Modifying Mendel Redux: Unbiased Approaches Can Find Modifiers

Kim L. McBride, MD[1] and **Stephanie M. Ware, MD PhD**[2]

[1]Center for Cardiovascular Research, Nationwide Children's Hospital, Department of Pediatrics, College of Medicine, Ohio State University, Columbus, OH

[2]Departments of Pediatrics and Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, IN

Journal Subject Terms

Basic Science Research; Clinical Studies; Mechanisms; Genetic; Association Studies; Genetics

Keywords

editorial; congenital cardiac defect; genome-wide association scan; genomics

Congenital heart disease (CHD) is one of the most common groups of birth defects[1], contributing to a major portion of mortality in early childhood and consuming large amounts of healthcare and family resources. They have a birth prevalence of 6–8/1000 live births, excluding late recognized defects such as BAV which has a population frequency of 1–2%[1].

Genetic epidemiology studies and reports of multiple recurrences of CHDs within families demonstrate a strong genetic component. Familial clustering of CHDs is particularly apparent when grouped by developmental mechanism[2]. A recent large study using hundreds of these multiplex families confirmed the concept of grouping CHD by developmental mechanism, and supporting animal data suggests these groupings are due to perturbations of genetic networks important in cardiogenesis[3]. Indeed, of all risk factors for CHD, a family history of CHD has the highest relative risk even over maternal diabetes or twinning[4]. More formal segregation analyses have confirmed the strong genetic component, also noting the inheritance pattern is likely complex and oligogenic[2].

This evidence for the genetic basis of CHD spurred investigators to search for responsible loci and genes. A few early successes occurred using the traditional genetic approach of linkage, identifying pathogenic variants in *NKX2-5*[5], *NOTCH1*[6], and *GATA4*[7] among multiplex families with CHDs. Unfortunately, further successes have been scarce, with the exception of CHDs occurring as part of a syndrome (such as the RASopathies). Genome-wide association studies (GWAS) for specific groups of CHD have added a few more loci,

Correspondence: Kim L. McBride, MD, Nationwide Children's Hospital, 700 Children's Dr, Columbus, OH 43205, Tel: 614 355 5734, Fax: 614 355 5725, Kim.McBride@NationwideChildrens.org.

but with limited replications in a second study[8, 9]. Copy number variant (CNV) studies have identified novel genomic disorders in as many as 20% of syndromic cases, a few percent of nonsyndromic CHDs, including identification of individuals with multiple CNVs[10–12]. The advent of next-generation sequencing technology raised hopes that this bottleneck in gene discovery would be resolved. Early studies appeared exciting, with the identification of chromatin remodeling genes in CHD cases[13]. Unfortunately it appears at least some individuals may have had unrecognized or undefined syndromes that include CHDs, rather than an isolated CHD. Subsequent studies have found de novo changes in only 1–2% of those cases[14], thus the cause remains elusive for isolated CHD. These studies also highlight the significant difficulty in case ascertainment, the need for careful (and longitudinal) phenotyping, and the challenges of variant interpretation.

While the causes of syndromes that include CHD as part of the phenotype have been answered more successfully, an interesting question has arisen: Why does only a portion of individuals with the syndrome have a CHD? Conversely, individuals with specific syndromes offer a unique opportunity to study a population with susceptibility to CHDs. This topic was highlighted in our previous editorial on Modifying Mendel[15] in which Li et al. used a trisomy 21 (Down syndrome) mouse (Ts65dn) in an attempt to answer the question of CHD risk using this sensitized model[16]. Their work implicated *Creld1* and *Gata4* in mouse, supported by data from human T21 subjects, providing one of the first studies on modification of a major susceptibility locus (in this case an extra chromosome 21) in the causation of CHD.

The article by Guo et al.[17] in this issue is among the first to use an unbiased approach to identifying a modifier locus for CHD in humans. Using GWAS in individuals with a deletion of 22q11.2 (22q11DS), the authors combined bioinformatics of the human locus with additional studies in mice to narrow down to a potential candidate gene. 22q11DS is the most common genomic disorder that causes CHD. Individuals may have characteristic dysmorphic features, velopalatal abnormalities (cleft or insufficiency), hypocalcemia, immunodeficiency, and a variety of neurodevelopmental disorders. Most individuals have CHD related to second heart field developmental anomalies, including truncus arteriosus, interrupted aortic arch type B, and most commonly, tetralogy of Fallot (TOF).

Dr. Morrow's group was able to assemble an impressive cohort of nearly 1500 carefully phenotyped individuals with 22q11DS. Over 1200 samples were SNP genotyped on an Affymetrix 6.0 platform. The analysis was careful and thorough, including proper quality control and accounting for possible biases from population stratification. Comparing 22q11DS subjects with TOF (n=326) to 22q11DS subjects with normal cardiac anatomy (n=566) in the cohort, they identified a locus on 5q14.3 with multiple SNPs (genotyped and imputed) exceeding the threshold for genome-wide significance. The highest associated SNP was rs12519770 ($p=2.98\times10^{-8}$) with an odds ratio of 1.69. This locus was further narrowed to a ~100kb interval within an intron of *GPRC78* using additional whole genome sequencing of just under 400 22q11DS subjects. Bioinformatic analysis of chromatin confirmation (Hi-C) showed that this region lay in a topological associated domain of 2.3 Mb size, containing six genes: *GPR98, MEF2C, CETN3, MBLAC2, POLR3G*, and *LYSMD3*. Expression studies in mouse were performed using in-situ hybridization, and only *Mef2c* demonstrated

expression in cardiac progenitor cells. *MEF2C* is a very biologically plausible candidate gene, as it is an important transcription factor during cardiogenesis, acting in a network that regulates second heart field development.

Several weaknesses in this study are important to note, despite the huge effort by this group. Relatively speaking, this study is small by GWAS standards, thus it was powered to only be able to identify loci associated with large effect sizes. They did not have a second, replication cohort; however, assembling this size of initial cohort at all is a striking accomplishment. Finally, although the authors performed appropriate analyses to control for population stratification, inclusion of a variety of populations (European, Hispanic, African American, Asian) does create some concern for false positive results. Further work will also be required to identify whether the SNPs in the intron of *GPR78* are themselves functional or tag an as yet unidentified variant, and to determine the mechanistic basis by which the functional variants modify or cause CHDs in 22q11DS. Whether this study data can be generalized to other second heart field malformations not associated with 22q11DS will be interesting to see. While further validation of these data will be important, the results provide initial data on a potential modifier that causes CHD in 22q11DS individuals.

The previous study using T21 as a susceptibility group and known mouse CHD genes provided proof of principle that variant burden threshold influences disease penetrance. This current study by Dr. Morrow's group advances the field in an important way by demonstrating that an unbiased approach in a human susceptibility population is able to identify a modifier locus. Other syndromes with CHD as part of the phenotype, such as Noonan syndrome and other RASopathies, are additional candidates for this type of study.

Identifying major susceptibility loci, let alone modifiers, remains a challenge for isolated CHDs. Here, several approaches are promising. First, multiplex families are also a susceptibility population. Rates of identifying major susceptibility loci are much higher, with identification of pathogenic variants in 25–30% of families, compared to 1–2% of sporadic cases[14, 18]. Genome approaches in multiplex families also offer an unbiased screen for finding multiple loci. Second, brute force animal models, such as using mouse forward genetic screens (inducing mutation followed by high throughput phenotypic screening), can find not only major susceptibility loci but also digenic inheritance or modifier loci. This has already been successfully accomplished by Dr. Cecilia Lo's group, who found two genes causing hypoplastic left heart syndrome (HLHS) in a forward mouse screen followed by identification of variants in these genes in human subjects with HLHS[19].

Many difficulties lie ahead. Phenotyping remains a problem not only for anatomic features of a specific individual, but also for the correct grouping of distinct, but related, malformations. Statistical modeling has yet to be successful given the small numbers of subjects with CHDs that can be collected. Even large groups, such as the Pediatric Cardiovascular Genetics Consortium, would be unable to collect the thousands of samples necessary to identify significant results in rare variant analyses by current statistical methods. Methods for multiplex families for narrowing the list of potential variants in oligogenic inheritance models are lacking. In addition, studying multiple variants in multiple genes in animal models is slow, laborious and expensive. Newer techniques, such as use of

CRISPR/Cas9 to more quickly create animal models, and use of fruit fly to rapidly and cheaply study variants in multiple loci via multiple crosses, may offer some hope[20].

The genetics of CHD appears to be more complex than previously thought. The field will need to move away from simple Mendelian genetic models to those with multiple loci, possibly a major locus with one or more modifiers, or multiple loci acting together. This concept is not unique to the CHD genetics field. Other cardiovascular genetic diseases have well established complex patterns of inheritance that are already altering how patients with these diseases are diagnosed and managed in clinical practice. Arrhythmogenic right ventricular cardiomyopathy, where up to 10% of individuals have two disease causing variants, and long QT syndrome where two pathogenic variants lead to more severe phenotype that includes deafness, are two examples.

Dr. Morrow's group shows that modifiers of Mendelian CHD traits can be found. These methods should become part of future study designs in CHD genetics.

References

1. Hoffman JI, Kaplan S, Liberthson RR. Prevalence of congenital heart disease. American heart journal. 2004; 147:425–39. [PubMed: 14999190]

2. McBride KL, Pignatelli R, Lewin M, Ho T, Fernbach S, Menesses A, et al. Inheritance analysis of congenital left ventricular outflow tract obstruction malformations: Segregation, multiplex relative risk, and heritability. American journal of medical genetics Part A. 2005; 134A:180–6. [PubMed: 15690347]

3. Ellesøe SG, Workman CT, Bouvagnet P, Loffredo CA, McBride KL, Hinton RB, et al. Familial co-occurrence of congenital heart defects follows distinct patterns. European Heart Journal. 2017:ehx314.

4. Liu S, Joseph KS, Lisonkova S, Rouleau J, Van den Hof M, Sauve R, Kramer MS, Canadian Perinatal Surveillance S. Association between maternal chronic conditions and congenital heart defects: a population-based cohort study. Circulation. 2013; 128:583–9. [PubMed: 23812182]

5. Schott JJ, Benson DW, Basson CT, Pease W, Silberbach GM, Moak JP, et al. Congenital heart disease caused by mutations in the transcription factor NKX2-5. Science. 1998; 281:108–11. [PubMed: 9651244]

6. Garg V, Muth AN, Ransom JF, Schluterman MK, Barnes R, King IN, et al. Mutations in NOTCH1 cause aortic valve disease. Nature. 2005; 437:270–4. [PubMed: 16025100]

7. Garg V, Kathiriya IS, Barnes R, Schluterman MK, King IN, Butler CA, Rothrock CR, Eapen RS, Hirayama-Yamada K, Joo K, Matsuoka R, Cohen JC, Srivastava D. GATA4 mutations cause human congenital heart defects and reveal an interaction with TBX5. Nature. 2003; 424:443–7. [PubMed: 12845333]

8. Cordell HJ, Topf A, Mamasoula C, Postma AV, Bentham J, Zelenika D, et al. Genome-wide association study identifies loci on 12q24 and 13q32 associated with tetralogy of Fallot. Hum Mol Genet. 2013; 22:1473–81. [PubMed: 23297363]

9. Hanchard NA, Swaminathan S, Bucasas K, Furthner D, Fernbach S, Azamian MS, et al. A Genome Wide Association Study of Congenital Cardiovascular Left-Sided Lesions Shows Association with a Locus on Chromosome 20. Hum Mol Genet. 2016

10. Greenway SC, Pereira AC, Lin JC, DePalma SR, Israel SJ, Mesquita SM, et al. De novo copy number variants identify new genes and loci in isolated sporadic tetralogy of Fallot. Nat Genet. 2009; 41:931–5. [PubMed: 19597493]

11. Lander J, Ware SM. Copy Number Variation in Congenital Heart Defects. Current Genetic Medicine Reports. 2014; 2:168–178.

12. Hanchard NA, Umana LA, D'Alessandro L, Azamian M, Poopola M, Morris SA, et al. Assessment of large copy number variants in patients with apparently isolated congenital left-sided cardiac

lesions reveals clinically relevant genomic events. American journal of medical genetics Part A. 2017; 173:2176–2188. [PubMed: 28653806]

13. Zaidi S, Choi M, Wakimoto H, Ma L, Jiang J, Overton JD, et al. De novo mutations in histone-modifying genes in congenital heart disease. Nature. 2013; 498:220–3. [PubMed: 23665959]

14. Sifrim A, Hitz MP, Wilsdon A, Breckpot J, Turki SH, Thienpont B, et al. Distinct genetic architectures for syndromic and nonsyndromic congenital heart defects identified by exome sequencing. Nat Genet. 2016; 48:1060–5. [PubMed: 27479907]

15. McBride KL, Ware SM. Modifying Mendel: approaches for identification of susceptibility alleles for human cardiovascular malformations. Circulation Cardiovascular genetics. 2012; 5:274–6. [PubMed: 22715278]

16. Li H, Cherry S, Klinedinst D, DeLeon V, Redig J, Reshey B, Chin MT, Sherman SL, Maslen CL, Reeves RH. Genetic modifiers predisposing to congenital heart disease in the sensitized Down syndrome population. Circulation Cardiovascular genetics. 2012; 5:301–8. [PubMed: 22523272]

17. Guo T, Repetto GM, McDonald McGinn DM, Chung JH, Nomaru H, Campbell CL. Genome-wide Association Study to Find Modifiers for Tetralogy of Fallot in the 22q11.2 Deletion Syndrome Identifies Variants in the GPR98 locus on 5q14.3. Circulation Cardiovascular genetics. 2017; 10:e001690. [PubMed: 29025761]

18. LaHaye S, Corsmeier D, Basu M, Bowman JL, Fitzgerald-Butt S, Zender G, et al. Utilization of Whole Exome Sequencing to Identify Causative Mutations in Familial Congenital Heart Disease. Circulation Cardiovascular genetics. 2016; 9:320–9. [PubMed: 27418595]

19. Liu X, Yagi H, Saeed S, Bais AS, Gabriel GC, Chen Z, et al. The complex genetics of hypoplastic left heart syndrome. Nat Genet. 2017; 49:1152–1159. [PubMed: 28530678]

20. Zhu JY, Fu Y, Nettleton M, Richman A, Han Z. High throughput in vivo functional validation of candidate congenital heart disease genes in Drosophila. Elife. 2017; 6