2013
<i>SOS1</i>	2	Roberts <i>et al.</i> , 2007 (PMID: 17143285)	ONC	Roberts <i>et al.</i> , 2007 (PMID: 17143285)
<i>STK11</i>	1	Garber and Offit, 2005	TSG	Vogelstein <i>et al.</i> , 2013
<i>TERT</i>	1	Horn <i>et al.</i> , 2013 (PMID: 23348503)	UND	-
<i>TMEM127</i>	1	Qin <i>et al.</i> , 2010 (PMID: 20154675)	TSG	Qin <i>et al.</i> , 2010 (PMID: 20154675)
<i>TP53</i>	1	Garber and Offit, 2005	TSG	Vogelstein <i>et al.</i> , 2013
<i>TSC1</i>	1	Garber and Offit, 2005	TSG	Vogelstein <i>et al.</i> , 2013
<i>TSC2</i>	1	Garber and Offit, 2005	TSG	Vogelstein <i>et al.</i> , 2013
<i>VHL</i>	1	Garber and Offit, 2005	TSG	Vogelstein <i>et al.</i> , 2013
<i>WAS</i>	1	Garber and Offit, 2005	UND	-
<i>WRN</i>	2	Garber and Offit, 2005	TSG	Moser <i>et al.</i> , 1999 (PMID: 10220139)
<i>WT1</i>	1	Garber and Offit, 2005	TSG	Vogelstein <i>et al.</i> , 2013
<i>XPA</i>	1	Garber and Offit, 2005	TSG	Kraemer and DiGiovanna, 2003 (PMID: 20301571)
<i>XPC</i>	1	Garber and Offit, 2005	TSG	Kraemer and DiGiovanna, 2003 (PMID: 20301571)
<i>XRCC3</i>	1	Winsey <i>et al.</i> , 2000 (PMID: 11059748)	TSG	Winsey <i>et al.</i> , 2000 (PMID: 11059748)

ONC, oncogene; TSG, tumor suppressor gene; UND, "undetermined function" gene.

^a The following criteria were used to select the cancer predisposition genes included in the current study: (1) genes associated with genetic syndromes in which the sole or main clinical phenotype is hereditary predisposition to cancer (classic hereditary cancer syndromes and certain types of familial cancer); (2) genes associated with syndromes/conditions in which one of the phenotypes is increased tumor predisposition; and (3) genes associated with an increased risk for certain sporadic tumor type(s); ^b "Undetermined function" (UND) genes did not fit into tumor suppressor or oncogene categories and/or without sufficient evidence about their role in tumor development;

^c The main reference related to this classification was the recent study of Vogelstein and colleagues (2013). In this study, genes were classified as Oncogenes if they had an "Oncogene Score" >20% and classified as a Tumor Suppressor Gene (TSG) if the "TSG Score" was >20% (the 20/20 rule). The "Oncogene Score" was defined as the number of clustered mutations (i.e., missense mutations at the same amino acid or identical in-frame insertions or deletions) divided by the total number of mutations. The "TSG Score" was defined as the number of truncating mutations divided by the total number of mutations. Truncating mutations included nonsense mutations, insertions or deletions that alter the reading frame, splice-site mutations, or mutations at the normal stop codon predicted to result in a longer protein. When a gene had an "Oncogene score" >20% and a "TSG Score" >5%, it was classified as a TSG because well-studied oncogenes rarely harbor stop codons. The major data source for this classification was the COSMIC database (www.sanger.ac.uk/genetics/CGP/cosmic/).

S2 Table. Background set composition used in the over-representation statistical analysis of regulation by miRNAs for oncogenes and tumor suppressor groups.

Database type^a	Human miRNA-gene interactions	miRNAs	Genes
Validated	410.270	2.587	15.120
Predicted	919.117	1.860	18.925
Both	1.206.936	1.588	19.252

^a The validated interactions were collected from miRTarBase, starBase, TarBase and miRecords sources, and computationally predicted data from TargetScan, Diana MicroT-CDS and miRanda-mirSVR algorithms. Interactions involving the studied cancer predisposition genes were not filtered in this background set.

Other Supplementary Tables:

The complete raw data from the microRNA-mediated regulation analysis will be available in the format of Excel Tables during article submission.

CAPÍTULO V – DISCUSSÃO

A discussão específica e completa referente aos resultados obtidos no presente estudo encontra-se no manuscrito apresentado no Capítulo IV. Neste capítulo serão mencionados aspectos mais gerais referentes ao tema, retomando questões não discutidas anteriormente, bem como novas hipóteses e perguntas de pesquisa que são suscitadas pelos resultados encontrados. Entretanto, alguma sobreposição de conteúdo é inevitável, visto a escassez de dados prévios sobre o tema na literatura.

O processo de clivagem e poliadenilação desempenha um papel fundamental na expressão gênica permitindo a maturação do pré-mRNA em um mRNA poliadenilado maduro, sendo interconectado com outras modificações pós-transcricionais, como os mecanismos de encadeamento (*splicing*) para remoção dos íntrons e junção dos éxons não contíguos (Zhao *et al.*, 1999; Maniatis e Reed, 2002; Millevoi e Vagner, 2010). Sabe-se que transcritos processados inadequadamente na extremidade 3' devem ser transportados do núcleo para o citoplasma com menor frequência a fim de serem traduzidos nos ribossomos, em comparação com aqueles mRNAs eficientemente processados. Isso ocorre porque a adição da cauda poli-A influencia a estabilidade do mRNA, sendo que mRNAs não-poliadenilados são rapidamente degradados por 3'-5' exonucleases (Sachs, 1990; Lewis *et al.*, 1995; Jacobson e Peltz, 1996; Lemay *et al.*, 2010).

A formação da extremidade 3' poliadenilada da maioria dos mRNAs eucarióticos ocorre essencialmente através do reconhecimento de elementos regulatórios em *cis* pelo complexo proteico de clivagem/poliadenilação, dentre os quais destaca-se o sinal de poliadenilação (SP), geralmente um hexâmero/hexanucleotídeo AAUAAA. Esses elementos regulatórios são determinantes para a eficiência do processo (Proudfoot e Brownlee, 1976; Chen *et al.*, 1995; Millevoi e Vagner, 2010; Hollerer *et al.*, 2014). O SP constitui o motivo mais frequentemente encontrado na região 3'UTR de genes codificadores de proteínas e está presente em cerca de 90% de todos os elementos de poliadenilação conhecidos até o momento, sendo reconhecido como uma das sequências mais altamente conservadas em transcritos de eucariotos (Graber *et al.*, 1999; Zhao *et al.*, 1999; Beaudoin *et al.*, 2000; Tian *et al.*, 2005; Xie *et al.*, 2005; Proudfoot, 2011). Sendo assim, não é surpreendente que um número considerável de variantes deletérias situadas

nessa sequência já tenham sido descritas (**Tabela 2** apresentada no Capítulo I). Corroborando esses achados e abordando a questão da poliadenilação alternativa tecido-específica, um estudo de Derti e colaboradores (2012) sugeriu que o SP AAUAAA está sob seleção purificadora em mamíferos, bem como a utilização de determinados sítios de poliadenilação alternativos preferenciais parece mais conservada no mesmo tecido em diferentes espécies do que em tecidos diferentes dentro da mesma espécie. Nesse mesmo estudo, o hexâmero canônico AAUAAA foi fortemente enriquecido e conservado quanto à sua posição em todas as espécies de mamíferos investigadas (Derti *et al.*, 2012). Além disso, a conservação evolutiva do SP constitutivo AAUAAA do gene *TP53* foi avaliada recentemente pelo nosso grupo de pesquisa e não foi observado qualquer tipo de variação nessa sequência em 63 espécies de mamíferos analisadas, incluindo dados de três espécies de homínídeos: o genoma humano, Neandertal e Denisova (Macedo *et al.*, 2016). No presente estudo, a alta conservação do SP também foi observada em um conjunto abrangente de genes de predisposição ao câncer (GPC), sendo a sequência consenso AAUAAA encontrada em 74,4% dos SP descritos para esses genes no banco de dados do NCBI.

Curiosamente, nos últimos anos, dois estudos independentes descreveram variantes patogênicas localizadas no SP e elementos adjacentes da região 3'UTR em GPC, especificamente nos genes supressores de tumor *TP53* (rs78378222) e *MSH6* (duplicação denominada S20) (Stacey *et al.*, 2011; Decorsière *et al.*, 2012). Entretanto, poucos estudos posteriores abordaram estratégias para avaliação da variabilidade genética nas sequências regulatórias de poliadenilação (SRP) em outros GPC. Essa escassez de estudos está relacionada ao fato de que a região 3'UTR e demais sequências regulatórias (região promotora e 5'UTR) podem ser negligenciadas na rotina de diagnóstico molecular das síndromes de predisposição hereditária ao câncer (SPHC), uma vez que as análises abrangem principalmente regiões codificantes e junções éxon-íntron (Garber e Offit, 2005; Cruz-Correa *et al.*, 2017). Conceitualmente, a região 3'UTR do mRNA maduro inicia no códon de parada, localizado no último éxon, e se estende até o sítio de clivagem (SC), não necessariamente situado dentro do último éxon e a partir do qual os pré-mRNAs correspondentes são poliadenilados (Edwalds-Gilbert *et al.*, 1997; An

et al., 2013). Portanto, os protocolos de sequenciamento que abrangem até o último éxon de determinado GPC podem não cobrir todas as SRP. Nesse contexto, tem sido observado que parcela significativa das famílias que preenchem critérios clínicos para SPHC não possui mutações patogênicas identificadas em genes de alta penetrância associados. Por exemplo, cerca de 20% e 60% das famílias com fenótipo clínico da Síndrome de Li-Fraumeni e sua variante, a Síndrome de Li-Fraumeni-like, não apresentam mutação causal detectada no gene *TP53*, respectivamente (Varley, 2003; Malkin, 2011). A identificação de variantes deletérias situadas nas SRP poderia contribuir para a explicação dessa “herdabilidade perdida” em SPHC, bem como seria importante para auxiliar o manejo terapêutico e o monitoramento dos indivíduos afetados e seus familiares.

Ainda em uma perspectiva de aprimoramento do diagnóstico molecular, nosso estudo ressaltou a diversidade de sítios de poliadenilação alternativos tecido-específicos presente no transcrito derivado do gene *PTEN*, no qual mutações germinativas estão associadas com a Síndrome de Cowden (Liaw *et al.*, 1997). Adicionalmente, foi destacado que dois sítios específicos são preferencialmente utilizados no processamento desse transcrito. Esse achado exemplifica o papel dos eventos de poliadenilação alternativa na regulação pós-transcricional da expressão gênica (aspecto revisado em Elkon *et al.*, 2013), bem como aponta uma possível aplicabilidade em mapear as regiões da 3'UTR contendo os sítios de poliadenilação com maior importância funcional em cada GPC, a qual seria incluir esses elementos nas análises moleculares e possibilitar a detecção de novas variantes potencialmente patogênicas e/ou com efeito modificador de fenótipo.

No presente estudo, outro elemento regulatório de poliadenilação caracterizado em GPC foi o SC, também chamado sítio de poliadenilação. Embora seja descrito que a maioria dos pré-mRNAs em eucariotos sejam clivados de modo preferencial imediatamente 3' de um dinucleotídeo “CA” (Chen *et al.*, 1995), o dinucleotídeo “AA” foi identificado como o SC mais frequente dentre os genes analisados. Esse resultado contraditório pode ser explicado por evidências experimentais que já associaram o dinucleotídeo “CA” com uma maior eficiência no processo de poliadenilação, mas não como uma sequência necessariamente exigida para a clivagem correta do pré-mRNA

(Sheets *et al.*, 1990; Chen *et al.*, 1995; Gehring *et al.*, 2001). A ocorrência de poliadenilação alternativa em grande parte dos GPC analisados (~90%) também foi um achado surpreendente, já que uma estimativa anterior em escala genômica indicava que apenas cerca de 50% dos genes humanos apresentavam mais de um sítio de poliadenilação funcional (Tian *et al.*, 2005). Tais resultados, bem como a observação de uma distância entre o SP e o SC maior que 30 nucleotídeos na sequência de alguns genes, nos levaram a questionar se essas estimativas fornecidas por estudos clássicos sobre o mecanismo de poliadenilação podem ser aplicadas a todos os genes/transcritos humanos conhecidos.

Por outro lado, nosso estudo não incluiu a caracterização de outro elemento regulatório central da poliadenilação situado 14-70 nucleotídeos a jusante do SP e reconhecido pela alta densidade de uracilas e repetições GU, denominado região rica em U/GU (Bagga *et al.*, 1995; Beyer *et al.*, 1997; Lemay *et al.*, 2010). Enquanto o SP AAUAAA e suas variantes funcionais são reconhecidas pelo fator CPSF (*Cleavage and Polyadenylation Specificity Factor*), que é a endonuclease responsável pela clivagem dos transcritos primários, a região rica em U/GU apresenta interação com uma subunidade específica do fator heterotrimérico CstF (*Cleavage Stimulation Factor*), o qual promove a eficiência desse processamento 3' (Beyer *et al.*, 1997; Wahle e Rügsegger, 1999; Ryan *et al.*, 2004; Danckwardt *et al.*, 2008). Essa sequência rica em U/GU e demais elementos auxiliares já descritos a montante e a jusante do SC não foram estudados por serem pouco conservados evolutivamente e possivelmente redundantes (Gil e Proudfoot, 1987; Chen e Wilusz, 1998; Tian e Graber, 2012; Hollerer *et al.*, 2014).

Por fim, nossas análises relacionadas à regulação da expressão gênica por miRNAs em GPC identificaram uma série de famílias de miRNAs com funções potencialmente oncogênicas (miR-192) e supressoras de tumor (miR-128, miR-1471, miR-483, miR-3170 e miR-218). Para a maioria dessas famílias, ensaios funcionais e análises de expressão gênica realizados anteriormente em amostras e linhagens celulares de diversos tipos tumorais corroboram as funções aqui sugeridas, exceto para miR-1471 e miR-3170 que apresentam poucas evidências na literatura sobre o seu papel na carcinogênese. Embora tenha sido utilizada uma abordagem *in silico* e estudos

adicionais sejam necessários para validar as funções atribuídas a essas famílias de miRNAs, nossos resultados contribuem para uma melhor compreensão da modulação da expressão por miRNAs compartilhada entre grupos de genes supressores de tumor e oncogenes envolvidos em diferentes mecanismos de reparo do DNA, vias de sinalização celular e demais processos fisiológicos.

CAPÍTULO VI – CONCLUSÕES

Considerando os resultados encontrados no presente estudo, podemos concluir que:

- a) Uma parcela substancial dos GPC estudados (~18%) não apresenta sequência de SP indicada em um banco de dados referência (NCBI), destacando a necessidade do estabelecimento de ferramentas de bioinformática e bancos de dados atualizados para detectar hexâmeros que podem funcionar como SP em genes humanos;
- b) Os SP da maioria dos GPC são caracterizados pela sequência consenso AAUAAA, corroborando o conhecimento prévio de que esse hexâmero representa o SP mais comum em genes humanos e uma das sequências mais altamente conservadas em eucariotos;
- c) O dinucleotídeo “AA” constitui o SC mais comum considerando os GPC analisados, em contraste com estimativas anteriores de que a maioria dos pré-mRNAs em eucariotos seriam clivados preferencialmente 3’ de um dinucleotídeo “CA”, o qual apresenta maior reconhecimento pelo complexo proteico de clivagem/poliadenilação e seria fisiologicamente mais eficiente;
- d) Aproximadamente 11% dos GPC incluídos nesse estudo exibiram uma distância entre o SP e SC maior que 30 nucleotídeos, demonstrando que certas estimativas de estudos experimentais clássicos sobre poliadenilação não se aplicam necessariamente a todos os genes/transcritos humanos;
- e) Eventos de poliadenilação alternativa ocorrem em grande parte dos GPC avaliados (~90%), em comparação com uma estimativa anterior da sua frequência em cerca de metade dos genes humanos (~54%), o que sugere uma maior complexidade na regulação da poliadenilação em transcritos especificamente derivados de GPC;

- f) Alguns GPC apresentam um número elevado de sítios de poliadenilação alternativos diferencialmente utilizados entre os vários tecidos humanos, dentre os quais destaca-se o gene supressor tumoral *PTEN* que apresenta 61 SP tecido-específicos, sendo dois deles caracterizados pelo hexâmero variante AUUAAA e preferencialmente reconhecidos pela maquinaria de clivagem/poliadenilação durante o processamento do transcrito correspondente;
- g) A família de miRNA miR-192 foi significativamente super-representada como reguladora da expressão de genes supressores de tumor, apresentando sítios de ligação experimentalmente validados na região 3'UTR de genes que codificam proteínas essenciais em diversas vias de reparo do DNA, sugerindo uma função potencialmente oncogênica. Essa função também é sugerida por uma série de estudos publicados anteriormente envolvendo diversos tipos de câncer;
- h) Embora o número de oncogenes estudados fosse menor, várias famílias de miRNAs fortemente associadas com a regulação desse grupo de genes foram identificadas a partir das análises *in silico*, dentre elas miR-128, miR-1471, miR-483, miR-3170 e miR-218. Evidências experimentais recentes corroborando uma função supressora de tumor já foram descritas para as famílias miR-128, miR-483 e miR-218;
- i) Os principais achados obtidos a partir das análises de regulação por miRNAs foram consistentes com evidências anteriores da literatura, sugerindo que a combinação de fontes de dados de interações miRNA-mRNA experimentalmente validadas e preditas computacionalmente representa uma metodologia adequada para a identificação de miRNAs que podem atuar como supressores de tumor e oncogenes;
- j) Estudos funcionais devem ser realizados para validar as potenciais funções supressoras de tumor e oncogênicas atribuídas a partir das análises *in silico* às famílias de miRNAs mencionadas anteriormente;

- k) Nossos dados fornecem um mapeamento das SRP em GPC, o qual pode ser utilizado no desenvolvimento de análises moleculares mais refinadas incluindo esses elementos da região 3'UTR frequentemente negligenciados na rotina de diagnóstico molecular das síndromes de câncer hereditário;

- l) Ao nosso conhecimento, o presente estudo é o primeiro a abordar esse tipo de caracterização abrangente (elementos regulatórios de poliadenilação e modulação por miRNAs) em genes de suscetibilidade ao câncer.

CAPÍTULO VII – PERSPECTIVAS

Como perspectivas e sugestões de análises adicionais destacamos as seguintes:

- Aumentar o número de genes de predisposição ao câncer (GPC) avaliados, especialmente expandindo o grupo de oncogenes para as análises de modulação da expressão por miRNAs;
- Investigar a presença de variações de sequência germinativas já descritas nas sequências regulatórias de poliadenilação (SRP) dos GPC estudados, consultando bancos de dados genéticos populacionais humanos, como o Projeto 1000 Genomas, *Exome Aggregation Consortium* (ExAC), entre outros;
- Avaliar a presença de variantes somáticas localizadas nas SRP dos GPC analisados utilizando o banco de dados genômicos de diferentes tipos tumorais denominado TCGA (*The Cancer Genome Atlas*);
- Analisar os dados de expressão de miRNAs em amostras de tecido tumoral disponibilizados pelo TCGA, a fim de identificar o padrão de expressão das famílias de miRNAs que podem desempenhar funções supressoras de tumor e oncogênicas destacadas no presente estudo;
- Realizar análises de enriquecimento funcional utilizando o banco de dados relacionado a mapas de vias biológicas *Kyoto Encyclopedia of Genes and Genomes (KEGG Pathway)*, com o objetivo de determinar as vias de sinalização celular reguladas pelas famílias de miRNAs aqui identificadas como potencialmente supressoras de tumor e oncogênicas;
- Caracterizar as variações de sequência na região 3'UTR do gene *TP53* em pacientes com fenótipo clínico de uma síndrome de câncer hereditário amplamente estudada pelo nosso grupo de pesquisa, a Síndrome de Li-Fraumeni (SLF), visando a identificação de variantes

patogênicas ou com potencial efeito modificador de fenótipo localizadas em sítios de ligação para miRNAs ou nas SRP do gene;

- Estudar o papel de SNPs funcionais localizados no gene *MIR34A* (rs35301225, rs6577555 e rs72631823), que codifica o miRNA supressor de tumor miR-34a, como potenciais modificadores de fenótipo na síndrome de SLF/LFL, considerando que nossas análises mostraram que este miRNA regula diretamente o gene *TP53* e um dos seus principais reguladores negativos, o oncogene *MDM4*;
- Avaliar os níveis de metilação no promotor do gene *MIR34A* na linhagem germinativa e em tecido tumoral (tumor primário) de pacientes SLF/LFL, a fim de investigar se alterações epigenéticas nesse gene podem atuar como modificadores de fenótipo da síndrome;
- Elaborar um painel-piloto de sequenciamento de nova geração baseado na análise das regiões 3'UTR em diferentes genes associados com síndromes de predisposição hereditária ao câncer (SPHC) altamente prevalentes na nossa população e/ou genes contendo variantes funcionais já descritas afetando sítios de ligação para miRNAs e SRP, cuja principal aplicabilidade seria a testagem de pacientes que preenchem critérios clínicos para determinada SPHC, mas não apresentam mutações patogênicas identificadas nas regiões codificantes e junções éxon-íntron dos genes associados.

REFERÊNCIAS BIBLIOGRÁFICAS

Agarwal V, Bell GW, Nam J e Bartel DP (2015) Predicting effective microRNA target sites in mammalian mRNAs. *eLife* 4:e05005.

Ambros V (2004) The functions of animal microRNAs. *Nature* 431:350-355.

An J, Zhu X, Wang H e Jin X (2013) A dynamic interplay between alternative polyadenylation and microRNA regulation: implications for cancer (Review). *Int J Oncol* 43:995-1001.

Bagga PS, Ford LP, Chen F e Wilusz J (1995) The G-rich auxiliary downstream element has distinct sequence and position requirements and mediates efficient 3' end and pre-mRNA processing through a trans-acting factor. *Nucleic Acids Res* 23:1625-1631.

Balim Z, Kosova B, Falzon K, Bezzina Wettinger S e Colak Y (2003) Budd–Chiari syndrome in a patient heterozygous for the point mutation C20221T of the prothrombin gene. *J Thromb Haemost* 1:852-853.

Bartel DP (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 116:281-297.

Bartel DP (2009) MicroRNAs: target recognition and regulatory functions. *Cell* 136:215-233.

Bauer NC, Corbett AH e Doetsch PW (2015) The current state of eukaryotic DNA base damage and repair. *Nucleic Acids Res* 43:10083-100101.

Beaudoing E, Freier S, Wyatt JR, Claverie JM e Gautheret D (2000) Patterns of variant polyadenylation signal usage in human genes. *Genome Res* 10:1001-1010.

Beilharz TH, Humphreys DT, Clancy JL, Thermann R, Martin DI, Hentze MW e Preiss T (2009) microRNA-mediated messenger RNA deadenylation contributes to translational repression in mammalian cells. *PLoS One* 4:e6783.

Bennett CL, Brunkow ME, Ramsdell F, O'Briant KC, Zhu Q, Fuleihan RL, Shigeoka AO, Ochs HD e Chance PF (2001) A rare polyadenylation signal mutation of the FOXP3 gene (AAUAAA-->AAUGAA) leads to the IPEX syndrome. *Immunogenetics* 53:435-439.

Beyer K, Dandekar T e Keller W (1997) RNA ligands selected by cleavage stimulation factor contain distinct sequence motifs that function as downstream elements in 3'-end processing of pre-mRNA. *J Biol Chem* 272:26769-26779.

Biesecker BB e Garber JE (1995) Testing and counseling adults for heritable cancer risk. *J Natl Cancer Inst Monogr* 115-118.

Bonapace G, Concolino D, Formicola S e Strisciuglio P (2003) A novel mutation in a patient with insulin-like growth factor 1 (IGF1) deficiency. *J Med Genet* 40:913-917.

Brady RO, Gal AE, Bradley RM, Martensson E, Warshaw AL e Laster L (1967) Enzymatic defect in Fabry's disease. Ceramidetrihexosidase deficiency. *N Engl J Med* 276:1163-1167.

Calin GA, Sevignani C, Dumitru CD, Hyslop T, Noch E, Yendamuri S, Shimizu M, Rattan S, Bullrich F, Negrini M, *et al.* (2004) Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. *Proc Natl Acad Sci U S A* 101:2999-3004.

Chen F, MacDonald CC e Wilusz J (1995) Cleavage site determinants in the mammalian polyadenylation signal. *Nucleic Acids Res* 23:2614-2620.

Chen F e Wilusz J (1998) Auxiliary downstream elements are required for efficient polyadenylation of mammalian pre-mRNAs. *Nucleic Acids Res* 26:2891-2898.

Chen JM, Férec C e Cooper DN (2006) A systematic analysis of disease-associated variants in the 3' regulatory regions of human protein-coding genes II: the importance of mRNA secondary structure in assessing the functionality of 3' UTR variants. *Hum Genet* 120:301-333.

Chin LJ, Ratner E, Leng S, Zhai R, Nallur S, Babar I, Muller RU, Straka E, Su L, Burki EA, *et al.* (2008) A SNP in a let-7 microRNA complementary site in the KRAS 3' untranslated region increases non-small cell lung cancer risk. *Cancer Res* 68:8535-8540.

Chou CH, Chang NW, Shrestha S, Hsu SD, Lin YL, Lee WH, Yang CD, Hong HC, Wei TY, Tu SJ, *et al.* (2016) miRTarBase 2016: updates to the experimentally validated miRNA-target interactions database. *Nucleic Acids Res* 44:D239-247

Conway L e Wickens M (1987) Analysis of mRNA 3' end formation by modification interference: the only modifications which prevent processing lie in AAUAAA and the poly(A) site. *EMBO J* 6:4177-4184.

Croce CM (2008) Oncogenes and cancer. *N Engl J Med* 358:502-511.

Cruz-Correa M, Pérez-Mayoral J, Dutil J, Echenique M, Mosquera R, Rivera-Román K, Umpierre S, Rodriguez-Quilichini S, Gonzalez-Pons M, Olivera MI, *et al.* (2017) Hereditary cancer syndromes in Latino populations: genetic characterization and surveillance guidelines. *Hered Cancer Clin Pract* 15:3.

Danckwardt S, Gehring NH, Neu-Yilik G, Hundsdoerfer P, Pforsich M, Frede U, Hentze MW e Kulozik AE (2004) The prothrombin 3' end formation

signal reveals a unique architecture that is sensitive to thrombophilic gain-of-function mutations. *Blood* 104:428-435.

Danckwardt S, Hartmann K, Katz B, Hentze M, Levy Y, Eichele R, Deutsch V, Kulozik A e Ben-Tal O (2006) The prothrombin 20209 C4T mutation in Jewish–Moroccan Caucasians: molecular analysis of gain-of-function of 3' end processing. *J Thromb Haemost* 4:1078-1085.

Danckwardt S, Hentze MW e Kulozik AE (2008) 3' end mRNA processing: molecular mechanisms and implications for health and disease. *EMBO J* 27:482-498.

de la Chapelle A e Peltomäki P (1998) The genetics of hereditary common cancers. *Curr Opin Genet Dev* 8:298-303.

Decorsière A, Toulas C, Fouque F, Tilkin-Mariamé AF, Selves J, Guimbaud R, Chipoulet E, Delmas C, Rey JM, Pujol P, *et al.* (2012) Decreased efficiency of MSH6 mRNA polyadenylation linked to a 20-base-pair duplication in Lynch syndrome families. *Cell Cycle* 11:2578-2580.

Derti A, Garrett-Engele P, Macisaac KD, Stevens RC, Sriram S, Chen R, Rohl CA, Johnson JM e Babak T (2012) A quantitative atlas of polyadenylation in five mammals. *Genome Res* 22:1173-1183.

Diederichs S, Bartsch L, Berkmann JC, Fröse K, Heitmann J, Hoppe C, Iggena D, Jazmati D, Karschnia P, Linsenmeier M, *et al.* (2016) The dark matter of the cancer genome: aberrations in regulatory elements, untranslated regions, splice sites, non-coding RNA and synonymous mutations. *EMBO Mol Med* 8:442-457.

Dweep H, Sticht C e Gretz N (2013) In-Silico Algorithms for the Screening of Possible microRNA Binding Sites and Their Interactions. *Curr Genomics* 14:127-136.

Edwards-Gilbert G, Veraldi KL e Milcarek C (1997) Alternative poly(A) site selection in complex transcription units: means to an end? *Nucleic Acids Res* 25:2547-2561.

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

Elton TS e Yalowich JC (2015) Experimental procedures to identify and validate specific mRNA targets of miRNAs. *EXCLI J* 14:758-790.

Endale Ahanda ML, Fritz ER, Estellé J, Hu ZL, Madsen O, Groenen MA, Beraldi D, Kapetanovic R, Hume DA, Rowland RR, *et al.* (2012) Prediction of altered 3'- UTR miRNA-binding sites from RNA-Seq data: the swine leukocyte antigen complex (SLA) as a model region. *PLoS One* 7:e48607.

Esquela-Kerscher A e Slack FJ (2006) Oncomirs - microRNAs with a role in cancer. *Nat Rev Cancer* 6:259-269.

Fabian MR, Sonenberg N e Filipowicz W (2010) Regulation of mRNA translation and stability by microRNAs. *Annu Rev Biochem* 79:351-379.

Friedman RC, Farh KK, Burge CB e Bartel DP (2009) Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res* 19:92-105.

Garber JE e Offit K (2005) Hereditary cancer predisposition syndromes. *J Clin Oncol* 23:276-292.

Gehring NH, Frede U, Neu-Yilik G, Hundsdoerfer P, Vetter B, Hentze MW e Kulozik AE (2001) Increased efficiency of mRNA 3' end formation: a new genetic mechanism contributing to hereditary thrombophilia. *Nat Genet* 28:389-392.

Gil A e Proudfoot NJ (1987) Position-dependent sequence elements downstream of AAUAAA are required for efficient rabbit beta-globin mRNA 3' end formation. *Cell* 49:399-406.

Giordano PC, Bouva MJ, Van Delft P, Akkerman N, Kappers-Klunne MC e Hartevelde CL (2005) A new polyadenylation site mutation associated with a mild beta-thalassemia phenotype. *Haematologica* 90:551-552.

Graber JH, Cantor CR, Mohr SC e Smith TF (1999) In silico detection of control signals: mRNA 3'-end-processing sequences in diverse species. *Proc Natl Acad Sci U S A* 96:14055-14060.

Grimson A, Farh KK, Johnston WK, Garrett-Engele P, Lim LP e Bartel DP (2007) MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol Cell* 27:91-105.

Groden J, Thliveris A, Samowitz W, Carlson M, Gelbert L, Albertsen H, Joslyn G, Stevens J, Spirio L, Robertson M, *et al.* (1991) Identification and characterization of the familial adenomatous polyposis coli gene. *Cell* 66:589-600.

Hanahan D e Weinberg RA (2011) Hallmarks of cancer: the next generation. *Cell* 144:646-674.

Higgs DR, Goodbourn SE, Lamb J, Clegg JB, Weatherall DJ e Proudfoot NJ (1983) Alpha-thalassaemia caused by a polyadenylation signal mutation. *Nature* 306:398-400.

Hollerer I, Grund K, Hentze MW, Kulozik AE (2014) mRNA 3'end processing: A tale of the tail reaches the clinic. *EMBO Mol Med* 6: 16-26.

Jacobson A e Peltz SW (1996) Interrelationships of the pathways of mRNA decay and translation in eukaryotic cells. *Annu Rev Biochem* 65:693-739.

Kinzler KW e Vogelstein B (1998) Landscaping the cancer terrain. *Science* 280:1036-1037.

Knudson AG (1971) Mutation and cancer: statistical study of retinoblastoma. *Proc Natl Acad Sci U S A* 68:820-823.

Kuwahara M, Kurachi S e Kurachi K (2004) Molecular mechanism of prothrombin G20210A variant: critical new role of exon splicing enhancer in poly(A) tailing. *Blood* 104:1944 (abstract).

Leach FS, Nicolaidis NC, Papadopoulos N, Liu B, Jen J, Parsons R, Peltomäki P, Sistonen P, Aaltonen LA, Nyström-Lahti M, *et al.* (1993) Mutations of a mutS homolog in hereditary nonpolyposis colorectal cancer. *Cell* 75:1215-1225.

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

Lemay JF, Lemieux C, St-André O e Bachand F (2010) Crossing the borders: poly(A)-binding proteins working on both sides of the fence. *RNA Biol* 7:291-295.

Lewis JD, Gunderson SI e Mattaj JW (1995) The influence of 5' and 3' end structures on pre-mRNA metabolism. *J Cell Sci Suppl* 19:13-19.

Li JH, Liu S, Zhou H, Qu LH e Yang JH (2014) starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Res* 42:D92-97.

Liaw D, Marsh DJ, Li J, Dahia PL, Wang SI, Zheng Z, Bose S, Call KM, Tsou HC, Peacocke M, *et al.* (1997) Germline mutations of the PTEN gene in Cowden disease, an inherited breast and thyroid cancer syndrome. *Nat Genet* 16:64-67.

Liu B, Parsons R, Papadopoulos N, Nicolaides NC, Lynch HT, Watson P, Jass JR, Dunlop M, Wyllie A, Peltomäki P, *et al.* (1996) Analysis of mismatch repair genes in hereditary non-polyposis colorectal cancer patients. *Nat Med* 2:169-174.

Long D, Lee R, Williams P, Chan CY, Ambros V e Ding Y (2007) Potent effect of target structure on microRNA function. *Nat Struct Mol Biol* 14:287-294.

Lu J, Getz G, Miska EA, Alvarez-Saavedra E, Lamb J, Peck D, Sweet-Cordero A, Ebert BL, Mak RH, Ferrando AA, *et al.* (2005) MicroRNA expression profiles classify human cancers. *Nature* 435:834-838.

Lutz CS (2008) Alternative polyadenylation: a twist on mRNA 3' end formation. *ACS Chem Biol* 3:609-617.

MacDonald CC e Redondo JL (2002) Reexamining the polyadenylation signal: were we wrong about AAUAAA? *Mol Cell Endocrinol* 190:1-8.

Macedo GS, Araujo Vieira I, Brandalize AP, Giacomazzi J, Inez Palmero E, Volc S, Rodrigues Paixão-Côrtés V, Caleffi M, Silva Alves M, Achatz MI, *et al.* (2016) Rare germline variant (rs78378222) in the TP53 3' UTR: Evidence for a new mechanism of cancer predisposition in Li-Fraumeni syndrome. *Cancer Genet* 209:97-106.

Malkin D, Li FP, Strong LC, Fraumeni JF, Nelson CE, Kim DH, Kassel J, Gryka MA, Bischoff FZ, Tainsky MA, *et al.* (1990) Germ line p53 mutations in a familial syndrome of breast cancer, sarcomas, and other neoplasms. *Science* 250:1233-1238.

Malkin D (2011) Li-fraumeni syndrome. *Genes Cancer* 2:475-484.

Maniatis T e Reed R (2002) An extensive network of coupling among gene expression machines. *Nature* 416:499–506.

Manley JL, Yu H e Ryner L (1985) RNA sequence containing hexanucleotide AAUAAA directs efficient mRNA polyadenylation in vitro. *Mol Cell Biol* 5:373-379.

Maragkakis M, Alexiou P, Papadopoulos GL, Reczko M, Dalamagas T, Giannopoulos G, Goumas G, Koukis E, Kourtis K, Simossis VA, *et al.* (2009) Accurate microRNA target prediction correlates with protein repression levels. *BMC Bioinformatics* 10:295.

Marín RM e Vanícek J (2011) Efficient use of accessibility in microRNA target prediction. *Nucleic Acids Res* 39:19-29.

Mathonnet G, Fabian MR, Svitkin YV, Parsyan A, Huck L, Murata T, Biffo S, Merrick WC, Darzynkiewicz E, Pillai RS, *et al.* (2007) MicroRNA inhibition of translation initiation in vitro by targeting the cap-binding complex eIF4F. *Science* 317:1764-1767.

Mayr C e Bartel DP (2009) Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. *Cell* 138:673-684.

Michalova E, Vojtesek B e Hrstka R (2013) Impaired pre-mRNA processing and altered architecture of 3' untranslated regions contribute to the development of human disorders. *Int J Mol Sci* 14:15681-15694.

Miki Y, Swensen J, Shattuck-Eidens D, Futreal PA, Harshman K, Tavtigian S, Liu Q, Cochran C, Bennett LM, Ding W, *et al.* (1994) A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science* 266:66-71.

Millevoi S e Vagner S (2010) Molecular mechanisms of eukaryotic pre-mRNA 3' end processing regulation. *Nucleic Acids Res* 38: 2757-2774.

miRBase versão 21.0 (2014) Browse miRBase by species. Disponível em: <<http://mirbase.org/cgi-bin/browse.pl>>. (Acessado em 21 de fevereiro de 2017).

Moore CL e Sharp PA (1984) Site-specific polyadenylation in a cell-free reaction. *Cell* 36:581-591.

Moore CL e Sharp PA (1985) Accurate cleavage and polyadenylation of exogenous RNA substrate. *Cell* 41:845-855.

Morris AR, Bos A, Diosdado B, Rooijers K, Elkon R, Bolijn AS, Carvalho B, Meijer GA e Agami R (2012) Alternative cleavage and polyadenylation during colorectal cancer development. *Clin Cancer Res* 18:5256-5266.

Murga M e Fernández-Capetillo O (2007) Genomic instability: on the birth and death of cancer. *Clin Transl Oncol* 9:216-20.

Nilsen TW (2007) Mechanisms of microRNA-mediated gene regulation in animal cells. *Trends Genet* 23:243-249.

Orkin SH, Cheng TC, Antonarakis SE e Kazazian HH.Jr (1985) Thalassemia due to a mutation in the cleavage-polyadenylation signal of the human beta-globin gene. *EMBO J* 4:453-456.

Osborne C, Wilson P e Tripathy D (2004) Oncogenes and tumor suppressor genes in breast cancer: potential diagnostic and therapeutic applications. *Oncologist* 9:361-377.

Paraskevopoulou MD, Georgakilas G, Kostoulas N, Vlachos IS, Vergoulis T, Reczko M, Filippidis C, Dalamagas T e Hatzigeorgiou AG (2013) DIANAmicroT web server v5.0: service integration into miRNA functional analysis workflows. *Nucleic Acids Res* 41:W169-73.

Ponder B (1997) Genetic testing for cancer risk. *Science* 278:1050-1054.

Poort SR, Rosendaal FR, Reitsma PH e Bertina RM (1996) A common genetic variation in the 3' untranslated region of the prothrombin gene is

associated with elevated plasma prothrombin levels and an increase in venous thrombosis. *Blood* 88:3698-3703.

Portnoy V, Huang V, Place RF e Li LC (2011) Small RNA and transcriptional upregulation. *Wiley Interdiscip Rev RNA* 2:748-760.

Proudfoot N (1991) Poly(A) signals. *Cell* 64:671-674.

Proudfoot NJ (2011) Ending the message: poly(A) signals then and now. *Genes Dev* 25:1770-1782.

Proudfoot NJ e Brownlee GG (1976) 3' non-coding region sequences in eukaryotic messenger RNA. *Nature* 263:211-214.

Rachagani S, Kumar S e Batra SK (2010) MicroRNA in pancreatic cancer: pathological, diagnostic and therapeutic implications. *Cancer Lett* 292:8-16.

Rehmsmeier M, Steffen P, Hochsmann M e Giegerich R (2004) Fast and effective prediction of microRNA/target duplexes. *RNA* 10:1507-1517.

Retelska D, Iseli C, Bucher P, Jongeneel CV e Naef F (2006) Similarities and differences of polyadenylation signals in human and fly. *BMC Genomics* 7:176.

Rocha JCC (2010) Síndromes de Câncer Hereditário. In: Ferreira CG e Rocha JCC (eds) *Oncologia Molecular*. 2ª edição. Editora Atheneu, São Paulo, pp 445-450.

Ruan K, Fang X e Ouyang G (2009) MicroRNAs: novel regulators in the hallmarks of human cancer. *Cancer Lett* 285:116-126.

Rund D, Dowling C, Najjar K, Rachmilewitz EA, Kazazian HH Jr e Oppenheim A (1992) Two mutations in the beta-globin polyadenylation signal

reveal extended transcripts and new RNA polyadenylation sites. *Proc Natl Acad Sci USA* 89:4324-4328.

Ryan K, Calvo O e Manley JL (2004) Evidence that polyadenylation factor CPSF-73 is the mRNA 3' processing endonuclease. *RNA* 10:565-573.

Sachs A (1990) The role of poly(A) in the translation and stability of mRNA. *Curr Opin Cell Biol* 2:1092-1098.

Sadofsky M e Alwine JC (1984) Sequences on the 3' side of hexanucleotide AAUAAA affect efficiency of cleavage at the polyadenylation site. *Mol Cell Biol* 4:1460-1468.

Sandberg R, Neilson JR, Sarma A, Sharp PA e Burge CB (2008) Proliferating cells express mRNAs with shortened 3' untranslated regions and fewer microRNA target sites. *Science* 320:1643-1647.

Schwarzenbach H, Nishida N, Calin GA e Pantel K (2014) Clinical relevance of circulating cell-free microRNAs in cancer. *Nat Rev Clin Oncol* 11:145-156.

Sethupathy P, Corda B e Hatzigeorgiou AG (2006) TarBase: A comprehensive database of experimentally supported animal microRNA targets. *RNA* 12:192-197.

Sheets MD, Ogg SC e Wickens MP (1990) Point mutations in AAUAAA and the poly (A) addition site: effects on the accuracy and efficiency of cleavage and polyadenylation in vitro. *Nucleic Acids Res* 18:5799-5805.

Sidransky D (1997) Nucleic acid-based methods for the detection of cancer. *Science* 278:1054-1059.

Stacey SN, Sulem P, Jonasdottir A, Masson G, Gudmundsson J, Gudbjartsson DF, Magnusson OT, Gudjonsson SA, Sigurgeirsson B,

Thorisdottir K, *et al.* (2011) A germline variant in the TP53 polyadenylation signal confers cancer susceptibility. *Nat Genet* 43:1098-1103.

Tian B e Graber JH (2012) Signals for pre-mRNA cleavage and polyadenylation. *Wiley Interdiscip Rev RNA* 3:385-396.

Tian B, Hu J, Zhang H e Lutz CS (2005) A large-scale analysis of mRNA polyadenylation of human and mouse genes. *Nucleic Acids Res* 33:201-212.

Varley JM (2003) Germline TP53 mutations and Li-Fraumeni syndrome. *Hum Mutat* 21:313-320.

Wahle E e Rügsegger U (1999) 3'-End processing of pre-mRNA in eukaryotes. *FEMS Microbiol Rev* 23:277-295.

Weinstein IB e Joe A (2008) Oncogene addiction. *Cancer Res* 68:3077-3080.

Wickens M e Stephenson P (1984) Role of the conserved AAUAAA sequence: four AAUAAA point mutants prevent messenger RNA 3' end formation. *Science* 226:1045-1051.

Winter J, Jung S, Keller S, Gregory RI e Diederichs S (2009) Many roads to maturity: microRNA biogenesis pathways and their regulation. *Nat Cell Biol* 11: 228-234.

Witkos TM, Koscianska E e Krzyzosiak WJ (2011) Practical Aspects of microRNA Target Prediction. *Curr Mol Med* 11:93-109.

Wooster R, Neuhausen SL, Mangion J, Quirk Y, Ford D, Collins N, Nguyen K, Seal S, Tran T, Averill D, *et al.* (1994) Localization of a breast cancer susceptibility gene, BRCA2, to chromosome 13q12-13. *Science* 265:2088-2090.

Xiao F, Zuo Z, Cai G, Kang S, Gao X e Li T (2009) miRecords: an integrated resource for microRNAtarget interactions. *Nucleic Acids Res* 37:D105-110.

Xie X, Lu J, Kulbokas EJ, Golub TR, Mootha V, Lindblad-Toh K, Lander ES e Kellis M (2005) Systematic discovery of regulatory motifs in human promoters and 3' UTRs by comparison of several mammals. *Nature* 434:338-345.

Yasuda M, Shabbeer J, Osawa M e Desnick RJ (2003) Fabry disease: novel alpha-galactosidase A 3'-terminal mutations result in multiple transcripts due to aberrant 3'-end formation. *Am J Hum Genet* 73:162-173.

Zarkower D, Stephenson P, Sheets M e Wickens M (1986) The AAUAAA sequence is required both for cleavage and for polyadenylation of simian virus 40 pre-mRNA in vitro. *Mol Cell Biol* 6:2317-2323.

Zhao J, Hyman L e Moore C (1999) Formation of mRNA 3' ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis. *Microbiol Mol Biol Rev* 63:405-445.