



Limited diagnostic impact of duplications <1 Mb of uncertain clinical significance: a 10-year retrospective analysis of reporting practices at the Mayo Clinic

Cherisse A. Marcou, PhD¹, Beth Pitel, MS¹, Clinton E. Hagen, MS², Nicole J. Boczek, PhD¹, Ross A. Rowsey, PhD¹, Linda B. Baughn, PhD¹, Nicole L. Hoppman, PhD¹, Erik C. Thorland, PhD¹ and Hutton M. Kearney, PhD¹

Purpose: Copy-number variants (CNVs) of uncertain clinical significance are routinely reported in a clinical setting only when exceeding predetermined reporting thresholds, typically based on CNV size. Given that very few genes are associated with triplosensitive phenotypes, it is not surprising that many interstitial duplications <1 Mb are found to be inherited and anticipated to be of limited or no clinical significance.

Methods: In an effort to further refine our reporting criteria to maximize diagnostic yield while minimizing the return of uncertain variants, we performed a retrospective analysis of all clinical microarray cases reported in a 10-year window. A total of 1112 reported duplications had parental follow-up, and these were compared by size, RefSeq gene content, and inheritance pattern. *De novo* origin was used as a rough proxy for pathogenicity.

Results: Approximately 6% of duplications 500 kb–1 Mb were *de novo* observations, compared with approximately 14% for 1–2 Mb duplications ($p = 0.0005$). On average, *de novo* duplications had higher gene counts than inherited duplications.

Conclusion: Our data reveal limited diagnostic utility for duplications of uncertain significance <1 Mb. Considerations for revised reporting criteria are discussed and are applicable to CNVs detected by any genome-wide exploratory methodology, including exome/genome sequencing.

Genetics in Medicine (2020) 22:2120–2124; <https://doi.org/10.1038/s41436-020-0932-0>

Keywords: copy-number variant; reporting criteria; duplication; triplosensitivity; microarray

INTRODUCTION

The use of genomic microarrays in the diagnostic laboratory setting has enhanced the ability to detect copy-number variation (CNV) that underlies the pathogenicity of human disease. Genomic microarrays for copy-number assessment, both array-based comparative genomic hybridization (aCGH) and single-nucleotide polymorphism (SNP)-based microarrays, are collectively described as “chromosomal microarrays” (CMA). These assays have evolved over time, from low density targeted bacterial artificial chromosome (BAC) arrays to the now high density copy-number plus SNP arrays that not only detect CNVs but also absence of heterozygosity.^{1,2} The appreciation for the diagnostic utility of CMA is reflected in numerous consensus statements establishing these assays as first-tier tests for many clinical indications.^{3,4} Additionally, reporting of CNVs from genome-wide sequencing is increasingly becoming a standard approach, with performance for CNV detection near or exceeding CMA.⁵ An incidental casualty of such comprehensive genomic screens is the detection of variants of uncertain clinical significance. The goal of clinical interpretation and reporting strategies for

discovered variants is to maximize return of useful diagnostic findings, while minimizing the return of uncertain variants, especially those that are likely benign.

When first introduced, professional guidelines for CMA required that positive findings be orthogonally confirmed, usually by fluorescence in situ hybridization (FISH);⁶ however, current guidance no longer recommends this routine practice.⁷ Given that most FISH assays have a lower limit of detection of ~200 kb for deletions and ~500 kb for duplications, reporting thresholds based on these, or similar, sizes were adopted by the majority of clinical laboratories for CMA findings, and have persisted for more than a decade.

Interpretation of the clinical significance of CNVs is complex and thoroughly addressed by recently updated American College of Medical Genetics and Genomics (ACMG)/ClinGen guidance.⁸ Interpretive considerations for nonrecurrent duplications can be broadly grouped into three categories: (1) evaluation of fully encompassed genes for potential triplosensitive phenotypes; (2) evaluation of genes mapping to duplication breakpoint(s), for potential haploinsufficient phenotypes; and (3) consideration of more rare

¹Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN, USA; ²Department of Health Sciences Research, Mayo Clinic, Rochester, MN, USA. Correspondence: Cherisse A. Marcou (marcou.cherisse@mayo.edu)

Submitted 11 March 2020; revised 26 July 2020; accepted: 28 July 2020
Published online: 21 August 2020

pathogenic mechanisms such as gene disruption at other loci due to insertional translations^{9,10} or positional effects on gene regulation.¹¹ Considering the first category, triplosensitivity is an uncommon cause of abnormal phenotypes.¹² Therefore, it is not surprising that even moderately sized (500 kb–1 Mb) duplications are frequently observed in the general population without observed clinical consequence.¹³ Considering the second category is less straightforward. Intragenic duplications with *both* breakpoints within the same gene are generally expected to disrupt coding sequence and result in loss-of-function variants. However, as the majority of interstitial duplications have a direct tandem orientation, partial gene duplications with *only a single breakpoint* in the gene of interest most often occur without associated gene disruption.¹⁴ Therefore, small duplications (<500 kb) involving only a single breakpoint in dosage-sensitive genes are almost always inherited and inferred to have an intact copy of the gene(s) at the breakpoint; data supporting this observation are presented below. Considerations involving the third category are more complex, and often outside the available scope of a routine clinical study. However, technological and knowledge-base advancements continue to evolve, allowing for increasing diagnostic clarity in such complex cases.

After more than 15 years of experience interpreting, reporting, characterizing, and performing parental/familial follow-up testing for duplication CNVs, our laboratory has appreciated limited clinical impact for duplications less than 1–2 Mb in size initially assessed to have uncertain clinical significance. We designed a retrospective analysis of our experience to provide evidence for this assumption and to support a clinically appropriate refinement of our laboratory reporting criteria.

MATERIALS AND METHODS

A retrospective analysis of all clinical postnatal chromosomal microarray cases reported from 2006 to 2016 was performed to identify all cases with reported duplication CNVs ($n = 4059$). During this time period, using standard interpretive approaches,^{15,16} all pathogenic or likely pathogenic duplications were reported, regardless of size. Duplications with breakpoints in genes with known or strongly suspected haploinsufficiency phenotypes were also reported, regardless of size. Other duplications of uncertain clinical significance were only reported when exceeding 500 kb in size. Duplications of any size without protein-encoding genes or known

functional elements were not reported, nor were events representing common benign polymorphism, therefore this series does not include such duplications. Of the 4059 cases with reported duplications, 1624 cases were analyzed using the CytoScan HD microarray (Thermo Fisher Scientific, Waltham, MA) and 2435 cases were analyzed using a custom 44K or 180K Agilent microarray (Agilent, Santa Clara, CA). Pathogenic and likely pathogenic intragenic duplications in known haploinsufficient genes with both breakpoints within the gene (high likelihood of pathogenic loss-of-function variants) were excluded from further analysis. Duplications mapping to the sex chromosomes were excluded from further analysis given the frequency of maternal inheritance and the complexity in deciphering pathogenicity and clinical significance. We further restricted our focus to those cases with available parental inheritance information, either both parents tested, or a single parent tested with documented inheritance. These cases were binned by size (<500 kb, 500 kb–1 Mb, 1 Mb–2 Mb, >2 Mb) to roughly determine the proportion of *de novo* (and thus more likely pathogenic) duplications in each size range. Statistical analysis was performed using Fisher's exact test; two-sided P value <0.05 was considered statistically significant.

RESULTS

Of the 4059 total cases reported with duplications, there were a total of 708 duplication CNVs less than 500 kb (reporting in this size range already is restricted to known or suspected pathogenic findings, therefore this is an underrepresentation of observed duplications in this size range), 1503 cases with duplication CNVs in the size range of 500 kb to 1 Mb, 689 cases with duplication CNVs in the size range of 1 Mb to 2 Mb, and 1159 cases with duplication CNVs greater than 2 Mb (Table 1). Of these total reported cases, 1112 met our inclusion criteria defined in "Materials and Methods": 197 duplication CNVs less than 500 kb, 515 cases with duplication CNVs in the size range of 500 kb to 1 Mb, 235 cases with duplication CNVs in the size range of 1 Mb to 2 Mb, and 165 cases with duplication CNVs greater than 2 Mb (Table 1). Using inheritance pattern as a rough proxy for likelihood of pathogenicity, we demonstrated a *de novo* origin in 4% (8/197) of selected duplications less than 500 kb, 6% (30/515) of duplications in the size range of 500 kb to 1 Mb, 14% (32/235) of duplications in the size range of 1 Mb to 2 Mb, and 28% (46/165) of duplications greater than 2 Mb. We observed a statistically significant increase in the occurrence

Table 1 Cohort of duplication cases included in the study binned by size.

Size	Total number of cases ^a	Number of cases with parental follow-up (%)	Number maternally inherited (%)	Number paternally inherited (%)	Number <i>de novo</i> (%)
<500 kb	708	197 (42.5)	107 (54.3)	82 (41.6)	8 (4.06)
500 kb–1 Mb	1503	515 (40.3)	269 (52.2)	216 (41.9)	30 (5.83)
1–2 Mb	689	235 (35.8)	114 (48.7)	89 (37.8)	32 (13.6)
>2 Mb	1159	165 (17.1)	78 (47.3)	41 (25.0)	46 (27.9)

^aRepresents total number of reported cases in the cohort before exclusion criteria were applied (see "Materials and Methods").

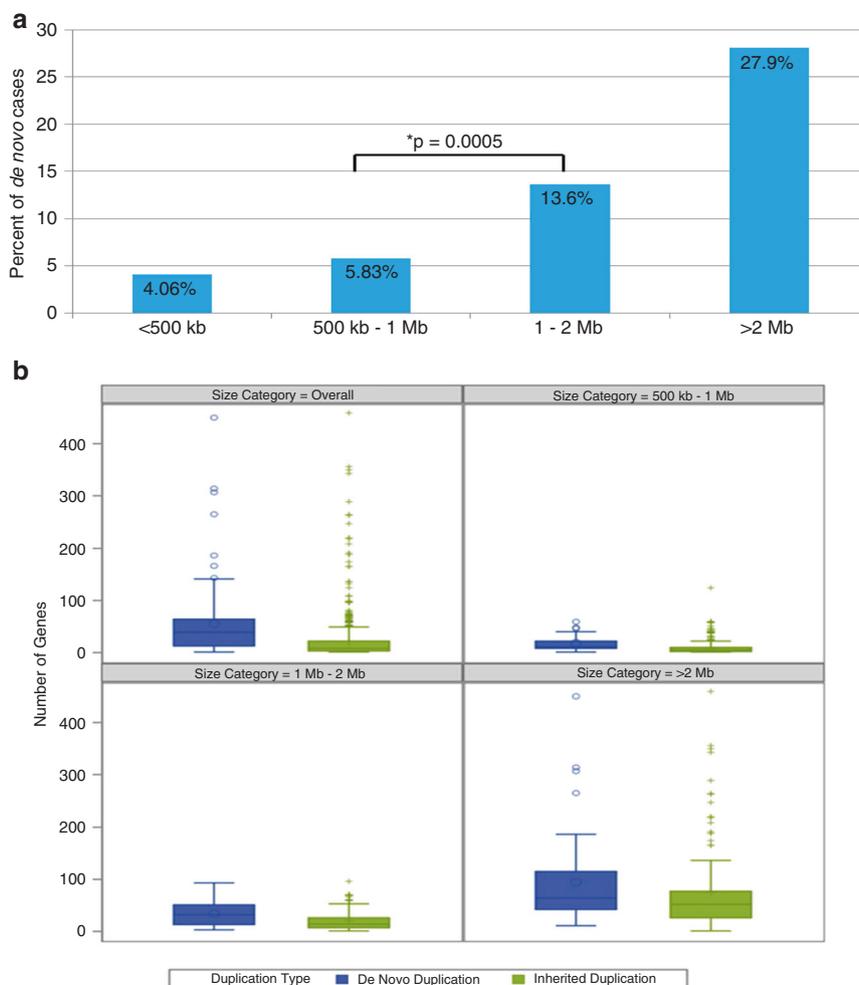


Fig. 1 Size distribution of de novo duplications and RefSeq gene content assessment. (a) A statistically significant increase in the occurrence of de novo duplications at the size cutoff of 1 Mb. (b) Boxplot distribution of RefSeq gene content of de novo and inherited duplications binned by size.

of de novo duplications between those binned at 500 kb–1 Mb and those binned at 1 Mb–2 Mb ($p = 0.0005$) (Fig. 1a). Of note, there is an estimated background rate of de novo CNVs in control populations (estimated at ~1%)¹⁷ that could be represented in the data set, as well as a rate of misattributed paternity (estimated at ~1%), which is known to result in the false assessment of a small proportion of de novo cases;¹⁸ however, this should apply to a similar proportion of cases in each size category.

As discussed previously, duplications less than 500 kb are reported at the Mayo Clinic when (1) one or more genes fully contained within the interval has documented triplosensitivity; (2) they are fully intragenic (both breakpoints within the gene), with potential/expected loss of function of a haploinsufficient gene; or (3) there is a concern for potential gene disruption of a haploinsufficient gene harboring a single duplication breakpoint. Pathogenic or likely pathogenic small duplications meeting criteria 1 or 2 were removed from this analysis, as these are always reported in our laboratory. A total of 197 cases were reported with concern for disruption of a haploinsufficient gene (due to a single duplication breakpoint

within that gene) and had parental follow-up (Table 1). More than 95% of these duplications were inherited, suggesting that the vast majority of these events are not pathogenic. This is consistent with the demonstration by Newman *et al.* that the majority of such duplications are direct tandem events that are not disruptive to genes at the breakpoint.¹⁴

To further assess the utility and clinical impact of parental follow-up for small duplications, we retrospectively reviewed the 30 duplications 500 kb to 1 Mb in size found to have de novo inheritance after reporting (See Supplemental Table 1). Eleven of the 30 duplications (37%) would continue to meet reporting criteria in our laboratory, regardless of size, due to documented or high likelihood of pathogenicity or known recurrent regions that have documented reduced penetrance or variable expressivity. The remaining 19 duplications were classified as uncertain at the time of original reporting, and remain classified as uncertain today leveraging the current knowledgebase and ACMG/ClinGen guidelines, despite de novo inheritance.

Lastly, in an attempt to further refine the reporting thresholds, we used our data set to compare the RefSeq gene

content of de novo and inherited duplications within each size bin. We identified a median number of 5 genes for inherited duplications and a median number of 11 genes for de novo duplications in the size range of 500 kb to 1 Mb, a median number of 14 genes for inherited duplications and a median number of 32 genes for de novo duplications in the size range of 1 Mb to 2 Mb, and a median number of 52 genes for inherited duplications and a median number of 64 genes for de novo duplications greater than 2 Mb. The variability around the median, however, was pronounced (Fig. 1b).

DISCUSSION

Here we report the retrospective assessment of a large series of duplication CNVs reported in a diagnostic laboratory setting in an attempt to evaluate and refine reporting criteria. Our data suggest that, apart from fully intragenic duplications expected to result in loss-of-function variants in haploinsufficient genes and those fully encompassing known triplosensitive genes, duplication CNVs less than 1 Mb are very unlikely to be clinically significant in the absence of additional evidence of pathogenicity. Approximately 95% of duplications less than 1 Mb were shown to be inherited from a carrier parent, suggesting that the majority of these reported duplications are likely benign familial variants. Our retrospective assessment of the small number of de novo duplications in this size range supports our assertion of minimal clinical utility of reporting duplications <1 Mb of uncertain clinical significance, as those that were classified as uncertain remained so even after the finding of de novo inheritance and maturation of our collective knowledgebase (see Supplemental Table 1). There are caveats to this assumption that are important to acknowledge. For example, a small proportion of the remaining 5% of cases with de novo duplications 500–1 kb in size actually represent inherited variants with misattributed paternity (estimated at ~1%),¹⁸ and a small proportion of de novo cases will simply represent spurious new variants unrelated to the proband's phenotype of interest.¹⁷ Additionally, incomplete penetrance or variable expressivity of abnormal phenotypes complicates the assumption that CNVs inherited from unaffected parents are not clinically significant. Therefore, we are making the assumption of pathogenicity in aggregate across the large data set, with the understanding that a small proportion of inherited cases do have relevant clinical significance and a small proportion of de novo cases do not have relevant clinical significance.

Our assessment of the number of genes in inherited versus de novo duplications supports the general assertion that duplications containing larger numbers of protein-encoding genes are more likely to have clinical significance; however, the significant variability around the median in these data did not allow for further confident refinement of reporting thresholds beyond those established on the genomic size of the duplication. Regardless, the number of genes in a CNV interval is now included in the current pathogenicity

assessment rubric per ACMG/ClinGen guidance,⁸ therefore the total gene count will influence reporting decisions.

In conclusion, this retrospective analysis supports a more clinically appropriate reporting threshold of ≥ 1 Mb for duplications of uncertain clinical significance. This revised reporting threshold is applicable to duplication CNVs detected by any methodology, including CMA and exome/genome sequencing. Importantly, any duplication <1 Mb will continue to be reviewed and will be reported in our laboratory when there is known or highly suspected clinical significance. This threshold change is estimated to impact approximately 5% of CMA cases, which are currently reported as uncertain findings and, after family studies, presumed to have unlikely significance. This reporting policy change will result in significant savings of health-care costs and eliminate the uncertainty and burden of further testing for the families involved. As variants not meeting reporting criteria are maintained in laboratory databases, they will continue to contribute to the growth of our collective knowledgebase such that when the pathogenicity of smaller CNVs is more clear, they can be reported with clinical confidence.

SUPPLEMENTARY INFORMATION

The online version of this article (<https://doi.org/10.1038/s41436-020-0932-0>) contains supplementary material, which is available to authorized users.

DISCLOSURE

The authors declare no conflicts of interest.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

REFERENCES

- Schwartz S. Clinical utility of single nucleotide polymorphism arrays. *Clin Lab Med.* 2011;31:581–594.
- Kearney HM, South ST, Wolff DJ, et al. American College of Medical Genetics recommendations for the design and performance expectations for clinical genomic copy number microarrays intended for use in the postnatal setting for detection of constitutional abnormalities. *Genet Med.* 2011;13:676–679.
- Miller DT, Adam MP, Aradhya S, et al. Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies. *Am J Hum Genet.* 2010;86:749–764.
- Manning M, Hudgins L, Professional Practice and Guidelines Committee. Array-based technology and recommendations for utilization in medical genetics practice for detection of chromosomal abnormalities. *Genet Med.* 2010;12:742–745.
- Stavropoulos DJ, Merico D, Jobling R, et al. Whole genome sequencing expands diagnostic utility and improves clinical management in pediatric medicine. *NPJ Genom Med.* 2016;1:15012.
- Shaffer LG, Beaudet AL, Brothman AR, et al. Microarray analysis for constitutional cytogenetic abnormalities. *Genet Med.* 2007;9:654–662.
- South ST, Lee C, Lamb AN, et al. ACMG Standards and Guidelines for constitutional cytogenomic microarray analysis, including postnatal and prenatal applications: revision 2013. *Genet Med.* 2013;15:901–909.
- Riggs ER, Andersen EF, Cherry AM, 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). *Genet Med.* 2020;22:245–257.

9. Kang SH, Shaw C, Ou Z, et al. Insertional translocation detected using FISH confirmation of array-comparative genomic hybridization (aCGH) results. *Am J Med Genet A*. 2010;152A:1111–1126.
10. Neill NJ, Ballif BC, Lamb AN, et al. Recurrence, submicroscopic complexity, and potential clinical relevance of copy gains detected by array CGH that are shown to be unbalanced insertions by FISH. *Genome Res*. 2011;21:535–544.
11. Zepeda-Mendoza CJ, Ibn-Salem J, Kammin T, et al. Computational prediction of position effects of apparently balanced human chromosomal rearrangements. *Am J Hum Genet*. 2017;101:206–217.
12. Clinical Genome Resource (ClinGen). <https://www.ncbi.nlm.nih.gov/projects/dbvar/clingen/index.shtml>. Accessed March 2020.
13. MacDonald JR, Ziman R, Yuen RK, et al. The Database of Genomic Variants: a curated collection of structural variation in the human genome. *Nucleic Acids Res*. 2014;42:D986–992.
14. Newman S, Hermetz KE, Wechselblatt B, et al. Next-generation sequencing of duplication CNVs reveals that most are tandem and some create fusion genes at breakpoints. *Am J Hum Genet*. 2015;96:208–220.
15. Kearney HM, Thorland EC, Brown KK, et al. American College of Medical Genetics standards and guidelines for interpretation and reporting of postnatal constitutional copy number variants. *Genet Med*. 2011;13:680–685.
16. Riggs ER, Church DM, Hanson K, et al. Towards an evidence-based process for the clinical interpretation of copy number variation. *Clin Genet*. 2012;81:403–412.
17. Itsara A, Wu H, Smith JD, et al. De novo rates and selection of large copy number variation. *Genome Res*. 2010;20:1469–1481.
18. Brock DJ, Shrimpton AE. Non-paternity and prenatal genetic screening. *Lancet*. 1991;338:1151.