

Review

Genetic Profiling for Risk Reduction in Human Cardiovascular Disease

Megan J. Puckelwartz¹ and Elizabeth M. McNally^{1,2,*}

¹ Department of Medicine, University of Chicago, Chicago, IL 60637, USA;
E-Mail: mjroy@uchicago.edu

² Department of Human Genetics, University of Chicago, Chicago, IL 60637, USA

* Author to whom correspondence should be addressed; E-Mail: emcnally@uchicago.edu;
Tel.: +1-773-702-2672; Fax: +1-773-702-2681.

Received: 16 January 2014; in revised form: 26 February 2014 / Accepted: 27 February 2014 /
Published: 12 March 2014

Abstract: Cardiovascular disease is a major health concern affecting over 80,000,000 people in the U.S. alone. Heart failure, cardiomyopathy, heart rhythm disorders, atherosclerosis and aneurysm formation have significant heritable contribution. Supported by familial aggregation and twin studies, these cardiovascular diseases are influenced by genetic variation. Family-based linkage studies and population-based genome-wide association studies (GWAS) have each identified genes and variants important for the pathogenesis of cardiovascular disease. The advent of next generation sequencing has ushered in a new era in the genetic diagnosis of cardiovascular disease, and this is especially evident when considering cardiomyopathy, a leading cause of heart failure. Cardiomyopathy is a genetically heterogeneous disorder characterized by morphologically abnormal heart with abnormal function. Genetic testing for cardiomyopathy employs gene panels, and these panels assess more than 50 genes simultaneously. Despite the large size of these panels, the sensitivity for detecting the primary genetic defect is still only approximately 50%. Recently, there has been a shift towards applying broader exome and/or genome sequencing to interrogate more of the genome to provide a genetic diagnosis for cardiomyopathy. Genetic mutations in cardiomyopathy offer the capacity to predict clinical outcome, including arrhythmia risk, and genetic diagnosis often provides an early window in which to institute therapy. This discussion is an overview as to how genomic data is shaping the current understanding and treatment of cardiovascular disease.

Keywords: next generation sequencing; cardiovascular disease; genetic profiling; whole genome sequencing

1. Introduction

Next generation sequencing has revolutionized the study of human genome variation and has the capacity to greatly influence health care decision making. The Human Genome Project, as conceived, was to sequence the first human genome in ~15 years at a cost of almost \$3 billion using traditional dideoxy chain termination sequencing. In under 10 years, massively parallel next generation sequencing has led to the routine sequencing of exomes and whole genomes. Now achieved in weeks and at a cost that is multiple orders of magnitude less than the first genome, personalized genetic information is now widely available. These rapid advances in sequencing technology require new ways of collecting, analyzing, and disseminating genomic data. Herein, we discuss the ways that genomic information is currently being applied and how that data is shaping the ability to understand and treat cardiovascular disease (CVD).

Genetic variation is considered a contributory component for nearly all disease, whether single gene familial disorders or more common, complex traits with multiple gene involvement. Single gene or “Mendelian” disorders can be attributed to one gene as both necessary and sufficient to cause a large component of the disease phenotype. With complex traits, the gene-gene and gene-environment interactions are multifactorial. CVD consists of both single gene-familial disorders and common, complex disease. CVD is a major health concern affecting over 80,000,000 people in the U.S. alone [1]. CVD extends to heart failure and cardiomyopathy, heart rhythm disorders, atherosclerosis and thromboembolic events, aneurysm and others disorders. Familial aggregation and twin studies demonstrate that most, if not all, of CVD is heavily influenced by a genetic component [2–4].

2. Genetic Variation in CVD

Beginning in the 1980s, family-based linkage analysis was used to identify regions of the genome responsible for monogenic disease. The success of these methods required large families with penetrant phenotypes. Polymorphic genetic markers segregated with the phenotype of interest in large multi-generational families to identify chromosomal regions bearing the causal genes of interest [5]. Such familial linkage studies were highly successful in identifying genes for multiple forms of CVD. In 1989, linkage analysis defined the chromosomal location responsible for hypertrophic cardiomyopathy [6]. The next year, this data was used to identify mutations in the causative gene, *MYH7*, encoding β -myosin heavy chain [7]. Genetic determinants for Long QT syndrome, multiple cardiomyopathies, Marfan’s disease, and forms of congenital heart disease were identified highlighting both extensive locus and allelic heterogeneity [8–14]. However, these methods remain limited by the need for large families, a feature often not available since CVD confers survival disadvantage. Furthermore, much of CVD is under the influence of multiple genetic loci, and therefore requires alternative statistical methods and larger phenotypically and genetically characterized cohorts [15]. The HapMap project annotated the location of millions of single-nucleotide polymorphisms (SNPs)

and took advantage of the long haplotype structure of the human genome [16]. Concurrently, commercially available platforms such as SNP arrays were developed that allowed simultaneous sampling of hundreds of thousands of SNPs paving the path for genome-wide association studies (GWAS). GWAS, which correlates SNPs with disease phenotypes, does not require a specific mode of inheritance and takes advantage of the extensive linkage disequilibrium (LD) in the human genome. In order to have enough statistical power to detect correlation, these large-scale association studies typically assess thousands to millions of SNPs across the genome in hundreds to thousands of cases and controls. According to the National Human Genome Research Institute (NHGRI) Catalog of Published Genome-Wide Association Studies [17] over 2800 strong SNP associations have been identified ($p < 1 \times 10^{-8}$) to date, and many of these are CVD-associated traits.

2.1. The Overlap between GWAS Hits and Monogenic Disease in CVD

CVD related phenotypes are well suited for GWAS because many CVDs have readily quantifiable discernable traits. Intriguingly, many GWAS “hits” overlap considerably with the same genes already linked to the disease through familial linkage studies. For example, Newton-Cheh and colleagues conducted a meta-analysis of three GWAS from ~14,000 individuals to examine the duration of QT interval from surface electrocardiograms [18]. QT duration reflects electrical depolarization and repolarization of the cardiac ventricles. A long QT interval is a biomarker for arrhythmias and a risk factor for sudden death. Non-familial QT disorders are still highly heritable ($h^2 \approx 0.35$), indicating a genetic component. GWAS identified 10 loci with significant ($p < 5 \times 10^{-8}$) association with QT interval. Five loci were those known to be involved in Mendelian long-QT syndromes, while the other five loci were genes that offer additional insights into variation at the QT interval. In total, the variation at these 10 loci accounted for 5.4%–6.5% of variation in the QT interval, which is quite high by GWAS standards. Genetic testing for the Mendelian form of long QT currently identifies mutations in ~75% of probands, so the additional GWAS loci may represent new candidate genes for mutation screening in familial long QT disorders.

BAG3 (B-cell lymphoma 2-associated athanogene 3) is another example of GWAS results informing rare, Mendelian disease. In 2011, Villard and colleagues performed GWAS to identify loci contributing to sporadic dilated cardiomyopathy. Dilated cardiomyopathy (DCM) is exemplified by left ventricular dilation and systolic dysfunction, and is a major cause of heart failure and the principle indicator for heart transplant [19]. DCM has a high heritable component with 20%–35% of DCM patients having an affected first-degree relative [20]. More than 50 genes have been implicated in familial monogenic DCM [21–23]. GWAS was performed with DNA from 1179 sporadic (non-familial) DCM patients and 1108 controls using ~500,000 SNPs [19]. The authors identified a DCM-associated non-synonymous SNP (p. C151R) in the coding region of *BAG3*. *BAG3* is a co-chaperone that regulates HSP70 [24]. Further analysis of *BAG3* non-synonymous SNPs found another higher frequency SNP also associated with DCM. The authors investigated *BAG3* variation in familial DCM based on both the apparent association of *BAG3* non-synonymous SNPs with sporadic DCM and previously reported linkage with familial DCM between markers on chromosome 10 that include the *BAG3* locus [25]. In a cohort of 168 cases from DCM families, the authors identified additional likely pathogenic mutations in *BAG3*. Features of DCM were identified in 16 of 18 mutation carriers in the

cohort [19]. In the same year, Norton and colleagues identified *BAG3* mutations in familial DCM [26]. The authors also created a *BAG3* knockdown zebrafish model that recapitulated the DCM and heart failure found in patients [26]. Together, the GWAS and familial data implicate *BAG3* in DCM and indicate that genes can harbor common variation that influences risk of disease in sporadic cases and rare variation that accounts for familial disease.

2.2. The Missing Heritability of GWAS

Despite the overlap of GWAS findings with monogenic disease, GWAS associations often account for only a small proportion of genetic variation. Also, the vast majority of variants identified by GWAS does not explain the high heritability or reveal the causal mechanism for the cardiovascular phenotype in question [27–29]. Using GWAS, McPherson and colleagues identified an interval on chromosome 9q21 that consistently associated with coronary heart disease in more than 23,000 participants from 6 independent cohorts [30]. Homozygotes for the risk allele have a 30%–40% increased risk for coronary artery disease (CAD). This finding remains perplexing as this region on 9q21 has no annotated genes and is not associated with known CAD risk factors [30]. This same chromosome 9q21 region has also been associated with myocardial infarction in a GWAS with 4587 cases and 12,676 controls [31]. Homozygotes for the allele have 1.64 times increased risk for myocardial infarction compared to noncarriers. Despite the evidence that this region is important for CVD pathophysiology, no disease mechanism has yet been identified.

GWAS is ultimately based on the idea that common diseases are caused by common genetic variants, each with small effect. Through additive and interactive effects, in conjunction with the environment, GWAS variants explain disease [27,32,33]. Recently, the common disease common variant hypothesis has been called into question due to the observation of missing heritability [34,35]. Missing heritability refers to the proportion of genetic variance that is not explained by the effect of common variants identified by GWAS. Several explanations have been suggested to explain missing heritability. It is possible that the number of variants responsible for a trait has been significantly underestimated, and that many more yet identified variants with very small effect sizes must be discovered. Another possibility is the presence of rare variants with larger effect size. Such rare variants are undetectable using present day SNP arrays, which are biased towards SNPs with allele frequencies close to 50%. There is also a possibility that missing heritability arises from variation caused by structural variants in the genome, also difficult to detect with SNP arrays. Lastly, there is also the possibility that gene-gene interactions and gene-environment interactions are of major importance, but are not appropriately modeled by current methods.

GWAS typically interrogate SNPs with a minor allele frequency (MAF) > 5% while largely ignoring variants with lower population-based frequencies (MAF: 0.5%–5%) and those SNPs that are rare (MAF < 0.5%). The rationale for ignoring these variants relates to the strong linkage disequilibrium (LD) in the human genome. The human genome has an estimated mutation rate of approximately 1.4×10^{-8} or approximately 40 new mutations per generation. Projected over the current population of ~7 billion, the world currently has 300–400 million new mutations this generation [36]. Within the coding region alone, there are ~13,000 nonsynonymous variants per genome [37,38]. The National Heart Lung and Blood Institute sponsored exome sequencing project

(NHLBI ESP) sequenced ~15,000 human-protein coding genes in >2000 individuals [39]. This study revealed an abundance of rare variants that were often population specific, potentially offering some support that rare variation explains some component of missing heritability. Alleles that confer high risk of disease are subject to negative selection pressure and would not reach high population frequencies. However, many different rare alleles in the same genes or gene pathways would induce the same phenotype across the population despite the “rare” genetic etiology. Johansen and colleagues recently performed GWAS and resequencing to determine mutational burden of rare variants in individuals with hypertriglyceridemia *versus* control subjects [40]. Hypertriglyceridemia is polygenic in nature and confers risk for cardiovascular disease. In general, GWAS variants explain <10% of variation for lipid traits [41,42]. Loci associated with GWAS signals were assessed for rare variation by focusing on protein-coding sequences of four GWAS-associated genes [40]. The authors identified a significant number of rare variants in individuals with hypertriglyceridemia compared to controls. An additional more restricted study, which analyzed only rare variants unique to cases and removing all reported variants without functional deficits, also confirmed a greater mutational burden in cases compared to controls [40]. These data indicate that GWAS-identified genes may carry rare variation that contributes to the heritability of a complex trait, and rare SNPs with relatively large effects on common disease may not be identified by GWAS studies. The underlying assumption that genotypes can be inferred between common alleles in strong LD may be flawed. Inferring genotypes based on strong LD may misestimate variation across the genome. While haplotype structure may be maintained with several to many common SNPs still inherited together, rare variation may occur between these SNPs, reducing the predictive value of the common SNPs. This rare variation may be the result of a higher than expected mutation rate and/or population structure.

2.3. Rare Variation as a Cause of CVD

Next generation sequencing (NGS) provides a method to identify rare genetic variation. Massively parallel, array-based sequencing dramatically reduced cost and increased efficiency of DNA sequencing. Depending on the platform, sequencing reads range from ~35 base pairs to up to ~1000 base pairs. As the generation of sequence has become far more facile with NGS, large-scale alignment and interpretation has become the rate limiting step. Bioinformatics tools are available for alignment to the referent genome and calling variants (reviewed in [43,44]). A number of efforts are currently underway to catalog human genetic variation. The 1000 Genomes project has sequenced 1092 individuals from 14 populations using a combination of low coverage whole genome sequencing and exome sequencing [45]. It is estimated that this dataset captures 98% of SNPs at a frequency of 1%, finding 38 million SNPs, 59% of which were novel [45]. The NHLBI ESP provided whole exome sequencing on >6500 individuals, including 180,000 exons in 23,000 genes. The ESP is derived from diverse, well-phenotyped populations, including the Atherosclerosis Risk in Communities (ARIC) study, the Coronary Artery Risk Development in Young Adults (CARDIA) study, the Cardiovascular Health Study, the Framingham Heart Study, the Jackson Heart Study, and the Multi-Ethnic Study of Atherosclerosis. The data from this project will examine the genetic contribution to early-onset myocardial infarction, low-density lipoprotein cholesterol, body mass index/type 2 Diabetes mellitus, blood pressure, and ischemic stroke.

These databases are not only providing a rich dataset to identify rare variation, but they are also providing better allele frequencies across diverse populations for alleles once thought to be rare and pathogenic. Jabbari and colleagues examined the NHLBI ESP database ($n = 6503$) for variants previously associated with Catecholaminergic Polymorphic Ventricular Tachycardia (CPVT) and compared the frequency of those variants with the expected prevalence of CPVT in the population [46]. CPVT is a rare, lethal, hereditary cardiac disease characterized by fatal ventricular arrhythmias in the absence of structural defects of the heart or abnormal electrocardiographic findings [47]. Eleven percent of variants previously associated with CPVT were found in the ESP population, corresponding to a 1:150 prevalence of CPVT in the population, much higher than the known 1:10,000 prevalence. The 1000 Genomes (1KG) database was evaluated for the presence of predicted and previously reported pathogenic variation in three genes associated with cardiomyopathy (*MYH7*, *MYBPC3* and *TTN*) [48]. Nine percent of the population was identified as having a pathogenic variant (9%), which exceeds population prevalence estimates for dilated and hypertrophic cardiomyopathy (0.04%–0.2%, respectively). Similar studies have been performed for other cardiovascular diseases including Brugada Syndrome, cardiac channelopathies and arrhythmogenic right ventricular cardiomyopathy [49–51]. Since in every case the predicted pathogenic variation vastly exceeds the known disease prevalence, it suggests that on their own, these variants are not sufficient to cause disease. Whether these variants represent “at risk” genotypes for developing milder forms of disease is not known and requires prospective studies.

3. The Genetics of Dilated Cardiomyopathy

Cardiomyopathy is marked by a morphologically abnormal ventricle and frequently is associated with heart failure. Cardiomyopathy is divided into four groups: dilated (DCM), hypertrophic (HCM), restrictive (RCM), and arrhythmogenic right ventricular (ARVC), and the most common mode of inheritance is autosomal dominant. Depending on screening methods, nonischemic DCM, defined as DCM not arising from myocardial infarct or ischemia, is familial in 25%–50% of cases [52]. Greater than 50 genes have been implicated in familial DCM, and the majority of mutations are phenocopies as there are no outward clinical signs that predict the specific gene mutation. Identification of the genetic cause of cardiomyopathy is clinically important because of the high incidence of sudden death and clinical progression, which can be medically managed. In 2007, genetic testing relied on commercially available panels that interrogated only 5 genes using traditional sequencing. In 2011, Meder and colleagues developed an array-based panel that enriched for coding regions of 29 known and 18 novel, potential cardiomyopathy genes, followed by next-generation sequencing [53]. Current panels now include >50 genes [54], and the estimated sensitivity in DCM for detecting a pathogenic mutation is under 50% (Partners Healthcare [55]).

3.1. Next Generation Sequencing Identifies *TTN* as a Major Contributor to DCM

Next-generation sequencing facilitated the screening of *TTN* for DCM-causing variants. The *TTN* gene includes >350 exons and encodes the giant sarcomere protein titin, which ranges in size from ~27,000 to ~33,000 amino acids depending on isoform, making it the largest human protein [56]. Together, two titin molecules span the sarcomere, providing both passive and active contractile forces [57–60].

Previous work has linked *TTN* to dilated cardiomyopathy in families, but extensive screening was limited due to its large size [61–63]. Herman and colleagues developed an array to capture *TTN* exons and sequence *TTN* in patients with DCM ($n = 312$) [64]. *TTN* truncating mutations accounted for approximately 25% of familial DCM, but had a minimal contribution to hypertrophic cardiomyopathy (~1%). Approximately 30% of *TTN* truncating variants identified were putative splice site disrupting mutations whose effect on function can be difficult to assess *in silico*.

3.2. Beyond Panel Based Sequencing for Cardiomyopathy and Beyond

Despite the inclusion of *TTN*, the sensitivity for detecting a DCM mutation remains at just under 50%. There are several explanations for the missing variation. First, there are likely novel genes not yet associated with cardiomyopathy; second, certain genetic variation may not be readily detectable with NGS and SNP analysis. For example, nucleotide repeat expansions and structural variation is not commonly determined by NGS, as analytic methods are biased toward SNP analysis and small insertions and deletions; Third, pathogenic variation may arise from combinations of pathogenic variation, and analysis is generally biased towards finding a single pathogenic variant. This bias reflects that most families with inherited cardiomyopathy have autosomal dominant inheritance; Fourth, pathogenic variation may be non-coding and at this point, these regions are not captured by gene panels. To combat these problems and provide a more comprehensive variant profile, whole exome sequencing (WES) and whole genome sequencing (WGS) are being applied to identify disease-causing variation for many different diseases (Table 1). WES interrogates the coding portion of the genome; approximately 1%–2% of nuclear DNA, although at higher coverage than-comparably priced WGS [65]. Interrogating only a small portion of the genome, as in WES, is less expensive than WGS, and it is currently the most-readily interpreted, as approximately 85% of Mendelian-disease causing mutations cause changes in the coding sequence of the genome [66]. WES relies on commercially available sequence-capture arrays to enrich for the coding subset of genomic DNA, followed by massively parallel, next-generation sequencing of the enriched fragments. The choice of exome kit is an important consideration as the exome is approximately 30 megabases, and exome capture kits interrogate anywhere from 50 to 100 megabases, depending on the provider. Most of the additional sequences are untranslated regions (UTRs) that may be important for disease pathogenesis.

Table 1. Comparison of Panel, whole exome sequencing (WES) and whole genome sequencing (WGS).

	Panel	WES	WGS
Variation in Known Genes	yes	yes	yes
Novel Gene Identification	no	yes	yes
Structural Variation	no	limited	yes
Non-coding Variation	no	limited	yes
Repeat testing required if first pass negative	yes	yes	no

3.3. Exome Sequencing of Multiple Family Members Improves Identification of Pathogenic Variation

Campbell and colleagues used exome sequencing on three members of a large multi-generational family with classical DCM. After sequencing, variants were filtered by frequency and protein prediction algorithms [67]. Eight potentially causative mutations were shared across the three family members, significantly reducing the potential variants to be considered. Variants were tested for segregation across the other family members and only one variant segregated with disease, *TNNT2* R173W, a known cardiomyopathy gene. These data are particularly convincing as the variant segregates across all affected members of the family, including fourth-degree relatives [67]. The underlying assumption with this analysis is that a single variant accounts for disease in all family members, an assumption that may or may not hold true. In 2013, Wells and others used WES in a large, multi-generational family with DCM of unknown etiology [68]. The proband had undergone extensive unrevealing panel testing for DCM. The authors selected 3 distantly related affected family members for WES. Distantly related family members that are obligate carriers of the same mutation should share fewer variants by descent than closely related members, allowing for easier filtering of potentially causative variants. Variants were then filtered for rarity, functional significance, conservation and autosomal-dominant inheritance [68]. Variants were also prioritized using the VAAST tool, a probabilistic search tool that combines conservation, amino acid substitution chemistry, and frequency data to build a unified likelihood-framework to identify damaged genes and disease-causing variants [69]. Heterozygous, nonsense, nonsynonymous and splice site variants shared between the 3 affected candidates were filtered based on rare frequency in 1 KG and ESP, leaving 26 candidates for analysis. Wells and colleagues then compared these variants to variants identified in ~70 exomes previously sequenced by their laboratory. Variants identified in multiple exomes were removed, leaving 2 putative variants. This comparison allowed for removal of false positives that may be inherent to some sequencing platforms and variant calling pipelines. An *RBM20* variant was identified in an unrelated patient with familial-DCM, consistent with its role in causing disease. The authors confirmed segregation within the larger pedigree of the variant in *RBM20*, a recently identified cardiomyopathy gene [70], providing statistical support for *RBM20* causing DCM. It is interesting to note that this finding is largely based on frequency in both the general population and in a cohort already sequenced by the laboratory. The authors did perform filtering that considered evolutionary conservation (Genomic Evolutionary Rate Profiling) and functional effects (Polyphen2) but these tools did not reduce the list as extensively as population and cohort frequency combined (8 variants *versus* 2 variants, respectively [69,71,72]). These data indicate that exome sequencing followed by extensive filtering, in conjunction with segregation analysis can identify rare, DCM-causing variation in known cardiomyopathy genes.

3.4. Identifying Cardiomyopathy Modifier Loci Using Broad Based Sequencing

With large gene panels or WES/WGS additional, disease-modifying variation can be identified. Intra-familial variability including age of onset, severity and penetrance is a hallmark of DCM, but the loci that modify DCM phenotypes have not been well elucidated [73,74]. Roncarati and colleagues used WES to investigate clinical variability in an extended family with 14 subjects that included four family members with severe DCM that required heart transplant in early adulthood [75]. WES was

performed on three severely affected and one unaffected family member. Variants were filtered for rarity, predicted pathogenicity and inheritance. The filtering process left a list of only 28 variants that were further filtered through the Human Phenotype Ontology project, which uses formal ontology to capture phenotypic information to identify relationships between different genes and phenotypes [76]. Through this analysis, the authors identified eight genes with variation associated with Mendelian disease [75], and two of eight *LMNA* and *TTN*, were known to cause DCM. A missense *LMNA* mutation, previously identified in an unrelated DCM patient, was confirmed in all affected family members [77]. A *TTN* variant was identified in five family members, four of whom were severely affected, while the fifth is likely too young to yet be symptomatic. Doubly heterozygous family members had a more severe clinical course than the *LMNA*-only family members, indicating that the *TTN* variant modifies the clinical progression [75]. This mutational stratification is expected to prove useful for assessing clinical risk and guiding treatment. Broad based sequencing through gene panels or WES/WGS is now positioned to outline the contribution of multiple variants to DCM development.

3.5. WES/WGS Can Identify New Genes for Cardiomyopathy

In 2011, Theis and colleagues used genome-wide mapping and exome sequencing in a consanguineous family with autosomal recessive DCM [78]. Genome-wide linkage analysis was performed in nineteen family members and a significant LOD score was identified on chromosome 7q21, in a region containing >250 genes. Exome sequencing was performed on 2 affected siblings, and variants were called and filtered without taking into account the linkage peak on 7q21. Synonymous, intergenic and intronic variants were removed from further consideration. Variants were filtered based on presence in 1 KG, HapMap, and in the authors' collection of exome sequences. Heterozygous SNPs were excluded due to the autosomal recessive inheritance mode. This extensive filtering left only 3 homozygous missense variants and only 1 was not present in unaffected family members, a mutation in *GATAD1* that maps to the already identified linkage region on 7q21. *GATAD1* encodes the GATA zinc finger domain-containing protein 1 which is ubiquitously expressed and is thought to bind to a histone modification site that regulates gene expression [79]. Immunohistochemistry with an antibody to GATAD1 revealed an abnormal staining pattern in the proband heart compared to control heart [78]. These data implicate *GATAD1* in the pathogenesis of DCM indicating that exome sequencing can be used to identify novel DCM genes.

3.6. Limitations of WES

While fruitful, exome sequencing does have limitations. There are inherent technical limitations associated with the method. WES requires a capture step, which is limited by design of capture oligonucleotides. Not all genes or exons are adequately annotated and therefore will not be properly included in the methods to capture exons. Furthermore, there can be inconsistencies in capture resulting in poorly covered exons and off-target sequencing. Capture efficiency only approximates 70%–80% in part due to the high GC content of exonic sequence. Probably the most notable limitation is that only 1%–2% of the entire genome is evaluated. Because approximately 85% of described Mendelian mutations occur in the coding regions of genes, it is assumed that Mendelian disease is more likely to be caused by mutations in protein coding exons than in non-coding sequences.

However, over a third of Mendelian diseases reported in OMIM have no known molecular basis. It is reasonable to conclude that some of these missing mutations are either structural variants or that they occur in non-coding regions of the genome, exome sequencing is not suited to interrogate either of these possibilities.

Copy number variants (CNVs) are regions, >50 bp in length, that differ from the expected diploid status [80]. CNVs are an important component of genomic variation in humans and some contribute to disease including cardiovascular disease [81–83]. A recent study by Norton and colleagues identified a large deletion in *BAG3* in a large multi-generational family with DCM [26]. Whole exome sequencing was performed on 4 of the affected family members and comparative genomic hybridization was performed on the proband to detect copy number variations. The technical limitations of WES prevented identification of the large *BAG3* deletion (>8.7 kb). Algorithms are being created to aid in the use of exome sequencing for the detection of copy number variation [84,85]. However, these methods come with a variety of limitations and cannot detect other types of structural variation such as uniparental disomies or chromosomal rearrangements, both exceedingly important for disease pathogenesis. WGS, as opposed to WES, may offer a better method to detect some structural variants, but may require improved analytic tools for specificity and sensitivity of structural variant detection.

4. WGS as a Tool to Investigate Non-Coding Variation for CVD

Perhaps even more important to understanding disease etiology is the investigation of non-coding variation. Over 98% of the genome is non-coding, and WGS captures nearly 100 fold more of this information compared to WES. However, the interpretation of non-coding variation is currently far more challenging than the interpretation of coding variants. Often referred to as the “dark matter” of the genome, these regions can include microRNAs, long non-coding RNAs, splice variants and regulatory elements that can directly cause or modulate disease phenotypes [37]. miRNAs are important regulators of heart function and recent studies have revealed miRNA misexpression in human cardiac disease and animal models of heart failure [86–88]. For example, miR-208 is encoded by an intron within *MYH6*, which encodes α myosin heavy chain and is in close proximity to *MYH7* [87,88]. miR-208 null mice do not hypertrophy in response to cardiac stress and null mice do not upregulate *Myh7* [89,90]. Silencing of miR-208 reduces cardiac remodeling, deterioration of heart function, and improves survival in a rat model of heart failure, while overexpression of miR-208 in cardiomyocytes leads to cardiomyocyte hypertrophy [90,91]. Another miR, miR-1 is the most highly expressed miRNA in the murine heart [88,92]. miR-1 targets *HAND2*, a transcription factor important for expansion of ventricular cardiomyocytes. Deletion of miR-1 results in 50% perinatal lethality due to ventricular septal defects [88]. The majority of surviving mice exhibit sudden death due to conduction defects. Overexpression of miR-1 in embryonic cardiomyocytes caused thin-walled ventricles leading to death at embryonic day 13.5 [86]. These data underscore the importance of miRNAs in cardiac phenotypes. Variation in these and other non-coding regions of the genome may play a vital role in the disease process. Annotation of the non-coding genome is currently underway to aid in the interpretation of these variants. The ENCODE project (Encyclopedia of DNA Elements) has assigned biochemical functions to 80% of the genome [93]. Only twelve percent of SNPs identified by GWAS as disease-associated are located in the vicinity of a protein-coding region even though SNPs in

coding regions are over-represented on SNP arrays. However, over 60% of disease-associated SNPs identified by GWAS lie within functional, non-coding regions, especially in promoters and enhancers [94].

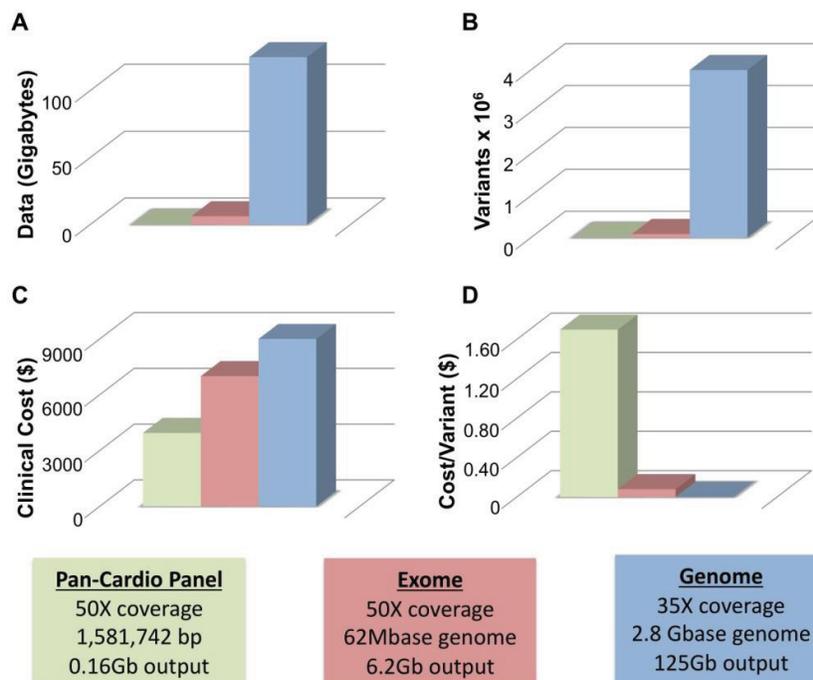
4.1. WGS Has Greater Sensitivity than WES

WGS is only now emerging as an alternative to WES since higher cost and more complex analysis limited uptake of WGS. Recently, WGS was used to identify a putative causative variant in a family with two children affected by a previously unreported disease defined by cardiomyopathy and progressive muscle weakness [95]. Wang and colleagues performed WES on one sibling and WGS on the other sibling. WES was performed to a depth of 118× with >90% target regions covered by ≥ 10 reads, while WGS was performed to a depth of 81×. Variants were filtered using ANNOVAR, which relies on frequency and functional variation [96]. After validation with Sanger sequencing and transmission pattern testing, only two genes remained as candidates. One, TAF1L, is homologous to TAF(II)250 and is specifically expressed in testis. The other candidate, *RBCK1*, codes for an E3 ubiquitin-protein ligase. In this family, *RBCK1* had two truncating mutations, each inherited from one parent [95]. *RBCK1* was considered a good candidate for both its rarity and the involvement of other ubiquitin-ligase proteins in muscle disease. The WES data set failed to reveal the *RBCK1* variants despite good coverage over the targeted regions including the exons of *RBCK1*. Upon reanalysis, coverage was very low (2 and 4 reads respectively) for the two mutations, with only one read containing a mutation [95]. Further investigation revealed a high GC content surrounding these mutations [97]. Previous studies have also noted that uneven exome coverage has resulted in filtering of disease genes [98]. This study serves as a proof of principle that whole genome sequencing can be used to identify rare Mendelian, cardiomyopathy phenotypes, and, in some instances, may be more sensitive than WES.

4.2. Limitations of WGS

Two of the major limitations of WGS are size and cost. To achieve average coverage $\sim 35\text{--}40\times$ with WGS requires approximately 125 Gb of generated sequencing data. Figure 1A compares the amount of data generated from panel (blue), exome (red) and whole genome sequencing (green). WGS produces an order of magnitude more data than WES. All three technologies call variants proportionate to the amount of data generated with WGS calling ~ 4 million variants per genome, WES calling ~ 90 K, and gene panels calling far fewer (Figure 1B). At this time, clinical WGS is more expensive than either WES or panel sequencing at approximately \$9000–\$9500 for WGS, \sim \$7000 for WES and \sim \$4000 for a pan-cardiomyopathy panel. The price for both WES and WGS is considerably more than the cost of a panel. However, this only remains true if a pathogenic mutation is identified in the first panel. The value of each test can be thought of in terms of the cost per variant identified, and with this metric WGS is a better value (Figure 1D). Panel sequencing is \sim \$1.70, WES is \sim \$0.08 (8¢) and WGS is \sim \$0.002 (0.2¢) per variant detected. While only a handful of variants may be germane to identifying the cause of an individual's primary cardiomyopathy, the other sequence data remains available where it may provide useful guidance for life-style and medical decisions.

Figure 1. Size and Cost Considerations of Next-Generation Sequencing. (A) The amount of data generated by a typical cardiomyopathy gene panel of ~50 genes (green), whole exome sequencing (red) and whole genome Sequencing (blue) is shown; (B) The approximate number of variants produced by each method is indicated; (C) The Clinical cost of each method ranges from ~\$4000 (cardiomyopathy gene panel, green) to ~\$9500 (whole genome sequencing, blue); (D) The cost per variant is greatly reduced for WGS (\$0.002, blue) *versus* WES (\$0.08, red) and gene panel-based sequencing (\$1.70, green). Boxes indicate parameters used to calculate values in A–D including coverage, base pairs interrogated and total output.

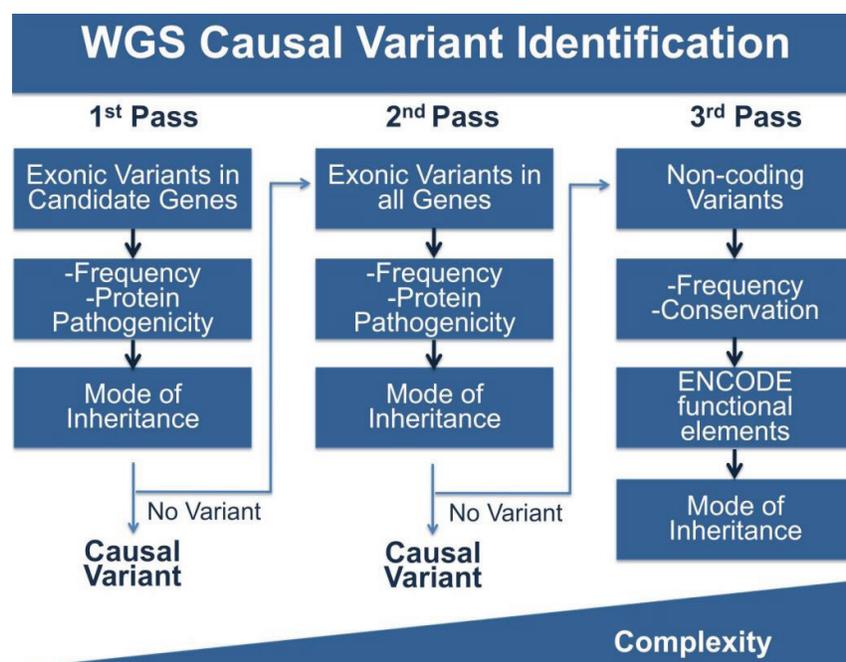


4.3. Multi-Pass Filtering Methods Allow for More Efficient Variant Identification

WGS produces vastly more data than either panels or WES, and this is the double-edged sword of broad based sequence analysis. To cope with this problem we (and others) have adopted stepwise, multi-pass filtering methods (Figure 2). In the first pass, candidate genes are analyzed for the phenotype of interest, in this case cardiomyopathy. Typically, exonic variants in candidate genes are filtered for frequency and for potential protein pathogenicity. A number of tools are freely available that predict the impact of amino acid changes on protein structure and function including PolyPhen-2, GERP, SIFT, PhastCons, Panther and Conseq [72,99–103]. MaxEnt can be used to score the strength of splice site variants [104]. If filtering candidate genes does not produce meaningful variation, the search can be expanded to less attractive candidates or to all rare protein coding variation and finally to non-coding variation. Non-coding variants can be filtered for frequency, however this is more challenging as many frequency databases are biased towards coding regions. Conservation of sequences across multiple species may provide clues about the selection acting on a sequence. The data generated by the ENCODE project will also provide information about variation in functional elements. The complexity of analysis increases with each pass. While this approach approximates panel and WES analysis, it is

an improvement in several ways. With this model, WGS only needs to be performed once, while patients that are panel negative will need additional sequencing. In the case of WES, exome capture kits are often updated due to changing gene annotations and the inclusion of newly understood non-coding sequence, requiring retesting with new kits. WGS is less likely to need additional sequencing, and moreover, may provide additional data that over a lifetime can be used to inform not only the primary health concern for which data was collected, but medical choices throughout a patient's life.

Figure 2. Pipeline for WGS Variant Identification. WGS produces ~4 million variants per genome and requires extensive filtering to identify variants of interest. Shown here is a potential pipeline to identify variants. The first pass of the pipeline entails only reviewing variants in the coding regions of genes of interest and filtering by frequency, protein pathogenicity, and mode of inheritance (segregation in available family members). If no variant is identified, a second pass includes the same filtering steps, but on variants in all coding regions. The third pass includes analysis of non-coding variation using frequency, conservation and ENCODE annotation, along with mode of inheritance. The complexity of analysis increases with each pass.



5. Incidental Findings and Their Importance for CVD Related Phenotypes

Incidental findings are a concern with all genetic assessment and especially so for WES and WGS. However, incidental findings are not unique to genetic testing and are part of medical decision making for any mode of testing, including imaging and blood tests. WGS may provide many more incidental findings than any other tests available, and this has led to new recommendations for delivering results of incidental findings from genetic research and testing. In 2006, an NHLBI working recommended reporting research results to study participants when the risk of disease is significant and has important

health implications including sudden death or considerable risk of morbidity especially when therapeutic interventions are available [105]. In 2013, the American College of Medical Genetics and Genomics (ACMG) made recommendations for reporting incidental findings in exome and genome sequencing [106]. The ACMG recommended that laboratories performing sequencing should identify and report mutations in genes included on their minimal list. Notably, this list includes 24 phenotypes of which a third are cardiovascular disorders for which penetrance is high and clinical interventions are available [106]. Of the 57 genes for which recommendations were made to report incidental findings, more than half (34) were CVD-associated genes.

6. Conclusions

In the case of CVD genetic profiling, there are often medical management decisions that can reduce risk. This is the case whether the initial genetic profiling was done for assessing CVD risk or whether the genetic profiling was done to assess risk of other inherited diseases. For example, risks for cardiomyopathy and especially arrhythmias can be managed medically, with increased surveillance or even with device insertion. Risks for developing atherosclerosis or aneurysms can be mitigated through drug or even surgical intervention. Importantly, since CVD disorders can be associated with sudden cardiac death, the capacity to intervene based on genetic risk profiles is evident. With the improvement in genetic databases that are accompanied by robust phenotyping, it should be possible to more accurately predict risk for CVD.

Acknowledgments

This work was supported in part by National Institutes of Health NIH AR052646, NIH HL61322, NIH NS072027, and the Doris Duke Charitable Foundation.

Author Contributions

Megan J. Puckelwartz researched the topics and drafted the manuscript. Elizabeth M. McNally revised the manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Go, A.S.; Mozaffarian, D.; Roger, V.L.; Benjamin, E.J.; Berry, J.D.; Borden, W.B.; Bravata, D.M.; Dai, S.; Ford, E.S.; Fox, C.S.; *et al.* Executive summary: Heart disease and stroke statistics—2013 update: A report from the American Heart Association. *Circulation* **2013**, *127*, 143–152.
2. Marenberg, M.E.; Risch, N.; Berkman, L.F.; Floderus, B.; de Faire, U. Genetic susceptibility to death from coronary heart disease in a study of twins. *N. Engl. J. Med.* **1994**, *330*, 1041–1046.
3. Post, W.S.; Larson, M.G.; Myers, R.H.; Galderisi, M.; Levy, D. Heritability of left ventricular mass: The Framingham Heart Study. *Hypertension* **1997**, *30*, 1025–1028.

4. Adams, T.D.; Yanowitz, F.G.; Fisher, A.G.; Ridges, J.D.; Nelson, A.G.; Hagan, A.D.; Williams, R.R.; Hunt, S.C. Heritability of cardiac size: An echocardiographic and electrocardiographic study of monozygotic and dizygotic twins. *Circulation* **1985**, *71*, 39–44.
5. Botstein, D.; White, R.L.; Skolnick, M.; Davis, R.W. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.* **1980**, *32*, 314–331.
6. Jarcho, J.A.; McKenna, W.; Pare, J.A.; Solomon, S.D.; Holcombe, R.F.; Dickie, S.; Levi, T.; Donis-Keller, H.; Seidman, J.G.; Seidman, C.E. Mapping a gene for familial hypertrophic cardiomyopathy to chromosome 14q1. *N. Engl. J. Med.* **1989**, *321*, 1372–1378.
7. Geisterfer-Lowrance, A.A.; Kass, S.; Tanigawa, G.; Vosberg, H.P.; McKenna, W.; Seidman, C.E.; Seidman, J.G. A molecular basis for familial hypertrophic cardiomyopathy: A beta cardiac myosin heavy chain gene missense mutation. *Cell* **1990**, *62*, 999–1006.
8. Basson, C.T.; Bachinsky, D.R.; Lin, R.C.; Levi, T.; Elkins, J.A.; Soultz, J.; Grayzel, D.; Kroumpouzou, E.; Traill, T.A.; Leblanc-Straceski, J.; *et al.* Mutations in human TBX5 [corrected] cause limb and cardiac malformation in Holt-Oram syndrome. *Nat. Genet.* **1997**, *15*, 30–35.
9. Curran, M.E.; Splawski, I.; Timothy, K.W.; Vincent, G.M.; Green, E.D.; Keating, M.T. A molecular basis for cardiac arrhythmia: HERG mutations cause long QT syndrome. *Cell* **1995**, *80*, 795–803.
10. Dietz, H.C.; Cutting, G.R.; Pyeritz, R.E.; Maslen, C.L.; Sakai, L.Y.; Corson, G.M.; Puffenberger, E.G.; Hamosh, A.; Nanthakumar, E.J.; Curristin, S.M.; *et al.* Marfan syndrome caused by a recurrent de novo missense mutation in the fibrillin gene. *Nature* **1991**, *352*, 337–339.
11. Garg, V.; Kathiriya, I.S.; Barnes, R.; Schluterman, M.K.; King, I.N.; Butler, C.A.; Rothrock, C.R.; Eapen, R.S.; Hirayama-Yamada, K.; Joo, K.; *et al.* GATA4 mutations cause human congenital heart defects and reveal an interaction with TBX5. *Nature* **2003**, *424*, 443–447.
12. 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.
13. Schott, J.J.; Benson, D.W.; Basson, C.T.; Pease, W.; Silberbach, G.M.; Moak, J.P.; Maron, B.J.; Seidman, C.E.; Seidman, J.G. Congenital heart disease caused by mutations in the transcription factor NKX2-5. *Science* **1998**, *281*, 108–111.
14. Tartaglia, M.; Mehler, E.L.; Goldberg, R.; Zampino, G.; Brunner, H.G.; Kremer, H.; van der Burgt, I.; Crosby, A.H.; Ion, A.; Jeffery, S.; *et al.* Mutations in PTPN11, encoding the protein tyrosine phosphatase SHP-2, cause Noonan syndrome. *Nat. Genet.* **2001**, *29*, 465–468.
15. Lander, E.; Kruglyak, L. Genetic dissection of complex traits: Guidelines for interpreting and reporting linkage results. *Nat. Genet.* **1995**, *11*, 241–247.
16. International HapMap Consortium. The International HapMap Project. *Nature* **2003**, *426*, 789–796.
17. National Human Genome Research Institute (NHGRI) Catalog of Published Genome-Wide Association Studies. Available online: <http://www.genome.gov/gwasstudies/> (accessed on 25 February 2014).
18. Newton-Cheh, C.; Eijgelsheim, M.; Rice, K.M.; de Bakker, P.I.; Yin, X.; Estrada, K.; Bis, J.C.; Marciante, K.; Rivadeneira, F.; Noseworthy, P.A.; *et al.* Common variants at ten loci influence QT interval duration in the QTGEN Study. *Nat. Genet.* **2009**, *41*, 399–406.

19. Villard, E.; Perret, C.; Gary, F.; Proust, C.; Dilanian, G.; Hengstenberg, C.; Ruppert, V.; Arbustini, E.; Wichter, T.; Germain, M.; *et al.* A genome-wide association study identifies two loci associated with heart failure due to dilated cardiomyopathy. *Eur. Heart J.* **2011**, *32*, 1065–1076.
20. Jefferies, J.L.; Towbin, J.A. Dilated cardiomyopathy. *Lancet* **2010**, *375*, 752–762.
21. Hershberger, R.E.; Norton, N.; Morales, A.; Li, D.; Siegfried, J.D.; Gonzalez-Quintana, J. Coding sequence rare variants identified in MYBPC3, MYH6, TPM1, TNNC1, and TNNI3 from 312 patients with familial or idiopathic dilated cardiomyopathy. *Circ. Cardiovasc. Genet.* **2010**, *3*, 155–161.
22. Rampersaud, E.; Kinnamon, D.D.; Hamilton, K.; Khuri, S.; Hershberger, R.E.; Martin, E.R. Common susceptibility variants examined for association with dilated cardiomyopathy. *Ann. Hum. Genet.* **2010**, *74*, 110–116.
23. Tiret, L.; Mallet, C.; Poirier, O.; Nicaud, V.; Millaire, A.; Bouhour, J.B.; Roizes, G.; Desnos, M.; Dorent, R.; Schwartz, K.; *et al.* Lack of association between polymorphisms of eight candidate genes and idiopathic dilated cardiomyopathy: The CARDIGENE study. *J. Am. Coll. Cardiol.* **2000**, *35*, 29–35.
24. Takayama, S.; Xie, Z.; Reed, J.C. An evolutionarily conserved family of Hsp70/Hsc70 molecular chaperone regulators. *J. Biol. Chem.* **1999**, *274*, 781–786.
25. Ellinor, P.T.; Sasse-Klaassen, S.; Probst, S.; Gerull, B.; Shin, J.T.; Toepfel, A.; Heuser, A.; Michely, B.; Yoerger, D.M.; Song, B.S.; *et al.* A novel locus for dilated cardiomyopathy, diffuse myocardial fibrosis, and sudden death on chromosome 10q25–26. *J. Am. Coll. Cardiol.* **2006**, *48*, 106–111.
26. Norton, N.; Li, D.; Rieder, M.J.; Siegfried, J.D.; Rampersaud, E.; Zuchner, S.; Mangos, S.; Gonzalez-Quintana, J.; Wang, L.; McGee, S.; *et al.* Genome-wide studies of copy number variation and exome sequencing identify rare variants in BAG3 as a cause of dilated cardiomyopathy. *Am. J. Hum. Genet.* **2011**, *88*, 273–282.
27. Reich, D.E.; Lander, E.S. On the allelic spectrum of human disease. *Trends Genet.* **2001**, *17*, 502–510.
28. Wang, W.Y.; Barratt, B.J.; Clayton, D.G.; Todd, J.A. Genome-wide association studies: Theoretical and practical concerns. *Nat. Rev. Genet.* **2005**, *6*, 109–118.
29. Manolio, T.A.; Collins, F.S.; Cox, N.J.; Goldstein, D.B.; Hindorff, L.A.; Hunter, D.J.; McCarthy, M.I.; Ramos, E.M.; Cardon, L.R.; Chakravarti, A.; *et al.* Finding the missing heritability of complex diseases. *Nature* **2009**, *461*, 747–753.
30. McPherson, R.; Pertsemlidis, A.; Kavaslar, N.; Stewart, A.; Roberts, R.; Cox, D.R.; Hinds, D.A.; Pennacchio, L.A.; Tybjaerg-Hansen, A.; Folsom, A.R.; *et al.* A common allele on chromosome 9 associated with coronary heart disease. *Science* **2007**, *316*, 1488–1491.
31. Helgadóttir, A.; Thorleifsson, G.; Manolescu, A.; Gretarsdóttir, S.; Blondal, T.; Jonasdóttir, A.; Sigurdsson, A.; Baker, A.; Palsson, A.; Masson, G.; *et al.* A common variant on chromosome 9p21 affects the risk of myocardial infarction. *Science* **2007**, *316*, 1491–1493.
32. Lander, E.S. The new genomics: Global views of biology. *Science* **1996**, *274*, 536–539.
33. Pritchard, J.K.; Cox, N.J. The allelic architecture of human disease genes: Common disease-common variant ... or not? *Hum. Mol. Genet.* **2002**, *11*, 2417–2423.
34. Gibson, G. Rare and common variants: Twenty arguments. *Nat. Rev. Genet.* **2011**, *13*, 135–145.

35. Manolio, T.A. Cohort studies and the genetics of complex disease. *Nat. Genet.* **2009**, *41*, 5–6.
36. Sun, J.X.; Helgason, A.; Masson, G.; Ebenesersdottir, S.S.; Li, H.; Mallick, S.; Gnerre, S.; Patterson, N.; Kong, A.; Reich, D.; *et al.* A direct characterization of human mutation based on microsatellites. *Nat. Genet.* **2012**, *44*, 1161–1165.
37. Marian, A.J.; Belmont, J. Strategic approaches to unraveling genetic causes of cardiovascular diseases. *Circ. Res.* **2011**, *108*, 1252–1269.
38. Ng, P.C.; Levy, S.; Huang, J.; Stockwell, T.B.; Walenz, B.P.; Li, K.; Axelrod, N.; Busam, D.A.; Strausberg, R.L.; Venter, J.C. Genetic variation in an individual human exome. *PLoS Genet.* **2008**, *4*, e1000160.
39. Tennessen, J.A.; Bigham, A.W.; O'Connor, T.D.; Fu, W.; Kenny, E.E.; Gravel, S.; McGee, S.; Do, R.; Liu, X.; Jun, G.; *et al.* Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science* **2012**, *337*, 64–69.
40. Johansen, C.T.; Wang, J.; Lanktree, M.B.; Cao, H.; McIntyre, A.D.; Ban, M.R.; Martins, R.A.; Kennedy, B.A.; Hassell, R.G.; Visser, M.E.; *et al.* Excess of rare variants in genes identified by genome-wide association study of hypertriglyceridemia. *Nat. Genet.* **2010**, *42*, 684–687.
41. Aulchenko, Y.S.; Ripatti, S.; Lindqvist, I.; Boomsma, D.; Heid, I.M.; Pramstaller, P.P.; Penninx, B.W.; Janssens, A.C.; Wilson, J.F.; Spector, T.; *et al.* Loci influencing lipid levels and coronary heart disease risk in 16 European population cohorts. *Nat. Genet.* **2009**, *41*, 47–55.
42. Kathiresan, S.; Willer, C.J.; Peloso, G.M.; Demissie, S.; Musunuru, K.; Schadt, E.E.; Kaplan, L.; Bennett, D.; Li, Y.; Tanaka, T.; *et al.* Common variants at 30 loci contribute to polygenic dyslipidemia. *Nat. Genet.* **2009**, *41*, 56–65.
43. Li, H.; Homer, N. A survey of sequence alignment algorithms for next-generation sequencing. *Brief. Bioinforma.* **2010**, *11*, 473–483.
44. Nielsen, R.; Paul, J.S.; Albrechtsen, A.; Song, Y.S. Genotype and SNP calling from next-generation sequencing data. *Nat. Rev. Genet.* **2011**, *12*, 443–451.
45. Abecasis, G.R.; Auton, A.; Brooks, L.D.; DePristo, M.A.; Durbin, R.M.; Handsaker, R.E.; Kang, H.M.; Marth, G.T.; McVean, G.A. An integrated map of genetic variation from 1092 human genomes. *Nature* **2012**, *491*, 56–65.
46. Jabbari, J.; Jabbari, R.; Nielsen, M.W.; Holst, A.G.; Nielsen, J.B.; Haunso, S.; Tfelt-Hansen, J.; Svendsen, J.H.; Olesen, M.S. New exome data question the pathogenicity of genetic variants previously associated with catecholaminergic polymorphic ventricular tachycardia. *Circ. Cardiovasc. Genet.* **2013**, *6*, 481–489.
47. Priori, S.G.; Napolitano, C.; Memmi, M.; Colombi, B.; Drago, F.; Gasparini, M.; DeSimone, L.; Coltorti, F.; Bloise, R.; Keegan, R.; *et al.* Clinical and molecular characterization of patients with catecholaminergic polymorphic ventricular tachycardia. *Circulation* **2002**, *106*, 69–74.
48. Golbus, J.R.; Puckelwartz, M.J.; Fahrenbach, J.P.; Dellefave-Castillo, L.M.; Wolfgeher, D.; McNally, E.M. Population-based variation in cardiomyopathy genes. *Circ. Cardiovasc. Genet.* **2012**, *5*, 391–399.
49. Risgaard, B.; Jabbari, R.; Refsgaard, L.; Holst, A.G.; Haunso, S.; Sadjadieh, A.; Winkel, B.G.; Olesen, M.S.; Tfelt-Hansen, J. High prevalence of genetic variants previously associated with Brugada syndrome in new exome data. *Clin. Genet.* **2013**, *84*, 489–495.

50. Kapplinger, J.D.; Landstrom, A.P.; Salisbury, B.A.; Callis, T.E.; Pollevick, G.D.; Tester, D.J.; Cox, M.G.; Bhuiyan, Z.; Bikker, H.; Wiesfeld, A.C.; *et al.* Distinguishing arrhythmogenic right ventricular cardiomyopathy/dysplasia-associated mutations from background genetic noise. *J. Am. Coll. Cardiol.* **2011**, *57*, 2317–2327.
51. Andreassen, C.; Refsgaard, L.; Nielsen, J.B.; Sajadieh, A.; Winkel, B.G.; Tfelt-Hansen, J.; Haunso, S.; Holst, A.G.; Svendsen, J.H.; Olesen, M.S. Mutations in genes encoding cardiac ion channels previously associated with sudden infant death syndrome (SIDS) are present with high frequency in new exome data. *Can. J. Cardiol.* **2013**, *29*, 1104–1109.
52. Petretta, M.; Pirozzi, F.; Sasso, L.; Paglia, A.; Bonaduce, D. Review and metaanalysis of the frequency of familial dilated cardiomyopathy. *Am. J. Cardiol.* **2011**, *108*, 1171–1176.
53. Meder, B.; Haas, J.; Keller, A.; Heid, C.; Just, S.; Borries, A.; Boisguerin, V.; Scharfenberger-Schmeer, M.; Stahler, P.; Beier, M.; *et al.* Targeted next-generation sequencing for the molecular genetic diagnostics of cardiomyopathies. *Circ. Cardiovasc. Genet.* **2011**, *4*, 110–122.
54. Zimmerman, R.S.; Cox, S.; Lakdawala, N.K.; Cirino, A.; Mancini-DiNardo, D.; Clark, E.; Leon, A.; Duffy, E.; White, E.; Baxter, S.; *et al.* A novel custom resequencing array for dilated cardiomyopathy. *Genet. Med.* **2010**, *12*, 268–278.
55. Partners Healthcare. Available online: <http://pcpgm.partners.org/> (accessed on 25 February 2014).
56. Opitz, C.A.; Kulke, M.; Leake, M.C.; Neagoe, C.; Hinssen, H.; Hajjar, R.J.; Linke, W.A. Damped elastic recoil of the titin spring in myofibrils of human myocardium. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 12688–12693.
57. Granzier, H.L.; Irving, T.C. Passive tension in cardiac muscle: Contribution of collagen, titin, microtubules, and intermediate filaments. *Biophys. J.* **1995**, *68*, 1027–1044.
58. Horowitz, R.; Kempner, E.S.; Bisher, M.E.; Podolsky, R.J. A physiological role for titin and nebulin in skeletal muscle. *Nature* **1986**, *323*, 160–164.
59. Muhle-Goll, C.; Habeck, M.; Cazorla, O.; Nilges, M.; Labeit, S.; Granzier, H. Structural and functional studies of titin's fn3 modules reveal conserved surface patterns and binding to myosin S1—A possible role in the Frank-Starling mechanism of the heart. *J. Mol. Biol.* **2001**, *313*, 431–447.
60. Cazorla, O.; Wu, Y.; Irving, T.C.; Granzier, H. Titin-based modulation of calcium sensitivity of active tension in mouse skinned cardiac myocytes. *Circ. Res.* **2001**, *88*, 1028–1035.
61. Siu, B.L.; Niimura, H.; Osborne, J.A.; Fatkin, D.; MacRae, C.; Solomon, S.; Benson, D.W.; Seidman, J.G.; Seidman, C.E. Familial dilated cardiomyopathy locus maps to chromosome 2q31. *Circulation* **1999**, *99*, 1022–1026.
62. Gerull, B.; Atherton, J.; Geupel, A.; Sasse-Klaassen, S.; Heuser, A.; Frenneaux, M.; McNabb, M.; Granzier, H.; Labeit, S.; Thierfelder, L. Identification of a novel frameshift mutation in the giant muscle filament titin in a large Australian family with dilated cardiomyopathy. *J. Mol. Med.* **2006**, *84*, 478–483.
63. Gerull, B.; Gramlich, M.; Atherton, J.; McNabb, M.; Trombitas, K.; Sasse-Klaassen, S.; Seidman, J.G.; Seidman, C.; Granzier, H.; Labeit, S.; *et al.* Mutations of TTN, encoding the giant muscle filament titin, cause familial dilated cardiomyopathy. *Nat. Genet.* **2002**, *30*, 201–204.
64. Herman, D.S.; Lam, L.; Taylor, M.R.; Wang, L.; Teekakirikul, P.; Christodoulou, D.; Conner, L.; DePalma, S.R.; McDonough, B.; Sparks, E.; *et al.* Truncations of titin causing dilated cardiomyopathy. *N. Engl. J. Med.* **2012**, *366*, 619–628.

65. Ng, S.B.; Turner, E.H.; Robertson, P.D.; Flygare, S.D.; Bigham, A.W.; Lee, C.; Shaffer, T.; Wong, M.; Bhattacharjee, A.; Eichler, E.E.; *et al.* Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* **2009**, *461*, 272–276.
66. Majewski, J.; Schwartzentruber, J.; Lalonde, E.; Montpetit, A.; Jabado, N. What can exome sequencing do for you? *J. Med. Genet.* **2011**, *48*, 580–589.
67. Campbell, N.; Sinagra, G.; Jones, K.L.; Slavov, D.; Gowan, K.; Merlo, M.; Carniel, E.; Fain, P.R.; Aragona, P.; di Lenarda, A.; *et al.* Whole exome sequencing identifies a troponin T mutation hot spot in familial dilated cardiomyopathy. *PLoS One* **2013**, *8*, e78104.
68. Wells, Q.S.; Becker, J.R.; Su, Y.R.; Mosley, J.D.; Weeke, P.; D’Aoust, L.; Ausborn, N.L.; Ramirez, A.H.; Pfothenauer, J.P.; Naftilan, A.J.; *et al.* Whole exome sequencing identifies a causal RBM20 mutation in a large pedigree with familial dilated cardiomyopathy. *Circ. Cardiovasc. Genet.* **2013**, *6*, 317–326.
69. Yandell, M.; Huff, C.; Hu, H.; Singleton, M.; Moore, B.; Xing, J.; Jorde, L.B.; Reese, M.G. A probabilistic disease-gene finder for personal genomes. *Genome Res.* **2011**, *21*, 1529–1542.
70. Li, D.; Morales, A.; Gonzalez-Quintana, J.; Norton, N.; Siegfried, J.D.; Hofmeyer, M.; Hershberger, R.E. Identification of novel mutations in RBM20 in patients with dilated cardiomyopathy. *Clin. Transl. Sci.* **2010**, *3*, 90–97.
71. Davydov, E.V.; Goode, D.L.; Sirota, M.; Cooper, G.M.; Sidow, A.; Batzoglou, S. Identifying a high fraction of the human genome to be under selective constraint using GERP++. *PLoS Comput. Biol.* **2010**, *6*, e1001025.
72. 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.
73. Millat, G.; Bouvagnet, P.; Chevalier, P.; Sebbag, L.; Dulac, A.; Dauphin, C.; Jouk, P.S.; Delrue, M.A.; Thambo, J.B.; Le Metayer, P.; *et al.* Clinical and mutational spectrum in a cohort of 105 unrelated patients with dilated cardiomyopathy. *Eur. J. Med. Genet.* **2011**, *54*, e570–e575.
74. Lakdawala, N.K.; Dellefave, L.; Redwood, C.S.; Sparks, E.; Cirino, A.L.; Depalma, S.; Colan, S.D.; Funke, B.; Zimmerman, R.S.; Robinson, P.; *et al.* Familial dilated cardiomyopathy caused by an alpha-tropomyosin mutation: The distinctive natural history of sarcomeric dilated cardiomyopathy. *J. Am. Coll. Cardiol.* **2010**, *55*, 320–329.
75. Roncarati, R.; Viviani Anselmi, C.; Krawitz, P.; Lattanzi, G.; von Kodolitsch, Y.; Perrot, A.; di Pasquale, E.; Papa, L.; Portararo, P.; Columbaro, M.; *et al.* Doubly heterozygous LMNA and TTN mutations revealed by exome sequencing in a severe form of dilated cardiomyopathy. *Eur. J. Hum. Genet.* **2013**, *21*, 1105–1111.
76. Robinson, P.N.; Kohler, S.; Bauer, S.; Seelow, D.; Horn, D.; Mundlos, S. The Human Phenotype Ontology: A tool for annotating and analyzing human hereditary disease. *Am. J. Hum. Genet.* **2008**, *83*, 610–615.
77. Perrot, A.; Hussein, S.; Ruppert, V.; Schmidt, H.H.; Wehnert, M.S.; Duong, N.T.; Posch, M.G.; Panek, A.; Dietz, R.; Kindermann, I.; *et al.* Identification of mutational hot spots in LMNA encoding lamin A/C in patients with familial dilated cardiomyopathy. *Basic Res. Cardiol.* **2009**, *104*, 90–99.

78. Theis, J.L.; Sharpe, K.M.; Matsumoto, M.E.; Chai, H.S.; Nair, A.A.; Theis, J.D.; de Andrade, M.; Wieben, E.D.; Michels, V.V.; Olson, T.M. Homozygosity mapping and exome sequencing reveal GATAD1 mutation in autosomal recessive dilated cardiomyopathy. *Circ. Cardiovasc. Genet.* **2011**, *4*, 585–594.
79. Vermeulen, M.; Eberl, H.C.; Matarese, F.; Marks, H.; Denissov, S.; Butter, F.; Lee, K.K.; Olsen, J.V.; Hyman, A.A.; Stunnenberg, H.G.; *et al.* Quantitative interaction proteomics and genome-wide profiling of epigenetic histone marks and their readers. *Cell* **2010**, *142*, 967–980.
80. Alkan, C.; Coe, B.P.; Eichler, E.E. Genome structural variation discovery and genotyping. *Nat. Rev. Genet.* **2011**, *12*, 363–376.
81. Iafrate, A.J.; Feuk, L.; Rivera, M.N.; Listewnik, M.L.; Donahoe, P.K.; Qi, Y.; Scherer, S.W.; Lee, C. Detection of large-scale variation in the human genome. *Nat. Genet.* **2004**, *36*, 949–951.
82. Tuzun, E.; Sharp, A.J.; Bailey, J.A.; Kaul, R.; Morrison, V.A.; Pertz, L.M.; Haugen, E.; Hayden, H.; Albertson, D.; Pinkel, D.; *et al.* Fine-scale structural variation of the human genome. *Nat. Genet.* **2005**, *37*, 727–732.
83. Kidd, J.M.; Cooper, G.M.; Donahue, W.F.; Hayden, H.S.; Sampas, N.; Graves, T.; Hansen, N.; Teague, B.; Alkan, C.; Antonacci, F.; *et al.* Mapping and sequencing of structural variation from eight human genomes. *Nature* **2008**, *453*, 56–64.
84. Krumm, N.; Sudmant, P.H.; Ko, A.; O’Roak, B.J.; Malig, M.; Coe, B.P.; Quinlan, A.R.; Nickerson, D.A.; Eichler, E.E. Copy number variation detection and genotyping from exome sequence data. *Genome Res.* **2012**, *22*, 1525–1532.
85. Magi, A.; Tattini, L.; Cifola, I.; D’Aurizio, R.; Benelli, M.; Mangano, E.; Battaglia, C.; Bonora, E.; Kurg, A.; Seri, M.; *et al.* EXCAVATOR: Detecting copy number variants from whole-exome sequencing data. *Genome Biol.* **2013**, *14*, R120.
86. Zhao, Y.; Samal, E.; Srivastava, D. Serum response factor regulates a muscle-specific microRNA that targets Hand2 during cardiogenesis. *Nature* **2005**, *436*, 214–220.
87. Thum, T.; Galuppo, P.; Wolf, C.; Fiedler, J.; Kneitz, S.; van Laake, L.W.; Doevendans, P.A.; Mummery, C.L.; Borlak, J.; Haverich, A.; *et al.* MicroRNAs in the human heart: A clue to fetal gene reprogramming in heart failure. *Circulation* **2007**, *116*, 258–267.
88. Zhao, Y.; Ransom, J.F.; Li, A.; Vedantham, V.; von Drehle, M.; Muth, A.N.; Tsuchihashi, T.; McManus, M.T.; Schwartz, R.J.; Srivastava, D. Dysregulation of cardiogenesis, cardiac conduction, and cell cycle in mice lacking miRNA-1-2. *Cell* **2007**, *129*, 303–317.
89. Van Rooij, E.; Sutherland, L.B.; Qi, X.; Richardson, J.A.; Hill, J.; Olson, E.N. Control of stress-dependent cardiac growth and gene expression by a microRNA. *Science* **2007**, *316*, 575–579.
90. Callis, T.E.; Pandya, K.; Seok, H.Y.; Tang, R.H.; Tatsuguchi, M.; Huang, Z.P.; Chen, J.F.; Deng, Z.; Gunn, B.; Shumate, J.; *et al.* MicroRNA-208a is a regulator of cardiac hypertrophy and conduction in mice. *J. Clin. Invest.* **2009**, *119*, 2772–2786.
91. Montgomery, R.L.; Hullinger, T.G.; Semus, H.M.; Dickinson, B.A.; Seto, A.G.; Lynch, J.M.; Stack, C.; Latimer, P.A.; Olson, E.N.; van Rooij, E. Therapeutic inhibition of miR-208a improves cardiac function and survival during heart failure. *Circulation* **2011**, *124*, 1537–1547.
92. Rao, P.K.; Toyama, Y.; Chiang, H.R.; Gupta, S.; Bauer, M.; Medvid, R.; Reinhardt, F.; Liao, R.; Krieger, M.; Jaenisch, R.; *et al.* Loss of cardiac microRNA-mediated regulation leads to dilated cardiomyopathy and heart failure. *Circ. Res.* **2009**, *105*, 585–594.

93. Bernstein, B.E.; Birney, E.; Dunham, I.; Green, E.D.; Gunter, C.; Snyder, M. An integrated encyclopedia of DNA elements in the human genome. *Nature* **2012**, *489*, 57–74.
94. Hindorff, L.A.; Sethupathy, P.; Junkins, H.A.; Ramos, E.M.; Mehta, J.P.; Collins, F.S.; Manolio, T.A. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 9362–9367.
95. Wang, K.; Kim, C.; Bradfield, J.; Guo, Y.; Toskala, E.; Otieno, F.G.; Hou, C.; Thomas, K.; Cardinale, C.; Lyon, G.J.; *et al.* Whole-genome DNA/RNA sequencing identifies truncating mutations in RBCK1 in a novel Mendelian disease with neuromuscular and cardiac involvement. *Genome Med.* **2013**, *5*, 67.
96. Wang, K.; Li, M.; Hakonarson, H. ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **2010**, *38*, e164.
97. Benjamini, Y.; Speed, T.P. Summarizing and correcting the GC content bias in high-throughput sequencing. *Nucleic Acids Res.* **2012**, *40*, e72.
98. Sirmaci, A.; Edwards, Y.J.; Akay, H.; Tekin, M. Challenges in whole exome sequencing: An example from hereditary deafness. *PLoS One* **2012**, *7*, e32000.
99. Siepel, A.; Bejerano, G.; Pedersen, J.S.; Hinrichs, A.S.; Hou, M.; Rosenbloom, K.; Clawson, H.; Spieth, J.; Hillier, L.W.; Richards, S.; *et al.* Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* **2005**, *15*, 1034–1050.
100. Ng, P.C.; Henikoff, S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res.* **2003**, *31*, 3812–3814.
101. Cooper, G.M.; Stone, E.A.; Asimenos, G.; Green, E.D.; Batzoglou, S.; Sidow, A. Distribution and intensity of constraint in mammalian genomic sequence. *Genome Res.* **2005**, *15*, 901–913.
102. Thomas, P.D.; Kejariwal, A.; Campbell, M.J.; Mi, H.; Diemer, K.; Guo, N.; Ladunga, I.; Ulitsky-Lazareva, B.; Muruganujan, A.; Rabkin, S.; *et al.* PANTHER: A browsable database of gene products organized by biological function, using curated protein family and subfamily classification. *Nucleic Acids Res.* **2003**, *31*, 334–341.
103. Ashkenazy, H.; Erez, E.; Martz, E.; Pupko, T.; Ben-Tal, N. ConSurf 2010: Calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res.* **2010**, *38*, W529–W533.
104. Yeo, G.; Burge, C.B. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J. Comput. Biol.* **2004**, *11*, 377–394.
105. Bookman, E.B.; Langehorne, A.A.; Eckfeldt, J.H.; Glass, K.C.; Jarvik, G.P.; Klag, M.; Koski, G.; Motulsky, A.; Wilfond, B.; Manolio, T.A.; *et al.* Reporting genetic results in research studies: Summary and recommendations of an NHLBI working group. *Am. J. Med. Genet. A* **2006**, *140*, 1033–1040.
106. Green, R.C.; Berg, J.S.; Grody, W.W.; Kalia, S.S.; Korf, B.R.; Martin, C.L.; McGuire, A.L.; Nussbaum, R.L.; O’Daniel, J.M.; Ormond, K.E.; *et al.* ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. *Genet. Med.* **2013**, *15*, 565–574.