

A gene-centric strategy for identifying disease-causing rare variants in dilated cardiomyopathy

Claire Horvat, PhD¹, Renee Johnson, PhD¹, Lien Lam, PhD^{2,3}, Jacob Munro, BSc, MIT⁴, Francesco Mazzarotto, PhD⁵, Angharad M. Roberts, MRCP⁵, Daniel S. Herman, MD, PhD², Michael Parfenov, PhD², Alireza Haghghi, MD, PhD^{2,3,6}, Barbara McDonough, RN², Steven R. DePalma, PhD², Anne M. Keogh, MD^{7,8,9}, Peter S. Macdonald, MBBS, PhD^{7,8,9}, Christopher S. Hayward, MD^{7,8,9}, Amy Roberts, MD¹⁰, Paul J. R. Barton, PhD⁵, Leanne E. Felkin, PhD⁵, Eleni Giannoulatou, DPhil⁴, Stuart A. Cook, PhD, MRCP^{5,11}, J. G. Seidman, PhD^{2,3}, Christine E. Seidman, MD^{2,6} and Diane Fatkin, MD^{1,8,9}

Purpose: We evaluated strategies for identifying disease-causing variants in genetic testing for dilated cardiomyopathy (DCM).

Methods: Cardiomyopathy gene panel testing was performed in 532 DCM patients and 527 healthy control subjects. Rare variants in 41 genes were stratified using variant-level and gene-level characteristics.

Results: A majority of DCM cases and controls carried rare protein-altering cardiomyopathy gene variants. Variant-level characteristics alone had limited discriminative value. Differentiation between groups was substantially improved by addition of gene-level information that incorporated ranking of genes based on literature evidence for disease association. The odds of DCM were increased to nearly 9-fold for truncating variants or high-impact missense variants in the subset of 14 genes that had the strongest biological links to DCM ($P < 0.0001$). For some of these genes,

DCM-associated variants appeared to be clustered in key protein functional domains. Multiple rare variants were present in many family probands, however, there was generally only one “driver” pathogenic variant that cosegregated with disease.

Conclusion: Rare variants in cardiomyopathy genes can be effectively stratified by combining variant-level and gene-level information. Prioritization of genes based on their a priori likelihood of disease causation is a key factor in identifying clinically actionable variants in cardiac genetic testing.

Genetics in Medicine (2019) 21:133–143; <https://doi.org/10.1038/s41436-018-0036-2>

Keywords: Dilated cardiomyopathy; pathogenic variant Genetic testing; Next-generation sequencing

INTRODUCTION

Determining variant pathogenicity is *the* major challenge in cardiomyopathy genetic testing, spawning debate in clinics worldwide. Sequence data interpretation has been biased historically by a focus on affected patients and a full appreciation of the normal spectrum of variation in cardiomyopathy “disease genes” has been lacking. In recent years, it has become apparent that many rare protein-altering variants reported to be pathogenic are also seen in the general population.^{1–4} These disturbing findings raise doubts about the discriminative efficacy of variant annotation pipelines used in literature reports and the reliability of genetics results provided to patients. Given the increasing role of personal genome sequencing in clinical practice, there is a critical need for an improved strategy for

identifying the subset of rare variants that are truly disease-causing.

Dilated cardiomyopathy (DCM) is the most common cardiomyopathy and frequently has a genetic etiology. More than 100 genes have been implicated to date.^{5–8} Genetic testing panels generally contain sets of putative DCM-associated genes and have expanded over time to include genes with direct and indirect links to other cardiac and skeletal myopathies. This inclusive approach potentially reduces negative screening results but magnifies the problem of rare variant interpretation.

The aim of this study was to investigate parameters that might improve discrimination between rare cardiomyopathy gene variants in patients with DCM and healthy control subjects. Our analysis included evaluation of variant-level and

¹Molecular Cardiology Division, Victor Chang Cardiac Research Institute, Sydney, New South Wales, Australia; ²Department of Genetics, Harvard Medical School, Boston, Massachusetts, USA; ³Howard Hughes Medical Institute, Boston, Massachusetts, USA; ⁴Molecular, Structural and Computational Biology Division, Victor Chang Cardiac Research Institute, Sydney, New South Wales, Australia; ⁵NIHR Cardiovascular Biomedical Research Unit, Royal Brompton & Harefield NHS Foundation Trust, and Imperial College London, London, UK; ⁶Cardiovascular Division, Brigham and Women’s Hospital, Boston, Massachusetts, USA; ⁷Cardiac Physiology and Transplantation Division, Victor Chang Cardiac Research Institute, Sydney, New South Wales, Australia; ⁸Cardiology Department, St Vincent’s Hospital, Sydney, New South Wales, Australia; ⁹Faculty of Medicine, University of New South Wales, Sydney, New South Wales, Australia; ¹⁰Boston Children’s Hospital, Boston, Massachusetts, USA; ¹¹Duke-National University of Singapore, and National Heart Center, Singapore, Singapore. Correspondence: Diane Fatkin (d.fatkin@victorchang.edu.au)

Submitted 14 December 2017; accepted: 27 March 2018

Published online: 11 June 2018

gene-level characteristics, as well as assessment of the total numbers of rare variants in each individual. This type of information about personal burden of rare variants is unable to be determined from studies of single genes or small gene panels in DCM patients, and is generally not available in population databases where single variants in de-identified subjects are listed separately. Our findings provide a new framework for variant prioritization and highlight the key role of a subset of genes in DCM pathogenesis.

MATERIALS AND METHODS

Study subjects

Participants provided informed written consent, and protocols were approved by the institutional human ethics committees. Patients with familial or sporadic idiopathic DCM ($n = 532$, aged 47 ± 19 years, 69% males) and healthy control subjects ($n = 527$, aged 49 ± 15 years, 38% males) were recruited from Brigham and Women's Hospital, Boston Children's Hospital, Victor Chang Cardiac Research Institute, Royal Brompton & Harefield NHS Foundation Trust, and the London Institute of Medical Sciences, Imperial College, London (Supplementary Methods, Supplementary Fig. S1). The absence of cardiovascular disease in control subjects was ascertained by medical history, and was confirmed by cardiac magnetic resonance imaging in 319 individuals (61%). All subjects had self-reported European ancestry and population stratification was evaluated using principal component analysis (Supplementary Methods, Supplementary Table S1).

A DCM replication cohort was comprised of Australian familial DCM probands ($n = 101$, aged 45 ± 16 years, 67% males). Replication control data were obtained for subjects aged >70 years drawn from the Alzheimer's Disease Sequencing Project ($n = 2971$) and the Medical Genome Reference Bank ($n = 1144$). These replication control cohorts were selected specifically to account for the age-related penetrance of DCM and because per-person data were available.

Gene sequencing

Genomic DNA libraries were constructed using standard library preparation protocols. Fragments were ligated to adaptors, amplified, purified, then hybridized to custom arrays enriched for coding regions of genes associated with DCM and other inherited cardiac disorders. A 69-gene panel was used by the Boston laboratory ($n = 203$ DCM cases, 208 controls), and a 64-gene panel was used by the UK laboratory ($n = 320$ DCM cases, 319 controls) (Supplementary Fig. S1, Supplementary Table S2). Sequencing was performed using the Illumina HiSeq 2000 or SOLiD 5500xl platforms. Selected variants were evaluated in probands and family members using Sanger sequencing.

Sequence data analysis

Sequence data were processed and aligned to the hg19 (GRCh37) human genome reference using Novoalign and the Genome Analysis Toolkit, or Lifescope v2.5.1. Data for the 41 genes that were represented on both the 69-gene and 64-gene

panels, and for which variants were detected in cases, were included in this study (Supplementary Methods, Supplementary Table S2). Coding-sequence variants that were truncating (stop gain, splice donor or acceptor site gain or loss, frameshift indels) or missense, and that met quality metrics for mapping, read depth, and allelic balance were evaluated. Population data for variant minor allele frequency (MAF) were obtained from the Exome Sequencing Project (ESP) and Exome Aggregation Consortium (ExAC) v1 databases. In silico pathogenicity predictions for missense variants were made using PolyPhen2, SIFT, PROVEAN, and MetaSVM.^{9–11}

Gene group allocation

A PubMed literature search was undertaken to assess available evidence supporting roles in DCM pathogenesis for each of the 41 genes evaluated. A semiquantitative scoring system was devised to rank genes using grades of A (strong) to D (weak) based on genetic, in vitro, and animal model data (Supplementary Methods).

Statistical analysis

Statistical comparisons between groups were made using Fisher's exact or chi-squared (2×2 , 2×4) tests. To identify the cells that contributed most to 2×4 chi-squared tests, standardized residuals were calculated and deviations of absolute value >2 were considered significant. To evaluate variant distribution in annotated domains of group A genes, missense variants identified in DCM cases were pooled with those of Walsh *et al.*¹² and compared with ExAC missense variants (MAF $<0.1\%$) within the same protein domains using Fisher's exact test. $P < 0.05$ was used to denote statistical significance.

RESULTS

Rare variants in cardiomyopathy gene are common in cases and controls

Genetic screening was performed in 532 DCM patients and 527 healthy control subjects. Study participants were of European ancestry, and no significant population stratification effects were detected on a principal component analysis (Supplementary Fig. S2). Variants that were protein-altering (truncating, missense) and rare (defined here by MAF $<0.1\%$ in the European [EA] subgroup in the ESP database) in a set of 41 cardiomyopathy-associated genes were evaluated. Variants that met these criteria were found in 407 (77%) DCM cases and in 348 (66%) control subjects ($P = 0.0002$), with the number of rare variants per person ranging from 0 to 13 (mean 1.63) in DCM cases and from 0 to 8 (mean 1.24) in controls ($P < 0.0001$). To explore strategies for identifying disease-associated variants, we compared 770 different variants found in DCM cases and 589 variants in control subjects (Supplementary Table S3).

TTN variants

There was an excess of truncating *TTN* variants (*TTN*tv) in DCM cases (83 [11%] variants versus controls, 6 [1%] variants; $P < 0.0001$; Supplementary Table S4). *TTN* missense

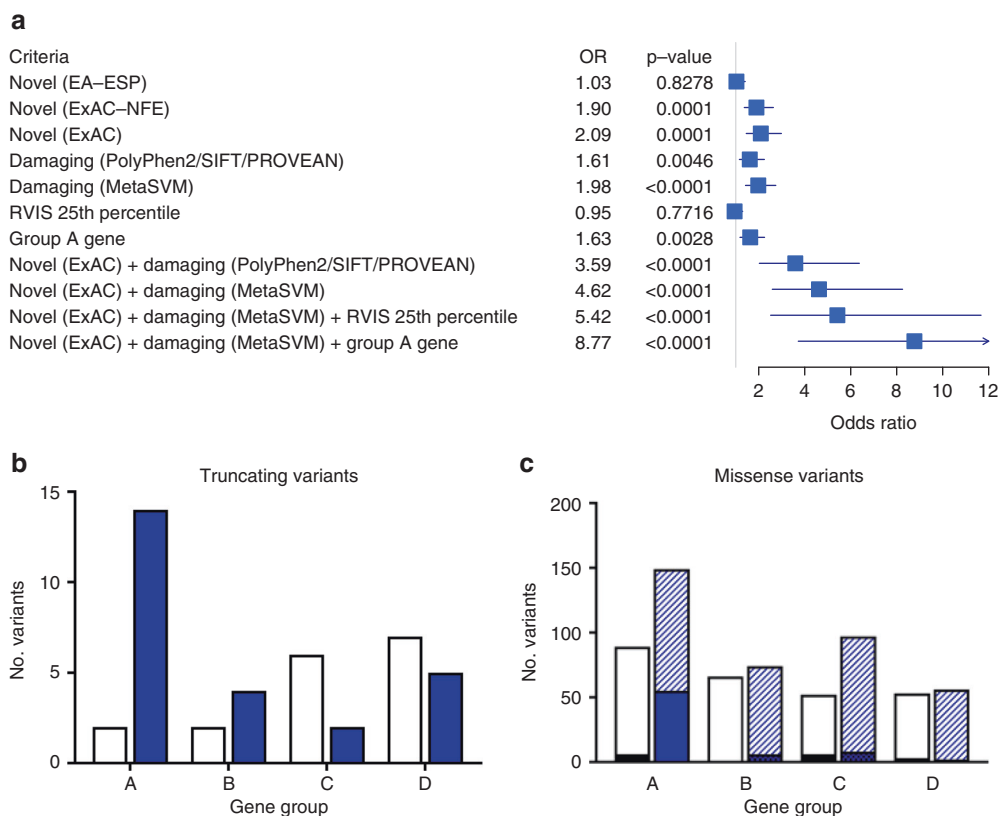


Fig. 1 Criteria for variant prioritization. (a) Missense rare variants in cardiomyopathy genes in DCM cases and controls were compared using a number of criteria including novelty with respect to the European subgroup of the Exome Sequencing Project (EA-ESP) and non-Finnish European (NFE) or all subjects in the Exome Aggregation Consortium (ExAC), in silico functional predictions (PolyPhen2, SIFT, PROVEAN, MetaSVM), Residual Variation Intolerance Score (RVIS), and gene group. The odds ratios (OR) for DCM are displayed in a forest plot. (b) Distribution of truncating variants across gene groups in controls (white bars) and DCM cases (blue bars). (c) Distribution of missense variants across gene groups in controls (white bars) and DCM cases (blue bars). The subset of “novel (ExAC) + deleterious (MetaSVM)” missense variants are denoted in the solid sections of bars. There were significant effects of clinical status and gene group for truncating variants (b, $P = 0.01$; 2×4 chi-squared test) and missense variants (c; $P = 0.018$); see Supplementary Table S8 for statistical analysis. For all panels, *TTN* truncating and missense variants were excluded

variants comprised ~40% of all variants found in cases and controls. Unlike missense variants in other genes (see below), *TTN* missense variants were not differentiated between groups by assessment of novelty or in silico functional predictions (Supplementary Table S5), and these were excluded from subsequent variant-level, gene-level, and per-person analyses.

Limited discriminative value of variant characteristics

There were relatively few truncating variants in cardiomyopathy genes other than *TTN*, and these were equally seen in DCM cases (25 [3%] variants) and controls (17 [3%] variants; $P = 0.75$). To start to stratify the larger numbers of missense variants found in both groups, we looked at novelty and in silico functional predictions.

Variant novelty

The absence of a variant in healthy subjects has been used as a criterion of pathogenicity in many DCM pathogenic variant reports.^{13–15} However, assessment of novelty varies with the size of the reference cohort, and differences between groups were only found by using the ExAC database

(>60,000 subjects), with DCM cases having more novel variants than controls ($P < 0.0001$; Supplementary Table S5). Surprisingly, more than half of the missense variants in our data set have been identified previously in genetic testing studies and are listed in ClinVar (Supplementary Table S5).

Variant functional predictions

To assess variant function, we evaluated subsets of variants that were (1) uniformly predicted to be deleterious in PolyPhen2, SIFT, and PROVEAN,^{9,10} or (2) predicted deleterious by MetaSVM, which provides an ensemble score derived from 10 separate algorithms.¹¹ Both of these approaches yielded significant differences between DCM cases and controls, with modestly better results for MetaSVM’s ensemble score than for the three-program consensus score (Supplementary Table S5). When novelty and functional predictions were combined, discrimination between groups was increased: DCM cases were twice as likely as controls to have variants that were absent from ExAC or deleterious (MetaSVM), and nearly five-fold more likely to have variants that were absent from ExAC and deleterious (MetaSVM) (Fig. 1a).

Importance of gene-based parameters

Because variant-level parameters incompletely differentiated DCM cases from controls, we next sought to determine whether there might be a hierarchy for pathogenicity between genes.

ExAC metrics of expected genetic variation

ExAC provides two gene constraint metrics for expected variation in the general population.¹⁶ The “probability of being loss-of-function intolerant” (pLI) scores for our 41 genes ranged from 0 to 1,¹⁶ with only 10 genes assessed as extremely intolerant (pLI > 0.9; Table 1). DCM-associated truncating variants were enriched in genes with high pLI scores (11 of 23 [48%] variants versus controls, 1 of 17 [6%] variants, $P = 0.005$), but >50% variants were in genes with intermediate or low pLI scores. Twenty-six genes had expected intolerance to missense variants, indicated by positive Z scores (Table 1).¹⁶ There was a statistically significant but relatively modest excess of missense variants in these genes in DCM cases (182 [52%] variants versus controls, 129 [43%] variants; $P = 0.018$).

Residual Variation Intolerance Score (RVIS)

RVIS is another metric of expected genetic variation.¹⁷ In this scoring system, genes that cause Mendelian disorders should have little population variability (25th percentile RVIS), while genes associated with complex disorders are likely to be highly variable (100th percentile RVIS). In keeping with the Mendelian model, we found that 18 of our genes were in the 25th percentile bin (Table 1). However, while DCM-associated truncating variants were preferentially distributed in 25th percentile genes ($P = 0.01$, 2×4 chi-squared test; Supplementary Table S6), this was not the case for missense variants ($P = 0.67$).

Variant burden per gene

There were only three genes, *RBM20*, *MYH7*, and *LMNA*, in which there were significantly more rare variants in DCM cases than in controls (Supplementary Table S3). These findings strongly suggest that these genes harbor pathogenic rare variants, but do not directly inform the interpretation of any single variant. For most genes, the relatively small numbers of variants limited statistical comparisons.

DCM gene ranking

We used a semiquantitative method to rank genes according to literature evidence for roles in DCM pathogenesis (Table 1, Supplementary Table S7). Only 14 of 41 genes had robust evidence for disease association and were classified as group A. There was a prominent peak of DCM-associated truncating variants in these genes, with control-associated truncating variants mostly found in gene groups C and D ($P = 0.01$, 2×4 chi-squared test; Fig. 1b, Supplementary Table S8). DCM-associated missense variants also showed a peak in group A genes ($P = 0.018$), with marked enrichment of variants that were novel (ExAC) + deleterious (MetaSVM) ($P = 0.001$, Fig. 1c). Adding the “group A” gene parameter to these

variant parameters increased the odds ratio for DCM from approximately 5-fold to nearly 9-fold (Fig. 1a).

A similar pattern was seen for the subset of ClinVar-listed truncating and missense variants, with a majority of pathogenic/likely pathogenic variants occurring in DCM cases and in group A genes ($P = 0.0007$, Supplementary Fig. S3A). Variants annotated specifically as pathogenic/likely pathogenic for DCM were exclusive to group A genes, comprising 17 of the 94 (18%) group A variants in DCM cases and 1 of 65 (1.5%) group A variants in controls. Of the 30 total pathogenic/likely pathogenic group A variants in DCM cases, 24 (80%) were also identified as “damaging” by our criteria (Supplementary Figure S3B). The 6 ClinVar variants that were not captured by our method included 5 very rare variants (1 to 5 alleles in ExAC), 3 of which were annotated pathogenic/likely pathogenic for DCM, and one variant (40 alleles in ExAC) that was likely pathogenic for HCM.

Yield of variants per person

We next looked at the yield per person of prioritized variants (Table 2). For this analysis, truncating variants and “novel (ExAC) + deleterious (MetaSVM)” missense variants are referred to as “damaging” group A gene variants (Supplementary Table S9). Damaging variants were present in 65 (12.2%) DCM cases and in 8 (1.5%) control subjects ($P < 0.0001$). Similar patterns for the prevalence of damaging group A variants were seen in an independent cohort of familial DCM cases (17 probands [16.8%]), and in two replication control cohorts from the Alzheimer’s Disease Sequencing Project (54 subjects [1.8%]) and the Medical Genome Reference Bank (33 subjects [2.9%]).

When *TTNtv* were also considered, 151 (28.4%) of our DCM cases were positive for damaging variants and/or *TTNtv* compared with 14 (2.7%) controls ($P < 0.0001$). Damaging variants and/or *TTNtv* were present in 32 (31.7%) probands in the familial DCM replication cohort, and in 61 (2.1%) and 46 (4.0%) subjects, respectively, in the two control replication cohorts. It has been questioned whether *TTNtv* are sufficient alone to cause DCM or require “second hit” genetic and/or acquired factors for DCM manifestation. We found no differences in the background burden of variants in *TTNtv* + DCM cases (1.58 ± 1.74 variants, range 0–9) when compared with *TTNtv*-/damaging + DCM cases (1.17 ± 1.22 variants, range 0–5), *TTNtv*-/damaging–DCM cases (1.30 ± 1.55 variants, range 0–13), or *TTNtv*-/damaging–control subjects (1.20 ± 1.29 , range 0–8; $P = 0.3819$, Kruskal–Wallis test). These data suggest that *TTNtv* carriers generally do not show an excess of second damaging variants.

Family segregation

To further test our criteria for variant prioritization, we performed cosegregation analysis in 28 DCM families in which DNA samples from ≥ 3 informative individuals were available. All rare variants that were identified in each family proband were evaluated in the respective family members.

Table 1 Gene characteristics and evidence for role in DCM pathogenesis

Gene	Protein	Gene group	RVIS ^a score ^b	pLI score ^b	Z score ^b	Genetic data ^c	In vitro defects	Animal model ^d	Other phenotypes
<i>MYH7</i>	Myosin heavy chain 7	A	0.28	0.00	6.54	+++	Yes	+++ ^e	HCM, RCM, LVNC, CHD, DM, MSM, SPM
<i>DSP</i>	Desmoplakin	A	1.32	1.00	0.91	+++	Yes	+++	ARVC, LVNC, palmoplantar keratoderma, skin-fragility woolly hair syndrome
<i>SCN5A</i>	Cardiac sodium channel	A	1.99	1.00	2.53	+++	Yes	+++	CD, AF, LQTS, BS, VF, LVNC
<i>LMNA</i>	Lamin A/C	A	9.96	0.99	3.37	+++	Yes	+++	CD, EDMD, >10 additional phenotypes/overlap syndromes.
<i>LDB3</i>	LIM domain binding protein 3	A	10.16	0.00	0.32	+++	Yes	+++	HCM, LVNC, ARVC, MM
<i>DMD</i>	Dystrophin	A	11.28	1.00	-4.82	+++	Yes	+++	DMD, BMD
<i>DES</i>	Desmin	A	21.56	0.00	2.34	+++	Yes	+++	RCM, ARVC, HCM, CD, MM, LGMD-2R, NSS
<i>TPM1</i>	Tropomyosin α -1 chain	A	25.15	0.80	3.42	+++	Yes	+++	HCM, RCM, LVNC
<i>TMNC1</i>	Troponin C	A	39.68	0.51	2.22	+++	Yes	+++	HCM
<i>TMNT2</i>	Troponin T	A	41.64	0.01	1.54	+++	Yes	+++	HCM, RCM, LVNC
<i>BAG3</i>	BCL-2 associated athanogene	A	62.14	0.53	-1.01	+++	Yes	+++	MM
<i>PLN</i>	Phospholamban	A	62.38	0.11	0.57	+++	Yes	+++	HCM, ARVC
<i>RBM20</i>	RNA-binding protein 20	A	90.98	NA	NA	+++	Yes	+	
<i>TTN</i>	Titin	A	98.04	0.00	-5.48	+++	Yes	++	HCM, ARVC, RCM, LVNC, CM, CNM, HMERF, LGMD-2J, TMD
<i>VCL</i>	Vinculin	B	2.23	0.99	3.11	++	Yes	++	HCM
<i>LAMA4</i>	Laminin subunit α -4	B	4.68	0.00	-1.14	+	Yes	++	
<i>ILK</i>	Integrin-linked protein kinase	B	14.97	0.04	0.96	+	Yes	++	
<i>MYBPC3</i>	Cardiac myosin binding protein 3	B	20.54	0.00	0.69	++	Yes	++	HCM, LVNC
<i>CSRP3</i>	Muscle LIM protein	B	28.93	0.00	-0.66	++	Yes	++	HCM
<i>CRYAB</i>	α β -crystallin	B	35.42	0.01	0.38	++	Yes	+++	MM, DM, cataracts
<i>ACTC1</i>	Cardiac actin	B	38.28	0.95	5.25	++	Yes	+++	HCM, RCM, LVNC, CHD
<i>NEXN</i>	Nexilin	B	50.34	0.00	-1.32	++	NA	++	HCM, CHD
<i>TAZ</i>	Tafazzin	B	59.76	0.97	2.42	+++	Yes	++	Barth syndrome (may present as infantile X-linked DCM), LVNC, EFE
<i>LAMP2</i>	Lysosome-associated membrane glycoprotein 2	B	61.73	0.95	0.41	++	Yes	+	Danon disease (may present as HCM or DCM)
<i>TMNB</i>	Troponin I	B	64.11	0.17	1.88	++	Yes	+	HCM, RCM, AF
<i>ANKRD1</i>	Cardiac ankyrin repeat protein	B	65.76	0.00	-0.01	++	Yes	No	HCM, CHD
<i>DSG2</i>	Desmoglein-2	B	98.32	0.00	-1.20	++	NA	+++	ARVC
<i>MYH6</i>	Myosin heavy chain 6	C	0.66	0.00	2.87	++	NA	++ ^e	HCM, CHD, SSS
<i>ABCC9</i>	K _{ATP} channel SUR2A subunit	C	1.82	0.00	4.89	+	Yes	++	AF, BS, Cantu syndrome, coronary vasospasm
<i>ACTN2</i>	α -actinin	C	2.47	1.00	1.76	+	Yes	+	HCM
<i>PDLIM3</i>	PDZ and LIM domain protein 3	C	20.70	0.00	-0.04	+	NA	++	
<i>SGCD</i>	δ -sarcoglycan	C	34.32	0.00	-0.23	++	Yes	++	LGMD-2F
<i>SYNE1</i>	Nesprin-1	C	41.65	0.00	-0.95	++	Yes	+	EDMD, cerebellar ataxia, autism, arthrogyriosis
<i>FXN</i>	Frxatin	C	76.67	0.82	0.47	+	NA	++	Friedreich's ataxia (may develop LVH + late DCM)

Table 1 (continued)

Gene	Protein	Gene group	RVIS ^a	pLI score ^b	Z score ^b	Genetic data ^c	In vitro defects	Animal model ^d	Other phenotypes
CAV3	Caveolin-3	C	77.70	0.34	1.19	+	Yes	++	HCM, LQTS, SIDS, LGMD-1C, RMD-2, DM, HCK
TCAP	Telethonin	C	78.28	0.08	0.45	+	Yes	+	HCM, LGMD-2G
FKTN	Fukutin	C	82.25	0.00	-0.64	++	Yes	NA	LGMD-2M (may present as DCM)
SDHA	Succinate dehydrogenase flavoprotein subunit	D	10.95	0.00	2.32	++	Yes	NA	Leigh syndrome (rarely associated with pediatric recessive DCM)
DTNA	α-dystrobrevin	D	21.65	0.92	1.17	+	NA	No	LVNC, CHD
LAMA2	Laminin subunit α-2	D	79.64	0.00	-1.58	+	NA	NA	Merosin-deficient congenital MD/LGMD (rarely includes DCM)
SYMM	Synemin	D	NA	0.00	-0.15	NA	NA	NA	Accumulates in skeletal myopathy

^a Residual Variation Intolerance Score (RVIS)—see text and reference¹⁷

^b Probability of being loss-of-function intolerant (pLI) and missense Z scores—see text and reference¹⁶

^c Genetic evidence: +++ cosegregation of variants in multiple (3+) families, with at least one family showing variant cosegregation in 5 or more affected individuals; ++ cosegregation of variant in 1 or more small families (<5 affected); + variant demonstrated in single cases (with/without positive family history), or single-nucleotide polymorphisms (SNPs) associated with DCM susceptibility

^d Animal models: +++ spontaneous development of DCM in animal model with a heterozygous gene loss-of-function or expressing a human disease-associated nonsynonymous variant; ++ spontaneous DCM present in homozygous gene loss-of-function model, or stress-induced DCM in model of a human nonsynonymous variant; + no DCM but other relevant phenotypes, e.g., isolated ventricular dilation or contractile dysfunction, evidence of reduced contractile reserve, skeletal myopathy. See Supplementary Methods and Supplementary Table S7 for more details of gene group allocation and animal data

^e α-MHC is major myosin isoform in murine ventricle: human MYH7 pathogenic variants modeled in murine myh6 gene

AF atrial fibrillation, ARVC arrhythmogenic right ventricular cardiomyopathy, BMD Becker muscular dystrophy, BS Brugada syndrome, CD conduction-system defects, CHD congenital heart disease, CM congenital myopathy, CMM centronuclear myopathy, DCM dilated cardiomyopathy, DM distal myopathy, DMD Duchenne muscular dystrophy, EDMD Emery-Dreifuss muscular dystrophy, EFE endomyocardial fibroelastosis, HCK high creatine kinase, HCM hypertrophic cardiomyopathy, HMERF hereditary myopathy with early respiratory failure, LGMD limb girdle muscular dystrophy, LQTS long QT syndrome, LVH left ventricular hypertrophy, LVWC left ventricular noncompaction, MD muscular dystrophy, MM myofibrillar myopathy, MSM myosin storage myopathy, NA not available, NSS neurogenic scapuloperoneal syndrome, RCM restrictive cardiomyopathy, RMD rippling muscle disease, SIDS sudden infant death syndrome, SPM scapuloperoneal myopathy, SSS sick sinus syndrome, TMD tibial muscular dystrophy, VF ventricular fibrillation

Table 2 Yield per person of *TTN*tv and damaging^a group A gene variants in the primary cohort and in replication cohorts

Variants	DCM cases		Control subjects			
	Primary cohort (n = 532 subjects)	FDCM (n = 101 subjects)	Primary cohort (n = 527 subjects)	Alzheimer's (n = 2971 subjects)	MGRB (n = 1144 subjects)	
Single <i>TTN</i> tv	85 (16.0%)	15 (14.9%)	6 (1.1%)	6 (0.2%)	13 (1.1%)	
Two <i>TTN</i> tv	1 (0.2%)	0	0	0	0	
<i>TTN</i> tv + group A trunc	1 (0.2%)	0	0	0	0	
<i>TTN</i> tv + group A missense	5 (0.9%)	1 (1.0%)	0	0	1 (0.1%)	
Group A trunc only	13 (2.4%)	6 (5.9%)	2 (0.4%)	9 (0.3%)	3 (0.3%)	
Group A trunc + missense	0	1 (1.0%)	0	0	0	
Group A missense only	46 (8.6%) ^b	9 (8.9%)	6 (1.1%)	46 (1.5%)	29 (2.5%)	
None	381 (71.6%)	69 (68.3%)	513 (97.3%)	2910 (97.9%)	1098 (96.0%)	

^a Damaging variants included truncating (trunc) or "novel (ExAC) + deleterious (MetaSVM)" missense variants in group A genes (*TTN* missense variants were excluded)

^b Three cases had 2 variants

ADSP Alzheimer Disease Sequencing Project, FDCM familial dilated cardiomyopathy, MGRB Medical Genome Reference Bank

Truncating group A gene variants

Group A gene truncating variants were evaluated in three families (Supplementary Figure S4). In family BY, the proband carried a *BAG3* stop codon with segregation analysis consistent with linkage (odds ratio 1/18, logarithm of the odds (LOD) score 1.3). The truncated protein would lack the signature BAG domain that binds to HSP70 and is required for chaperone activity.¹⁸ Numerous *BAG3* truncating pathogenic variants have been reported in DCM patients, and reduced levels of *BAG3* protein have been seen in ventricular tissues from patients with heart failure.^{19–21}

The family DF proband and her affected brother carried a *MYH7* splice acceptor site variant. This variant was absent from three unaffected individuals aged >40 years, but the family size was insufficient to show statistically significant linkage (odds ratio 1/6, LOD score 0.8). *MYH7* truncating variants were found in two additional sporadic DCM cases in our cohort (Supplementary Table S9) but have rarely been reported in DCM cases,^{12,22} and there is uncertainty about their clinical significance.

The family C proband carried a *SCN5A* stop codon that would truncate the C-terminus with loss of a PY-motif (binding site for Nedd4 ubiquitin ligase) and a PDZ domain binding motif (interacts with the cytoskeletal adapter protein, syntrophin), potentially giving rise to protein degradation or trafficking defects.^{23,24} The cardiac phenotype associated with *SCN5A* pathogenic variants often includes arrhythmias and conduction-system abnormalities as well as DCM,^{25,26} and atrial fibrillation was present in all affected individuals in this kindred. The apparent nonsegregation of this variant with DCM (LOD score –2.5) may be confounded by two possible phenocopies who had other plausible acquired causes of disease.

Missense variants

Segregation of 30 missense variants was evaluated in 14 kindreds (Supplementary Table S10). Only 4 of the 30 variants had LOD scores >1 (suggestive of linkage relative to family size) and all of these were damaging group A gene variants (Supplementary Fig. S4). Affected individuals in family CZ carried a p.Arg369Gln *MYH7* variant and had both DCM and left ventricular noncompaction. This variant, in the myosin motor domain, has been associated with both phenotypes.^{6,27} The p.Arg634Trp *RBM20* variant, present in all affected individuals in family AB, lies within the arginine–serine-rich (RS) domain, which is a putative DCM pathogenic variant hotspot.^{28,29} There were also two cosegregating *DES* variants, p.Leu398Pro (family FK) and p.Lys449Thr (family FG), the latter associated with myofibrillar myopathy.³⁰ Four additional kindreds (families BG, BK, FR, KS) had damaging group A gene variants with LOD scores <1 (Supplementary Fig. S5). The damaging variants were present in 15 of 16 affected individuals in these families, with genotype–phenotype discordance driven mainly by nonpenetrant variant carriers. In all families tested, there was no evidence for cosegregation of any variants that did not meet

the criteria for being “damaging” or that were in genes other than group A.

Multiple variants

With expanded genetic testing panels, it is not uncommon to find several rare variants in DCM cases, prompting hypotheses of multiple pathogenic variants.^{7,31} In an extension of the current analysis, we reviewed data for individuals who had undergone testing with the 69-gene testing panel and looked at all 69 genes as well as increasing the threshold level of MAF to <1%. Family cosegregation analysis was then performed in five kindreds in which the proband was found to carry ≥ 5 variants. In four of these families (families BY, FK, AB, BA), only one of the proband’s multiple variants segregated with DCM (Supplementary Figure S6). In the remaining family (family GX), there were no cosegregating variants. Four of the proband’s five rare variants were inherited from her unaffected father, and none were damaging group A variants. These studies demonstrate the value of family analysis and support the expectation that a single “driver” variant will be present when DCM appears as a Mendelian trait.

Location of variants in group A genes

To explore potential effects of variant location within group A genes, we compared the distribution of missense variants identified in DCM cases (derived from our own cohorts and two clinical laboratories¹²) with rare (MAF < 0.1%) missense variants in ExAC (Fig. 2, Supplementary Table S11). In *MYH7*, DCM variants occurred in all protein domains but were preferentially located in the myosin motor (Fig. 2a). DCM-associated variants in *RBM20* were significantly enriched in the RS domain that is associated with altered titin splicing and myocardial compliance (Fig. 2b).^{28,29} More than half of the damaging *RBM20* variants resided in the RS domain or in a glutamate-rich region that is also associated with titin splicing defects.³² DCM-associated *LMNA* variants were mostly located in the coiled-coil rod domain, particularly in coil 2, which is critical for dimer formation (Fig. 2c),³³ and there was a cluster of DCM-associated variants in the S4 voltage sensor, repeat I, of *SCN5A* (Fig. 2e). Variants in the S4 transmembrane segments cause arrhythmic forms of DCM and have been associated with gain-of-function effects and gating pore currents^{25,26,34–36}. Clustering of DCM variants was also apparent in the α -tropomyosin binding domain of *TNNT2*, and several damaging variants in *TPMI* resided in the cardiac troponin T binding domain (Fig. 2g,h).

DISCUSSION

Our data show that rare protein-altering cardiomyopathy gene variants are not unique to DCM patients but also occur commonly in apparently healthy individuals. Because of this, effective discriminative strategies are clearly needed to derive meaningful results from genetic testing. Here we provide a new approach for assessment of rare variants that combines both variant-level and gene-level information. We propose

that ranking genes based on their a priori likelihood of disease causation is a key factor in identifying clinically actionable variants.

Only 14 of 41 genes achieved “group A” status. Our literature-based gene grading method was subsequently validated by the predilection for damaging variants in DCM cases to occur in group A genes and by the similar clustering of pathogenic/likely pathogenic ClinVar variants. Gene group is a dynamic parameter, and is highly dependent on the contemporary knowledge base. Assessment of genetic evidence is intrinsically biased by the size of the kindreds evaluated, the frequency in which genes have been screened, and published reports, and definitive conclusions cannot be drawn if there is insufficient information. Genes in which variants cosegregated with DCM in multiple large kindreds (≥ 5 affected) scored highly but for many genes, only small families or single cases have been studied, raising the possibility that variants, even if function-altering, might cosegregate by chance or be incidental findings. Similarly, in vitro functional data are helpful if these are available but are often lacking from the literature. Group A genes typically had animal data showing spontaneous development of DCM in models expressing heterozygous loss-of-function alleles or human missense pathogenic variants. This level of evidence was missing from many studies in which only homozygous loss-of-function animals have been studied. While such models can implicate genes in normal cardiac function, it cannot be assumed that there is a direct correlation with gene “dose” and that heterozygous counterparts will have a similar, albeit less severe, phenotype. In homozygous loss-of-function animals, there may also be profound effects on cardiac development that independently predispose to contractile impairment.

The group A genes, *MYH7*, *RBM20*, and *LMNA*, had significantly more rare variants in DCM cases than controls, and frequently harbored damaging variants. *MYH7* and *LMNA* have consistently appeared on lists of the “most frequently mutated” genes in DCM genetics studies, and were the top two genes (after *TTN*) showing an excess of rare variants (MAF < 0.0001) in a recent analysis of 1315 DCM patients referred to diagnostic genetic testing laboratories.^{6,7,12,22} *RBM20* is a more recent addition to the disease gene list and has been less frequently screened. Despite their overall enrichment in DCM cases, not all rare variants in *MYH7*, *RBM20*, and *LMNA* are necessarily deleterious and additional information is required for clinical interpretation of any specific variant. Interestingly, variants in these genes had a nonrandom distribution in DCM cases: *MYH7* variants clustered in the myosin motor, *RBM20* variants in the RS domain, and *LMNA* variants in the coiled-coil rod. As the numbers of reported DCM-associated variants expands, additional patterns of variant distribution patterns may become apparent.

Our method for stratifying rare variants provides a framework for future studies and further refinements are anticipated. We expect that over time, more genes will

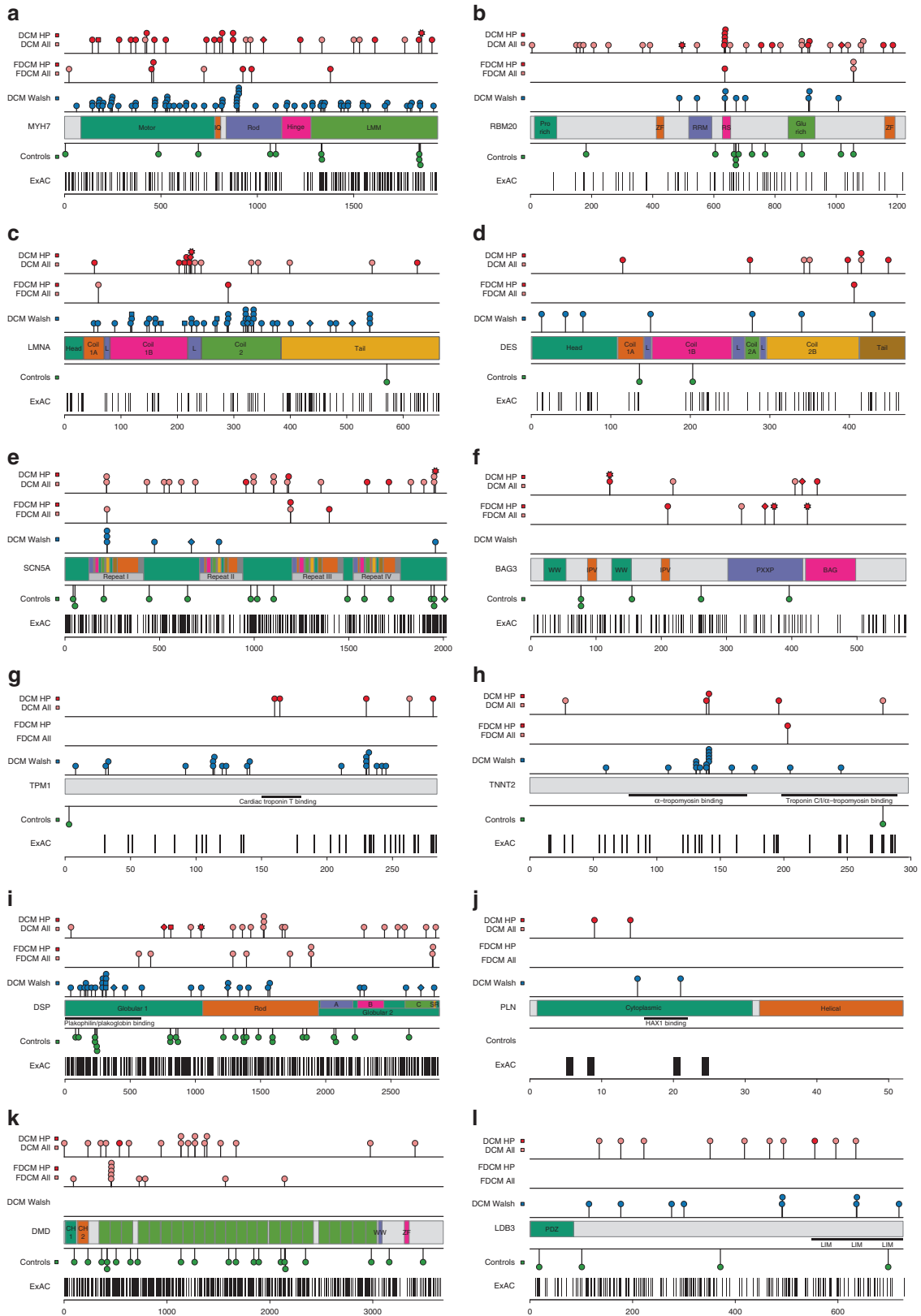


Fig. 2 Rare variant distribution in DCM cases and control subjects. The location of variants in different protein domains of group A genes is shown in panels (a–l). In each panel, variants identified in DCM patients in the primary cohort (top row) and in the familial dilated cardiomyopathy (FDCM) replication cohort (second row) are shown: all variants (pink, “DCM all”), damaging variants (red, “DCM HP”). Variant types are indicated as follows: circle = missense, square = splice site change, diamond = frameshift, star = stop codon. The third row shows DCM variants identified by two clinical diagnostic laboratories (“DCM Walsh”).¹² Missense variants identified in control subjects are shown below the protein schematic: control subjects in this study (Controls), Exome Aggregation Consortium (ExAC) database. See Supplementary Table S11 for protein domain coordinates and statistical analysis

achieve group A status, and ongoing rigorous gene ranking is warranted. For missense variants, the criterion of absence from ExAC may underestimate the yield of deleterious variants and varying MAF threshold levels need to be evaluated. It is notable, for example, that the few ClinVar-annotated pathogenic/likely pathogenic group A gene variants that were not captured by our method were mostly extremely rare in ExAC. Many of the ClinVar-listed variants are annotated as pathogenic/likely pathogenic for disorders other than DCM. This phenotypic heterogeneity suggests that at least some of these variants might be modifiers rather than directly causative of disease. Another explanation is that the ultimate clinical manifestations of any given variant may be flavored by the context of each person’s genetic and lifestyle factors. The detection of pathogenic/likely pathogenic variants for diverse disorders suggests that our prioritization method could be applied to a range of genetic testing indications, with the caveat that disease-specific gene lists would be needed.

The lack of cardiac investigations in ~40% of control subjects is a limitation of this study. We focused on subjects of European ancestry to avoid confounding effects of ethnicity-related background genetic variation, and the applicability of our findings to other populations has yet to be determined. In assessing individual burden of genetic variation, only rare cardiomyopathy gene variants were considered. The extent to which rare, low-frequency, and common variants in a broad range of cardiac genes might cumulatively contribute to a myopathic substrate is unknown.

Ongoing efforts by the American College of Medical Genetics to standardize variant annotation methods are an important step toward establishing international guidelines for genetic testing in DCM and other inherited human disorders.³⁷ The cost-efficacy of first-line screening of a core set of key DCM disease genes (e.g., group A genes) versus extended panels or genome-wide testing warrants further analysis. As genome sequencing is poised to become part of mainstream health care, there is a pressing need for curation of clinical and genetic information in databases such as ClinGen,³⁸ functional evaluation of variants, and clinical trials in genotyped patients. Advancements in these areas should expedite implementation of personalized medicine.

In summary, our data suggest that interpretation of genetic testing results cannot be made on the basis of variant characteristics alone, and that the effect size of protein-altering variants is primarily determined by the biological importance of the gene involved and relevance to DCM pathogenesis.

ELECTRONIC SUPPLEMENTARY MATERIAL

The online version of this article (<https://doi.org/10.1038/s41436-018-0036-2>) contains supplementary material, which is available to authorized users.

ACKNOWLEDGMENTS

We thank Erica Mazaika, Monique Ohanian, Santiago Pineda, Celine Santiago, Magdalena Soka, and Gunjan Trivedi for assistance with data collection, collaborating physicians who referred patients to the study, and participating families. See Supplementary Acknowledgments for the Medical Genome Research Bank.

FUNDING

This study was supported by grants from the National Health and Medical Research Council of Australia (D.F., P.S.M.), Harvard Club of Australia Foundation (D.F.), the estate of the late RT Hall (D.F.), Simon Lee Foundation (D.F.), British Heart Foundation (S.A.C.), National Institute for Health Research Cardiovascular BRU at the Brompton and Harefield & Harefield NHS Foundation Trust and Imperial College London (S.A.C., P.J.B.), Howard Hughes Medical Institute (C.E.S.), National Institutes of Health (D.S.H., J.G.S.), and the Leducq Foundation (S.A.C., J.G.S., C.E.S.).

DISCLOSURE

The authors declare no conflicts of interest.

REFERENCES

- Pan S, Caleshu CA, Dunn KE, et al. Cardiac structural and sarcomere genes associated with cardiomyopathy exhibit marked intolerance of genetic variation. *Circ Cardiovasc Genet*. 2012;5:602–10.
- Norton N, Robertson PD, Rieder MK, et al. Evaluating pathogenicity of rare variants from dilated cardiomyopathy in the exome era. *Circ Cardiovasc Genet*. 2012;5:167–74.
- Golbus JR, Puckelwartz MJ, Fahrenbach JP, Dellefave-Castillo LM, Wolfgeher D, McNally EM. Population-based variation in cardiomyopathy genes. *Circ Cardiovasc Genet*. 2012;5:391–9.
- Haggerty CM, James CA, Calkins H, et al. Electronic health record phenotype in subjects with genetic variants associated with arrhythmogenic right ventricular cardiomyopathy: a study of 30,716 subjects with exome sequencing. *Genet Med*. 2017;19:1245–52.
- Fatkin D, Seidman CE, Seidman JG. Genetics and disease of ventricular muscle. *Cold Spring Harb Perspect Med*. 2014;4:a021063.
- Pugh TJ, Kelly MA, Gowrisankar S, et al. The landscape of genetic variation in dilated cardiomyopathy as surveyed by clinical DNA sequencing. *Genet Med*. 2014;16:601–8.
- Haas J, Frese KS, Peil B, et al. Atlas of the clinical genetics of human dilated cardiomyopathy. *Eur Heart J*. 2015;36:1123–35.
- McNally EM, Mestroni L. Dilated cardiomyopathy: genetic determinants and mechanisms. *Circ Res*. 2017;121:731–48.
- Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. *Nat Methods*. 2010;7:248–9.

10. Sim NL, Kumar P, Hu J, Henikoff S, Schneider G, Ng PC. SIFT web server: predicting effects of amino acid substitutions on proteins. *Nucleic Acid Res.* 2012;40:W452–W457.
11. Dong C, Wei P, Jian X, et al. Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. *Hum Mol Genet.* 2015;24:2125–37.
12. Walsh R, Thomson KL, Ware JS, et al. Reassessment of Mendelian gene pathogenicity using 7855 cardiomyopathy cases and 60,706 reference samples. *Genet Med.* 2017;19:192–203.
13. Olson TM, Michels VV, Thibodeau SN, Tai YS, Keating MT. Actin mutations in dilated cardiomyopathy, a heritable form of heart failure. *Science.* 1998;280:750–2.
14. Fatkin D, MacRae C, Sasaki T, et al. Missense mutations in the rod domain of the lamin A/C gene as causes of dilated cardiomyopathy and conduction-system disease. *N Engl J Med.* 1999;341:1715–24.
15. Kamisago M, Sharma SD, dePalma SR, et al. Mutations in sarcomere protein genes as a cause of dilated cardiomyopathy. *N Engl J Med.* 2000;343:1688–96.
16. Lek M, Karczewski KJ, Minikel EV, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature.* 2016;536:285–91.
17. Petrovski S, Wang Q, Heinzen EL, Allen AS, Goldstein DB. Genic intolerance to functional variation and the interpretation of personal genomes. *PLoS Genet.* 2013;9:e1003709.
18. Hishiya A, Kitazawa T, Takayama S. BAG3 and Hsc70 interact with actin capping protein CapZ to maintain myofibrillar integrity under mechanical stress. *Circ Res.* 2010;107:1220–31.
19. Villard E, Perret C, Gary F, et al. A genome-wide association study identified two loci associated with heart failure due to dilated cardiomyopathy. *Eur Heart J.* 2011;32:1065–76.
20. Norton N, Li D, Rieder MJ, 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;88:273–82.
21. Feldman AM, Begay RL, Knezevic T, et al. Decreased levels of BAG3 in a family with a rare variant and in idiopathic dilated cardiomyopathy. *J Cell Physiol.* 2014;229:1697–702.
22. Kayvanpour E, Sedaghat-Hamedani F, Amr A, et al. Genotype-phenotype associations in dilated cardiomyopathy: meta-analysis on more than 8000 individuals. *Clin Res Cardiol.* 2017;106:127–39.
23. Van Bemmelen MX, Rougier JS, Gavillet B, et al. Cardiac voltage-gated sodium channel Nav1.5 is regulated by Nedd4-2 mediated ubiquitination. *Circ Res.* 2004;95:284–91.
24. Gavillet B, Rougier JS, Domenighetti AA, et al. Cardiac sodium channel Nav1.5 is regulated by a multiprotein complex composed of syntrophins and dystrophin. *Circ Res.* 2006;99:407–14.
25. McNair WP, Ku L, Taylor MR, et al. *SCN5A* mutation associated with dilated cardiomyopathy, conduction disorder, and arrhythmia. *Circulation.* 2004;110:2163–7.
26. Olson TM, Michels VV, Ballew JD, et al. Sodium channel mutations and susceptibility to heart failure and atrial fibrillation. *JAMA.* 2005;293:447–54.
27. Dellefave LM, Pytel P, Mewborn S, et al. Sarcomere mutations in cardiomyopathy with left ventricular hypertrabeculation. *Circ Cardiovasc Genet.* 2009;2:442–9.
28. Brauch KM, Karst ML, Herron KJ, et al. Mutations in ribonucleic acid binding protein gene cause familial dilated cardiomyopathy. *J Am Coll Cardiol.* 2009;54:930–41.
29. Guo W, Schafer S, Greaser ML, et al. RBM20, a gene for hereditary cardiomyopathy, regulates titin splicing. *Nat Med.* 2012;18:766–73.
30. Selcen D, Ohno K, Engel AG. Myofibrillar myopathy: clinical, morphological and genetic studies in 63 patients. *Brain.* 2004;127:439–51.
31. Hershberger RE, Hedges DJ, Morales A. Dilated cardiomyopathy: the complexity of a diverse genetic architecture. *Nat Rev Cardiol.* 2013;10:531–47.
32. Beqqali A, Bollen IA, Rasmussen TB, et al. A mutation in the glutamate-rich region of RNA-binding motif protein 20 causes dilated cardiomyopathy through missplicing of titin and impaired Frank-Starling mechanism. *Cardiovasc Res.* 2016;112:452–63.
33. Gruenbaum Y, Foisner R. Lamins: nuclear intermediate filament proteins with fundamental functions in nuclear mechanics and genome regulation. *Annu Rev Biochem.* 2015;84:131–64.
34. McNair WP, Sinagra G, Taylor MR, et al. *SCN5A* mutations associate with arrhythmic dilated cardiomyopathy and commonly localize to the voltage-sensing mechanism. *J Am Coll Cardiol.* 2011;57:2160–8.
35. Mann SA, Castro ML, Ohanian M, et al. R222Q *SCN5A* mutation is associated with reversible ventricular ectopy and dilated cardiomyopathy. *J Am Coll Cardiol.* 2012;60:1566–73.
36. Moreau A, Gosselin-Badaroudine P, Boutjdir M, Chahine M. Mutations in the voltage sensors of domain I and II of Nav1.5 that are associated with arrhythmias and dilated cardiomyopathy generate gating pore currents. *Front Pharmacol.* 2015;6:301.
37. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med.* 2015;17:405–24.
38. Rehm HL, Berg JS, Brooks LD, et al. ClinGen—the clinical genome resource. *N Engl J Med.* 2015;372:2235–42.