Copy Number Variants and the Genetic Enigma of Congenital Heart Disease

Ali J. Marian

For years, many reasoned, perhaps not so naïvely, that genetics of congenital heart diseases (CHD) would be the last frontier in efforts to elucidate the genetic causes of cardiovascular disease. This impression was based on 2 main reasons. First, CHD seldom exhibits a clear familial inheritance pattern, as opposed to many single gene disorders. This is despite the strong evidence for familial aggregation of CHD and a higher risk of recurrence in the offspring, which denote a clear genetic cause.1,2 CHD are often sporadic or part of the aneuploidy syndromes with pleomorphic noncardiac phenotypes.1,3 Thus, an approach distinct from the genetic linkage analysis in large families, which was commonly applied to delineate the genetic basis of hereditary cardiovascular diseases such as cardiomyopathies and arrhythmia syndromes,4–8 was needed to define the genetic cause of CHD. Second, the extreme phenotypic assortment of CHD,3,9–11 which is far beyond the scope of the disease—specific gene disorders, as well as the phenotypic variability of single gene disorders, as well as the diversity of common complex diseases, such as coronary artery disease and systemic arterial hypertension. Consequently, it was difficult envisioning how apparently the simple phenotype of atrial septal defects and the complex phenotype of tetralogy of Fallot, which share no anatomic and physiological similarities, would causally share a common class of genetic networks, let alone arise from mutations in a single gene.

Recent discoveries are transforming the landscape of molecular genetics of CHD and diluting the aforementioned antediluvian impression. The pioneering work of Basson et al12 and Schott et al13 in late 1990s led to the identification of loss-of-function mutations in TBX5 and NKX2-5 as causes of Holt–Oran syndrome and atrial septal defect, respectively. An intriguing finding was the assortment of clinical phenotypes in the mutation carriers of NKX2-5, ranging from atrial septal defect to tetralogy of Fallot and hypoplastic left heart syndrome, often in conjunction with conduction defects, among others.13 During the course of the next several years, small-scale studies led to the identification of about 3-dozen genes, encoding transcription factors, cell signaling molecules, and structural proteins in patients with CHD (Fahed et al,3 Table 3). Moreover, a genome-wide association study comprised 1995 cases with a variety of CHD and 5159 controls was conducted with the goal of identifying susceptibility loci for atrial septal defect.14 Despite these discoveries, the genetic causes of CHD in ≈80% of the patients had remained unknown.15 The convergence of 4 sets of advances, indicated below, is gradually changing the landscape and accelerating the pace of new discoveries of genetic causes of CHD.

1. Initial genetic discoveries by pointing to mutations in cardiac transcription factors as causes of CHD provided the framework for subsequent genetic studies.12,13,15
2. Delineation of the regulatory genetic networks involved in cardiac development offered a biological context for the genetic discoveries.16–21
3. Availability of large repositories of patients with CHD, such as the Pediatric Cardiac Genomics Consortium (PCGC), afforded large-scale genetic studies.22–24
4. Advent of newer genetic technologies enabled defining the transcriptional state of the genome, as well as genome-wide sequence variations, including single nucleotide variants and copy number variants (CNVs).25–30

Among the first fruits of these advances was a whole exome sequencing project that involved 362 parent–offspring trios with severe CHD in the PCGC population.11 It led to the identification of premature truncation, frameshift, and splice site de novo mutations in 28 gene encoding histone-modifying proteins.31 The de novo variants collectively contributed to ≈10% of severe cases of CHD in the PCGC population. The findings not only broadened the spectrum of the disease-causing mechanisms to include the epigenetic machinery but also advocated for the gene dosage mechanism, both increased and reduced gene expression, in the pathogenesis of CHD.

In a recent issue of Circulation Research, Glessner et al12 report another large-scale whole-genome study designed to delineate the causal role of CNVs in CHD in the PCGC population. CNVs are structural genomic variations, typically larger than 1000 base pairs that through duplication or deletion lead to gain or loss of chromosomal segments that often contain multiple contiguous genes. The results of the study by Glessner et al12 are notable for a 4-fold increase in the frequency of all de novo CNVs and a 2-fold increase in the frequency of novel de novo CNVs in trios with CHD when compared with controls.32 Approximately 10% of the CHD population had rare de novo CNVs, resulting from deletions or duplication, the former being more common. Combining the sequencing data and CNVs, the authors identified ETS1, encoding the ETS1 transcription factor,33 and CTBP2, which codes for a transcriptional corepressor,34 as the likely pathogenic genes affected in the 11q24.2-q25 (Jacobsen syndrome) and 10q subtelomeric
deletions, respectively.\textsuperscript{32} The findings are in accord with the prevailing gene dosage mechanism and the pathogenic role of CNVs, particularly de novo CNVs, in CHD.\textsuperscript{3,9,35,36}

The study by Glessner et al\textsuperscript{32} benefits from a robust family-based trio study design composed of probands, who did not have the known cytogenetic anomalies and pathogenic CNVs. It also uses state-of-the-art technology that includes detection of CNVs by 2 independent and complementary methods of single-nucleotide polymorphism arrays and whole exome sequencing in a subset of 233 trios and validation of the CNVs by digital droplet polymerase chain reaction. The use of 2 independent CNVs detection platforms also afforded the opportunity to compare detection sensitivity of each platform, which was calculated to be ≈65% to 70% (30%–35% false-negative rate), on the condition of ≥10 adjacent single-nucleotide polymorphisms for calling CNVs by the single-nucleotide polymorphism arrays and involvement of ≥3 adjacent exons in the whole exome sequencing approach. This finding suggests considerable underdetection of the CNVs if only one of the detection methods (single-nucleotide polymorphism arrays and whole exome sequencing) is used. It also has direct implications in the design of future studies.

The findings of the study by Glessner et al\textsuperscript{32} and the existing data identify CNVs as important causes of CHD and imply that CNVs are likely to contribute to a larger fraction of CHD that has been demonstrated to date. Several CNVs, identified by Glessner et al\textsuperscript{32} affected only a single gene, hence, rendering them potentially causal genes. However, the pathogenic role of the individual CNVs identified in the study by Glessner et al,\textsuperscript{32} with the exception of a few whose causality in CHD already has been established, such as CNVs affecting \textit{NKX2-5} and \textit{GATA4}, cannot be ascertained and will require additional experimentation. The majority of identified CNVs affected large chromosomal segments involving up to several million base pairs of DNA and multiple contiguous genes, rendering the identification of the specific causal gene more tedious. It also merits noting that singleton de novo CNVs identified by Glessner et al,\textsuperscript{32} as well as those identified previously, cannot be considered causal pending replication of the findings in independent populations and validation through experimentation, as clearly stated by the authors.

The collective results of 2 large-scale studies on the PCGC population support the gene dosage hypothesis in the pathogenesis of a subset of CHD because mutations in genes encoding histone-modifying proteins and the CNVs are expected to change expression levels of the affected proteins. As for the pleiotropic phenotypic expression of CHD that also includes the extracardiac phenotypes, one could speculate several plausible explanations that might operate in isolation or cooperatively as follows:

1. Altered expression levels of multiple proteins resulting from CNVs affecting multiple contiguous genes.
2. Altered regulatory elements in the nonprotein coding regions, including noncoding RNAs and enhancers, affected by the CNVs.
3. Altered expression of multiple gene targets of the mutant transcription factors.
4. Key position of the mutant protein in the regulatory networks, resulting in altered biological functions of multiple target proteins.
5. Multiple hit (digenic or multigeneic) hypothesis, with 1 mutation serving as the variant with the largest effect size (causal) and multiple others exerting a gradient of effect sizes (modifiers).
6. Concomitant epigenetic variants, modified partly by the environmental factors, exerting additional changes (modifiers) in the background of the main causal mutation.

Figure. Genetic cause of congenital heart disease (CHD): The glass is half empty.\textsuperscript{3} The known genetic cause of CHD, namely chromosomal aneuploidy, rare familial CHD with Mendelian patterns of inheritance, mutations affecting proteins involved in histone modifiers (epigenetics), and copy number variants (CNVs) contribute to minority of the CHD cases. The genetic cause of CHD in the majority of the cases is unknown.

Recent discoveries have hastened elucidation of the molecular genetic basis of CHD. Despite these advances, however, genetic cause of CHD has remained inadequately defined, eloquently described by the leading scientists in the field as the glass half empty\textsuperscript{3} (Figure). To shift the paradigm, elaborate studies would be required to define the genetic cause of CHD further, elucidate the underpinning mechanisms, and delineate the molecular basis of phenotypic plasticity. Perhaps, not so naïvely, many still consider elucidation of the molecular genetic basis of CHD as the last frontier in human cardiovascular genetics.

Acknowledgments

We acknowledge the funding support from National Institutes of Health (NIH)/NIH Heart, Lung and Blood Institute (R01 HL088498 and R34 HL105563), Leducq Foundation, Roderick MacDonald Foundation (13RDM005), TexGen Fund from Greater Houston Community Foundation and George and Mary Josephine Hanman Foundation.

Disclosures

None.
References


Keywords: Editorsials ■ copy number variants, DNA ■ genetics ■ mutation
Copy Number Variants and the Genetic Enigma of Congenital Heart Disease
Ali J. Marian

_Circ Res._ 2014;115:821-823
doi: 10.1161/CIRCRESAHA.114.305243
_Circulation Research_ is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2014 American Heart Association, Inc. All rights reserved.
Print ISSN: 0009-7330. Online ISSN: 1524-4571

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://circres.ahajournals.org/content/115/10/821

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in _Circulation Research_ can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to _Circulation Research_ is online at:
http://circres.ahajournals.org/subscriptions/