TẠP CHÍ HÓA HỌC T. 53(1) 90-94 THÁNG 2 NĂM 2015
DOI: 10.15625/0866-7144.2015-0094

XÂY DỰNG MÔ HÌNH THỬ NGHIỆM THỐNG KÊ ĐA BIẾN NHẬN BIẾT UNG THƯ ĐẠI TRỰC TRÀNG DỰA TRÊN DỮ LIỆU PHỔ NMR MẪU NƯỚC TIỂU

Nguyễn Tiến Tài[1*], Vũ Thị Kim Ngọc[2], Vũ Hùng Thái[2], Lê Thành Tụ[3], Nguyễn Quang Tâm[1]

[1]*Viện Hóa học, Viện Hàn lâm Khoa học và Công nghệ Việt Nam*

[2]*Trung tâm Dịch vụ Phân tích Thí nghiệm Thành phố Hồ Chí Minh*

[3]*Phòng Thí nghiệm phân tích trung tâm, Trường Đại học Khoa học tự nhiên Thành phố Hồ Chí Minh*

Đến Tòa soạn 20-10-2014; Chấp nhận đăng 22-12-2014

Abstract

Statistical models for Colorectal Cancer (CRC) identification were built by metabolomics, based on the ^{1}H NMR data of urine and used statistical multivariate analysis (MVA) tools, named principle component analysis (PCA) and partial least square (PLS). Demonstration of PCA and PLS application on urine NMR data of 17 cases and 41 controls collected at MEDIC centre (HoChiMinh city) showed remarkable differences between CRC cases and controls; Tryptophan and Glutamate were profiled as metabolic CRC markers.

Keywords. Metabolomics, CRC, 1H NMR, statistics, MVA, PCA, PLS.

1. MỞ ĐẦU

Phương pháp sàng lọc chuyển hóa chất (metabolomics) phát triển mạnh từ những năm đầu thế kỷ này và đã ứng dụng thành công trong nhiều lĩnh vực, trong đó có y dược [1]. Các nghiên cứu sàng lọc chuyển hoá chất trên động vật và con người được thực hiện theo nguyên tắc so sánh vân tay bằng cách phân tích dịch sinh học nhờ các công cụ phổ có độ phân giải cao, sau đó dữ liệu phổ được xử lý bằng các công cụ phân tích đa biến (Multivariate Analysis - MVA) thích hợp [2].

Các dạng bệnh ung thư có liên quan rất chặt chẽ và nhạy cảm với chuyển hóa chất, do vậy phát triển ứng dụng sàng lọc chuyển hóa chất trong nghiên cứu ung thư là cách tiếp cận tự nhiên và hợp lý [3]. Ung thư đại trực tràng (Colorectal Cancer - CRC) là căn bệnh ung thư rất phổ biến trên thế giới với tỷ lệ tử vong cao. Hiện tại các bệnh viện thường phát hiện CRC bằng nội soi trực tràng, sinh thiết tổn thương ruột, kiểm tra máu trong phân hoặc kiểm tra các kháng nguyên ác tính đặc thù cho CRC.

Tổng hợp thông tin từ các công bố gần đây nghiên cứu CRC dựa trên sàng lọc chuyển hóa chất được cho trên bảng 1. Thống kê cho thấy gần như toàn bộ các nghiên cứu CRC bằng sàng lọc chuyển hóa chất đều được thực hiện trên người, rất ít trên động vật; Bốn dạng dịch sinh học được lấy phân tích là tế bào, máu, nước tiểu và phân; Các công cụ phổ được sử dụng là NMR, GC-MS, LC-MS, CE-MS và FT ICR-MS; Dữ liệu phổ được xử lý bằng các công cụ thống kê phân loại biến, phân tích thành phần chính (Principal Component Analysis - PCA), bình phương tối thiểu riêng phần (Partial Least Squares - PLS) hay thống kê R; Cách thức tiến hành nghiên cứu là so sánh dữ liệu phổ dịch sinh học của bệnh nhân CRC và đối chứng khoẻ mạnh, so sánh trước và sau khi phẫu thuật hay hóa trị, xạ trị, so sánh các giai đoạn phát triển CRC, so sánh ảnh hưởng của các tác nhân nghi ngờ liên quan tới CRC, đặc biệt là khói thuốc lá.

CRC cũng được đánh giá là căn bệnh ung thư phổ biến và có tỷ lệ tử vong cao ở Việt Nam, đứng hàng thứ 3 về tỷ lệ tử vong. Việc chẩn đoán CRC tại các bệnh viện ở Việt Nam vẫn chủ yếu dựa trên nội soi trực tràng và sinh thiết tổn thương trực tràng. Việc ứng dụng phương pháp sàng lọc chuyển hóa chất trong lĩnh vực y học tại Việt Nam thực tế mới bắt đầu với các nội dung và kết quả mang tính thăm dò trên đối tượng bệnh nhân nhi [10].

Mục đích công trình này là thử nghiệm xây dựng mô hình thống kê chẩn đoán CRC bằng phương pháp sàng lọc chuyển hóa chất dựa trên số liệu phổ ^{1}H NMR các mẫu nước tiểu người Việt Nam, định hướng tới ứng dụng phân loại, nhận biết, chẩn đoán sớm CRC, đánh giá hiệu quả các giải pháp điều trị

CRC, nghiên cứu ảnh hưởng các yếu tố môi trường, dinh dưỡng tới sự hình thành và phát triển CRC.

Bảng 1: Tổng hợp thông tin về một số công bố sàng lọc CRC dựa trên nghiên cứu chuyển hóa chất

Tài liệu	Mẫu phân tích			Công cụ	
	Loại mẫu	Số mẫu bệnh	Số mẫu đối chứng	Phổ	Thống kê
[4] (2010)	Nước tiểu	6	6	HSQC NMR	PCA
[5] (2011)	Tế bào	31	32	GC TOF	PLS, PCA
		22	25	NMR	
[6] (2012)	Huyết thanh	153	139	^{1}H NMR	PLS
[7] (2014)	Huyết thanh	832	295	MALDY TOF	PCA
[8] (2014)	Phân	255	254	UHPLC-MS/MS	Phân loại biến
[9] (2014)	Huyết thanh	48	102	HPLC, MS-MS	Thống kê R

2. THỰC NGHIỆM

2.1. Lấy và xử lý mẫu nước tiểu

Mẫu nước tiểu kèm theo các thông tin liên quan tới đối tượng lấy mẫu khám nghiệm, như tên, tuổi, giới tính, nơi cư ngụ, tiền căn cá nhân và gia đình được lấy tại Trung tâm Chẩn đoán Y khoa MEDIC thành phố Hồ Chí Minh trong thời gian từ tháng 2 đến tháng 7 năm 2014 theo quy trình lấy mẫu nước tiểu sạch giữa dòng vào buổi sáng sớm. Việc phân loại người mắc bệnh CRC dựa trên kết quả nội soi đại trực tràng kết hợp sinh thiết tổn thương trực tràng. Số lượng mẫu nước tiểu: 58 mẫu, bao gồm 17 mẫu của bệnh nhân CRC và 41 mẫu đối chứng. Các mẫu nước tiểu được đánh số kèm theo chữ cái C phía trước đối với các mẫu bệnh nhân CRC.

Các mẫu nước tiểu được lưu giữ trong điều kiện đông lạnh (–80^{o}C) trước khi xử lý theo đợt. Quy trình chuẩn bị mẫu nước tiểu cũng như đo phổ NMR cho xử lý thống kê đa biến có một số điểm khác biệt so với phân tích phổ nước tiểu thông thường [10]: Chỉ cần lọc sơ bộ mẫu nước tiểu để loại bỏ cặn rắn, không cần tách chất và làm sạch nhưng việc chuẩn bị, xử lý mẫu cũng như chế độ đo phổ NMR phải cố gắng để đạt được sự tương đương tối đa về nồng độ mẫu và chất lượng phổ giữa toàn bộ các mẫu và phổ. Quy trình xử lý gồm các bước: Rã đông mẫu nước tiểu đến nhiệt độ phòng trong 20÷30 phút; Ly tâm 6.000 vòng/phút trong 5÷10 phút để tách cặn rắn. Chuyển 540 µl lớp dung dịch trong ở phía trên cho vào ống nghiệm 1.500 µl. Bổ sung 60 µl dung dịch đệm 1,5 M KH_2PO_4 trong D_2O (pH 7,4) có chứa chất chuẩn TSP 0,1 % 3-(trimetylsilyl)-1-(propanesulfonate). Bổ sung 0,1 mM Sodium azide (NaN_3) để diệt khuẩn. Dung dịch sau xử lý được lắc đều rồi rút cho vào ống đo NMR 5mm sạch, đậy nắp, dán kín bằng băng paraphin có đánh số, chờ chạy phổ NMR.

2.2. Đo và xử lý thống kê phổ 1H-NMR

Phổ ^{1}H NMR được ghi trên máy Bruker AVANCE III 500 (500 MHz) tại Phòng Thí nghiệm phân tích trung tâm, Trường Đại học Khoa học Tự nhiên thành phố Hồ Chí Minh, sử dụng đầu dò dải rộng 5 mm thu gián tiếp (BBI), chương trình xung: NOESY-1D có xung khử tín hiệu nước, số lần quét tích phổ NS = 8, độ rộng xung proton 8,5 µs, đo ở nhiệt độ phòng.

Dữ liệu phổ ^{1}H NMR mẫu nước tiểu được thử nghiệm bằng 2 công cụ thống kê đa biến là PCA và PLS. Trong các phân tích sàng lọc thống kê với mục tiêu trợ giúp chẩn đoán y học, PCA là phương pháp phân tích thống kê không định hướng (untargeted, unsupevised), cho phép phân loại các kết quả quân tích theo mức độ giống nhau hay khác nhau về đặc trưng phổ một cách khách quan, không áp đặt, trong khi đó PLS là phương pháp phân tích thống kê có định hướng (targeted, supevised), nhằm đánh giá xem dữ liệu phổ của một hay vài mẫu phân tích nào đó có thuộc một mô hình thống kê đã được xây dựng trước hay không. PCA hay PLS được sử dụng tuỳ thuộc mục đích phân tích, nhưng cũng thường được sử dụng đồng thời để bổ sung, hỗ trợ thông tin cho nhau.

Để xử lý thống kê PCA và PLS, trước tiên phổ ^{1}H NMR được chia thành các biến (variation) hay còn gọi là các bucket, là các khoảng phổ được chia nhỏ theo thang độ dịch chuyển hóa học (ppm). Trong nghiên cứu này, việc chia biến được thực hiện theo các thông số và thuật toán sau: Vùng phổ 9,50÷0,5 ppm được chia đều thành 180 bucket rộng 0,05 ppm, trong đó loại bỏ 28 bucket vùng tín hiệu nước và Urea (6,00÷4,55 ppm). Giá trị của các biến được xác định theo quy tắc tích phân chữ nhật đơn giản (Simple rectangular).

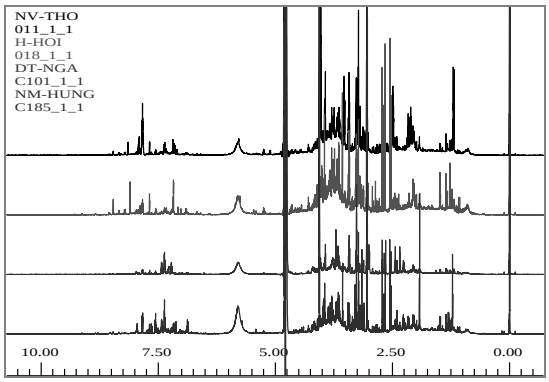
PCA được tiến hành theo chế độ "Không chuẩn hóa các biến" (No scaling of bucket variables) để thể hiện được nguyên trạng nồng độ tự nhiên các chất chuyển hóa trong từng mẫu. Thông số phân tích PLS: 58 phổ ^{1}H-NMR được gán nhãn (labeled) 0 và 1, tương ứng với phổ của mẫu nước tiểu bệnh nhân và mẫu khỏe mạnh đối chứng. Việc xác định các

chất chuyển hóa tương ứng với các đỉnh phổ NMR được thực hiện dựa theo công bố [11], kết hợp với phần mềm Chenomx NMR Suite 7.0.

Kết quả chính nhận được từ PCA được thể hiện trên giản đồ Score với sự phân bố thành nhóm của các điểm phổ có đặc trưng phổ gần giống nhau. Các butket đặc thù tương hỗ với các nhóm điểm phổ trên giản đồ Score được chỉ ra trên giản đồ Loading tương ứng. Kết hợp giản đồ Score và Loading cho phép chúng ta phân loại các phổ thực nghiệm một cách khách quan thành các nhóm gần nhau về đặc trưng phổ và xác định được đặc trưng phổ của mỗi nhóm phổ.

3. KẾT QUẢ VÀ THẢO LUẬN

Phổ ^{1}H NMR của 02 mẫu nước tiểu bệnh nhân CRC và 02 mẫu nước tiểu đối chứng được nêu trên hình 1. Dạng phổ trên hình rất đặc trưng cho phổ ^{1}H NMR của mẫu nước tiểu nguyên khai, với các pic cộng hưởng điển hình ứng với các chất chuyển hoá như Creatinine (4,08 và 3,08 ppm), Urea (5,78 ppm), Trimethylamine–N–oxide (3,25 ppm), Alanine (1,47-3,77 ppm), Hippurate (7,54-7,67-7,83 ppm) [11]. Cũng có thể nhận ra sự khác biệt giữa phổ của mẫu nước tiểu bệnh nhân CRC và đối chứng khi trực tiếp so sánh dữ liệu phổ trên hình 1, nhưng sẽ rất khó để rút ra kết luận định tính hay định lượng có tính hệ thống và quy luật đối với lượng rất lớn chất chuyển hoá từ hàng chục tới hàng trăm phổ NMR.



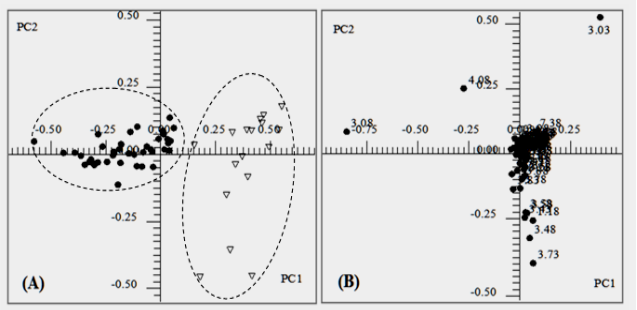
Hình 1: Phổ ^{1}H NMR của 02 mẫu nước tiểu đối chứng (011, 018) và 02 mẫu nước tiểu bệnh nhân CRC (C101, C185)

3.1. Sử dụng công cụ thống kê PLS sàng lọc dữ liệu phổ ^{1}HNMR bệnh nhân CRC

Kết quả PCA dữ liệu phổ ^{1}H NMR của 58 mẫu nước tiểu bệnh nhân CRC và đối chứng cho thấy, số thành phần chính tối thiểu để đạt độ tin cậy 95% là 4 PC (PC1 đến PC4), trong đó trọng số giải thích của 2 thành phần chính PC1 và PC2 tương ứng là 59,30 % và 9,94 %.

Trên giản đồ Score (hình 2A), 41 điểm phổ của các mẫu đối chứng (ký hiệu hình tròn) phân bố về phía trái của trục PC1 (từ −5,08 tới 0,08) và tương đối chụm theo PC2 (từ −0,13 tới 0,10) trong khi các điểm phổ của 17 mẫu CRC (ký hiệu tam giác) phân bố ở phía phải của trục PC1 (từ −0,15 tới 0,57) và phân tán tương đối rộng theo PC2 (từ −0,48 tới 0,25), tạo thành 2 nhóm điểm phổ tách biệt nhau rõ rệt.

Kết hợp kết quả phân tích giản đồ Score PC1-PC2 với kết quả phân tích giản đồ Loading (hình 2B) cho phép chúng ta xác định một số chất chuyển hoá đặc thù cho nhóm CRC và nhóm đối chứng. Cụ thể, các bucket tương ứng tới các pic cộng hưởng của Creatinine (4,08 và 3,08 ppm) là đặc trưng cho nhóm phổ đối chứng trong khi các bucket tương ứng với các chất chuyển hóa Tryptophan và Glutamate là đặc trưng cho nhóm phổ CRC. Kết quả sơ bộ này hoàn toàn phù hợp với kết luận từ các công bố [12, 13]. Cũng cần lưu ý rằng: Khi xem xét, đánh giá các điểm phổ lệch xa khỏi phân bố chung và các bucket liên quan tương ứng cần phải chú ý tới thông tin về chế độ dinh dưỡng (ăn theo chế độ kiêng, nghiện) hay chế độ hoá trị, là các yếu tố có thể ảnh hưởng mạnh tức thời tới thành phần chất trao đổi trong cơ thể nói chung, trong nước tiểu nói riêng [8].



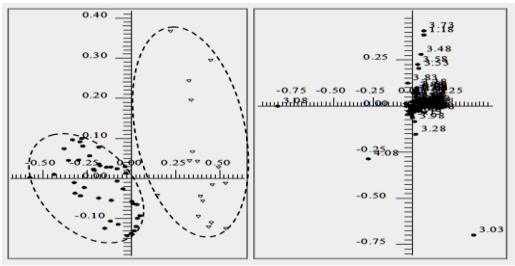
Hình 1: Kết quả PCA của 58 mẫu nước tiểu. A: Giản đồ Score PC1-PC2 (hình 2A); B: Giản đồ Loading 2 chiều PC1-PC2

3.2. Sử dụng công cụ thống kê PLS sàng lọc dữ liệu phổ ^{1}HNMR bệnh nhân CRC

Đối với phương pháp PLS, một hay vài biến số bổ sung được đưa thêm vào, như nồng độ chất chuyển hóa, pH hoặc là thông tin về tình trạng của đối tượng mẫu (giới tính, bệnh tính, …). Trong nghiên cứu này, biến số bổ sung, ký hiệu Y là trạng thái bệnh, nhận 2 giá trị "0" và "1", tương ứng là mẫu mắc CRC và mẫu đối chứng.

Kết quả PLS dữ liệu phổ ^{1}H NMR của 17 mẫu nước tiểu bệnh nhân CRC và 41 mẫu nước tiểu đối

chứng cho thấy, số thành phần chính tối thiểu để đạt độ tin cậy 90% là 4 PC, các điểm phổ tách thành 2 nhóm rất rõ rệt, tương ứng với 2 trạng thái mắc CRC và khỏa mạnh (đối chứng).



Hình 3: Kết quả PLS đối với dữ liệu phổ 1H NMR của 58 mẫu nước tiểu; A – Giản đồ Score PC1-PC2, B- Giản đồ Loading 2 chiều PC1-PC2

Phương pháp PLS có nhiều điểm giống với PCA. Sự khác biệt quan trọng nhất là PLS xây dựng mô hình thống kê một cách áp đặt, có định hướng dựa trên dữ liệu phổ của bệnh nhân, vì vậy, khi mô hình đã được xây dựng với độ tin tưởng chấp nhận được, chúng ta sẽ có thể đánh giá phổ của một hay các đối tượng mới xem có thuộc mô hình không, cách bao xa so với tâm mô hình. Độ tin cậy của mô hình sẽ càng cao khi lượng mẫu thống kê càng nhiều.

4. KẾT LUẬN

1. Kết quả sơ bộ xây dựng và ứng dụng phương pháp sàng lọc chuyển hoá chất dựa trên xử lý thống kê đa biến dữ liệu phổ ^{1}H NMR mẫu nước tiểu không tách chiết, để xây dựng mô hình thống kê nhận biết CRC cho thấy tính khả thi và triển vọng ứng dụng cao của phương pháp.

2. Còn nhiều nội dung và khả năng cần thực hiện trong các bước tiếp theo để hoàn thiện và nâng cao hiệu quả của mô hình thử nghiệm chẩn đoán CRC như: Chuẩn hóa quy trình thu và xử lý mẫu nước tiểu, chuẩn hóa các thông số và điều kiện đo phổ ^{1}H NMR, thử nghiệm các công cụ và phần mềm thống kê đa biến cập nhật khác, tăng số lượng mẫu thống kê.

3. Những kết quả ban đầu này mới mang tính kỹ thuật, nhằm xác lập và chuẩn hóa quy trình sàng lọc chuyển hóa chất, chưa có ý nghĩa chẩn đoán y học cụ thể, nhưng mở ra triển vọng ứng dụng trên thực tế, mở rộng cho chẩn đoán các căn bệnh liên quan nhiều tới chuyển hóa chất.

Lời cám ơn. *Các tác giả trân trọng cảm ơn sự tài trợ của Sở Khoa học & Công nghệ Thành phố HCM thông qua đề tài Nghiên cứu ứng dụng Cộng hưởng từ hạt nhân (NMR) phân tích nước tiểu bệnh nhân ung thư đại-trực tràng tại Việt Nam, mã số YT_2013_38 và sự hỗ trợ kỹ thuật của hãng Bruker-BioSpin-Asia Pacific.*

TÀI LIỆU THAM KHẢO

1. Robert Powers. *The Current State of Drug Discovery and a Potential Role for NMR Metabolomics*, J. Med. Chem., **57**, 5860-5870 (2014).
2. Jeremy K. Nicholson and John C. Lindon. *Metabonomics*, Nature, **453**, 1054-1056 (2008).
3. Richard D. Beger,.*A Review of Applications of Metabolomics in Cancer*, Metabolites, **3**, 552-574 (2013).
4. Young Kee Chae, Woo-Young Kang, Seong Hwan Kim, Jong Eun Joo, Joon Kil Han, Boo Whan Hong. *Metabolic Profiling of Urine Samples from Colorectal Cancer Patients Before and After Surgical Treatments*, Journal of the Korean Magnetic Resonance Society, **14**, 28-37 (2010).
5. Chan EC, Koh PK, Mal M, et al. *Metabolic profiling of human colorectal cancer using high resolution MAS nuclear magnetic resonance spectroscopy and gas chromatography mass spectrometry (GC/MS)*, J. Proteome Res, **8**, 352-61 (2009).
6. Ivano Bertini, Stefano Cacciatore, Benny V. Jensen, Jakob V. Schou, Julia S. Johansen, Mogens Kruhøffer, Claudio Luchinat, Dorte L. Nielsen and Paola Turano. *Metabolomic NMR Fingerprinting to Identify and Predict Survival of Patients with Metastatic Colorectal Cancer*, Cancer Res, **72(1)**, 356-364 (2012).
7. Jun Hwa Lee, Kyung-Hee Kim, Ji-Won Park, Hee Jin Chang, Byung Chang Kim, Sun Young Kim, Kwang Gi Kim, Eun Sook Lee, Dae Yong Kim, Jae Hwan Oh, Byong Chul Yoo and In-Hoo Kim. *Low-mass-ion discriminant equation: A new concept for colorectal cancer screening*, Int. J. Cancer, **134**, 1844-1853 (2014).
8. Amanda J. Cross, Simina Boca, Neal D. Freedman, Neil E. Caporaso, Wen-Yi Huang, Rashmi Sinha, Joshua N. Sampson and Steven C. Moore. *Metabolites of tobacco smoking and colorectal cancer risk*, Carcinogenesis, **35(7)**, 1516-1522 (2014).
9. Mirnezami, R.; Jimenez, B.; Li, J. V., Kinross, J. M., Veselkov, K., Goldin, R. D., Holmes, E., Nicholson, J. K., Darzi, A. *Rapid diagnosis and staging of colorectal cancer via HR-MAS NMR spectroscopy of intact tissue biopsies*, Ann. Surg., **259(6)**, 1138-1149 (2014).
10. Trần Thị Thanh Thủy, Thành Thị Thu Thủy, Nguyễn Tiến Tài, Đặng Vũ Lương. *Biomaker detection of inborn error of metabolism in baby urine and serum based on 1H-NMR spectroscopy*, Journal of Chemistry, **49(5)**, 639-642 (2011).

11. Souhaila Bouatra, Farid Aziat, David S. Wishart. *The Human Urine Metabolome*, PLoS ONE, **8(9)**, e73076 (2013).
12. Y. Qiu, G. Cai, M. Su, T. Chen, Y. Liu, Y. Xu, Y. Ni, A. Zhao, S. Cai, L. X. Xu, W. Jia. *Urinary metabonomic study on colorectal cancer*, J. Proteome. Res., **9(3)**, 1627-1634 (2010).
13. Doreen Jia Yi Ng et al. *Trend analysis of metabolomics and systematic review of metabonomics-derived cancer marker metabolites,* Metabolomics, **7**, 155-178 (2011).

Liên hệ: **Nguyễn Tiến Tài**
Viện Hóa học
Viện Hàn lâm Khoa học và Công nghệ Việt Nam
Số 18, Hoàng Quốc Việt, Cầu Giấy, Hà Nội
E-mail: nttai@ich.vast.ac.vn.