

1.06.04 - Química Analítica

ANÁLISE ESPECTROQUÍMICA DE PLASMA SANGÜÍNEO ATRELADA À ANÁLISE MULTIVARIADA NO RASTREAMENTO DO CÂNCER DE MAMA

Ingrid M. Câmara¹, Daniel L.D. Freitas¹, Priscila P. Silva¹, Maria B.C. Alves², Camilo L.M. Morais³, Francis L. Martin⁴, Tirezah B.P. Lajus⁵, Kassio M.G. Lima⁶

¹ Estudante de Química da Universidade Federal do Rio Grande do Norte (UFRN)

² Estudante do Departamento de Biologia celular e Genética da UFRN

³ Estudante de Química da Universidade Central de Lancashire, UK

⁴ Professor da Escola de Farmácia e Ciências Biomédicas, UCLan, UK

⁵ Professora/Pesquisadora da UFRN – Depto de Genética e Biologia Celular

⁶ Professor/Pesquisador da UFRN - Instituto de Química/Orientador

Resumo

A mortalidade do câncer de mama pode ser reduzida por meio de programas de rastreamento; sua introdução em uma população saudável utilizando tecnologias novas e menos invasivas para a detecção precoce do câncer de mama é altamente desejável. Com isso, análises espectroquímicas combinadas a técnicas de classificação multivariada são usadas como uma ferramenta bioanalítica para um Programa de Rastreamento do Câncer de Mama, utilizando-se biópsia líquida na forma de amostras de plasma sanguíneo. Esta metodologia baseia-se na aquisição e análise da impressão digital espectroquímica de amostras de plasma por ATR-FTIR; os espectros derivados refletem a composição bioquímica intrínseca, gerando informações sobre ácidos nucleicos, carboidratos, lipídios e proteínas. Excelentes resultados em termos de sensibilidade (94%) e especificidade (91%) foram obtidos usando esse método em comparação com a mamografia (88-93% e 85-94%, respectivamente). Vantagens adicionais, como melhor prognóstico da doença, tratamento mais eficaz, menor morbidade associada, menos resultados falso-positivos e falso-negativos, menor custo e maior frequência analítica tornam esse método atraente para a instalação no ambiente clínico.

Autorização legal: 44113115.1.1001.5292.

Palavras-chave: Técnicas multivariadas; Quimiometria; Ambiente hospitalar.

Apoio financeiro: CNPq, PROPESQ/UFRN.

Trabalho selecionado para a JNIC: UFRN.

Introdução

O câncer de mama é a segunda mais comum e a principal causa de morte relacionada ao câncer entre as mulheres¹. Segundo o Sistema de Informação sobre Mortalidade (SIM), 14.206 mulheres morreram em 2013 devido a esta doença². Em 2014, a estimativa era de 49.240 casos e, em 2018, esperava-se atingir 59.700 novos apenas no Brasil¹. É um grande problema de saúde pública, levando em consideração os custos de detecção e tratamento^{3,4,5}. Apenas um em cada três casos de câncer de mama pode ser curado se descoberto em um estágio inicial² e não há maneiras eficazes de reduzir a incidência desta doença⁶. A melhor abordagem alternativa para combater o câncer de mama é o conceito de que quanto mais cedo a doença for detectada, mais eficaz será o tratamento. A detecção precoce através do rastreamento é o único método que provou ser eficaz na redução da mortalidade¹. Os programas de rastreamento são uma prática importante da política de saúde, onde a fase assintomática da doença é longa o suficiente para permitir a detecção direta ou indireta de lesões pré-cancerígenas. Tal teste de rastreamento que diagnostica doenças precoces precisa ser aceitável pelos pacientes e estar disponível a um custo razoável⁵.

A mamografia é o método recomendado para o rastreamento de rotina do câncer de mama em todo o mundo⁶. A exposição a essa radiação raramente causa câncer, a menos que seja realizada com uma frequência periódica alta pela qual o risco aumentará. Além de ser considerada dolorosa, relativamente cara e fonte de muito desconforto e até constrangimento para as pacientes, sua sensibilidade varia de 88% a 93%, enquanto sua especificidade varia de 85% a 94%⁶. Essas métricas estatísticas demonstram a proporção de mulheres com câncer de mama que apresentarão uma mamografia positiva sinalizando presença de doença e a taxa de mulheres sem câncer de mama que terão mamografia normal, respectivamente⁶. O tempo entre o exame inicial da paciente e o diagnóstico pode ser muito longo.

A Espectroscopia de Infravermelho é uma técnica vibracional capaz de analisar biomoléculas, como ácidos nucleicos, carboidratos, proteínas e lipídios, que apresentam características na região de IR⁷. A espectroscopia no infravermelho por transformada de Fourier no modo de reflexão total atenuada (ATR-FTIR)

vem sendo usada para analisar vários biofluidos devido à sua rápida aquisição espectral, mínima preparação e volume da amostra, e sua natureza não destrutiva para a amostra⁸. Pesquisas recentes estão progredindo gradualmente, em que excelentes resultados diagnósticos, comparados aos métodos tradicionais, foram obtidos em vários tipos de câncer como: ovário⁹, cervical¹⁰ e próstata¹¹; além disso, para diagnosticar doenças neurodegenerativas como o Alzheimer¹². Sendo assim, apresentamos os resultados do uso da espectroscopia ATR-FTIR juntamente com a quimiometria para classificação de pacientes com câncer de mama em um programa de rastreamento em larga escala utilizando plasma sanguíneo.

Metodologia

Neste estudo, avaliamos dois grupos de mulheres. O primeiro, Câncer de Mama (*Breast Cancer* (BC)), refere-se a um grupo de mulheres diagnosticadas com câncer de mama, com ou sem tratamento neoadjuvante, e foram coletadas por profissionais treinados no Hospital Liga Contra o Câncer (Natal/RN, Brasil), durante um período de dois anos. O segundo, Controle Saudáveis (*Healthy Controls* (HC)), refere-se a um grupo de mulheres sem diagnóstico prévio ou atual de câncer de mama, coletadas na Prontoclinica Dr. Paulo Gurgel (Natal/RN, Brasil), no mesmo período. As amostras de ambos os grupos foram obtidas após a leitura do Termo de Consentimento Livre e Esclarecido (Aprovação Ética: 44113115.1.1001.5292) e assinatura das pacientes. Em seguida, as amostras foram centrifugadas por 10 minutos e congeladas a aproximadamente -20°C até o momento da análise. Um total de 476 amostras foi obtido. As amostras foram removidas do freezer 15 minutos antes da análise para permitir o descongelamento. As amostras foram randomizadas e, para minimizar os efeitos temporais ou instrumentais, um número semelhante de amostras de ambos os grupos foi medido a cada dia. Os espectros de absorção foram obtidos utilizando um ATR-FTIR modelo IRAffinity-1S (Shimadzu Corp., Kyoto, Japão). Os espectros foram obtidos na faixa de 600 a 4000 cm⁻¹. Uma amostragem de 10 µL foi realizada. Este procedimento foi repetido em triplicata. O tempo de medição para cada amostra foi de aproximadamente 5 minutos.

Três espectros foram coletados por amostras e inicialmente, foi realizada a média destes. Em seguida, foi aplicado ao conjunto de dados os seguintes pré-processamentos: corte na região de impressão digital biológica (900-1800 cm⁻¹), suavização de Savitzky-Golay (SG) para remover ruídos aleatórios, correção de linha de base via mínimos quadrados automáticos ponderados e normalização para o pico de Amida I (1650 cm⁻¹). A importação de dados espectrais, pré-processamento e construção de modelos de classificação multivariados foram realizados no ambiente MATLAB R2014b versão 8.4 (MathWorks, Inc., Natick, EUA) com o PLS-Toolbox versão 7.9.3 (Eigenvector Research, Inc., Manson, EUA) e rotinas feitas em laboratório. As amostras foram divididas em três subconjuntos diferentes pelo algoritmo de seleção de amostras de Kennard-Stone (KS)¹³: treinamento (70%), validação (15%) e teste (15%).

A análise computacional consistiu em testar três algoritmos para extração e seleção de recursos: PCA¹⁴, SPA¹⁵ ou GA¹⁶; seguido de classificadores de análise discriminante: LDA¹⁷, QDA¹⁷ ou SVM¹⁸. Esses algoritmos foram acoplados como extração/seleção e classificação de recursos como: PCA-LDA, PCA-QDA e PCA-SVM; SPA-LDA, SPA-QDA e SPA-SVM; e GA-LDA, GA-QDA e GA-SVM.

Resultados e Discussão

Neste trabalho, a espectroscopia ATR-FTIR em conjunto com técnicas quimiométricas foi usada para detectar câncer de mama em um total de 476 pacientes recrutadas por dois anos para um programa de rastreamento de câncer de mama em estágio inicial em Natal, Brasil. A detecção de câncer de mama entre amostras normais foi realizada com sucesso com base nos espectros de plasma sanguíneo com 93% de acurácia (94% de sensibilidade, 91% de especificidade, AUC = 0,929) em um grupo externo de 71 pacientes, utilizando o algoritmo SPA-SVM.

Aplicações de plasma sanguíneo para detecção de câncer de mama foram realizadas por espectroscopia de infravermelho por Backhaus *et al.*¹⁹, onde 98% de sensibilidade e 95% de especificidade (usando análise de agrupamentos) e 92% de sensibilidade e 100% de especificidade (usando redes neurais artificiais (ANN)) foi obtido em um estudo realizado com 196 pacientes. Da mesma forma, Elmi *et al.*²⁰ detectou câncer de mama em espectroscopia de infravermelho com base em plasma sanguíneo com sensibilidade de 76% e especificidade de 72% para casos de câncer de mama usando a análise de componentes principais e análise discriminante linear (PCA-LDA) em um estudo com 86 amostras (43 câncer de mama, 43 controle saudáveis). Os resultados aqui relatados são mais altos, levando em consideração o grande número de pacientes, onde a sensibilidade e especificidade são maiores que 90%; sendo comparável aos resultados obtidos por métodos mais sofisticados, como o uso de imagens IR baseado em laser de cascata quântica, onde sensibilidade e especificidade foram relatadas em 94% e 86%, respectivamente, usando um classificador baseado em árvore de decisões aleatórias^{21,22}.

Conclusões

Os dados espectrais do FTIR na região de impressão digital (900-1800 cm^{-1}) foram pré-processados e os dados espectrais brutos e pré-processados, onde sobreposições visuais entre os espectros de câncer de mama e controle saudáveis estão presentes em toda a região espectral, indicando a necessidade de técnicas quimiométricas para distinguir amostras com matrizes tão complexas.

Entre as técnicas de classificação testadas, o SPA-SVM apresentou o melhor desempenho de classificação com acurácia de 92,9% (94% de sensibilidade e 91% de especificidade) para detectar amostras de câncer de mama com base em um conjunto de testes externo. Aproximadamente 70% das amostras ($n = 334$ pacientes) foram usadas para a construção do modelo e outros 15% para validação interna ($n = 71$ pacientes). O desempenho geral da classificação representado pelos valores do F-Score e G-Score foi bom (93%), indicando desempenho igual com ou sem considerar os dados desequilibrados. A melhor curva ROC (área sob a curva [AUC] = 0,929) foi encontrada para o SPA-SVM, indicando um excelente desempenho preditivo. PCA-SVM (AUC = 0,886) e GA-SVM (AUC = 0,871) foram, respectivamente, o segundo e o terceiro melhor algoritmo de classificação, demonstrando um bom desempenho de classificação.

No total, 16 números de onda (901, 959, 980, 999, 1018, 1277, 1364, 1402, 1464, 1489, 1582, 1311, 1626, 1643, 1661 e 1742 cm^{-1}) foram responsáveis pela diferenciação das classes na região de impressão digital biológica usando SPA-SVM. As tentativas bioquímicas dessas variáveis com base em Movasaghi *et al.*²³ são mostradas na tabela 1 abaixo.

Números de onda selecionados (cm^{-1})	Tentativa de atribuição
901	Fosfodiéster (absorbâncias devido ao colágeno e glicogênio)
959	Vibração simétrica de estiramento do n_1PO_4
980	OCH_3 (polissacarídeos)
999	Vibrações de estiramento de anel misturadas fortemente com CH
1018	$\text{n}(\text{CO}), \text{n}(\text{CC}), \text{d}(\text{OCH})$, anel (polissacarídeos, pectina)
1277	Modos vibracionais do colágeno
1311	Componentes de proteínas (banda de amida III)
1364	Estiramento C-O, deformação C-H, deformação N-H
1402	Modos simétricos de flexão de CH_3 dos grupos de metil de proteínas
1464	Modo de tesoura CH_2 da cadeia acil dos lipídios
1489	Vibração de flexão de CH no plano
1582	Estiramento do anel C-C de fenil
1626	Pico de ácidos nucleicos devido ao estiramento da base de carbonila e modo normal do anel
1643	Banda de amida I (surge das ligações de estiramento de C=O)
1661	$\text{n}(\text{C}=\text{O})_{\text{cis}}$ em lipídios e ácidos graxos
1742	C=O modo de estiramento de lipídios

Tabela 1. Números de onda selecionados pelo SPA-SVM para distinguir as amostras de câncer de mama e controle saudáveis.

Referências bibliográficas

1. BRASIL. Instituto Nacional de Câncer José Alencar Gomes da Silva/Ministério da Saúde. Estimativa 2018: incidência de câncer no Brasil. Rio de Janeiro: INCA, 2017. Disponível em <https://www.inca.gov.br/sites/ufu.sti.inca.local/files//media/document//estimativa-incidencia-de-cancer-no-brasil-2018.pdf> (acessado em 26/12/2019).
2. BRASIL. Instituto Nacional de Câncer José Alencar Gomes da Silva/Ministério da Saúde. Câncer de mama: é preciso falar disso. Rio de Janeiro: INCA, 2014. Disponível em <https://www.inca.gov.br/sites/ufu.sti.inca.local/files//media/document//cartilha-cancer-de-mama-vamos-falar-sobre-isso2014.pdf> (acessado em 26/12/2019).
3. Castro, R. Câncer na Mídia: uma Questão de Saúde Pública. *Revista Brasileira de Cancerologia* **55**, 41–48 (2009).
4. Facina, T. Estimativa 2014 – Incidência de Câncer no Brasil. *Revista Brasileira de Cancerologia* **60**, 63 (2014).
5. BRASIL. Plano de ações estratégicas para o enfrentamento das doenças crônicas não transmissíveis (DCNT) no Brasil 2011-2022. Brasília: Ministério de Saúde, 2011. Disponível em http://bvsmms.saude.gov.br/bvs/publicacoes/plano_acoes_enfrent_dcnt_2011.pdf (acessado em 26/12/2019).
6. BRASIL. Ministério da Saúde. Instituto Nacional de Câncer. Mamografia: da prática ao controle. Rio de Janeiro: INCA, 2007. http://bvsmms.saude.gov.br/bvs/publicacoes/qualidade_mamografia.pdf (acessado em 26/12/2019).
7. Baker, M. J. et al. Using Fourier transform IR spectroscopy to analyze biological materials. *Nat. Protoc.* **9**, 1771–1791, DOI: <https://doi.org/10.1038/nprot.2014.110> (2014).
8. Mitchell, A. L., Gajjar, K. B., Theophilou, G., Martin, F. L. & Martin-Hirsch, P. L. Vibrational spectroscopy of biofluids for disease screening or diagnosis: translation from the laboratory to a clinical setting. *J. Biophotonics* **7**, 153–165, DOI: <https://doi.org/10.1002/jbio.201400018> (2014).
9. Theophilou, G., Lima, K. M. G., Martin-Hirsch, P. L., Stringfellow, H. F. & Martin, F. L. ATR-FTIR spectroscopy coupled with chemometric analysis discriminates normal, borderline and malignant ovarian tissue: classifying subtypes of human cancer. *Analyst* **141**, 585–594, DOI: <https://doi.org/10.1039/C5AN00939A> (2016).
10. Neves, A. C. O., Silva, P. P., Morais, C. L. M., Miranda, C. G., Crispim, J. C. O. & Lima, K. M. G. ATR-FTIR and multivariate analysis as a screening tool for cervical cancer in women from northeast Brazil: a biospectroscopic approach. *RSC Adv.* **6**, 99648–99655, DOI: <https://doi.org/10.1039/C6RA21331F> (2016).
11. Siqueira, L. F. S. & Lima, K. M. G. A decade (2004 – 2014) of FTIR prostate cancer spectroscopy studies: An overview of recent advancements. *Trends Analyt. Chem.* **82**, 208–221, DOI: <https://doi.org/10.1016/j.trac.2016.05.028> (2016).
12. Paraskevaidi, M. et al. Differential diagnosis of Alzheimer's disease using spectrochemical analysis of blood. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E7929–E7938, DOI: <https://doi.org/10.1073/pnas.1701517114> (2017).
13. Kennard, R. W. & Stone, L. A. Computer Aided Design of Experiments. *Technometrics* **11**, 137–148, DOI: <https://doi.org/10.1080/00401706.1969.10490666> (1969).
14. Bro, R. & Smilde, A. K. Principal component analysis. *Anal. Methods* **6**, 2812–2831, DOI: <https://doi.org/10.1039/C3AY41907J> (2014).
15. Soares, S. F. C., Gomes, A. A., Araujo, M. C. U., Galvão Filho, A. R. & Galvão, R. K. H. The successive projections algorithm. *Trends Analyt. Chem.* **42**, 84–98, DOI: <https://doi.org/10.1016/j.trac.2012.09.006> (2013).
16. McCall, J. Genetic algorithms for modelling and optimisation. *J. Comput. Appl. Math.* **184**, 205–222, DOI: <https://doi.org/10.1016/j.cam.2004.07.034> (2005).
17. Morais, C. L. M. & Lima, K. M. G. Principal Component Analysis with Linear and Quadratic Discriminant Analysis for Identification of Cancer Samples Based on Mass Spectrometry. *J. Braz. Chem. Soc.* **29**, 472–481, DOI: <http://dx.doi.org/10.21577/0103-5053.20170159> (2018).
18. Cortes, C. & Vapnik, V. Support-Vector Networks. *Mach. Learn.* **20**, 273–297, DOI: <https://doi.org/10.1023/A:1022627411411> (1995).
19. Smith, R. A. et al. American Cancer Society Guidelines for Breast Cancer Screening: Update 2003. *CA Cancer J. Clin.* **53**, 141–169, DOI: <https://doi.org/10.3322/canjclin.53.3.141> (2003).
20. Elmi, F., Movaghar, A. F., Elmi, M. M., Alinezhad, H. & Nikbakhsh, N. Application of FT-IR spectroscopy on breast cancer serum analysis. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* **187**, 87–91, DOI: <https://doi.org/10.1016/j.saa.2017.06.021> (2017).
21. Pilling, M. J., Henderson, A. & Gardner, P. Quantum Cascade Laser Spectral Histopathology: Breast Cancer Diagnostics Using High Throughput Chemical Imaging. *Anal. Chem.* **89**, 7348–7355, DOI: <https://doi.org/10.1021/acs.analchem.7b00426> (2017).

22. Kuepper, C., Kallenbach-Thieltges, A., Juette, H. *et al.* Quantum Cascade Laser-Based Infrared Microscopy for Label-Free and Automated Cancer Classification in Tissue Sections. *Sci Rep* **8**, 7717 (2018). <https://doi.org/10.1038/s41598-018-26098-w>.
23. Movasaghi, Z., Rehman, S. & ur Rehman, I. Fourier Transform Infrared (FTIR) Spectroscopy of Biological Tissues. *Appl. Spectrosc. Rev.* **43**, 134–179, DOI: <https://doi.org/10.1080/05704920701829043> (2008).