

Bước đầu xác định mối liên quan giữa SNP rs3937033 trên DNA ty thể và bệnh ung thư vú người Việt Nam

- Dương Thị Hồng Hạnh
- Nguyễn Thị Ngọc Thanh
- Nguyễn Thị Huệ

Trường Đại học Khoa học Tự nhiên, ĐHQG-HCM
Email: honghanh167@gmail.com

(Bài nhận ngày 10 tháng 02 năm 2017, nhận đăng ngày 17 tháng 06 năm 2017)

TÓM TẮT

Ung thư vú là ung thư thường gặp nhất ở phụ nữ và chiếm tỷ lệ tử vong cao nhất ở các nước đang phát triển, trong đó có Việt Nam. D-loop là vùng không mã hóa, kiểm soát sao chép và phiên mã các gene ty thể. Đột biến trên D-loop làm rối loạn chức năng chuỗi hô hấp từ đó đóng góp vào sự phát triển ung thư vú. SNP rs3937033 (T/C) trên D-loop đã được báo cáo là có liên quan đến ung thư vú ở quần thể Âu Mỹ (OR [95%CI]= 1,98 [1,25–3,12], $P=0,036$), và người da Trắng (OR[95%CI]= 21 [2,15–204,6], $P=0,003$). Trong nghiên cứu này, phương pháp High Resolution Melt (HRM) được tối ưu để tầm soát kiểu gene của 100/100 mẫu.

Từ khóa: mtDNA, D-loop, ung thư vú, SNP T16519C, rs3937033

MỞ ĐẦU

Ung thư vú là ung thư thường gặp nhất ở phụ nữ trên thế giới, chiếm tỷ lệ tử vong cao nhất ở phụ nữ với gần 1,7 tỷ ca mắc mới và có hơn 5,20/100.000 ca tử vong được thống kê năm 2012 [1]. Tại Việt Nam, tỷ lệ mắc bệnh ung thư vú liên tục tăng theo các năm. Năm 2012, theo ước tính của IARC có khoảng 9,91/100.000 ca tử vong trong số 22,96/100.000 ca mắc ung thư vú trên cả nước [2]. Nguyên nhân chủ yếu dẫn đến tỷ lệ tử vong cao được xác định là do bệnh được chẩn đoán và phát hiện chủ yếu vào giai đoạn trễ của bệnh, chủ yếu là giai đoạn II (61,2 %), III (19,4 %) [3]. Theo báo cáo của tổ chức WHO năm 2012 cho biết

bệnh/chứng, sau đó xác định mối liên quan của SNP với bệnh. Tần số kiểu gene TT, TC và CC trong nhóm bệnh (36 %, 3 % và 61 %) và nhóm chứng (23 %, 1 % và 76 %) được tính toán dựa trên kết quả khảo sát kiểu gene bộ mẫu bằng phương pháp HRM với nhiệt độ bắt cặp tối ưu là 66 °C. Tiến hành phân tích mối liên quan cho thấy allele T làm tăng nguy cơ ung thư vú lên 1,41 lần ở người bệnh (OR_{alen} [95 % CI] =1,41 [1,03–1,92], $P=0,03$). Do vậy, SNP rs3937033 có thể là chỉ thị di truyền tiềm năng cho chẩn đoán sớm ung thư vú Việt Nam, tuy nhiên nghiên cứu này cần tiếp tục được thực hiện với cỡ mẫu lớn hơn để đạt được độ tin cậy cao hơn.

nếu bệnh được phát hiện sớm thì sẽ giảm được nguy cơ tử vong xuống 1/3 so với tỷ lệ mắc, từ đó cho thấy việc chẩn đoán và phát hiện sớm ung thư vú đóng một vai trò hết sức quan trọng trong hướng điều trị để nâng cao khả năng sống cho bệnh nhân. Những biến đổi di truyền được xem là một trong những nguyên nhân gây nên sự tiến triển ung thư vú. Bên cạnh những đột biến nghiêm trọng xảy ra trong nhân thì đột biến trên mtDNA đặc biệt là vùng D-loop được quan sát thấy ở nhiều loại ung thư [4]. Gần đây DNA ty thể (mtDNA) được xem là chỉ thị mới có thể được sử dụng trong chẩn đoán nguy cơ mắc bệnh ở giai đoạn sớm. Ty thể là bào quan đóng vai trò quan trọng trong chuyển hóa năng lượng, lão hóa và apoptosis [5]. mtDNA

người có cấu trúc sợi đôi, phân tử đóng vòng, có kích thước 16,569 base pairs mã hóa cho 13 polypeptide của chuỗi chuyển điện tử (ETC), 22 transfer RNAs (tRNA) và 2 ribosomal RNAs (12S và 16 S rRNA) [6]. mtDNA có tỷ lệ đột biến cao hơn DNA nhân (nDNA) gấp 10-20 lần [7] và dễ dàng bị tổn thương oxy hóa bởi thiếu sự bảo vệ của protein giống histon, không có introns và cơ chế hoạt động kém hiệu quả của hệ thống sửa chữa sao chép và nhạy cảm với mức độ ROS cao được tạo ra trong tế bào [8]. Displacement loop (D-loop) là vùng không mã hóa của mtDNA có kích thước 1124bp (16024–576np) được biết đến là điểm nóng thường xảy ra đột biến, chịu trách nhiệm kiểm soát sự sao chép và phiên mã gene ty thể [9]. Do vậy, sự biến đổi xảy ra trong vùng này sẽ làm thay đổi quá trình sao chép và phiên mã các gene, gây ảnh hưởng đến chức năng của các thành phần trong chuỗi ETC bởi các gốc oxy hóa (ROS) được tạo ra quá mức trong tế bào, từ đó góp phần vào sự khởi phát và tiến triển của ung thư, trong đó có ung thư vú [10, 11]. Không chỉ vậy, sự tăng quá mức ROS trong tế bào còn liên quan đến sự tổn thương nDNA, ảnh hưởng đến con đường truyền tín hiệu dẫn đến sự hoạt hóa một số oncogen và ức chế các gene ức chế khối u từ đó ức chế quá trình apoptosis, tăng sinh mạch và thúc đẩy sự di căn của ung thư [12].

Sự đa hình rs3937033 xảy ra bởi sự thay đổi allele từ Thymine (T) thành Cytosine (C) ở vị trí 16519np trên D-loop mtDNA, do đó còn gọi là SNP T16519C, nó được quan sát thấy là một trong những điểm nóng thường xảy ra biến đổi liên quan đến sự tiến triển ung thư vú [13, 14]. Hơn nữa, đã có nhiều nghiên cứu trên thế giới cho thấy SNP rs3937033 có mối liên quan mật thiết đến nguy cơ mắc bệnh ung thư vú như quần thể Âu Mỹ và Người da trắng [14, 15]. Tuy nhiên, mỗi quần thể, dân tộc khác nhau sẽ có đặc điểm di truyền khác nhau nên việc xác định mối liên quan của SNP rs3937033 với bệnh ung thư vú cho từng dân tộc là hết sức cần thiết. Tại Việt Nam việc xác định

mối liên quan SNP rs3937033 với bệnh là chưa được thực hiện. Do vậy, trong nghiên cứu này chúng tôi bước đầu tiến hành xác định mối liên quan của SNP rs3937033 và nguy cơ phát triển ung thư vú ở phụ nữ Việt Nam với công cụ tầm soát kiểu gene là phương pháp High Resolution Melt (HRM) tối ưu, nhằm mục đích tìm kiếm các chỉ thị di truyền đặc trưng phục vụ cho chẩn đoán sớm ung thư vú người Việt Nam trong tương lai.

VẬT LIỆU VÀ PHƯƠNG PHÁP

Quần thể nghiên cứu

Quần thể được sử dụng cho nghiên cứu này là phụ nữ người Việt Nam. 100 mẫu máu từ bệnh nhân được chẩn đoán là ung thư vú được thu từ Khoa Ngoại 4 của Bệnh viện Ung Bướu Thành phố Hồ Chí Minh. 100 mẫu máu từ những người khỏe mạnh được thu nhận từ phòng khám Sức Sống và Khoa Ngoại 6 của bệnh Viện Ung Bướu. Tất cả các mẫu được sử dụng đều có sự đồng thuận của bệnh nhân tham gia nghiên cứu. Mẫu sau khi được thu nhận được lưu trữ ở -20°C sẵn sàng cho tách chiết.

Tách chiết DNA

DNA tổng số từ mẫu máu được tách chiết bằng phương pháp muối theo quy trình của N.T. Huệ và cộng sự [16]. Đầu tiên, các tế bào máu trắng sẽ được phân lập và thu nhận lại từ máu tổng số, sau đó được xử lý bằng nhiều hóa chất. DNA tổng số sẽ được thu nhận bằng cách tủa cồn. Sau khi loại bỏ cồn, phần cặn DNA được huyền phù trong nước phân tử dùng cho PCR (nuclease-free water) và lưu trữ ở -20 °C chuẩn bị cho các bước tiếp theo.

Thiết kế và tối ưu hóa phương pháp HRM

HRM là phương pháp xác định kiểu gene của SNP dựa vào sự khác biệt về Tm của đoạn sản phẩm DNA. Hình dạng đường cong nóng chảy với kiểu gene đặc trưng sẽ được hiển thị trên biểu đồ phân tích nhờ vào máy đọc tín hiệu màu huỳnh quang phát ra từ các mạch đôi DNA bị biến tính.

Và do đó, khi tăng dần nhiệt độ, DNA sẽ dần được tách mạch cùng với sự giảm dần tín hiệu huỳnh quang. Các kiểu gene khác nhau sẽ cho các đường cong nóng chảy đặc trưng. Do vậy, các cặp mồi cho HRM được thiết kế nằm gần vị trí SNP để làm giảm kích thước trình tự khuếch đại, tăng sự khác biệt về T_m của các kiểu gene, và do đó sự tách biệt của chúng càng rõ ràng hơn.

Thiết kế mồi

Việc xác định vị trí và thu nhận trình tự nucleotide của SNP rs3937033 được lấy từ NCBI. Các cặp mồi được thiết kế bằng phần mềm Primer3Plus, kiểm tra độ đặc hiệu mồi bằng nhiều phần mềm chuyên dụng như Primer Blast, NCBI Blast tool để chắc chắn không có sản phẩm PCR không mong muốn. Cuối cùng kiểm tra đặc hiệu mồi dựa trên sự hình thành cấu trúc thứ cấp hoặc dimer mồi bằng phần mềm OligoAnalyzer 3.1 (<https://www.idtdna.com/calc/analyzer>).

Dự đoán sản phẩm HRM bằng phần mềm in silico

Trình tự sản phẩm HRM được xác nhận và lấy từ *in silico* PCR của cơ sở dữ liệu UCSC (<https://genome.ucsc.edu/cgi-bin/hgPcr>).

Dự đoán kiểu gene dựa trên đường cong nóng chảy từ phần mềm UMELT HETs (<https://www.dna.utah.edu/hets/umh.php>)

Đoạn trình tự sản phẩm HRM truy xuất từ *in silico* được sử dụng để dự đoán đường cong nóng chảy tương ứng với 3 kiểu gene chuẩn cho SNP rs3937033. Nồng độ $MgCl_2$ và DMSO được khảo sát để kiểm tra sự ảnh hưởng của chúng lên sự tách biệt 3 dạng đường cong nóng chảy.

Tối ưu hóa nhiệt độ bắt cặp mồi (T_a)

Khảo sát trong khoảng nhiệt độ từ 60–70°C bởi chu trình nhiệt của máy PCR Eppendorf và Toptaq Master Mix (QIAGEN). Thành phần cho 25 μ L/phản ứng PCR bao gồm: Toptaq Master Mix 1X, 0,2 μ M mỗi mồi, DNA 10 ng/ μ L và nước phân tử (nuclease-free water) cho đủ thể tích.

Xác định mẫu chứng

Việc xác định các mẫu chứng đại diện cho 3 kiểu gene chuẩn của SNP rs3937033 được thực hiện thông qua phân tích HRM một số mẫu DNA ngẫu nhiên. Thành phần trong 5 μ L/phản ứng HRM bao gồm: Brilliant HRM Ultra-Fast Loci Master Mix 1X (AGILENT), 0,2 μ M cho mỗi mồi, 5 ng/ μ L DNA tổng số và thêm nước cho đủ thể tích.

Xác định kiểu gene

Sau khi có các mẫu chứng đại diện cho 3 kiểu gene, điều kiện của phản ứng tiếp tục được tối ưu hóa để có được sự tách biệt rõ ràng nhất của 3 nhóm đường cong đại diện cho 3 kiểu gene. Điều kiện HRM tốt nhất sẽ được lựa chọn để tầm soát kiểu gene của 100/100 mẫu bệnh/chứng. Phân tích kết quả HRM sẽ được thực hiện bằng phần mềm chuyên dụng LightCycler® 96 SW 1.1. Từ kết quả phân tích kiểu gene từ HRM, tần số các allele C ($f_{(C)}$) và T ($f_{(T)}$) sẽ được tính theo công thức: $f_{(C)} = f_{(CC)} + \frac{1}{2} f_{(CT)}$ và $f_{(T)} = f_{(TT)} + \frac{1}{2} f_{(CT)}$, trong đó f (CC, CT hoặc TT) bằng số cá thể mang kiểu gene cụ thể chia cho tổng số mẫu trong nhóm khảo sát.

Phân tích thống kê

Để xác định tần số các allele và kiểu gene có sự khác biệt đáng kể hay không, chúng tôi sử dụng phần mềm phân tích chuyên dụng STATA ver.12 để thử nghiệm mức ý nghĩa thống kê Chi-Square. Với mức ý nghĩa thống kê đạt giá trị $P < 0,05$ cho thấy SNP rs3937033 có mối liên quan với ung thư vú. Hơn nữa OR [95%CI] cũng được tính toán để đánh giá sự mạnh yếu của mối liên quan SNP rs 3937033 với bệnh.

Ước tính độ tin cậy nghiên cứu

Trong các nghiên cứu di truyền quần thể, cỡ mẫu sử dụng là yếu tố ảnh hưởng đến độ tin cậy của kết quả nghiên cứu. Việc sử dụng đúng cỡ mẫu sẽ mang lại độ tin cậy cao và tiết kiệm được kinh phí thực hiện.

Độ tin cậy của nghiên cứu (power), cũng như số lượng mẫu (sample size) cần thiết cho nghiên cứu trong tương lai được ước tính bằng phần mềm

phân tích của Philippe Glaziou (<http://samsize.sourceforge.net/iface/s3.html>) [17].

KẾT QUẢ VÀ THẢO LUẬN

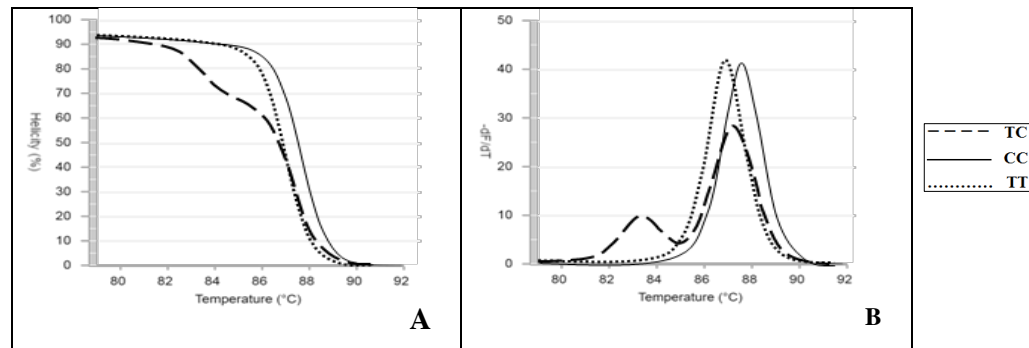
Thiết kế mồi và dự đoán *in silico*

Cặp mồi thiết kế được lựa chọn thỏa mãn các tiêu chí về độ đặc hiệu trên các phần mềm

Primer3Plus, Primer Blast, NCBI Blast tool, OligoAnalyzer 3.1 và dự đoán đường cong nóng chảy rõ ràng từ UMELT HETs (Hình 1). Cặp mồi sử dụng cho giải trình tự để xác nhận kiểu gene chứng từ phân tích HRM được thiết kế sao cho SNP ở vị trí vùng trung tâm của trình tự sản phẩm mục tiêu. Các cặp mồi thiết kế tốt nhất được lựa chọn thể hiện ở Bảng 1.

Bảng 1. Các cặp mồi HRM và giải trình tự

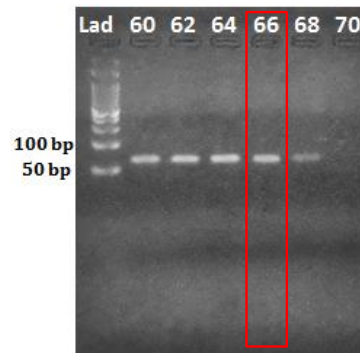
Mồi		Trình tự mồi	Tm mồi	Kích thước mồi	Kích thước sản phẩm
Mồi giải trình tự	Xuôi	CTATCACACATCAACTGCAACTCCAA	67,9 °C	26bp	496bp
	Ngược	ATAGGATGAGGCAGGAATCAAAGACA	68,4 °C	26bp	
Mồi HRM	Xuôi	AGTGAAGTGTATCCGACATCTGGTTCC	69,7 °C	27bp	70bp
	Ngược	AAGGGGAACGTGTGGGCTATTTAGG	70,9 °C	25bp	



Hình 1. Kết quả phân tích bằng biểu đồ nhằm dự đoán đường cong nóng chảy. A: đường cong nóng chảy, B: định nóng chảy trên UMELT HETs ở nồng độ Mg^{2+} 1,5mM, DMSO 0% của SNP rs3937033

Tối ưu hóa nhiệt độ bắt cặp mồi

Để tối ưu hóa nhiệt độ bắt cặp mồi, khoảng nhiệt độ từ 60 đến 70 °C được khảo sát, kết quả được phân tích trên băng điện di gel agarose 1,5% (Hình 2). Kết quả điện di trên gel cho thấy rằng các vạch sản phẩm hiện diện ở nhiệt độ từ 60–68°C. Tuy nhiên để đảm bảo độ đặc hiệu tối đa và lượng sản phẩm mục tiêu đủ cho phân tích HRM tốt nhất chúng tôi chọn 66 °C là nhiệt độ bắt cặp tối ưu cho phân tích HRM tiếp theo.

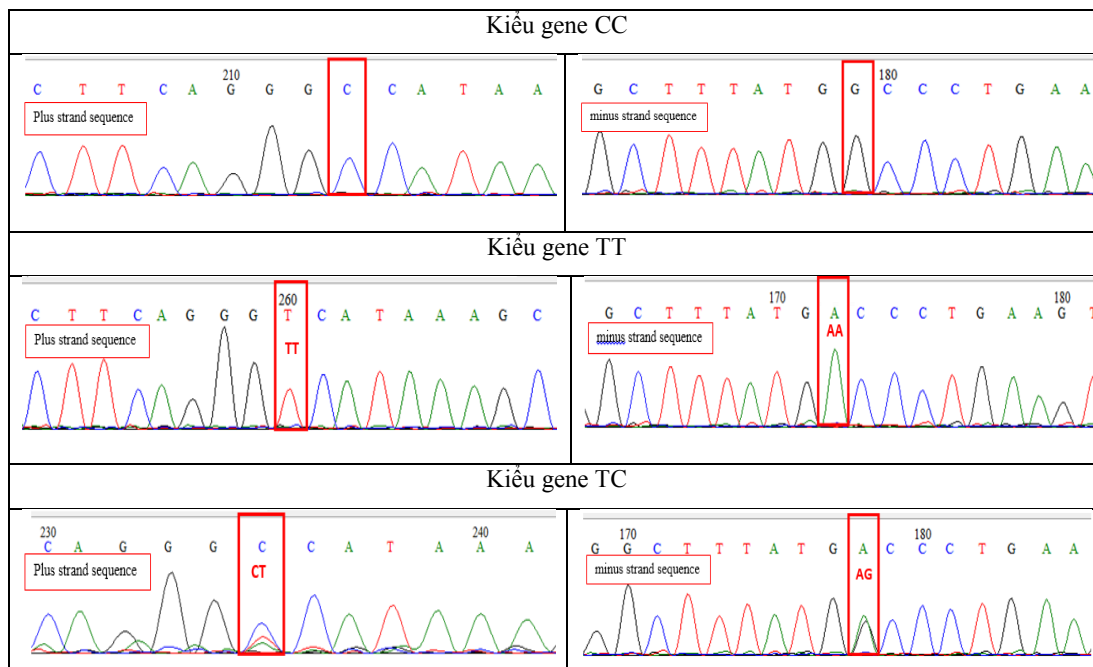


Hình 2. Kết quả điện di sản phẩm PCR nhằm xác định nhiệt độ bắt cặp mồi tối ưu

Xác định mẫu chứng

Với điều kiện phản ứng HRM được lựa chọn: Brilliant HRM Ultra-Fast Loci Master Mix 1X và Ta tối ưu 66⁰C, một số mẫu DNA ngẫu nhiên được tiến hành phân tích HRM để tìm 3 mẫu chứng đại diện cho 3 kiểu gene chuẩn của SNP rs3937033. Dựa trên sự tách biệt rõ ràng nhất của các nhóm đường cong nóng chảy được hiển thị từ phần mềm LightCycler® 96 SW 1.1, một số mẫu đại diện cho

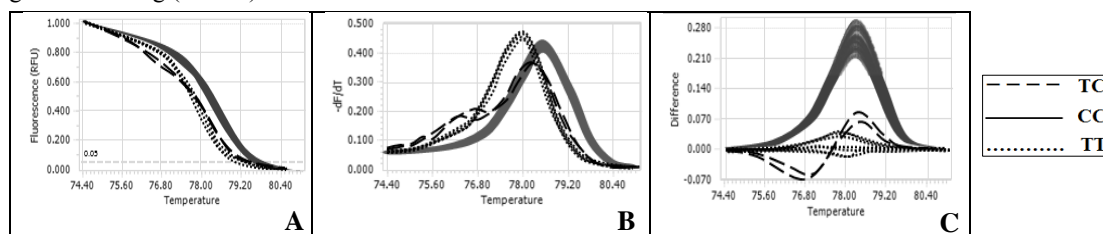
mỗi nhóm đường cong nóng chảy sẽ được tiến hành giải trình tự để xác nhận kiểu gene được dự đoán từ phân tích HRM. Kết quả giải trình tự được phân tích bởi phần mềm Sequencing Analysis 5.3. Kết quả nhận được đúng như mong đợi, với 3 kiểu đường cong nóng chảy đặc trưng. Chúng tôi nhận được 3 kiểu gene của SNP: CC, TT và TC (Hình 3).



Hình 3. Kết quả giải trình tự xác nhận ba mẫu chứng

Xác định kiểu gene cho bộ mẫu bệnh chứng

Dựa trên điều kiện HRM đã tối ưu và các mẫu chứng dương đại diện cho 3 kiểu gene chuẩn TT, CC, và TC của SNP vừa xác định, 100/100 mẫu bệnh/chứng được tiến hành phân tích HRM để xác định kiểu gene của chúng (Hình 4).



Hình 4. Kết quả phân tích HRM xác định kiểu gene 100/100 mẫu bệnh chứng. A) đường cong nóng chảy, B) đỉnh nóng chảy, C) cụm biểu đồ khác biệt

Xác định tần số kiểu gen, allele và xác định mối liên quan của SNP rs3937033 với bệnh

Dựa trên kết quả xác định kiểu gene từ phân tích HRM, việc tính toán tần số kiểu gene và allele được thực hiện. Kết quả cho thấy tần số của kiểu gene TT, TC và CC tương ứng ở nhóm bệnh là 36 %, 3 % và 61 % và ở nhóm người khỏe mạnh là 23 %, 1 % và 76 %. Tần số allele T và C được xác định trên nhóm bệnh là 37,5 % và 62,5 % và trong nhóm chứng là 23,5 % và 76,5 %. Kết quả cho thấy kiểu gene CC và allele C xuất hiện thường xuyên hơn ở nhóm chứng. Bên cạnh đó, phân tích Chi-square (χ^2) chỉ rõ sự khác biệt đáng kể giữa nhóm

bệnh và nhóm chứng ở những người mang allele T khi kiểm định $\chi^2=4,77$ đạt mức ý nghĩa thống kê ($P=0,03$), tỷ số nguy cơ $OR_{alen T}$ [95% CI] =1,41 [1,03 – 1,92], $P = 0,03$). Kiểm định χ^2 của kiểu gene chứa allele T bằng 5,58, gần đạt mức ý nghĩa thống kê khi $P=0,06$ chứng tỏ rằng kiểu gene có chứa allele T có thể mang nguy cơ liên quan đến bệnh. Phân tích cụ thể từng mô hình di truyền kiểu gene cho thấy: TT vs. CC: OR [95% CI] =1,95 [1,05 – 3,63], $P = 0,03$); TT+TC vs. CC: OR [95% CI] =2,03 [1,10– 3,73], $P = 0,02$ (Bảng 2). Với những kết quả phân tích ở trên độ tin cậy của nghiên cứu đạt được là 18,67 %.

Bảng 2. Tần số kiểu gene và allele của rs3937033

	Kiểu gen			Alen	
	TT	TC	CC	T	C
100 mẫu ung thư vú	36 (36%)	3 (3%)	61 (61%)	75 (37,5%)	125(62,5%)
100 người khỏe mạnh	23 (23%)	1 (1%)	76 (76%)	47 (23,5%)	153(76,5%)
χ^2 P-value	$\chi^2 = 5,58$ $P = 0,06$			$\chi^2 = 4,77$ $P = 0,03$	

Bảng 3. Phân tích mối liên quan allele và kiểu gene của rs3937033 với bệnh

Bệnh/chứng 100/100	T vs. C	TC vs. CC	TT vs. CC	TT+TC vs CC
OR [95%CI]	1,41[1,03-1,92]	3,74[0,38-36,84]	1,95[1,05-3,63]	2,03[1,10-3,73]
P-value	0,03	0,23	0,03	0,02
Độ tin cậy	18,67%			

Ước tính cỡ mẫu để tăng độ tin cậy nghiên cứu

Dựa trên kết quả phân tích ở bộ mẫu 100 bệnh/chứng ban đầu cho thấy SNP có mối liên quan với bệnh ung thư vú người Việt Nam, nhưng độ tin cậy nghiên cứu là khá thấp 18,67 %, do vậy chúng tôi tiến hành ước tính sự gia tăng độ tin cậy của nghiên cứu khi tăng bộ mẫu nghiên cứu. Kết quả phân tích cho thấy khi tăng cỡ mẫu lên trên 428 mẫu bệnh/chứng độ tin cậy nghiên cứu được cải thiện đáng kể lên đến ≥ 60 %. Để đạt được độ

tin cậy 90 % cỡ mẫu ước tính cần có là 916 mẫu bệnh/chứng (Bảng 4).

Bảng 4. Ước tính cỡ mẫu cho SNP rs3937033 với sự tăng độ tin cậy cho nghiên cứu

SNPs	Power			
	60 %	70 %	80 %	90 %
Rs3937033	428/428	539/539	685/685	916/916

Trong nghiên cứu này, chúng tôi tiến hành xác định mối liên hệ của SNPs rs3937033 với bệnh ung thư vú ở người Việt Nam sử dụng công cụ tầm soát kiểu gene là phương pháp HRM. Đây là một

phương pháp xác định kiểu gene với độ phân giải cao, có thể phân biệt được các trình tự DNA chỉ khác nhau 1 nucleotide dựa nhiệt độ nóng chảy (T_m) của sản phẩm PCR. Các cặp mồi được thiết kế đặc hiệu cho phân tích HRM bởi các phần mềm chuyên dụng như Primer3Plus, Primer Blast, NCBI Blast tool và OligoAnalyzer3.1. Cặp mồi được lựa chọn chỉ khi đáp ứng đầy đủ điều kiện về độ đặc hiệu và phân biệt rõ 3 dạng đường cong nóng chảy từ dự đoán UMELT HETs. Trong vùng nhiệt độ bắt cặp mồi từ 60 – 70 °C được khảo sát cho phân tích HRM cho thấy 66 °C là nhiệt độ thích hợp được lựa chọn. Với điều kiện ban đầu này, một số mẫu DNA ngẫu nhiên được tiến hành phân tích HRM để tìm các mẫu chứng đại diện cho 3 kiểu gene chuẩn dựa trên 3 kiểu đường cong nóng chảy của HRM. Các mẫu DNA đại diện cho 3 nhóm đường cong nóng chảy của HRM được tiến hành giải trình tự tự động để xác nhận kiểu gen. Với 3 mẫu chuẩn vừa được xác định và 1 mẫu chứng âm (sử dụng nuclease-free water làm chứng âm) chúng tôi tiến hành tầm soát kiểu gene cho bộ mẫu 100/100 mẫu bệnh/chứng. Kết quả phân tích cho thấy tần số kiểu gene TT, TC và CC tương ứng ở nhóm bệnh là 36 %, 3 % và 61 % và ở nhóm người khỏe mạnh là 23 %, 1 % và 76 %, tần số allele T và C tương ứng ở nhóm bệnh là 37,5 % và 62,5 % và nhóm chứng là 23,5 % và 76,5 %. Phân tích thống kê kiểm định Chi-square cho thấy giá trị OR khi phân tích trên allele bằng 1,41 với 95% CI = 1,3 – 1,92 và có ý nghĩa về mặt thống kê ($P=0,03$) (Bảng3). Điều này cho thấy rằng allele T là allele nguy cơ đóng góp vào sự phát triển ung thư vú, với sự thay đổi allele T thành C cho thấy làm giảm nguy cơ mắc bệnh. Bên cạnh đó, kết quả phân tích trên kiểu gene đồng hợp TT so với CC [OR [95% CI] =1,95 [1,05 – 3,63], $P=0,03$] và sự kết hợp của kiểu gene TT và TC với CC [OR [95% CI] =2,03 [1,10– 3,73], $P=0,02$] tăng nguy cơ mắc ung thư vú người Việt Nam. Khác với kết luận của chúng tôi, một nghiên cứu di truyền quần thể trên 156 bệnh nhân ung thư vú

và 260 người khỏe mạnh người Maryland (Âu Mỹ) cho thấy sự thay đổi trình tự T>C ở SNP rs3937033 có mối liên quan đến nguy cơ tăng ung thư vú ($P=0,036$; OR=1,98; 95%CI=1,25 – 3,12) [15]. Một nghiên cứu tiếp tục trên quần thể Âu Mỹ 156/260 bệnh/chứng cho thấy khi có sự tương tác đồng thời của 2 biến đổi T>C của rs3937033 và rs2853826 A>G cho thấy làm tăng nguy cơ ung thư vú ($P=0,02$) [18]. Một nghiên cứu khác trên quần thể người da trắng cho thấy sự thay đổi T>C chiếm 90% ở người mang đột biến *BRCA1* có mối liên quan với bệnh nhân ung thư vú bị đột biến *BRCA1* (OR = 21, 95%CI [2,15 – 204,6], $P=0,003$) khi so sánh với những người không mang đột biến *BRCA1* (30 %), bên cạnh đó rs3937033 còn cho thấy là có liên quan đến giai đoạn biệt hóa của khối u khi xuất hiện 57 % giai đoạn 3 (G3 – tumor grade 3 differentiation) và 12,5 % ở giai đoạn 2 (G2) ($P=0,05$) [14] và được tìm thấy trước đó là có liên quan đến nguy cơ làm tăng ung thư vú gia đình [15]. Điều này cho thấy rằng, rs3937033 là một trong những nóng thường xảy ra đột biến đóng vai trò quan trọng trong nguy cơ phát triển ung thư vú [14]. Hơn nữa rs3937033 là SNP nằm trong vùng quan trọng của mtDNA – D-LOOP chịu trách nhiệm khởi động và kiểm soát quá trình sao chép và phiên mã gen. Đột biến xảy ra trong vùng D-Loop dẫn đến thay đổi biểu hiện gen, mất khả năng điều hòa hệ thống OXPHOS và quá trình chuyển hóa các chất trong ty thể [7, 19]. Do vậy sự thay đổi di truyền T>C ở SNP rs3937033 rất có thể là một trong những nguyên nhân góp phần làm rối loạn quá trình sao chép và phiên mã, từ đó góp phần vào nguy cơ phát triển ung thư vú. Tuy nhiên, trong nghiên cứu của chúng tôi, việc xác định mối liên quan rs3937033 với nguy cơ mắc ung thư vú trên quần thể Việt Nam với bộ mẫu 100 bệnh/chứng mặc dù cho thấy khả năng rất cao có mối liên quan với bệnh nhưng chỉ đạt giá trị tin cậy (Power) là 18,67 %, chứng tỏ kết quả nghiên cứu có độ tin cậy còn quá thấp. Và do đó, rất có thể cỡ mẫu sử dụng cho nghiên cứu

chính là nguyên nhân làm giảm giá trị tin cậy của nghiên cứu, việc xác định mối liên quan của rs3937033 cần được thực hiện với cỡ mẫu lớn hơn để đạt độ tin cậy nghiên cứu cao hơn (power 90 % được đề xuất). Dựa trên khác biệt về tần số allele giữa nhóm bệnh và nhóm chứng (odd ratio), ước tính số cỡ mẫu bệnh/chứng cần sử dụng để tăng độ tin cậy nghiên cứu đã được thực hiện. Kết quả phân tích cho thấy rõ ràng việc tăng độ tin cậy của nghiên cứu phụ thuộc vào số lượng cỡ mẫu: với bộ mẫu 100 bệnh/chứng độ tin cậy chỉ đạt 18,67 %, nhưng nếu cỡ mẫu được tăng lên 916 bệnh/chứng thì độ tin cậy nghiên cứu sẽ đạt đến 90 %. Điều đó cho thấy rằng việc ước tính để dự đoán cỡ mẫu sử dụng và độ tin cậy nghiên cứu là hết sức cần thiết cho một nghiên cứu di truyền quần thể, một cỡ mẫu thích hợp được sử dụng có vai trò quan trọng trong nghiên cứu bệnh/chứng nhằm tiết kiệm kinh phí, cũng như mang lại hiệu quả cao nhất và chính xác nhất có thể.

KẾT LUẬN

Trong nghiên cứu bệnh/chứng này cho thấy rằng SNP rs3937033 là một chỉ thị nhạy cảm với nguy cơ mắc ung thư vú quần thể Việt Nam với $OR=1,41$; 95 % $CI = [1,3 - 1,92]$, $P=0,03$. Tuy nhiên, giá trị tin cậy nghiên cứu này là khá thấp chỉ có 18,67 %, do đó cần thiết phải tăng cỡ mẫu lên trên 916 mẫu bệnh/chứng để đảm bảo đạt đến độ tin cậy 90 %, cần thiết cho một nghiên cứu di truyền quần thể. Như vậy, với kết quả giới hạn trong nghiên cứu 100 bệnh/chứng này của chúng tôi cung cấp thông tin ban đầu để định hướng cho việc thực hiện một nghiên cứu lớn hơn và sâu sắc hơn để có thể xác định một cách chính xác và tin cậy nhất về mối liên hệ của SNP rs3937033 với bệnh ung thư vú, từ đó cung cấp những thông tin đầy đủ hơn cho việc chẩn đoán và chữa trị bệnh ung thư vú phụ nữ Việt Nam trong tương lai.

Lời cảm ơn: Chúng tôi cảm ơn tất cả các bác sĩ và nhân viên của Bệnh viện Ung Bướu thành phố Hồ Chí Minh, Việt Nam đã cung cấp mẫu máu cho nghiên cứu này. Nghiên cứu này được tài trợ bởi Đại học Quốc Gia thành phố Hồ Chí Minh – C2016-18-01.

Initial determination of the association between SNP rs3937033 on the D-LOOP mtDNA and breast cancer in Vietnam

- **Duong Thi Hong Hanh**
- **Nguyen Thi Ngoc Thanh**
- **Nguyen Thi Hue**
University of Science, VNU-HCM

ABSTRACT

Breast cancer (BC) is the most common cancer among women and accounts for the highest mortality rates in the developing countries, including Vietnam. D-loop, is a non-coding region in mitochondrial circular DNA molecules, related to control the replication and transcription of mitochondrial genes. D-loop mutations cause the dysfunction of the respiratory chain, which contributes to the process of BC. SNP rs3937033

(T/C) on D-loop was reported to have the association with BC in Europe America population ($OR [95\%CI]= 1.98 [1.25-3.12]$, $P=0.036$), and Caucasian ($OR[95\%CI]= 21 [2.15-204.6]$, $P=0.003$). In this study, the High Resolution Melting (HRM) method is optimized for genotyping 100/100 cases/controls samples, then to determine the association between the disease and this SNP. TT, TC and CC genotype

frequencies in patient group (36 %, 3 % and 61 %) and control group (23 %, 1 % and 76 %) were calculated based on the results of genotyping by optimal HRM method with the annealing temperature is 66 °C. Association analysis result showed that T allele increases the risk of breast

cancer patients up to 1.41 times (OR [95% CI] = 1.41 [1.3-1.92], $P_{allele} = 0.03$). Therefore, SNP rs3937033 could be a potential biomarker for early BC diagnosis in Vietnamese. However, this research needs to be conducted with a larger sample size to reach the greater confidence.

Keyword: mtDNA, D-loop, breast cancer, SNP T16519C, rs3937033

TÀI LIỆU THAM KHẢO

- [1]. L.A.Torre, et al., *Global cancer statistics, 2012*. CA Cancer J Clin, 2015.
- [2]. GLOBOCAN, *GLOBOCAN 2012: Estimated Incidence, Mortality and Prevalence Worldwide in 2012; Available from: http://globocan.iarc.fr/Pages/fact_sheets_cancer.aspx*.
- [3]. N.H.Lan, W. Laohasiriwong, J.F. Stewart, Survival probability and prognostic factors for breast cancer patients in Vietnam, *Global Health Action*, 17, 6, 1–9 (2013).
- [4]. M.Yu, et al., Sequence variations of mitochondrial DNA D-loop region are highly frequent events in familial breast cancer, *Journal of Biomedical Science*, 15, 4, 535–543 (2008).
- [5]. D.C.Chan, Mitochondria: dynamic organelles in disease, aging, and development, *Cell*, 125, 7, 1241–1252 (2006).
- [6]. A.Smith, R. Staden, I. Young, Sequence and organization of the human mitochondrial genome, *Nature*, 290, 5806, 457–465 (1981).
- [7]. H.C. Lee, et al., Mitochondrial genome instability and mtDNA depletion in human cancers, *Annals of the New York Academy of Sciences*, 1042, 1, 109–122 (2005).
- [8]. D.L. Croteau, V.A. Bohr, Repair of oxidative damage to nuclear and mitochondrial DNA in mammalian cells, *Journal of Biological Chemistry*, 272, 41, 25409–25412 (1997).
- [9]. Clayton, D.A., Transcription and replication of mitochondrial DNA, *Human Reproduction*, 15(suppl 2), 11–17 (2000).
- [10]. N.O.Bianchi, M.S. Bianchi, S.M. Richard, Mitochondrial genome instability in human cancers, *Mutation Research/Reviews in Mutation Research*, 488, 1, 9–23 (2001).
- [11]. A.Lièvre, et al., Clinical value of mitochondrial mutations in colorectal cancer, *Journal of Clinical Oncology*, 23, 15, 3517–3525 (2005).
- [12]. P.C.Goswami, Mutant mitochondria and cancer cell metastasis: quest for a mechanism, *Cancer Biology & Therapy*, 8, 14, 1386–1388 (2009).
- [13]. M.Kulawiec, K.M. Owens, K.K. Singh, Cancer cell mitochondria confer apoptosis resistance and promote metastasis, *Cancer Biology & Therapy*, 8, 14, 1378–1385 (2009).
- [14]. S.Tommasi, et al., Mitochondrial DNA variants and risk of familial breast cancer: an exploratory study, *International Journal of Oncology*, 44, 5, 1691–1698 (2014).
- [15]. R.K.Bai, et al., Mitochondrial genetic background modifies breast cancer risk, *Cancer Research*, 67, 10, 4687–4694 (2007).
- [16]. N.T.Hue, et al., Extraction of human genomic DNA from dried blood spots and hair roots, *International Journal of Bioscience, Biochemistry and Bioinformatics*, 2, 1, 21, 2012..
- [17]. Glaziou, P. *Sample Size*. Available from: <http://sampsizemethod.sourceforge.net/iface/s3.html>, 2005 [cited 2015 August 25th].
- [18]. D.Covarrubias, et al., Mitochondrial DNA variant interactions modify breast cancer risk, *Journal of Human Genetics*, 53, 10, 924–928 (2008).
- [19]. K.K. Singh, M. Kulawiec, Mitochondrial DNA polymorphism and risk of cancer, *Cancer Epidemiology*, 291–303 (2009).