

Genomic Variation: Lessons Learned from Whole-Genome CNV Analysis

Erin Rooney Riggs · David H. Ledbetter ·
Christa Lese Martin

Published online: 18 July 2014

© The Author(s) 2014. This article is published with open access at Springerlink.com

Abstract One of the most fundamental goals of the study of human genetics was to determine the relationship between genomic variation and human disease. The effects of large-scale structural variation, such as aneuploidy and other cytogenetically visible imbalances, as well as sequence-level variation, have been studied for several decades. However, compared to these, the impact of submicroscopic copy number variants (CNV) has only recently been appreciated. Despite this, lessons learned from the study of CNVs have already proven significant and broadly applicable. From expanding the concept of normal human variation to providing concrete examples of the utility of genomics in clinical care and challenging notions of the genetic architecture of complex disease, CNVs have provided valuable insights into the genomics of human health and development.

Keywords Copy number variants · Structural variation · Gene dosage · Chromosomal microarray

Introduction

The physical map and DNA sequence derived from the Human Genome Project [1] revolutionized cytogenetic

testing by providing the ability to detect submicroscopic imbalances across the genome. Technologic advances, namely the advent of comparative genomic hybridization, allowed for the detection of CNVs in the research setting in the 1990s [2–5]; subsequent improvements in the test design gave chromosomal microarray (CMA) more utility in the clinical diagnostic setting. These advances included the addition of large genomic clones (e.g., bacterial artificial chromosomes, or BACs) to allow for the detection of single-copy losses or gains [6] followed by the replacement of genomic clones with synthetic oligonucleotides [7] for genome-wide interrogation and more precise identification of breakpoints. By the late 2000s, oligonucleotide CMA designs were implemented for clinical testing that included both targeted (representing known clinically relevant regions) and genome-wide backbone coverage [8]. This design schema enabled CMA to identify all imbalances detectable by karyotype plus submicroscopic CNVs, thus surpassing the diagnostic yield of a G-banded karyotype [9••].

At the time CMA was first being implemented in the clinic, the role of CNVs in human disease was still largely unclear—the mechanisms of formation were not completely understood, and the clinical significance of many novel findings was frustratingly uncertain. Over time, through intense examination and data sharing, the roles, mechanisms, and significance of CNVs have become clearer, laying the foundation for discovery in the next-generation sequencing era.

Expanding the Scope of Normal Human Variation

Although structural variation in normal individuals has been appreciated for decades at the microscopic level [10], the extent to which the human genome was subjected to

E. R. Riggs · C. L. Martin (✉)
Autism and Developmental Medicine Institute, Geisinger Health System, 120 Hamm Drive, Suite 2A, Lewisburg, PA 17837, USA
e-mail: clmartin1@geisinger.edu

E. R. Riggs
e-mail: eriggs@geisinger.edu

D. H. Ledbetter
Geisinger Health System, 100 N Academy Ave., Danville, PA 17822, USA
e-mail: dhledbetter@geisinger.edu

submicroscopic copy number variation was not realized until 2004. Studies using CMA technology revealed CNVs throughout the genomes of normal individuals, several of which were present in >10 % of the individuals studied [11, 12]. Indeed, CNVs are thought to account for ~1 % of the variation between two individuals; in contrast, single nucleotide polymorphisms (SNPs) are thought to account for approximately 0.1 % [13].

Studies documenting common areas of normal structural variation continue to serve as valuable resources for those evaluating the clinical significance of CNVs; it has historically been assumed that CNVs identified in “normal” populations could be classified as “likely benign” or “benign.” Early guidelines for the clinical interpretation of CNVs proposed that CNVs inherited from reportedly normal parents could “probably” be considered benign [9•]. However, assumptions such as these should be made with caution and in the context of the family’s clinical presentation. There are many examples of “pathogenic” CNVs exhibiting reduced penetrance and/or variable expressivity, and many instances in which a more severely affected child has inherited a CNV from a seemingly normal parent. The 22q11.2 deletion (del) syndrome is a classic example of this phenomenon [14], though few would argue about the pathogenicity of this particular CNV.

A more fitting example might be del 15q11.2, including the region between breakpoints (BP) 1–2. A number of case-control studies have demonstrated that this deletion is enriched in cases as compared to controls [15–17]. Nonetheless, the fact that it has been observed in control individuals and unaffected relatives, coupled with the broad spectrum of associated phenotypes (developmental delay [15•, 16], schizophrenia [18], epilepsy [17, 19], etc.), has anecdotally resulted in some hesitations to classify it as “pathogenic.” However, a recent study of control individuals found to carry CNVs previously associated with neuropsychiatric disorders, such as autism spectrum disorders (ASD) and schizophrenia (including del 15q11.2 BP 1–2), showed that these individuals performed at a level between that of schizophrenic patients and population controls on a series of cognitive measures, even though they had never received a formal neuropsychiatric diagnosis [20•]. This observation gives more credence to the interpretation of del 15q11.2 as pathogenic with variable expressivity but, more importantly, puts forth the idea that CNVs observed in seemingly normal populations could indeed be conferring varying levels of clinical effects, challenging the notion that variation found in normal populations is predominantly benign. Therefore, quantitative measures assessing neurodevelopmental phenotypes, such as cognition and behavior, may be more helpful than broad categorical diagnoses (e.g., affected versus unaffected) when trying to establish the effect of genomic variation [21].

A Paradigm Shift in Clinical Genetic Care

While certain CNVs are common within the “normal” population, others have been associated with disorders of human health and development. Several genomic disorders were identified with the advent of high-resolution chromosome banding and fluorescence in situ hybridization (FISH) technologies (Prader Willi/Angelman syndromes [OMIM 176270/105830]; Miller-Dieker syndrome [OMIM 247200], Williams-Beuren syndrome [OMIM 194050], etc.); however, the widespread use of CMA allowed for the identification of numerous others at a remarkable pace [22–26], often before a clear phenotypic picture had emerged.

As CMA design evolved to include targeted coverage of clinically relevant regions and uniformly spaced backbone coverage throughout the euchromatic regions of each chromosome [8], CMA truly became a “genome-wide” assay. Although the G-banded karyotype was essentially the first genome-wide assay in the most basic sense of the term, CMA provided clinicians with a way to interrogate the entire genome with a single, high-resolution assay. Before this, genetic diagnoses were made based upon a clinician’s observations of the patients presenting phenotype, and that phenotype’s consistency (or lack thereof) with previously described genetic syndromes. Even with the advent of sequence-based genetic testing, diagnoses still relied on the clinician’s ability to deduce a plausible set of differential diagnoses from the observed phenotype and select the correct gene(s) to test, if clinically available. Diagnostic testing under this paradigm required an a priori idea of the underlying diagnosis and causative mechanism, and testing each of the possible differential diagnoses was undertaken separately. This approach was ineffective if the suspected clinical diagnosis was incorrect or if the causative mechanism was unknown, leaving many patients without a confirmed genetic diagnosis. It naturally followed that those presenting with classical symptoms of well-described genetic conditions had the best chance of obtaining a diagnosis, while those with more ambiguous symptoms or with rarely or never described conditions remained undiagnosed or were misdiagnosed.

With CMA, no priori predictions of the patient’s diagnosis are required, making the test particularly appealing for use in those with non-specific symptoms, including developmental delays, ASD, and congenital anomalies. By evaluating the entire genome at once, both previously described syndromes and novel etiologies could be identified. CMA becoming recognized as a first-tier test for these groups of individuals [27•] represented a paradigm shift in the diagnosis of genetic disorders from “phenotype-first,” where clinicians used the patient’s phenotype to guide decisions about which genetic tests to order, to “genotype-first,” where clinicians used the patient’s genotype, to guide their evaluation and management.

CMA results, expressed in specific genomic coordinates, also afforded clinicians with the ability to truly integrate information garnered from personal genotype into medical care. With more precise breakpoints for patients' CNVs, laboratories and clinicians can determine which genes are involved in a CNV. Using available information about the gene(s) phenotypic effects and dosage sensitivity, one can extrapolate which of the patient's presenting features could be explained by the CNV, as well as health issues that the patient could be at risk for in the future. Such health issues may be associated with specific management recommendations which could be implemented for a patient before they were even symptomatic.

For example, an individual could present to medical attention for developmental delay, and CMA results demonstrate a large deletion on chromosome 7q involving the *KCNH2* gene associated with long-QT syndrome 2 (OMIM 613688) [28]. Such results would indicate that the individual is at high risk for developing this disorder, which is associated with cardiac arrhythmia and sudden death. A referral to a cardiologist is then warranted for evaluation and management of this risk, something that was likely not expected in the context of the original presenting symptoms (i.e., an incidental finding). Examples such as this one are not infrequent, particularly given the number of potentially actionable genes with dosage sensitivity, such as cancer predisposition genes, which are covered as part of standard array designs. One study using data from the International Standards for Cytogenomic Arrays (ISCA) consortium database estimated that ~7 % of reported cases involved a region of the genome associated with some types of published medical management recommendations, demonstrating the clinical actionability of CMA test results [29].

Key Contributor to Complex Conditions

Additionally, CNVs have been identified as important contributors to complex conditions, such as ASD. Heritability estimates have varied for ASD, but have been reported as high as 90 % [30]. Identifying the genetic basis of these types of conditions through traditional methods (such as linkage analysis) has been challenging due to their extensive genetic and phenotypic heterogeneity. Linkage studies identified several SNPs associated with ASD, but all with relatively low effect size [30]. Likewise, genome-wide association studies (GWAS) undertaken to identify common risk variants have been largely unsuccessful. Although several common risk variants have been identified, the results have been difficult to replicate and were associated with odds ratios less than 1.23 [1, 32], making it clear that these methods were missing variants of moderate–high effect.

As these populations began to be evaluated for CNVs, the significant contribution of structural variation to ASD became

apparent. Many CNVs have been identified among individuals with ASD [33–35], and some recurrent CNVs (mediated by segmental duplications) have even reached statistical significance in large-scale case–control studies [34]. Although none of these CNVs individually account for more than 1 % of ASD cases, as a group, structural variants play a significant role in the development of this disorder, providing a genetic diagnosis in 5–10 % of cases. CNVs have been reported to confer more than 3 times the level of risk attributed to SNPs identified through GWAS [32, 36].

Similar observations have been made in other complex disorders, such as intellectual disability, schizophrenia, epilepsy, and cardiac defects (presented in this issue). Interestingly, many of the same CNVs are being identified among phenotypes which were previously considered distinct. These observations have led some to consider whether these phenotypes may actually represent aspects of an etiologically related continuum, such as the developmental brain dysfunction model described for neurodevelopmental disorders [21].

Conclusion

Identifying an underlying genetic etiology for any human phenotype is invaluable, both to the individual patient and to the research community as a whole. For the individual patient, receiving a specific genetic diagnosis can end the taxing diagnostic odyssey, contribute to current and future medical management, and impact family planning considerations. For the research community, molecularly defining previously uncharacterized disorders provides the opportunity to learn more about gene function, gene–gene interactions, and genotype–phenotype correlations, which will ultimately lead to targeted therapeutics.

The lessons learned from CMA and copy number variation have contributed greatly to both the ability to diagnose individuals and to the knowledge base surrounding the mechanisms of human disease. As whole-exome and whole-genome technologies become more accessible, the cycle of discovery and knowledge assimilation will continue to accelerate; new variants will be discovered, new mechanisms will be deduced, and our perception of clinical genomics will evolve. These types of genome-wide assays have already and will continue to identify genomic variants with immediately appreciable effects on human health and development. However, they will also continue to identify variants of uncertain clinical significance, as our ability to identify variation is currently beyond our ability to accurately interpret its possible phenotypic consequences.

Various databases designed to catalog and make publicly available structural variation data from both normal and affected populations (Database of Genomic Variation [37] for normal populations; the ISCA Consortium [9••],

DECIPHER [38], ECARUCA [39], etc. for affected populations) have been established. These databases continue to serve as valuable resources for laboratories and clinicians as they interpret CNVs observed in patients on a daily basis. As technologic advances move toward the ability to detect both structural- and sequence-level data from the same testing platform, similar resources are needed to make information about both types of variation readily accessible to both the research and clinical communities. ClinVar is a resource housed within the National Center for Biotechnology Information (NCBI) that collects information about the relationships between human variation (both structural and sequence level) and human disease. Large-scale efforts are underway through the International Collaboration for Clinical Genomics (ICCG) [40] to facilitate the submission of clinical laboratory data from CMA- and sequencing-based tests into this database, with plans to ultimately include whole-exome and whole-genome data. These efforts are part of a larger collaboration, the Clinical Genomics Resource (ClinGen), which aims to use this information, coupled with manual and machine learning-based curation efforts, to create a “clinical genome,” or catalog of variants known to be relevant to clinical care.

The scope of knowledge regarding the nature of structural variation and its relationship to human health has expanded dramatically over the last several decades; with increased usage of whole-genome and whole-exome sequencing, it is expected that the same will occur for sequence-level variation. Increased data sharing and collaboration will lead to substantial progress in understanding the relationships between variants and disease, making personalized genomic medicine a reality.

Acknowledgments This work was funded in part by NIH Grants RO1MH074090 (to DH Ledbetter and CL Martin) and U41HG006834 (to DH Ledbetter and CL Martin). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Disclosure ER Riggs, DH Ledbetter, and CL Martin all declare no conflicts of interest.

Human and Animal Rights and Informed Consent This article does not contain any studies with human or animal subjects performed by any of the authors.

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

References

Papers of particular interest, published recently, have been highlighted as

- Of importance
- Of major importance

1. Lander ES, Linton LM, Birren B, et al. Initial sequencing and analysis of the human genome. *Nature*. 2001;409(6822):860–921.
2. Fodor SP, Read JL, Pirrung MC, Stryer L, Lu AT, Solas D. Light-directed, spatially addressable parallel chemical synthesis. *Science*. 1991;251(4995):767–73.
3. Schena M, Shalon D, Davis RW, Brown PO. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science*. 1995;270(5235):467–70.
4. Solinas-Toldo S, Lampel S, Stilgenbauer S, et al. Matrix-based comparative genomic hybridization: biochips to screen for genomic imbalances. *Genes Chromosom Cancer*. 1997;20(4):399–407.
5. Kallioniemi A, Kallioniemi OP, Sudar D, et al. Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. *Science*. 1992;258(5083):818–21.
6. Pinkel D, Seagraves R, Sudar D, et al. High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nat Genet*. 1998;20(2):207–11.
7. Ylstra B, van den Ijssel P, Carvalho B, Brakenhoff RH, Meijer GA. BAC to the future! or oligonucleotides: a perspective for micro array comparative genomic hybridization (array CGH). *Nucleic Acids Res*. 2006;34(2):445–50.
8. Baldwin EL, Lee JY, Blake DM, et al. Enhanced detection of clinically relevant genomic imbalances using a targeted plus whole genome oligonucleotide microarray. *Genet med*. 2008;10(6):415–29.
9. •• 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(5):749–64. *The authors provide an evidence-based summary of the yield of clinical cytogenomic microarray testing as compared to the diagnostic yield of traditional G-banded karyotyping. This paper also discusses the technical advantages and limitations, and issues that affect test interpretation.*
10. Harrison CJ, Jack EM, Allen TD, Harris R. Investigation of human chromosome polymorphisms by scanning electron microscopy. *J Med Genet*. 1985;22(1):16–23.
11. Sebat J, Lakshmi B, Troge J, et al. Large-scale copy number polymorphism in the human genome. *Science*. 2004;305(5683):525–8.
12. Iafrate AJ, Feuk L, Rivera MN, et al. Detection of large-scale variation in the human genome. *Nat Genet*. 2004;36(9):949–51.
13. Pang AW, MacDonald JR, Pinto D, et al. Towards a comprehensive structural variation map of an individual human genome. *Genome Biol*. 2010;11(5):R52.
14. Bassett AS, McDonald-McGinn DM, Devriendt K, et al. Practical guidelines for managing patients with 22q11.2 deletion syndrome. *J pediatr*. 2011;159(2):332–9.
15. • Cooper GM, Coe BP, Girirajan S, et al. A copy number variation morbidity map of developmental delay. *Nat Genet*. 2011;43(9):838–46. *Case-control analysis of copy number variants identified amongst 15,767 individuals referred for clinical chromosomal microarray testing and 8,329 unaffected adult controls.*
16. Mefford HC, Cooper GM, Zerr T, et al. A method for rapid, targeted CNV genotyping identifies rare variants associated with neurocognitive disease. *Genome Res*. 2009;19(9):1579–85.
17. de Kovel CG, Trucks H, Helbig I, et al. Recurrent microdeletions at 15q11.2 and 16p13.11 predispose to idiopathic generalized epilepsies. *Brain*. 2010;133(Pt 1):23–32.
18. Rees E, Walters JT, Georgieva L, et al. Analysis of copy number variations at 15 schizophrenia-associated loci. *Br j psychiatry*. 2014;204(2):108–114.
19. Mefford HC, Muhle H, Ostertag P, et al. Genome-wide copy number variation in epilepsy: novel susceptibility loci in idiopathic generalized and focal epilepsies. *PLoS Genet*. 2010;6(5):e1000962.

20. •• Stefansson H, Meyer-Lindenberg A, Steinberg S, et al. CNVs conferring risk of autism or schizophrenia affect cognition in controls. *Nature*. 2013;505(7483):361–6. *The authors demonstrate that controls harboring copy number variants previously reported in affected populations may exhibit subtle differences in cognitive profile compared to controls without these CNVs, challenging previously held ideas of reduced penetrance.*
21. Moreno-De-Luca A, Myers SM, Challman TD, Moreno-De-Luca D, Evans DW, Ledbetter DH. Developmental brain dysfunction: revival and expansion of old concepts based on new genetic evidence. *Lancet Neurol*. 2013;12(4):406–14.
22. Shaffer LG, Theisen A, Bejjani BA, et al. The discovery of microdeletion syndromes in the post-genomic era: review of the methodology and characterization of a new 1q41q42 microdeletion syndrome. *Genet med*. 2007;9(9):607–16.
23. Koolen DA, Vissers LE, Pfundt R, et al. A new chromosome 17q21.31 microdeletion syndrome associated with a common inversion polymorphism. *Nat Genet*. 2006;38(9):999–1001.
24. Sharp AJ, Hansen S, Selzer RR, et al. Discovery of previously unidentified genomic disorders from the duplication architecture of the human genome. *Nat Genet*. 2006;38(9):1038–42.
25. Klopocki E, Graul-Neumann LM, Grieben U, et al. A further case of the recurrent 15q24 microdeletion syndrome, detected by array CGH. *Eur J Pediatr*. 2008;167(8):903–8.
26. Ballif BC, Hornor SA, Jenkins E, et al. Discovery of a previously unrecognized microdeletion syndrome of 16p11.2–p12.2. *Nat Genet*. 2007;39(9):1071–3.
27. • Manning M, Hudgins L. Array-based technology and recommendations for utilization in medical genetics practice for detection of chromosomal abnormalities. *Genet med*. 2010;12(11):742–5. *American College of Medical Genetics and Genomics (ACMG) practice guideline officially endorsing cytogenomic microarray as a first-tier test for individuals with developmental delay, autism spectrum disorders, and/or other congenital anomalies.*
28. Ledbetter DH, Riggs E, Martin CL. Clinical applications of whole-genome chromosomal microarray analysis. In: Ginsburg GS, Willard HF, editors. *Genomic and Personalized Medicine*, vol. 1. 2nd ed. Waltham: Elsevier Academic Press; 2012. p. 133–44.
29. Riggs E, Wain K, Riethmaier D, et al. Chromosomal microarray impacts clinical management. *Clin Genet*. 2013;85(2):147–53.
30. Malhotra D, Sebat J. CNVs: harbingers of a rare variant revolution in psychiatric genetics. *Cell*. 2012;148(6):1223–41.
31. Buxbaum JD, Daly MJ, Devlin B, Lehner T, Roeder K, State MW. The autism sequencing consortium: large-scale, high-throughput sequencing in autism spectrum disorders. *Neuron*. 2012;76(6):1052–6.
32. Devlin B, Melhem N, Roeder K. Do common variants play a role in risk for autism? Evidence and theoretical musings. *Brain Res*. 2011;1380:78–84.
33. Pinto D, Pagnamenta AT, Klei L, et al. Functional impact of global rare copy number variation in autism spectrum disorders. *Nature*. 2010;466(7304):368–72.
34. Sanders SJ, Ercan-Sencicek AG, Hus V, et al. Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. *Neuron*. 2011;70(5):863–85.
35. Glessner JT, Wang K, Cai G, et al. Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. *Nature*. 2009;459(7246):569–73.
36. Moreno-De-Luca D, Sanders SJ, Willsey AJ, et al. Using large clinical data sets to infer pathogenicity for rare copy number variants in autism cohorts. *Mol psychiatry*. 2013;18(10):1090–5.
37. Macdonald JR, Ziman R, Yuen RK, Feuk L, Scherer SW. The database of genomic variants: a curated collection of structural variation in the human genome. *Nucleic acids research*. 2013;42(D1):D986–92.
38. Firth HV, Richards SM, Bevan AP, et al. DECIPHER: Database of chromosomal imbalance and phenotype in humans using ensembl resources. *Am J Hum Genet*. 2009;84(4):524–33.
39. Feenstra I, Fang J, Koolen DA, et al. European cytogeneticists association register of unbalanced chromosome aberrations (E-CARUCA); an online database for rare chromosome abnormalities. *Eur J Med Genet*. 2006;49(4):279–91.
40. Riggs ER, Wain KE, Riethmaier D, et al. Towards a universal clinical genomics database: the 2012 international standards for cytogenomic arrays consortium meeting. *Hum Mutat*. 2013;34(6):915–9.