

Xây dựng phương pháp xác định kiểu gene rs7107217 và bước đầu xác định mối tương quan của gene này với nguy cơ ung thư vú ở phụ nữ Việt Nam

Nguyễn Thị Ngọc Thanh*, Mai Thị Ngọc Giàu, Nguyễn Thị Huệ

Tóm tắt—Ung thư vú là loại ung thư thường gặp nhất ở phụ nữ trên toàn thế giới. Việc xuất hiện các điểm đa hình ngay trên gene hoặc các vùng gần các gene nhạy cảm với ung thư vú có thể ảnh hưởng đến sự điều hòa biểu hiện của các gene này, từ đó làm tăng hoặc giảm nguy cơ ung thư vú. BARX2 được tìm thấy điều hòa biểu hiện của gene ERS1 từ đó có thể tham gia vào sự hình thành ung thư vú. Điểm đa hình rs7107217, nằm cách gene BARX2 152kb về phía hạ nguồn, có khả năng ảnh hưởng đến biểu hiện của BARX2 và đã được chứng minh có liên quan với nguy cơ ung thư vú ở các quần thể gần với người Việt Nam là Trung Quốc và Hàn Quốc. Trong nghiên cứu này, rs7107217 được khảo sát tần số và bước đầu xác định mối liên quan với nguy cơ ung thư vú ở người Việt Nam. Phương pháp Real-time PCR HRM được tối ưu điều kiện và được sử dụng để khảo sát tần số kiểu gene của rs7107217 trong 117 phụ nữ ung thư vú và 105 phụ nữ khỏe mạnh. Từ đó, mối tương quan của SNP này với nguy cơ ung thư vú bước đầu được xác định bằng phân tích sự khác biệt trong tần số allele và kiểu gene giữa nhóm bệnh và nhóm đối chứng. Kết quả cho thấy phương pháp xác định kiểu gene của rs7107217 đã được xây dựng thành công với độ nhạy, độ đặc hiệu và độ ổn định cao. SNP rs7107217 có độ đa hình cao với tần số allele thiếu số C chiếm 29,9% và 35,3% lần lượt trong nhóm bệnh và đối chứng. SNP rs7107217 chưa cho thấy sự liên quan (C vs A: $P = 0,23$, OR (95% CI) = 0,79 (0,53 – 1,17)) với nguy cơ ung thư vú. Tuy nhiên với độ tin cậy của phân tích thấp (11,71%) và tiềm năng cao liên quan đến sự hình thành ung thư vú, mối liên quan của rs7107217 với nguy cơ ung thư vú ở người Việt Nam cần được tiếp tục thực hiện trên cỡ mẫu lớn hơn để đạt độ tin cậy của phân tích cao hơn.

Từ khóa—rs7107217, ung thư vú, Việt Nam, phương pháp xác định kiểu gen, High Resolution Melting

1. GIỚI THIỆU

Ung thư vú là ung thư phổ biến nhất ở phụ nữ trên toàn thế giới. Có nhiều nguyên nhân ảnh hưởng đến nguy cơ ung thư vú liên quan với các yếu tố ngoại sinh như béo phì, lượng estrogen cao, lượng progesterone cao, và tăng cân khi sinh [1, 2]; và các yếu tố nội sinh như các yếu tố di truyền [3]. Trong đó, yếu tố di truyền mặc dù chỉ chiếm tỉ lệ nhỏ nhưng lại rất có ý nghĩa trong tiên lượng và chẩn đoán sớm ung thư vú [4]. Nghiên cứu trước đây đã xác định rằng sự phát triển của khối u vú thông qua sự tích tụ các biến thể di truyền [5], một số được biểu hiện ở giai đoạn sớm [6]. Dựa vào nguy cơ mắc bệnh và tần số xuất hiện trong quần thể, những biến thể này có thể được phân thành ba nhóm chi thị di truyền [7]. Trong số đó, nhóm gene nhạy cảm thấp là nhóm gene có tần số xuất hiện biến dị di truyền cao trong quần thể, đặc biệt là các điểm đa hình (Single Nucleotide Polymorphisms (SNPs)) [8]. Các điểm đa hình này có tác động riêng lẻ là rất nhỏ, tuy nhiên khi kết hợp với nhau, hoặc với các biến đổi khác trên gene nhạy cảm cao với ung thư vú, các điểm đa hình này có khả năng làm tăng đáng kể nguy cơ ung thư [9]. Chính vì thế, các SNP đang tiềm ẩn trong quần thể này có tiềm năng rất lớn để trở thành các chi thị di truyền quần thể đặc trưng cho ung thư vú [10].

Ngày nhận bản thảo 23-08-2017, ngày chấp nhận đăng 18-04-2018, ngày đăng 20-11-2018

Nguyễn Thị Ngọc Thanh*, Mai Thị Ngọc Giàu, Nguyễn Thị Huệ – Trường Đại học Khoa học Tự nhiên, ĐHQG-HCM

*Email: ngtntanh@hcmus.edu.vn

Trong số các gene nhạy cảm thấp, gene BARX2 có ảnh hưởng đến các quá trình tế bào như sự tái tổ chức các sợi actin và quá trình tạo sụn [11]. Đặc biệt, BARX2 còn được phát hiện là có sự biểu hiện cao trong các dòng tế bào ung thư vú phụ thuộc Estrogen như MCF7 và T47D [12]. Điểm đa hình rs7107217, nằm ở 152Kb về phía hạ nguồn của gene BARX2, có tần số allele thiếu số cao trong quần thể người Trung Quốc (C% = 31,6%), Nhật Bản (A% = 48,6%), và Việt Nam (C% = 27,8%) trên cơ sở dữ liệu 1000Genome. SNP rs7107217 đã được tìm thấy có liên quan với nguy cơ mắc ung thư vú ở người Trung Quốc (CC vs AA: OR (95% CI) = 1,14 (1,05–1,25); P = 2,2 × 10⁻⁴) và Hàn Quốc (OR (95% CI) = 1,19 (1,06–1,34); P = 7,1 × 10⁻⁷) nhưng không liên quan ở người Nhật Bản (P = 0,33) [13]. Từ đó cho thấy, điểm đa hình này là một chỉ thị tiềm năng đặc trưng cho từng quần thể. Vì vậy, để xác định chỉ thị đặc trưng cho ung thư vú ở người Việt Nam, cần phải khảo sát mối liên quan của rs7107217 với nguy cơ mắc bệnh. Nghiên cứu này được thực hiện nhằm xây dựng phương pháp xác định kiểu gene và khảo sát tần số kiểu gene của rs7107217 trên bệnh nhân ung thư vú, từ đó đưa ra một số thông tin ban đầu về mối liên quan của SNP này với nguy cơ mắc bệnh ở người Việt Nam.

2. VẬT LIỆU VÀ PHƯƠNG PHÁP

Mẫu nghiên cứu

Mẫu nghiên cứu gồm mẫu máu toàn phần từ bệnh nhân nữ được chẩn đoán mắc ung thư vú và chuẩn bị được mổ cắt khối u tại bệnh viện Ung Bướu Thành phố Hồ Chí Minh. Độ tuổi trung bình của nhóm bệnh là 47,8±4,7. Nhóm mẫu máu đối chứng được thu từ các tình nguyện viên nữ đã được xác nhận là không mắc ung thư thông qua kiểm tra sức khỏe hằng năm. Độ tuổi trung bình của nhóm đối chứng là 46,3±5,0. Như vậy, tất cả mẫu máu được thu nhận từ những đối tượng tham gia có cùng độ tuổi, cùng giới tính nữ và là dân tộc Kinh, Việt Nam. Nghiên cứu này được thông qua bởi Hội đồng y đức của bệnh viện Ung Bướu Thành phố Hồ Chí Minh với quyết định số 177/HĐĐĐ-CĐT, 18.11.2014.

Các mẫu máu đã nhận được sự đồng thuận cung cấp từ phía bệnh nhân. Các mẫu máu thu nhận được chứa trong các ống dung dịch EDTA và được vận chuyển về phòng thí nghiệm trong ngày và được bảo quản ở -20 °C trước khi thực

hiện tách chiết DNA. DNA bộ gene sẽ được tách chiết mẫu máu toàn phần bằng phương pháp muối theo quy trình được xây dựng bởi PGS.TS Nguyễn Thị Huệ [10] với một số điều chỉnh. DNA sau khi tách chiết sẽ được tiến hành xác định nồng độ và độ tinh sạch bằng cách đo các giá trị OD₂₆₀ và OD₂₈₀, sử dụng máy NanoDrop 1000 Spectrophotometer (Thermo Scientific, USA). Các mẫu DNA có tỉ lệ OD₂₆₀/OD₂₈₀ trong khoảng từ 1,7–2,0 sẽ được sử dụng để pha loãng xuống nồng độ 10 ng/μL để chuẩn bị cho khảo sát kiểu gene.

Thiết kế môi cho phương pháp Real-time PCR kết hợp phân tích nhiệt độ nóng chảy (Real-time PCR HRM) nhằm xác định kiểu gene của rs7107217

Môi PCR được thiết kế nhằm nhân bản đoạn trình tự mục tiêu có chứa SNP rs7107217 và phải đáp ứng các điều kiện sau: (1) Cặp môi thiết kế phải chuyên biệt cho vùng trình tự mục tiêu; (2) Sản phẩm nhân bản có độ dài từ 50–120 bp; (3) Chênh lệch nhiệt độ nóng chảy (ΔT_m) dự đoán giữa sản phẩm PCR của kiểu gene đồng hợp AA và CC không nhỏ hơn 0,4 °C; (4) Hình dạng đường cong nóng chảy dự đoán kiểu gene AC phải có sự khác biệt rõ ràng với hai kiểu gene còn lại.

Thông tin vùng trình tự chứa rs7107217 được tham khảo từ cơ sở dữ liệu của Ngân hàng gene (NCBI) với mã số NC_000011.9, SNP nằm ở vị trí 129473690. Môi được thiết kế bằng phần mềm Primer3plus (<http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi>). Độ đặc hiệu của môi được kiểm tra bằng phần mềm Primer-Blast (<https://www.ncbi.nlm.nih.gov/tools/primer-blast/>). Đường cong nóng chảy đặc trưng cho từng kiểu gene của sản phẩm PCR được dự đoán bằng phần mềm uMelt HETS (<https://www.dna.utah.edu/hets/umh.php>). Các cấu trúc thứ cấp có thể có của môi được kiểm tra thông qua OligoAnalyzer (<https://sg.idtdna.com/calc/analyzer>).

Tối ưu điều kiện phản ứng xác định kiểu gene của rs7107217

Dựa vào nhiệt độ nóng chảy (T_m) thực tế của môi (được cung cấp bởi nhà sản xuất), nhiệt độ bắt cặp (T_a) được dự đoán thấp hơn T_m 5°C, từ đó thực hiện PCR khảo sát khoảng T_a tối ưu từ thấp hơn T_a dự đoán 5 °C đến cao hơn T_a dự đoán 5 °C bằng máy PCR Mastercycler® pro (Eppendorf). Phản ứng PCR dùng cho việc khảo

sát Ta được tiến hành với tổng thể tích là 10 μ L với các thành phần phản ứng theo khuyến cáo của nhà sản xuất bộ kit Hotstart taq Master Mix (QIAGEN) gồm Buffer 1X; $MgCl_2$ 2 mM; dNTP 0,2 mM; mỗi xuôi 0,2 μ M; mỗi ngược 0,2 μ M; 0,05 μ L Hotstart Taq Polymerase, 20 ng DNA; và nước cho đến 10 μ L. Chu trình nhiệt của phản ứng PCR được tiến hành như sau: (1) 94 °C trong 15 phút; (2) 95 °C trong 30 giây; (3) khoảng Ta khảo sát trong 30 giây; (4) 72 °C trong 1 phút; lặp lại bước (2) đến (4) 40 chu kỳ; (5) 72 °C trong 10 phút; và (6) bảo quản ở 4 °C. Sản phẩm PCR được kiểm tra định tính bằng phương pháp điện di (90 V, 30 phút) trên gel agarose 2%. Phân tích các vạch điện di dựa trên thang chuẩn 50bp DNA (Invitrogen™) được thực hiện cùng trên bảng gel với các sản phẩm điện di. Nhiệt độ bắt cặp tối ưu được chọn cần phải nhân bản được đặc hiệu trình tự mục tiêu và có nhiều sản phẩm nhân bản nhất.

Với Ta tối ưu, phản ứng PCR được thực hiện cho một số mẫu DNA ngẫu nhiên và phân tích nhiệt độ nóng chảy với độ phân giải cao (High Resolution Melting (HRM)) trên hệ thống máy Real - time PCR Light Cycler 96 (Roche) và kết quả được phân tích bằng phần mềm Light Cycler® 96 Version 1.1. Dựa vào kết quả ban đầu này có thể xác định được ba dạng đường cong nóng chảy đặc trưng cho ba kiểu gene AA, AC, và CC. Từ mỗi nhóm đường cong nóng chảy riêng biệt, một mẫu được chọn để giải trình tự xác định kiểu gene làm mẫu chứng cho việc xác định kiểu gene của các mẫu DNA cần khảo sát. Ba mẫu dùng làm mẫu chứng sẽ được tiếp tục sử dụng để khảo sát nồng độ $MgCl_2$ bổ sung tối ưu cho phản ứng xác định kiểu gene của rs7107217 sao cho ba kiểu gene được phân biệt rõ nhất, đồng thời không có sản phẩm phụ. Phản ứng Real-time PCR HRM được tiến hành với tổng thể tích là 5 μ L với các thành phần phản ứng theo khuyến cáo của nhà sản xuất bộ kit Brilliant HRM Ultra-Fast Loci Master Mix (Agilent) gồm Buffer 1X (đã có sẵn một lượng $MgCl_2$ không biết nồng độ); bổ sung thêm một nồng độ $MgCl_2$ nhất định tối ưu; mỗi xuôi 0,2 μ M; mỗi ngược 0,2 μ M; 20 ng DNA; và nước cho đến 5 μ L. Chu trình nhiệt cho phản ứng PCR được tiến hành như sau: (1) 95 °C trong 300 s; (2) 95 °C trong 30 s; (3) Ta tối ưu trong 30 s; (4) 72 °C trong 30 s; lặp lại bước (2) đến (4) 40 chu kỳ. Tiếp ngay sau đó là chu trình nhiệt cho HRM được tiến hành như sau: 95 °C trong 90 s; 40 °C

trong 60 s; 65 °C trong 30 s; và 95 °C trong 1 s. Cuối cùng là bảo quản trong 37 °C trong 30 s.

Tầm soát kiểu gene của rs7107217 trên bệnh nhân ung thư vú Việt Nam

Mẫu DNA thu nhận từ các bệnh nhân ung thư vú được xác định kiểu gene bằng phương pháp Real-time PCR HRM cùng với ba mẫu chứng dương và một mẫu chứng âm. Dựa trên cơ sở là hình dạng đường cong nóng chảy của các mẫu chứng đặc trưng cho ba kiểu gene, các mẫu DNA có đường cong nóng chảy hợp thành một nhóm với mẫu chứng sẽ được xác định kiểu gene tương ứng. Dựa vào số lượng kiểu gene trên tổng số mẫu, tần số allele và kiểu gene của rs7107217 trên bệnh nhân ung thư vú Việt Nam được xác định.

Đánh giá phương pháp xác định kiểu gene của rs7107217

Độ nhạy của phương pháp được đánh giá dựa vào tỉ lệ số mẫu dương tính về DNA được xác định kiểu gene thành công ở lần thực hiện đầu tiên và không lặp lại. Độ ổn định của phương pháp được khảo sát dựa trên sự dao động của T_m các mẫu chứng cùng kiểu gene so với T_m trung bình của chúng và được thực hiện bằng kiểm định T-Test sử dụng phần mềm thống kê R (R i386 3.4.0). Độ đặc hiệu của phương pháp được khảo sát dựa trên khả năng phân biệt ba kiểu gene của các mẫu chứng và của các mẫu cần xác định kiểu gene. Khả năng phân biệt kiểu gene giữa các mẫu chứng được thực hiện bằng cách so sánh ba T_m trung bình của các mẫu chứng qua các lần chạy của ba kiểu gene với nhau. Khả năng phân biệt kiểu gene giữa các mẫu khảo sát được thực hiện bằng cách so sánh ba T_m trung bình của tất cả các mẫu của ba kiểu gene với nhau. Sự khác biệt về ba T_m trung bình trong khảo sát độ đặc hiệu của phương pháp sẽ được phân tích thông qua kiểm định Anova-test sử dụng phần mềm thống kê R (R i386 3.4.0).

Dự đoán mối liên quan giữa rs7107217 với nguy cơ ung thư vú

Nhằm dự đoán mối liên quan giữa rs7107217 với nguy cơ ung thư vú, sự khác biệt giữa tần số allele và kiểu gene giữa nhóm đối chứng (tham khảo từ cơ sở dữ liệu 1000 Genomes) và nhóm ung thư vú người Việt Nam được xác định bằng kiểm định Chi-Test. Các giá trị OR [95% CI] và p-value được xác định bằng phân tích hồi quy logistic bằng phần mềm R (R i386 3.4.0).

Tầm soát kiểu gene của rs7107217 trên nhóm đối chứng người Việt Nam khỏe mạnh

Mẫu DNA thu nhận từ nhóm đối chứng được xác định kiểu gene bằng phương pháp Real-time PCR HRM đã được tối ưu. Dựa trên cơ sở là hình dạng đường cong nóng chảy của các mẫu chứng đặc trưng cho ba kiểu gen, các mẫu DNA có đường cong nóng chảy hợp thành một nhóm với mẫu chứng sẽ được xác định kiểu gene tương ứng. Dựa vào số lượng kiểu gene trên tổng số mẫu, tần số allele và kiểu gene của rs7107217 trên nhóm đối chứng người Việt Nam được xác định.

Xác định mối liên quan giữa rs7107217 với nguy cơ ung thư vú

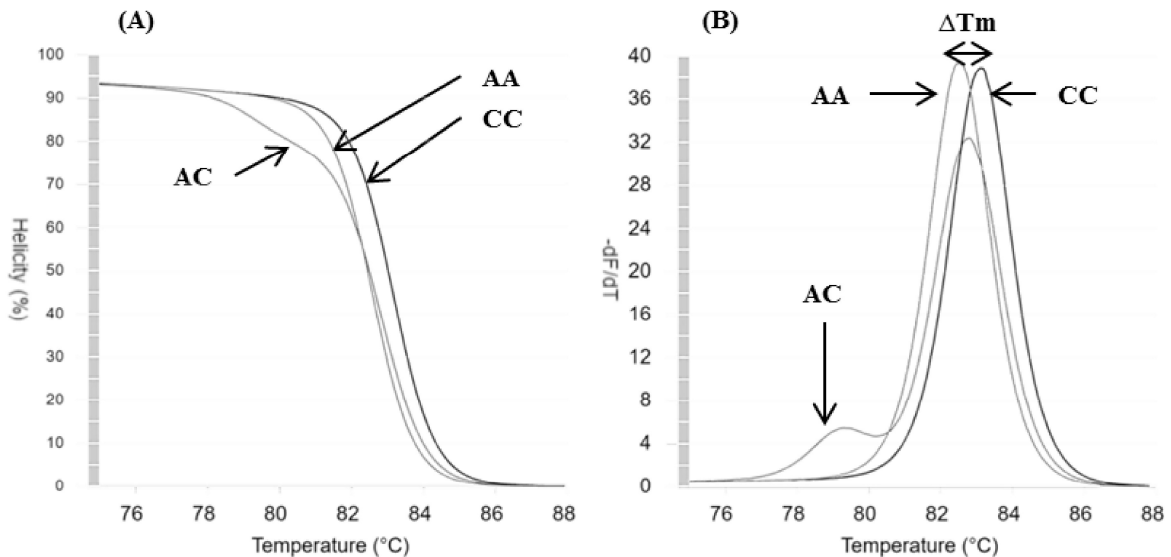
Nhằm xác định mối liên quan giữa rs7107217 với nguy cơ ung thư vú, sự khác biệt giữa tần số allele và kiểu gene giữa nhóm đối chứng trong đồng về giới tính, độ tuổi, và dân tộc với nhóm ung thư vú được xác định bằng kiểm định Chi-Test. Các giá trị OR [95% CI] và p-value được xác định bằng phân tích hồi quy logistic bằng phần mềm R (R i386 3.4.0).

3. KẾT QUẢ VÀ THẢO LUẬN

SNPs được xem như là các chỉ thị sinh học tiềm năng trong việc nghiên cứu tác động của chúng đến các con đường sinh bệnh học và trong chuẩn đoán sớm bệnh, đặc biệt là bệnh ung thư [5]. SNP rs7107217 nằm trên gene BARX2 có khả năng

tham gia điều hòa biểu hiện của BARX2 từ đó gián tiếp điều hòa biểu hiện của ESR1 liên quan đến hình thành ung thư vú. Nghiên cứu xây dựng phương pháp khảo sát tần số kiểu gene của rs7107217 trong quần thể bệnh nhân ung thư vú Việt Nam là một bước đầu trong nghiên cứu mối liên quan của SNP này với con đường bệnh sinh và chẩn đoán bệnh sớm thông qua chỉ thị di truyền đặc trưng cho người Việt Nam.

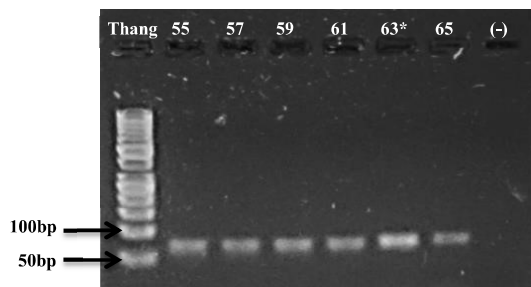
Trong nhiều cặp môi được thiết kế cho phương pháp Real-time PCR HRM xác định kiểu gene của rs7107217, cặp môi được lựa chọn là cặp môi tạo sản phẩm đặc hiệu và có đường cong nóng chảy dự đoán đại diện ba kiểu gene trên phần mềm Umelt phân biệt rõ nhất. Kết quả dự đoán hình dạng đường cong nóng chảy và đỉnh nóng chảy của ba kiểu gene đặc trưng cho từng kiểu gene ở nồng độ $MgCl_2$ dự đoán là 2 mM (Hình 1) trong dãy nồng độ khảo sát từ 1,5 đến 3 mM. Sự chênh lệch T_m của hai đỉnh nóng chảy của kiểu gene đồng hợp là lớn ($\Delta T_m = 0,6$) (Hình 1B), giúp phân biệt rõ ràng hai kiểu gene AA và CC. Kiểu gene dị hợp AC có hai đỉnh nóng chảy rõ ràng (Hình 1B) và đường cong nóng chảy cắt chính giữa đường cong nóng chảy của kiểu gene đồng hợp AA (Hình 1A). Như vậy, cặp môi được chọn gồm môi xuôi với trình tự 5'AGTACATGTATGCCAGGAGACA CA3' và môi ngược với trình tự 5'TTGCTATT TTGGCAAGTTCAACTATGA3'.



Hình 1. Đường cong nóng chảy (A) và đỉnh nóng chảy (B) dự đoán của ba kiểu gene của rs7107217 ở nồng độ $MgCl_2$ dự đoán là 2 mM

Dựa trên T_m thực tế của mỗi xuôi là 64,5 °C và mỗi ngược là 67 °C, nhiệt độ bắt cặp (T_a) được khảo sát từ 55 °C đến 65 °C. Kết quả cho thấy sản phẩm mục tiêu đã được nhân bản thành công với kích thước 65 bp nằm giữa hai vạch 50 bp và 100 bp của thang 50 bp DNA (Invitrogen™) (Hình 2). Ở nhiệt độ 63 °C, vạch điện di sáng nhất chứng tỏ ở nhiệt độ này phản ứng PCR nhân bản được lượng sản phẩm nhiều nhất. Do đó, 63 °C được chọn là T_a tối ưu cho phản ứng Real-time PCR HRM.

Một số mẫu DNA ngẫu nhiên được tiến hành xác định kiểu gene bằng phương pháp Real-time PCR HRM với T_a được chọn. Nồng độ $MgCl_2$ bổ sung vào thành phần phản ứng được chọn dựa vào nồng độ $MgCl_2$ dự đoán từ phần mềm Umelt. Mặc dù nồng độ $MgCl_2$ dự đoán là 2 mM, nhưng vì trong Buffer của Brilliant HRM Ultra-Fast Loci Master Mix (Agilent) đã có sẵn một lượng $MgCl_2$ không xác định nên nồng độ $MgCl_2$ được bổ sung vào lần chạy Real-time PCR HRM đầu tiên là tối thiểu theo khuyến cáo của nhà sản xuất (1,5 mM). Kết quả xuất hiện ba dạng đường cong nóng chảy khác biệt rõ ràng (Hình 3) và được dự đoán là đại diện cho ba kiểu gene của rs7107217.



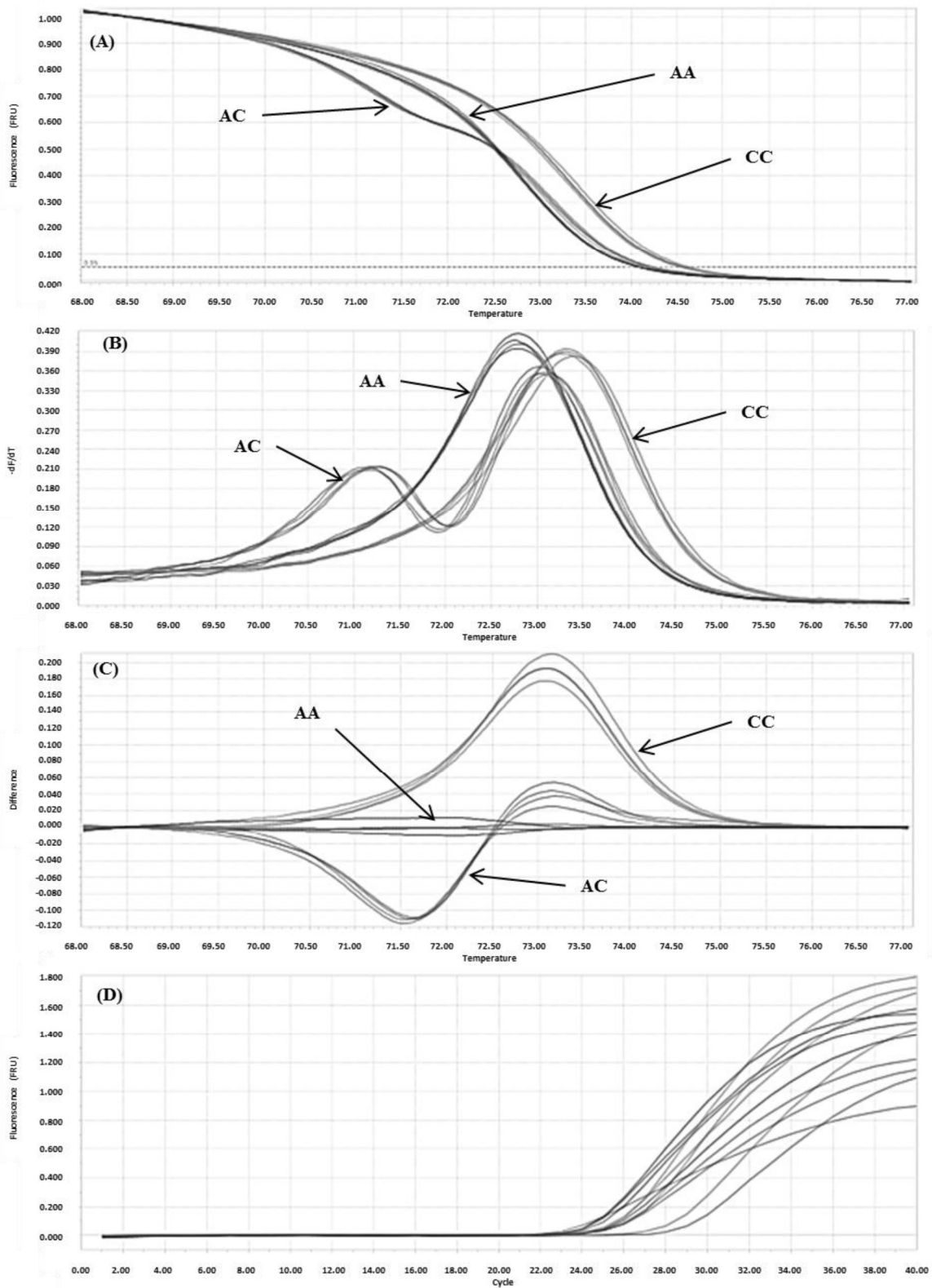
Hình 2. Khảo sát nhiệt độ bắt cặp tối ưu của phản ứng Real-time PCR HRM. * Nhiệt độ tối ưu được chọn

Ba mẫu bất kỳ đại diện cho ba dạng đường cong nóng chảy được giải trình tự xác nhận kiểu gene. Kết quả giải trình tự cho thấy ba mẫu DNA được chọn có ba kiểu gene khác nhau tương ứng với ba dạng đường cong nóng chảy đã phân tích trước đó.

Nồng độ $MgCl_2$ là yếu tố quan trọng cần được khảo sát vì đây là cofactor của Taq polymerase và giúp ổn định mạch DNA từ đó giúp quá trình tách

mạch ở bước HRM diễn ra ổn định hơn. Nếu nồng độ $MgCl_2$ quá thấp có thể làm giảm hiệu quả hoạt động của Taq polymerase, cũng như làm cho khả năng tách mạch DNA kém ổn định dẫn đến sự tách biệt của 3 nhóm kiểu gene kém hơn; ngược lại nếu nồng độ $MgCl_2$ quá cao làm tăng hoạt động của Taq polymerase dẫn đến dễ tạo sản phẩm phụ ảnh hưởng kết quả đường cong nóng chảy. Nồng độ $MgCl_2$ thường được sử dụng trong Real-time PCR HRM cũng như được nhà sản xuất bộ kit đề nghị là từ 1,5 đến 3,0 mM để khảo sát sự phân biệt ba dạng đường cong nóng chảy giữa ba kiểu gene. Với nồng độ $MgCl_2$ bổ sung vào phản ứng Real-time PCR HRM là 1,5 mM, ba kiểu gene của rs7107217 đã có sự phân biệt rõ ràng trong ba kiểu phân tích (Hình 3). Phân tích đường cong nóng chảy (Hình 3A) cho thấy các mẫu trong cùng một kiểu gene nhóm vào nhau và nhóm vào với mẫu chứng dương một cách ổn định. Hình dạng của các đường cong nóng chảy giữa các kiểu gene được phân biệt rõ ràng. Dựa vào phân tích đỉnh nóng chảy (Hình 3B), ba kiểu gene có thể được nhận diện một cách dễ dàng bằng hai đỉnh của mẫu dị hợp AC và cách biệt lớn giữa hai đỉnh T_m ($\Delta T_m = 0,59$ °C) của mẫu đồng hợp (73,36 °C đối với CC và 72,77 °C đối với AA). Trong phân tích đường cong nóng chảy với mẫu đồng hợp AA là chuẩn (Hình 3C), ba kiểu gene được phân biệt một cách rõ ràng nhất. Phân tích đường cong tăng trưởng (Hình 3D) cho thấy giá trị C_t luôn trong khoảng từ 24 đến 28, điều này chỉ ra rằng, sản phẩm PCR là phù hợp cho điều kiện phân tích nhiệt độ nóng chảy trong hình 3-A, B, C. Do đó, nồng độ $MgCl_2$ bổ sung được giữ nguyên là 1,5 mM để tiến hành xác định kiểu gene cho các mẫu DNA khảo sát.

Để khảo sát độ ổn định của điều kiện Real-time PCR HRM xác định kiểu gene của rs7107217, sự khác biệt T_m của các mẫu chứng trong cùng một kiểu gene qua tất cả các lần chạy so với T_m trung bình của chúng được kiểm tra bằng thuật toán T-test. Kết quả cho thấy sự khác biệt này không có ý nghĩa thống kê (P -value > 0,05) (Bảng 1). Qua đó chứng tỏ điều kiện Real-time PCR HRM được xây dựng trong nghiên cứu này có độ ổn định cao giữa các lần chạy.



Hình 3. Xác định kiểu gene của một số mẫu ngẫu nhiên. (A) Đường cong nóng chảy. (B) Định nóng chảy. (C) Đường cong nóng chảy so với đường chuẩn của kiểu gene AA. (D) Đường cong tăng trưởng PCR với Ct từ 24 đến 28

Để đánh giá độ đặc hiệu trong phân biệt ba kiểu gene của rs7107217, đầu tiên dựa trên ba T_m trung bình của các mẫu chứng qua các lần thực hiện của ba kiểu gen, ba T_m trung bình này được so sánh với nhau bằng kiểm định Anova Test (Bảng 1). Kết quả cho thấy có sự khác biệt có ý nghĩa thống kê giữa ba T_m trung bình của ba kiểu

gene (P -value < 0,001). Bên cạnh đó, dựa trên T_m trung bình của các mẫu được khảo sát kiểu gene trong nghiên cứu, ba T_m trung bình của ba kiểu gene được so sánh với nhau bằng kiểm định Anova Test (Bảng 1). Kết quả tương tự cho thấy có sự khác biệt có ý nghĩa thống kê giữa ba kiểu gene AA, AC và CC.

Bảng 1. Độ ổn định nhiệt độ nóng chảy của các mẫu chứng qua các lần chạy

Khảo sát độ ổn định của phương pháp	Kiểu gene	T_m trung bình của chứng \pm SD ($^{\circ}$ C)	P-value T-test
	AA	72,78 \pm 0,23	0,65
	AC	73,12 \pm 0,19	0,77
	CC	73,30 \pm 0,21	0,97
Khảo sát độ đặc hiệu của phương pháp	Kiểu gene	T_m trung bình của chứng \pm SD ($^{\circ}$ C)	P-value Anova test
	AA	72,78 \pm 0,13	<0,001
	AC	73,12 \pm 0,14	
	CC	73,30 \pm 0,09	
	Kiểu gene	T_m trung bình của mẫu \pm SD ($^{\circ}$ C)	P-value Anova test
	AA	72,85 \pm 0,23	<0,001
	AC	73,12 \pm 0,19	
CC	73,35 \pm 0,21		

Từ những kết quả phân tích thống kê trên cho thấy điều kiện Real-time PCR HRM được sử dụng trong nghiên cứu này để phân biệt ba kiểu gene của rs7107217 có độ đặc hiệu cao. Hai kiểu gene đồng hợp với hình dạng đường cong nóng chảy tương tự nhau nhưng dựa vào T_m có thể dễ dàng được phân biệt. Kiểu gene dị hợp có dạng đỉnh nóng chảy gồm 2 đỉnh khác biệt rõ ràng với hai dạng của đồng hợp, đồng thời T_m của sản phẩm dị hợp cũng có sự tách biệt với hai sản phẩm đồng hợp. Như vậy hình dạng đường nóng chảy và T_m của mỗi kiểu gene của rs7107217 là đặc hiệu cho từng kiểu gene khi sử dụng điều kiện Real-time PCR HRM trong nghiên cứu này.

Kết quả khảo sát tần số kiểu gene và allele của rs7107217 trên 117 bệnh nhân ung thư vú người Việt Nam được trình bày ở Bảng 2. Sự phân bố kiểu gene và allele của rs7107217 trong quần thể ung thư vú Việt Nam nằm trong cân bằng Hardy–Weinberg ($P_{HWE} = 0,51$). So sánh với dữ liệu tần số kiểu gene và allele của người Việt Nam dân tộc Kinh trên cơ sở dữ liệu 1000 Genome, sự tăng tần số của allele C trong nhóm bệnh nhân ung thư vú

(29,9%) so với nhóm người khảo sát trên 1000 Genome (27,7%) cùng với chỉ số OR là 1,113 cho thấy allele C có xu hướng gây tăng nguy cơ ung thư vú (Bảng 2). Tuy nhiên, chỉ số 95% CI lại có độ dao động trong một khoảng tương đối rộng xung quanh giá trị 1 (0,734–1,687) (Bảng 2) cho thấy dự đoán xu hướng làm tăng ung thư vú của allele C chưa rõ ràng. Đồng thời sự khác biệt về tần số kiểu gene và allele giữa quần thể người Việt Nam mắc ung thư vú trong nghiên cứu này và quần thể người Việt Nam tham khảo từ dự án 1000 Genome cũng được tìm thấy là chưa có ý nghĩa thống kê ($P_{kiểu\ gene} = 0,48$; $P_{allele} = 0,61$) (Bảng 2). Độ tin cậy của kết quả dự đoán mối liên quan giữa rs7107217 và nguy cơ ung thư vú trong nghiên cứu này là thấp, chỉ 5,6%. Do đó, kết quả dự đoán này vẫn chưa thể xác định một cách chính xác và rõ ràng mối tương quan giữa rs7107217 và nguy cơ ung thư vú ở người Việt Nam.

Bảng 2. Tần số kiểu gene và allele của nhóm bệnh nhân ung thư vú Việt Nam được khảo sát và nhóm đối chứng từ 1000 Genome. Dự đoán mối liên quan của rs7107217 (C/A) với nguy cơ ung thư vú ở người Việt Nam

Nhóm mẫu	Kiểu gen			P	Allele			OR (95% CI)	P _{HWE}
	AA	AC	CC		A	C	P		
Bệnh (n=117)	59 (50,4%)	46 (39,3%)	12 (10,3%)	0,48	164 (70,1%)	70 (29,9%)	0,61	1,113 (0,734 – 1,687)	0,51
*Chứng (n=101)	51 (50,1%)	44 (43,6%)	6 (5,9%)		146 (72,3%)	56 (27,7%)			

*Cơ sở dữ liệu 1000Genome

Để tìm hiểu chính xác hơn về tần số allele và kiểu gene cũng như mối tương quan của rs7107217 với nguy cơ ung thư vú, nhóm đối chứng là các tình nguyện viên người Việt Nam được khảo sát kiểu gen. Kết quả khảo sát tần số kiểu gene và allele của rs7107217 trên 105 tình nguyện viên nhóm đối chứng người Việt Nam được trình bày ở Bảng 3. Sự phân bố kiểu gene và allele của rs7107217 trong quần thể này cũng nằm trong cân bằng Hardy–Weinberg ($P_{HWE} = 0,83$). So sánh với dữ liệu tần số kiểu gene và allele của người Việt Nam dân tộc Kinh trên cơ sở dữ liệu

1000Genome (Bảng 2) cho thấy có sự tương đồng về allele và kiểu gene thiếu số đều là allele C và kiểu gene CC. Tuy nhiên, có sự thay đổi ở kiểu gene đa số với kiểu gene AA chiếm đa số trên 1000 Genome (50,1%) và kiểu gene AC chiếm đa số trong nghiên cứu này (43,6%). Sự khác biệt này ở hai quần thể khảo sát có thể do sự chọn mẫu, trong nghiên cứu này mẫu đối chứng được chọn là các đối tượng ở miền nam Việt Nam, trong khi trên 1000 Genome mẫu được chọn ngẫu nhiên ở các miền của Việt Nam.

Bảng 3. Tần số kiểu gene và allele của nhóm đối chứng người Việt Nam. Mối liên quan của rs7107217 (C/A) với nguy cơ ung thư vú ở người Việt Nam.

Nhóm mẫu	Kiểu gene			P	Allele			OR (95% CI)	P _{HWE}
	AA	AC	CC		A	C	P		
Bệnh (n=117)	59 (50,4%)	46 (39,3%)	12 (10,3%)	0,36	164 (70,1%)	70 (29,9%)	0,23	0,79 (0,53 – 1,17)	0,51
Đối chứng (n=105)	43 (45%)	50 (47,6%)	12 (11,4%)		136 (64,7%)	74 (35,3%)			

Phân tích mối tương quan với nguy cơ ung thư vú ở cỡ mẫu 117 ca và 105 chứng cho thấy allele C có xu hướng làm giảm nguy cơ ung thư vú khoảng 0,79 lần (OR = 0,79) (Bảng 3). Đây là một sự thay đổi đáng chú ý trong phân tích này. Khi dự đoán mối liên quan với nguy cơ ung thư vú sử dụng cơ sở dữ liệu trên 1000 Genome, allele C lại được cho thấy có xu hướng làm tăng nguy cơ ung thư vú (OR > 1) (Bảng 2). Giá trị OR bị đảo ngược này có thể do sự khác biệt của 2 nhóm mẫu đối chứng. Đối với số liệu từ 1000 Genome, mẫu trong nhóm này chỉ tương đồng với nhóm bệnh ở yếu tố dân tộc. Đối với số liệu trong nghiên cứu này, mẫu đối chứng được chọn có sự tương đồng

với nhóm bệnh ở các yếu tố về giới tính, tuổi, và dân tộc. Tuy nhiên, các giá trị p-value của phân tích này đều lớn hơn 0,05 (Bảng 3). Điều này chỉ ra rằng ảnh hưởng của SNP này với nguy cơ ung thư vú ở người Việt Nam là chưa có ý nghĩa thống kê.

Tiếp tục đánh giá về độ tin cậy của phân tích mối liên quan này, với cỡ mẫu hiện tại là 222 mẫu, kết quả mối tương quan chỉ đạt độ tin cậy là 11,71%. Mặc dù rs7107217 chưa được tìm thấy có liên quan đến nguy cơ ung thư vú ở người Việt Nam nhưng vì độ tin cậy thấp (11,71%) nên phân tích này cần được thực hiện thêm trên cỡ mẫu lớn hơn. Để tăng độ tin cậy của mối tương quan lên

50%, cần tiến hành trên cỡ mẫu là 632 ca / 632 chứng. Với độ tin cậy mong muốn là 80%, cỡ mẫu cần thực hiện vào khoảng 1289 ca / 1289 chứng.

Mặc dù chưa tìm thấy mối liên quan giữa rs7107217 với nguy cơ ung thư vú trong nghiên cứu này, nhưng với vị trí nằm ở 152Kb về phía hạ nguồn của gene BARX2 cho thấy có nhiều khả năng SNP này nằm trong vùng điều hòa gene và có thể ảnh hưởng đến biểu hiện của protein BARX2. Các nghiên cứu cho thấy rằng BARX2 cùng với Estrogen Receptor α (ESR1) góp phần vào việc kiểm soát quá trình tăng sinh và xâm lấn của tế bào ung thư vú đồng thời điều hòa tạo ra các isoform khác nhau của ESR1. Con đường đầu tiên, BARX2 có thể tương tác về mặt chức năng với estrogen trong sự điều hòa một vài gene mục tiêu có liên quan đến các con đường tăng sinh và xâm lấn của các tế bào ung thư vú như gene MMP (matrix metalloproteinase) và gene TIMP (tissue inhibitor of metalloproteinase) từ đó có thể dẫn đến sự hình thành và phát triển ung thư vú [12]. Ngoài ra, BARX2 cũng có thể ảnh hưởng đến ung thư vú thông qua sự đồng điều hòa với ESR1 dưới sự hoạt hóa của estrogen, gene ESR1 được phiên mã thành các mRNA có các vùng khởi đầu phiên mã khác nhau, từ đó được dịch mã thành các loại protein có chiều dài khác nhau gồm dạng đầy đủ ESR1-66 kDa và dạng thiếu exone 1 ESR1-46 kDa. Tuy nhiên, khi BARX2 và ESR1 protein bám vào những vùng promoter khác nhau của gene ESR1, chúng sẽ điều hòa sự biểu hiện của ESR1-46 và ESR1-66 kDa dẫn đến làm thay đổi tỷ lệ của hai protein này trong tế bào. Chính sự thay đổi tỉ lệ của hai dạng protein ESR1 đã dẫn đến sự hình thành và phát triển của ung thư vú [12]. BARX2 làm tăng biểu hiện của cả hai loại protein ESR1, nhưng ảnh hưởng mạnh hơn với dạng ESR1-46 kDa - dạng được biểu hiện trong các dòng tế bào ung thư vú, ung thư tuyến vú, ung thư xương [12]. Với tiềm năng liên quan đến sự hình thành ung thư vú, SNP rs7107217, một lần nữa cho thấy, cần được khảo sát trên cỡ mẫu lớn hơn để xác định chính xác hơn mối liên quan với nguy cơ ung thư vú ở người Việt Nam.

4. KẾT LUẬN

Nghiên cứu này đã xây dựng thành công phương pháp Real-time PCR HRM xác định kiểu gene của 117 mẫu DNA từ bệnh nhân ung thư vú

Việt Nam với độ nhạy, độ ổn định và độ đặc hiệu cao. Chính vì thế, phương pháp Real-time PCR HRM đã xây dựng có thể áp dụng trong việc tầm soát kiểu gene của rs7107217. Trong quần thể ung thư vú người Việt Nam, tần số xuất hiện của kiểu gene AA là 50,4%, AC là 39,3%, CC là 29,9% và tần số của allele thiếu số C chiếm 29,9%. Trong quần thể đối chứng người Việt Nam, tần số xuất hiện của kiểu gene AA, AC, và CC lần lượt là 45%, 47,6%, và 11,4%; và tần số của allele thiếu số C chiếm 35,3% cho thấy rs7107217 có độ đa hình cao. Bên cạnh đó, với cơ chế điều hòa rõ ràng trong hình thành ung thư vú thông qua điều hòa biểu hiện ESR1 của BARX2, SNP rs7107217, ở hạ nguồn gene BARX2, có khả năng ảnh hưởng đến nguy cơ hình thành ung thư vú một cách độc lập hay kết hợp với các SNP khác trên cùng gene hoặc khác gene. Mặc dù trong nghiên cứu này, rs7107217 chưa cho thấy có liên quan với nguy cơ ung thư vú ở người Việt Nam (p -value > 0,05) nhưng với độ tin cậy của mối tương quan này còn thấp (11,71%) và tiềm năng ảnh hưởng đến nguy cơ hình thành ung thư vú đã nêu trên, nghiên cứu này cần được thực hiện thêm trên cỡ mẫu lớn hơn, trên 632 ca và 632 chứng để đạt độ tin cậy trên 50%.

Lời cảm ơn: Nghiên cứu này được tài trợ bởi Trường Đại học Khoa học Tự nhiên, ĐHQG-HCM với mã số đề tài T2017-25. Các tác giả xin cảm ơn Bệnh viện Ung bướu – TP.HCM trong việc thu mẫu tại bệnh viện.

TÀI LIỆU THAM KHẢO

- [1] D.S.S. Isabel, D.S. Bianca, M. Valerie, "Birth size and breast cancer risk: re-analysis of individual participant data from 32 studies", *PLoS Med*, vol. 5, no. 9, e193, pp. 1372–1386, 2008.
- [2] G. Berclaz, et al., "Body mass index as a prognostic feature in operable breast cancer: the International Breast Cancer Study Group experience", *Annals of Oncology*, vol. 15, no. 6, pp. 875–884, 2004.
- [3] J. Pei, F. Li, B. Wang, "Single nucleotide polymorphism 6q25, 1 rs2046210 and increased risk of breast cancer", *Tumor Biology*, vol. 34, no. 6, pp. 4073–4079, 2013.
- [4] J.A. Ludwig, J.N. Weinstein, "Biomarkers in cancer staging, prognosis and treatment selection", *Nature Reviews, Cancer*, vol. 5, no. 11, pp. 845–856, 2005.
- [5] Z. Herceg, P. Hainaut, "Genetic and epigenetic alterations as biomarkers for cancer detection, diagnosis and prognosis", *Molecular Oncology*, vol. 1, no. 1, 26–41, 2007.
- [6] C. Garnis, T.P. Buys, W.L. Lam, Genetic alteration and gene expression modulation during cancer progression, *Molecular Cancer*, vol. 3, no. 1, 9, 2004.
- [7] T.J. Harris, F. McCormick, "The molecular

- pathology of cancer”, *Nature Reviews Clinical oncology*, vol. 7, no. 5, pp. 251–265, 2010.
- [8] M. Kanehisa, S. Goto, “KEGG: kyoto encyclopedia of genes and genomes”, *Nucleic Acids Research*, vol. 28, no. 1, pp. 27–30, 2000.
- [9] P.D. Pharoah, et al., “Polygenic susceptibility to breast cancer and implications for prevention”, *Nat Genet*, vol. 31, no. 1, pp. 33–36, 2002.
- [10] A.M. Martin, B.L. Weber, “Genetic and hormonal risk factors in breast cancer”, *Journal of the National Cancer Institute*, vol. 92, no. 14, pp. 1126–1135, 2000.
- [11] R. Meech, et al., “The homeodomain protein Barx2 promotes myogenic differentiation and is regulated by myogenic regulatory factors”, *Journal of Biological Chemistry*, vol. 278, no. 10, pp. 8269–8278, 2003.
- [12] T. Stevens, R. Meech, “BARX2 and estrogen receptor- α (ESR1) coordinately regulate the production of alternatively spliced ESR1 isoforms and control breast cancer cell growth and invasion”, *Oncogene*, vol. 25, no. 39, pp. 5426–5435, 2006.
- [13] J. Long, et al., “Genome-wide association study in east Asians identifies novel susceptibility loci for breast cancer”, *PLoS Genet*, vol. 8, no. 2, e1002532, 2012.

The development of rs7107217 genotyping method and initial study of the association of this gene with the breast cancer risk in Vietnamese women

Nguyen Thi Ngoc Thanh*, Mai Thi Ngoc Giau, Nguyen Thi Hue

University of Science, VNU-HCM

Corresponding author: ngtnthanh@hcmus.edu.vn

Received 23-08-2017; Accepted 18-04-2018; Published 20-11-2018

Abstract—Breast cancer is the most common cancer for women around the world. The presence of single nucleotide polymorphisms (SNP) on or near the coding region of breast cancer susceptibility genes can affect the regulation of gene expression, which may increase or decrease the risk of breast cancer. BARX2 was showed to stimulate the expression of ESR1, which involved in the development of breast cancer. SNP rs7107217 on 152kb downstream of the BARX2 could affect the level of protein BARX2 and had been proved to associate with the breast cancer risk in populations similar to Vietnamese, including Chinese and Korean. In this study, rs7107217 was genotyped and initially determined the association with the breast cancer risk in Vietnamese. Real-time PCR HRM was optimized and used to genotype rs7107217 in 117 breast cancer cases and 105 healthy controls. Thereafter, the correlation of this SNP with the risk of breast cancer was initially determined by

analyzing the differences in allelic and genotypic frequencies between cases and control groups. The results showed the optimal rs7107217 genotyping condition was successfully developed with the high sensitivity, specificity, and consistency. SNP rs7107217 had high polymorphism with the frequency of minor allele C of 29.9% and 35.3% in case and control, respectively. SNP rs7107217 had been found no association with the breast cancer risk (C vs A: $P = 0.23$, OR (95% CI) = 0.79 (0.53 – 1.17)). However with the low reliability of the analysis (11.71%) and the high potential related to the formation of breast cancer, the association between rs7107217 and breast cancer risk in Vietnamese population should be further conducted on a larger sample size to get higher accuracy.

Keywords—rs7107217, breast cancer, Vietnam, SNP genotyping method, High Resolution Melting