

# PROTEÔMICA DO CÂNCER: ISOFORMAS DE SPLICING ALTERNATIVO COMO POTENCIAIS MARCADORES MOLECULARES DE TECIDOS TUMORAIS\*

**Morais Gama, L. C.**

Complexo Educacional Faculdades  
Metropolitanas Unidas

A. C. Camargo

Centro Internacional de Pesquisa (CIPE)

**Ramalho, R. F.**

A. C. Camargo

Centro Internacional de Pesquisa (CIPE), Rua  
Taguá, 440, Liberdade, 01508-010

**E-mail:** rramalho@cipe.accamargo.org.br

## RESUMO

Resultados recentes indicam que a espectrometria de massas (MS) é capaz de detectar proteínas que diferem daquelas descritas pela anotação referência do genoma humano (proteínas alternativas). Em trabalho anterior, usamos dados de MS e identificamos, para 19 genes, proteínas alternativas geradas pela tradução de transcritos contendo *exon-skipping*. O estudo do proteoma em linhagens celulares tumorais revelou um número de proteínas alternativas aproximadamente 10 vezes maior do que em tecidos normais. No presente trabalho buscamos em dados públicos de espectrometria de massa de adenocarcinoma de colon e colon normal pela presença de 14 isoformas geradas por *splicing* alternativo previamente detectadas em tecidos humanos normais. Assumindo uma taxa de falso-positivo menor do que 1%, observamos que variantes dos genes HNRNPAB (sem o exon 6) e FN1 (sem o exon 25) foram detectadas nas amostras de cólon tumoral, mas não em cólon normal e portanto representam potenciais marcadores de adenocarcinoma de cólon.

**Palavras-chave:** câncer; exon-skipping; marcadores; proteômica.

---

\* Apresentado e premiados na XVI Jornada da Biomedicina - 2016.

## 1. INTRODUÇÃO

O *splicing* alternativo é o mecanismo celular pelo qual o transcrito primário (RNA primário) de um gene é processado de formas distintas resultando em diferentes transcritos maduros (transcritos alternativos) para um mesmo gene (1). Existem muitos tipos diferentes de *splicing* mas em mamíferos o mais abundante é o *exon-skipping* (2). Apesar desta ampla ocorrência, a relevância funcional desse fenômeno ainda é controversa (3, 4) e ainda não se sabe com precisão qual fração dos transcritos alternativos é traduzida em proteínas.

Atualmente a espectrometria de massas (MS) é o método de larga escala mais utilizado para detectar proteínas em uma amostra (para revisão do método ver (5)). Diversos trabalhos utilizaram essa técnica para detectar proteínas resultantes da tradução de transcritos alternativos, aqui chamadas de proteínas alternativas. Recentemente, Lawrence et al., 2015 (6), usou MS e detectou 10 vezes mais proteínas alternativas em linhagens celulares de câncer do que previamente observado em tecidos normais. Até o momento, poucos estudos avaliaram a presença de proteínas alternativas em amostras de tecidos tumorais.

## 2. OBJETIVOS

O objetivo principal deste projeto é avaliar em amostras tumorais e normais de cólon a presença de variantes de *splicing* para 14 proteínas previamente detectadas em tecidos normais (7). Isso é importante pois poderá revelar marcadores tumorais que auxiliarão no diagnóstico de adenocarcinoma de cólon e poderão servir de alvos terapêuticos.

## 3. MATERIAL E MÉTODOS

Os dados de espectrometria de massas de cólon (tumor e normal) serão obtidos do banco de dados público CPTAC (*Clinical Proteomic Tumor Analysis Consortium*). O conjunto de sequências de proteínas alternativas será o mesmo utilizado por Ramalho et. al., 2015 (7). Essas sequências foram criadas a partir de tradução *in-silico* de transcritos contendo eventos de *exon-skipping*.

A busca por essas proteínas alternativas será feita através do programa MS-GF+ (*Universal Database Tool for Mass Spectrometry*) (8) que realiza o pareamento (*match*) entre os espectros de massas e um banco de sequências de aminoácidos. Para cada PSM (*peptide-spectrum match*) encontrado, o programa calcula um valor de qualidade (*MSGFScore*), assim como a significância estatística do pareamento (*SpecEvaluate*).

#### 4. DESENVOLVIMENTO

Para estimar a taxa de falso-positivo (*False-Discovery rate, FDR*) para cada valor de *MSGFScore*, utilizamos uma abordagem conhecida como *target-decoy* (9), em que a busca computacional é feita em um conjunto de proteínas reais misturadas à proteínas falsas (simuladas). Com essa técnica calcula-se o FDR para um dado *MSGFScore* (*i*) através da fórmula:  $FDR_i = N_s / N_f$ , onde  $N_s$  é o total acumulado de PSMs falsos e  $N_f$  é o total acumulado de PSMs verdadeiros, ambos com *MSGFScore* acima de um determinado valor *i*.

Para isso, usamos como base de dados verdadeira, o conjunto de todas as proteínas referência do genoma humano disponível no banco de dados UNIPROT ([www.uniprot.org](http://www.uniprot.org)). A base de dados de proteínas falsas foi criada com o *script* decoy.pl disponível na internet.

#### 5. RESULTADOS E DISCUSSÃO

Até o momento foram analisados 12 milhões de espectros de cólon do CPTAC, sendo 6 milhões tumorais e 6 milhões normais. Primeiramente definimos que na busca por proteínas alternativas apenas PSMs com FDR menor que 1% seriam selecionados para análises posteriores. Usando a estratégia *target-decoy* definimos em 78 e 88, os limiares (valores mínimos) de *MSGFScore* para os conjuntos de espectros de cólon normal e tumoral, respectivamente.

A aplicação destes limiares aos PSMs resultantes da busca de proteínas alternativas revelou PSMs significativos para 3 proteínas (HNRNPAB, SPTAN1 e FN1). Estes PSMs reportam, em cada gene, a exclusão de um

exon alternativo previamente descrito na literatura e nos bancos de dados referência. No presente trabalho mostramos que estas variantes sem o exon alternativo, foram encontradas exclusivamente na amostra de cólon tumoral.

## 6. CONCLUSÕES

Através da investigação de dados de espectrometria de massas este trabalho traz evidências de tradução para variantes de *splicing* de 3 genes HNRNPAB, SPTAN1 e FN1 em amostras de cólon. Assumindo uma taxa de falso-positivo menor do que 1%, observamos que as variantes dos genes HNRNPAB e FN1 foram detectadas nas amostras de cólon tumoral, mas não em cólon normal e que a variante do SPTAN1 provavelmente está mais expressa em adenocarcinoma de colon do que em colon normal. Vale ressaltar que variantes de *splicing* do gene FN1, geradas por *skipping* do exon 25, já foram descritas como diferencialmente expressas em adenocarcinoma de cólon em relação ao colon normal (10) . Portanto, nossos resultados sugerem que essas variantes de *splicing* dos genes HNRNPAB, SPTAN1 e FN1 são traduzidas em proteínas e representam potenciais marcadores moleculares para adenocarcinoma de cólon.

## 7. REFERÊNCIAS BIBLIOGRÁFICAS

1. Kim E, Goren A, Ast G. Alternative splicing: current perspectives. *Bioessays*. 2008;30(1):38-47.
2. Wang ET, Sandberg R, Luo S, Khrebtkova I, Zhang L, Mayr C, et al. Alternative isoform regulation in human tissue transcriptomes. *Nature*. 2008;456(7221):470-6.
3. Kelemen O, Convertini P, Zhang Z, Wen Y, Shen M, Falaleeva M, et al. Function of alternative splicing. *Gene*. 2013;514(1):1-30.
4. Pickrell JK, Pai AA, Gilad Y, Pritchard JK. Noisy splicing drives mRNA isoform diversity in human cells. *PLoS Genet*. 2010;6(12):e1001236.

5. Angel TE, Aryal UK, Hengel SM, Baker ES, Kelly RT, Robinson EW, et al. Mass spectrometry-based proteomics: existing capabilities and future directions. *Chemical Society reviews*. 2012;41(10):3912-28.
6. Lawrence RT, Perez EM, Hernandez D, Miller CP, Haas KM, Irie HY, et al. The proteomic landscape of triple-negative breast cancer. *Cell reports*. 2015;11(4):630-44.
7. Ramalho R, Li S, Radivojac P, Hahn M. Proteomic evidence for in-frame and out-of-frame alternatively spliced isoforms in human and mouse. *IEEE/ACM transactions on computational biology and bioinformatics / IEEE, ACM*. 2015.
8. Kim S, Pevzner PA. MS-GF+ makes progress towards a universal database search tool for proteomics. *Nature communications*. 2014;5:5277.
9. Elias JE, Gygi SP. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nature methods*. 2007;4(3):207-14.
10. Gardina PJ, Clark TA, Shimada B, Staples MK, Yang Q, Veitch J, et al. Alternative splicing and differential gene expression in colon cancer detected by a whole genome exon array. *BMC Genomics*. 2006;7:325.