

ỨNG DỤNG PHƯƠNG PHÁP GIẢI TRÌNH TỰ TOÀN BỘ VÙNG GEN MÃ HÓA TRONG VIỆC XÁC ĐỊNH SƠ BỘ BIẾN THỂ DI TRUYỀN Ở BỆNH NHÂN MẮC DỊ TẬT VAN TIM BẨM SINH

USING WHOLE EXOME SEQUENCING TO PRELIMINARY ASSESSMENT OF GENETIC VARIATIONS IN PATIENTS WITH CONGENITAL HEART VALVE DEFECTS

*Nguyễn Hoàng Thanh Trang**, *Nguyễn Thị Kim Liên†*, *Nguyễn Văn Tụng‡*,
Nguyễn Huy Hoàng§, *Trần Đức Đại¶*

Ngày tòa soạn nhận được bài báo: 03/11/2021

Ngày nhận kết quả phản biện đánh giá: 04/05/2022

Ngày bài báo được duyệt đăng: 27/05/2022

Tóm tắt: Dị tật van tim bẩm sinh đặc trưng bởi một hoặc nhiều van tim phát triển bất thường. Có một số nguyên nhân phổ biến gây ra bệnh như nhiễm độc và nhiễm bệnh trong thời gian thai kỳ đặc biệt là do di truyền. Giải trình tự toàn bộ vùng gen mã hóa cho phép xác định biến thể di truyền trên đồng thời nhiều gen được coi là phương pháp thích hợp trong nghiên cứu di truyền dị tật van tim bẩm sinh. Nghiên cứu này thực hiện giải trình tự toàn bộ vùng gen mã hóa của một bệnh nhân mắc dị tật van tim bẩm sinh từ đó phân tích xác định được 82.556 đột biến dạng thay thế nucleotide và 11.334 đột biến thêm bớt nucleotide trên toàn bộ vùng mã hóa bao gồm cả những đột biến đã được báo cáo trên cơ sở dữ liệu dbSNP và đột biến mới. Kết quả của nghiên cứu cho thấy tiềm năng của việc sử dụng công nghệ giải trình tự toàn bộ vùng gen mã hóa trong nghiên cứu và chẩn đoán dị tật van tim bẩm sinh.

Từ khóa: Dị tật van tim bẩm sinh, đột biến gen, giải trình tự toàn bộ vùng mã hóa, giải trình tự thế hệ mới, tin sinh học

Abstract: Congenital valvular heart valve defects are characterized by abnormality of the heart valves, such as any valve in the heart that has damage or missing. There are several causes of this disease such as infections, degenerative conditions and genetic variants. Whole exome sequencing (WES) allows simultaneous analysis of variants of multiple or even all

* Trường Đại học Mở Hà Nội

† Viện Nghiên cứu hệ gen – Viện Hàn lâm Khoa học và Công nghệ Việt Nam

‡ Viện Nghiên cứu hệ gen – Viện Hàn lâm Khoa học và Công nghệ Việt Nam

§ Viện Nghiên cứu hệ gen – Viện Hàn lâm Khoa học và Công nghệ Việt Nam

¶ Bệnh viện E

genes, thereby reducing the time needed to diagnose for patients. Therefore, WES has been considered as an effective tool for the detection of novel causal genes in the study of genetics of the heart valve defects. In this study, by applying whole exome sequencing in 01 patient with congenital heart valve defects, we detected 82,556 missense and 11.334 indel including variants were reported in the database of single nucleotide polymorphisms (dbSNP) and novel variants. The result of this study shows potential of WES in genetic research, particularly in the identification of inherited genetic disorders.

Keywords: Bioinformatics, genetic variant, next generation sequencing, valvular heart disease, whole exome sequencing, bioinformatics.

I. Đặt vấn đề

Dị tật van tim bẩm sinh (Valvular heart disease – VHD) là tình trạng khi một hoặc nhiều van trong bốn van tim không được phát triển đúng cách trong thời gian còn là phôi thai, gây nên những khiếm khuyết về cấu trúc tim và làm ảnh hưởng đến quá trình lưu thông máu. Dị tật van tim bao gồm cả hai dạng bẩm sinh và mắc phải là một vấn đề sức khỏe cộng đồng quan trọng và ngày càng tăng. Dựa trên các nghiên cứu dịch tễ học ở Hoa Kỳ, tỷ lệ mắc bệnh là 2,5%, và tỷ lệ mắc bệnh tăng theo tuổi tác. Tỷ lệ dị tật van tim bẩm sinh chiếm 10% trong tổng số dị tật tim bẩm sinh được phát hiện. Các dị tật van tim bẩm sinh thường gặp nhất và là nguyên nhân gây tử vong hàng đầu trong số các dị tật bẩm sinh ở trẻ. Ngày nay với những tiến bộ trong chẩn đoán và điều trị đã làm tăng đáng kể tỷ lệ sống của những trường hợp tim bẩm sinh phức tạp.

Vai trò quan trọng của các yếu tố di truyền trong bệnh van tim của con người ngày càng trở nên rõ ràng. Nguyên nhân di truyền của các dị tật van tim bẩm sinh được xác định là do đột biến trên nhiều gen gây ra như *NOTCH1*, *GATA5*, *TGFBR1* và *TGFBR2* [1], [2]. Năm 2016, nhóm nghiên cứu Dargis và các đồng tác giả sử dụng phương pháp giải trình tự gen

thể hệ mới qua đó xác định được 9 gen liên quan đến dị tật van tim bẩm sinh gồm: *NOTCH1*, *AXINI*, *EGFR*, *ENG*, *GATA5*, *NKX2-5*, *NOS3*, *PDIA2*, *TGFBR2* [3]. Sự phát triển của phương pháp giải trình tự thể hệ mới đã tạo điều kiện thuận lợi cho giải trình tự gen một cách nhanh chóng trong y học. Giải trình tự gen thể hệ mới cho phép phân tích đồng thời nhiều hoặc thậm chí tất cả các gen do đó giảm thời gian chẩn đoán cho nhiều bệnh nhân. Giải trình tự vùng mã hóa - Whole exome sequencing (WES) là một ứng dụng của công nghệ giải trình tự thể hệ mới để xác định các biến thể trên tất cả các vùng mã hóa, hoặc exon của gen được biết đến. Vì thế WES đã được sử dụng rộng rãi trong các nghiên cứu lâm sàng vài năm gần đây, đặc biệt trong việc xác định các gen bệnh di truyền. Hàng chục nghìn biến thể gen có thể được xác định trong exome và WES đang được coi là hướng đi đúng đắn để nghiên cứu di truyền. WES đã được ứng dụng trong nghiên cứu di truyền đặc biệt là đối với những bệnh có tính không đồng nhất cao như tim mạch và rối loạn cơ xương [4], bệnh thần kinh [5].

Dị tật van tim bẩm sinh do biến dị di truyền trên nhiều gen gây ra. Với số lượng và kích thước gen lớn, việc nghiên cứu từng gen riêng lẻ đòi hỏi nhiều thời gian, chi phí cao. Giải trình tự toàn bộ vùng mã

hóa bằng công nghệ giải trình tự thế hệ mới cho phép xác định biến thể xảy ra đồng thời trên nhiều gen, qua đó rút ngắn quá trình phân tích đã trở thành phương pháp thay thế hiệu quả khi tiến hành các nghiên cứu di truyền. Trong nghiên cứu này chúng tôi sử dụng phương pháp giải trình tự toàn bộ vùng gen mã hóa nhằm xác định biến đổi di truyền ở bệnh nhân mắc dị tật van tim bẩm sinh, qua đó đánh giá tính khả thi và hiệu quả của phương pháp này trong nghiên cứu.

II. Phương pháp nghiên cứu

2.1. Vật liệu nghiên cứu

Vật liệu nghiên cứu là mẫu máu của bệnh nhân mắc dị tật tim bẩm sinh được cung cấp bởi Trung tâm tim mạch - Bệnh viện E.

2.2. Phương pháp nghiên cứu

Phương pháp nghiên cứu bao gồm các bước như sau: thu thập mẫu máu của bệnh nhân, tách chiết DNA tổng số từ mẫu máu của bệnh nhân sau khi thu nhận, giải trình tự toàn bộ vùng gen mã hóa, phân tích số liệu thu được và sàng lọc ra các biến thể/ đột biến có khả năng gây bệnh.

2.2.1. Tách chiết DNA tổng số

DNA tổng số được tách chiết từ mẫu máu toàn phần của bệnh nhân bằng bộ kit QIAamp DNA Blood Mini Kit (QIAGEN, Đức).

2.2.2. Phương pháp giải trình tự WES

DNA tổng số sẽ được cắt, lai và tinh sạch bằng bộ kit SureSelect All Exon V7 để tạo thư viện. Quá trình tạo thư viện được thực hiện theo 4 bước chính: phân cắt DNA tổng số thành những đoạn nhỏ; gắn mỗi giải trình tự và index; khuếch đại

thư viện; tinh sạch và kiểm tra chất lượng. Nồng độ và kích thước thư viện được kiểm tra bằng máy Bioanalyzer 2100.

Thư viện exome sau khi được xử lý, được tiến hành giải trình tự trên máy giải trình tự thế hệ mới Illumina theo hướng dẫn của nhà sản xuất.

2.2.3. Phân tích tin sinh học để dò tìm các biến thể.

Sau khi giải trình tự trên máy Illumina, chất lượng đọc được kiểm tra bằng phần mềm FastQC. Dữ liệu trình tự được sắp xếp và so sánh với trình tự hệ gen người (hg19) bằng phần mềm BWA 0.7.10 [6]. Công cụ Picard được sử dụng để xử lý dữ liệu sau khi giống hàng. Các biến thể được phát hiện bằng phần mềm Genome Analysis Toolkit v3.4 [7]. Ảnh hưởng của biến thể được xác định bằng các phần mềm SnpEff v4.1 [8]. Đây là công cụ chú thích và dự báo ảnh hưởng của các biến thể gen (như thay đổi axit amin). Dữ liệu đầu vào của công cụ này là các biến thể được dự đoán (SNPs, chèn, xóa và MNPs), là kết quả của giải trình tự, và có định dạng VCF (Variant Call Format). Ảnh hưởng của đột biến đến chức năng protein được đánh giá bằng các phần mềm SIFT [9] và Polyphen2 [10].

III. Kết quả và thảo luận

3.1. Thu thập mẫu máu

Đề tài thu thập mẫu máu của 01 bệnh nhân và các thành viên trong gia đình bao gồm bố và mẹ. Bố mẹ của bệnh nhân đều bình thường, bệnh nhân được chẩn đoán hẹp trên van động mạch phổi kèm theo các biểu hiện lâm sàng như bộ mặt bất thường, chậm phát triển thần kinh vận động và bị suy tim cấp độ 1. Bệnh nhân được chẩn

đoán bị tim bẩm sinh, và ở lại bệnh viện để điều trị cho tới hiện tại. Việc phân tích di truyền được tiến hành với sự đồng ý của bố mẹ bệnh nhân.

3.2. Tách chiết DNA tổng số

Mẫu máu của các bệnh nhân được tách chiết bằng bộ kit QIAamp DNA. Sản phẩm DNA tổng số của các mẫu được điện di trên gel agarose 1% để kiểm tra chất lượng và độ tinh sạch.

Hình ảnh trên điện di đồ cho thấy DNA tổng số tách từ mẫu máu bệnh nhân và các thành viên trong gia đình có chất lượng tốt, hiện băng rõ ràng, không bị đứt gãy và không lẫn tạp chất (Hình 1). DNA tổng số được lưu giữ và bảo quản trong ở nhiệt độ -20°C .

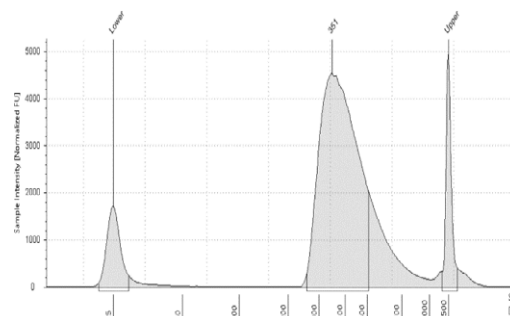


Hình 1: Điện di đồ đồ sản phẩm DNA tách chiết từ mẫu máu bệnh nhân

3.3. Kết quả tạo thư viện và giải trình tự toàn bộ vùng gen mã hóa

Thư viện DNA được tạo bằng kit SureSelect All Exon V6 theo hướng dẫn của nhà sản xuất. Để kiểm tra kích thước của mảnh PCR khuếch đại, chúng tôi kiểm tra sự phân bố kích thước mẫu bằng cách chạy mẫu trên máy Agilent Technologies 2100 Bioanalyzer sử dụng một chip DNA

1000. Thư viện DNA tạo ra có kích thước 351 bp, nồng độ 122,27 ng/ μl (Hình 2).



Hình 2. Phân bố kích thước thư viện DNA

Kết quả kiểm định chất lượng đọc trình tự ban đầu bằng phần mềm FastQC cho thấy mẫu thu được có số lượng trình tự đọc lớn, bao gồm 62.552.002 đoạn đọc. Tỷ lệ trình tự có điểm chất lượng Phred trên 30 (Q30) là 93,3%. Trung vị độ sâu bao phủ vùng quan tâm của các mẫu nằm trong khoảng 90,4–100,1X. Tỷ lệ % với độ sâu bao phủ $>30\text{X}$ của các mẫu đều lớn hơn 80%. Các kết quả trên cho thấy các trình tự thu được là có chất lượng cao, đủ điều kiện để tiến hành phân tích nhằm xác định đột biến di truyền ở bệnh nhân. Thông tin chất lượng đọc trình tự được thể hiện ở Bảng 1.

Bảng 1. Thông tin chất lượng đọc trình tự

Tổng số nucleotide	9.378.354.365
Tổng số trình tự đọc	62.552.002
Q20 (%)	97,6
Q30 (%)	93,3
Trung vị độ sâu bao phủ vùng quan tâm (x)	91,6
% với độ sâu bao phủ $>30\text{X}$	89,8

3.4. Kết quả phân tích số liệu và sàng lọc biến thể

Sau khi xác định và chú giải biến thể, một số lượng lớn các biến thể thay thế một nucleotide (82.556 biến thể) và các biến thể thêm mất nucleotide (11.334

biến thể) được phát hiện trong dữ liệu WES (Bảng 2). Các biến thể được chia thành các nhóm theo mức độ ảnh hưởng chức năng của biến thể như biến thể đồng nghĩa, biến thể sai nghĩa, thêm/mất một bộ ba mã hóa kết thúc, biến thể dịch khung, thêm/mất bộ ba mã hóa...

Trong tổng số 82,556 biến thể được tìm thấy ở bệnh nhân tham gia nghiên cứu, có 8 biến thể trên 8 gen gây bệnh có tần suất alen <0,01 (Bảng 3). Trong đó có 6 biến thể ở trạng thái dị hợp tử bao gồm c.17_18delCC trên gen *NOTCH2*, c.751C>T trên gen *ABCG5*, c.8069C>T trên gen *TTN*, c.1675G>A trên gen *ROR2*,

c.389A>G trên gen *MAP2K1*, c.2101A>G trên gen *FANCA*; 2 biến thể ở trạng thái đồng hợp tử c.163_166dupGATG trên gen *LFNG*, c.85G>A trên gen *AMERI*.

Bảng 2. Kết quả xác định và chú giải biến thể

Tổng SNP	82.556
Biến thể đồng nghĩa	12.008
Biến thể sai nghĩa	11.500
Thêm bộ ba mã hóa kết thúc	114
Mất bộ ba mã hóa kết thúc	34
Tổng số biến thể Indel	11.334
Biến thể dịch khung	332
Thêm bộ ba mã hóa	196
Mất bộ ba mã hóa	231
Đã được báo cáo trên dbSNP142 (%)	96,0

Bảng 3. Kết quả lọc các đột biến có ảnh hưởng chức năng có khả năng liên quan đến bệnh trên bệnh nhân mắc dị tật van tim bẩm sinh

Gen	Di truyền	Trạng thái	Ảnh hưởng	Exon	Thay đổi DNA	Thay đổi protein
<i>NOTCH2</i>	Dị hợp tử	Trội	Dịch khung	1/34	c.17_18delCC	p.Pro6fs
<i>ABCG5</i>	Dị hợp tử	Lặn	Vô nghĩa	6/13	c.751C>T	p.Gln251*
<i>TTN</i>	Dị hợp tử	Trội	Sai nghĩa	34/363	c.8069C>T	p.Thr2690Ile
<i>LFNG</i>	Đồng hợp tử	Lặn	Chèn	2/9	c.163_166dupGATG	p.Glu56fs
<i>ROR2</i>	Dị hợp tử	Lặn	Sai nghĩa	9/9	c.1675G>A	p.Gly559Ser
<i>MAP2K1</i>	Dị hợp tử	Trội	Sai nghĩa	3/11	c.389A>G	p.Tyr130Cys
<i>FANCA</i>	Dị hợp tử	Lặn	Sai nghĩa	23/43	c.2101A>G	p.Lys701Glu
<i>AMERI</i>	Đồng hợp tử	X-linked dominant	Sai nghĩa	2/2	c.85G>A	p.Ala29Thr

Ảnh hưởng của các đột biến đến chức năng protein được đánh giá bằng các phần mềm SIFT và Polyphen_2 thông qua giá trị SIFT Score và Polyphen Score tương ứng (Bảng 4). Phần mềm SIFT đánh giá mức độ ảnh hưởng của đột biến đến chức năng protein theo thang điểm SIFT Score từ 0 đến 1. Trong đó, giá trị càng nhỏ thì đột biến càng có khả năng gây hại. SIFT đánh giá một đột biến là lành tính – T (Tolerated) nếu điểm SIFT Score > 0,05, gây bệnh – D (Deleterious) nếu SIFT

Score < 0,05. Phần mềm PolyPhen 2 có giá trị PolyPhen Score nằm trong khoảng từ 0 đến 1, giá trị này càng lớn thì đột biến càng có khả năng gây hại. Polyphen2 đánh giá đột biến là lành tính – B (Benign) nếu đột biến có PolyPhen Score < 0,452, có khả năng gây bệnh – P (Possibly Damaging) nếu PolyPhen Score nằm trong khoảng từ 0,453 đến 0,956, gây bệnh – D (probably damaging) nếu PolyPhen Score lớn hơn 0,956. Tuy nhiên 1 số đột biến dịch khung hoặc đột biến vô nghĩa chưa đánh giá được

ảnh hưởng chức năng của protein nên đã không thể cho điểm đánh giá chính xác (được đánh dấu “-” trong bảng 4).

Kết quả phân tích bằng các phần mềm tin sinh cho thấy đột biến c.389A>G dạng dị hợp tử trên gen trội *MAP2K1* được

cả hai phần mềm đánh giá là có khả năng gây bệnh (SIFT Score: 1.00; Polyphen Score: 0) có tiềm năng là nguyên nhân gây ra tình trạng bệnh ở bệnh nhân. Đột biến này đã được báo cáo trên cơ sở dữ liệu dbSNP với mã số rs121908595.

Bảng 4. Kết quả đánh giá đột biến trên các phần mềm tin sinh

Gen	Thay đổi DNA	Polyphen2	SIFT
<i>NOTCH2</i>	c.17_18delCC	-	-
<i>ABCG5</i>	c.751C>T	-	-
<i>TTN</i>	c.8069C>T	Damaging Score:0.991	Deleterious Score:0.044
<i>LFNG</i>	c.163_166dupGATG	-	-
<i>ROR2</i>	c.1675G>A	Benign score : 0.384	Tolerated Score: 0.102
<i>MAP2K1</i>	c.389A>G	Damaging Score : 1.000	Damaging Score: 0
<i>FANCA</i>	c.2101A>G	Benign score : 0.006	Tolerated Score: 0,262
<i>AMER1</i>	c.85G>A	Benign score : 0.000	Damaging Score: 0.009

Phương pháp giải trình tự WES là một ứng dụng của công nghệ NGS nhằm xác định các biến thể trên tất cả các vùng mã hóa trong hệ gen. Ưu điểm của việc giải trình tự hệ gen biểu hiện (WES) là phương pháp tiếp cận hợp lý đối với những vùng gen quan tâm hay những vùng exome chứa gen đột biến, ví dụ exome chỉ chiếm 2% trong hệ gen người nhưng biến đổi trong đó lại gây nên 85% các căn bệnh đã biết [11]a pattern of repetitive behavior and/or restricted interests in early childhood. The prevalence is higher in male children than in female children. As a complex neurodevelopmental disorder, the phenotype and severity of autism are extremely heterogeneous with differences from one patient to another. Genetics has a key role in the etiology of autism. Environmental factors are also interacting

with the genetic profile and cause abnormal changes in neuronal development, brain growth, and functional connectivity. The term of exome represents less than 1% of the human genome, but contains 85% of known disease-causing variants. Whole-exome sequencing (WES). Giải trình tự gen thể hệ mới đã được xem như là một công cụ hữu hiệu cho phát hiện các gen bệnh mới, đặc biệt giải trình tự whole exome có thể sẽ trở thành công cụ phổ biến nhất được sử dụng để xác định gen bệnh cho những năm tới. Với các tiến bộ không ngừng trong việc hạ giá thành giải trình tự và phân tích trình tự, giải trình tự thể hệ mới sẽ tiến tới khả năng trở thành một công cụ gần như thông lệ trong chẩn đoán di truyền cho bệnh nhân. Giải trình tự gen thể hệ mới đã nhanh chóng được công nhận là có thể vượt qua những hạn chế của

phương pháp giải trình tự Sanger. Trong nghiên cứu này, từ 82.556 đột biến dạng thay thế nucleotide và 11.334 đột biến thêm bớt nucleotide trên toàn bộ vùng mã hóa thông qua các bước phân tích tin sinh có thể sàng lọc về 8 đột biến tiềm năng gây bệnh cho thấy tính khả thi của phương pháp này trong việc xác định biến thể di truyền là nguyên nhân gây bệnh.

Tuy nhiên, giải trình tự toàn bộ vùng gen mã hóa còn tồn tại một số điểm hạn chế. Lượng dữ liệu tạo ra từ WES lớn do đó đòi hỏi kỹ thuật phân tích phức tạp. Ngoài ra, dữ liệu giải trình tự toàn bộ vùng mã hóa cần được xử lý bằng các phần mềm tin sinh học chuyên sâu trên hệ thống tính toán hiệu năng cao. Hơn nữa, việc phát hiện và sàng lọc biến thể bằng cách so sánh dữ liệu bệnh nhân với dữ liệu tham chiếu là chưa đủ để đánh giá mức độ ảnh hưởng của biến thể đó. Cần có những phân tích sâu hơn như giải trình tự Sanger để kiểm chứng đột biến, đánh giá ảnh hưởng của đột biến đến cấu trúc và chức năng protein hoặc khảo sát trên số lượng mẫu bệnh và đối chứng lớn để đánh giá mối tương quan của biến thể đến nguy cơ mắc bệnh nếu biến thể đó là đa hình nucleotide đơn.

IV. Kết luận

Giải trình tự toàn bộ vùng gen mã hóa đang ngày càng phát triển với nhiều ưu điểm như giá thành rẻ, thông lượng cao, chất lượng đọc trình tự tốt. Những ưu điểm đó khiến WES trở thành công nghệ được nhiều nhà khoa học trên thế giới áp dụng để phân tích toàn bộ vùng gen mã hóa hoặc toàn bộ hệ gen từ đó tìm ra các biến thể di truyền của đối tượng nghiên cứu. Đối với các hội chứng dị tật van

tim bẩm sinh, do số lượng và kích thước gen liên quan đến hội chứng này rất lớn, việc nghiên cứu từng gen riêng lẻ đòi hỏi nhiều thời gian, chi phí cao. Do đó, việc giải trình tự toàn bộ vùng gen mã hóa cho phép phân tích đồng thời nhiều gen qua đó giảm chi phí và thời gian phân tích trở thành phương pháp thay thế hiệu quả và khả thi khi tiến hành nghiên cứu di truyền ở hội chứng này.

Lời cảm ơn

Công trình nghiên cứu này thực hiện với sự tài trợ kinh phí của Bộ Khoa học và Công nghệ cho đề tài với mã số ĐTĐL. CN-45/21. Các tác giả xin gửi lời cảm ơn tới bệnh nhân và gia đình đã tham gia nghiên cứu này.

Tài liệu tham khảo:

- [1]. Shi L.-M., Tao J.-W., Qiu X.-B., et al. (2014). GATA5 loss-of-function mutations associated with congenital bicuspid aortic valve. *Int J Mol Med*, **33**(5), 1219–1226.
- [2]. Foffa I., Ait Ali L., Panesi P., et al. (2013). Sequencing of NOTCH1, GATA5, TGFBR1 and TGFBR2 genes in familial cases of bicuspid aortic valve. *BMC Med Genet*, **14**, 44.
- [3]. Dargis N., Lamontagne M., Gaudreault N., et al. (2016). Identification of Gender-Specific Genetic Variants in Patients With Bicuspid Aortic Valve. *Am J Cardiol*, **117**(3), 420–426.
- [4]. Yang Y., Muzny D.M., Reid J.G., et al. (2013). Clinical whole-exome sequencing for the diagnosis of mendelian disorders. *N Engl J Med*, **369**(16), 1502–1511.
- [5]. Lee H., Deignan J.L., Dorrani N., et al. (2014). Clinical Exome Sequencing for Genetic Identification of Rare Mendelian Disorders. *JAMA*, **312**(18), 1880–1887.

- [6]. Li H. and Durbin R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinforma Oxf Engl*, **25(14)**, 1754–1760.
- [7]. McKenna A., Hanna M., Banks E., et al. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*, **20(9)**, 1297–1303.
- [8]. Cingolani P., Platts A., Wang L.L., et al. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)*, **6(2)**, 80–92.
- [9]. Ng P.C. and Henikoff S. (2003). SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res*, **31(13)**, 3812–3814.
- [10]. Adzhubei I., Jordan D.M., and Sunyaev S.R. (2013). Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet*, **Chapter 7**, Unit7.20.
- [11]. Sener E.F., Canatan H., and Ozkul Y. (2016). Recent Advances in Autism Spectrum Disorders: Applications of Whole Exome Sequencing Technology. *Psychiatry Investig*, **13(3)**, 255–264.

Địa chỉ tác giả: Trường Đại học Mở Hà Nội
Email: nhhoang@igr.ac.vn

