Review Article

J Genet Med 2023;20(2):46-51
https://doi.org/10.5734/JGM.2023.20.2.46
ISSN 1226-1769 (Print) 2383-8442 (Online)

JGM Journal of Genetic Medicine

Check for updates

Exonic copy number variations in rare genetic disorders

Man Jin Kim*

Department of Genomic Medicine, Seoul National University Hospital, Seoul, Korea

Exonic copy number variation (CNV), involving deletions and duplications at the gene's exon level, presents challenges in detection due to their variable impact on gene function. The study delves into the complexities of identifying large CNVs and investigates less familiar but recurrent exonic CNVs, notably enriched in East Asian populations. Examining specific cases like *DRC1, STX16, LAMA2*, and *CFTR* highlights the clinical implications and prevalence of exonic CNVs in diverse populations. The review addresses diagnostic challenges, particularly for single exon alterations, advocating for a strategic, multi-method approach. Diagnostic methods, including multiplex ligation-dependent probe amplification, droplet digital PCR, and CNV screening using next-generation sequencing data, are discussed, with whole genome sequencing emerging as a powerful tool. The study underscores the crucial role of ethnic considerations in understanding specific CNV prevalence and ongoing efforts to unravel subtle variations. The ultimate goal is to advance rare disease diagnosis and treatment through ethnically-specific therapeutic interventions.

Key words: DNA copy number variations, Genetic diseases, inborn, High-throughput nucleotide sequencing, Pathology, molecular.

Introduction

Copy number variations (CNVs) refer to segments of DNA where variations in copy numbers are observed when comparing different genomes [1]. CNVs can manifest as gains (insertions or duplications) or losses (deletions or null genotypes) compared to a reference genome. Exonic CNVs specifically involve deletions and duplications at the exon level within specific genes, affecting genes differently based on the degree of overlap [2]. Large CNVs, especially those involving multiple genes, are relatively easily detected, while their association with syndromes and establishing genotype-phenotype correlations pose challenges. This review aims to explore the relationship between exon level CNVs in well-characterized genes and rare genetic diseases. Additionally, the study investigates less familiar but recurrently reported CNVs, focusing on the enrichment of CNVs in East Asians.

The subsequent sections delve into specific genes with recurrent exonic CNVs, such as *DRC1, STX16, LAMA2*, and *CFTR*. Each case highlights the clinical implications and prevalence of these CNVs in different populations. The challenges in diagnosing exonic CNVs, particularly single exon alterations, are discussed, emphasizing the importance of a strategic, multi-method approach for accurate diagnosis. Diagnostic methods, including multiplex ligation-dependent probe amplification (MLPA), droplet digital PCR (ddPCR), and CNV screening using next-generation sequencing (NGS) data, are explored, each method's strengths and limitations delineated. Whole genome sequencing (WGS) emerges as a powerful tool for comprehensive CNV detection, with specific algorithms and tools recommended for accurate analysis.

Received: 23 November 2023, Revised: 1 December 2023, Accepted: 4 December 2023, Published: 31 December 2023
*Corresponding author: Man Jin Kim, M.D., Ph.D. https://orcid.org/0000-0002-9345-6976
Department of Genomic Medicine, Seoul National University Hospital, 101 Daehak-ro, Jongno-gu, Seoul 03080, Korea.
Tel: +82-2-2072-4265, Fax: +82-2-2072-5260, E-mail: number1@snu.ac.kr
Conflict of interest: I declare that I do not have any conflicts of interests.

The Concept and Characteristics of Exonic CNVs

CNV denotes segments of DNA where variations in copy numbers are detected when comparing two or more genomes. In the absence of additional annotations, CNV does not imply any specific information about the frequency relative to others or the phenotypic impact. These structural variants in quantity can manifest as gains in genomic copy numbers (insertions or duplications) or losses (deletions or null genotypes) compared to a specified reference genome sequence [3]. Here, exonic CNV specifically refers to deletions and duplications at the exon level within specific genes. CNVs affect genes in different ways depending on the degree of overlap with them [2]. Some CNVs cover entire genes (from now on, whole gene CNVs), other CNVs overlap with part of the coding sequence but not the whole gene (exonic CNVs) and other CNVs are found within purely intronic regions (intronic CNVs), not overlapping with any exon from any annotated isoform.

This suggests that detecting large CNVs involving multiple genes (greater than 100 kb) is relatively straightforward. Furthermore, large CNVs are associated with syndromes that induce multiple anomalies, and establishing a genotype-phenotype correlation is comparatively challenging. Therefore, in this review, we aim to explore the relationship between exon level CNVs in well-characterized genes and rare genetic diseases, given their known phenotypes.

Interrogating Recurrent Exonic CNVs

Numerous well-established genes, such as *DMD* [4,5] and *SMN1* [6,7], have been documented for exonic CNVs. In this section, our objective is to investigate CNVs that are comparatively less familiar but recurrently reported in scientific literature. Among these, the enrichment of CNVs in East Asians is particularly intriguing.

1. DRC1

Recently, the exon 1-4 deletion of the *DRC1* (dynein regulatory complex subunit 1) has been documented in a number of Japanese individuals diagnosed with primary ciliary dyskinesia (PCD) [8-10]. Building upon these findings in Japan, a reassessment was carried out in Korea, given its similar ethnic makeup. In Korea, three cases of *DRC1* exon 1-4 deletion were identified, all of which were confirmed as PCD-negative [9]. These three patients, experiencing persistent respiratory symptoms since childhood, such as coughing and sputum production, exhibited homozygous or heterozygous deletions in exon 1-4 of the *DRC1* gene. Despite the presence of chronic sinusitis and bronchiectasis, they did not manifest situs inversus, a common symptom of primary ciliary syndrome. Importantly, there was no reported family history of similar symptoms in these patients.

2. STX16

Pseudohypoparathyroidism type Ib (PHP-1b) is a rare imprinting disorder, characterized by renal parathyroid hormone resistance, but the absence of physical features of Albright hereditary osteodystrophy. A common heterozygous 3-kb deletion of *STX16* gene was previously reported in multiple unrelated probands [11-17]. This deletion causes isolated loss of methylation at *GNAS* exon A/B. The deletion range was from g.57,243,566 to g.57,246,545 (2,979 bp). Since the most common cause of PHP-1b is maternal 3-kb *STX16* deletion, it is recommended to check for the presence of the *STX16* 3-kb-deletion in patients suspected of having PHP-1b.

3. LAMA2

Laminin α2-related muscular dystrophy (LAMA2 MD) is a rare autosomal-recessive genetic disorder, impacting an estimated 0.7 to 2.5 individuals per 100,000 in predominantly European cohorts. An NGS-based CNV profiling was conducted on 114 individuals clinically diagnosed with *LAMA2*-related muscular dystrophy, encompassing 96 individuals with *LAMA2* mutations and 34 with intragenic rearrangements. In total, we identified 18 distinct *LAMA2* CNVs, previously exclusively documented in the Chinese population, with 10 of them being novel. The overall frequency of CNVs in the cohort was 19.3%. Notably, exon 4 deletion was identified in 10 alleles of eight patients, constituting 27% of all CNVs. These patients, of Han Chinese descent, exhibited identical haplotypes and sequences at the breakpoint junction, indicating that exon 4 deletion serves as a founder mutation in the Chinese Han population and represents a mutation hotspot [18].

4. CFTR

Cystic fibrosis stands out as one of the most prevalent autosomal recessively inherited disorders among Caucasians, attributed to pathogenic variants in the *CFTR* (cystic fibrosis transmembrane conductance regulator) gene [19]. However, the incidence of this disease is notably lower in Asian populations. Among a cohort of six Korean patients, five exhibited at least one allele featuring a deletion spanning exons 16-17b, with four displaying a heterozygous deletion and one demonstrat-

ing a homozygous deletion. Notably, 50% of the alleles (six out of 12) manifested a multi-exon deletion spanning 16-17b. All six patients presented with a classical cystic fibrosis phenotype, characterized by chronic steatorrhea and malabsorption since infancy, leading to growth failure and persistent recurrent respiratory symptoms, including chronic sinusitis, mucus plugging, and bronchiectasis. Despite these challenges, all patients survived with supportive care. Recognizing the importance of early diagnosis and management is crucial for enhancing the clinical outcomes of cystic fibrosis patients. Given the high prevalence of multi- or single-exon deletions in *CFTR* among Asian populations, including those in Korea and Japan, molecular investigations aimed at identifying exon deletions should be routinely performed to establish a timely and definitive diagnosis [20].

Diagnosis and Genetic Testing of Exonic CNVs

The diagnosis of exonic CNVs, particularly those encompassing a single exon, poses challenges due to their inherent complexity. Despite the availability of various diagnostic methods, it is recommended to adopt a strategic approach by combining these methods judiciously. Cross-checking the results through the application of multiple diagnostic techniques becomes crucial to ensure the accuracy of the diagnosis, given the intricacies involved in detecting CNVs at the exonic level. This comprehensive and multi-method approach enhances the reliability of identifying and characterizing exonic CNVs, contributing to more precise and informed clinical assessments.

1. MLPA

MLPA is highly sensitive and can detect small changes in DNA copy number, making it effective for identifying both small and large genomic alterations. MLPA allows the simultaneous analysis of multiple targets in a single reaction, enabling the assessment of several genomic loci in a cost-effective and time-efficient manner. MLPA provides quantitative information about the copy number of specific DNA sequences, offering precise insights into genetic variations. MLPA is versatile and can be applied to various sample types, including genomic DNA extracted from blood, tissues, or other biological samples. MLPA is a well-established and reliable technique, widely used in clinical diagnostics and research settings for identifying and characterizing genomic imbalances. However MLPA is targeted and designed for specific genomic regions, which means it may not provide a comprehensive overview of the entire genome. Whole genome approaches may be more suitable for global copy number analysis. The success of MLPA relies on the design of specific probes, and variations outside the targeted regions can be missed. MLPA may face challenges when applied to regions with novel genes or poorly characterized loci, as the probes need prior design based on known genomic sequences. While MLPA provides quantitative data, it is semi-quantitative and may not be as precise as some other quantitative techniques.

2. Droplet Digital PCR

ddPCR is a molecular biology technique that allows for the absolute quantification of nucleic acid targets. One of the major strengths of ddPCR is its ability to provide absolute quantification of nucleic acid targets, offering precise and accurate measurements of target concentrations without the need for standard curves [21,22]. ddPCR is highly sensitive and can detect low-abundance targets, making it suitable for applications where high sensitivity is crucial, such as detecting rare pathogenic variants or monitoring minimal residual disease [23,24]. ddPCR allows for the simultaneous detection and quantification of multiple targets in the same reaction, enhancing its efficiency and reducing the amount of sample needed. The digital nature of ddPCR, where individual reactions are partitioned into thousands of droplets, reduces the impact of reaction inhibitors and provides robust and accurate results. However, ddPCR may have a more limited dynamic range compared to quantitative PCR (qPCR), which can affect its ability to quantify targets across a wide range of concentrations [25]. The initial investment for ddPCR instrumentation can be relatively high, making it less accessible for some laboratories compared to more traditional PCR methods. Analyzing ddPCR data can be more complex compared to conventional PCR, particularly for users who are not familiar with the digital nature of the technique. Designing and optimizing ddPCR assays may require more effort compared to qPCR, and the flexibility to modify assays on the fly may be limited.

3. CNV Screening Using NGS Data from Whole-Exome Sequencing or Targeted Panel Sequencing

The continuous emergence of algorithms for CNV analysis using NGS data reflects the dynamic evolution and advancement in genomic research [26-28]. As NGS technologies progress, researchers and bioinformaticians are developing and refining algorithms to enhance the accuracy and efficiency of CNV detection. This ongoing trend signifies the commitment of the scientific community to harness the potential of NGS data for unraveling genomic complexities, contributing to a deeper understanding of genetic variations, and ultimately advancing pre-

cision medicine and diagnostic capabilities. The challenge lies in detecting germline CNVs from targeted NGS data, especially for single and multi-exon alterations. One study evaluates five CNV calling tools (DECoN, CoNVaDING, panelcn.MOPS, ExomeDepth, and CODEX2) using four genetic diagnostics datasets, totaling 495 samples with 231 validated CNVs [26]. The evaluation, conducted with default and sensitivity-optimized parameters, reveals that most tools exhibit high sensitivity and specificity, although performance varies based on the dataset. In the diagnostic scenario, DECoN and panelcn.MOPS emerge as effective for CNV screening, with DECoN showing superior specificity. The study highlights the importance of tool selection and parameter optimization for accurate CNV detection in genetic diagnostics. However, the absence of a standardized approach remains a prevailing reality within the discipline.

4. Whole Genome Sequencing

WGS is a powerful method for detecting CNVs across the entire genome. By employing high-throughput sequencing techniques, WGS provides a comprehensive view of genomic alterations, making it effective in identifying CNVs of various sizes. Additionally, structural variant analysis tools such as Delly and Manta are commonly utilized in conjunction with WGS data to identify complex rearrangements, including insertions, deletions, and inversions [29,30]. These tools contribute to a more thorough understanding of genomic architecture and aid in the accurate detection of CNVs, further enhancing the capabilities of WGS in unraveling structural variations at the nucleotide level.

However, accurately calling CNVs from WGS remains challenging, lacking a consensus. Another study explores practical calling options, highlighting the complementary results obtained from callers based on different signals (paired-end reads, split reads, coverage depth). The authors propose a combined approach using four selected callers (Manta, Delly, ERDS, CNVnator) and a regenotyping tool (SV2), demonstrating its applicability in routine practice in terms of computation time and interpretation. The study showcases the superiority of these approaches over array-based comparative genomic hybridization (aCGH), particularly in breakpoint definition resolution and the detection of potentially relevant CNVs. The findings are confirmed on benchmark and clinically validated genomes, suggesting that WGS presents a timely and economically viable alternative to the combination of aCGH and whole-exome sequencing [31].

5. Gap PCR

Confirmatory PCR is the most reliable method for confirming deletions and duplications. However, it requires a relatively precise understanding of the CNV range, this involves the inconvenience of designing new primers for experimentation. Specifically defining the CNV range allows for the design of specific primers for conducting Gap PCR [32].

Conclusion

Large-sized CNVs are typically identified using chromosomal microarray (CMA). However, for small-sized CNVs, understanding their prevalence in specific genes is crucial for utilizing methods such as MLPA or long-read sequencing. Some genes may have hotspots for exonic CNVs, and certain populations may show enrichment of these CNVs, possibly due to founder effects [8-10,33]. Detection methods such as MLPA, qPCR, ddPCR are preferred over traditional methods like CMA or karyotyping. WGS can also be considered, especially if the target gene is well-defined. When NGS data is available, various algorithms can be employed to detect exonic CNVs; however, these algorithms may be less robust in identifying single exon deletions or duplications. In a broader context, CNVs, particularly those of smaller scale such as exonic CNVs, may conceal latent genetic variations. The discernment and therapeutic targeting of these specific CNVs, potentially influenced by ethnic considerations, have the potential to substantially advance the realms of rare disease diagnosis and treatment. The continuous effort to discover and understand these subtle CNVs is highly significant, as it holds the potential to lead to the development of ethnically-specific treatments for rare diseases in the future.

Acknowledgements

None.

Funding

No fundings to declare.

References

1. Henrichsen CN, Chaignat E, Reymond A. Copy number variants, diseases and gene expression. Hum Mol Genet 2009;18:R1-8.
2. Rigau M, Juan D, Valencia A, Rico D. Intronic CNVs and gene expression variation in human populations. PLoS Genet 2019;15:e1007902.

3. Lee C, Scherer SW. The clinical context of copy number variation in the human genome. Expert Rev Mol Med 2010;12:e8.
4. Kozareva V, Stroff C, Silver M, Freidin JF, Delaney NF. Clinical analysis of germline copy number variation in DMD using a non-conjugate hierarchical Bayesian model. BMC Med Genomics 2018;11:91.
5. Wei X, Dai Y, Yu P, Qu N, Lan Z, Hong X, et al. Targeted next-generation sequencing as a comprehensive test for patients with and female carriers of DMD/BMD: a multi-population diagnostic study. Eur J Hum Genet 2014;22:110-8.
6. Passon N, Dubsky de Wittenau G, Jurman I, Radovic S, Bregant E, Molinis C, et al. Quick MLPA test for quantification of SMN1 and SMN2 copy numbers. Mol Cell Probes 2010;24:310-4.
7. Vijzelaar R, Snetselaar R, Clausen M, Mason AG, Rinsma M, Zegers M, et al. The frequency of SMN gene variants lacking exon 7 and 8 is highly population dependent. PLoS One 2019;14:e0220211.
8. Keicho N, Hijikata M, Morimoto K, Homma S, Taguchi Y, Azuma A, et al. Primary ciliary dyskinesia caused by a large homozygous deletion including exons 1-4 of DRC1 in Japanese patients with recurrent sinopulmonary infection. Mol Genet Genomic Med 2020;8:e1033.
9. Kim MJ, Kim S, Chae SW, Lee S, Yoon JG, Kim B, et al. Prevalence and founder effect of DRC1 exon 1-4 deletion in Korean patients with primary ciliary dyskinesia. J Hum Genet 2023;68:369-74.
10. Morimoto K, Hijikata M, Zariwala MA, Nykamp K, Inaba A, Guo TC, et al. Recurring large deletion in DRC1 (CCDC164) identified as causing primary ciliary dyskinesia in two Asian patients. Mol Genet Genomic Med 2019;7:e838.
11. Chu X, Zhu Y, Wang O, Nie M, Quan T, Xue Y, et al. Clinical and genetic characteristics of Pseudohypoparathyroidism in the Chinese population. Clin Endocrinol (Oxf) 2018;88:285-94.
12. Danzig J, Li D, Jan de Beur S, Levine MA. High-throughput molecular analysis of pseudohypoparathyroidism 1b patients reveals novel genetic and epigenetic defects. J Clin Endocrinol Metab 2021;106:e4603-20.
13. Elli FM, de Sanctis L, Peverelli E, Bordogna P, Pivetta B, Miolo G, et al. Autosomal dominant pseudohypoparathyroidism type Ib: a novel inherited deletion ablating STX16 causes loss of imprinting at the A/B DMR. J Clin Endocrinol Metab 2014;99:E724-8.
14. Jüppner H. Molecular definition of pseudohypoparathyroidism variants. J Clin Endocrinol Metab 2021;106:1541-52.
15. Kiuchi Z, Reyes M, Brickman AS, Jüppner H. A distinct variant of pseudohypoparathyroidism (PHP) first characterized some 41 years ago is caused by the 3-kbSTX16 deletion. JBMR Plus 2021;5:e10505.
16. Linglart A, Gensure RC, Olney RC, Jüppner H, Bastepe M. A novel STX16 deletion in autosomal dominant pseudohypoparathyroidism type Ib redefines the boundaries of a cis-acting imprinting control element of GNAS. Am J Hum Genet 2005;76:804-14. Erratum in: Am J Hum Genet 2007;81:196.
17. Ramalho E Silva JD, da Rocha GFMA, Oliveira MJM. An intricate case of sporadic pseudohypoparathyroidism type 1B with a review of literature. Arch Endocrinol Metab 2021;65:112-6.
18. Ge L, Liu A, Gao K, Du R, Ding J, Mao B, et al. Deletion of exon 4 in LAMA2 is the most frequent mutation in Chinese patients with laminin α2-related muscular dystrophy. Sci Rep 2018;8:14989.
19. Palomaki GE, FitzSimmons SC, Haddow JE. Clinical sensitivity of prenatal screening for cystic fibrosis via CFTR carrier testing in a United States panethnic population. Genet Med 2004;6:405-14.
20. Sohn YB, Ko JM, Jang JY, Seong MW, Park SS, Suh DI, et al. Deletion of exons 16-17b of CFTR is frequently identified in Korean patients with cystic fibrosis. Eur J Med Genet 2019;62:103681.
21. Hindson BJ, Ness KD, Masquelier DA, Belgrader P, Heredia NJ, Makarewicz AJ, et al. High-throughput droplet digital PCR system for absolute quantitation of DNA copy number. Anal Chem 2011;83:8604-10.
22. Hindson CM, Chevillet JR, Briggs HA, Gallichotte EN, Ruf IK, Hindson BJ, et al. Absolute quantification by droplet digital PCR versus analog real-time PCR. Nat Methods 2013;10:1003-5.
23. Coccaro N, Tota G, Anelli L, Zagaria A, Specchia G, Albano F. Digital PCR: a reliable tool for analyzing and monitoring hematologic malignancies. Int J Mol Sci 2020;21:3141.
24. Zhou B, Haney MS, Zhu X, Pattni R, Abyzov A, Urban AE. Detection and quantification of mosaic genomic DNA variation in primary somatic tissues using ddPCR: analysis of mosaic transposable-element insertions, copy-number variants, and single-nucleotide variants. Methods Mol Biol 2018;1768:173-90.
25. Zhao Y, Xia Q, Yin Y, Wang Z. Comparison of droplet digital PCR and quantitative PCR assays for quantitative detection of Xanthomonas citri Subsp. citri. PLoS One 2016;11:e0159004.
26. Moreno-Cabrera JM, Del Valle J, Castellanos E, Feliubadaló L, Pineda M, Brunet J, et al. Evaluation of CNV detection tools for NGS panel data in genetic diagnostics. Eur J Hum Genet 2020;28:1645-55.
27. Singh AK, Olsen MF, Lavik LAS, Vold T, Drabløs F, Sjursen W. Detecting copy number variation in next generation sequencing data from diagnostic gene panels. BMC Med Genomics 2021;14:214.
28. Zhao M, Wang Q, Wang Q, Jia P, Zhao Z. Computational tools for copy number variation (CNV) detection using next-generation sequencing data: features and perspectives. BMC Bioinformatics 2013;14 Suppl 11:S1.
29. Chen X, Schulz-Trieglaff O, Shaw R, Barnes B, Schlesinger F, Källberg M, et al. Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. Bioinformatics 2016;32:1220-2.
30. Rausch T, Zichner T, Schlattl A, Stütz AM, Benes V, Korbel JO. DELLY:

structural variant discovery by integrated paired-end and split-read analysis. Bioinformatics 2012;28:i333-9.

31. Coutelier M, Holtgrewe M, Jäger M, Flöttman R, Mensah MA, Spielmann M, et al. Combining callers improves the detection of copy number variants from whole-genome sequencing. Eur J Hum Genet 2022;30:178-86.
32. Fan DM, Yang X, Huang LM, Ouyang GJ, Yang XX, Li M. Simultaneous detection of target CNVs and SNVs of thalassemia by multiplex PCR and next-generation sequencing. Mol Med Rep 2019;19:2837-48.
33. Takeuchi K, Xu Y, Kitano M, Chiyonobu K, Abo M, Ikegami K, et al. Copy number variation in DRC1 is the major cause of primary ciliary dyskinesia in the Japanese population. Mol Genet Genomic Med 2020;8:e1137.