ORIGINAL RESEARCH

Genomic Context Differs Between Human Dilated Cardiomyopathy and Hypertrophic Cardiomyopathy

Megan J. Puckelwartz [iD], PhD*; Lorenzo L. Pesce, PhD*; Lisa M. Dellefave-Castillo [iD], MS; Matthew T. Wheeler [iD], MD, PhD; Tess D. Pottinger, PhD; Avery C. Robinson, BS; Samuel D. Kearns, BS; Anthony M. Gacita, PhD; Zachary J. Schoppen [iD], MD; Wenyu Pan, MD; Gene Kim, MD; Jane E. Wilcox, MD, MSc; Allen S. Anderson [iD], MD; Euan A. Ashley [iD], MD, PhD; Sharlene M. Day [iD], MD; Thomas Cappola, MD, ScM; Gerald W. Dorn, II, MD; Elizabeth M. McNally [iD], MD, PhD

BACKGROUND: Inherited cardiomyopathies display variable penetrance and expression, and a component of phenotypic variation is genetically determined. To evaluate the genetic contribution to this variable expression, we compared protein coding variation in the genomes of those with hypertrophic cardiomyopathy (HCM) and dilated cardiomyopathy (DCM).

METHODS AND RESULTS: Nonsynonymous single-nucleotide variants (nsSNVs) were ascertained using whole genome sequencing from familial cases of HCM (n=56) or DCM (n=70) and correlated with echocardiographic information. Focusing on nsSNVs in 102 genes linked to inherited cardiomyopathies, we correlated the number of nsSNVs per person with left ventricular measurements. Principal component analysis and generalized linear models were applied to identify the probability of cardiomyopathy type as it related to the number of nsSNVs in cardiomyopathy genes. The probability of having DCM significantly increased as the number of cardiomyopathy gene nsSNVs per person increased. The increase in nsSNVs in cardiomyopathy genes significantly associated with reduced left ventricular ejection fraction and increased left ventricular diameter for individuals carrying a DCM diagnosis, but not for those with HCM. Resampling was used to identify genes with aberrant cumulative allele frequencies, identifying potential modifier genes for cardiomyopathy.

CONCLUSIONS: Participants with DCM had more nsSNVs per person in cardiomyopathy genes than participants with HCM. The nsSNV burden in cardiomyopathy genes did not correlate with the probability or manifestation of left ventricular measures in HCM. These findings support the concept that increased variation in cardiomyopathy genes creates a genetic background that predisposes to DCM and increased disease severity.

Key Words: dilated cardiomyopathy ■ hypertrophic cardiomyopathy ■ modifier genes ■ variable expressivity ■ variant burden

Heart failure affects >5 million Americans and is of growing health and economic concern.[1] A leading cause of heart failure is cardiomyopathy, a disease with a strong heritable component. The 2 most common forms of cardiomyopathy are hypertrophic cardiomyopathy (HCM) and dilated cardiomyopathy (DCM). HCM occurs in 1:200 - 500 adults and is characterized by gross hypertrophy of the septum and left ventricular (LV) free wall.[2,3] DCM is found in ≈1:250 - 500 adults and presents with reduced LV ejection fraction (LVEF) and dilation.[4,5] Over 100 genes are implicated in the pathogenesis of cardiomyopathy, and most cardiomyopathy is inherited in an autosomal dominant manner. Two genes, *MYH7*

CLINICAL PERSPECTIVE

What Is New?

- Dilated and hypertrophic cardiomyopathy are highly variable in onset and progression of disease severity, indicating that modifiers, including genetic modifiers, play a role in disease expression.
- We evaluated and compared genome-wide coding variation in dilated cardiomyopathy and hypertrophic cardiomyopathy cases to query whether genomic backgrounds differ.

What Are the Clinical Implications?

- We evaluated missense single-nucleotide variation in genes associated with cardiomyopathy and found that the number of genetic changes per person increased the probability to having dilated cardiomyopathy but not hypertrophic cardiomyopathy and correlated with more severe disease.
- The distinct genetic landscapes between hypertrophic cardiomyopathy and dilated cardiomyopathy suggest that greater genetic variation in cardiomyopathy genes provokes unfavorable disease expression in dilated cardiomyopathy.

Nonstandard Abbreviations and Acronyms

DCM	dilated cardiomyopathy
gnomAD	Genome Aggregation Database
HCM	hypertrophic cardiomyopathy
IVSd	interventricular septal end diastole
LVIDd	left ventricle internal diameter end diastole
LVPWd	left ventricular posterior wall end diastole
nsSNV	nonsynonymous single-nucleotide variant
PC	principal component
PCA	principal component analysis
WGS	whole genome sequencing

and *MYBPC3,* are responsible for >80% of genetic HCM, making HCM largely a disease of sarcomere dysfunction.[2,6–8] DCM is more genetically heterogeneous, with mutations in one gene, *TTN*, accounting for ≈15% to 20% of inherited DCM; the other mutations are found in genes encoding cytoskeletal, nucleoskeletal, mitochondrial, myofilament, and calcium handling proteins.[5,9,10] A feature of both HCM and DCM is variable expressivity and penetrance.[11]

Genetic heterogeneity and variable expressivity imply a more complex inheritance, and some studies support oligogenic contributions to pathogenesis.[12–14]

Genetic testing for cardiomyopathies has emerged as a useful clinical tool for both disease diagnosis and risk stratification.[15,16] Rare variants account for most primary mutations in inherited cardiomyopathy, with few hot spots or recurrent mutations. Variant interpretation considers population frequency, in silico tools of pathogenicity, and previous reports of clinical and functional outcomes. Current cardiac genetic testing samples 20 to 100 genes, and depending on the primary indication, gene panel testing has ≈50% sensitivity. This reduced sensitivity or "missing heritability" for cardiomyopathy may be attributable to multiple factors, including (1) undiscovered primary or "driver" gene mutations and/or (2) an oligogenic or modifier genetic mechanism involving the interplay between highly penetrant variants and the genomic context in which they are expressed.

Whole genome sequencing (WGS) is an effective means to determine both rare and common variation. Herein, we applied WGS to 126 subjects with either familial HCM or DCM from whom echocardiographic measurements were available. We examined variation in 102 cardiomyopathy genes routinely assayed in clinical gene testing panels. Both linear and logistic regression models revealed that subjects with more nonsynonymous coding variants per person in the 102 cardiomyopathy genes were significantly more likely to express a DCM clinical phenotype as opposed to an HCM phenotype. These data also held true for high-frequency gene variants in the cardiomyopathy cohort. The number of cardiomyopathy gene variants per person also associated with reduced ejection fraction and increased LV diameter in subjects with DCM, but not subjects with HCM. These results suggest that distinct genetic landscapes exist between HCM and DCM and that greater genetic variation in cardiomyopathy genes may be partially responsible for the unfavorable ventricular remodeling with reduced systolic function in DCM compared with HCM.

METHODS

Detailed methods available in Data S1. The study was approved by the Institutional Review Boards at University of Chicago, Stanford University, University of Michigan, and Northwestern University. All subjects provided written informed consent.

Because of the sensitive nature of the data collected for this study, requests to access the data set from qualified researchers trained in human subject confidentiality protocols may be sent to Megan Puckelwartz at Northwestern University.

Study Subjects

Subjects with nonischemic DCM or HCM with familial disease were selected for WGS.

Generation of WGS Data

Genomic DNA was determined using the Illumina HiSeq2000, 2500, or XTen and mapped to National Center for Biotechnology Information hg19 with MegaSeq.[17] Only genomes with coverage >30× were included. Variant effects were predicted using snpEff.[18]

Ancestry Principal Component Analysis

Principal component analysis (PCA) was used to estimate global ancestry through dimensional reduction. PCAs were conducted using singular-value decomposition of shared variants using ≈5 million biallelic variants selected from across the genome using PLINK v1.9 and R 3.2.3.

Echocardiography PCA

PCA was performed using the R function prcomp (R statistical software version 3.4.1) on scaled and centered LVEF, LV internal diameter end diastole (LVIDd), interventricular septal end diastole (IVSd), and LV posterior wall end diastole (LVPWd). All dimensions were adjusted to body surface area (BSA).

Linear Models

Multivariate and univariate linear models were fit using the lm function of R (R statistical software version 3.4.1). Model significance was assessed using a χ^2 test on reduction in the residual sum of squares. The response variables were echocardiographic measures or their principal components, and the independent variables were nonsynonymous single-nucleotide variants (nsSNVs), genetic ancestry, and platform.

Generalized Linear Model

Multivariate and univariate logistic regressions were fit using the generalized linear model R function, glm, with binomial link function. The response variable was whether a patient had DCM as opposed to HCM. The number of nsSNVs per patient was used as an independent variable in the primary analysis. Variant frequency subsets, defined by Genome Aggregation Database (gnomAD) exome allele frequency in both the cardiomyopathy genes and across all high-expression heart genes, were used for secondary analysis. P values correspond to the R analysis of deviance test function and therefore refer to comparison between nested models.

Resampling-Based Estimation of Excess Nonsynonymous Variation

A bootstrap approach was designed where each random sample with replacement was taken from the 126-subject set without constraining for cardiomyopathy subtype, sex, or ancestry to mimic the original sample collection method. We calculated excess allele counts, subtracting baseline values computed assuming ancestral gnomAD exome frequencies for African ancestry and non-Finnish European ancestry as reference values, including the sex chromosome number in each subject in the sample. Once corrected, allele numbers were summed over each of the cardiomyopathy genes, then in high-expression heart genes, creating observed excess cumulative allele numbers. The empirical CIs for these cumulative allele frequencies were estimated using 5000 bootstrap samples. The bootstrap samples were created by sampling with replacement using subjects as the sampling unit. These calculations were performed using in-house functions in R. Significance was assessed using a false discovery rate <0.25.

RESULTS

Cardiomyopathy Subject Selection and WGS

Subjects with nonischemic DCM (n=70) or HCM (n=56) with familial disease were selected for WGS analysis (Table 1). Detailed family history of cardiomyopathy was available for all subjects. Subjects were retrospectively recruited from specialized clinics at the University of Chicago, Stanford University, University of Michigan, and Northwestern University. Genetic ancestry was determined using PCA of ≈5 million biallelic markers, and individuals were classified as African and non-African based largely on principal component (PC) 1. Genetic analysis revealed that one pair of individuals was distantly related (half uncle relation), and subjects were otherwise unrelated. At the outset, 35 individuals with *MYH7* mutations were selected for the HCM cohort. These *MYH7*-mutation carriers served as an internal control when identifying genes/pathways enriched for genetic burden.

To focus on missense protein coding variation, only nsSNVs were studied. After excluding insertion/deletions, missense variants accounted for 96% of total variants, and nonsense variants constituted the remaining 4%. More than one platform was used for WGS, and the number of single-nucleotide variants identified across platforms did not differ (t test; P=0.32; R²=0.008). A cardiomyopathy gene list (102 genes) was generated from commercial testing panels, and only 89 of the 102 genes had variation in the

Table 1. Demographic Characteristics of Cohort Subjects

Demographics	Dilated Cardiomyopathy	Hypertrophic Cardiomyopathy	*P* Value
No.	70	56	
Male sex, n (%)	42 (60)	32 (57)	0.48
Age, mean±SD, y*	39±16	45±15	0.08†
Genetic race, n (%)			
African ancestry	18 (25)‡	6 (11)	0.04
Non-African ancestry	53 (75)	50 (89)	0.04
Ascertained for *MYH7*‡	0 (0)	35 (62)	<0.0001

The χ² test was performed, unless otherwise noted.

*Age at presentation.

†Nonparametric unpaired Mann-Whitney test was performed. Individuals classified as Hispanic and other were reclassified as non-African ancestry on the basis of genetically determined best-fit ancestry.

‡Individuals selected on the basis of clinical genetic testing that returned a variant in *MYH7*.

sequenced cohort (Table S1). When evaluating these 102 genes, 10 357 nsSNVs were identified across the 126 genomes, and missense variation accounted for 97.4% of all variants. Table S2 provides the distribution of nsSNVs per person in 102 cardiomyopathy genes for DCM and HCM. Pathogenic, likely pathogenic, and suspicious variants of uncertain significance were identified in each subject and curated using evidence from ClinVar and expert input (Table S3).

PCA of Echocardiographic Measures

Echocardiographic LV measurements were used to correlate genetic data with phenotype. Cardiac dimensions were normalized to BSA. LVEF, LVIDd, IVSd, and LVPWd were significantly different between the DCM and HCM cohorts, consistent with the primary diagnosis (Figure S1 and Table 2). PCA was performed on measures of LVEF, LVIDd/BSA, IVSd/BSA, and LVPWd/BSA. Figure S2A shows that the first PC of echocardiographic measures (PC1) accounted for 59% of the variance, whereas PC2 accounted for 25%, PC3 accounted for 11%, and PC4 accounted for 5% (Table S4 and Figure S2A). Figure 1A illustrates that echocardiographic PC1 reliably separated cardiomyopathy subtypes, in addition to explaining most of the variance.

Component loadings of each phenotype demonstrated the contribution of LV functional measures to the first 2 components (Figure S2B). Collectively, this suggests that echocardiographic PC1 is a simple quantitative variable for illustrating the echocardiographic differences between DCM and HCM.

To determine if the variability of echocardiographic PC1 could be partially explained by cumulative genetic factors, we queried the number of nsSNVs per person in the 102 cardiomyopathy genes. Regression of echocardiographic PC1 against the number of nsSNVs per person in cardiomyopathy genes was significant (n=126; *P*=0.0019), indicating that the number of nsSNVs per person contributes to the echocardiographic differences between DCM and HCM (Figure 1B). This model using total cardiomyopathy gene nsSNVs per person (n=10 357 total variants), as a linear predictor, accounted for 11.5% of the variance of echocardiographic PC1. To determine the contribution of pathogenic variants/ likely pathogenic variants/suspicious variants of uncertain significance, the analysis was performed adjusting for pathogenic variants/likely pathogenic variants/suspicious variants of uncertain significance (Table S3; n=71), and this did not affect significance or size of effect. In addition, a supporting subgroup

Table 2. LV Measurements Derived From Echocardiograms in Cardiomyopathy Cohort

Measure	Dilated Cardiomyopathy	Hypertrophic Cardiomyopathy	*P* Value
No.	70	56	
Age at echocardiogram, mean±SD, y (N)	41±16 (61)	47±14 (48)	0.06*
Ejection fraction, %, median (IQR) (N)	21 (15–35) (61)	65 (60–70) (48)	<0.0001
Left ventricle internal diameter, diastole/BSA, median (IQR), cm/m² (N)	3.1 (2.9–3.7) (43)	2.3 (2.0–2.6) (45)	<0.0001
Interventricular septum, diastole/BSA, median (IQR), cm/m² (N)	0.5 (0.47–0.60) (39)	0.85 (0.70–1.1) (46)	<0.0001
Left ventricle posterior wall thickness, diastole/BSA, median (IQR), cm/m² (N)	0.51 (0.44–0.59) (39)	0.61 (0.52–0.72) (45)	0.002

Measurements were normalized to BSA (m²), where indicated. The *t* test was performed, unless otherwise noted. BSA indicates body surface area; IQR, interquartile range (first and third quartiles); and LV, left ventricular.

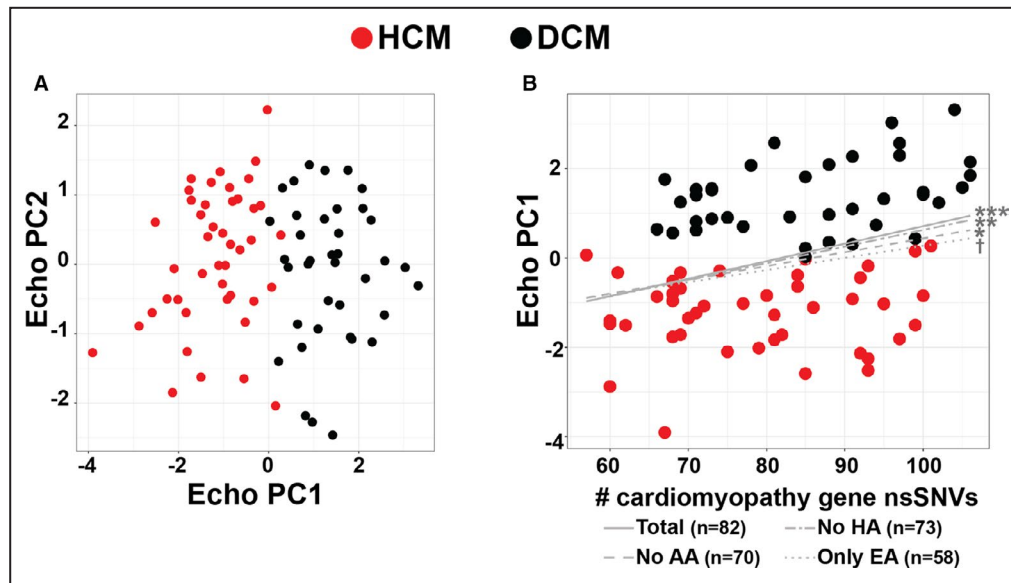*Nonparametric unpaired Mann-Whitney test was performed.

Figure 1.   Principal component analysis (PCA) of echocardiographic data separates hypertrophic cardiomyopathy (HCM) and dilated cardiomyopathy (DCM) and is predicted by the number of nonsynonymous single-nucleotide variants (nsSNVs) in cardiomyopathy genes.

A, PCA of echocardiography data shows that echocardiographic principal component 1 (Echo PC1) summarizes the difference in echocardiography data between HCM (red) and DCM (black). **B**, Regression of Echo PC1 against the number of cardiomyopathy gene nsSNVs/person was significant and effectively separated HCM and DCM, as seen by the solid gray line (n=82). To account for genetic ancestry, linear regression was repeated in the absence of African ancestry (n=70) subjects (dashed gray line) or Hispanic ancestry (n=73) subjects (dot-dash gray line) or using only the European ancestry (dotted gray line; n=58) subjects. *P=0.024, **P=0.005, ***P=0.002, †P=0.069.

analysis was performed to determine if the trend remained consistent within ethnically homogeneous groups. When subjects with either Hispanic or African ancestry were removed, the estimates remained significant (P=0.005 and n=73 and P=0.024 and n=70, respectively; Table S5). When the sample was restricted to only European ancestry subjects, reducing the number of subjects by a third, the general trend remained unaffected, with a borderline significant P value (P=0.069). Together, these data indicate that the variance in echocardiographic PC1 is, in part, explained by the number of cardiomyopathy gene nsSNVs per person. As a control, we also considered sequencing platform in addition to echocardiographic PC1 and nsSNVs, and this resulted in no change in the coefficient (0.039, for both models) slope or significance (P=0.002) (Figure S3). Together, these data indicate that greater protein coding variation in cardiomyopathy genes accounts, in part, for the differences in LV measures between DCM and HCM.

Probability of DCM Relative to HCM Increases With the Number of Cardiomyopathy Gene nsSNVs

To illustrate the association between these cardiomyopathy subtypes and the number of cardiomyopathy gene nsSNVs per person, a simple generalized multivariate linear model was fitted using a standard stepwise procedure based on analysis of deviance and Akaike Information Criterion. Because DCM and HCM genome sequencing was imbalanced across sequencing platform, it was included as an adjustment to the model (Table S6). Adding the number of cardiomyopathy gene nsSNVs per person to the model significantly improved the fit and the ability to predict DCM in this cohort (Table S6; P=0.021), and this analysis demonstrates a cumulative genetic factor model where having more nsSNVs predisposes to DCM (Figure 2). Adding genetic ancestry to this model neither improved the fit nor negated the contribution of the number of nsSNVs, thus showing a robust dependency between cardiomyopathy subtype and number of nsSNVs (Table S6).

We assessed the contribution of moderate population frequency nsSNVs, defined by gnomAD exome global frequency (Figure 2, right panel). We found that the per-person number of cardiomyopathy gene nsSNVs in the 25% to 50% population frequency spectrum predicted DCM versus HCM using a nested model with platform and number of nsSNVs, similar to the above analysis (analysis of deviance=0.023; Table S6). Although adding ancestry to the model improved fit, it did not change the
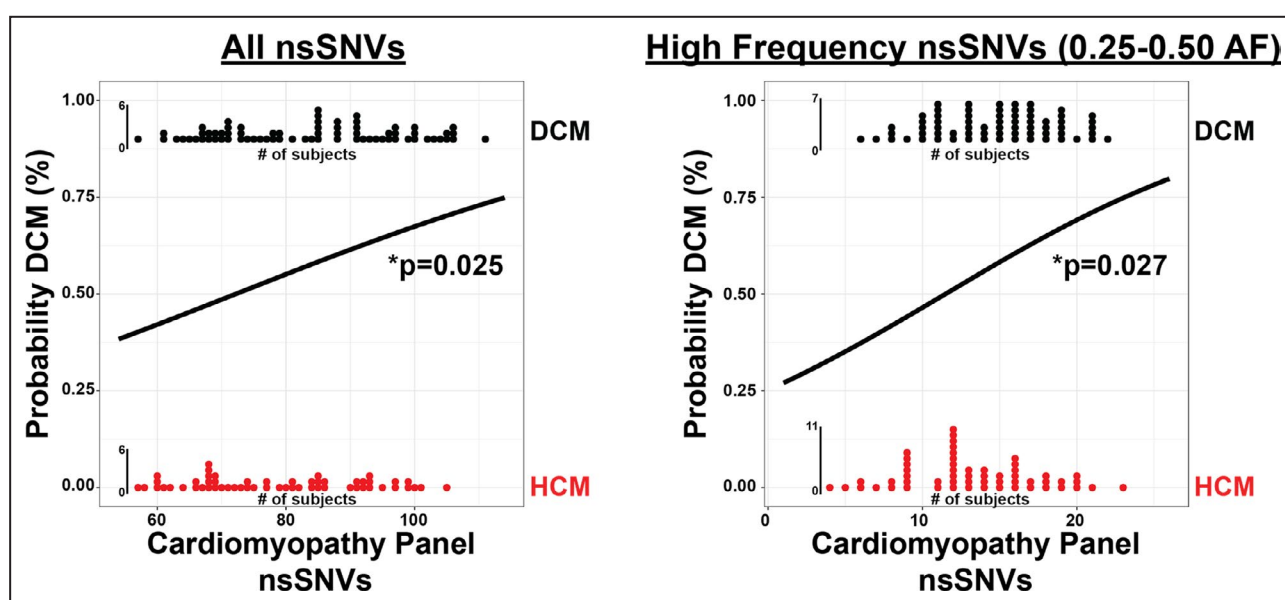
Figure 2. Dilated cardiomyopathy (DCM) probability is increased with cardiomyopathy gene nonsynonymous single-nucleotide variant (nsSNV) number compared with hypertrophic cardiomyopathy (HCM).
Multivariate generalized linear models demonstrate that the probability of DCM is increased with the total number of cardiomyopathy gene nsSNVs per person in this cohort (left panel). The red and black dots represent individual participants and their number of nsSNVs in cardiomyopathy genes. The analysis on the left considered all nsSNVs in cardiomyopathy genes, including rare and high-frequency variants. The right-hand panel shows the same analysis when considering only high-frequency cardiomyopathy gene nsSNVs, where the same trend was evident (allele frequency, 0.25–0.50 variants included). DCM is black, and HCM is red. *P* values after adjustment for platform prevalence imbalance are shown (see Table S5). Number on x axis indicates the number of nsSNVs per subject.

coefficient (0.110 for number of nsSNVs). Using nsSNVs from genes with low expression in the heart showed no ability to predict the probability of DCM and HCM (Table S6). We also completed an analysis using additional frequency bins and found no qualitative difference in the direction of the effect (Table S7 and Figure S4). Although underpowered to evaluate all frequency subsets, these data suggest that the number of nsSNVs per person predicts DCM compared with HCM across the frequency spectrum. Together, these data indicate that the number of cardiomyopathy gene nsSNVs associates with the DCM cardiomyopathy subtype.

Number of Cardiomyopathy Gene nsSNVs Associates With Disease Severity in DCM

To interrogate the relationship between the number of cardiomyopathy gene nsSNVs and disease severity, we regressed individual LV measures against the number of nsSNVs per person in cardiomyopathy genes (Figure 3A and 3B). Reduced LVEF and increased LVIDd were each significantly dependent on the number of cardiomyopathy gene nsSNVs per subject with DCM (Table S8). In subjects with HCM, the number of cardiomyopathy gene nsSNVs was not significantly associated with either LVEF or LVIDd (Table S8). IVSd

and LVPWd, both hallmarks of HCM, also showed no association with cardiomyopathy gene nsSNV number (Table S8).

To determine if these results were specific for cardiomyopathy gene nsSNVs and not the result of overall genomic burden, we separated all genes into either high or low cardiac expression groups using gene expression levels derived from genotype-tissue expression (see Data S1 and Figure S5). We used low expression heart genes to determine if the association between LV measures and cardiomyopathy gene number nsSNV was specific, reasoning that low expression heart genes should play a lesser role in LV function. We considered variation per person in 89 randomly selected genes from the low-expression heart genes to match the 89 cardiomyopathy genes that carry variation in the HCM/DCM cohort, and repeated this process 1000 times. Using both total and 25% to 50% allele frequency, nsSNVs from these randomly selected low-expression heart genes revealed minor qualitative trends; however, no statistical association between LVEF, LVIDd, IVSd, or LVPWd with nsSNV count was observed (Figure 3C and 3D and Table S9). We cannot exclude that a type II error is not driving these results; however, total nsSNVs encompass cardiomyopathy nsSNVs and therefore will provide a weaker, but nonzero, predictor.
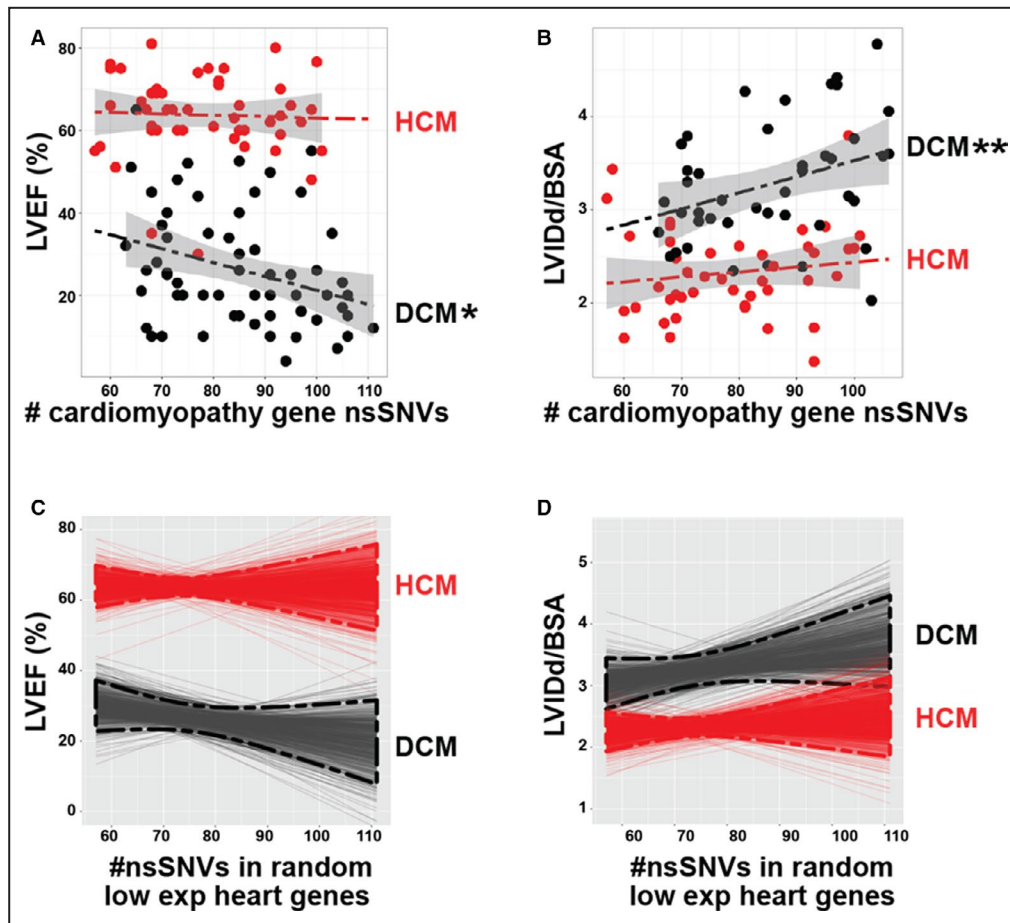
Figure 3. The number of cardiomyopathy gene nonsynonymous single-nucleotide variants (nsSNVs) per person associates with reduced cardiac function and increased left ventricular (LV) diameter.

LV ejection fraction (LVEF) (**A**) and left ventricular internal diameter end diastole (LVIDd), normalized to body surface area (BSA) (**B**), were regressed against the total number of cardiomyopathy gene nsSNVs per person. Cardiomyopathy gene nsSNV number significantly correlated with reduced LVEF and increased LVIDd in subjects with dilated cardiomyopathy (DCM), and this correlation was not seen for subjects with hypertrophic cardiomyopathy (HCM) (*LVEF DCM $P=0.01$; HCM $P=0.78$; **LVIDd/BSA DCM $P=0.02$; HCM $P=0.34$); gray shading represents 95% CIs (see Table S6 for values). A randomly selected group of genes with comparable variant numbers was chosen from the low-expression heart genes and similarly tested. This process was repeated 1000 times with different groups of genes. LVEF (**C**) and LVIDd/BSA (**D**) were regressed against low-expression heart genes, and the regression lines did not reveal any significant association with LV measures. Dashed lines represent 95% CIs (see Table S8 for values).

Together, these results indicate that the probability and severity for each subject with DCM, relative to subjects with HCM, are associated with increasing nonsynonymous variant load in cardiomyopathy genes.

Genes With Deviant Cumulative Variant Frequency in DCM and HCM

To identify candidate genes that contribute to the differences in variant load between DCM and HCM, we investigated if any genes had aberrant cumulative variant frequencies compared with gnomAD exome frequency data. Ancestral allele frequencies were compared with the allele frequencies in each bootstrap

sample, generating excess cumulative allele frequencies for each gene, and this process was repeated 5000 times. The resulting values for DCM and HCM were subtracted from each other to produce Delta (schematic shown in Figure S6).

Three cardiomyopathy genes had excess cumulative frequencies that were statistically significant at false discovery rate <0.25 (Figure 4).[19] We originally selected 35 of 56 subjects with HCM based on *MYH7* variant carrier status, so, if reliable, the bootstrap method should identify *MYH7* variation as enriched in the HCM cohort and serve as an internal control. *MYH7* was significantly enriched for variation in the HCM cohort. *LMNA* was significantly enriched in DCM,
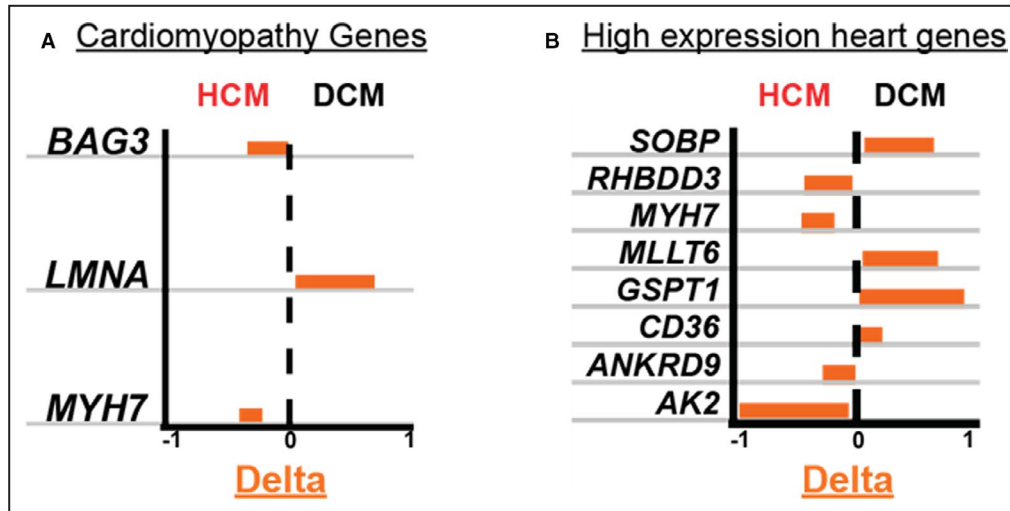
Figure 4.   Resampling identifies genes with deviant cumulative missense allele frequencies that may modify dilated cardiomyopathy (DCM) or hypertrophic cardiomyopathy (HCM).

The bootstrap method (5000 resampling with replacement tests) was used to identify excess nonsynonymous single-nucleotide variant burden in DCM compared with HCM (Delta) in either cardiomyopathy genes (**A**) or high-expression heart genes (**B**). When conducting this analysis on cardiomyopathy genes, *MYH7* and *BAG3* were identified as having increased cumulative variation in HCM. *MYH7* was expected to appear in this analysis because *MYH7* mutations were enriched in the HCM cohort, and thus serve as an internal control for this approach. From known cardiomyopathy genes, *BAG3* was identified as being enriched in HCM over DCM, and *LMNA* variation was enriched in DCM over HCM. This bootstrap method was applied across all high expressed cardiac genes and identified potential novel modifiers of HCM and DCM.

fitting well with *LMNA*'s known role in the pathogenesis of DCM.[20] Together, these data support this method of calculating deviant allele frequencies while adjusting for ancestry and sex.

We repeated the bootstrap analysis on genes with high heart expression (as defined in Figure S5). In this exploratory analysis of 7306 cardiac genes, 8 had excess cumulative frequencies that differed between HCM and DCM with a false discovery rate <0.25 (Figure 4B). Table S10 provides, in ranked order, the genotype-tissue expression levels in heart for these genes and any previous genome-wide association study from the genome-wide association study catalog (https://www.ebi.ac.uk/gwas). Two of these genes, *ANKRD9,* encoding ankyrin repeat domain 9, and *CD36*, encoding the thrombospondin receptor, have previously been associated with cardiac phenotypes of QT interval and LV mass, respectively.[21,22]

DISCUSSION

Cardiomyopathy Gene Variation Differs Between DCM and HCM

Cascade family testing for primary gene mutations highlights the range of clinical expression seen with genetic variants.[23] Environmental factors and other genetic variants contribute to this variable expressivity. In this report, we identified that the number of nsSNVs in cardiomyopathy genes correlated with DCM but not HCM. Notably, we found the number of nonsynonymous variants in cardiomyopathy genes correlated not only for DCM diagnosis compared with HCM, but also for specific aspects of DCM, including reduced LV function and increased LV size. These findings were evident in cardiomyopathy genes and were not present in low-expression heart genes.

Identifying common variation that contributes to cardiomyopathy disease expression would aid in the development of genomic risk scores for cardiomyopathy. A large-scale genome-wide association study and multitrait analysis in HCM, DCM, and LV traits from UK Biobank participants with healthy hearts revealed loci associated with each cardiomyopathy subtype and with LV measures.[24] The study identified strong genetic correlations with the cardiomyopathies and LV traits. We also identified correlations with cardiac dimensions and cardiomyopathy subtype. The authors further generated polygenic risk scores and found that for subjects with HCM carrying a rare, disease-causing variant, common variation accounted for phenotypic variability in subjects with HCM. Our data are consistent with this concept (namely, that common variation is contributing to disease variability in the cardiomyopathies).

Oligogenic Inheritance in Cardiomyopathies and Implications for Genetic Testing

In light of the reduced penetrance and expressivity of familial cardiomyopathies, modifying factors of disease

have been postulated.[11] We now identified potential modifiers of cardiomyopathy using a random resampling method. We identified expected genes, including *MYH7* in HCM and *LMNA* in DCM. The identification of additional variants that may act in concert to cause disease or affect severity suggests oligogenic inheritance. Functional experiments using CRISPR-Cas9 now allow testing of multiple variants in concert. Recent work by Gifford and colleagues identified a family with asymptomatic parents and 3 children with early-onset heart disease.[25] Exome sequencing revealed a complex inheritance, with 3 variants likely contributing to disease. In vivo gene editing techniques revealed that variants in *MKL2*, *MYH7*, and *NKX2-5* act together to cause LV noncompaction. The 2 variants in *MKL2* and *MYH7* were paternally inherited and unique to the family, whereas the maternal *NKX2-5* allele was rare. Experimental modeling confirmed the role of these variants in disease and established *NKX2-5* as a modifier of disease. These data support our hypothesis that modifying variants may make a significant contribution to disease phenotype.

Despite progress, likely pathogenic or pathogenic variants are found in less than half of cases, depending on the type of cardiomyopathy.[26] Recent work by Haas et al used deep sequencing of 76 cardiomyopathy genes in a large cohort of 639 people with DCM and found that >38% patients had compound or combined rare mutations, further supporting oligogenic contribution.[20] Cowan and colleagues reexamined 19 pedigrees with *LMNA*-associated cardiomyopathy with cardiomyopathy-positive family members who did not have the "causal" *LMNA* gene variant, suggesting additional genetic causes of disease.[27] In a large DCM cohort with 1040 subjects, Mazzarotto et al sequenced 56 cardiomyopathy genes and found robust disease association with 12 of those genes, explaining 17% of cases and 26% of cases in a validation cohort with 1498 subjects.[28] Our data indicate that variant load plays an important role in both the manifestation and severity of DCM. The clinical utility of knowing variant load and how this relates to an individual's disease progression is not established and requires additional study. However, the concept of oligogenic inheritance in DCM provides one path to better understanding variable expressivity. Highly penetrant, rare variants provide risk knowledge, and additional variants may ultimately help refine that risk.

In this study, we examined the genetic landscape that differentiates DCM and HCM using a cohort in which individual-level data, both clinical and genetic, were available. To define genetic drivers of cardiomyopathy, a large well-phenotyped and genotyped control data set is required. Current data sets, such as gnomAD and others, provide a rich source of population-level allele frequency information. However, these data sets,

by design, cannot be used to parse the contributions of allele combinations on disease state because they provide aggregate data. Moving the forefront of precision medicine requires deep sequencing and phenotype information while protecting subjects' privacy.

REFERENCES

1. Mozaffarian D, Benjamin EJ, Go AS, Arnett DK, Blaha MJ, Cushman M, de Ferranti S, Despres JP, Fullerton HJ, Howard VJ, et al. Heart disease and stroke statistics–2015 update: a report from the American Heart Association. *Circulation*. 2015;131:e29–e322. DOI: 10.1161/CIR.0000000000000152

2. Marian AJ, Braunwald E. Hypertrophic cardiomyopathy: genetics, pathogenesis, clinical manifestations, diagnosis, and therapy. *Circ Res*. 2017;121:749–770. DOI: 10.1161/CIRCRESAHA.117.311059.

3. Geske JB, Ommen SR, Gersh BJ. Hypertrophic cardiomyopathy: clinical update. *JACC Heart Fail*. 2018;6:364–375. DOI: 10.1016/j.jchf.2018.02.010.

4. Hershberger RE, Hedges DJ, Morales A. Dilated cardiomyopathy: the complexity of a diverse genetic architecture. *Nat Rev Cardiol*. 2013;10:531–547. DOI: 10.1038/nrcardio.2013.105.

5. McNally EM, Mestroni L. Dilated cardiomyopathy: genetic determinants and mechanisms. *Circ Res*. 2017;121:731–748. DOI: 10.1161/CIRCRESAHA.116.309396.

6. Walsh R, Buchan R, Wilk A, John S, Felkin LE, Thomson KL, Chiaw TH, Loong CCW, Pua CJ, Raphael C, et al. Defining the genetic architecture of hypertrophic cardiomyopathy: re-evaluating the role of non-sarcomeric genes. *Eur Heart J*. 2017;38:3461–3468. DOI: 10.1093/eurheartj/ehw603.

7. Wilcox JE, Hershberger RE. Genetic cardiomyopathies. *Curr Opin Cardiol*. 2018;33:354–362. DOI: 10.1097/HCO.0000000000000512.

8. Garfinkel AC, Seidman JG, Seidman CE. Genetic pathogenesis of hypertrophic and dilated cardiomyopathy. *Heart Fail Clin*. 2018;14:139–146. DOI: 10.1016/j.hfc.2017.12.004.

9. Herman DS, Lam L, Taylor MRG, Wang L, Teekakirikul P, Christodoulou D, Conner L, DePalma SR, McDonough B, Sparks E, et al. Truncations of titin causing dilated cardiomyopathy. *N Engl J Med*. 2012;366:619–628. DOI: 10.1056/NEJMoa1110186.

10. Roberts AM, Ware JS, Herman DS, Schafer S, Baksi J, Bick AG, Buchan RJ, Walsh R, John S, Wilkinson S, et al. Integrated allelic, transcriptional, and phenomic dissection of the cardiac effects of titin truncations in health and disease. *Sci Transl Med*. 2015;7:270ra276. DOI: 10.1126/scitranslmed.3010134.

11. Morales A, Hershberger R. Clinical application of genetic testing in heart failure. *Curr Heart Fail Rep*. 2017;14:543–553. DOI: 10.1007/s11897-017-0366-4.

12. Li L, Bainbridge MN, Tan Y, Willerson JT, Marian AJ. A potential oligogenic etiology of hypertrophic cardiomyopathy: a classic single-gene disorder. *Circ Res*. 2017;120:1084–1090. DOI: 10.1161/CIRCRESAHA.116.310559.

13. Kelly M, Semsarian C. Multiple mutations in genetic cardiovascular disease: a marker of disease severity? *Circ Cardiovasc Genet*. 2009;2:182–190. DOI: 10.1161/CIRCGENETICS.108.836478.

14. Priest JR, Osoegawa K, Mohammed N, Nanda V, Kundu R, Schultz K, Lammer EJ, Girirajan S, Scheetz T, Waggott D, et al. De novo and rare variants at multiple loci support the oligogenic origins of atrioventricular septal heart defects. *PLoS Genet*. 2016;12:e1005963.–10.1371/journal.pgen.1005963.

15. Cirino AL, Harris S, Lakdawala NK, Michels M, Olivotto I, Day SM, Abrams DJ, Charron P, Caleshu C, Semsarian C, et al. Role of genetic testing in inherited cardiovascular disease: a review. *JAMA Cardiol*. 2017;2:1153–1160. DOI: 10.1001/jamacardio.2017.2352.

16. Hershberger RE, Givertz MM, Ho CY, Judge DP, Kantor PF, McBride KL, Morales A, Taylor MRG, Vatta M, Ware SM. Genetic evaluation of cardiomyopathy—a Heart Failure Society of America Practice Guideline. *J Cardiac Fail*. 2018;24:281–302. DOI: 10.1016/j.cardfail.2018.03.004.

17. Puckelwartz MJ, Pesce LL, Nelakuditi V, Dellefave-Castillo L, Golbus JR, Day SM, Cappola TP, Dorn GW II, Foster IT, McNally EM. Supercomputing for the parallelization of whole genome analysis. *Bioinformatics*. 2014;30:1508–1513. DOI: 10.1093/bioinformatics/btu071.

18. Cingolani P, Platts A, le Wang L, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of drosophila melanogaster strain w1118; iso-2; iso-3. *Fly (Austin)*. 2012;6:80–92. DOI: 10.4161/fly.19695.

19. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Roy Stat Soc: Ser B (Methodol)*. 1995;57:289–300. DOI: 10.1111/j.2517-6161.1995.tb02031.x.

20. Haas J, Frese KS, Peil B, Kloos W, Keller A, Nietsch R, Feng Z, Müller S, Kayvanpour E, Vogel B, et al. Atlas of the clinical genetics of human dilated cardiomyopathy. *Eur Heart J*. 2015;36:1123–1135a. DOI: 10.1093/eurheartj/ehu301.

21. Arking DE, Pulit SL, Crotti L, van der Harst P, Munroe PB, Koopmann TT, Sotoodehnia N, Rossin EJ, Morley M, Wang X, et al. Genetic association study of QT interval highlights role for calcium signaling pathways in myocardial repolarization. *Nat Genet*. 2014;46:826–836. DOI: 10.1038/ng.3014.

22. Hall D, Mayosi BM, Rahman TJ, Avery PJ, Watkins HC, Keavney B. Common variation in the CD36 (fatty acid translocase) gene is associated with left-ventricular mass. *J Hypertens*. 2011;29:690–695. DOI: 10.1097/HJH.0b013e3283440115.

23. Wilde AA, Behr ER. Genetic testing for inherited cardiac disease. *Nat Rev Cardiol*. 2013;10:571–583. DOI: 10.1038/nrcardio.2013.108.

24. Tadros R, Francis C, Xu X, Vermeer AMC, Harper AR, Huurman R, Kelu Bisabu K, Walsh R, Hoorntje ET, te Rijdt WP, et al. Shared genetic pathways contribute to risk of hypertrophic and dilated cardiomyopathies with opposite directions of effect. *Nat Genet*. 2021;53:128–134. DOI: 10.1038/s41588-020-00762-2.

25. Gifford CA, Ranade SS, Samarakoon R, Salunga HT, de Soysa TY, Huang YU, Zhou P, Elfenbein A, Wyman SK, Bui YK, et al. Oligogenic inheritance of a human heart disease involving a genetic modifier. *Science*. 2019;364:865–870. DOI: 10.1126/science.aat5056.

26. Sweet M, Taylor MR, Mestroni L. Diagnosis, prevalence, and screening of familial dilated cardiomyopathy. *Expert Opin Orphan Drugs*. 2015;3:869–876. DOI: 10.1517/21678707.2015.1057498.

27. Cowan JR, Kinnamon DD, Morales A, Salyer L, Nickerson DA, Hershberger RE. Multigenic disease and bilineal inheritance in dilated cardiomyopathy is illustrated in nonsegregating LMNA pedigrees. *Circ Genom Precis Med*. 2018;11:e002038. DOI: 10.1161/CIRCGEN.117.002038.

28. Mazzarotto F, Tayal U, Buchan RJ, Midwinter W, Wilk A, Whiffin N, Govind R, Mazaika E, de Marvao A, Dawes TJW, et al. Reevaluating the genetic contribution of monogenic dilated cardiomyopathy. *Circulation*. 2020;141:387–398. DOI: 10.1161/CIRCULATIONAHA.119.037661.

29. Efron B, Tibshirani R. *An Introduction to the Bootstrap*. New York, NY: Chapman & Hall; 1993.

SUPPLEMENTAL MATERIAL

Data S1.

EXPANDED MATERIALS & METHODS

Study subjects. Non-ischemic dilated or hypertrophic cardiomyopathy subjects with familial disease were selected for whole genome sequencing (WGS). Detailed family history of cardiomyopathy was available for all subjects. Subjects were retrospectively recruited from specialized clinics at the University of Chicago, Stanford University, University of Michigan and Northwestern University. Subjects were selected based on a diagnosis of familial non-ischemic dilated cardiomyopathy or familial hypertrophic cardiomyopathy. Preference was given to subjects who had previous genetic testing and for whom a primary mutation was not identified, except in the case of selected *MYH7* variants and for whom echocardiographic data was available. Echocardiographic measurements were excluded if age at test was ≤ 17 years old or if the subject had a myectomy.

Generation of whole genome sequence data. Genomic DNA was extracted from peripheral blood. Sequencing was performed using the Illumina HiSeq2000, 2500 or XTen to greater than 30X coverage. Paired-end reads were mapped to NCBI reference genome hg19. Reads were aligned and variants were called using MegaSeq, a supercomputing pipeline using the open source Burrows Wheeler Aligner (BWA) and Genome Analysis Tool Kit (GATK).[18] Variant effects were predicted using snpEff.[19] Variant frequencies were annotated using frequencies from the gnomAD database (http://gnomad.broadinstitute.org/) accessed in March 2018. Allele frequencies are reported as global, EA (non-Finnish European ancestry) or AA (African ancestry) for each variant. When population allele frequency was used, only dataset variants with gnomAD exome coverage ≥ 8X were included in the analysis. Indels and multi-allelic variants were removed from the working dataset. Relatedness was accounted for using PLINK.

Ancestry principal component analysis (PCA). PCA was used to estimate global ancestry through dimensional reduction. The 1st and 3rd component illustrated ancestry. PCAs were conducted using singular-value decomposition of shared variants using ~5 million biallelic variants selected from across the genome using PLINK v1.9 and R 3.2.3.

Cardiomyopathy gene selection. Cardiomyopathy genes were identified based on their inclusion in one or more Cardiomyopathy gene testing panels of Partners Healthcare Laboratory for Molecular Medicine, GeneDx, and Invitae.

Dichotomizing heart genes into high and low expression. The Genotype-Tissue Expression (GTEx) database version 6p was downloaded Dec 2016 (https://www.gtexportal.org). The GTEx dataset includes RNA sequence analysis of human left atrial appendage (referred to as AA in GTEX, n=264 samples, and referred to as LA in this report) and left ventricle (LV, n=272). Expression values in RPKM (Reads Per Kilobase Million) for left ventricle (LV) and atrial appendage (AA) were used for genes identified as having variation in the DCM and HCM WGS dataset (17,058 genes with a total of 1,744,423 variants). Chromosome Y was removed that included 13 genes with 105 variants leaving 17,045 genes and 1,744,318 variants. Genes without cardiac GTEx values were removed leaving 16,825 genes and 1,729,723 variants. Variants with gnomAD exome coverage less than 8X were excluded leaving 16,353 genes and 1,635,325 variants. Genes with known segmental duplications were also removed (genome.ucsc.edu; GenomicSuperDups Table, assessed March 2018), leaving 14,915 genes and 1,252,213 variants. For these genes, RPKM values for LV were plotted against RPKM

values for AA as logarithms. We overlaid the cardiomyopathy genes on this plot to create an expression space that encompasses genes expressed in the heart.

Echocardiographic Data. 2-D echocardiographic measures were collated retrospectively from health records. Left ventricular ejection fraction (EF), left ventricular internal diameter end diastole (LVIDd), interventricular septal end diastole (IVSd), and left ventricular posterior wall end diastole (LVPWd) and body surface area were collected when available. If multiple echocardiograms were available, the study with the most abnormal measures was used. For DCM subjects, lowest EF was selected while for HCM subjects thickest IVSd was selected. All measures were taken from only one study date.

Echocardiography PCA. Principal component analysis (PCA) was performed using the R function prcomp (R statistical software version 3.4.1) on scaled and centered left ventricular ejection fraction, left ventricular internal diameter end diastole (LVIDd), interventricular septal end diastole (IVSd), and left ventricular posterior wall end diastole (LVPWd). All dimensions were adjusted to body surface area (BSA). Variables were centered by subtracting the mean and scaled by dividing by the standard deviation to prevent measurement units from determining their importance.

Linear models. Multivariate and univariate linear models were fit using the lm function of R (R statistical software version 3.4.1). Models were ranked in order of complexity (from one to four explanatory variables). Models of the same complexity were compared using t-tests provided by R and estimated r-squared values, while nested models were compared using the R anova function to estimate whether additional variables improved the fit significantly. Model significance was assessed using a chi-squared test on reduction in the residual sum of squares. Regression against random samples of genes were performed by randomly selecting 89 genes from the noncardiac gene list. 89 genes were selected to match the number of genes in the cardiomyopathy genes with nsSNVs. The sample was checked to ensure that the minimum and maximum number of variants in the randomly selected set were within 1 standard deviation from the minimum and the maximum number of variants of the 89 cardiomyopathy genes. This was repeated until 1000 sets met criteria. Linear regression of phenotype against these gene sets was performed as above. Confidence intervals were estimated directly from the sample.

Generalized linear model (GLM). Multivariate and univariate generalized linear models (GLM) were fit using the R function, glm, with binomial link function, fit to a logistic model. Models were generated using all nonsynonymous single nucleotide variants (nsSNVs) as primary analysis as well as variant frequency subsets defined by gnomAD exome allele frequency in both the cardiomyopathy genes and across all low-expression heart genes. To control for the potentially confounding effects of sequencing platform and genetic ancestry, a standard step-wise approach was used selecting the best-fit model and adding additional variables (sequencing platform, race and number of nsSNVs) until either the Akaike information criterion (AIC) increased or Analysis of Deviance (R function anova) showed non-significant improvements. Additional variables were excluded if they did not qualitatively change the contribution of variants to the fit by comparing the estimated coefficient values. P values correspond to the R anova (Analysis of Deviance test) function and therefore refer to the overall model.

Resampling based estimation of excess nonsynonymous variation. See **Figure S5** for schematic. Allele counts were aggregated by variant, sex, race, and cardiomyopathy type to create relatively homogeneous subsets. Differences in variant frequencies due to race were compensated for using gnomAD exome allele frequencies for African (AFR) or European ancestry (NFE). Sex was corrected for by only counting one allele and one chromosome per

male for chromosome X variants. Based on the number of chromosomes in the race and cardiomyopathy subset of the cohort, the expected allele numbers using ancestral gnomAD frequencies were calculated and subtracted from the observed alleles in the subset to re-center allele number around the expected value. These re-centered allele numbers were summed over each of the cardiomyopathy genes, then in high-expression heart genes set (as defined in **Figure S5**) creating observed excess cumulative allele numbers. Allele numbers were divided by the total number of chromosomes in each bootstrap. The empirical confidence intervals for these cumulative allele frequencies was estimated using 5,000 bootstrap samples.[30] As proof of principle, we analyzed cardiomyopathy genes and identified genes with significant differences between the subtypes. The bootstrap samples were created by sampling with replacement using subjects as the sampling unit. These calculations were performed using in-house functions in R. Significance for the difference between the cardiomyopathy types was estimated using the empirical quantiles of the distribution of the bootstrap sample differences, corresponding to a false discovery rate of 0.25, following the Benjamini-Hochberg method.[20]

Table S1. Cardiomyopathy Genes.

A2ML1	DOLK	JUP	NEXN	SLC22A5
ABCC9	DSC2	KRAS	NF1	SOS1
ACTC1	DSG2	LAMA4	NKX2-5	SOS2
ACTN2	DSP	LAMP2	NPPA	SPRED1
AGL	DTNA	LDB3	NRAS	TAZ
ALPK3	EMD	LMNA	PDLIM3	TCAP
ANKRD1	EYA4	LRRC10	PKP2	TGFB3
BAG3	FHL1	MAP2K1	PLEKHM2	TMEM43
BRAF	FHL2	MAP2K2	PLN	TMPO
CACNA1C	FKRP	MIB1	PRDM16	TNNC1
CALR3	FKTN	MURC	PRKAG2	TNNI3
CASQ2	FLNC	MYBPC3	PTPN11	TNNT2
CAV3	GAA	MYH6	RAF1	TPM1
CBL	GATA4	MYH7	RASA1	TRDN
CHRM2	GATA6	MYL2	RBM20	TTN
CRYAB	GATAD1	MYL3	RIT1	TTR
CSRP3	GLA	MYLK2	RRAS	TXNRD2
CTF1	HCN4	MYOM1	RYR2	VCL
CTNNA3	HRAS	MYOZ2	SCN5A	
DES	ILK	MYPN	SGCD	
DMD	JPH2	NEBL	SHOC2	

HUGO Gene names.

Table S2. Distribution of cardiomyopathy nsSNVs.

Phenotype	1st quartile	Median	3rd quartile
DCM	71	85	95.8
HCM	68	78	91

Table S3. Pathogenic/Likely Pathogenic/Suspicious VUS in HCM and DCM subjects.

Gene	Variant	Variant Designation	Segregation†
ACTA1	p.W358S	sVUS	N/A
BAG3	p.Q244*	LP	Yes
DES	c.735+3A>G	P	Yes
DMD	dup exons 3-9	P	Yes
DSP	p.E1245_N1246fs	LP	Yes
	p.Y1065*	P	Yes
FLNC	c.3791G>C	sVUS	N/A
	p.C1013*	P	N/A
	p.T291_V292fs	P/LP	N/A
	p.S2457fs	P	N/A
GLA	p.R118C	sVUS	Yes
LMNA	p.R433H	LP	N/A
MYBPC3	p.D1063_K1064fs	LP	Yes
	c.1224T>G	P	N/A
	p.W1097*	P	N/A
	p.L156P	sVUS	N/A
	p.P16S	LP	N/A
	p.E541Q	P	Yes
	p.R495Q	P/LP	Yes
MYBPHL	R255*	LP	Yes
MYH6	p.A1004S	P	N/A
MYH7	p.R1250W	LP	Yes
	p.M982T	sVUS	Yes
	p.L908V	P	Yes
	p.A355T	P	N/A
	p.H576R	LP	N/A
	p.G1057S	sVUS	N/A
	p.R663H	P	N/A
	p.E1468K	sVUS	N/A
	p.G716R	P	N/A
	p.K847E	LP	N/A
	p.A797T	P	N/A
	p.G1057D	sVUS	N/A
	p.R1606C	sVUS	N/A
	p.R719W	P	N/A
	p.R869H	sVUS	N/A

MYH7	p.P731A	sVUS	N/A
	p.N1327K	sVUS	N/A
	p.I736T	P	N/A
	p.V320M	LP	N/A
	p.E497D	P	N/A
	p.D906G	P	N/A
	p.D469N	sVUS	N/A
	p.R204H	LP	Yes
	p.T1377M	LP	N/A
	p.G741W	P	N/A
	p.R1712Q	LP	N/A
	p.E1455K	sVUS	N/A
	p.R652G	LP	Yes
	p.R403Q	P	N/A
	p.R783H	LP	N/A
RBM20	p.R634Q	P	Yes
	p.R636H	P	Yes
SCN5A	p.R814W	LP	Yes
	p.G1318V	P/LP	Yes
	p.Arg1316*	LP	Yes
TNNT2	K210del	sVUS	Yes
	p.N274D	sVUS	N/A
TPM1	p.D230N	P	Yes
	p.E54K	P/LP	N/A
	p.Leu254fs	sVUS	N/A
TTN	p.E3707*	LP	Yes
	p.Phe20604_Leu20605fs	P	Yes
	c.42521–5 C>G	sVUS	Yes
	p.Tyr15330*	sVUS	N/A
	p.R4386*	sVUS	N/A
	c.17087G>A	sVUS	Yes
	p.Val14761fs	sVUS	N/A
	p.Leu18170fs	sVUS	N/A
	p.Gly14825*	sVUS	N/A
	p.R12136*	LP	N/A

*P=pathogenic, LP=likely pathogenic, sVUS= suspicious variant of uncertain significance; as adjudicated using ClinVar, frequency data, subject phenotype and expert opinion. †Proband and at least one other affected relative carry the variant. N/A= not available

Table S4. PCA Weights for Left Ventricular Measures.

	PC1	PC2	PC3	PC4
IVSd_BSA	0.52	-0.25	-0.81	-0.08
LVPWd_BSA	0.29	-0.84	0.44	0.13
LVIDd_BSA	-.056	-0.39	-0.17	-0.71
EF (%)	0.57	0.28	0.36	-0.68
Proportion of Variance	0.59	0.25	0.11	0.05
Cumulative Proportion	0.59	0.84	0.95	1.00

PCA weights obtained by applying the R function prcomp to echocardiographic measures.
PCA= principal component analysis

Table S5. Linear regression of PC1 against ancestry subsets.

Model	Number of subjects	Coefficient	p value
V(all)*	82	0.0391	0.0019
No HA	73	0.0372	0.0055
No AA	70	0.0314	0.0243
Only EA	58	0.0272	0.0691

*V(all) = total cardiomyopathy gene nsSNVs; Genetic ancestry abbreviations: HA = Hispanic ancestry; AA = African ancestry; EA = European ancestry.

Table S6. Nested Generalized Linear Model Fits Adjusting for Ancestry and Sequencing Platform.

Model Variables	Degree of Model	Coefficient Ancestry	Coefficient nsSNV	p value*	Analysis of Deviance	AIC
Total nsSNVs in cardiomyopathy genes						
Platform	1			4.15×10^{-5}		**160.2**
nsSNVs	1		0.026	0.053		174.41
Ancestry	1	1.040		0.042		173.69
Platform & nsSNVs	2		**0.034**	**0.025** (nsSNVs)	**0.021**	**156.85**
Platform & Ancestry	2	1.11		0.043 (Ancestry)	0.034	157.71
Platform & nsSNVs & Ancestry	3	0.719	0.025		0.224	157.37
nsSNVs in cardiomyopathy genes with gnomAD Frequency 0.25-0.50						
Platform	1			4.15×10^{-5}		**160.2**
nsSNVs	1		0.095	0.037		173.76
Ancestry	1	1.040		0.042		173.69
Platform & nsSNVs	2		**0.110**	**0.027** (nsSNVs)	**0.023**	**157.03**
Platform & Ancestry	2	1.11		0.043 (Ancestry)	0.034	157.71
Platform & nsSNVs & Ancestry	3	1.11	**0.110**		0.037	154.67
Total nsSNVs in low expression heart genes						
Platform	1	**-1.59**		**4.15×10^{-5}**		**160.2**
nsSNVs	1		0.0008	0.027		172.83
Ancestry	1	1.04		0.042		173.69
Platform & nsSNVs	2		0.0007	0.055 (nsSNVs)	0.045	158.17
Platform & Ancestry	2	**1.11**		**0.043** (Ancestry)	**0.034**	**157.71**
Platform & nsSNVs & Ancestry	3	1.02	6.11×10^{-5}		0.494	159.17
nsSNVs in low expression heart genes with gnomAD Frequency 0.25-0.50						
Platform	1	**-1.59**		**4.15×10^{-5}**		**160.2**
nsSNVs	1		-0.002	0.28		177.09
Ancestry	1	1.04		0.042		173.69
Platform & nsSNVs	2		-0.003	0.205 (nsSNVs)	0.045	160.55
Platform & Ancestry	2	**1.11**		**0.043** (Ancestry)	**0.034**	**157.71**
Platform & nsSNVs & Ancestry	3	1.68	0.003		0.064	159.12

*p value cutoff of 0.1 used to determine variables for nested models. Model with best fit is shaded gray for each variable grouping. AIC=Akaike Information Criterion.

Table S7. Number of nsSNVs in Each Frequency Bin.

nsSNV Frequency Bin	Number of nsSNVs
Total	10357
0.25-0.50	1788
0.1-0.25	1795
<0.1	2472
<0.01	1075
<0.0025	626

Table S8. Linear regression of phenotype against V(all)*

Phenotype	Coefficient	P value	95% CI
LVEF			
DCM	-0.338	0.012	-0.593 - -0.082
HCM	-0.032	0.779	-0.258 – 0.193
LVIDd/BSA			
DCM	0.017	0.024	0.003 – 0.031
HCM	0.005	0.344	-0.005 – 0.016
IVSd/BSA			
DCM	0.0001	0.934	-0.003 – 0.003
HCM	-0.0008	0.830	0.008 – 0.007
LVPWd/BSA			
DCM	0.002	0.174	-0.235 – 0.239
HCM	0.001	0.473	-0.002 – 0.004

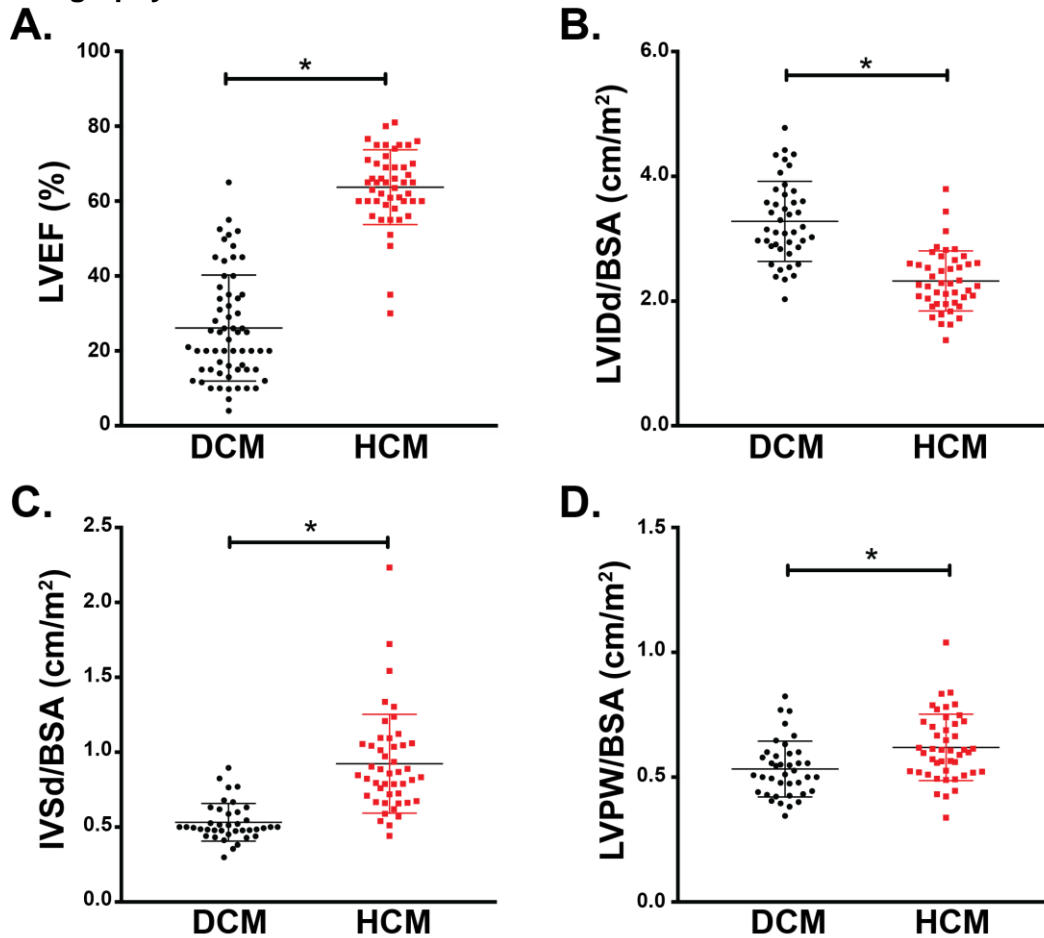* V(all) = total nsSNVs in cardiomyopathy genes. LVEF, Left ventricular ejection fraction; LVIDd/BSA, Left ventricular internal diameter at end diastole normalized to body surface area; IVSd/BSA, Interventricular septal end diastole normalized to BSA; LVPWd/BSA, Left ventricular posterior wall thickness at end diastole normalized to BSA. 95% CI = 95% confidence intervals

Table S9. Confidence Intervals for Simulated Linear Regression.

Phenotype	95% CI
LVEF	
DCM	-0.52 - 0.13
HCM	-0.33 - 0.32
LVIDd/BSA	
DCM	-0.01 - 0.03
HCM	-0.01 - 0.02
IVSd/BSA	
DCM	-0.003 - 0.003
HCM	-0.01 - 0.01
LVPWd/BSA	
DCM	-0.002 - 0.004
HCM	-0.004 - 0.005

LVEF, Left ventricular ejection fraction; LVIDd/BSA, Left ventricular internal diameter at end diastole normalized to body surface area; IVSd/BSA, Interventricular septal end diastole normalized to BSA; LVPWd/BSA, Left ventricular posterior wall thickness at end diastole normalized to BSA. 95% CI = 95% confidence intervals

Table S10. Cardiac expression levels and GWAS Associations of Cardiac Genes.

| GENE NAME	GENE	Atria		Ventricle		GWAS Reported Trait
		TPM	GTEx Rank*	TPM	GTEx Rank*	
MYH7	Myosin heavy chain 7	563.7	3	4552	1	Resting Heart Rate
ANKRD9	Ankyrin repeat domain 9	42.15	4	47.05	3	QT Interval; Blood Protein Levels
CD36	thrombospondin receptor	85.38	6	147.5	4	LV mass; RBC width; Mean corpuscular volume; RBC count; Platelet count; Mean platelet volume; HDL; Response to fenofibrate
SOBP	Sine oculis binding protein homolog	9.15	32	7.23	38	Airway Responsiveness in COPD; Myopia; Alzheimer Age of Onset
AK2	Adenylate kinase 2	23.88	40	18.05	43	Gut microbiota; Cerebral amyloid deposition in APOEe4 non-carriers
GSPT1	G1 to S phase transition	21.56	39	17.16	44	Menopause (age of onset); Testicular Germ Cell Tumor
PC	Pyruvate carboxylase	3.53	46	2.21	51	HIV-1 susceptibility
MLLT6	PHD finger containing	29.73	43	20.89	52	
RHBDD3	Rhomboid domain containing 3	10.55	48	7.19	52	Breast Cancer; Pancreatic Cancer

*Rank relative expression in relation to 53 tissues included in GTEx dataset
GWAS traits from the GWAS catalog (https://www.ebi.ac.uk/gwas). TPM= Transcripts Per Million GTEx database (https://www.gtexportal.org, assessed June 2018)

Figure S1. Left ventricle measurements of DCM and HCM subjects determined by echocardiography.


(**A**) Left ventricular ejection fraction (LVEF), (**B**) Left ventricular internal diameter at end diastole normalized to body surface area (LVIDd/BSA), (**C**) Interventricular septal end diastole normalized to BSA (IVSd/BSA), and (**D**) Left ventricular posterior wall thickness at end diastole normalized to BSA (LVPWd/BSA) are shown for subjects with DCM (black) and HCM (red), (*p<0.05, t-test). DCM, dilated cardiomyopathy; HCM, hypertrophic cardiomyopathy. DCM participants had lower LVEF and larger LVIDd measurements, while HCM subjects had increased IVS and LVPW compared to DCM, consistent with the primary diagnosis.

**Figure S2. PCA metrics of cardiac phenotype.**
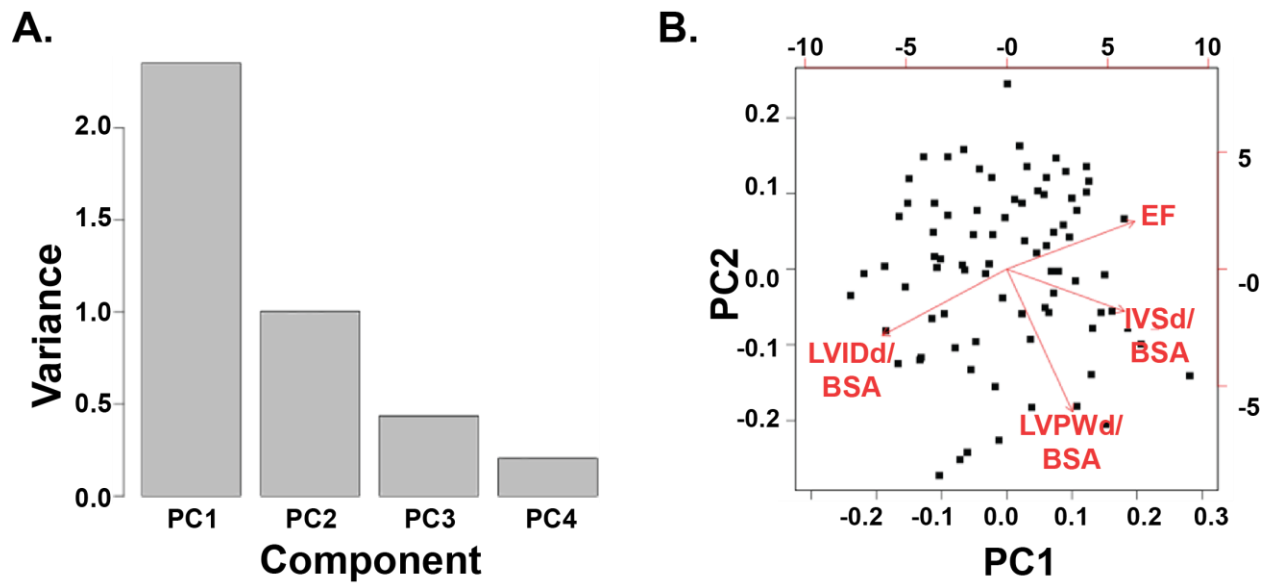
PCA on echocardiographic measurements for the cardiomyopathy cohorts. **A**) Scree plot showing the variance associated with each component. **B**) Component loadings of each phenotype measurement on the principal components.

Figure S3. Platform does not change the dependency between Echo PC1 and nsSNVs.
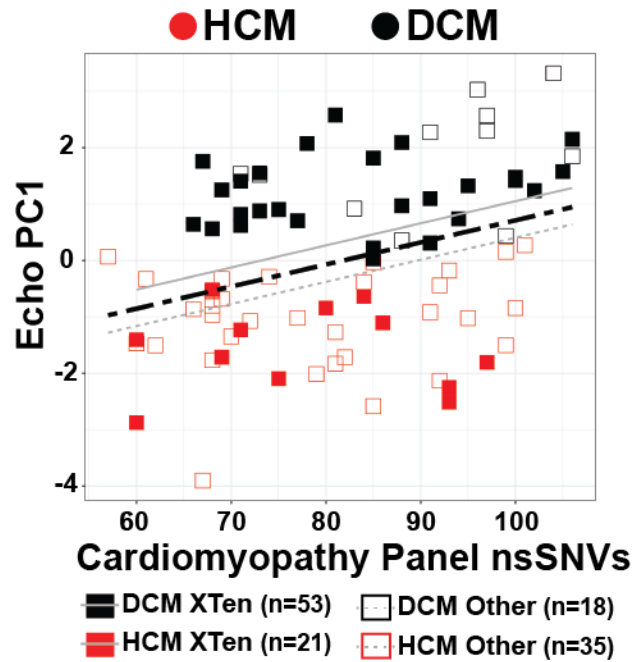

PCA of echocardiography data revealed that PC1 summarizes the difference between HCM and DCM. Regression of Echo PC1 against the number of cardiomyopathy gene nsSNVs per person demonstrates SNV number distinguishes cardiomyopathy subtype. Platform was tested to determine if the either of the two platforms used in this study were responsible for the dependence of Echo PC1 on nsSNV number. These data indicate that platform does not cause the dependency. Squares represent subjects. DCM, dilated cardiomyopathy (black); HCM, hypertrophic cardiomyopathy (red). Solid gray line, Xten platform; dashed gray line, 2000/2500 platform.

Figure S4. The number of cardiomyopathy gene nsSNVs per person qualitatively predicts DCM across frequency bins.
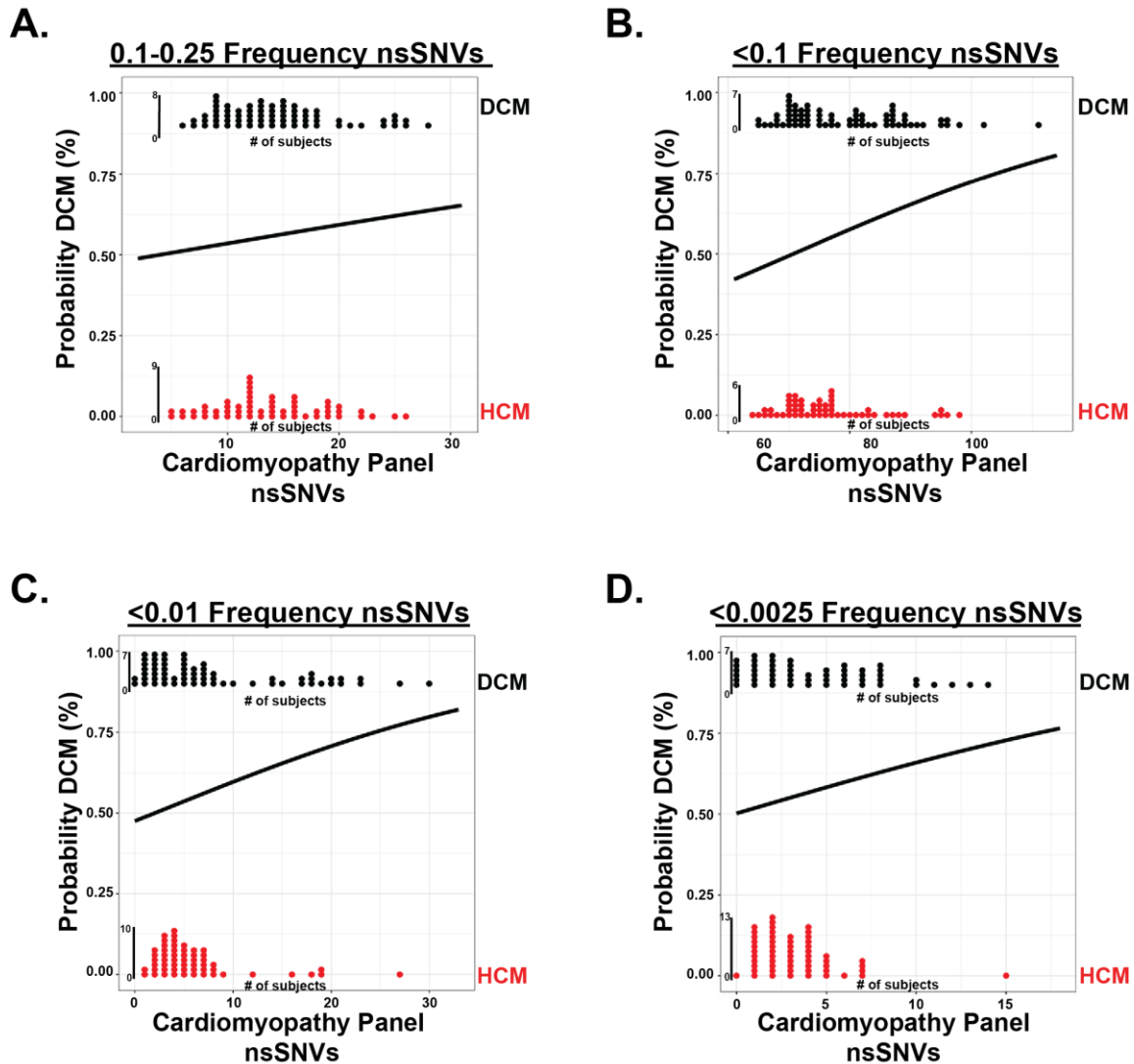

Multivariate generalized linear models suggest that the probability of DCM increases with the number of nsSNVs per person across frequency bins in this cohort. The red and black dots represent individual participants and their number of nsSNVs in cardiomyopathy genes. (**A**) 0.1-0.25; (**B**) <0.1 (**C**) <0.01 and (**D**) <0.0025. DCM (black), HCM (red). Number on X axis indicates the number of nsSNVs per subject.

Figure S5. Dichotomizing heart genes into high expression and low expression.
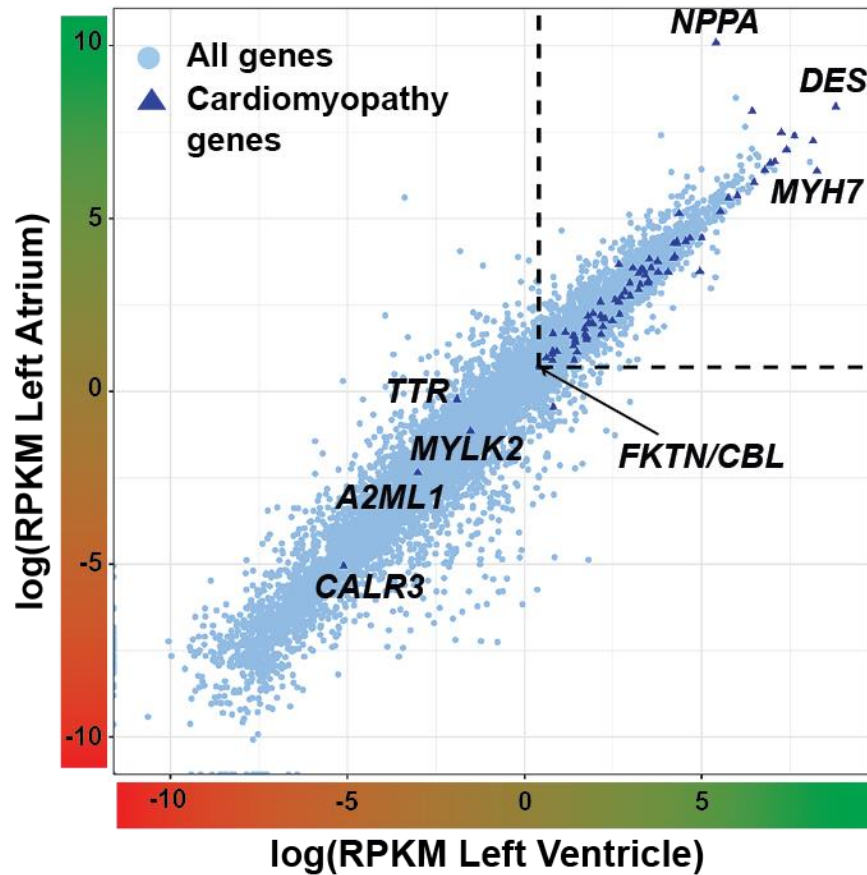
The expression of all genes (light blue circles) and genes linked to cardiomyopathies (dark blue triangles) was assessed in the Genotype-Tissue Expression (GTEx) database. Genes linked to cardiomyopathy were those found in commercial and academic gene testing panels (n=102 genes). The logs of RPKM (Reads Per Kilobase Million) for the left atrium (left atrial appendage, n=264 samples) and left ventricle (n=272 samples) were plotted against each other for all genes with at least one variant in the HCM/DCM cohort (n=14915 genes in total). Ninety-eight genes of the 102 cardiomyopathy genes were contained in a well-defined subset of the expression space bounded from below by the expression values for *FKTN* (GTEx LA = 2.00) and *CBL* (GTEx LV = 1.49); denoted by the black dotted lines. All genes with GTEx values below these cutoffs were considered low-expression heart genes (7,609 genes). All genes with values equal to or greater than both of the above cutoffs constitute high-expression heart genes (7,306) genes. Expression heat maps at each axis show increasing expression from red to green. RPKM, reads per kilobase million.

Figure S6. Workflow to correct allele frequencies.
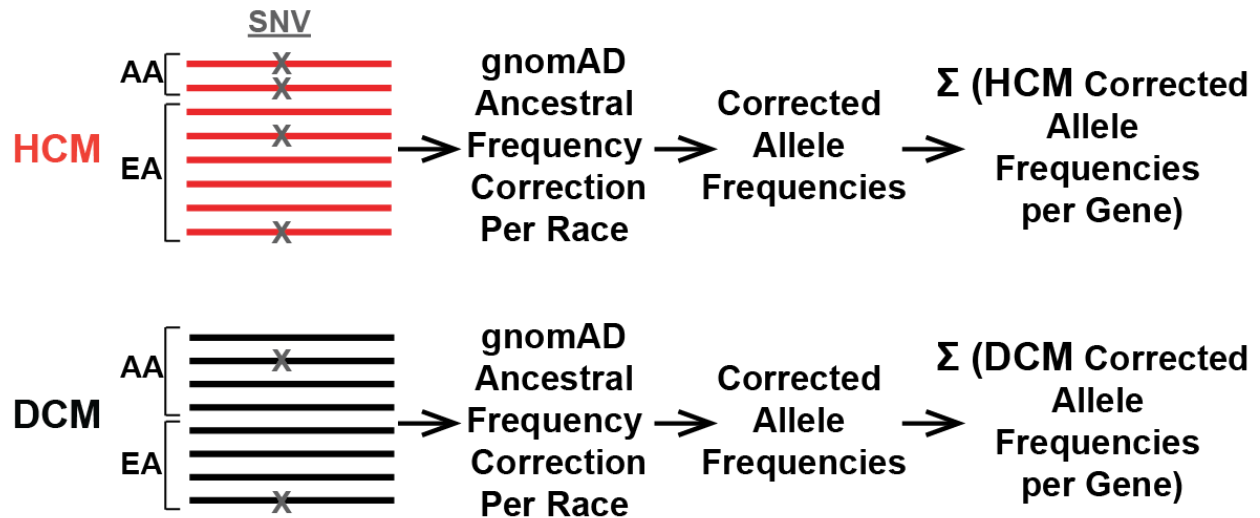
Schematic illustrating ancestral allele frequency correction performed on each nsSNV. nsSNV count was determined in cardiomyopathy and ancestry subgroups, and adjusted by gnomAD ancestral allele frequency. Corrected allele frequencies were summed across each gene by cardiomyopathy subgroup and ancestry. Delta is the difference in corrected allele frequency per gene between DCM and HCM. snSNV=nonsynonymous sinlge