

# Xác định phân nhóm biến thể số bản sao được phát hiện bằng kỹ thuật giải trình tự thế hệ mới theo tiêu chuẩn của ACMG/ClinGen

*Ngô Thị Diệu Hương, Lê Phan Tường Quỳnh, Hà Thị Minh Thi\**

*Trường Đại học Y - Dược, Đại học Huế*

## Tóm tắt

**Đặt vấn đề:** Đề tài này nhằm hai mục tiêu: 1) Khảo sát đặc điểm vùng genome bị biến đổi của các CNV được phát hiện bằng kỹ thuật giải trình tự thế hệ mới; 2) Xác định các phân nhóm CNV được phát hiện bằng kỹ thuật giải trình tự thế hệ mới theo tiêu chuẩn của ACMG/ClinGen. **Đối tượng và phương pháp nghiên cứu:** 33 biến thể số bản sao của 31 trường hợp được phát hiện bằng giải trình tự thế hệ mới đến khám và tư vấn tại Bộ môn Di truyền Y học trong thời gian từ tháng 01 năm 2023 đến tháng 12 năm 2024. **Kết quả:** Về đặc điểm genome của các CNV, mất đoạn và nhân đoạn lần lượt chiếm 51,5% và 48,5%. Phần lớn các CNV là các vi mất đoạn/nhân đoạn, kích thước dưới 5 Mb (chiếm 75,8%). Số lượng gene mã hoá protein trung bình là  $33,4 \pm 49,4$ . Các CNV phân bố trên hầu hết các nhiễm sắc thể. Về các phân nhóm CNV theo tiêu chuẩn ACMG/ClinGen, CNV gây bệnh và có khả năng gây bệnh lần lượt chiếm tỷ lệ 36,3% và 12,1%, CNV lành tính và có khả năng lành tính đều chiếm tỷ lệ 6,1%, CNV chưa chắc chắn ý nghĩa lâm sàng chiếm 39,4%. Có sự khác biệt về kích thước trung bình của CNV nhóm gây bệnh/có khả năng gây bệnh so với hai nhóm còn lại là CNV VUS và lành tính/có khả năng lành tính. **Kết luận:** Các CNV được chẩn đoán bao gồm cả mất đoạn và nhân đoạn, hầu hết có kích thước nhỏ (dưới 5 Mb). Chúng tôi đã phân loại được các CNV theo ý nghĩa lâm sàng dựa trên bộ tiêu chí của ACMG/ClinGen, từ đó cung cấp thông tin quan trọng trong tư vấn di truyền.

**Từ khóa:** Biến thể số bản sao, giải trình tự thế hệ mới, tiêu chuẩn ACMG/ClinGen, CNV gây bệnh, VUS

## Classification of copy-number variants detected by next-generation sequencing using ACMG/ClinGen standards

*Ngo Thi Dieu Huong, Le Phan Tuong Quynh, Ha Thi Minh Thi\**

*University of Medicine and Pharmacy, Hue University*

## Abstract

**Background:** This study aimed to: 1) Investigate the genomic content of CNVs detected by next-generation sequencing; 2) Determine CNV classification detected by NGS using ACMG/ClinGen standards. **Materials and methods:** 33 CNVs detected by next-generation sequencing of 31 individuals examined and consulted at the Department of Medical Genetics from January 2023 to December 2024. **Results:** Regarding the genomic characteristics of CNVs, deletions and duplications accounted for 51.5% and 48.5%, respectively. The majority of CNVs were microdeletions/microduplications, less than 5 Mb in size (75.8%). The average number of protein-coding genes was  $33.4 \pm 49.4$ . CNVs were found to be distributed on almost all chromosomes. In terms of CNV classification using ACMG/ClinGen criteria, pathogenic CNV (pCNV) and likely pathogenic CNV (lpCNV) accounted for 36.3% and 12.1%, respectively. Meanwhile, benign CNV (bCNV) and likely benign CNV (lbCNV) both accounted for 6.1%. Notably, variant of uncertain significance (VUS) accounted for 39.4%. There was a difference in size of p/lpCNVs compared to the VUS and b/lbCNV. **Conclusion:** The detected CNVs included deletion and duplication, most of which were minor in size. We classified CNVs based on clinical significance using the ACMG/ClinGen criteria, providing valuable information for genetic counseling.

**Key words:** Copy-number variant, next-generation sequencing, ACMG/ClinGen criteria, pathogenic CNV, VUS

## 1. ĐẶT VẤN ĐỀ

Bất thường nhiễm sắc thể là một trong những nguyên nhân hàng đầu gây dị tật bẩm sinh, chậm phát triển tâm thần và bất thường thai sản. Việc xác định các rối loạn nhiễm sắc thể là một phần quan trọng

trong chẩn đoán bệnh lý di truyền bao gồm cả trước sinh và sau sinh. Đến nay, nhiều kỹ thuật chẩn đoán các bệnh lý di truyền đã ra đời, trong đó, lập karyotype là kỹ thuật kinh điển trong phát hiện các bất thường nhiễm sắc thể. Mặc dù có thể khảo sát toàn bộ bộ

\* Tác giả liên hệ: Hà Thị Minh Thi, Email: htmthi@huemed-univ.edu.vn

Ngày nhận bài: 13/02/2025; Ngày đồng ý đăng: 20/07/2025; Ngày xuất bản: 25/09/2025

nhiễm sắc thể, nhưng kỹ thuật truyền thống này có nhược điểm là mất nhiều thời gian và không phát hiện được các bất thường cấu trúc kích thước dưới 5 Mb [1]. Hiện nay, với sự phát triển của các kỹ thuật sinh học phân tử, đặc biệt là kỹ thuật giải trình tự thế hệ mới (Next-generation sequencing: NGS) có nhiều ưu điểm vượt trội đã tạo nên kỷ nguyên mới trong chẩn đoán bệnh lý di truyền. Một cách tiếp cận của NGS là dựa trên việc giải trình tự biến thể số bản sao (copy-number variant: CNV) - vùng trình tự DNA có sự biến đổi (tăng thêm hoặc giảm đi) trong genome người. Bằng kỹ thuật này, các CNV trong genome có thể được phân tích nhằm chẩn đoán bất thường cả cấu trúc và số lượng nhiễm sắc thể, đặc biệt có thể phát hiện các vi mất đoạn, vi nhân đoạn với kích thước đoạn thay đổi chỉ từ 50 - 100 kb trở lên [2]. Ngoài ra, với các hệ thống giải trình tự NGS hiện nay, việc phân tích CNV có thể được thực hiện tự động hoá và có hiệu suất cao. Bên cạnh các ưu điểm, phân tích CNV bằng kỹ thuật NGS có thể cho ra các kết quả bao gồm cả biến thể gây bệnh, biến thể lành tính và cả biến thể chưa rõ ý nghĩa lâm sàng. Đây là một thách thức lớn đối với các nhà di truyền lâm sàng trong việc phiên giải kết quả và tư vấn cho bệnh nhân và gia đình.

Hiệp hội Di truyền và hệ gene y học Hoa Kỳ (The American College of Medical Genetics and Genomics: ACMG) đã phân các CNV thành 5 nhóm bao gồm lành tính, có khả năng lành tính, chưa rõ ý nghĩa lâm sàng, có khả năng gây bệnh và gây bệnh. Năm 2019, đồng thuận giữa ACMG và dự án Nguồn hệ gene lâm sàng (Clinical Genome Resource: ClinGen) đã thống nhất các tiêu chí phân nhóm CNV như sau: (1) Có chứa gene mã hoá protein hay các thành phần có chức năng quan trọng hay không; (2) Mức độ bao phủ các gene nhạy cảm liều hoặc gene lành tính đã được công nhận; (3) Số lượng gene mã hoá protein có trong CNV; (4) Đánh giá chi tiết thành phần genome của CNV dựa trên các ca lâm sàng từ các bài báo khoa học, cơ sở dữ liệu chung và/hoặc dữ liệu nội bộ của phòng thí nghiệm; (5) Đánh giá phương thức di truyền/tiền sử gia đình của bệnh nhân [3].

Việc áp dụng các tiêu chuẩn của ACMG/ClinGen để phân loại một CNV cụ thể là cần thiết để phiên giải ý nghĩa lâm sàng của kết quả CNV trong chẩn đoán bệnh lý di truyền và tư vấn cho bệnh nhân và gia đình. Trong đó, việc truy cập các cơ sở dữ liệu về genome và biến thể như UCSC, ClinGen, DECIPHER, OMIM (Online Mendelian Inheritance in Man), DGV (Database of genomic variants), ClinVar... để xác định các gene nằm trong CNV, chức năng của chúng, cũng như các kiểu hình liên quan là đặc biệt quan trọng. Vì vậy, chúng tôi tiến hành nghiên cứu đề tài này với hai mục tiêu sau:

1) Khảo sát đặc điểm vùng genome bị biến đổi của các CNV được phát hiện bằng kỹ thuật giải trình tự thế hệ mới.

2) Xác định các phân nhóm CNV được phát hiện bằng kỹ thuật giải trình tự thế hệ mới theo tiêu chuẩn của ACMG/ClinGen.

## 2. ĐỐI TƯỢNG VÀ PHƯƠNG PHÁP NGHIÊN CỨU

### 2.1. Đối tượng nghiên cứu

31 trường hợp có kết quả giải trình tự thế hệ mới phát hiện CNV cấu trúc đến khám và tư vấn tại Bộ môn Di truyền Y học trong thời gian từ tháng 01 năm 2023 đến tháng 12 năm 2024.

### 2.2. Phương pháp nghiên cứu

Nghiên cứu mô tả cắt ngang.

Các bước nghiên cứu được tiến hành như sau:

**Bước 1: Khảo sát đặc điểm vùng genome bị biến đổi của các CNV được phát hiện bằng kỹ thuật NGS**

- Xác định vùng genome chứa CNV được phát hiện bởi NGS.

+ Vị trí trên nhiễm sắc thể của CNV.

+ Kích thước CNV (Mb).

- Khảo sát sự mất cân bằng vật chất di truyền: mất đoạn, nhân đoạn.

**Bước 2: Tính điểm CNV (CNV-scoring) theo tiêu chuẩn của ACMG/ClinGen**

Theo hướng dẫn của ACMG/ClinGen năm 2019, CNV được phân nhóm dựa vào các tiêu chuẩn sau [3]:

(1) **Đánh giá sơ lược về thành phần chức năng của CNV**

- Xác định CNV có chứa các thành phần như gene mã hoá protein hoặc các yếu tố có chức năng quan trọng trong genome không.

- Sử dụng cơ sở dữ liệu của UCSC và DECIPHER.

(2) **Đánh giá mức độ bao phủ các gene nhạy cảm liều hoặc gene lành tính đã được công nhận**

- Xác định CNV có chứa hoàn toàn hoặc một phần các gene/vùng genome mà đã được xác nhận về khả năng gây bệnh khi mất đoạn hoặc nhân đoạn:

+ Gene có tính thiếu hụt đơn bội (haploinsufficient: HI)

+ Gene có khả năng gây bệnh khi có ba bản sao (triplosensitive: TS)

+ Vùng genome lành tính có trong CNV đã được công nhận.

- Sử dụng cơ sở dữ liệu của UCSC, ClinGen và DECIPHER, gnomAD.

(3) **Đánh giá số lượng gene trong CNV**

- Xác định số lượng gene mã hoá protein bao gồm những gene hiện diện một phần hoặc toàn bộ trong CNV.

- Tra cứu trên dữ liệu của DECIPHER.

(4) **Đánh giá chi tiết thành phần genome của CNV dựa trên các ca lâm sàng từ các bài báo khoa học**

đã công bố, các cơ sở dữ liệu chung và/hoặc nội bộ trong phòng thí nghiệm

- Tra cứu các công bố khoa học, các cơ sở dữ liệu như OMIM, Pubmed, ClinVar.

- Sử dụng cơ sở dữ liệu của quần thể trên DGV (Database of Genomic Variants Genomic Variants) và gnomAD structural variant.

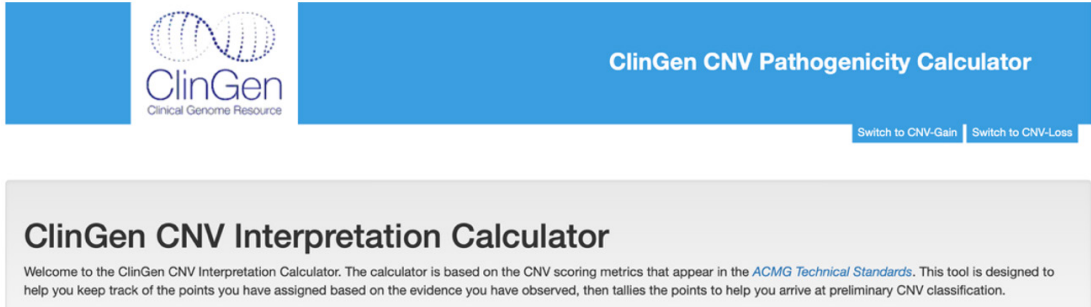
(5) Đánh giá phương thức di truyền/tiền sử gia đình đối với bệnh nhân đang được nghiên cứu:

- Khai thác yếu tố tiền sử, thăm khám lâm sàng, lập phả hệ;

- Khai thác kết quả phân tích CNV của các thành viên khác trong gia đình nếu có.

**Bước 3: Phân nhóm CNV theo tiêu chuẩn của ACMG/ClinGen**

Sử dụng công cụ trực tuyến ClinGen CNV Interpretation Calculator (<https://cnvcalc.clinicalgenome.org>) (Hình 1). Trong đó, có hai thang điểm CNV-Gain và CNV-Loss được sử dụng lần lượt cho trường hợp tăng/giảm số bản sao. Sử dụng điểm cắt để phân nhóm CNV theo tiêu chuẩn của ACMG/ClinGen (Bảng 1).



**Bảng 1.** Điểm cắt phân nhóm CNV theo tiêu chuẩn ACMG/ClinGen

| Điểm cắt         | Phân nhóm CNV                                             |
|------------------|-----------------------------------------------------------|
| $\leq -0,99$     | Lành tính (benign: bCNV)                                  |
| - 0,9 đến - 0,98 | Có khả năng lành tính (likely benign: lbCNV)              |
| - 0,89 đến 0,89  | Chưa rõ ý nghĩa (Variants of uncertain significance: VUS) |
| 0,9 đến 0,98     | Có khả năng gây bệnh (likely pathogenic: lpCNV)           |
| $\geq 0,99$      | Gây bệnh (pathogenic: pCNV)                               |

**3. KẾT QUẢ**

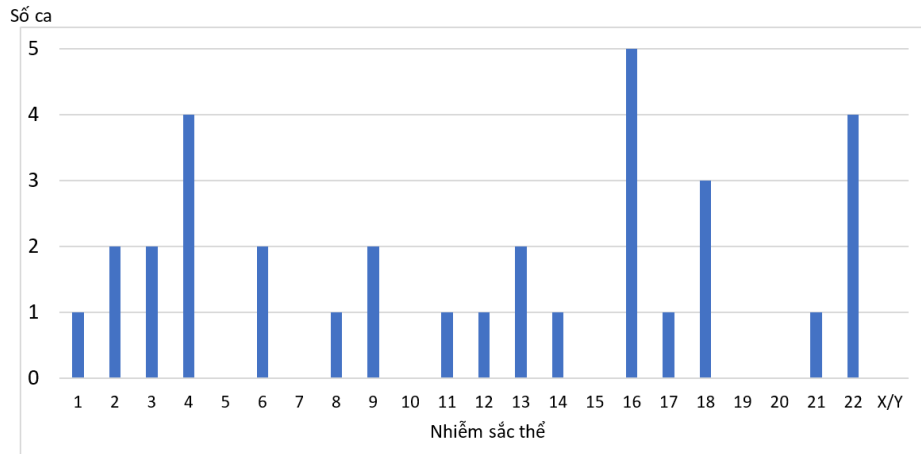
Kỹ thuật phân tích CNV dựa trên NGS (CNV-seq) đã phát hiện 33 CNV cấu trúc của 31 trường hợp tham gia chẩn đoán trước sinh, trong đó có hai trường hợp mang hai CNV.

**3.1. Đặc điểm vùng genome bị biến đổi của các CNV được phát hiện bằng kỹ thuật giải trình tự thế hệ mới**

**Bảng 2.** Đặc điểm vùng genome bị biến đổi của CNV

| Đặc điểm                           | Số lượng               | Tỷ lệ (%) |
|------------------------------------|------------------------|-----------|
| Sự biến đổi của vật chất di truyền |                        |           |
| Mất đoạn (deletion)                | 17                     | 51,5      |
| Nhân đoạn (duplication)            | 16                     | 48,5      |
| Kích thước (Mb)                    |                        |           |
| < 5 Mb                             | 25                     | 75,8      |
| ≥ 5 Mb                             | 8                      | 24,2      |
| X ± SD (Min - Max)                 | 4,5 ± 7,2 (0,4 - 36,8) |           |
| Số lượng gene mã hoá protein       |                        |           |
| Chung (X ± SD; Min - Max)          | 33,4 ± 49,4 (0 - 214)  |           |
| Mất đoạn                           | 32,2 ± 25,5            |           |
| Nhân đoạn                          | 34,6 ± 67,1            |           |
| Tổng                               | 33                     | 100,0     |

Các CNV là mất đoạn và nhân đoạn tương đương nhau, lần lượt chiếm 51,5% và 48,5%. Đa số các CNV có kích thước dưới 5 Mb (vi mất đoạn/nhân đoạn), chiếm 75,8%. Số lượng gene mã hoá protein trung bình là 33,4  $\pm$  49,4.



Các CNV phân bố trên hầu hết các nhiễm sắc thể, trong đó, các NST 4, 16, 22 có bốn đến năm trường hợp.

### 3.2. Các phân nhóm CNV được phát hiện bằng kỹ thuật giải trình tự thể hệ mới theo tiêu chuẩn của ACMG/ClinGen

**Bảng 3.** Đặc điểm của CNV theo tiêu chuẩn của ACMG/ClinGen

| Tiêu chí đánh giá CNV theo ACMG/ClinGen                                                              | Số lượng (N = 33) | Tỷ lệ (%) |
|------------------------------------------------------------------------------------------------------|-------------------|-----------|
| <b>Tiêu chí 1:</b> CNV có chứa gene mã hoá protein hoặc yếu tố có chức năng quan trọng               |                   |           |
| Có                                                                                                   | 31                | 93,9      |
| Không                                                                                                | 2                 | 6,1       |
| <b>Tiêu chí 2:</b> Mức độ bao phủ các gene nhạy cảm liều hoặc gene lành tính đã được công nhận       |                   |           |
| Chỉ bao phủ vùng genome lành tính                                                                    | 2                 | 6,1       |
| Bao phủ hoàn toàn ít nhất một gene HI                                                                | 11                | 33,3      |
| Bao phủ hoàn toàn ít nhất một gene TS (đối với nhân đoạn)                                            | 0                 | 0,0       |
| Chứa một phần gene HI hoặc TS                                                                        | 0                 | 0,0       |
| Không chứa các gene/vùng genome ở trên                                                               | 20                | 60,6      |
| <b>Tiêu chí 3:</b> Số lượng gene trong CNV                                                           |                   |           |
| 0 - 24 (đối với mất đoạn); 0 - 34 (đối với nhân đoạn)                                                | 20                | 60,6      |
| 25 - 34 (đối với mất đoạn); 35 - 49 (đối với nhân đoạn)                                              | 0                 | 0,0       |
| ≥ 35 (đối với mất đoạn); ≤ 50 (đối với nhân đoạn)                                                    | 13                | 39,4      |
| <b>Tiêu chí 4:</b> Ca lâm sàng liên quan từ các bài báo khoa học đã công bố, các cơ sở dữ liệu chung |                   |           |
| Có                                                                                                   | 19                | 57,6      |
| Không                                                                                                | 14                | 42,4      |
| <b>Tiêu chí 5:</b> Phương thức di truyền/tiền sử gia đình                                            |                   |           |
| CNV có nguồn gốc từ mẹ                                                                               | 2*                | 6,1       |
| Không xác định được nguồn gốc CNV                                                                    | 31                | 93,9      |

*Chú thích: \*Cả hai người mẹ đều không có kiểu hình bất thường. Kết quả CNV-seq ở hai người mẹ này lần lượt là dup(13)(q31.2-q31.3) và dup(21)(q21.1), tương ứng với các thai nhi P3 và P4 trong Bảng 4.*

Phần lớn các CNV cấu trúc được đánh giá có chứa các thành phần quan trọng, chiếm 93,9%. Các CNV có chứa các vùng lành tính, gene HI và gene TS lần lượt là 6,1%, 33,3% và 0%. Có 39,4% CNV có chứa rất nhiều gene (≥ 35 gene đối với mất đoạn; ≥ 50 gene đối với nhân đoạn). Có 57,6% trường hợp CNV có các ca lâm sàng liên quan đã được báo cáo trong các bài báo khoa học và/hoặc các cơ sở dữ liệu chung. Chỉ có 6,1% được xác định thừa hưởng CNV từ người mẹ không có kiểu hình bất thường.

**Bảng 4.** Phân nhóm CNV theo tiêu chuẩn của ACMG/ClinGen

| Mã số        | CNV                    | Kích thước (Mb) | Điểm | Phân nhóm             | Tỷ lệ %    |
|--------------|------------------------|-----------------|------|-----------------------|------------|
| P1           | del(4)(q13.2)          | 0,6             | -1   | Lành tính             | 2 (6,1%)   |
| P2           | dup(4)(q35.2)          | 1               | -1   |                       |            |
| P3*          | dup(13)(q31.2-q31.3)   | 2,3             | -0,9 | Có khả năng lành tính | 2 (6,1%)   |
| P4*          | dup(21)(q21.1)         | 2,4             | -0,9 |                       |            |
| P5           | dup(2)(q32.1-q32.2)    | 0,4             | -0,6 | VUS                   | 13 (39,4%) |
| P6           | del(16)(p12.2)         | 0,4             | 0    |                       |            |
| P7           | dup(16)(q23.1)         | 0,6             | 0    |                       |            |
| P8           | dup(16)(q23.1)         | 0,7             | 0    |                       |            |
| P9           | dup(9)(p24.3)          | 0,7             | 0    |                       |            |
| P10          | dup(8)(q11.21-q11.23)  | 0,7             | 0    |                       |            |
| P11          | del(2)(q21.1)          | 0,8             | 0    |                       |            |
| P12          | dup(6)(q14.2-q14.3)    | 0,9             | 0    |                       |            |
| P13          | dup(18)(q12.3)         | 1,0             | 0    |                       |            |
| P14          | dup(4)(q35.2)          | 1,1             | 0    |                       |            |
| P15          | dup(18)(p11.32-p11.31) | 1,2             | 0    |                       |            |
| P16          | del(11)(p14.3)         | 1,4             | 0    |                       |            |
| P17          | del(13)(q33.2-q33.3)   | 1,4             | 0,15 |                       |            |
| P18          | del(14)(q32.33)        | 1,3             | 0,9  |                       |            |
| P19          | dup(12)(q24.31)        | 4,2             | 0,9  | Có khả năng gây bệnh  | 4 (12,1%)  |
| P20-C1**     | dup(16)(p13.3-p13.13)  | 11,4            | 0,9  |                       |            |
| P21-C1***    | dup(9)(p24.3-p13.2)    | 36,8            | 0,9  | Gây bệnh              | 12 (36,3%) |
| P22          | del(16)(p13.11)        | 1,2             | 1    |                       |            |
| P23          | del(17)(q12)           | 1,4             | 1    |                       |            |
| P24,25,26,27 | del(22)(q11.21)        | 2,4 - 2,5       | 1    |                       |            |
| P21-C2***    | del(1)(q44)            | 5               | 1    |                       |            |
| P28          | del(3)(p26.3-p25.3)    | 9,9             | 1    |                       |            |
| P29          | del(6)(q25.3-q27)      | 11              | 1    |                       |            |
| P20-C2**     | del(3)(p26.3-p25.3)    | 8,7             | > 1  |                       |            |
| P30          | dup(18)(p11.31-p11.21) | 14,1            | > 1  |                       |            |
| P31          | del(4)(q32.2-q34.3)    | 15,8            | > 1  |                       |            |

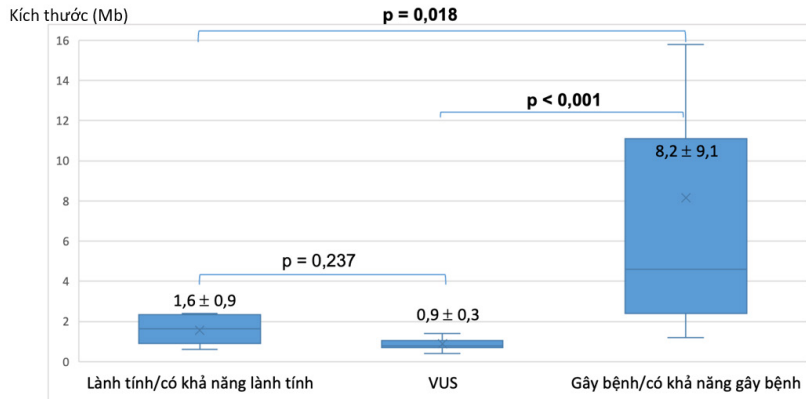
*Chú thích:*

\*Kiểu hình bình thường ở hai bà mẹ của cả hai trường hợp P3 và P4 đã giúp các CNV dup(13)(q31.2-q31.3) và dup(21)(q21.1) được phân nhóm là có khả năng lành tính thay vì phải phân nhóm là VUS.

\*\* P20-C1 và P20-C2 là hai CNV của bệnh nhân P20.

\*\*\* P21-C1 và P21-C2 là hai CNV của bệnh nhân P21.

Tỷ lệ các bCNV và lb CNV đều là 6,1%. Tỷ lệ pCNV và lp CNV lần lượt là 36,3% và 12,1%. Có hai bệnh nhân mang hai CNV phối hợp, gồm một lpCNV và một pCNV. Có 39,4% trường hợp không chắc chắn ý nghĩa lâm sàng (VUS).



Kích thước trung bình của CNV nhóm gây bệnh/có khả năng gây bệnh cao hơn có ý nghĩa so với hai nhóm còn lại là CNV VUS và lành tính/có khả năng lành tính ( $p < 0,05$ ).

#### 4. BÀN LUẬN

##### 4.1. Đặc điểm vùng genome bị biến đổi của các CNV được phát hiện bằng kỹ thuật giải trình tự thế hệ mới

Phân tích biến thể số bản sao (CNV) trong genome người nhằm phát hiện các vùng trình tự DNA có sự biến đổi tăng thêm (duplication, triplication) hoặc giảm đi (deletion) được khuyến cáo là phương pháp tiếp cận đầu tiên để đánh giá các dị tật bẩm sinh cũng như chẩn đoán trước sinh những thai nhi có bất thường hình thái được phát hiện bằng siêu âm thai [3]. Sự ra đời và ứng dụng rộng rãi trong lâm sàng của các kỹ thuật sinh học phân tử như phân tích CNV bằng microarray (CMA) hoặc giải trình tự thế hệ mới (NGS) đã giúp mở rộng phát hiện các CNV và nâng cao hiểu biết về ảnh hưởng của các biến thể đối với sức khỏe con người và trong cộng đồng nói chung.

Trong nghiên cứu này, các CNV là các mất đoạn/nhân đoạn được phát hiện bằng kỹ thuật phân tích CNV dựa trên NGS hay còn gọi là giải trình tự biến thể số bản sao (CNV-seq). Kết quả từ Bảng 2 cho thấy, trong số 33 CNV được phát hiện, tỷ lệ mất đoạn và nhân đoạn tương đương nhau, lần lượt là 51,5% và 48,5%. Một trong những ưu điểm của CNV-seq so với karyotype thường quy là khả năng phát hiện vi mất đoạn/nhân đoạn với kích thước dưới 5 Mb. Điều này phù hợp với kích thước trung bình của các CNV trong nghiên cứu này là  $4,5 \pm 7,2$  (Mb), trong đó có những CNV kích thước chỉ 0,4 Mb. Phần lớn các mất đoạn, nhân đoạn này có kích thước dưới 5 Mb, chiếm 75,8%. Đây là các vi mất đoạn/nhân đoạn không thể phát hiện bằng kỹ thuật lập karyotype thường quy. Các nghiên cứu tương tự cũng đã chứng minh được khả năng phát hiện các vi mất đoạn/nhân đoạn của kỹ thuật CNV-seq như nghiên cứu của Nguyễn Thị Hào (2022) cho thấy CNV-seq đã phát hiện thêm 4,5%

trường hợp bất thường nhiễm sắc thể so với kỹ thuật lập karyotype truyền thống trong chẩn đoán trước sinh [4]. Tuy nhiên, hậu quả lâm sàng không chỉ phụ thuộc vào kích thước của CNV. Để xác định ý nghĩa lâm sàng của các biến thể này, theo khuyến cáo của ACMG cần đánh giá các gene bị thay đổi số bản sao có trong CNV. Kết quả ở Bảng 2 còn cho thấy số lượng gene mã hoá protein trung bình trong vùng genome bị mất đoạn/nhân đoạn là  $33,4 \pm 49,4$  trong đó thấp nhất là CNV không chứa gene nào và nhiều nhất là có đến 214 gene mã hoá protein. Sự khác biệt số lượng gene của các CNV cũng phù hợp với kích thước CNV được phát hiện cũng thay đổi trong phạm vi khá lớn (từ 0,4 đến 36,8 Mb). Số lượng gene không có sự khác biệt giữa hai nhóm mất đoạn và nhân đoạn.

Khi khảo sát vị trí của các CNV, tác giả Zarrei (2015) đã lập bản đồ CNV của genome người và cho thấy sự phân bố của CNV rải rác và không đồng đều trong các vùng genome và trên các nhiễm sắc thể (NST). Đối với nhân đoạn, các NST 22, Y có tần suất cao nhất, tiếp theo là NST 16, 9, 15; trong khi đó, đối với mất đoạn, sự phân bố của CNV tập trung nhiều ở các NST 19, 22 và NST Y [5]. Mặc dù nghiên cứu của chúng tôi chỉ đánh giá một số ít là trong 33 CNV, tuy nhiên chúng tôi nhận thấy các CNV phân bố trên hầu hết các nhiễm sắc thể, đặc biệt các NST 4, 16, 22 có từ bốn đến năm trường hợp (Biểu đồ 1). Sự xuất hiện nhiều trường hợp trên các NST này phù hợp với các vùng genome có tần suất tái tổ hợp cao như các vùng 16p11.2 và 22q11.2 theo nhận định của Rein (2020) và Sue (2021) [6, 7].

##### 4.2. Các phân nhóm CNV được phát hiện bằng kỹ thuật giải trình tự thế hệ mới theo tiêu chuẩn của ACMG/ClinGen

Năm 2019, để hỗ trợ các phòng xét nghiệm trong việc phân loại và báo cáo CNV, ACMG cùng với dự án

ClinGen đã đưa ra bộ tiêu chí nhằm phân loại ý nghĩa CNV. Bộ tiêu chí này giới thiệu cách chấm điểm dựa trên bằng chứng và khuyến khích sử dụng rộng rãi hệ thống phân loại CNV theo năm cấp trong thực hành lâm sàng.

Bảng 3 cho thấy kết quả đánh giá CNV theo năm tiêu chí phân loại của ACMG/ClinGen. Trong đó, tiêu chí 1 khảo sát sơ lược về thành phần của CNV, kết quả cho thấy phần lớn các CNV được đánh giá đều có chứa các thành phần quan trọng như gene mã hoá protein hoặc các yếu tố chức năng quan trọng, chiếm 93,9%. Tuỳ thuộc vào chức năng của các protein được mã hoá mà các CNV có thể có các hậu quả lâm sàng khác nhau.

Do các CNV phản ánh sự biến đổi vật chất di truyền như mất đoạn (deletion) hoặc nhân đoạn (duplication), đồng thời tuỳ thuộc gene HI/TS có trong CNV mà hậu quả lâm sàng khác nhau, vì vậy, tiêu chí 2 chú trọng vào mức độ bao phủ với các gene nhạy cảm liều. Trong nghiên cứu này có đến 33,3% trường hợp bao phủ hoàn toàn ít nhất một gene có tính thiếu hụt đơn bội (HI), đây là các gene mà nếu chỉ còn một bản sao duy nhất còn hoạt động thì không đủ để đảm bảo chức năng bình thường. Vì vậy, những mất đoạn chứa những gene này sẽ được xếp vào phân nhóm gây bệnh trên lâm sàng. Đáng chú ý, có đến 60,6% CNV không chứa các gene/vùng genome nhạy cảm liều như HI, TS cũng như vùng genome lành tính đã được công nhận, vì vậy ACMG cũng đã đưa ra các tiêu chí khác để có thể đánh giá hậu quả lâm sàng ở những trường hợp này bao gồm số lượng gene, bằng chứng lâm sàng và phương thức di truyền/tiền sử gia đình (lần lượt là tiêu chí 3, 4, 5).

Về số lượng các gene có trong CNV, ACMG/ClinGen chia thành 3 mức độ như đã mô tả ở Bảng 3. Đặc biệt, có đến 39,4% CNV có chứa rất nhiều gene ( $\geq 35$  gene đối với mất đoạn;  $\geq 50$  gene đối với nhân đoạn), điều này có thể liên quan đến hậu quả lâm sàng ở mức độ trầm trọng. Ngoài ra, khi tra cứu bằng chứng lâm sàng liên quan, chỉ khoảng 57,6% trường hợp có các ca lâm sàng liên quan đã được báo cáo trong các bài báo khoa học và/hoặc các cơ sở dữ liệu chung. Do tính chất đặc thù của các bệnh lý di truyền - phần lớn là các bệnh hiếm nên việc thu thập dữ liệu các ca lâm sàng liên quan đến từng biến thể CNV trở nên đặc biệt khó khăn. Điều này tạo ra thách thức lớn trong quá trình tìm kiếm bằng chứng đáng tin cậy để phân loại CNV một cách chính xác. Vì vậy, một yếu tố quan trọng trong cải thiện độ tin cậy của phân tích ý nghĩa CNV là các phòng xét nghiệm di truyền cần chủ động và thường xuyên cập nhật, báo cáo các CNV được phát hiện kèm theo đặc điểm lâm sàng lên các cơ sở dữ liệu chung. Đây sẽ là nguồn tài liệu quý

giá giúp các nhà khoa học và bác sĩ lâm sàng có thêm căn cứ để tham khảo và đưa ra đánh giá chính xác trong phân loại biến thể CNV [3, 8].

Ngoài ra, trong phân loại CNV, việc xác định phương thức di truyền/tiền sử gia đình đóng vai trò quan trọng. Điều này có thể có thể cung cấp thêm bằng chứng hỗ trợ hoặc bác bỏ khả năng gây bệnh của biến thể được phát hiện, vì vậy, thực hiện xét nghiệm trio (xét nghiệm cho bố, mẹ và con/thai nhi) được khuyến cáo thực hiện nhằm tăng thêm bằng chứng cho hậu quả lâm sàng của các biến thể được phát hiện, đặc biệt trong chẩn đoán trước sinh [9]. Thực vậy, nghiên cứu của chúng tôi ghi nhận có hai người mẹ kiểu hình bình thường, lần lượt mang CNV dup(13)(q31.2-q31.3) và dup(21)(q21.1), cả hai CNV này đều được di truyền cho các thai nhi P3 và P4 trong Bảng 4. Dựa vào kiểu hình bình thường của hai bà mẹ này, chúng tôi đã phân nhóm cả hai CNV trên là có khả năng lành tính (-0,9 điểm), thay vì phân nhóm VUS (-0,6 điểm). Như vậy, có thể thấy việc xét nghiệm trio đóng vai trò quan trọng trong cung cấp bằng chứng để phân loại biến thể được phát hiện trước sinh. Tuy nhiên, trên thực tế, việc triển khai xét nghiệm này còn gặp nhiều khó khăn, đặc biệt tại các nước đang phát triển như Việt Nam khi mà nguồn lực kinh tế còn hạn chế và sự hiểu biết cộng đồng chưa thực sự cao. Trong nghiên cứu của chúng tôi, chỉ có 2 trường hợp (chiếm 6,1%) đã đề cập ở trên được phát hiện có nguồn gốc CNV từ người mẹ không có kiểu hình bất thường, tuy nhiên, hai trường hợp này không thực hiện xét nghiệm trio ban đầu mà chỉ tình cờ phát hiện qua thông qua xét nghiệm sàng lọc trước sinh không xâm lấn. Như vậy, việc khuyến cáo cho nhân viên y tế và truyền thông về vai trò của thực hiện xét nghiệm trio trong các bệnh lý di truyền hết sức quan trọng. Đặc biệt, các nhà di truyền y học và bác sĩ sản khoa cần phối hợp chặt chẽ, từng bước đưa xét nghiệm trio vào trong chẩn đoán trước sinh.

Theo ACMG/ClinGen, các CNV được phát hiện bằng các kỹ thuật dựa trên NGS hoặc microarray được chia thành 5 nhóm bao gồm pCNV (CNV gây bệnh), lpCNV (có khả năng gây bệnh), VUS CNV (không chắc chắn ý nghĩa lâm sàng), lbCNV (có khả năng lành tính), bcCNV (lành tính). Trong nghiên cứu này, có 12,2% trường hợp là không có hậu quả lâm sàng bao gồm bcCNV (6,1%) và lbCNV (6,1%). Đây là những biến thể có bằng chứng mạnh ("strong evidence") là không liên quan đến bệnh lý di truyền và được gặp với tần suất khá phổ biến trong quần thể. Vì vậy, ACMG đã khuyến cáo các phòng xét nghiệm có thể lựa chọn không thông báo những biến thể này. Mặt khác, các CNV gây bệnh và có khả năng gây bệnh đều là những biến thể có bằng chứng rất

mạnh (“very strong evidence”) gây nên hậu quả lâm sàng. Trong đó, những biến thể lpCNV mặc dù có số điểm tính theo tiêu chuẩn của ACMG thấp hơn so với pCNV, tuy nhiên trên thực hành lâm sàng, cả hai biến thể pCNV và lpCNV đều được xem là có nguyên nhân của bệnh lý di truyền, vì vậy thái độ quản lý và tư vấn di truyền trên phương diện lâm sàng của hai nhóm này là tương tự nhau [10]. Kết quả từ Bảng 4 cho thấy, tỷ lệ của nhóm có hậu quả lâm sàng này là 48,4%, bao gồm 36,3% pCNV và 12,1% lpCNV.

Theo khuyến cáo của ACMG/ClinGen, khi phân loại đối với trường hợp có nhiều CNV khác nhau, cần đánh giá đồng thời độc lập từng CNV được phát hiện và thái độ quản lý lâm sàng sẽ dựa trên phân loại hậu quả nặng nề nhất. Trong nghiên cứu của chúng tôi, có hai trường hợp có sự phối hợp hai CNV khác nhau, đó là trường hợp P20 và P21, cả hai đều mang một lpCNV và một pCNV. Đây là những trường hợp vừa mang monosomy một phần của NST này và trisomy một phần của NST kia, thường là con cái của các cặp vợ chồng mà một trong hai người mang chuyển đoạn tương hỗ giữa hai NST liên quan [11]. Chuyển đoạn tương hỗ là bất thường cấu trúc NST dạng cân bằng vì vậy người mang thường có kiểu hình bình thường, trên thực tế lâm sàng, họ chỉ được phát hiện khi đã có những lần sinh con bất thường và/hoặc sảy thai liên tiếp. Hai trường hợp của chúng tôi là đều là những cặp vợ chồng mang thai lần đầu, vì vậy cần lập karyotype của hai vợ chồng để xác định người mang chuyển đoạn tương hỗ và tư vấn di truyền. Như vậy, trong bốn trường hợp lpCNV, chúng tôi có hai trường hợp mang một CNV đơn độc. Thông tin lâm sàng của các biến thể từ hai trường hợp này trong nghiên cứu của chúng tôi là sẽ những bằng chứng lâm sàng góp phần bổ sung cho cơ sở dữ liệu di truyền để các nhà chuyên môn có thể tham khảo.

Bên cạnh đó, một tỷ lệ khá lớn (39,4%) CNV được phân nhóm VUS, đây là nhóm không chắc chắn về ý nghĩa lâm sàng. Phần lớn các biến thể VUS này là mất đoạn/nhân đoạn có kích thước nhỏ, dưới 5 Mb, không hoặc chứa ít gene mã hoá protein và chưa có nhiều công bố liên quan. Điều này đặt ra thách thức đối với các nhà lâm sàng và tư vấn di truyền trong phiên giải các biến thể này. Vì vậy, các bất thường được phát hiện trên siêu âm thai và/hoặc đặc điểm của trẻ được sinh ra là những dữ liệu quan trọng giúp cập nhật và tái phân nhóm ý nghĩa lâm sàng của CNV VUS này.

Khi đánh giá về kích thước của CNV theo phân nhóm, Biểu đồ 2 cho thấy kích thước của CNV nhóm p/lp CNV cao hơn có ý nghĩa thống kê so với nhóm b/lb CNV và VUS CNV, lần lượt là  $8,2 \pm 9,1$  Mb so với  $1,6 \pm 0,9$  Mb và  $0,9 \pm 0,3$  Mb. Kết quả này tương tự

như nghiên cứu của Tiago (2019) khi cho thấy nhóm CNV gây bệnh có kích thước trung bình 7,77 Mb, cao hơn so với nhóm CNV lành tính (0,483 Mb) và CNV VUS (0,666 Mb) [12]. Điều này phù hợp với việc kích thước CNV càng lớn có thể chứa càng nhiều gene. Tuy nhiên, chúng tôi chưa tìm thấy sự khác biệt có ý nghĩa về kích thước giữa hai nhóm b/lb CNV và VUS CNV, hầu như các CNV thuộc hai nhóm này đều có kích thước khá nhỏ, dưới 5 Mb. Vì vậy, một lần nữa cho thấy việc đánh giá chi tiết về thành phần cấu trúc gene trong CNV theo các tiêu chuẩn của ACMG là rất quan trọng để phân loại chính xác biến thể được phát hiện.

Tóm lại, nghiên cứu của chúng tôi đã ứng dụng bộ tiêu chuẩn có tính định lượng và dựa trên bằng chứng của ACMG/ClinGen trong phân loại ý nghĩa lâm sàng của các biến thể CNV. Đây là hệ thống phân loại hiệu quả trong diễn giải kết quả phát hiện bất thường cấu trúc NST, từ đó cung cấp thông tin quan trọng trong tư vấn di truyền. Tuy nhiên, nghiên cứu này vẫn còn hạn chế là cỡ mẫu tương đối nhỏ (33 CNV của 31 trường hợp thai nhi được chẩn đoán trước sinh).

## 5. KẾT LUẬN

Qua nghiên cứu trên 33 biến thể số bản sao của 31 trường hợp được phát hiện bằng kỹ thuật giải trình tự thế hệ mới, chúng tôi có một số kết luận như sau:

### 5.1. Về đặc điểm genome của các CNV:

- Mất đoạn và nhân đoạn lần lượt chiếm 51,5% và 48,5%.
- Phần lớn các CNV là các vi mất đoạn/nhân đoạn, kích thước dưới 5 Mb (chiếm 75,8%).
- Số lượng gene mã hoá protein trung bình là  $33,4 \pm 49,4$ . Các CNV phân bố trên hầu hết các nhiễm sắc thể.

### 5.2. Về các phân nhóm CNV theo tiêu chuẩn ACMG/ClinGen:

- Các CNV gây bệnh và có khả năng gây bệnh lần lượt chiếm tỷ lệ 36,3% và 12,1%. Các CNV lành tính và có khả năng lành tính đều chiếm tỷ lệ 6,1%. Các CNV chưa chắc chắn ý nghĩa lâm sàng chiếm 39,4%.
- Có sự khác biệt về kích thước trung bình của CNV nhóm gây bệnh/có khả năng gây bệnh so với hai nhóm còn lại là CNV VUS và lành tính/có khả năng lành tính.

### Lời cảm ơn:

Đây là kết quả của đề tài nghiên cứu khoa học cấp Trường Đại học Y - Dược, Đại học Huế, mã số đề tài 132/23.

## TÀI LIỆU THAM KHẢO

1. Zhang H, Xu Z, Chen Q, Chen H, Ding X, Liu L, et al. Comparison of the combined use of CNV-seq and karyotyping or QF-PCR in prenatal diagnosis: a retrospective study. *Sci Rep.* 2023 Dec 1;13(1).
2. Lan L, She L, Zhang B, He Y, Zheng Z. Prenatal diagnosis of 913 fetuses samples using copy number variation sequencing. *Journal of Gene Medicine.* 2021 May 1;23(5).
3. Riggs ER, Andersen EF, Cherry AM, Kantarci S, Kearney H, Patel A, et al. Technical standards for the interpretation and reporting of constitutional copy-number variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics (ACMG) and the Clinical Genome Resource (ClinGen). 2019;p.245–257.
4. Nguyễn Thị Hào, Tăng Xuân Hải, Nguyễn Xuân Chung, Đinh Thị Quỳnh, Ngô Văn Cảnh, Hoàng Minh Trường, et al. Chẩn đoán trước sinh các bất thường nhiễm sắc thể bằng kỹ thuật CNV-seq tại Bệnh viện Sản Nhi Nghệ An. *Tạp chí Phụ sản.* 2022 Oct 7;20(3):26–31.
5. Zarrei M, MacDonald JR, Merico D, Scherer SW. A copy number variation map of the human genome. Vol. 16, *Nature Reviews Genetics.* Nature Publishing Group. 2015;p. 172–83.
6. Rein B, Yan Z. 16p11.2 Copy Number Variations and Neurodevelopmental Disorders. Vol. 43. *Trends in Neurosciences.* Elsevier Ltd; 2020;p. 886–901.
7. Xue J, Shen R, Xie M, Liu Y, Zhang Y, Gong L, et al. 22q11.2 recurrent copy number variation-related syndrome: a retrospective analysis of our own microarray cohort and a systematic clinical overview of ClinGen curation. *Transl Pediatr.* 2021 Dec 1;10(12):3273–81.
8. Requena F, Abdallah HH, García A, Nitschké P, Romana S, Malan V, et al. CNVxplorer: A web tool to assist clinical interpretation of CNVs in rare disease patients. *Nucleic Acids Res.* 2021 Jul 2;49(1):W93–103.
9. Gabriel H, Korin D, Ritthaler M, Schulte B, Battke F, von Kaisenberg C, et al. Trio exome sequencing is highly relevant in prenatal diagnostics. *Prenat Diagn.* 2022 Jun 1;42(7):845–51.
10. East K, Chung W, Foreman K, Gilmore M, Gornick M, Hindorff L, et al. Guide to Interpreting Genomic Reports: A Genomics Toolkit. *Clinical Sequencing Exploratory Research.* 2017;p.1–36.
11. Nussbaum R. L., McInnes R. R., Willard H. F. Thompson & Thompson *Genetics in Medicine* 8th ed. Vol. 8th. Elsevier. 2016.
12. Chaves TF, Baretto N, Oliveira LF de, Ocampos M, Barbato IT, Anselmi M, et al. Copy Number Variations in a Cohort of 420 Individuals with Neurodevelopmental Disorders From the South of Brazil. *Sci Rep.* 2019 Dec 1;9(1).