

BIOMARCADORES EM CÂNCER DE PULMÃO, PRÓSTATA E MAMA¹

BIOMARKERS IN LUNG, PROSTATE AND BREAST CANCER

**Raul Javorsky da Costa², Carlos Eduardo Giuliani Baú³, Luiz Carlos Rodrigues Júnior⁴,
Cristhian Augusto Bugs⁵, Thiago Victorino Claus⁶ e Eder Maiquel Simão⁷**

RESUMO

O desenvolvimento do câncer ocorre através de danos no DNA. Segundo um estudo de levantamento estatístico, realizado pelo Instituto Nacional do Câncer, os tumores malignos mais frequentes são de pulmão, próstata e mama. O objetivo neste trabalho é verificar a expressão de biomarcadores em câncer e comparar esses resultados com a literatura. Neste trabalho, buscou-se, através do GEO2R, analisar o logaritmo da mudança de expressão de amostras de câncer de pulmão, próstata e mama, para localizar as proteínas significativamente expressas. Os genes mais expressos encontrados foram o *SPINK1*, *PRAC1*, *COL10A1*, *COMP*, *GJB2*, *MMP12* e *MMP1*. Esses genes estão relatados na literatura como estarem ativos em diferentes tipos de câncer, além de apresentarem funções, como degradação de colágeno e ligados ao transporte de substâncias através da parede celular.

Palavras-chave: GEO2R, microarranjos, proteínas, valor-p.

ABSTRACT

The development of cancer occurs through damage to the DNA. According to a statistical study conducted by the National Cancer Institute, the most common malignant tumors are lung, prostate and breast. The objective of this study is to assess the biomarker expression in cancer and to compare these results with the specialized literature. In this work, we sought, through GEO2R, to analyze the logarithm of the change of expression of lung, prostate and breast cancer samples to locate the significantly expressed proteins. The most expressed genes were SPINK1, PRAC1, COL10A1, COMP, GJB2, MMP12 and MMP1. These genes are reported in literature as being active in different types of cancer in addition to having functions such as collagen degradation and linked to the transport of substances through the cell wall.

Keywords: GEO2R, microarrays, protein, p-value.

¹ Trabalho de Iniciação Científica.

² Acadêmico do Curso de Física Médica - Centro Universitário Franciscano. E-mail: raul.javorskydacosta@gmail.com

³ Aluno de mestrado do Programa de Pós-graduação em Nanociências - Centro Universitário Franciscano. E-mail: contato.carlosbau@gmail.com

⁴ Colaborador. Professor do Programa de Pós-graduação em Nanociências - Centro Universitário Franciscano. E-mail: luizrdrgs@yahoo.com.br

⁵ Colaborador - Professor Pesquisador da Universidade Federal do Pampa. E-mail: cristhianbugs@gmail.com

⁶ Colaborador - Hospital Universitário de Santa Maria. E-mail: clausrx@gmail.com

⁷ Orientador - Centro Universitário Franciscano. E-mail: edersimao@gmail.com

INTRODUÇÃO

Os tumores são células que perderam a estabilidade genômica. A displasia é caracterizada como uma lesão celular reversível e pelo crescimento anormal de células que poderão evoluir ao câncer passando por três estágios: adenoma primário, intermediário e tardio (LUDWIG et al., 2014).

O desenvolvimento do câncer ocorre através do acúmulo de danos no DNA. Esses danos que ocorrem nas diferentes fases de desenvolvimento da doença, provocam a ativação de barreiras que impedem a evolução do câncer. A barreira ativa mecanismos de manutenção do genoma (MMG) conhecidos como reparo do DNA e morte celular programada. Se essa barreira não for ativada ocorre aumento dos danos e com isso, ativação de proteínas específicas envolvidas na proliferação celular. O acúmulo das mutações resultará em um tumor formado exclusivamente por uma massa de células malignas (SIMÃO et al., 2012).

A instabilidade genética é determinante para o desenvolvimento do câncer, percebendo-se assim a necessidade de análise dos processos pelos quais essa instabilidade se origina. A estabilidade das doenças genéticas se mantém devido à ativação dos MMG. A falha nesses mecanismos possibilita que as proteínas mais alteradas se ativem, provocando o desenvolvimento do câncer (MING-CHANG et al., 2009).

Estudos estatísticos, realizados pelo Instituto Nacional do Câncer (INCA), mostraram que o câncer de pulmão é um tumor prejudicial que acomete 1,2 milhões de vítimas por ano. Em 2014, obteve-se 16.400 casos de câncer de pulmão para os homens e 10.930 entre as mulheres. Cerca de 80% da ocorrência dessa doença é atribuída ao tabagismo, cuja fumaça do cigarro é caracterizada pela liberação de substâncias tóxicas, como o cianeto, o níquel e o benzeno (INCA, 2014).

O câncer de próstata é a causa mais comum de morte por câncer em homens com idade acima de 75 anos. Este tipo de câncer é raramente encontrado em homens com menos de 40 anos. As pessoas que estão no grupo de risco, incluem: homens afro-americanos, maiores de 60 anos e que têm um pai ou irmão com câncer de próstata. Ainda, o câncer de próstata como sendo a neoplasia maligna mais comum em brasileiros do sexo masculino, tem um risco estimado de aproximadamente 70 novos casos a cada 100 mil homens, para região sul representa 19,3% dos casos (INCA, 2014).

Dados da Agência Internacional de Pesquisa sobre o Câncer (IARC), mostraram que a incidência de câncer tem aumentado em todo o mundo. Com relação ao câncer de mama, as maiores taxas de incidência foram encontradas nos Estados Unidos e na Europa Ocidental, com 101 e 85 novos casos por 100.000 mulheres, respectivamente (INCA, 2014).

METODOLOGIA

AMOSTRAS DE MICROARRANJOS

Foram extraídas séries de amostras de microarranjos, do banco de dados GEO (*Gene Omnibus Expression*). Esse banco de dados armazena dados de expressão gênica de diversos tipos de doenças

genéticas, assim como o câncer. No banco são encontradas mais de 1 milhão de amostras de doenças. As amostras públicas no GEO podem ser obtidas por várias técnicas de análises de expressão gênica. Uma das técnicas é conhecida como microarranjos. Um *chip* de microarranjo é capaz de armazenar vários transcritos na forma de sondas, onde cada sonda representa um gene. Para esse trabalho foram utilizados microarranjos produzidos pela Affymetrix e pertencem a plataforma GPL570, contendo aproximadamente 54 mil sondas e 20 mil genes (GPL, 2003). Os microarranjos utilizados nesse trabalho estão listados abaixo:

- GSE27262: câncer de pulmão, contendo 25 amostras de tecido normal e 25 amostras de tecidos cancerosos. As amostras foram extraídas de pacientes com idades de 38 a 75 anos de ambos os sexos (GSE27262, 2013).
- GSE26910: câncer de mama e câncer de próstata, contendo 6 amostras de tecidos mamários normais, 6 tecidos mamários tumorais, 5 amostras de próstatas normais e 5 amostras de próstatas cancerosas (GSE26910, 2011).

ANÁLISES ESTATÍSTICAS

Para o desenvolvimento deste trabalho foi utilizado o aplicativo GEO2R. O aplicativo é uma ferramenta disponível em ambiente virtual (GEO2R, 2015), que permite comparar dois ou mais grupos de amostras (Tabela 1) de expressão gênica de uma série extraída do GEO. As análises foram baseadas em dois testes estatísticos: o logaritmo da mudança de expressão e o teste de significância estatística.

Tabela 1 - Grupos amostrais utilizados para calcular a diferenciação da expressão.

Série do GEO	Combinações (X_i)	N câncer*	N controle*
GSE27262	Câncer de pulmão x tecido normal	25	25
GSE26910	Câncer de próstata x tecido normal	6	6
	Câncer de mama x tecido normal	6	6

* N câncer, é o número de amostras de câncer e N controle é o número de amostras normais

Logaritmo da Mudança de Expressão

A análise estatística foi realizada através do logaritmo da mudança de expressão (logME), usada em grandes conjuntos de amostras, que tenham diferenças exponenciais entre a expressão dos genes presentes em microarranjos (RUIZ et al., 2004).

O método se baseia na identificação da expressão entre as amostras experimentais e de tecido normal. Para calcular a mudança de expressão de uma determinada combinação de amostras deve-se primeiramente agrupar as amostras de microarranjos através de uma média aritmética (SIMÃO, 2012).

Uma vez agrupada as amostras de tecidos pela média aritmética, deve-se calcular o logME, para cada combinação de amostras (Tabela 1). O logME consiste em calcular a razão da expressão para um determinado gene (i) alterado em uma doença genética (em relação ao seu valor controle). Esse método serve para identificar quantas vezes um gene está alterado em relação ao seu controle:

$$\log ME = \log_2 \left(\frac{X'_i}{X_i} \right)$$

Se for observado um valor de expressão para um gene em câncer (e essa expressão for igual ao seu valor normal), o logME será zero. Se, há um aumento da expressão do gene em câncer em relação ao normal, o logME será positivo e se o inverso for observado tem-se um logME negativo (GEO2R, 2015).

Teste de Significância

O valor-p é uma medida estatística usada para medir o nível de significância de um teste de hipótese, igual ou mais extremo que aquela observado em uma amostra. Esse método estatístico é usado em bioinformática para localizar os genes significativamente expressos em amostras de expressão gênica. O valor-p pode ser ajustado para evitar a presença de falsos positivos. Esse método é conhecido como ajuste do valor-p (adjP) (GEO2R, 2015). O valor-p é calculado através do teste Z. Esse teste corresponde a uma distribuição estatística realizada sobre um teste de hipótese que pode ser aproximada de uma distribuição normal (BLAIR; TAYLOR, 2013).

CRITÉRIOS DE SELEÇÃO

Para verificar os genes diferencialmente expressos e significativamente alterados foram utilizados alguns critérios de seleção. Os critérios de seleção permitem que sejam localizados nas combinações amostrais as diferenças de expressão entre os casos de câncer e tecido normal.

Os valores utilizados foram: logME: 2,0 e 3,0, que representam uma diferença de expressão de 4,0 e 8,0 vezes respectivamente entre o gene câncer (X'_i) *versus* o seu normal (X_i). Ou seja, um valor de logME igual a 2,0, que representa um gene i , está com expressão aumentada 4 vezes mais em câncer do que em normal. Para o corte dos valores significativos do adjP foi utilizado um valor de 5%.

RESULTADOS

As tabelas 2, 3 e 4, mostram os genes significativamente expressos para as combinações amostrais envolvendo os microarranjos de pulmão (Tabela 2), próstata (Tabela 3) e mama (Tabela 4). O valor de adjP considerado foi menor que 5% e os valores de logME considerados são maiores que 2,0.

Na tabela 2, estão listados os 76 genes diferencialmente expressos em câncer de pulmão. Dos genes com logME mais alterados observa-se o gene *SPINK1*. Esse gene é conhecido como um inibidor de serina peptidase encontrado em doenças inflamatórias do pâncreas. No pulmão ele é conhecido por produzir muco e sua função ainda não é conhecida (SAKATA et al., 2014). Segundo alguns autores o *SPINK1* é um biomarcador de várias doenças severas como, por exemplo, o câncer de pulmão, próstata e doenças relacionadas ao pâncreas (RINK et al., 2013).

A tabela 3 mostra os 12 genes diferencialmente expressos em câncer da próstata. O gene *PRAC1* é expresso na próstata, cólon e reto. De acordo com a literatura, ele pode desempenhar um papel regulador no núcleo da célula. A expressão da proteína *praca1* em câncer de próstata tem mostrado ainda uma heterogeneidade das células cancerosas em relação as células normais. Alguns resultados sugerem que o gene *PRAC1* pode atuar na patogênese e pode ser um biomarcador para o câncer da próstata (LENKA et al., 2013).

Tabela 2 - Resultado do logME para o câncer de pulmão.

GENE	logME	GENE	logME	GENE	logME	GENE	logME
SPINK1	4,04	UHRF1	2,81	CDH3	2,37	EFNA4	2,17
CST1	3,98	PPAP2C	2,76	BUB1B	2,33	UBE2C	2,14
MMP1	3,83	HABP2	2,73	TPX2	2,31	EPHX3	2,11
COL10A1	3,77	LINC01207	2,72	SGPP2	2,31	EEF1A2	2,11
CP	3,77	MMP7	2,66	KCNK5	2,30	BCL2L15	2,10
CTHRC1	3,60	KIF20A	2,57	CLIC6	2,29	KRT15	2,10
COMP	3,40	CD24	2,57	HMGB3	2,29	CRLF1	2,09
GREM1	3,28	UNC5CL	2,52	CDKN3	2,28	CENPK	2,09
CXCL13	3,15	UBE2T	2,51	E2F8	2,27	EZH2	2,08
MMP12	3,13	OCIAD2	2,47	CCNB1	2,26	KCNN4	2,08
GCNT3	3,09	TTK	2,47	NUSAP1	2,25	SERPIND1	2,06
LRRC31	3,02	ZWINT	2,47	ADAMDEC1	2,24	PROM1	2,06
HS6ST2	3,02	MELK	2,45	KIF11	2,23	NUF2	2,03
GJB2	2,97	CDC20	2,45	AGR2	2,22	PCP4	2,02
TMPRSS4	2,94	ANLN	2,43	LGSN	2,19	SLC22A18	2,02
CRABP2	2,88	CEP55	2,42	LOC100506918	2,19	AOC1	2,02
DLGAP5	2,85	GINS1	2,42	TM4SF4	2,19	KIF4A	2,00
ST6GALNAC1	2,84	C2CD4A	2,41	METTL7B	2,19	UGT8	2,00
RRM2	2,81	C1orf106	2,39	THBS2	2,18	KRT80	2,00

Tabela 3 - Resultado do logME para o câncer de próstata.

GENE	logME	GENE	logME	GENE	logME
PRAC1	4,00	LINC01207	3,12	GPR160	2,34
PLA1A	3,52	TRGC2	3,00	HIST1H1C	2,23
TARP	3,36	MAL2	2,60	TSPAN1	2,17

Os genes diferencialmente expressos nas análises feitas sobre os microarranjos do câncer de mama, mostraram 15 genes alterados. Na tabela 4, são apresentados os genes com valores de logME entre 5,38 e 2,19. O maior valor diferencial encontrado foi para o gene *COL10A1* (colágeno X). Esse gene foi identificado com expressão restrita na maioria dos tecidos normais e com expressão elevada

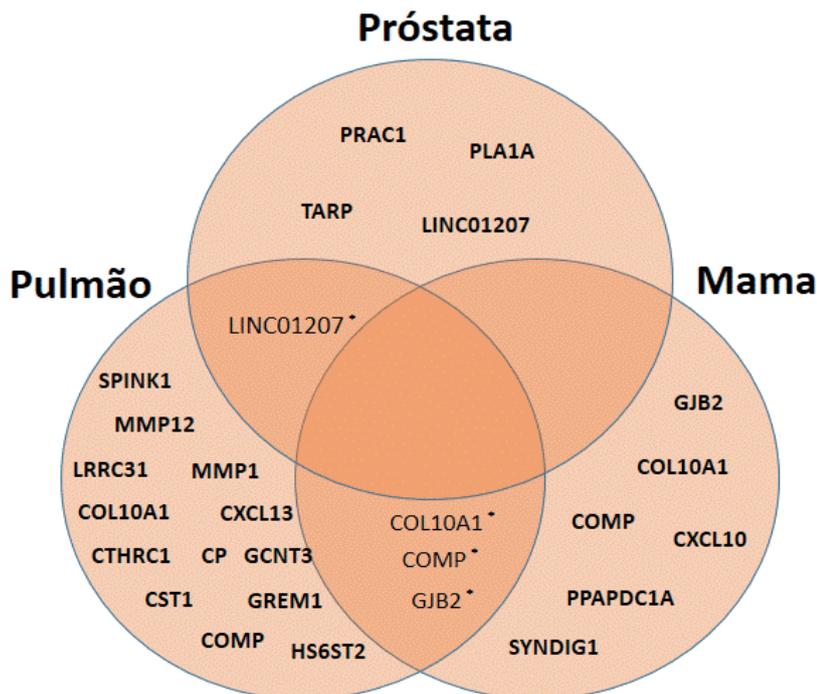
em diversos tipos de câncer. Ele também foi encontrado através de uma análise de imunofluorescência, na vascularização de tumores de mama enquanto que em tecidos normais os níveis de significância deste gene eram muito baixos. A expressão do gene colagênio X, pode facilitar o seu uso como um novo alvo para o diagnóstico e tratamento de diversos tipos de tumores sólidos (CHAPMAN et al., 2012).

Tabela 4 - Resultado do logME para o câncer de mama.

GENE	logME	GENE	logME	GENE	logME
COL10A1	5,38	SYNDIG1	3,11	NINJ2	2,37
COMP	4,85	LAMP5	2,99	SUGCT	2,30
PPAPDC1A	4,85	LOC101929122	2,81	C3orf80	2,24
GJB2	4,38	SULF1	2,73	SMCO4	2,21
CXCL10	3,89	NKX32	2,53	LOXL1	2,19

A análise dos genes diferencialmente expressos mencionados acima, possibilitou o desenvolvimento de arranjos entre as 3 doenças. O diagrama de venn da figura 1 mostra que há quatro genes (*) expressos em comum entre os cânceres de pulmão, próstata e mama. Os genes: *LINC01207*, *COL10A1*, *COMP* e *GJB2* possuem expressão diferencial acima de 2,0. O gene *LINC01207* é encontrado no cromossomo 4, os estudos ainda não mostraram a proteína que é codificada por esse gene e por esse motivo não há uma função conhecida. Com base nas tabelas anteriores observou-se que este gene aparece com expressão de 2,72 no câncer de pulmão e 3,12 em câncer de próstata.

Figura 1 - Gráfico de venn mostrando os genes diferencialmente expressos no consenso entre os cânceres de pulmão, próstata e mama (*). Os genes presentes em cada doença estão expressos com valores de logME superiores a 3,0.



CONCLUSÕES

Com as análises da mudança de expressão e teste de hipótese pode-se verificar a presença de alguns genes biomarcadores presentes nos cânceres de pulmão, próstata e mama. Um diferencial encontrado neste trabalho está relacionado aos seguintes biomarcadores: *SPINK1*, *PRAC1*, *COL10A1*, *COMP*, *GJB2*, *MMP12* e *MMP1*. Esses genes estão relatados na literatura como estarem ativos em diferentes tipos de câncer, além de apresentarem funções que não estão relacionadas aos principais mecanismos de reparação, proliferação e morte celular. Eles estão diretamente relacionados a degradação de colágeno e ao transporte de substâncias através da parede celular.

A análise de proteínas biomarcadores tem papel importante para o aprimoramento e desenvolvimento de novos dispositivos de tratamento em câncer. Uma das principais aplicações das nanociências e a nanotecnologias está voltada ao desenvolvimento de fármacos nanoencapsulados. Uma vez conhecidos os biomarcadores é possível desenvolver tecnologias aplicadas que permitem identificar e combater a expressão dessas proteínas de forma a minimizar os efeitos colaterais que aparecem no tratamento do câncer. Com o uso adequado dos “nanofármacos” os pacientes que sofrem desses tipos de câncer, poderão receber um tratamento personalizado afim de combater a proliferação das células tumorais e dos agentes patogênicos ligados a evolução do câncer.

REFERÊNCIAS

BLAIR, R. C.; TAYLOR, R. A. **Bioestatística para ciências da saúde**. São Paulo: PEARSON Editora, 2013.

CHAPMAN, K. B. et al. COL10A1 expression is elevated in diverse solid tumor types and is associated with tumor vasculature. **Future Oncol.**, v. 8, n. 8, p. 1031-1041, 2012.

GEO2R. **Gene Expression Omnibus for R**. 2015. Disponível em: <<http://1.usa.gov/1IYXcN4>>. Acesso em: 05 maio 2015.

GPL. **Affymetrix human genome U133 plus 2.0 array**. 2003. Disponível em: <<http://1.usa.gov/1TUwNR8>>. Acesso em: 20 abr. 2015.

GSE27262. **Gene expression profiling of non-small cell lung cancer in Taiwan**. 2013. Disponível em: <<http://1.usa.gov/1RtKfxI>>. Acesso em: 30 abr. 2015.

GSE26910. **Stromal molecular signatures of breast and prostatecancer**. 2011. Disponível em: <<http://1.usa.gov/1YmyqNW>>. Acesso em: 30 abr. 2015.

INCA – INSTITUTO NACIONAL DO CÂNCER. **Estatísticas do Câncer**. 2014. Disponível em: <<http://www1.inca.gov.br/vigilancia/>>. Acesso em: 18 maio 2015.

LENKA, G. et al. Aberrant expression of the PRAC gene in prostate câncer. **International Journal of Oncology**, v. 43, p. 1960-1966, 2013.

LUDWIG, J. K.; SANTOS, C. L.; SIMÃO, E. M. Radioterapia de glioblastoma: reorganização das vias de manutenção do genoma. **Scientia Plena**, Aracaju, v. 10, n. 07, p. 1-9, 2014.

MING-CHANG, C. et al. Genetics of brain fiber architecture and intellectual performance. **The Journal of Neuroscience**, v. 29, n. 7, p. 2212-2224, 2009.

RINK, M. et al. Loss of SPINK1 expression is associated with unfavorable outcomes in urothelial carcinoma of the bladder after radical cystectomy. **Urologic Oncology**, v. 31, p. 1716-1724, 2013.

RUIZ, C. et al. Differential gene expression analysis reveals activation of growth promoting signaling pathways by tenascin-C. **Cancer Research**, v. 64, p. 7377-7383, 2004.

SAKATA, K. et al. Generation and analysis of serine protease inhibitor kazal type 3-cre driver mice. **Exp. Anim**, v. 63, p. 45-53, 2014.

SIMÃO, E. M. **Dinâmica da transição pré-câncer para câncer**: estudo da expressão de vias de manutenção do genoma. 2012. 45f. Tese (Doutorado em Física) - Universidade Federal de Santa Maria, Santa Maria, 2012.

SIMÃO, E. M. et al. Induced genome maintenance pathways in pre-cancer tissues describe an anti-cancer barrier in tumor development. **Molecular BioSystems**, v. 8, p. 3003-3009, 2012.

XU, K. et al. Cartilage oligomeric matrix protein associates with granulin-epithelin precursor (GEP) and potentiates gep-stimulated chondrocyte proliferation. **J Biol Chem.**, v. 282, n. 15, p. 11347-11355, 2007.