Review

Focused Strategies for Defining the Genetic Architecture of Congenital Heart Defects

Lisa J. Martin [1,2,*] and D. Woodrow Benson [3]

1 Division of Human Genetics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH 45229, USA
2 Department of Pediatrics, University of Cincinnati School of Medicine, Cincinnati, OH 45229, USA
3 Department of Pediatrics, Medical College of Wisconsin, Wauwatosa, WI 53226, USA; dbenson@mcw.edu
* Correspondence: lisa.martin@cchmc.org; Tel.: +1-513-636-1244

Abstract: Congenital heart defects (CHD) are malformations present at birth that occur during heart development. Increasing evidence supports a genetic origin of CHD, but in the process important challenges have been identified. This review begins with information about CHD and the importance of detailed phenotyping of study subjects. To facilitate appropriate genetic study design, we review DNA structure, genetic variation in the human genome and tools to identify the genetic variation of interest. Analytic approaches powered for both common and rare variants are assessed. While the ideal outcome of genetic studies is to identify variants that have a causal role, a more realistic goal for genetic analytics is to identify variants in specific genes that influence the occurrence of a phenotype and which provide keys to open biologic doors that inform how the genetic variants modulate heart development. It has never been truer that good genetic studies start with good planning. Continued progress in unraveling the genetic underpinnings of CHD will require multidisciplinary collaboration between geneticists, quantitative scientists, clinicians, and developmental biologists.

Keywords: cardiovascular malformations; genetic variation; association; etiology; phenotyping

1. Introduction

The objective of this review is to provide an overview of strategies utilized in pursuit of defining the genetic origins of rare birth defects characterized by malformation of the heart, so-called congenital heart defects (CHD). CHD occurs during heart development, also known as cardiogenesis, which begins very early in gestation. The initial beating embryonic heart is tubular-shaped, and all the while functioning to support somatic growth of the fetus, the embryonic heart loops and is morphed to a complex four-chambered organ [1]. Identification of the regulatory networks controlling all stages of cardiogenesis has led to improved understanding of genes involved in heart development [2]. Conversely, complimentary, genetic studies of CHD patients have identified variants in genes essential to heart development [3]. We hope that review of focused strategies will facilitate studies that improve our understanding of the genetic architecture of human CHD.

2. Phenotypic Considerations

2.1. What Is CHD?

CHD refer to malformations of the heart present at birth. We will not consider malformations incompatible with fetal life but will limit our focus on CHD compatible with fetal development resulting in live birth. The anatomic details and clinical significance of the malformations are quite varied. Some may have a profound impact on postnatal clinical well-being and require intervention, usually surgical and often in the neonatal period. On the other hand, some CHD may have no clinical impact and go undetected until discovered incidentally later in life. While such defects may be of little clinical significance, they are highly relevant to the design of genetic studies. This is because statistical evaluation of

the co-occurrence between genotype and phenotype sought in a genetic study requires determination of the presence or absence of CHD.

2.2. What Is the Incidence of CHD?

CHD constitute a major portion of clinically significant birth defects. While an incidence of 10 per 1000 (~1%) is often cited [4], depending on the definition of what constitutes CHD, the incidence may be much higher. A history of clinically significant CHD, such as those requiring surgery, may be readily apparent during review of past medical history. However, some CHD may have been overlooked because they were not clinically significant; they cause no symptoms and may only be detected by a cardiac imaging study. Examples of such CHD include isolated aneurysm of the atrial septum, persistent left superior vena cava (LSVC), right aortic arch and bicuspid aortic valve (BAV). BAV, the most common congenital cardiac malformation, has an incidence of 10 to 20 per 1000 of the population [5]; BAV and other malformations of little apparent clinical significance are often excluded from estimates of CHD incidence. Taken together, CHD incidence may be as high as 50 per 1000 (~5%) [6]. This is an important consideration in genetic study design because even though CHD in a research participant lacks clinical significance it may be an indication of genetic abnormality and thereby be of profound genetic significance.

2.3. Tools to Determine CHD Phenotype

Classification of CHD phenotype is based on careful consideration of images of the heart position in the thorax, heart chambers, septa and valves as well as location, anatomy and relationships of vena cava, pulmonary veins and great arteries. A chest X-ray may be useful for determining the position of the heart in the thorax. Angiography, an invasive procedure, may provide useful images of abnormal cardiac anatomy. However, an echocardiogram which uses ultrasound technology has become the gold standard technique for clinical cardiac imaging as it is well adapted for cardiac imaging for patients of all ages including the fetus.

Details of pre- and postnatal medical history, family history and clinical exam may inform about the presence or absence of CHD; in genetic discovery studies, identification of the absence of CHD is as important as discovering the presence of CHD. Individuals with CHD may have a history of cardiac surgery, previous visits to a cardiologist, and/or records of past cardiac imaging, e.g., echocardiography, that reveal the CHD phenotype. However, the absence of such records does not equate to a phenotype of normal. Extracardiac features relevant to a CHD diagnosis include facial features, skeletal abnormalities including malformed vertebrae and limb anomalies, and abdominal viscera arrangement. For example, from the perspective of an echocardiographer, tetralogy of Fallot due to del22q11, also known as DiGeorge syndrome, may be indistinguishable from that CHD due to the NKX2.5 mutation, whereas in the same scenario a history of cleft palate would drastically alter the genetic focus [3]. A developmental assessment including gross and fine motor skills as well as cognitive development may lead to recognition of developmental delay which is more likely to be associated with certain CHD as part of a syndrome.

Family history can distinguish genetic conditions that are not usually inherited, e.g., Down syndrome (trisomy 21), from genetic conditions that exhibit familial clustering, e.g., BAV. The recognition of familial heart disease has been complicated by several genetic phenomena (Table 1) that obscure the familial nature [7]. Further, while most individuals believe family history is important, many are unfamiliar with important, relevant clinical details of familial CHD. Too often, in the hustle and bustle of a busy clinic, family history is asked on the initial visit, recorded and never revisited. This leads to a situation whereby family history is an under-utilized tool in the recognition of genetic etiology. Family history is dynamic, and a current account may require revisiting the questions on more than one occasion and obtaining information from more than one family member. A pedigree is a shorthand way to document and record family history and may give some indication as

to the mode of inheritance. However, with electronic medical records, pedigrees may be attached as images that can be reviewed manually and updated as necessary.

Table 1. Definition of genetic phenomena.

Phenomenon	Attribute
Genetic heterogeneity	Similar phenotypes, different genetic cause.
Variable expressivity	Individuals with same disease gene have different phenotypes.
Reduced penetrance	Disease absence in some individuals with disease gene.
Pleiotropy	Multiple phenotypes associated with the same genetic cause.

A genetic condition may be identified by recognizing signature cardiac and/or non-cardiac findings during evaluation. For example, tetralogy of Fallot is a signature cardiac malformation for 22q11 deletion syndrome (del22q11), but a physician evaluating a patient with right ventricular outflow tract malformation may overlook dysmorphic facial features characteristic of del22q11. The presence of syndromic features is strongly supportive of a genetic condition and may be an indication for genetic testing. Even with what appears to be isolated CHD, typical features of the cardiac phenotype may suggest a genetic etiology with known inheritance.

3. Genetic Considerations

3.1. What Is DNA?

Deoxyribonucleic acid (DNA) carries the information necessary to create and organize all cells and organs in the human body (Figure 1). DNA is a two-stranded molecule with a backbone made of sugar and phosphate. Attached to each sugar is one of four bases known as nucleotides (adenine, cytosine, guanine, and thymine). The strands are held together by bonds between the bases. The base sequences carry the information not only on how to make proteins but also when to make proteins. An individual's DNA carries complete information from each parent, and the paired set of information is known as the genome.

In humans, DNA is organized into 23 chromosomes which harbor three billion base pairs (bp). Genes are the DNA units that make proteins; the human genome has ~30,000 genes. Only ~1% of the genome codes for protein. The DNA sequences encoding these genes are known as exons (Figure 1). Most genes have multiple exons, with the interspersing region known as introns. Following transcription and RNA splicing, introns are removed, and exons join to form a contiguous coding sequence.

The sequences near an intron-exon boundary are critical for appropriate splicing to occur. While early work on the genome focused on the exons or protein coding regions, it is well recognized that DNA of the non-protein coding regions plays an important role in phenotypic variation among individuals [8,9]. For example, the regions upstream of the gene include the promotor sequence which contain the transcription initiation site which determines whether the DNA is copied to messenger RNA (the first step in creating a protein). Additionally, upstream and downstream elements may influence transcription [10]. Notably, except when copying itself, DNA is stored in a compact form wound around nucleosomes. As such, some regulatory elements may be long distances from the gene of interest [11]. While much of the focus on CHD genetics has been on variants which affect protein structure, work such as ENCODE [12] has demonstrated the importance of the non-protein coding sequences as well. Indeed, a recent paper by Gelb and colleagues demonstrated enrichment of de novo (not inherited from parents) variants in non-coding regions in CHD patients [13].
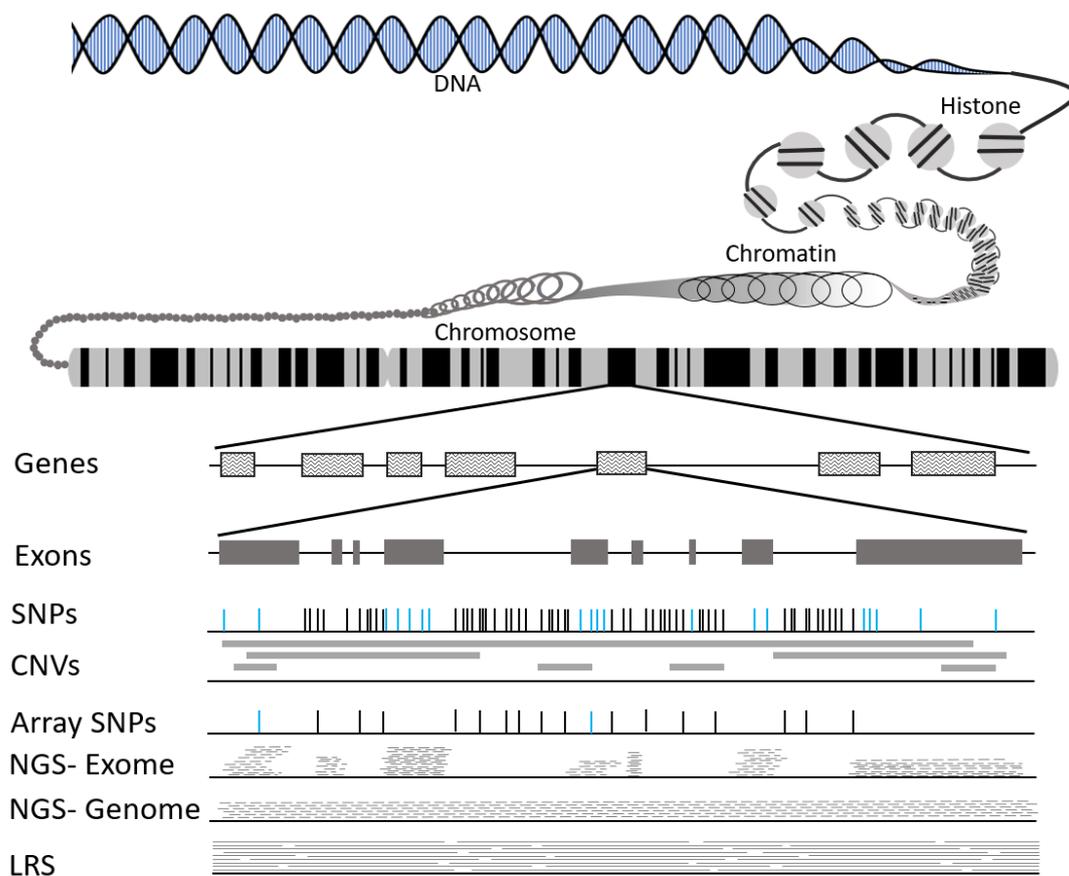
Figure 1. Illustration of DNA organization, types of variation present, and capture of variants using differing technologies.

3.2. Types of Genetic Variation

Changes (variation) in DNA can be classified according to size and type of variation (Table 2). Small-scale variation (<1 kilobase, kb) includes single nucleotide variants (SNV), short insertions or deletions (indels) and repetitive elements (RE; e.g., tandem repeats and transposable elements). Large structural variants (SV) involve DNA segments >1kb and include deletions, duplications, insertions, inversions, translocations, and copy number variation (CNV). Genetic variation is the rule rather than the exception; it is responsible for the phenotypic differences between individuals and may be beneficial or harmful. Genetic variation alters the phenotype by modifying (i) protein coding sequences, (ii) promoters and other regulatory elements, (iii) splice sites and other regions affecting transcript structures, and (iv) other genomic regions with unknown direct connections to known protein function. As new genetic techniques have become available, numerous examples of genetic variation have been identified in CHD patients (reviewed in [3]). Taken together, the study of genetic variation has improved our understanding of the underpinnings of CHD.

3.3. Tools to Identify Genetic Variation

Generally, the tools can be divided into two types: those that seek variation within the human genome and those that are targeted to specific genomic locations. While the karyotype has been widely used in the past century to identify numerous types of SV [14], in this paper we will focus on more recently developed high-throughput technologies. In the late 1990s and early 2000s, microarray genotyping chips were developed to capture known SNVs [15]. Notably, these genotyping chips capture only a fraction of the SNV variation; however, the seminal International HapMap project demonstrated that only a fraction of the variants would be required to capture common variation due to linkage

disequilibrium patterns [16]. Beyond SNV, these chips have been used to detect CNV [17], including in clinical laboratories [18].

Table 2. Characteristics of types of genetic variation.

Type of Variant	Description	Consequence	Laboratory Methods	Examples in CHD
Single Nucleotide Variation (SNV)	Substitution of single bp for another bp	• Individuals harbor ~3 million SNV [19,20]. • Many have no known functional effect. • Can alter protein structure or regulation	Array +++ NGS +++ LRS +++	Commonly identified in BAV and HLHS patients, e.g., NOTCH1 [21,22], GATA4 [23], GJA1 [24], and LRP2 [25]
Small Insertion/ Deletion (indel)	1–50 bp duplicated or deleted	• Multiple mechanisms of mutagenesis possible	NGS ++ LRS +++	SMAD4 [26] ETS1 [27]
Tandem Repeats (TR)	Repeats (1–100 bp) occur at single locus	• Present in at least 1/3 of human protein sequence [28]. • Contribution to disease [29–31] related to alterations of gene expression [32].	LRS +++	Emerging area of focus. The number of repeats in Fragile X associated with cardiovascular outcomes [33].
Transposable Elements (TE)	Repeats (100 bp–20 kbp) occur at multiple loci	• Common, accounting for more than 1/3 of the mammalian genome. • • Can impact protein structure [34] or gene regulation [35–38].	LRS +++	Emerging area of focus.
Copy Number Variation (CNV)	Duplication or deletion covering 1 Kb or greater.	• Major source of human genome variation [39]. • May alter gene expression [40,41]. • Plant studies suggest that up to 50% overlap genes/ gene regulatory regions [42]. • Account for up to 1 in 8 CHD cases [3,43].	Array ++ NGS ++ LRS +++	Aneuploidies, trisomies, and large SV are often associated with CHD [44,45]. Rare CNV are enriched in both BAV and HLHS [46–50].

+ = effectiveness of method to capture.

Another option to capture genetic variation is next generation sequencing (NGS). Briefly, NGS reads millions of small fragments of DNA in parallel [51]; these reads must be pieced back together using bioinformatics tools, which include quality control, alignment to a reference genome, and variant calling [52,53]. NGS can be used to capture SNV and some indels as well as CNV. NGS can be targeted to specific genomic regions, capture the exomes (WES), or capture the whole genome (WGS). The benefit of WES is a major reduction in costs (both from the laboratory perspective as well as data processing and storage). It has been shown that WES provides less even and more biased coverage than WGS (Figure 1) [54,55]. SV detection using WES is still a challenging area [55]. Finally, the NGS approaches are not assessable to all areas of the human genome, thus some variation will be missed [56].

Whereas NGS can produce reads up to 600 bp, long-read sequencing (LRS) routinely captures reads in excess of 10,000 bp (Figure 1), originating from single DNA molecules. Current technologies include Single Molecule Real Time (SMRT) by Pacific Biosciences of California (PacBio, Menlo Park, CA, USA) and Nanopore by Oxford Nanopore Technologies (ONT, Oxford, UK) [57]. As noted above, in NGS the small fragments have to be realigned using a reference genome, which is comprised of multiple individuals but regions which capture a single individual can create bias [58]. LRS is amenable to de novo assembly resulting in fewer missed regions [59] and identification of more structural

variation [60–62]. Further, LRS is the only high throughput technology to effectively capture RE [32]. A major challenge for LRS is that it requires ultra-long, high molecular weight DNA; thus, specialized DNA isolation protocols applied to fresh sample or intact cells is necessary [57]. It is important to note that the bioinformatic processes for LRS differ markedly from NGS approaches. While both SMRT and ONT are excellent at capturing structural variation, only SMRT LRS capture SNV and indels with high accuracy [63,64].

4. Design and Analytic Considerations

4.1. Strategies to Identify Genetic Variants of Interest

The goal of genetic studies is to find variants that contribute to the etiology of the phenotype of interest, in our case CHD. The initial steps in this process are i) identify the study subjects, ii) specify the technique to capture genetic information, and iii) specify the analytic approach to establish genetic variants of interest. When designing a study, it is important to consider analytic approaches for both common and rare variants [65]. However, it is important to remember that analytic evidence, in and of its own, is not sufficient to demonstrate that a variant causes CHD. Further, demonstrating that using independent participants to identify the same variants, e.g., replication, supports the confidence that a variant may contribute to CHD. However, given the correlated structure of the genome, replication should not be considered sufficient to implicate a variant in CHD etiology.

Types of study design considerations are summarized in Table 3. At the crux of all genetic discovery studies is the collection of study subjects (Case Type) and specification of a method (Study Design) to distinguish affected from unaffected (Control Type) subjects. As described above, the decision related to phenotype definition is not trivial. Inclusion of too broad a phenotype may introduce heterogeneity in the underlying etiology, thereby reducing power. However, based on animal models and human studies, CHD exhibits phenotypic plasticity [66–69]; thus, too narrow a phenotype may miss individuals with shared etiology. Additionally, most CHD studies do not consider extra-cardiac features. While on one hand the extra-cardiac features may be part of the underlying etiology, e.g., pleiotropy, it may also be simply by chance that an individual has causal variants for CHD as well as causal variants for extra-cardiac features. When studying unrelated cases, disentangling phenotypic plasticity and pleiotropy can be challenging. However, studying multiple affected individuals from the same family may alleviate these challenges. For example, for phenotypic plasticity, one can study how a broader phenotype segregates with a variant [66,68]. Further, for pleiotropy, family studies can allow researchers to determine how often the two traits co-occur and the number of variants of interest shared in common [70].

Table 3. Study designs, analysis types, and limitations.

Study Design	Case Type	Control Type	Analysis Type	Limitations
Case Control	Unrelated	Unrelated -local	Association	Heterogeneity
				Heterogeneity
		Unrelated—Out of study	Association	Differences between cases and controls
				Differences in genotype generation
Trio	Unrelated	Parents of cases	Family Based Association (TDT)	Heterogeneity
			Linkage analyses	Heterogeneity
			Filtering	Generalizability
Family	Related—may be a series of families	Family members of cases	Family based Association	Heterogeneity
			Linkage analysis	Generalizability
			Filtering	Generalizability

The selection of controls is also critical as biases in controls may impact power to identify variants. Ideally, controls should be (i) phenotyped using similar protocols to the

cases, (ii) genotyped with the same platform at the same time as the cases, and (iii) have a similar ancestral background (at minimum race and ethnicity) to the cases. Above, we emphasized the importance of appropriately phenotyping the cases. It is essential that controls are phenotyped equivalently to the cases. Failure to phenotype controls means that affected individuals may be included as a control; this will reduce power. While CHD is considered a rare condition, some phenotypes may go undetected in "normal, healthy" individuals. For example, BAV occurs in 1–2% of the population and may be unrecognized short of a cardiac imaging study. Thus, for optimal study power, it is essential that controls are as rigorously phenotyped as cases. Ideally, differences in genotype are due to phenotype differences between cases and controls and not, for example, because of variation in laboratory procedures, e.g., differences in genotyping chips, sequencing platforms, or genotyping chip version. For sequencing-based data, the version (build) of the human genome to which the data is aligned is also critical as not all regions are captured equally across all builds [71]. Matching of controls based on ancestry is essential as ancestral differences between cases and controls can cause spurious association [72]. Ancestral mismatching can be identified by evaluating the distribution of test statistics across the genome, the genomic inflation factor [73,74]. If mismatching is occurring, principal components analyses can be used to identify cases and controls with similar background genetic composition.

There are several sources of controls. Local controls recruited from the same geographic area as the cases are likely to be similar in ancestral distribution. Unfortunately, out of study controls who do not meet these criteria are commonly used. For example, genetic data from millions of individuals is available to researchers through portals such as dbGAP [75,76]. Although their use as out of study controls can markedly reduce the cost of genetic studies, dissimilarity of ancestral background may introduce noise that obscures key findings. Lastly, there are family-based controls. Unaffected parents are often included in studies of rare conditions to help exclude variants not contributing to disease. In extended family designs, multiple unaffected family members may be included. For rare conditions, the inclusion of extended family members may yield a large number of unaffected individuals. In these cases, sequential sampling where all first-degree relatives are phenotyped, and only when additional family members are identified, is the pedigree expanded. The combinations of types of cases and controls establishes the study design (Table 3).

4.2. Analytic Approaches

Analytic approaches can be separated into those which are powered for common variants (present in at least 1% of the population) or those optimal for rare variants (present in less than 1% of the population). We will first describe common variant approaches followed by a discussion of rare variants. For common variants, several assumptions should be evaluated prior to analysis. First, the variant should be evaluated for deviations from Hardy Weinberg Equilibrium (HWE). For variants which are not exhibiting selective pressure, deviations from HWE often are the result of erroneous genotypes [77]. Second, the genetic composition of the cases and controls should be evaluated by using principal component analyses [78] and the genomic inflation factor [73,74]. Differences in the genetic composition can be accounted for by selection of comparable groups (genetic matching) or by inclusion of principal components as covariates in the analyses [79]. To minimize risk for population stratification, most genetic analyses are stratified by continental ancestry, e.g., European, Asian, or African. For studies which evaluate variants across the genome, the risk of false positive association is high unless appropriate multiple testing correction is applied [80].

For common variants, association-based testing is used to test whether genotype predicts phenotype. For unrelated cases and controls, a Cochran–Armitage test for trend is often employed to test whether a specific genetic variant occurs more often in cases than controls. However, if covariates, such as age, sex, or adjustments for population

stratification, need to be incorporated in the model, then logistic regression is used, with the variant re-coded as the number of minor alleles present. Use of genome-wide association (GWAS) has been reported for various CHD [81]. While most association studies utilize unrelated cases and controls, family-based association tests are based on the distribution of Mendelian transmissions from parents to their offspring [82]. While considered less powerful than using unrelated individuals, they are robust to population stratification [83] and are considered more powerful for rare variant analyses [84,85]. When testing for association with common variants, a thousand or more cases are often required with larger numbers of controls for reasonably well powered studies as effect sizes of common variants are often modest.

For rare variants, three approaches can be used: gene collapsing/burden tests, linkage analyses, and filtering. The gene collapsing method uses an association framework, but the variants are not considered individually, either by generating a binary score (presence or absence) or a continuous score. A common approach for gene collapsing is SNP-set (Sequence) Kernel Association Test (SKAT) which allows covariate adjustment [86]. While burden analyses provide evidence that variation in a gene is associated with CHD, it is not based on specific variants [87,88]. While simulations suggest that thousands of cases are required for well powered studies [89], work using burden analysis in CHD suggests that cases sample sizes below 1000 may be sufficient [87,88]. When trio data are avaialble, de novo varaints can also be identified. Notably, researchers can also evaluate whether specific genes or classes of de novo variants enriched in individuals with CHD compared to expectations based on probalistic modeling [90]. This approach has been used to demonstrate the importance of de novo variants in sydromic CHD, while non-syndromic CHD did not exhibit enrichment [91,92]. Similar to the gene collapsing methods, these approaches have been used with case sample sizes below 1000 [91,92]. Another approach for rare variant assessment is linkage analyses. Linkage analyses tests the hypothesis that the phenotype is inherited (segregates) with a variant more often than expected by chance. Early work in gene discovery for CHD used linkage analyses [68,93–97]. A challenge with linkage analyses is that the structure of the family impacts the power of discovery. Large families with multiple affected individuals across generations can be powerful as effect sizes are typically large and there is less background noise. But, such families may be difficult to find. Thus, many studies use many smaller families, but this can introduce genetic heterogeneity. It is important to recognize that linkage analyses leverage recombination events, and thus they identify a region rather than a variant. These regions may span millions of base pairs and harbor variants segregating with disease.

Beyond analytic approaches, many investigators simply use a process of prioritizing rare variants using a filtering approach first used by Ng and colleagues to evaluate exome sequencing data [98]. The filtering approach restricts variants based on the proposed inheritance model, population frequency, and the putative impact of a variant. With respect to inheritance, models include recessive (two copies of the alternative allele, or for compound heterozygotes two alternative alleles within a gene), dominant (one copy of the alternative allele), and de novo (a new mutation not found in the parents). For recessive and dominant inheritance, filtering can be done using cases only. Inclusion of unaffected parents (trios) can be used to further narrow the number of variants. Unaffected parents are required for detection of de novo rare variants. With respect to population frequency, the assumption is that a highly penetrant variant of a serious medical condition is not likely to be seen in the general population [99]. Aggregate databases like ExAC and gnomAD [100,101] provide access to large numbers of sequenced individuals. As these data were from studies primarily of adults, the assumption is that rare conditions with serious medical implications such as CHD are unlikely to be represented in the cohort. However, as noted above, conditions like BAV, which share genetic etiology with other CHD, often do not exhibit the morbidities until mid-late life, thus they could be in those datasets. Additionally, the use of ExAC and gnomAD is similar to using out of study controls, so care should be taken to ensure data comparability. Lastly, putative impact of

variants is an important filtering criterion. Most NGS studies restrict variants to those that alter the protein and are bioinformatically predicted to impact protein function. Many tools have been developed to predict whether a protein change is likely to have functional impact [102–105]. It is important to recognize that these tools suffer from both false positives and negatives [106,107]. Further, the focus exclusively on variants which change protein structure ignores the potential role for regulatory variants. As heart development is a carefully orchestrated process where genes are turned on and off at the appropriate moments [108], it seems logical that regulatory variants could also contribute to CHD. The filtering strategy is an excellent approach for rare, highly penetrant conditions where the number of subjects is highly limited.

4.3. Implicating Variants in Disease Etiology

The ideal outcome of genetic studies is to identify variants that have a causal role, meaning they are necessary and sufficient for disease. However, as noted above, analytic approaches never achieve this goal. For CHD, which may be due to either multiple genetic hits [69] or genetic hits plus environmental insults [109–111], demonstrating causality may be an unattainable goal. A more realistic goal for genetic analytics is to identify variants in specific genes that influence occurrence of a phenotype which provide keys to open biologic doors that inform how the genetic variants modulate heart development. Further, we suggest that biologic studies should focus on evaluating the role of variants in CHD etiology rather than causality. The etiologic role can be explored with in vitro and/or in vivo studies. In vitro studies are powerful approaches for many conditions, but the evidence gained related to CHD should be interpreted with caution. This is because the developing heart is composed of both myocytes and endothelial cells and this cellular communication is critical for appropriate structural development [112]. In vivo models capture the cellular dynamics but there are still concerns for off-target effects with genome editing [113] and the potential differences in regulatory regions across species [114].

5. Clinical Considerations

Genetic testing of severe CHD has recently identified a genetic cause in only ~35% of the cases [115]. Having a genetic diagnosis can be useful clinically as it can help identify syndromic CHD which often requires different clinical management than non-syndromic CHD. Further, knowing the cause of CHD can help families understand their recurrence risk [116]. Given the familial data, it is likely that CHD will exhibit Mendelian inheritance in only a few families with highly penetrant variants and complex inheritance for most other families. These complex inheritance CHD will likely be missed by current clinical genetic testing as the focus is largely on highly penetrant rare variants. To address prediction in complex inheritance, use of polygenic risk scores has been proposed but the utility of such an approach is still uncertain [117,118]. However, we recognize that the long-term prognosis of CHD, even within specific types of CHD, is highly variable, thus the development of predictive models for outcomes, which may include biologic and genetic markers, rather than simply seeking to identify causes of CHD would be highly valuable. Ultimately, it is the hope that a genetic diagnosis can inform personalized medicine, but further work is required.

6. Conclusions

As recently reviewed [3], while considerable progress has been made in defining the genetic underpinnings of CHD, significant work remains. Available tools, such as high throughput genotyping and state-of-the-art analytic methods, will facilitate future studies advancing knowledge of CHD genetic etiology. However, it is becoming more evident that good genetic studies start with good planning. Continued progress in unraveling the genetic underpinnings of CHD will require multidisciplinary collaboration between geneticists, quantitative scientists, clinicians, and developmental biologists.

References

1. Tan, C.M.J.; Lewandowski, A.J. The Transitional Heart: From Early Embryonic and Fetal Development to Neonatal Life. *Fetal Diagn. Ther.* **2020**, *47*, 373–386. [CrossRef]
2. Combs, M.D.; Yutzey, K.E. Heart valve development: Regulatory networks in development and disease. *Circ. Res.* **2009**, *105*, 408–421. [CrossRef]
3. Pierpont, M.E.; Brueckner, M.; Chung, W.K.; Garg, V.; Lacro, R.V.; McGuire, A.L.; Mital, S.; Priest, J.R.; Pu, W.T.; Roberts, A.; et al. American Heart Association Council on Cardiovascular Disease in the Young; Council on Cardiovascular and Stroke Nursing; and Council on Genomic and Precision Medicine. (2018). Genetic basis for congenital heart disease: Revisited: A scientific statement from the American Heart Association. *Circulation* **2018**, *138*, e653–e711. [CrossRef]
4. Hoffman, J.I.; Kaplan, S. The incidence of congenital heart disease. *J. Am. Coll. Cardiol.* **2002**, *39*, 1890–1900. [CrossRef]
5. Cripe, L.; Andelfinger, G.; Martin, L.J.; Shooner, K.; Benson, D.W. Bicuspid aortic valve is heritable. *J. Am. Coll. Cardiol.* **2004**, *44*, 138–143. [CrossRef] [PubMed]
6. Benson, D.W. The genetics of congenital heart disease: A point in the revolution. *Cardiol. Clin.* **2002**, *20*, 385–394. [CrossRef]
7. Benson, D.W.; Sharkey, A.; Fatkin, D.; Lang, P.; Basson, C.T.; McDonough, B.; Strauss, A.W.; Seidman, J.G.; Seidman, C.E. Reduced penetrance, variable expressivity, and genetic heterogeneity of familial atrial septal defects. *Circulation* **1998**, *97*, 2043–2048. [CrossRef] [PubMed]
8. Devanna, P.; Chen, X.S.; Ho, J.; Gajewski, D.; Smith, S.D.; Gialluisi, A.; Francks, C.; Fisher, S.E.; Newbury, D.F.; Vernes, S.C. Next-gen sequencing identifies non-coding variation disrupting miRNA-binding sites in neurological disorders. *Mol. Psychiatry* **2018**, *23*, 1375–1384. [CrossRef] [PubMed]
9. Zhang, F.; Lupski, J.R. Non-coding genetic variants in human disease. *Hum. Mol. Genet.* **2015**, *24*, R102–R110. [CrossRef] [PubMed]
10. Moriya, T.; Yamamura, M.; Kiga, D. Effects of downstream genes on synthetic genetic circuits. *BMC Syst. Biol.* **2014**, *8*, 1–10. [CrossRef]
11. Smemo, S.; Tena, J.J.; Kim, K.H.; Gamazon, E.R.; Sakabe, N.J.; Gomez-Marin, C.; Aneas, I.; Credidio, F.L.; Sobreira, D.R.; Wasserman, N.F.; et al. Obesity-associated variants within FTO form long-range functional connections with IRX3. *Nature* **2014**, *507*, 371–375. [CrossRef]
12. Consortium, E.P. An integrated encyclopedia of DNA elements in the human genome. *Nature* **2012**, *489*, 57–74. [CrossRef]
13. Richter, F.; Morton, S.U.; Kim, S.W.; Kitaygorodsky, A.; Wasson, L.K.; Chen, K.M.; Zhou, J.; Qi, H.; Patel, N.; DePalma, S.R.; et al. Genomic analyses implicate noncoding de novo variants in congenital heart disease. *Nat. Genet.* **2020**, *52*, 769–777. [CrossRef] [PubMed]
14. Ferguson-Smith, M.A. History and evolution of cytogenetics. *Mol. Cytogenet.* **2015**, *8*, 1–8. [CrossRef] [PubMed]
15. Bumgarner, R. Overview of DNA microarrays: Types, applications, and their future. *Curr. Protoc. Mol. Biol.* **2013**, *101*, 22.1.1–22.1.11. [CrossRef]
16. A haplotype map of the human genome. *Nature* **2005**, *437*, 1299–1320. [CrossRef]
17. Carter, N.P. Methods and strategies for analyzing copy number variation using DNA microarrays. *Nat. Genet.* **2007**, *39*, S16–S21. [CrossRef]
18. Coughlin, C.R.; Scharer, G.H.; Shaikh, T.H. Clinical impact of copy number variation analysis using high-resolution microarray technologies: Advantages, limitations and concerns. *Genome Med.* **2012**, *4*, 1–12. [CrossRef]
19. Shen, H.; Li, J.; Zhang, J.; Xu, C.; Jiang, Y.; Wu, Z.; Zhao, F.; Liao, L.; Chen, J.; Lin, Y.; et al. Comprehensive characterization of human genome variation by high coverage whole-genome sequencing of forty four Caucasians. *PLoS ONE* **2013**, *8*, e59494. [CrossRef] [PubMed]
20. Vidal, E.A.; Moyano, T.C.; Bustos, B.I.; Perez-Palma, E.; Moraga, C.; Riveras, E.; Montecinos, A.; Azocar, L.; Soto, D.C.; Vidal, M.; et al. Whole Genome Sequence, Variant Discovery and Annotation in Mapuche-Huilliche Native South Americans. *Sci. Rep.* **2019**, *9*, 1–11. [CrossRef]
21. Garg, V.; Muth, A.N.; Ransom, J.F.; Schluterman, M.K.; Barnes, R.; King, I.N.; Grossfeld, P.D.; Srivastava, D. Mutations in NOTCH1 cause aortic valve disease. *Nature* **2005**, *437*, 270–274. [CrossRef] [PubMed]
22. Foffa, I.; Ali, L.A.; Panesi, P.; Mariani, M.; Festa, P.; Botto, N.; Vecoli, C.; Andreassi, M.G. Sequencing of NOTCH1, GATA5, TGFBR1 and TGFBR2 genes in familial cases of bicuspid aortic valve. *BMC Med. Genet.* **2013**, *14*, 1–8. [CrossRef] [PubMed]
23. Yang, B.; Zhou, W.; Jiao, J.; Nielsen, J.B.; Mathis, M.R.; Heydarpour, M.; Lettre, G.; Folkersen, L.; Prakash, S.; Schurmann, C.; et al. Protein-altering and regulatory genetic variants near GATA4 implicated in bicuspid aortic valve. *Nat. Commun.* **2017**, *8*, 1–10. [CrossRef]

24. Dasgupta, C.; Martinez, A.M.; Zuppan, C.W.; Shah, M.M.; Bailey, L.L.; Fletcher, W.H. Identification of connexin43 (alpha1) gap junction gene mutations in patients with hypoplastic left heart syndrome by denaturing gradient gel electrophoresis (DGGE). *Mutat. Res.* **2001**, *479*, 173–186. [CrossRef]

25. Theis, J.L.; Vogler, G.; Missinato, M.A.; Li, X.; Nielsen, T.; Zeng, X.I.; Martinez-Fernandez, A.; Walls, S.M.; Kervadec, A.; Kezos, J.N.; et al. Patient-specific genomics and cross-species functional analysis implicate LRP2 in hypoplastic left heart syndrome. *eLife* **2020**, *9*, e59554. [CrossRef] [PubMed]

26. Park, J.E.; Park, J.S.; Jang, S.Y.; Park, S.H.; Kim, J.W.; Ki, C.S.; Kim, D.K. A novel SMAD6 variant in a patient with severely calcified bicuspid aortic valve and thoracic aortic aneurysm. *Mol. Genet. Genom. Med.* **2019**, *7*, e620. [CrossRef] [PubMed]

27. Tootleman, E.; Malamut, B.; Akshoomoff, N.; Mattson, S.N.; Hoffman, H.M.; Jones, M.C.; Printz, B.; Shiryaev, S.A.; Grossfeld, P. Partial Jacobsen syndrome phenotype in a patient with a de novo frameshift mutation in the ETS1 transcription factor. *Mol. Case Stud.* **2019**, *5*, a004010. [CrossRef]

28. Torresen, O.K.; Star, B.; Mier, P.; Andrade-Navarro, M.A.; Bateman, A.; Jarnot, P.; Gruca, A.; Grynberg, M.; Kajava, A.V.; Promponas, V.J.; et al. Tandem repeats lead to sequence assembly errors and impose multi-level challenges for genome and protein databases. *Nucleic Acids Res.* **2019**, *47*, 10994–11006. [CrossRef]

29. Usdin, K. The biological effects of simple tandem repeats: Lessons from the repeat expansion diseases. *Genome Res.* **2008**, *18*, 1011–1019. [CrossRef]

30. Tang, H.; Kirkness, E.F.; Lippert, C.; Biggs, W.H.; Fabani, M.; Guzman, E.; Ramakrishnan, S.; Lavrenko, V.; Kakaradov, B.; Hou, C.; et al. Profiling of Short-Tandem-Repeat Disease Alleles in 12,632 Human Whole Genomes. *Am. J. Hum. Genet.* **2017**, *101*, 700–715. [CrossRef]

31. Sun, J.H.; Zhou, L.; Emerson, D.J.; Phyo, S.A.; Titus, K.R.; Gong, W.; Gilgenast, T.G.; Beagan, J.A.; Davidson, B.L.; Tassone, F.; et al. Disease-Associated Short Tandem Repeats Co-localize with Chromatin Domain Boundaries. *Cell* **2018**, *175*, 224–238. [CrossRef] [PubMed]

32. Sulovari, A.; Li, R.; Audano, P.A.; Porubsky, D.; Vollger, M.R.; Logsdon, G.A.; Variation, C.H.G.S.; Warren, W.C.; Pollen, A.A.; Chaisson, M.J.P.; et al. Human-specific tandem repeat expansion and differential gene expression during primate evolution. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 23243–23253. [CrossRef] [PubMed]

33. Tassanakijpanich, N.; Cohen, J.; Cohen, R.; Srivatsa, U.N.; Hagerman, R.J. Cardiovascular Problems in the Fragile X Premutation. *Front. Genet.* **2020**, *11*, 1244. [CrossRef]

34. Kazazian, H.H., Jr.; Wong, C.; Youssoufian, H.; Scott, A.F.; Phillips, D.G.; Antonarakis, S.E. A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man. *Nature* **1988**, *332*, 164–166. [CrossRef]

35. Ye, M.; Goudot, C.; Hoyler, T.; Lemoine, B.; Amigorena, S.; Zueva, E. Specific subfamilies of transposable elements contribute to different domains of T lymphocyte enhancers. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 7905–7916. [CrossRef]

36. Faulkner, G.J.; Billon, V. L1 retrotransposition in the soma: A field jumping ahead. *Mob. DNA* **2018**, *9*, 1–18. [CrossRef]

37. Pehrsson, E.C.; Choudhary, M.N.K.; Sundaram, V.; Wang, T. The epigenomic landscape of transposable elements across normal human development and anatomy. *Nat. Commun.* **2019**, *10*, 1–16. [CrossRef]

38. Diehl, A.G.; Ouyang, N.; Boyle, A.P. Transposable elements contribute to cell and species-specific chromatin looping and gene regulation in mammalian genomes. *Nat. Commun.* **2020**, *11*, 1–18. [CrossRef]

39. Abel, H.J.; Larson, D.E.; Regier, A.A.; Chiang, C.; Das, I.; Kanchi, K.L.; Layer, R.M.; Neale, B.M.; Salerno, W.J.; Reeves, C.; et al. Mapping and characterization of structural variation in 17,795 human genomes. *Nature* **2020**, *583*, 83–89. [CrossRef] [PubMed]

40. Chiang, C.; Scott, A.J.; Davis, J.R.; Tsang, E.K.; Li, X.; Kim, Y.; Hadzic, T.; Damani, F.N.; Ganel, L.; Consortium, G.T.; et al. The impact of structural variation on human gene expression. *Nat. Genet.* **2017**, *49*, 692–699. [CrossRef] [PubMed]

41. Han, L.; Zhao, X.; Benton, M.L.; Perumal, T.; Collins, R.L.; Hoffman, G.E.; Johnson, J.S.; Sloofman, L.; Wang, H.Z.; Stone, M.R.; et al. Functional annotation of rare structural variation in the human brain. *Nat. Commun.* **2020**, *11*, 1–13. [CrossRef] [PubMed]

42. Alonge, M.; Wang, X.; Benoit, M.; Soyk, S.; Pereira, L.; Zhang, L.; Suresh, H.; Ramakrishnan, S.; Maumus, F.; Ciren, D.; et al. Major Impacts of Widespread Structural Variation on Gene Expression and Crop Improvement in Tomato. *Cell* **2020**, *182*, 145–161. [CrossRef] [PubMed]

43. Costain, G.; Silversides, C.K.; Bassett, A.S. The importance of copy number variation in congenital heart disease. *NPJ Genom. Med.* **2016**, *1*, 1–11. [CrossRef] [PubMed]

44. Liu, Y.; Chang, X.; Glessner, J.; Qu, H.; Tian, L.; Li, D.; Nguyen, K.; Sleiman, P.M.A.; Hakonarson, H. Association of Rare Recurrent Copy Number Variants with Congenital Heart Defects Based on Next-Generation Sequencing Data from Family Trios. *Front. Genet.* **2019**, *10*, 819. [CrossRef] [PubMed]

45. Prakash, S.K.; Bondy, C.A.; Maslen, C.L.; Silberbach, M.; Lin, A.E.; Perrone, L.; Limongelli, G.; Michelena, H.I.; Bossone, E.; Citro, R.; et al. Autosomal and X chromosome structural variants are associated with congenital heart defects in Turner syndrome: The NHLBI GenTAC registry. *Am. J. Med. Genet. A* **2016**, *170*, 3157–3164. [CrossRef]

46. Hitz, M.P.; Lemieux-Perreault, L.P.; Marshall, C.; Feroz-Zada, Y.; Davies, R.; Yang, S.W.; Lionel, A.C.; D'Amours, G.; Lemyre, E.; Cullum, R.; et al. Rare copy number variants contribute to congenital left-sided heart disease. *PLoS Genet.* **2012**, *8*, e1002903. [CrossRef]

47. Glidewell, S.C.; Miyamoto, S.D.; Grossfeld, P.D.; Clouthier, D.E.; Coldren, C.D.; Stearman, R.S.; Geraci, M.W. Transcriptional Impact of Rare and Private Copy Number Variants in Hypoplastic Left Heart Syndrome. *Clin. Transl. Sci.* **2015**, *8*, 682–689. [CrossRef]

48. Carey, A.S.; Liang, L.; Edwards, J.; Brandt, T.; Mei, H.; Sharp, A.J.; Hsu, D.T.; Newburger, J.W.; Ohye, R.G.; Chung, W.K.; et al. Effect of copy number variants on outcomes for infants with single ventricle heart defects. *Circ. Cardiovasc. Genet.* **2013**, *6*, 444–451. [CrossRef]

49. Luyckx, I.; Kumar, A.A.; Reyniers, E.; Dekeyser, E.; Vanderstraeten, K.; Vandeweyer, G.; Wunnemann, F.; Preuss, C.; Mazzella, J.M.; Goudot, G.; et al. Copy number variation analysis in bicuspid aortic valve-related aortopathy identifies TBX20 as a contributing gene. *Eur. J. Hum. Genet.* **2019**, *27*, 1033–1043. [CrossRef]

50. Prakash, S.; Kuang, S.Q.; GenTAC Registry Investigators; Regalado, E.; Guo, D.; Milewicz, D. Recurrent Rare Genomic Copy Number Variants and Bicuspid Aortic Valve Are Enriched in Early Onset Thoracic Aortic Aneurysms and Dissections. *PLoS ONE* **2016**, *11*, e0153543. [CrossRef]

51. Behjati, S.; Tarpey, P.S. What is next generation sequencing? *Arch. Dis. Child. Educ. Pract.* **2013**, *98*, 236–238. [CrossRef] [PubMed]

52. Kanzi, A.M.; San, J.E.; Chimukangara, B.; Wilkinson, E.; Fish, M.; Ramsuran, V.; de Oliveira, T. Next Generation Sequencing and Bioinformatics Analysis of Family Genetic Inheritance. *Front. Genet.* **2020**, *11*, 1250. [CrossRef] [PubMed]

53. Rehm, H.L.; Bale, S.J.; Bayrak-Toydemir, P.; Berg, J.S.; Brown, K.K.; Deignan, J.L.; Friez, M.J.; Funke, B.H.; Hegde, M.R.; Lyon, E. ACMG clinical laboratory standards for next-generation sequencing. *Genet. Med.* **2013**, *15*, 733–747. [CrossRef] [PubMed]

54. Trudso, L.C.; Andersen, J.D.; Jacobsen, S.B.; Christiansen, S.L.; Congost-Teixidor, C.; Kampmann, M.L.; Morling, N. A comparative study of single nucleotide variant detection performance using three massively parallel sequencing methods. *PLoS ONE* **2020**, *15*, e0239850. [CrossRef]

55. Belkadi, A.; Bolze, A.; Itan, Y.; Cobat, A.; Vincent, Q.B.; Antipenko, A.; Shang, L.; Boisson, B.; Casanova, J.L.; Abel, L. Whole-genome sequencing is more powerful than whole-exome sequencing for detecting exome variants. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 5473–5478. [CrossRef]

56. Merker, J.D.; Wenger, A.M.; Sneddon, T.; Grove, M.; Zappala, Z.; Fresard, L.; Waggott, D.; Utiramerur, S.; Hou, Y.; Smith, K.S.; et al. Long-read genome sequencing identifies causal structural variation in a Mendelian disease. *Genet. Med.* **2018**, *20*, 159–163. [CrossRef]

57. Mantere, T.; Kersten, S.; Hoischen, A. Long-Read Sequencing Emerging in Medical Genetics. *Front. Genet.* **2019**, *10*, 426. [CrossRef] [PubMed]

58. Ballouz, S.; Dobin, A.; Gillis, J.A. Is it time to change the reference genome? *Genome Biol.* **2019**, *20*, 1–9. [CrossRef]

59. Porubsky, D.; Ebert, P.; Audano, P.A.; Vollger, M.R.; Harvey, W.T.; Marijon, P.; Ebler, J.; Munson, K.M.; Sorensen, M.; Sulovari, A.; et al. Fully phased human genome assembly without parental data using single-cell strand sequencing and long reads. *Nat. Biotechnol.* **2021**, *39*, 302–308. [CrossRef] [PubMed]

60. Stancu, M.C.; van Roosmalen, M.J.; Renkens, I.; Nieboer, M.M.; Middelkamp, S.; de Ligt, J.; Pregno, G.; Giachino, D.; Mandrile, G.; Valle-Inclan, J.E.; et al. Mapping and phasing of structural variation in patient genomes using nanopore sequencing. *Nat. Commun.* **2017**, *8*, 1–13. [CrossRef]

61. Tham, C.Y.; Tirado-Magallanes, R.; Goh, Y.; Fullwood, M.J.; Koh, B.T.H.; Wang, W.; Ng, C.H.; Chng, W.J.; Thiery, A.; Tenen, D.G.; et al. NanoVar: Accurate characterization of patients' genomic structural variants using low-depth nanopore sequencing. *Genome Biol.* **2020**, *21*, 1–15. [CrossRef] [PubMed]

62. Yan, J.; Lv, S.; Hu, M.; Gao, Z.; He, H.; Ma, Q.; Deng, X.W.; Zhu, Z.; Wang, X. Single-Molecule Sequencing Assists Genome Assembly Improvement and Structural Variation Inference. *Mol. Plant* **2016**, *9*, 1085–1087. [CrossRef] [PubMed]

63. Wenger, A.M.; Peluso, P.; Rowell, W.J.; Chang, P.C.; Hall, R.J.; Concepcion, G.T.; Ebler, J.; Fungtammasan, A.; Kolesnikov, A.; Olson, N.D.; et al. Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. *Nat. Biotechnol.* **2019**, *37*, 1155–1162. [CrossRef] [PubMed]

64. Krishnakumar, R.; Sinha, A.; Bird, S.W.; Jayamohan, H.; Edwards, H.S.; Schoeniger, J.S.; Patel, K.D.; Branda, S.S.; Bartsch, M.S. Systematic and stochastic influences on the performance of the MinION nanopore sequencer across a range of nucleotide bias. *Sci. Rep.* **2018**, *8*, 1–13. [CrossRef] [PubMed]

65. Teekakirikul, P.; Zhu, W.; Gabriel, G.C.; Young, C.B.; Williams, K.; Martin, L.J.; Hill, J.C.; Richards, T.; Billaud, M.; Phillippi, J.A.; et al. Common deletion variants causing protocadherin-deficiency contribute to the complex genetics of bicuspid aortic valve and left-sided congenital heart defects. *Hum. Genet. Genom. Adv.* **2021**, in press.

66. Liu, X.; Yagi, H.; Saeed, S.; Bais, A.S.; Gabriel, G.C.; Chen, Z.; Peterson, K.A.; Li, Y.; Schwartz, M.C.; Reynolds, W.T.; et al. The complex genetics of hypoplastic left heart syndrome. *Nat. Genet.* **2017**, *49*, 1152–1159. [CrossRef] [PubMed]

67. Schulkey, C.E.; Regmi, S.D.; Magnan, R.A.; Danzo, M.T.; Luther, H.; Hutchinson, A.K.; Panzer, A.A.; Grady, M.M.; Wilson, D.B.; Jay, P.Y. The maternal-age-associated risk of congenital heart disease is modifiable. *Nature* **2015**, *520*, 230–233. [CrossRef] [PubMed]

68. Hinton, R.B.; Martin, L.J.; Rame-Gowda, S.; Tabangin, M.E.; Cripe, L.H.; Benson, D.W. Hypoplastic left heart syndrome links to chromosomes 10q and 6q and is genetically related to bicuspid aortic valve. *J. Am. Coll. Cardiol.* **2009**, *53*, 1065–1071. [CrossRef]

69. McBride, K.L.; Pignatelli, R.; Lewin, M.; Ho, T.; Fernbach, S.; Menesses, A.; Lam, W.; Leal, S.M.; Kaplan, N.; Schliekelman, P.; et al. Inheritance analysis of congenital left ventricular outflow tract obstruction malformations: Segregation, multiplex relative risk, and heritability. *Am. J. Med. Genet. A* **2005**, *134*, 180–186. [CrossRef]

70. Egbe, A.; Uppu, S.; Lee, S.; Ho, D.; Srivastava, S. Prevalence of associated extracardiac malformations in the congenital heart disease population. *Pediatr. Cardiol.* **2014**, *35*, 1239–1245. [CrossRef] [PubMed]

71.	Pan, B.; Kusko, R.; Xiao, W.; Zheng, Y.; Liu, Z.; Xiao, C.; Sakkiah, S.; Guo, W.; Gong, P.; Zhang, C.; et al. Similarities and differences between variants called with human reference genome HG19 or HG38. *BMC Bioinform.* **2019**, *20*, 17–29. [CrossRef]

72.	Knowler, W.C.; Williams, R.C.; Pettitt, D.J.; Steinberg, A.G. Gm3;5,13,14 and type 2 diabetes mellitus: An association in American Indians with genetic admixture. *Am. J. Hum. Genet.* **1988**, *43*, 520–526.

73.	Zheng, G.; Freidlin, B.; Li, Z.; Gastwirth, J.L. Genomic control for association studies under various genetic models. *Biometrics* **2005**, *61*, 186–192. [CrossRef] [PubMed]

74.	Clayton, D.G.; Walker, N.M.; Smyth, D.J.; Pask, R.; Cooper, J.D.; Maier, L.M.; Smink, L.J.; Lam, A.C.; Ovington, N.R.; Stevens, H.E.; et al. Population structure, differential bias and genomic control in a large-scale, case-control association study. *Nat. Genet.* **2005**, *37*, 1243–1246. [CrossRef] [PubMed]

75.	National Center for Biotechnology Information. dbGaP/Database of Genotypes and Phenotypes. Available online: https://www.ncbi.nlm.nih.gov/gap (accessed on 10 March 2020).

76.	Mailman, M.D.; Feolo, M.; Jin, Y.; Kimura, M.; Tryka, K.; Bagoutdinov, R.; Hao, L.; Kiang, A.; Paschall, J.; Phan, L.; et al. The NCBI dbGaP database of genotypes and phenotypes. *Nat. Genet.* **2007**, *39*, 1181–1186. [CrossRef] [PubMed]

77.	Chen, B.; Cole, J.W.; Grond-Ginsbach, C. Departure from Hardy Weinberg Equilibrium and Genotyping Error. *Front. Genet.* **2017**, *8*, 167. [CrossRef] [PubMed]

78.	Price, A.L.; Patterson, N.J.; Plenge, R.M.; Weinblatt, M.E.; Shadick, N.A.; Reich, D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **2006**, *38*, 904–909. [CrossRef]

79.	Wang, K.; Hu, X.; Peng, Y. An analytical comparison of the principal component method and the mixed effects model for association studies in the presence of cryptic relatedness and population stratification. *Hum. Hered.* **2013**, *76*, 1–9. [CrossRef]

80.	McCarthy, M.I.; Abecasis, G.R.; Cardon, L.R.; Goldstein, D.B.; Little, J.; Ioannidis, J.P.; Hirschhorn, J.N. Genome-wide association studies for complex traits: Consensus, uncertainty and challenges. *Nat. Rev. Genet.* **2008**, *9*, 356–369. [CrossRef]

81.	Agopian, A.J.; Goldmuntz, E.; Hakonarson, H.; Sewda, A.; Taylor, D.; Mitchell, L.E.; Pediatric Cardiac Genomics Consortium. Genome-Wide Association Studies and Meta-Analyses for Congenital Heart Defects. *Circ. Cardiovasc. Genet.* **2017**, *10*, e001449. [CrossRef]

82.	Horvath, S.; Xu, X.; Laird, N.M. The family based association test method: Strategies for studying general genotype–phenotype associations. *Eur. J. Hum. Genet.* **2001**, *9*, 301–306. [CrossRef]

83.	Lewinger, J.P.; Bull, S.B. Validity, efficiency, and robustness of a family-based test of association. *Genet. Epidemiol.* **2006**, *30*, 62–76. [CrossRef]

84.	Hecker, J.; Townes, F.W.; Kachroo, P.; Laurie, C.; Lasky-Su, J.; Ziniti, J.; Cho, M.H.; Weiss, S.T.; Laird, N.M.; Lange, C. A unifying framework for rare variant association testing in family-based designs, including higher criticism approaches, SKATs, and burden tests. *Bioinformatics* **2020**. [CrossRef]

85.	Zhou, J.J.; Yip, W.K.; Cho, M.H.; Qiao, D.; McDonald, M.L.; Laird, N.M. A comparative analysis of family-based and population-based association tests using whole genome sequence data. *BMC Proc.* **2014**, *8*, 1–5. [CrossRef] [PubMed]

86.	Wu, M.C.; Lee, S.; Cai, T.; Li, Y.; Boehnke, M.; Lin, X. Rare-variant association testing for sequencing data with the sequence kernel association test. *Am. J. Hum. Genet.* **2011**, *89*, 82–93. [CrossRef] [PubMed]

87.	Blue, G.M.; Ip, E.; Walker, K.; Kirk, E.P.; Loughran-Fowlds, A.; Sholler, G.F.; Dunwoodie, S.L.; Harvey, R.P.; Giannoulatou, E.; Badawi, N.; et al. Genetic burden and associations with adverse neurodevelopment in neonates with congenital heart disease. *Am. Heart J.* **2018**, *201*, 33–39. [CrossRef] [PubMed]

88.	Izarzugaza, J.M.G.; Ellesoe, S.G.; Doganli, C.; Ehlers, N.S.; Dalgaard, M.D.; Audain, E.; Dombrowsky, G.; Banasik, K.; Sifrim, A.; Wilsdon, A.; et al. Systems genetics analysis identifies calcium-signaling defects as novel cause of congenital heart disease. *Genome Med.* **2020**, *12*, 1–13. [CrossRef] [PubMed]

89.	Zhang, X.; Basile, A.O.; Pendergrass, S.A.; Ritchie, M.D. Real world scenarios in rare variant association analysis: The impact of imbalance and sample size on the power in silico. *BMC Bioinform.* **2019**, *20*, 1–10. [CrossRef]

90.	Samocha, K.E.; Robinson, E.B.; Sanders, S.J.; Stevens, C.; Sabo, A.; McGrath, L.M.; Kosmicki, J.A.; Rehnstrom, K.; Mallick, S.; Kirby, A.; et al. A framework for the interpretation of de novo mutation in human disease. *Nat. Genet.* **2014**, *46*, 944–950. [CrossRef]

91.	Homsy, J.; Zaidi, S.; Shen, Y.; Ware, J.S.; Samocha, K.E.; Karczewski, K.J.; DePalma, S.R.; McKean, D.; Wakimoto, H.; Gorham, J.; et al. De novo mutations in congenital heart disease with neurodevelopmental and other congenital anomalies. *Science* **2015**, *350*, 1262–1266. [CrossRef]

92.	Sifrim, A.; Hitz, M.P.; Wilsdon, A.; Breckpot, J.; Turki, S.H.; Thienpont, B.; McRae, J.; Fitzgerald, T.W.; Singh, T.; Swaminathan, G.J.; et al. Distinct genetic architectures for syndromic and nonsyndromic congenital heart defects identified by exome sequencing. *Nat. Genet.* **2016**, *48*, 1060–1065. [CrossRef]

93.	Mathias, R.S.; Lacro, R.V.; Jones, K.L. X-linked laterality sequence: Situs inversus, complex cardiac defects, splenic defects. *Am. J. Med. Genet.* **1987**, *28*, 111–116. [CrossRef] [PubMed]

94.	Wilson, L.; Curtis, A.; Korenberg, J.R.; Schipper, R.D.; Allan, L.; Chenevix-Trench, G.; Stephenson, A.; Goodship, J.; Burn, J. A large, dominant pedigree of atrioventricular septal defect (AVSD): Exclusion from the Down syndrome critical region on chromosome 21. *Am. J. Hum. Genet.* **1993**, *53*, 1262–1268.

95.	Terrett, J.A.; Newbury-Ecob, R.; Cross, G.S.; Fenton, I.; Raeburn, J.A.; Young, I.D.; Brook, J.D. Holt-Oram syndrome is a genetically heterogeneous disease with one locus mapping to human chromosome 12q. *Nat. Genet.* **1994**, *6*, 401–404. [CrossRef] [PubMed]

96. Satoda, M.; Pierpont, M.E.; Diaz, G.A.; Bornemeier, R.A.; Gelb, B.D. Char syndrome, an inherited disorder with patent ductus arteriosus, maps to chromosome 6p12-p21. *Circulation* **1999**, *99*, 3036–3042. [CrossRef] [PubMed]

97. Martin, L.J.; Ramachandran, V.; Cripe, L.H.; Hinton, R.B.; Andelfinger, G.; Tabangin, M.; Shooner, K.; Keddache, M.; Benson, D.W. Evidence in favor of linkage to human chromosomal regions 18q, 5q and 13q for bicuspid aortic valve and associated cardiovascular malformations. *Hum. Genet.* **2007**, *121*, 275–284. [CrossRef] [PubMed]

98. Ng, S.B.; Buckingham, K.J.; Lee, C.; Bigham, A.W.; Tabor, H.K.; Dent, K.M.; Huff, C.D.; Shannon, P.T.; Jabs, E.W.; Nickerson, D.A.; et al. Exome sequencing identifies the cause of a mendelian disorder. *Nat. Genet.* **2010**, *42*, 30–35. [CrossRef] [PubMed]

99. Richards, S.; Aziz, N.; Bale, S.; Bick, D.; Das, S.; Gastier-Foster, J.; Grody, W.W.; Hegde, M.; Lyon, E.; Spector, E.; et al. Standards and guidelines for the interpretation of sequence variants: A joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med.* **2015**, *17*, 405–424. [CrossRef]

100. Collins, R.L.; Brand, H.; Karczewski, K.J.; Zhao, X.; Alfoldi, J.; Francioli, L.C.; Khera, A.V.; Lowther, C.; Gauthier, L.D.; Wang, H.; et al. A structural variation reference for medical and population genetics. *Nature* **2020**, *581*, 444–451. [CrossRef]

101. Lek, M.; Karczewski, K.J.; Minikel, E.V.; Samocha, K.E.; Banks, E.; Fennell, T.; O'Donnell-Luria, A.H.; Ware, J.S.; Hill, A.J.; Cummings, B.B.; et al. Exome Aggregation, Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **2016**, *536*, 285–291. [CrossRef]

102. Ng, P.C.; Henikoff, S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res.* **2003**, *31*, 3812–3814. [CrossRef]

103. Kircher, M.; Witten, D.M.; Jain, P.; O'Roak, B.J.; Cooper, G.M.; Shendure, J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* **2014**, *46*, 310–315. [CrossRef] [PubMed]

104. Adzhubei, I.A.; Schmidt, S.; Peshkin, L.; Ramensky, V.E.; Gerasimova, A.; Bork, P.; Kondrashov, A.S.; Sunyaev, S.R. A method and server for predicting damaging missense mutations. *Nat. Methods* **2010**, *7*, 248–249. [CrossRef] [PubMed]

105. Kim, S.; Jhong, J.H.; Lee, J.; Koo, J.Y. Meta-analytic support vector machine for integrating multiple omics data. *BioData Min.* **2017**, *10*, 1–14. [CrossRef]

106. Miosge, L.A.; Field, M.A.; Sontani, Y.; Cho, V.; Johnson, S.; Palkova, A.; Balakishnan, B.; Liang, R.; Zhang, Y.; Lyon, S.; et al. Andrews, Comparison of predicted and actual consequences of missense mutations. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, E5189–E5198. [CrossRef]

107. Ernst, C.; Hahnen, E.; Engel, C.; Nothnagel, M.; Weber, J.; Schmutzler, R.K.; Hauke, J. Performance of in silico prediction tools for the classification of rare BRCA1/2 missense variants in clinical diagnostics. *BMC Med. Genom.* **2018**, *11*, 1–10. [CrossRef]

108. Williams, K.; Carson, J.; Lo, C. Genetics of Congenital Heart Disease. *Biomolecules* **2019**, *9*, 879. [CrossRef] [PubMed]

109. Kuehl, K.S.; Loffredo, C.A. A cluster of hypoplastic left heart malformation in Baltimore, Maryland. *Pediatr. Cardiol.* **2006**, *27*, 25–31. [CrossRef] [PubMed]

110. Jenkins, K.J.; Correa, A.; Feinstein, J.A.; Botto, L.; Britt, A.E.; Daniels, S.R.; Elixson, M.; Warnes, C.A.; Webb, C.L. Noninherited risk factors and congenital cardiovascular defects: Current knowledge: A scientific statement from the American Heart Association Council on Cardiovascular Disease in the Young: Endorsed by the American Academy of Pediatrics. *Circulation* **2007**, *115*, 2995–3014. [CrossRef] [PubMed]

111. Strickland, M.J.; Klein, M.; Correa, A.; Reller, M.D.; Mahle, W.T.; Riehle-Colarusso, T.J.; Botto, L.D.; Flanders, W.D.; Mulholland, J.A.; Siffel, C.; et al. Ambient air pollution and cardiovascular malformations in Atlanta, Georgia, 1986–2003. *Am. J. Epidemiol.* **2009**, *169*, 1004–1014. [CrossRef]

112. Colliva, A.; Braga, L.; Giacca, M.; Zacchigna, S. Endothelial cell-cardiomyocyte crosstalk in heart development and disease. *J. Physiol.* **2020**, *598*, 2923–2939. [CrossRef] [PubMed]

113. Alkan, F.; Wenzel, A.; Anthon, C.; Havgaard, J.H.; Gorodkin, J. CRISPR-Cas9 off-targeting assessment with nucleic acid duplex energy parameters. *Genome Biol.* **2018**, *19*, 1–13. [CrossRef]

114. Berthelot, C.; Villar, D.; Horvath, J.E.; Odom, D.T.; Flicek, P. Complexity and conservation of regulatory landscapes underlie evolutionary resilience of mammalian gene expression. *Nat. Ecol. Evol.* **2018**, *2*, 152–163. [CrossRef] [PubMed]

115. van Nisselrooij, A.E.L.; Lugthart, M.A.; Clur, S.A.; Linskens, I.H.; Pajkrt, E.; Rammeloo, L.A.; Rozendaal, L.; Blom, N.A.; van Lith, J.M.M.; Knegt, A.C.; et al. The prevalence of genetic diagnoses in fetuses with severe congenital heart defects. *Genet. Med.* **2020**, *22*, 1206–1214. [CrossRef] [PubMed]

116. Chaix, M.A.; Andelfinger, G.; Khairy, P. Genetic testing in congenital heart disease: A clinical approach. *World J. Cardiol.* **2016**, *8*, 180–191. [CrossRef] [PubMed]

117. Sud, A.; Turnbull, C.; Houlston, R. Will polygenic risk scores for cancer ever be clinically useful? *NPJ Precis. Oncol.* **2021**, *5*, 1–5. [CrossRef]

118. Lewis, A.C.F.; Green, R.C. Polygenic risk scores in the clinic: New perspectives needed on familiar ethical issues. *Genome Med.* **2021**, *13*, 1–10. [CrossRef]