GENETIC MEDICINE

Integrated allelic, transcriptional, and phenomic dissection of the cardiac effects of titin truncations in health and disease

Angharad M. Roberts,[1,2]* James S. Ware,[2,3,4,5]* Daniel S. Herman,[4,5,6]* Sebastian Schafer,[7] John Baksi,[2] Alexander G. Bick,[4,5] Rachel J. Buchan,[2] Roddy Walsh,[2] Shibu John,[2] Samuel Wilkinson,[2] Francesco Mazzarotto,[2,3] Leanne E. Felkin,[2,3] Sungsam Gong,[2] Jacqueline A. L. MacArthur,[8] Fiona Cunningham,[8] Jason Flannick,[5,9] Stacey B. Gabriel,[5] David M. Altshuler,[5,10] Peter S. Macdonald,[11,12,13] Matthias Heinig,[7] Anne M. Keogh,[11,12,13] Christopher S. Hayward,[11,12,13] Nicholas R. Banner,[3,14] Dudley J. Pennell,[2,3] Declan P. O'Regan,[1] Tan Ru San,[15] Antonio de Marvao,[1] Timothy J. W. Dawes,[1] Ankur Gulati,[2] Emma J. Birks,[3,16] Magdi H. Yacoub,[3] Michael Radke,[17] Michael Gotthardt,[17,18] James G. Wilson,[19] Christopher J. O'Donnell,[20,21] Sanjay K. Prasad,[2] Paul J. R. Barton,[2,3] Diane Fatkin,[11,12,13] Norbert Hubner,[7,18,22] Jonathan G. Seidman,[4] Christine E. Seidman,[4,5,23,24†‡] Stuart A. Cook[1,3,15,25†‡]

The recent discovery of heterozygous human mutations that truncate full-length titin (TTN, an abundant structural, sensory, and signaling filament in muscle) as a common cause of end-stage dilated cardiomyopathy (DCM) promises new prospects for improving heart failure management. However, realization of this opportunity has been hindered by the burden of TTN-truncating variants (TTNtv) in the general population and uncertainty about their consequences in health or disease. To elucidate the effects of TTNtv, we coupled *TTN* gene sequencing with cardiac phenotyping in 5267 individuals across the spectrum of cardiac physiology and integrated these data with RNA and protein analyses of human heart tissues. We report diversity of *TTN* isoform expression in the heart, define the relative inclusion of *TTN* exons in different isoforms (using the *TTN* transcript annotations available at http://cardiodb.org/titin), and demonstrate that these data, coupled with the position of the TTNtv, provide a robust strategy to discriminate pathogenic from benign TTNtv. We show that TTNtv is the most common genetic cause of DCM in ambulant patients in the community, identify clinically important manifestations of TTNtv-positive DCM, and define the penetrance and outcomes of TTNtv in the general population. By integrating genetic, transcriptome, and protein analyses, we provide evidence for a length-dependent mechanism of disease. These data inform diagnostic criteria and management strategies for TTNtv-positive DCM patients and for TTNtv that are identified as incidental findings.

[1]Clinical Sciences Centre, Medical Research Council (MRC), Imperial College London, London W12 0NN, UK. [2]National Institute for Health Research (NIHR) Cardiovascular Biomedical Research Unit at Royal Brompton & Harefield National Health Service (NHS) Foundation Trust and Imperial College London, London SW3 6NP, UK. [3]National Heart & Lung Institute, Imperial College London, London SW3 6NP, UK. [4]Department of Genetics, Harvard Medical School, Boston, MA 02115, USA. [5]Broad Institute of Harvard and Massachusetts Institute of Technology, Cambridge, MA 02142, USA. [6]Department of Laboratory Medicine, University of Washington, Seattle, WA 98195, USA. [7]Cardiovascular and Metabolic Sciences, Max Delbrück Center for Molecular Medicine, 13125 Berlin, Germany. [8]European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton CB10 1SD, UK. [9]Department of Molecular Biology, Massachusetts General Hospital, Boston, MA 02114, USA. [10]Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA 02114, USA. [11]Cardiology Department, St Vincent's Hospital, Darlinghurst, New South Wales 2010, Australia. [12]Victor Chang Cardiac Research Institute, Darlinghurst, New South Wales 2010, Australia. [13]Faculty of Medicine, University of New South Wales, Kensington, New South Wales 2052, Australia. [14]Royal Brompton & Harefield NHS Foundation Trust, Harefield Hospital, Hill End Road, Harefield, Middlesex UB9 6JH, UK. [15]National Heart Centre Singapore, Singapore 169609, Singapore. [16]Department of Medicine, University of Louisville and Jewish Hospital, Louisville, KY 40202, USA. [17]Neuromuscular and Cardiovascular Cell Biology, Max Delbrück Center for Molecular Medicine, 13092 Berlin, Germany. [18]German Centre for Cardiovascular Research, 13347 Berlin, Germany. [19]Department of Physiology and Biophysics, University of Mississippi Medical Center, Jackson, MS 39216, USA. [20]National Heart, Lung, and Blood Institute's Framingham Heart Study, Framingham, MA 01702, USA. [21]Division of Intramural Research, National Heart, Lung, and Blood Institute, Bethesda, MD 20892, USA. [22]Charité-Universitätsmedizin, 10117 Berlin, Germany. [23]Cardiovascular Division, Brigham and Women's Hospital, Boston, MA 02115, USA. [24]Howard Hughes Medical Institute, Chevy Chase, MD 20815, USA. [25]Duke–National University of Singapore, Singapore 169857, Singapore.
*These authors contributed equally to this work.
†Corresponding author. E-mail: stuart.cook@nhcs.com.sg (S.A.C.); cseidman@genetics.med.harvard.edu (C.E.S.)
‡These authors are co-senior authors.

INTRODUCTION

Nonischemic dilated cardiomyopathy (DCM) has an estimated prevalence of 1:250, results in progressive cardiac failure, arrhythmia, and sudden death, and is the most frequent indication for cardiac transplantation (*1*, *2*). Despite a strong genetic basis for DCM (*2*) and the recent advent of affordable and comprehensive exome and genome sequencing techniques that permit screening of all DCM genes (*3*–*5*), the application of clinical molecular diagnostics in DCM management remains limited (*6*). Wider application is hindered by historically low mutational yield and a background prevalence of protein-altering variation of uncertain significance in the general population that make variant interpretation challenging (*7*–*9*).

TTN mutations can cause DCM (*10*, *11*), and heterozygous mutations that truncate full-length titin (TTNtv, titin-truncating variants) are the most common genetic cause of severe and familial DCM, accounting for about 25% of cases (*12*). TTNtv also occur in about 2% of individuals without overt cardiomyopathy (*12*–*14*), a value that exceeds the prevalence of nonischemic DCM by fivefold and poses significant challenges for the interpretation of TTNtv variants in the era of accessible genome sequencing. Critical parameters that distinguish pathogenic TTNtv and their mechanisms of disease remain unknown.

Titin is a highly modular protein with ~90% of its mass composed of repeating immunoglobulin (Ig) and fibronectin III (FN-III) modules

that are interspersed with nonrepetitive sequences with phosphorylation sites, PEVK motifs, and a terminal kinase (*15*). Two titin filaments with opposite polarity span each sarcomere, the contractile unit in striated muscle cells. The amino terminus of titin is embedded in the sarcomere Z-disc and participates in myofibril assembly, stabilization, and maintenance (*16*). The elastic I-band behaves as a bidirectional spring, restoring sarcomeres to their resting length after systole and limiting their stretch in early diastole (*17*). The inextensible A-band binds myosin and myosin-binding protein and is thought to be critical for biomechanical sensing and signaling. The M-band contains a kinase (*18*) that may participate in strain-sensitive signaling and affect gene expression and cardiac remodeling in DCM (*19*, *20*).

The *TTN* gene encodes 364 exons that undergo extensive alternative splicing to produce many isoforms ranging in size from 5604 to 34,350 amino acids. In the adult myocardium, two major full-length titin isoforms, N2BA and N2B, are robustly expressed along with low abundance, short novex isoforms (Fig. 1). N2BA and N2B isoforms span the sarcomere Z-disc to M-band but differ primarily in the I-band. The longer N2BA isoform contains both the N2A and N2B segments, whereas the N2B isoform lacks the unique N2A segment and contains fewer Ig domains and a smaller PEVK segment. The force required to stretch a titin molecule relates to its fractional extension (*21*), a parameter that shows nonlinear dependence on the I-band composition. For a given sarcomere length, the N2B isoform has a greater fractional extension and thus is stiffer than the longer N2BA isoform (*20*).

To explore further the spectrum of *TTN* genetic variation and transcript usage across the range of cardiac physiology, we studied five discovery cohorts, comprising healthy volunteers with full cardiovascular evaluations (n = 308), community-based cohorts with longitudinal clinical data [3603 participants in the Framingham (*22*) and Jackson (*23*) Heart Studies (FHS and JHS, respectively)], prospectively enrolled unselected ambulatory DCM patients (n = 374) (fig. S1), and end-stage DCM patients with left ventricular (LV) assist devices and/or considered for transplantation (n = 155). Integrated analyses of sequencing and transcriptional data yielded strategies for considerably narrowing the subsets of TTNtv that are likely to be pathogenic. We replicated these observations in two independent cohorts: patients with familial DCM (n = 163), and ethnicity-matched population controls from the Women's Health Initiative (*24*) (WHI; n = 667).

RESULTS

Burden of TTNtv in health and DCM

In the discovery cohorts, we identified 56 TTNtv affecting *TTN* isoforms that span the sarcomere in the 3911 controls: 9 in healthy volunteers (2.9%), 16 in FHS (1.0%), and 31 in JHS participants (1.6%). Eighty-three DCM patients carried TTNtv [versus controls, odds ratio (OR) = 13 (9 to 18), $P = 2.8 \times 10^{-43}$; Table 1]: 49 unselected DCM patients (13%) and 34 end-stage DCM patients (22%). Comparing variants found in healthy individuals and DCM patients, we found that nonsense, frameshift, and canonical splice site TTNtv were substantially enriched in DCM patients [OR = 17 (11 to 25), $P = 1.9 \times 10^{-45}$]. Additional variants predicted to alter noncanonical splice signals were also enriched in DCM [OR = 4.2 (1.8 to 9.7), $P = 0.0017$] but not as strongly (comparison: $P = 0.0068$), and these often occurred in combination with another TTNtv (tables S1 to S5).

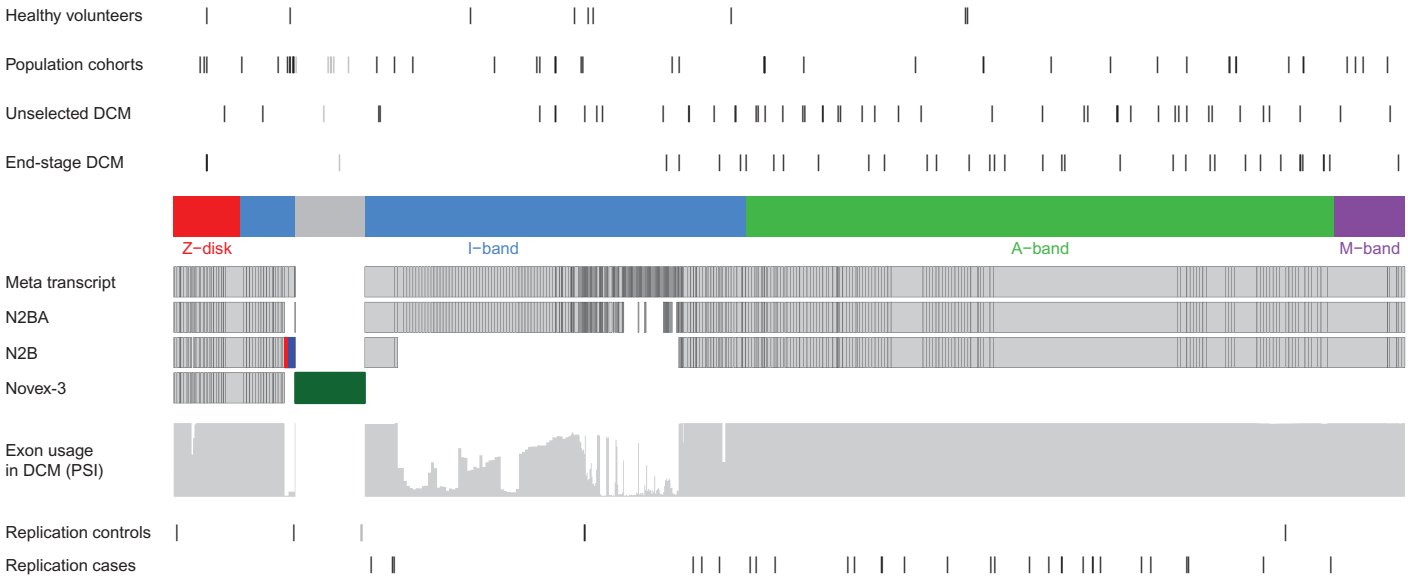

Fig. 1. Distribution of TTNtv in healthy individuals and DCM patients, and *TTN* exon usage in the heart. A schematic of the *TTN* meta-transcript, with sarcomere regions demarcated. The meta-transcript (LRG_391_t1/ENST00000589042) is a manually curated, inferred complete transcript, incorporating all exons of all known *TTN* isoforms (including fetal and non-cardiac isoforms) with the exception of the large alternative terminal exon 48 (dark green) that is unique to the novex-3 transcript (LRG_391_t2/ENST00000360870). Exon usage for the two principal adult cardiac isoforms, N2BA and N2B (ENST00000591111 and ENST00000460472), is shown, although exon usage in vivo is variable (see below). Novex-1 and novex-2 are rare cardiac isoforms that differ from N2B by the inclusion of a single unique exon each (red and blue, respectively, within the N2B track). Exon usage in human LV is depicted as the proportion spliced-in (PSI) (range, 0 to 1; gray bars): the proportion of transcripts that include a given exon. TTNtv are located more distally in cases compared with controls, with A-band and distal I-band enrichment in end-stage (n = 155) and unselected DCM patients (n = 374), and corresponding depletion in the population (n = 3603) and healthy volunteer (n = 308) cohorts.

Distribution of TTNtv in health and DCM

In our previous study of severe and familial DCM, we identified TTNtv that were located predominantly in the A-band (*12*). It is not known whether the A-band is more susceptible to truncating variation, whether TTNtv outside the A-band are excluded by alternate splicing, or whether the pathogenesis of TTN DCM is determined either by critical functional elements located in the A-band or by the length of a truncated TTN protein product. To explore these concepts, we examined the distribution of TTNtv across the spectrum of health and disease.

We observed that TTNtv were non-uniformly distributed within and between study groups (Fig. 1). TTNtv were more commonly located in the A-band in DCM than in controls [61 of 87 case variants located in A-band, versus 21 of 56 in controls, OR = 3.9 (1.8 to 8.3), $P = 1.4 \times 10^{-4}$], as a result of both an enrichment of A-band variants in DCM patients [compared with a uniform distribution; OR = 2.3 (1.4

Table 1. Number of TTNtv in DCM patients and controls. Numbers of subjects with a TTNtv are shown for each group. TTNtv are classified by type, the affected transcript, and expression level of the variant-encoding exon. Comparisons between groups were assessed by Fisher's exact test. Cohort ethnicity: Caucasian: healthy volunteer, 75%; FHS, 100%; JHS, 0%; unselected DCM, 88%; end-stage, DCM 85%; African American: healthy volunteer, 2%; FHS, 0%; JHS, 100%; unselected DCM, 4%; end-stage DCM, 6%.

	Discovery cohorts							Replication cohorts		
	Healthy volunteers (n = 308)	FHS (n = 1623)	JHS (n = 1980)	Unselected DCM (n = 371)*	End-stage DCM (n = 155)	P value DCM versus controls: Unselected DCM	P value DCM versus controls: All DCM	WHI (n = 667)	Familial DCM (n = 163)	P value DCM versus controls
Transcript affected by truncation										
N2BA and N2B	4	11	20	42	34	1.7×10^{-25}	5.5×10^{-46}	2	31	3.8×10^{-21}
N2BA only	4	2	8	7	0	0.0014	0.0083	2	0	1
Neither N2BA or N2B (novex-3 terminal exon only†)	1 (0)	8 (5)	4 (1)	1 (1)	1 (1)	1	0.7	5 (4)	0 (0)	1
Totals	*9*	*21*	*32*	*50*	*35*			*9*	*31*	
Sarcomere domain										
A-band (18,235 amino acids)	2	7	12	32	29	0.011	0.00014	1	25	0.015
Non–A-band (17,756 amino acids)	7	9	19	18	8			4	6	
Totals	*9*	*16*	*31*	*49‡*	*34‡*			*5*	*31*	
Usage§ of exon containing truncation										
Low (PSI <0.15)	1	4	4	2	0	0.25	0.38	1	0	1
Intermediate (PSI 0.15 to 0.9)	4	1	7	5	0	0.012	0.042	2	0	1
High (PSI >0.9)	4	11	20	42	34	1.7×10^{-25}	5.5×10^{-46}	2	31	3.8×10^{-21}
Totals	*9*	*16*	*31*	*49‡*	*34‡*			*5*	*31*	
Variant type										
Frameshift variant	2	3	7	22	15	4.9×10^{-16}	4.0×10^{-25}	1	15	1.9×10^{-10}
Stop gained	3	7	6	15	12	7.3×10^{-09}	2.3×10^{-15}	2	16	2.8×10^{-10}
Canonical splice sites¶	2	3	7	9	5	2.9×10^{-05}	2.3×10^{-07}	1	0	1
Splice variant predictions‖	2	3	11	4	5	0.089	0.0015	1	0	1
Totals	*9*	*16*	*31*	*49‡*	*34‡*	*1.4×10^{-25}*	*2.8×10^{-43}*	*5*	*31*	*1.6×10^{-18}*

*Three of 374 subjects were excluded from these analyses due to relatedness to other subjects. †Variants that only affect the alternative terminal exon of novex-3 are excluded elsewhere. ‡Total number of individuals with TTNtv; four individuals with DCM (one unselected, three end-stage) carry a second TTNtv, which is a splice variant prediction in three cases. §Exon usage levels are displayed categorically on the basis of PSI. ¶Canonical splice sites refer to the two intronic base pairs at the 5′ and 3′ splice junctions. ‖Variants close to canonical splice sites that are predicted in silico to alter splicing.

to 3.6), $P = 3.4 \times 10^{-4}$] and an opposing trend, towards A-band sparing, in controls [35 of 56 variants outside A-band, OR = 0.58 (0.34 to 1.0), $P = 0.06$]. A-band enrichment was most pronounced in end-stage DCM patients [OR = 3.5 (1.6 to 8.5), $P = 7.8 \times 10^{-4}$] with a concordant trend in less severe, unselected DCM [OR = 1.7 (0.97 to 3.1), $P = 0.07$] (Fig. 1 and Table 1). These distributions were not explained by DNA sequence susceptibility to truncating variation in the *TTN* gene or by differences in variant detection between cohorts (figs. S2 and S3).

The effect of TTNtv on different TTN isoforms

Distal exons, including those that encode the A-band of TTN, are constitutively expressed, whereas many proximal exons, particularly I-band exons, are variably spliced in different isoforms (tables S6 and S7). Because recent studies suggest that variants affecting only a subset of gene transcripts are less likely to cause loss of function than variants affecting all isoforms (*25*), we compared TTNtv among different isoforms. TTNtv that altered both N2BA and N2B were strongly enriched in DCM patients when compared to controls [OR = 19 (12 to 29), $P = 5.5 \times 10^{-46}$] and associated more strongly with DCM than TTNtv that affected only the N2BA isoform [OR = 3.8 (1.4 to 9.2), $P = 0.008$; Table 1].

By contrast, TTNtv found in controls were enriched in exons not incorporated into N2BA and N2B transcripts (such as exons in novex and fetal isoforms). Thirteen variants were found in N2BA/N2B-excluded exons [7406 base pairs (bp)], and 49 variants were found in N2BA/N2B-included exons [103,052 bp] [OR = 3.7 (2.0 to 6.7), $P = 2.0 \times 10^{-4}$]. The prevalence of truncations in the terminal exon unique to the novex-3 isoform was not significantly different between cohorts [0.37% DCM versus 0.15% controls, OR = 2.5 (0.36 to 13), $P = 0.24$], and the nominal about twofold excess in DCM was not robust to analyses that included only European subjects (0.26%, OR = 1.4). In addition, the novex-3 isoform only spans the sarcomere Z-disc and proximal I-band (*26*), and LV expression levels of novex-3 in 105 samples from the Genotype-Tissue Expression (GTEx) project (*27*) and in DCM patients (see below) were about 7.3 and 9.4% of N2BA and N2B isoform levels, respectively. Given these observations and the lack of evidence for pathogenicity of novex-3 truncations, mutations specific to this isoform were excluded from subsequent analyses.

Alternative splicing of TTN in the human heart

We generated RNA sequencing data from human LV samples (end-stage DCM hearts, $n = 84$), and determined the median usage of each *TTN* exon (table S7), denoted as the proportion of transcripts that incorporate each exon or PSI (see Materials and Methods). Identical exons were alternatively spliced in LV tissues from DCM patients and GTEx donors (global PSI: $R = 0.98$) (table S7 and fig. S4). There were important differences between observed exon usage and conventional transcript definitions (see table S7): 39 of 122 exons annotated as incorporated into the N2BA isoform were expressed in a small minority of transcripts (PSI, <0.15). Three exons annotated as constitutively expressed appeared to have variable usage (PSI, 0.15 to 0.9), as did two exons absent from the conventional N2BA/N2B descriptions. A summary of these transcript annotations, including PSI values, is available at http://cardiodb.org/titin.

The *TTN* gene structure is organized to accommodate extensive splicing events. Eighty-five percent of all *TTN* exons are symmetric, and consequently, their exclusion would not alter the translation frame, whereas exome-wide, only 68% of exons are symmetric ($P = 1.2 \times 10^{-11}$). Exon symmetry was correlated with PSI. Only three exons (103, 104, and 106) among 175 exons with PSI <0.99 are asymmetric, whereas 49 of 185 (27%) exons with PSI >0.99 are asymmetric (table S7; $P = 7 \times 10^{-13}$). Within the I-band, the domain with the overall lowest PSI, 93% of alternately spliced exons are symmetric. Moreover, the cassette of I-band exons 103 to 106 is symmetric because these exons are always spliced together. Hence, most I-band exons can be excluded without resulting in a frameshift, including exons that might include TTNtv.

We used the mean PSI scores from the end-stage DCM patients to annotate each *TTN* exon's usage. The usage of the exons containing TTNtv differed between cohorts. Treating all cohorts as an ordered variable with four levels (healthy volunteers, general population, ambulatory unselected DCM, and end-stage DCM), we noted a strong relationship between cohort and mutant exon usage (Kruskal-Wallis χ^2, $P = 4.9 \times 10^{-3}$), with the TTNtv-containing exons in controls having lower usage than the TTNtv-containing exons in DCM patients ($P = 2.5 \times 10^{-4}$) (Fig. 2A). On the basis of this observation, we suggest that many TTNtv in controls may be tolerated because they fall in exons that are spliced out of the majority of expressed transcripts.

Protein coordinates of truncating variants

TTNtv in DCM cases could also be described as occurring more distally in the *TTN* gene than TTNtv in controls. Treating all cohorts as an

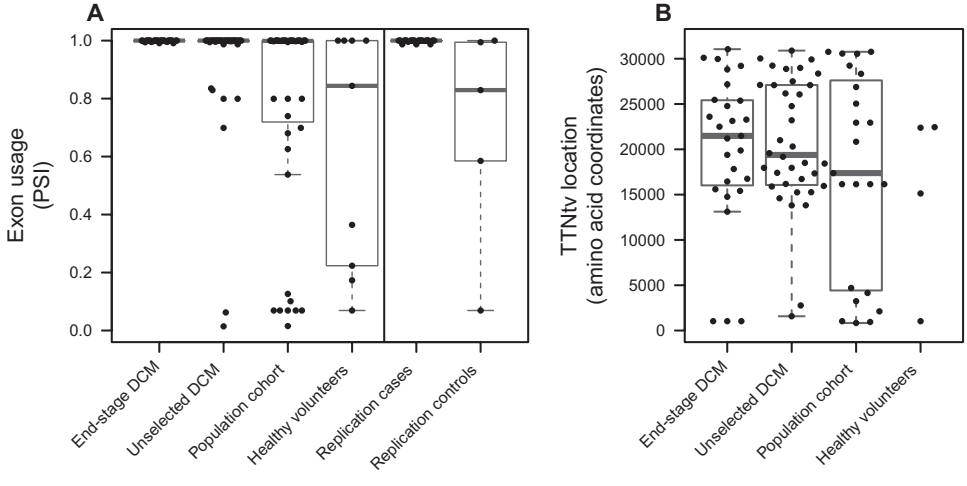

Fig. 2. Factors that discriminate TTNtv in health and disease. (**A**) Usage of *TTN* exons containing TTNtv in all cohorts. Exon usage is represented as PSI, which is an estimate of the proportion of transcripts that incorporate each exon. Each plotted data point represents the estimated PSI of an exon identified to have a TTNtv, grouped by cohort. There was a strong relationship between the PSI of exons containing TTNtv and disease status ($P = 4.9 \times 10^{-3}$, Kruskal-Wallis), with TTNtv in DCM cases found in more highly used exons ($P = 4.7 \times 10^{-4}$, Mann-Whitney). A similar difference was observed between the replication cohorts ($P = 7.5 \times 10^{-4}$). (**B**) Relationships between TTNtv location, PSI, and disease status. The positions of TTNtv (amino acid coordinates, reference transcript LRG_391_t1) are shown for constitutively expressed exons only (PSI = 1).

ordered variable (as above), we also observed a significant relationship between cohort phenotypes and TTNtv position (Kruskal-Wallis χ^2, $P = 3.1 \times 10^{-3}$).

Diagnostic interpretation of TTNtv

Because diagnostic sequencing of *TTN* will be most useful if causality can be confidently ascribed to individual variants, we estimated the probability of pathogenicity of TTNtv on the basis of their relative frequency between cohorts. Applying this framework to our discovery cohort, we estimated that TTNtv produced by nonsense, frameshift, or canonical splice site mutations that affect highly expressed exons (PSI, >0.9) had a 93% probability of pathogenicity [likelihood ratio (LR) = 14] when identified in an unselected patient with DCM, and an even higher probability of pathogenicity in end-stage disease (≥96%, LR = 24). When segregation data are available, we expect that probabilities will be even higher. These are conservative estimates, because we assumed an all or nothing model in which all TTNtv in controls were benign, giving an upper limit of the background noise.

About 50% of the TTNtv identified in healthy volunteers and community-based cohorts occurred in low PSI exons, including novex-specific exons. Analyses of publicly available genomic data sets showed similar results. TTNtv occur in 1.1% of alleles in the 1000 Genomes Project (*28*, *29*) and in 2.6% of alleles in the National Heart, Lung, and Blood Institute (NHLBI) GO Exome Sequencing Project (ESP; http://evs.gs.washington.edu/EVS/). Seven of 12 (58%) TTNtv in the 1000 Genomes Project and 83 of 168 (49%) TTNtv in ESP are in novex or other low PSI exons (tables S8 and S9).

To further explore the health effects of TTNtv in community-based cohorts, we examined longitudinal follow-up data. Most FHS and all JHS participants with TTNtv had normal cardiac parameters. However, among the small numbers of TTNtv-positive FHS subjects, we observed a higher lifetime incidence of DCM morphology [dilated LV with impaired ejection fraction (EF)] in the absence of coronary artery disease (CAD) [TTNtv-positive, 2 of 16 subjects; TTNtv-negative, 12 of 1574 subjects; relative risk (RR) = 16, $P = 0.008$; tables S10 and S11]. Although the two TTNtv (c.9727C>T and c.1245+3A>G) associated with DCM morphology are outside the A-band, both are in highly expressed exons (PSI = 1). The association between DCM morphology and TTNtv in highly expressed exons was even more marked (TTNtv-positive, 2 of 12 subjects; RR = 22, $P = 0.005$).

TTNtv-positive FHS subjects had no evidence of either early heart failure or early cardiovascular death (table S12 and fig. S5). Although some of these TTNtv are unlikely to be pathogenic (for example, TTNtv found in rare novex exons), others may cause DCM with reduced penetrance as a result of characteristics of the variant itself and/or the effects of additional protective and exacerbating genetic or environmental modifiers for DCM (*30*, *31*), factors that may also account for the mismatch between population prevalence of mutations in hypertrophic cardiomyopathy genes and overt disease (*8*).

Validation studies

The genetic and transcriptional analyses of the five discovery cohorts predicted that the pathogenicity of TTNtv was influenced by isoform, exon usage, and variant position. To validate this hypothesis, we identified TTNtv in an independent cohort of familial DCM patients ($n = 163$) and 667 healthy participants in the WHI (*24*), a cohort that excluded individuals with chronic disease. In comparison to controls, TTNtv affecting both the N2BA and N2B isoforms were enriched in

Table 2. Clinical characteristics of DCM patients with and without TTNtv. Unselected DCM cohort. Values are means ± SD. Measurements are indexed to body surface area where indicated. LV, left ventricle; RV, right ventricle; EDVi/ESVi, indexed end-diastolic/systolic volume; SVi, indexed stroke volume; EF, ejection fraction; LVMi, indexed LV mass; WTi, indexed wall thickness; VT, ventricular tachycardia; NYHA, New York Heart Association functional class. Groups were compared using Wilcoxon-Mann-Whitney test for continuous variables, and Fisher's exact test for categorical. *P* values not corrected for multiple testing, as variables were not independent.

CMR and clinical data		TTNtv-negative ($n = 277$)	TTNtv-positive ($n = 42$)	*P*
LV	EDVi	136 ± 38	137 ± 34.1	0.677
	ESVi	87.5 ± 39.1	93.7 ± 36.4	0.205
	SVi	48.2 ± 12.9	43.4 ± 14.1	0.041
	EF	37.5 ± 12.2	33.3 ± 13.3	0.047
RV	EDVi	89.4 ± 24.7	89.6 ± 26.1	0.972
	ESVi	45.2 ± 22.3	50.4 ± 25.9	0.234
	SVi	44.4 ± 12.6	39.2 ± 15.8	0.031
	EF	51.6 ± 14.1	45.2 ± 16	0.036
LVMi		95.4 ± 27.6	87.1 ± 18.3	0.106
Lateral WTi		3.13 ± 0.73	2.77 ± 0.713	0.003
Midwall fibrosis		94/270 (35%)	13/41 (32%)	0.869
Age at diagnosis (years)		53.4 ± 13.4	49.3 ± 13.7	0.115
NYHA status 1/2/3/4		116/104/36/1	19/16/4/1	0.692
Sustained VT		20/97 (21%)	9/14 (64%)	0.001
Conduction disease		82/227 (36%)	8/36 (22%)	0.130
Family history of DCM		24/218 (11%)	9/37 (24%)	0.034

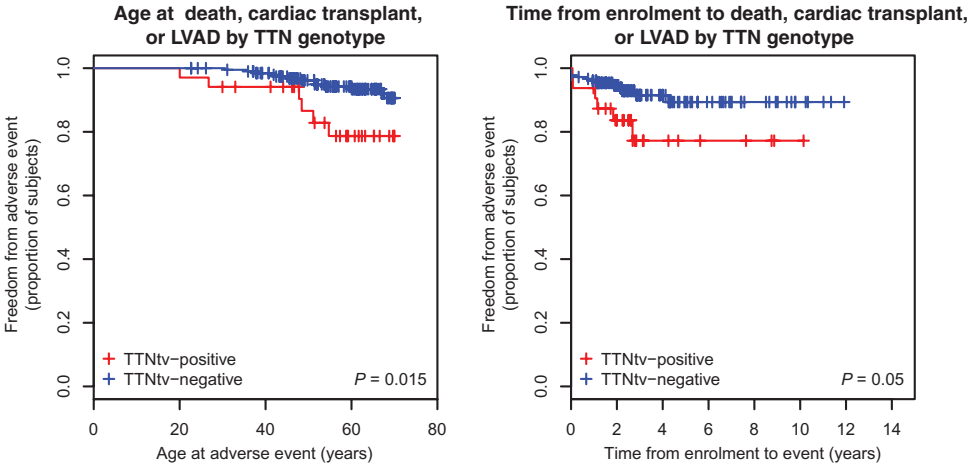

Fig. 3. TTNtv and survival in DCM. Outcomes in unselected DCM patients with (red) and without (blue) TTNtv. (**Left**) Age censored at adverse event [death, cardiac transplant, or left-ventricular assist device (LVAD)] or at age 70 years. (**Right**) Adverse events after enrollment, to control for ascertainment (interval censored from time of enrollment to age 70 years or adverse event). Event-free survival is reduced in TTNtv-positive DCM ($P = 0.015$) as a result of faster disease progression. A trend to younger presentation (Table 2) and worse outcomes after enrollment ($P = 0.05$) combine to give reduced survival overall.

the DCM replication cohort [OR = 78 (20 to 460), $P = 3.8 \times 10^{-21}$], and TTNtv in this cohort were located in more highly expressed exons ($P = 7.5 \times 10^{-4}$; Table 1, Figs. 1 and 2, and table S13). The predicted probability of pathogenicity of nonsense, frameshift, or canonical splice TTNtv in highly expressed exons in the DCM replication cohort was ≥98% (LR = 41).

Clinical stratification of DCM by TTN genotype

To better ascertain clinical phenotypes associated with TTNtv-positive DCM, we capitalized on quantitative cardiac magnetic resonance (CMR) imaging (*32*, *33*) in DCM patients. Among TTNtv-positive DCM patients, we observed more severely impaired LV function, lower stroke volumes, and thinner LV walls (Table 2) than in TTNtv-negative DCM patients. Multivariate regression confirmed that *TTN* genotype predicted phenotype severity after adjusting for important covariates (tables S14 and S15). Midwall fibrosis, an important prognostic factor in DCM (*34*, *35*), was similar in patients with and without TTNtv, but sustained ventricular tachycardia was more common in TTNtv-positive patients (OR = 6.7, $P = 0.001$) and robust to adjustment for LV EF. Consistent with these adverse intermediate phenotype associations, we observed a difference in the composite endpoint of LV assist device implantation, listing for cardiac transplantation, and all-cause mortality. TTNtv-positive DCM patients reached this endpoint at earlier ages ($P = 0.015$; Fig. 3) and sooner after prospective enrolment ($P = 0.05$; Fig. 3).

Length-dependent association between TTNtv position and DCM

Motivated by the association between TTNtv location and disease status (Fig. 1), which persists after controlling for PSI (Fig. 2B), we considered TTNtv in DCM patients as an allelic series to dissect positional effects and disease mechanism.

The distance from the N terminus of the TTN protein to the TTNtv correlated with CMR indices (Fig. 4 and fig. S6). Multivariate linear regression models showed that TTNtv location was significantly correlated with principle indices of heart function: EF and stroke volume ($P < 0.006$; tables S14 and S15). This positional effect on cardiac parameters was large, such that a C-terminal TTNtv would be associated with substantially reduced EF as compared with an N-terminal TTNtv [absolute reductions: LV −18 ± 7%, $P = 0.006$; right ventricle (RV) −21 ± 9%, $P = 7.3 \times 10^{-6}$] and SVi (absolute reductions: LV −22 ± 8 ml/m^2, $P = 0.0017$; RV −23 ± 8 ml/m^2, $P = 0.0013$). Among subjects with TTNtv, the variant position explained 19 to 23% of the observed variation (R^2) in phenotypic indices. Regression modeling of CMR data in FHS participants also suggested that the distance of the TTNtv from the N terminus correlated with cardiac morphology; there was a consistent direction of effect across a range of phenotypic indices (figs. S7 and S8 and tables S16 and S17).

We suggest that the phenotypic associations with exon usage and location of the truncation within the protein are potentially of clinical importance for diagnostic variant interpretation. In addition, we deduced that a linear positional effect of TTNtv implied that mutant proteins produced dominant negative effects.

Molecular studies and mechanistic implications

On the basis of the observation that TTNtv exhibited length-dependent effects, we studied allele-specific *TTN* transcript expression and protein levels in human LV tissue to further explore whether TTNtv caused DCM through dominant negative effects or through haploinsufficiency. RNA sequencing showed comparable total *TTN* transcript levels in patients with or without TTNtv (Fig. 5A). Moreover, the relative expression of TTNtv and of other single-nucleotide polymorphisms (SNPs) distributed throughout *TTN* transcripts showed robust expression of both alleles (Fig. 5B, table S18, and fig. S9). We also observed no discernible difference in the abundance of N2BA and N2B protein isoforms in DCM patients with or without TTNtv (Fig. 5C). Combined with the genetic data presented above, our analyses of *TTN* RNA and protein expression in LV tissues suggest that TTNtv may cause DCM by a dominant negative effect.

DISCUSSION

The integrated analyses of *TTN* sequence, protein and transcriptional data, and quantitative phenotypic assessment of more than 5200 healthy and DCM subjects define the spectrum of cardiac physiology associated with TTNtv. We demonstrate that TTNtv occur in ~2% of the general population, in 13% of ambulatory unselected DCM patients, and in 20% of end-stage DCM patients. We suggest that the clinical significance of TTNtv is largely determined by exon usage and variant location (the distance of the TTNtv from the protein N terminus). Incorporation of these data improved discrimination between pathogenic and benign variants in two independent study cohorts. That TTNtv exhibited length-dependent consequences and were highly expressed in human LV tissue suggest that these mutations may cause DCM through a dominant negative mechanism.

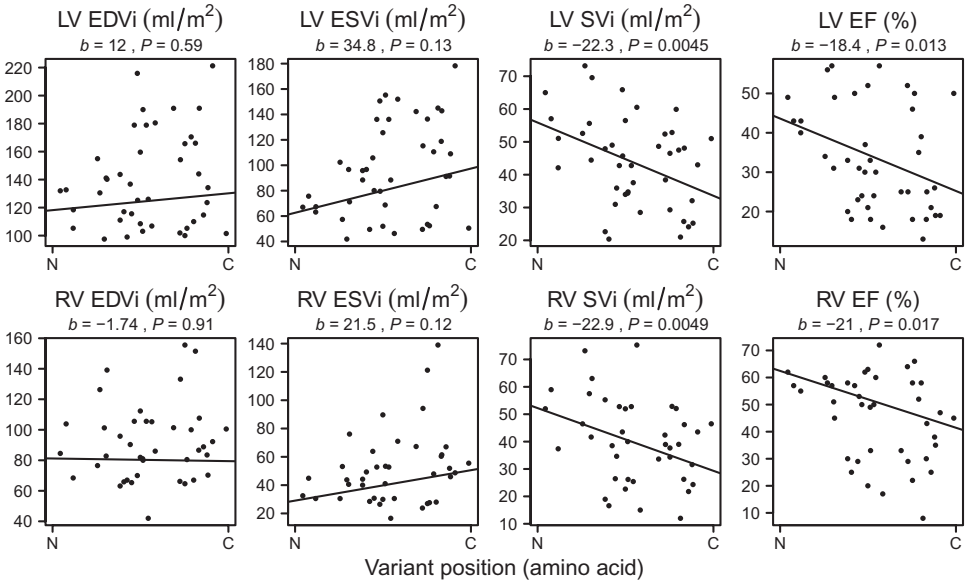

Fig. 4. Allelic dissection of the impact of TTNtv position on cardiac morphology and function. The relationships between TTNtv location and cardiac morphology and function assessed by CMR imaging in an allelic series of DCM cases. Genotype-phenotype relationships are shown for 43 TTNtv in unselected DCM patients. The TTNtv location (*x* axis) is plotted from the amino (N) to the carboxyl (C) end of the protein. Distal (C-terminal) TTNtv were associated with worse cardiac contractile performance, specifically diminished indexed stroke volume (SVi) and EF of both LV and RV as compared to proximal truncations. A regression line is shown for each variable (tables S14 and S15). EDVi, indexed end-diastolic stroke volume (ml/m^2); ESVi, indexed end-systolic volume (ml/m^2); SVi, indexed stroke volume (ml/m^2); EF, ejection fraction (%).

Cardiomyopathy genes feature prominently in the American College of Medical Genetics and Genomics' list of genes in which mutations should be reported to the patient regardless of the primary indication for sequencing that patient's genome (*36*). Accurate interpretation of such clinically actionable incidental findings in cardiomyopathy genes is both difficult and medically important because of the considerable population prevalence of protein-altering variants in cardiomyopathy genes (*8*, *36*), a combined population prevalence of cardiomyopathies of about 0.7%, and the associated important medical consequences including heart failure and sudden death (*37*–*39*).

The true frequency of TTNtv across the general population has been unclear, and the lack of penetrance of these variants is an issue of debate (*13*, *20*, *40*, *41*). From analyses of more than 4500 control subjects, we identified TTNtv in 1.6% individuals of African descent and 1.5% individuals of European descent. Using transcript and mean *TTN* exon expression in human heart tissue, we provide insight into why some of these TTNtv are phenotypically silent. Truncations in the control subjects were more likely to affect minor *TTN* isoforms as compared with DCM cases, including novex-3, a low-abundance isoform that does not span the cardiac sarcomere. Truncations that occurred specifically in novex-3 were not enriched in DCM patients. We also observed that, although canonical splice variants were enriched in *TTN* from DCM patients, other variants predicted in silico to alter splicing showed more modest enrichment (0.4% versus 1.7%, $P = 0.0015$; Table 1). This observation may reflect limitations in current prediction algorithms; interpretation of noncanonical splice variants should be cautious unless informed by RNA evidence or robust segregation data. In addition, the impact of TTNtv in exons with intermediate expression levels (PSI, 0.15 to 0.9) may be ameliorated because (i) symmetrical exons can be excluded without deleterious consequences, (ii) an isoform switch from N2BA to N2B can occur, or (iii) shorter mutant proteins may be less deleterious. Overall, TTNtv in low- to intermediate-expression exons (PSI, <0.9), novex *TTN* isoforms, and at noncanonical splice sites accounted for 50% of all TTNtv identified in controls cohorts. These variant types were associated with normal cardiac morphology and function and were not associated with DCM. We suggest that when TTNtv with these characteristics are identified in low-risk individuals, the clinical interpretation should not convey to the patient that he or she is at high risk for DCM.

A small number of FHS participants ($n = 12$) had TTNtv in highly expressed exons including two individuals with DCM morphology on cardiac imaging (RR = 22). There was no increased risk for DCM in TTNtv-positive JHS participants, possibly because of differences in phenotype ascertainment or ethnicity-specific genetics. Whether the differential risk associated with TTNtv between the DCM and population cohorts reflects differences in phenotypic assessment, the very small numbers of cases, an aggregation of additional genetic factors in phenotypically ascertained DCM families, or other factors is unknown.

Patient stratification is a cornerstone of precision medicine (*42*). We propose that DCM due to TTNtv represents a specific patient subgroup that may benefit from more tailored clinical management. In DCM, sustained ventricular tachycardia, LV wall thickness (*43*), and LV EF predict outcome (*44*). TTNtv-positive patients reported here had poorer cardiac indices and earlier onset of heart failure or death. Despite small differences in the functional indices between groups, we showed that, as compared to TTNtv-negative patients, TTNtv-positive DCM patients had substantially increased risk of sustained ventricular tachycardia (OR = 6.8, $P = 0.001$), perhaps related to increased wall stress (*45*). If these findings are replicated in prospective cohorts, TTNtv-positive DCM patients may benefit from a lower threshold for device therapy, as is practiced with *LMNA* DCM (*46*). Preliminary observations showed that five of six TTNtv-positive DCM patients who received mechanical unloading therapy support during this study had sustained recovery of cardiac function, raising the possibility that *TTN* DCM may prove amenable to targeted device therapy.

Mutations that lead to premature termination of encoded proteins often cause haploinsufficiency. By contrast, our analyses of allele-specific transcript expression and protein in human LV tissue indicate that *TTN* is highly and biallelically expressed: TTNtv containing transcripts were not subjected to substantial nonsense-mediated decay, and levels of the major titin protein isoforms were not diminished. Deletions within 2q31-q32 encompassing the entire *TTN* locus do not cause overt cardiac muscle disease (*47*), and patients with recessive TTN mutations exhibit truncated TTN in the sarcomeres of skeletal muscles by immunohistochemistry (*48*), supporting this postulate. Rather, the correlations between TTNtv position and cardiac function reported here suggest

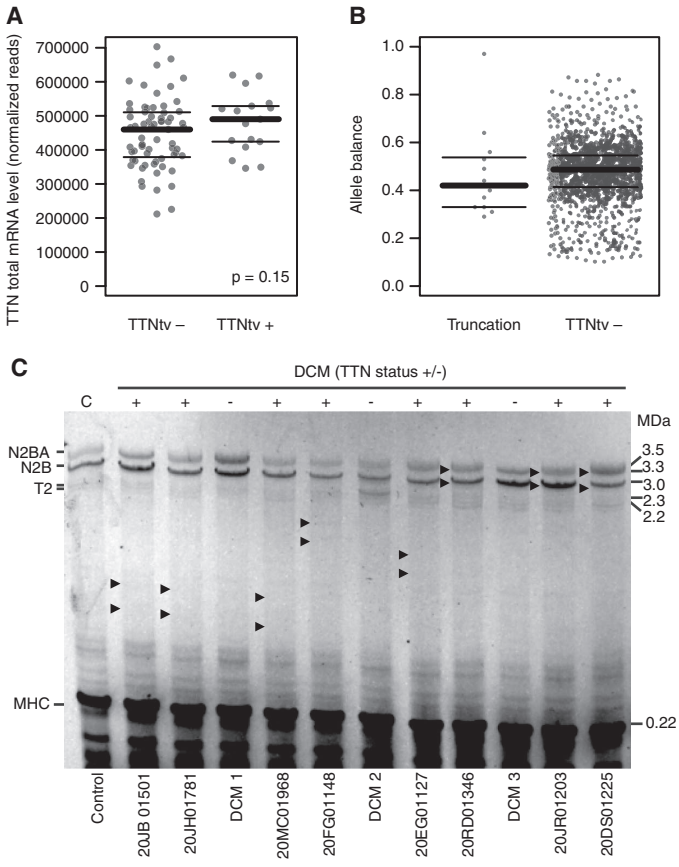

Fig. 5. *TTN* mRNA and protein expression in LV tissues from DCM patients with and without TTNtv. (**A**) *TTN* mRNA in TTNtv-positive (n = 18) and TTNtv-negative (n = 66) patients (quantile-normalized read counts). (**B**) Allelic balance of TTNtv compared to nontruncating *TTN* SNPs, a surrogate for the proportion of transcripts with variant alleles (TTNtv or SNPs) among DCM patients with and without TTNtvs. The comparable allelic expression of TTNtv and SNPs does not support substantial nonsense-mediated decay (see also fig. S9). Bars indicate median and quartiles. (**C**) Protein electrophoresis from a healthy LV (lane 1) and LV from DCM patients (lanes 2 to 12: +, TTNtv-positive; –, TTNtv-negative). Sample IDs are shown for subjects with TTNtv: variant details are shown in table S4. Truncated protein was not seen in TTNtv-positive samples. Arrowheads, approximate expected sizes of the truncated N2B and N2BA isoforms; T2, a TTN degradation product; MDa, megadaltons. Semiquantitative analysis of TTN protein relative to myosin (MHC, myosin heavy chain) showed no reduction of TTN in TTNtv-positive samples.

a dominant negative effect, as occurs with some *MYBPC3* truncations that cause hypertrophic cardiomyopathy (*49*, *50*), in which increasing DCM severity is associated with longer mutant proteins. Further studies are needed to determine whether this length relationship is valid and if it results from an increased energetic cost associated with generation and turnover of longer mutant proteins, deleterious sequestration of nonsarcomeric intracellular factors that is proportional to protein length, or increased propensity for disruptive sarcomere protein interactions with longer mutant proteins.

We recognize several potential limitations despite these extensive analyses. There are many exons in which TTNtv were not observed, and we can only extrapolate our findings to these regions. About 5% of the gene comprises repetitive exons with poor alignment: although coverage did not differ much among cohorts, additional TTNtv in these regions may have been missed. Newer sequencing assays with longer read lengths can address this. The lower burden of TTNtv in our population cohorts yielded a small allelic series that, combined with the limitations of screening-quality phenotype ascertainment, limited the power of genotype-phenotype analyses in these cohorts. We had arrhythmia data for only a subset of the unselected DCM cohort, and ongoing follow-up with replication is needed to further these assessments. Finally, molecular studies were limited by the scarcity of human cardiac tissue for study.

In conclusion, our data illuminate important features that determine TTNtv pathogenicity and begin to dissect the molecular mechanisms by which these cause DCM. Nonsense, frameshift, and canonical splice site TTNtv, particularly those that truncate both principal isoforms of *TTN* and/or reside towards the C terminus, cause DCM with severely impaired LV function and life-threatening ventricular arrhythmias. In contrast, truncations that occur in novex-specific exons or other infrequently used *TTN* exons are less likely to be deleterious. An immediate clinical use of our findings is improved variant interpretation that enables cascade screening of relatives, and gene- and genotype-guided stratified management of DCM. Further elucidation of the myocyte processes potentially altered by large dominant negative mutant TTN proteins will be important to direct the development of therapies that prevent or attenuate the progression of TTNtv-related DCM.

MATERIALS AND METHODS

Study design

We set out to compare the burden of rare *TTN* variants across five cohorts (detailed below) and to explore genotype-phenotype relationships within cohorts using standard cardiac investigations and techniques. There were no interventions. No genotype information was available at recruitment, so patient inclusion was blinded to genotype. Phenotype assessment was blinded to genotype but not to disease status. Study design and analyses underwent rigorous internal statistical review. All studies were carried out with protocols that were reviewed and approved by institutional ethics committees and with informed consent from all participants. Tissue studies complied with UK Human Tissue Act guidelines.

Cohort descriptions and subject selection

DCM cohorts. Three hundred seventy-four unselected prospective patients of predominantly European ancestry who were referred to the CMR unit of the Royal Brompton & Harefield Hospitals NHS Foundation Trust (RBHT) from July 2001 to August 2012 and diagnosed with idiopathic DCM were studied. DCM was diagnosed by CMR findings of EF >2 SD below and end-diastolic volume >2 SD above the mean normalized for age and sex (*33*, *51*) by two independent level 3–accredited CMR cardiologists (see fig. S1). Patients with clinical symptoms or signs of active myocarditis or CMR evidence of infiltrative disease were excluded. CAD was assessed either by coronary angiography (249 patients) or by non-invasive testing and clinical profiles (for example, young relatives of an individual with idiopathic DCM; 101 patients). Twenty-four patients had bystander CAD considered insufficient to produce CMR features of DCM (<2 myocardial segments with <25% late gadolinium enhancement).

One hundred fifty-five randomly selected end-stage nonischemic DCM patients who were listed for cardiac transplantation and/or LV device implantation between 1993 and 2011 at RBHT and prospectively

enrolled in a tissue bank were studied. Seventy-one of these patients were previously reported (*12*). Frozen LV samples from 84 patients were used for tissue studies.

One hundred sixty-three DCM patients of European ancestry who were referred to the genetics research program at St Vincent's Hospital and Victor Chang Cardiovascular Institute were studied as a replication cohort. Patients had a positive family history (DCM or sudden cardiac death in ≥2 family members), and three had subclinical skeletal myopathy (elevated creatine kinase blood levels). DCM was diagnosed at a mean age of 42 years based on presenting clinical symptoms (exertional dyspnea, palpitations) and echocardiographic findings of LV dilation [LV end-diastolic dimension (LVEDD), >56 mm] with reduced systolic performance (EF, <50%). DCM severity ranged from mild to severe: 32 patients (20%) required cardiac transplantation, and 2 patients died from advanced heart failure before transplantation.

Healthy volunteers and population cohorts. Three hundred eight clinically screened adult volunteers (age range, 18 to 72 years; mean, 40.3 years) of predominantly European ancestry were prospectively recruited via advertisement at the MRC Clinical Sciences Centre, Imperial College London. Participants with previously documented cardiovascular disease, hypertension (HTN), diabetes, or hypercholesterolemia were excluded.

Community-based cohorts. Unrelated participants in the FHS offspring cohort (1623) and 1980 unrelated randomly selected participants in the JHS were studied. The FHS is a multigeneration, prospective, population-based study aimed at identifying the causes of cardiovascular disease (*22*). In 1948, residents of Framingham, Massachusetts, of predominantly European Ancestry were enrolled. Between 1971 and 1975, the study enrolled a second generation, the offspring cohort, comprising 5124 children of the original cohort and their spouses. The offspring cohort has since been examined every 3 to 8 years, with the last exam reported here, exam 8, occurring between 2005 and 2008. All FHS phenotypic data were retrieved from National Center for Biotechnology Information (NCBI) dbGaP (accession: phs000007.v18.p7). *TTN* sequence data for FHS participants are available from NCBI dbGaP (accession: phs000307.v3.p7). The JHS is an African American population-based, prospective study of cardiovascular disease (*23*). Between 2000 and 2003, the study enrolled 5301 African Americans aged 35 to 84 years and living in the Jackson, Mississippi, metropolitan area. All JHS phenotypic data were retrieved from NCBI dbGaP (accession: phs000286.v3.p1). *TTN* sequence data for JHS participants are available from NCBI dbGaP (accession: phs000498.v1.p1).

Population control replication cohort. Six hundred sixty-seven women of European ancestry from the WHI participants with exome data (dbGaP study accession; phs000200.v1.p1) were studied. These individuals are a subset of 161,808 postmenopausal WHI participants (aged 50 to 79 years) who were recruited and followed from 40 clinical centers across the United States between 1993 and 1998 (*24*). The WHI eligibility criteria included the ability to complete study visits with expected survival and local residency for at least 3 years. Subjects with medical conditions that would limit full participation in the study including individuals with advanced heart failure were excluded from enrolling. Phenotype ascertainment details for all study cohorts are provided in Supplementary Materials.

TTN sequence data

DCM subjects and healthy volunteers. *TTN* was sequenced in prospective DCM cases, healthy volunteers, and 103 end-stage DCM cases by using a targeted approach. Custom hybridization capture probes were designed to target genes implicated in cardiovascular disease, including *TTN*. RNA baits were designed using Agilent's eArray platform. Baits targeted all exons of all Ensembl *TTN* transcripts (Ensembl version 54), including untranslated regions, with a 100-bp extension into adjacent introns, and 1.25 kb of upstream sequence (fig. S2 and table S7). A total of 6340 unique 120-mer RNA baits were generated with increased bait tiling across the target (fivefold), covering a target region of 168,369 bp, including 112,916 protein-coding bases. DNA library preparation and target capture were performed according to the manufacturers' protocols before paired-end sequencing on the SOLiD 5500xl (Life Technologies). Reads were demultiplexed and aligned to the human reference genome (hg19) in color space using LifeScope v2.5.1 "targeted.reseq.pe" pipeline. SOLiD Accuracy Enhancement Tool (SAET) was used to improve color call accuracy before mapping. All other LifeScope parameters were used as default. Duplicate reads and those mapping with a quality score <8 were removed. Variant calling was performed with diBayes (SNPs) and small indels modules, as well as GATK v1.5-2.7 (*52*) and SAMtools v0.1.18. Variants called by any of these methods were taken forward for Sanger validation. Alignment and coverage metrics were calculated using Picard v1.40, BEDTools v2.12, and in-house Perl scripts. GATK CallableLoci Walker was used to identify target genomic regions covered sufficiently for variant calling (minimum depth >4 with base quality >20 and mapping quality >10).

In a subset of end-stage patients ($n = 54$), *TTN* was studied by whole-genome sequencing (Complete Genomics), and variants were called using Complete Genomic Analysis Tools (www.completegenomics.com/analysis-tools/cgatools/version: 2.2.0.26). Variants were filtered out if they met any of the following conditions: (i) low confidence or incomplete calls flagged by the caller, (ii) read depth less than 10, and (iii) allele frequency less than 0.15 for heterozygous calls. All remaining putative TTNtv were taken forward for Sanger validation.

No genotype information was available at recruitment, so patient inclusion was blinded to genotype. In addition to sequencing *TTN*, we sequenced known DCM genes used in clinical practice and observed no difference in the prevalence of rare protein-altering variants in *LMNA*, *MYH6*, *MYH7*, *TNNT2*, *SCN5A*, and in *TTN* missense variants between TTNtv-positive and TTNtv-negative cases in the prospectively recruited unselected DCM cohort.

Community-based cohorts. A custom set of hybridization capture probes were designed that targeted cardiovascular disease genes including *TTN*. Genomic DNA libraries were constructed for each sample, and libraries were paired-end sequenced with an Illumina HiSeq2000 as described (*8*). Sequence reads were mapped to the hg19 human reference sequence with BWA (*53*). GATK v1.3 was used to recalibrate base quality scores, locally realign reads, call single-nucleotide variants and small indels, and filter variant calls. All *TTN* nonsense, frameshift, and splicing variants reported in FHS or JHS subjects were visually inspected using the Integrated Genomics Viewer, leading to the exclusion of 30 (FHS) and 4 (JHS) variants. FHS and JHS variants were not validated by an additional genotyping method.

Replication cohorts. The DCM replication cohort was studied by targeted sequencing using an Agilent custom capture that included *TTN*, followed by sequencing on the Illumina HiSeq 2000 platform. WHI exomes were captured and sequenced on the Illumina Genome Analyzer II and HiSeq platforms as described (*54*). Sequence reads for the replication cohorts were processed using an identical pipeline. Raw sequence reads for both WHI exomes and DCM replication

cohorts were aligned to the human reference genome hg19 using Novoalign (Novocraft). Duplicates were marked with Picard (Broad Institute). Indel realignment and base quality score recalibration were done with GATK v2.7. SNP and short insertions/deletions (indels) were called using a pipeline derived from GATK v2.7 best practices. Variants were simultaneously called on all replication DCM cases and WHI controls using the UnifiedGenotyper joint variant calling module.

LV tissue studies

Tissue studies were performed on LV tissue samples from the end-stage DCM cohort, which were snap-frozen in liquid nitrogen at the time of acquisition. RNA sequencing was performed on all 84 samples, and protein studies in a subset of these. Control LV tissue used for protein studies was from unused donor hearts with no known cardiac disease, from the RBHT transplant program, stored and prepared as for the DCM samples.

Transcript studies. Total RNA was extracted from frozen LV samples from 84 end-stage DCM cases using TRIzol (Life Technologies) following the manufacturer's protocol, and quantified by ultraviolet spectrophotometry. RNA quality was measured on the Agilent 2100 Bioanalyzer using Agilent's RNA 6000 reagents. RNA integrity numbers ranged between 6.3 and 8.7 with a mean of 7.6. Total RNA (4 μg) was used for library preparation with the TruSeq RNA Sample Preparation Kit (Illumina). Barcoded cDNA (complementary DNA) fragments of poly(A)+ RNA were then sequenced on a HiSeq 2000 (Illumina) using 2 × 100–bp paired-end chemistry. Pools of six samples were loaded on three lanes to avoid batch effects and obtain sufficient coverage for splicing analyses.

Reads were initially deconvoluted and aligned to the genome to detect and exclude multimapping sequences. The remaining sequences were then mapped against the GRCh37 reference genome and transcriptome using TopHat 1.4.1 (*55*) supplied with Ensembl (*56*) gene annotations. Splice junction detection was performed to allow split alignment across both known and novel splice sites.

RNAseq data were used to compare levels of N2BA, N2B, and novex-3 in DCM samples, and in 105 GTEx samples obtained from dbGaP (*27*) and processed using the same bioinformatic pipeline.

Reads from DCM samples were filtered stringently before calling point mutations in the RNAseq data. Only reads mapping to one unique position in either the genome or the transcriptome with at most two mismatches in 100 bp were considered for further analyses. The SAMtools suite (*57*) was used to call TTNtv at all positions covered by a minimum of 10 reads.

Sequencing was also performed on paired DNA samples for all 84 individuals, as described above and all putative variants taken forward for Sanger validation. Nineteen TTNtv were identified and confirmed in genomic DNA from 18 individuals. The allele balance of 12 TTNtv could be interrogated in both DNA and RNA (table S18 and fig. S9).

To assess allelic expression across the gene (Fig. 5B), base calls at SNP positions throughout *TTN* were quantified without applying any cutoff regarding the variant fraction. The read population supporting the SNP was then compared to the total number of reads to calculate the allele balance. SNPs found in RNAseq were compared against SNPs found in the same sample by targeted next-generation sequencing or whole-genome sequencing, and those SNPs found in both DNA and RNA were considered validated. There were no discordant zygosity calls.

To estimate *TTN* expression levels, all uniquely mapping reads that could be assigned to *TTN* unambiguously and did not intersect with any other known annotated transcripts were counted for each individual. These read numbers were then quantile-normalized to be compared across all samples.

To estimate exon usage, reads covering *TTN* exons (inclusion reads) and reads completely aligning before and after the exon but not within the exon borders (exclusion reads) were counted and normalized for exon length. The proportion of reads indicating incorporation of the exon compared to the number of reads deriving from isoforms excluding the particular exon indicates the proportion of transcripts that use the respective exon [PSI score (*58*)]. For each exon, the median value of PSI across 84 samples was taken as our estimate of exon usage. This value was then applied across all cohorts to give the estimated usage level of each exon containing a TTNtv (Fig. 2).

Novex-3 expression was estimated by comparing the number of reads mapping to the last kilobase of the novex-3 terminal exon (exon 48) and the last kilobase of the full-length terminal exon (364). The ratio was calculated for each sample from GTEx and DCM, and the mean was taken as an estimate of novex-3 abundance relative to other isoforms.

Protein studies. Analysis of titin isoform expression was by vertical 1% SDS agarose gel electrophoresis (VAGE). Protein samples from LVs were homogenized in sample buffer [8 M urea/2 M thiourea/0.05 M tris (pH 6.8)/75 mM DTT (dithiothreitol)/ 3% SDS/0.05% bromophenol blue], and titin isoforms were separated using an SDS/agarose gel electrophoresis system (*59*). Each sample was run on four independent gels, which were Coomassie-stained to visualize the titin isoforms N2BA (several sizes, including both the N2A and N2B regions), N2B, and the proteolytic fragment T2. In addition, each sample was validated in four Western blots probed with two independent titin-specific antibodies, so that each sample was interrogated eight times.

Variant annotation

To facilitate standardized variant annotation in accordance with international guidelines, we developed a Locus Reference Genomic (LRG) sequence (*60*) for *TTN* (www.lrg-sequence.org, LRG_391). Variants are described relative to an inferred complete meta-transcript (LRG_391_t1) manually curated by the HAVANA group that incorporates all *TTN* exons, with the exception of a single alternative terminal exon unique to the shorter novex-3 isoform (Fig. 1, fig. S2, and tables S6 and S7). Variants in the novex-3 terminal exon are reported relative to LRG_391_t2.

Variants were reported using the Human Genome Variation Society nomenclature. The functional consequences of variants were predicted using the Ensembl Perl API (*61*) Variant Effect Predictor (*62*). Variants were classified as truncating if their consequence included one of following sequence ontology terms: "transcript_ablation," "splice_donor_variant," "splice_acceptor_variant," "stop_gained," "stop_lost," or "frameshift_variant." To identify splice variants outside of the absolutely conserved two intron bases, Alamut (*63*) was used to calculate maximum entropy (*64*), and Neural Network (NNSplice), Splice Site Finder (SSF), Human Splicing Finder (HSF), and Gene Splicer (GS) scores for reference and alternate alleles. We used the FHS cohort variant frequencies to establish a threshold value for calling splice variant predictions: for each variant in the splicing region (donor: −3 to +6, acceptor: −20 to +3), the pairs of splicing scores were subtracted from one another and converted to percentiles. Variants scored ≥90th percentile by at least three algorithms and ≥70th percentile by all applied algorithms were considered conservative splicing variant predictions, and the minimum absolute

score change for each prediction algorithm was applied as threshold across all cohorts. These thresholds were selected to be more conservative than the previously applied maximum entropy score difference ≤ −2 threshold (*12*) and to exclude variants found more frequently than in 1 in 1000 individuals.

Statistical analyses

Statistical analyses were performed with R. Comparisons between groups were performed with Wilcoxon-Mann-Whitney or Fisher's exact tests, as appropriate, except where indicated. ORs are reported with 95% confidence intervals. Significance tests are two-tailed with $\alpha = 0.05$ unless otherwise indicated. Standard linear regressions were used to evaluate the relationship between *TTN* genotype and cardiovascular phenotypes. Multivariate models were generated using known clinical covariates and optimized to minimize Bayesian information criterion. The relationships between morphologic parameters and *TTN* genotype were assessed by analysis of variance (ANOVA) between nested linear models.

Determining the likelihood that a TTNtv found in an individual with DCM is pathogenic

Excluding novex, low-expression, and predicted splice site variants:
Total TTNtv frequency in controls = 31/3911 = 0.79%
Total TTNtv frequency in unselected DCM = 45/374 = 12%.

To estimate the proportion of variants in cases that are truly pathogenic, we take the frequency of TTNtv in controls as an estimate of the burden of benign variation in both cohorts (0.79%). The burden of pathogenic TTNtv in unselected DCM is therefore conservatively estimated at 11.21%.

In an individual with DCM, the likelihood that a TTNtv is pathogenic = 11.21/0.79 = 14.2, and the probability of pathogenicity = 11.21/12 = 93.4%. The calculation for end-stage DCM is equivalent.

Stratification of DCM and linear modeling

For prospectively recruited DCM subjects, linear modeling was used to more fully assess the relationship between *TTN* genotype and cardiac phenotype. For each phenotype, a model adjusting for age and sex was optimized using Bayesian information criteria, and then compared with a similarly optimized model that also included *TTN* genotype (defined by presence/absence of TTNtv and the distance of the TTNtv from the N terminus) using ANOVA. *TTN* genotype was a significant predictor of five phenotypic indices [LV EF, RV EF, LV SVi, RV SVi, and lateral WTi (indexed wall thickness)].

Stratification of community-based cohorts and linear modeling

FHS offspring subjects were studied by echocardiography as part of exams 2, 4, 5, 6, and 8. LV EF was estimated as $(LVEDD^2 - LVESD^2)/LVEDD^2$, where EDD and ESD are end-diastolic and end-systolic dimensions, respectively.

To adjust each measure, linear regression models were built including as potential covariates age, sex, weight, body surface area (BSA), height, systolic blood pressure (SBP), diastolic blood pressure (DBP), HTN status, diabetes status, and HTN treatment status (HTN_tx), as well as interactions between age or sex and each other covariate. Diabetes status and HTN treatment status were excluded from FHS exam 8 because these clinically assessed summary data were not available. Each model was built with all covariates and then stepwise-optimized to minimize Bayesian information criterion. Final models are denoted as the first model listed for each cohort, exam, and CMR/echocardiography combination (tables S16 and S17).

Atop these baseline models, we added TTNtv status or TTNtv status plus TTNtv exon usage (PSI). The latter model is detailed in the Supplementary Materials; in the tables, models including TTNtv status plus exon expression immediately follow each paired base model.

Although adding only TTNtv status to baseline models did not improve overall model prediction, adding TTNtv status plus TTNtv exon expression did appear to improve model performance for CMR data (tables S16 and S17).

DCM outcome analysis

Kaplan-Meier and Fleming-Harrinton estimates were used to compare time to events between unselected DCM cohort subgroups (TTNtv-positive and TTNtv-negative). Models were right-censored at age 70, last contact time if aged <70, or the earliest adverse event time recorded for that patient (fig. S5).

FHS outcome analyses

Cox proportional hazard models were used to compare time to events between FHS cohort subgroups. Models were left-censored at exam 5 and right-censored at "cardiovascular disease time" if there were no events recorded; "last contact time," if available; or the latest event time recorded across all FHS offspring subjects in the survival table. Multivariate models were generated and optimized, as described in stratification of community-based cohorts and linear modeling (above), initially considering potential covariates of age, sex, high-density lipoprotein cholesterol, total cholesterol, triglycerides, body mass index, SBP, DBP, antihyperlipidemia treatment status, anti-HTN treatment status, and all pairwise interactions between age or sex and the other considered covariates (table S13 and fig. S4).

SUPPLEMENTARY MATERIALS

www.sciencetranslationalmedicine.org/cgi/content/full/7/270/270ra6/DC1
Phenotype ascertainment methods for study cohorts
Fig. S1. Schematic representation of the unselected DCM cohort recruitment pathway and analyses.
Fig. S2. *TTN* sequencing coverage for each cohort.
Fig. S3. Sites susceptible to truncating events are non-uniformly distributed within the *TTN* gene but do not influence clustering effects in the A-band.
Fig. S4. Alternative splicing of *TTN* in the human heart.
Fig. S5. Time to events in FHS individuals grouped by TTNtv presence.
Fig. S6. Truncated transcript length is correlated with indices of cardiac impairment severity in DCM.
Fig. S7. FHS exam 7 CMR.
Fig. S8. FHS and JHS additional CMR and echocardiography exams.
Fig. S9. mRNA transcripts encoding truncated TTN proteins are expressed in human LV.
Table S1. TTNtv identified in UK prospective DCM cohort.
Table S2. TTNtv identified in the FHS offspring cohort.
Table S3. TTNtv identified in the JHS cohort.
Table S4. TTNtv identified in end-stage DCM.
Table S5. TTNtv identified in healthy volunteers.
Table S6. Titin reference transcript and protein identifiers.
Table S7. Overview of *TTN* transcripts and exon usage.
Table S8. TTNtv in publicly available control populations.
Table S9. Burden, type, and distribution of TTNtv in publicly available control populations.
Table S10. FHS exam 7 CMR phenotype grouped by TTNtv presence.
Table S11. Prevalence of DCM in FHS and JHS participants, grouped by TTNtv presence.
Table S12. Time to event empirical Cox proportional hazard models for the FHS cohort.

Table S13. TTNtv identified in replication cohorts.
Table S14. Linear modeling of the relationship between *TTN* genotype and phenotype for 14 continuous variables in the unselected DCM cohort.
Table S15. Full linear model describes impact of multivariate *TTN* genotype on phenotype for 14 continuous variables in the unselected DCM cohort.
Table S16. Linear models for FHS exam 7 CMR.
Table S17. Linear models for additional FHS and JHS exams.
Table S18. Allele-specific expression of exons containing TTNtv using RNA sequencing data.
References (*65*–*67*)

REFERENCES AND NOTES

1. E. M. McNally, J. R. Golbus, M. J. Puckelwartz, Genetic mutations and mechanisms in dilated cardiomyopathy. *J. Clin. Invest.* **123**, 19–26 (2013).
2. R. E. Hershberger, D. J. Hedges, A. Morales, Dilated cardiomyopathy: The complexity of a diverse genetic architecture. *Nat. Rev. Cardiol.* **10**, 531–547 (2013).
3. J. S. Ware, A. M. Roberts, S. A. Cook, Next generation sequencing for clinical diagnostics and personalised medicine: Implications for the next generation cardiologist. *Heart* **98**, 276–281 (2012).
4. N. K. Lakdawala, B. H. Funke, S. Baxter, A. L. Cirino, A. E. Roberts, D. P. Judge, N. Johnson, N. J. Mendelsohn, C. Morel, M. Care, W. K. Chung, C. Jones, A. Psychogios, E. Duffy, H. L. Rehm, E. White, J. G. Seidman, C. E. Seidman, C. Y. Ho, Genetic testing for dilated cardiomyopathy in clinical practice. *J. Card. Fail.* **18**, 296–303 (2012).
5. J. D. Siegfried, A. Morales, J. D. Kushner, E. Burkett, J. Cowan, A. C. Mauro, G. S. Huggins, D. Li, N. Norton, R. E. Hershberger, Return of genetic results in the familial dilated cardiomyopathy research project. *J. Genet. Couns.* **22**, 164–174 (2013).
6. M. J. Ackerman, S. G. Priori, S. Willems, C. Berul, R. Brugada, H. Calkins, A. J. Camm, P. T. Ellinor, M. Gollob, R. Hamilton, R. E. Hershberger, D. P. Judge, H. Le Marec, W. J. McKenna, E. Schulze-Bahr, C. Semsarian, J. A. Towbin, H. Watkins, A. Wilde, C. Wolpert, D. P. Zipes, HRS/EHRA expert consensus statement on the state of genetic testing for the channelopathies and cardiomyopathies this document was developed as a partnership between the Heart Rhythm Society (HRS) and the European Heart Rhythm Association (EHRA). *Heart Rhythm* **8**, 1308–1339 (2011).
7. P. S. Tarpey, R. Smith, E. Pleasance, A. Whibley, S. Edkins, C. Hardy, S. O'Meara, C. Latimer, E. Dicks, A. Menzies, P. Stephens, M. Blow, C. Greenman, Y. Xue, C. Tyler-Smith, D. Thompson, K. Gray, J. Andrews, S. Barthorpe, G. Buck, J. Cole, R. Dunmore, D. Jones, M. Maddison, T. Mironenko, R. Turner, K. Turrell, J. Varian, S. West, S. Widaa, P. Wray, J. Teague, A. Butler, A. Jenkinson, M. Jia, D. Richardson, R. Shepherd, R. Wooster, M. I. Tejada, F. Martinez, G. Carvill, R. Goliath, A. P. de Brouwer, H. van Bokhoven, H. Van Esch, J. Chelly, M. Raynaud, H. H. Ropers, F. E. Abidi, A. K. Srivastava, J. Cox, Y. Luo, U. Mallya, J. Moon, J. Parnau, S. Mohammed, J. L. Tolmie, C. Shoubridge, M. Corbett, A. Gardner, E. Haan, S. Rujirabanjerd, M. Shaw, L. Vandeleur, T. Fullston, D. F. Easton, J. Boyle, M. Partington, A. Hackett, M. Field, C. Skinner, R. E. Stevenson, M. Bobrow, G. Turner, C. E. Schwartz, J. Gecz, F. L. Raymond, P. A. Futreal, M. R. Stratton, A systematic, large-scale resequencing screen of X-chromosome coding exons in mental retardation. *Nat. Genet.* **41**, 535–543 (2009).
8. A. G. Bick, J. Flannick, K. Ito, S. Cheng, R. S. Vasan, M. G. Parfenov, D. S. Herman, S. R. DePalma, N. Gupta, S. B. Gabriel, B. H. Funke, H. L. Rehm, E. J. Benjamin, J. Aragam, H. A. Taylor Jr., E. R. Fox, C. Newton-Cheh, S. Kathiresan, C. J. O'Donnell, J. G. Wilson, D. M. Altshuler, J. N. Hirschhorn, J. G. Seidman, C. Seidman, Burden of rare sarcomere gene variants in the Framingham and Jackson Heart Study cohorts. *Am. J. Hum. Genet.* **91**, 513–519 (2012).
9. J. Flannick, N. L. Beer, A. G. Bick, V. Agarwala, J. Molnes, N. Gupta, N. P. Burtt, J. C. Florez, J. B. Meigs, H. Taylor, V. Lyssenko, H. Irgens, E. Fox, F. Burslem, S. Johansson, M. J. Brosnan, J. K. Trimmer, C. Newton-Cheh, T. Tuomi, A. Molven, J. G. Wilson, C. J. O'Donnell, S. Kathiresan, J. N. Hirschhorn, P. R. Njølstad, T. Rolph, J. G. Seidman, S. Gabriel, D. R. Cox, C. E. Seidman, L. Groop, D. Altshuler, Assessing the phenotypic effects in the general population of rare variants in genes for a dominant Mendelian form of diabetes. *Nat. Genet.* **45**, 1380–1385 (2013).
10. B. L. Siu, H. Niimura, J. A. Osborne, D. Fatkin, C. MacRae, S. Solomon, D. W. Benson, J. G. Seidman, C. E. Seidman, Familial dilated cardiomyopathy locus maps to chromosome 2q31. *Circulation* **99**, 1022–1026 (1999).
11. B. Gerull, M. Gramlich, J. Atherton, M. McNabb, K. Trombitas, S. Sasse-Klaassen, J. G. Seidman, C. Seidman, H. Granzier, S. Labeit, M. Frenneaux, L. Thierfelder, Mutations of *TTN*, encoding the giant muscle filament titin, cause familial dilated cardiomyopathy. *Nat. Genet.* **30**, 201–204 (2002).
12. D. S. Herman, L. Lam, M. R. Taylor, L. Wang, P. Teekakirikul, D. Christodoulou, L. Conner, S. R. DePalma, B. McDonough, E. Sparks, D. L. Teodorescu, A. L. Cirino, N. R. Banner, D. J. Pennell, S. Graw, M. Merlo, A. Di Lenarda, G. Sinagra, J. M. Bos, M. J. Ackerman, R. N. Mitchell, C. E. Murry, N. K. Lakdawala, C. Y. Ho, P. J. Barton, S. A. Cook, L. Mestroni, J. G. Seidman, C. E. Seidman, Truncations of titin causing dilated cardiomyopathy. *N. Engl. J. Med.* **366**, 619–628 (2012).
13. J. R. Golbus, M. J. Puckelwartz, J. P. Fahrenbach, L. M. Dellefave-Castillo, D. Wolfgeher, E. M. McNally, Population-based variation in cardiomyopathy genes. *Circ. Cardiovasc. Genet.* **5**, 391–399 (2012).
14. S. Pan, C. A. Caleshu, K. E. Dunn, M. J. Foti, M. K. Moran, O. Soyinka, E. A. Ashley, Cardiac structural and sarcomere genes associated with cardiomyopathy exhibit marked intolerance of genetic variation. *Circ. Cardiovasc. Genet.* **5**, 602–610 (2012).
15. A. Kontrogianni-Konstantopoulos, M. A. Ackermann, A. L. Bowman, S. V. Yap, R. J. Bloch, Muscle giants: Molecular scaffolds in sarcomerogenesis. *Physiol. Rev.* **89**, 1217–1267 (2009).
16. C. C. Gregorio, K. Trombitás, T. Centner, B. Kolmerer, G. Stier, K. Kunke, K. Suzuki, F. Obermayr, B. Herrmann, H. Granzier, H. Sorimachi, S. Labeit, The NH_2 terminus of titin spans the Z-disc: Its interaction with a novel 19-kD ligand (T-cap) is required for sarcomeric integrity. *J. Cell Biol.* **143**, 1013–1027 (1998).
17. W. A. Linke, H. Granzier, A spring tale: New facts on titin elasticity. *Biophys. J.* **75**, 2613–2614 (1998).
18. L. A. Leinwand, J. C. Tardiff, C. C. Gregorio, Mutations in the sensitive giant titin result in a broken heart. *Circ. Res.* **111**, 158–161 (2012).
19. E. M. Puchner, A. Alexandrovich, A. L. Kho, U. Hensen, L. V. Schäfer, B. Brandmeier, F. Grater, H. Grubmüller, H. E. Gaub, M. Gautel, Mechanoenzymatics of titin kinase. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 13385–13390 (2008).
20. M. M. LeWinter, H. L. Granzier, Titin is a major human disease gene. *Circulation* **127**, 938–944 (2013).
21. M. S. Kellermayer, S. B. Smith, C. Bustamante, H. L. Granzier, Complete unfolding of the titin molecule under external force. *J. Struct. Biol.* **122**, 197–205 (1998).
22. D. R. Govindaraju, L. A. Cupples, W. B. Kannel, C. J. O'Donnell, L. D. Atwood, R. B. D'Agostino Sr., C. S. Fox, M. Larson, D. Levy, J. Murabito, R. S. Vasan, G. L. Splansky, P. A. Wolf, E. J. Benjamin, Genetics of the Framingham Heart Study population. *Adv. Genet.* **62**, 33–65 (2008).
23. S. B. Wyatt, N. Diekelmann, F. Henderson, M. E. Andrew, G. Billingsley, S. H. Felder, S. Fuqua, P. B. Jackson, A community-driven model of research participation: The Jackson Heart Study Participant Recruitment and Retention Study. *Ethn. Dis.* **13**, 438–455 (2003).
24. The Women's Health Initiative Study Group, Design of the Women's Health Initiative clinical trial and observational study. *Control. Clin. Trials* **19**, 61–109 (1998).
25. D. G. MacArthur, S. Balasubramanian, A. Frankish, N. Huang, J. Morris, K. Walter, L. Jostins, L. Habegger, J. K. Pickrell, S. B. Montgomery, C. A. Albers, Z. D. Zhang, D. F. Conrad, G. Lunter, H. Zheng, Q. Ayub, M. A. DePristo, E. Banks, M. Hu, R. E. Handsaker, J. A. Rosenfeld, M. Fromer, M. Jin, X. J. Mu, E. Khurana, K. Ye, M. Kay, G. I. Saunders, M. M. Suner, T. Hunt, I. H. Barnes, C. Amid, D. R. Carvalho-Silva, A. H. Bignell, C. Snow, B. Yngvadottir, S. Bumpstead, D. N. Cooper, Y. Xue, I. G. Romero; 1000 Genomes Project Consortium, J. Wang, Y. Li, R. A. Gibbs, S. A. McCarroll, E. T. Dermitzakis, J. K. Pritchard, J. C. Barrett, J. Harrow, M. E. Hurles, M. B. Gerstein, C. Tyler-Smith, A systematic survey of loss-of-function variants in human protein-coding genes. *Science* **335**, 823–828 (2012).
26. M. L. Bang, T. Centner, F. Fornoff, A. J. Geach, M. Gotthardt, M. McNabb, C. C. Witt, D. Labeit, C. C. Gregorio, H. Granzier, S. Labeit, The complete gene sequence of titin, expression of an unusual approximately 700-kDa titin isoform, and its interaction with obscurin identify a novel Z-line to I-band linking system. *Circ. Res.* **89**, 1065–1072 (2001).
27. H. M. Moore, Acquisition of normal tissues for the GTEx program. *Biopreserv. Biobank* **11**, 75–76 (2013).
28. 1000 Genomes Project Consortium, G. R. Abecasis, D. Altshuler, A. Auton, L. D. Brooks, R. M. Durbin, R. A. Gibbs, M. E. Hurles, G. A. McVean, A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073 (2010).
29. R. E. Mills, K. Walter, C. Stewart, R. E. Handsaker, K. Chen, C. Alkan, A. Abyzov, S. C. Yoon, K. Ye, R. K. Cheetham, A. Chinwalla, D. F. Conrad, Y. Fu, F. Grubert, I. Hajirasouliha, F. Hormozdiari, L. M. Iakoucheva, Z. Iqbal, S. Kang, J. M. Kidd, M. K. Konkel, J. Korn, E. Khurana, D. Kural, H. Y. Lam, J. Leng, R. Li, Y. Li, C. Y. Lin, R. Luo, X. J. Mu, J. Nemesh, H. E. Peckham, T. Rausch, A. Scally, X. Shi, M. P. Stromberg, A. M. Stutz, A. E. Urban, J. A. Walker, J. Wu, Y. Zhang, Z. D. Zhang, M. A. Batzer, L. Ding, G. T. Marth, G. McVean, J. Sebat, M. Snyder, J. Wang, K. Ye, E. E. Eichler, M. B. Gerstein, M. E. Hurles, C. Lee, S. A. McCarroll, J. O. Korbel; 1000 Genomes Project Consortium, Mapping copy number variation by population-scale genome sequencing. *Nature* **470**, 59–65 (2011).
30. E. Villard, C. Perret, F. Gary, C. Proust, G. Dilanian, C. Hengstenberg, V. Ruppert, E. Arbustini, T. Wichter, M. Germain, O. Dubourg, L. Tavazzi, M. C. Aumont, P. DeGroote, L. Fauchier, J. N. Trochu, P. Gibelin, J. F. Aupetit, K. Stark, J. Erdmann, R. Hetzer, A. M. Roberts, P. J. Barton, V. Regitz-Zagrosek; Cardiogenics Consortium, U. Aslam, L. Duboscq-Bidot, M. Meyborg, B. Maisch, H. Madeira, A. Waldenström, E. Galve, J. G. Cleland, R. Dorent, G. Roizes, T. Zeller, S. Blankenberg, A. H. Goodall, S. Cook, D. A. Tregouet, L. Tiret, R. Isnard, M. Komajda, P. Charron, F. Cambien, A genome-wide association study identifies two loci associated with heart failure due to dilated cardiomyopathy. *Eur. Heart J.* **32**, 1065–1076 (2011).
31. K. Stark, U. B. Esslinger, W. Reinhard, G. Petrov, T. Winkler, M. Komajda, R. Isnard, P. Charron, E. Villard, F. Cambien, L. Tiret, M. C. Aumont, O. Dubourg, J. N. Trochu, L. Fauchier, P. Degroote, A. Richter, B. Maisch, T. Wichter, C. Zollbrecht, M. Grassl, H. Schunkert, P. Linsel-Nitschke, J. Erdmann, J. Baumert, T. Illig, N. Klopp, H. E. Wichmann, C. Meisinger, W. Koenig, P. Lichtner, T. Meitinger, A. Schillert, I. R. König, R. Hetzer, I. M. Heid, V. Regitz-Zagrosek, C. Hengstenberg,

Genetic association study identifies *HSPB7* as a risk gene for idiopathic dilated cardiomyopathy. *PLOS Genet.* **6**, e1001167 (2010).
32. N. G. Keenan, D. J. Pennell, CMR of ventricular function. *Echocardiography* **24**, 185–193 (2007).
33. C. Parsai, R. O'Hanlon, S. K. Prasad, R. H. Mohiaddin, Diagnostic and prognostic value of cardiovascular magnetic resonance in non-ischaemic cardiomyopathies. *J. Cardiovasc. Magn. Reson.* **14**, 54 (2012).
34. A. Gulati, A. Jabbour, T. F. Ismail, K. Guha, J. Khwaja, S. Raza, K. Morarji, T. D. Brown, N. A. Ismail, M. R. Dweck, E. Di Pietro, M. Roughton, R. Wage, Y. Daryani, R. O'Hanlon, M. N. Sheppard, F. Alpendurada, A. R. Lyon, S. A. Cook, M. R. Cowie, R. G. Assomull, D. J. Pennell, S. K. Prasad, Association of fibrosis with mortality and sudden cardiac death in patients with nonischemic dilated cardiomyopathy. *JAMA* **309**, 896–908 (2013).
35. R. G. Assomull, S. K. Prasad, J. Lyne, G. Smith, E. D. Burman, M. Khan, M. N. Sheppard, P. A. Poole-Wilson, D. J. Pennell, Cardiovascular magnetic resonance, fibrosis, and prognosis in dilated cardiomyopathy. *J. Am. Coll. Cardiol.* **48**, 1977–1985 (2006).
36. R. C. Green, J. S. Berg, W. W. Grody, S. S. Kalia, B. R. Korf, C. L. Martin, A. L. McGuire, R. L. Nussbaum, J. M. O'Daniel, K. E. Ormond, H. L. Rehm, M. S. Watson, M. S. Williams, L. G. Biesecker; American College of Medical Genetics and Genomics, ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. *Genet. Med.* **15**, 565–574 (2013).
37. R. E. Hershberger, J. Lindenfeld, L. Mestroni, C. E. Seidman, M. R. Taylor, J. A. Towbin; Heart Failure Society of America, Genetic evaluation of cardiomyopathy—A Heart Failure Society of America practice guideline. *J. Card. Fail.* **15**, 83–97 (2009).
38. J. D. Kushner, D. Nauman, D. Burgess, S. Ludwigsen, S. B. Parks, G. Pantely, E. Burkett, R. E. Hershberger, Clinical characteristics of 304 kindreds evaluated for familial dilated cardiomyopathy. *J. Card. Fail.* **12**, 422–429 (2006).
39. P. Richardson, W. McKenna, M. Bristow, B. Maisch, B. Mautner, J. O'Connell, E. Olsen, G. Thiene, J. Goodwin, I. Gyarfas, I. Martin, P. Nordet, Report of the 1995 World Health Organization/International Society and Federation of Cardiology Task Force on the Definition and Classification of Cardiomyopathies. *Circulation* **93**, 841–842 (1996).
40. E. M. McNally, Genetics: Broken giant linked to heart failure. *Nature* **483**, 281–282 (2012).
41. R. Roncarati, C. Viviani Anselmi, P. Krawitz, G. Lattanzi, Y. von Kodolitsch, A. Perrot, E. di Pasquale, L. Papa, P. Portararo, M. Columbaro, A. Forni, G. Faggian, G. Condorelli, P. N. Robinson, Doubly heterozygous *LMNA* and *TTN* mutations revealed by exome sequencing in a severe form of dilated cardiomyopathy. *Eur. J. Hum. Genet.* **21**, 1105–1111 (2013).
42. P. N. Robinson, Deep phenotyping for precision medicine. *Hum. Mutat.* **33**, 777–780 (2012).
43. E. Pahl, L. A. Sleeper, C. E. Canter, D. T. Hsu, M. Lu, S. A. Webber, S. D. Colan, P. F. Kantor, M. D. Everitt, J. A. Towbin, J. L. Jefferies, B. D. Kaufman, J. D. Wilkinson, S. E. Lipshultz; Pediatric Cardiomyopathy Registry Investigators, Incidence of and risk factors for sudden cardiac death in children with dilated cardiomyopathy: A report from the Pediatric Cardiomyopathy Registry. *J. Am. Coll. Cardiol.* **59**, 607–615 (2012).
44. S. D. Solomon, N. Anavekar, H. Skali, J. J. McMurray, K. Swedberg, S. Yusuf, C. B. Granger, E. L. Michelson, D. Wang, S. Pocock, M. A. Pfeffer; Candesartan in Heart Failure Reduction in Mortality (CHARM) Investigators, Influence of ejection fraction on cardiovascular outcomes in a broad spectrum of heart failure patients. *Circulation* **112**, 3738–3744 (2005).
45. O. Gjesdal, D. A. Bluemke, J. A. Lima, Cardiac remodeling at the population level—Risk factors, screening, and outcomes. *Nat. Rev. Cardiology* **8**, 673–685 (2011).
46. M. Pasotti, C. Klersy, A. Pilotto, N. Marziliano, C. Rapezzi, A. Serio, S. Mannarino, F. Gambarin, V. Favalli, M. Grasso, M. Agozzino, C. Campana, A. Gavazzi, O. Febo, M. Marini, M. Landolina, A. Mortara, G. Piccolo, M. Vigano, L. Tavazzi, E. Arbustini, Long-term outcome and risk stratification in dilated cardiolaminopathies. *J. Am. Coll. Cardiol.* **52**, 1250–1260 (2008).
47. P. Prontera, L. Bernardini, G. Stangoni, A. Capalbo, D. Rogaia, C. Ardisia, A. Novelli, B. Dallapiccola, E. Donti, 2q31.2q32.3 deletion syndrome: Report of an adult patient. *Am. J. Med. Genet. A* **149A**, 706–712 (2009).
48. O. Ceyhan-Birsoy, P. B. Agrawal, C. Hidalgo, K. Schmitz-Abe, E. T. DeChene, L. C. Swanson, R. Soemedi, N. Vasli, S. T. Iannaccone, P. B. Shieh, N. Shur, J. M. Dennison, M. W. Lawlor, J. Laporte, K. Markianos, W. G. Fairbrother, H. Granzier, A. H. Beggs, Recessive truncating titin gene, *TTN*, mutations presenting as centronuclear myopathy. *Neurology* **81**, 1205–1214 (2013).
49. Q. Yang, A. Sanbe, H. Osinska, T. E. Hewett, R. Klevitsky, J. Robbins, A mouse model of myosin binding protein C human familial hypertrophic cardiomyopathy. *J. Clin. Invest.* **102**, 1292–1300 (1998).
50. A. Sarikas, L. Carrier, C. Schenke, D. Doll, J. Flavigny, K. S. Lindenberg, T. Eschenhagen, O. Zolk, Impairment of the ubiquitin–proteasome system by truncated cardiac myosin binding protein C mutants. *Cardiovasc. Res.* **66**, 33–44 (2005).
51. A. M. Maceira, S. K. Prasad, M. Khan, D. J. Pennell, Normalized left ventricular systolic and diastolic function by steady state free precession cardiovascular magnetic resonance. *J. Cardiovasc. Magn. Reson.* **8**, 417–426 (2006).
52. A. McKenna, M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, M. A. DePristo, The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
53. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
54. J. A. Tennessen, A. W. Bigham, T. D. O'Connor, W. Fu, E. E. Kenny, S. Gravel, S. McGee, R. Do, X. Liu, G. Jun, H. M. Kang, D. Jordan, S. M. Leal, S. Gabriel, M. J. Rieder, G. Abecasis, D. Altshuler, D. A. Nickerson, E. Boerwinkle, S. Sunyaev, C. D. Bustamante, M. J. Bamshad, J. M. Akey; NHLBI Exome Sequencing Project, Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science* **337**, 64–69 (2012).
55. C. Trapnell, L. Pachter, S. L. Salzberg, TopHat: Discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105–1111 (2009).
56. P. Flicek, M. R. Amode, D. Barrell, K. Beal, K. Billis, S. Brent, D. Carvalho-Silva, P. Clapham, G. Coates, S. Fitzgerald, L. Gil, C. G. Girón, L. Gordon, T. Hourlier, S. Hunt, N. Johnson, T. Juettemann, A. K. Kähäri, S. Keenan, E. Kulesha, F. J. Martin, T. Maurel, W. M. McLaren, D. N. Murphy, R. Nag, B. Overduin, M. Pignatelli, B. Pritchard, E. Pritchard, H. S. Riat, M. Ruffier, D. Sheppard, K. Taylor, A. Thormann, S. J. Trevanion, A. Vullo, S. P. Wilder, M. Wilson, A. Zadissa, B. L. Aken, E. Birney, F. Cunningham, J. Harrow, J. Herrero, T. J. Hubbard, R. Kinsella, M. Muffato, A. Parker, G. Spudich, A. Yates, D. R. Zerbino, S. M. Searle, Ensembl 2014. *Nucleic Acids Res.* **42**, D749–D755 (2014).
57. H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin; 1000 Genome Project Data Processing Subgroup, The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
58. E. T. Wang, R. Sandberg, S. Luo, I. Khrebtukova, L. Zhang, C. Mayr, S. F. Kingsmore, G. P. Schroth, C. B. Burge, Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**, 470–476 (2008).
59. C. M. Warren, P. R. Krzesinski, M. L. Greaser, Vertical agarose gel electrophoresis and electroblotting of high-molecular-weight proteins. *Electrophoresis* **24**, 1695–1702 (2003).
60. J. A. MacArthur, J. Morales, R. E. Tully, A. Astashyn, L. Gil, E. A. Bruford, P. Larsson, P. Flicek, R. Dalgleish, D. R. Maglott, F. Cunningham, Locus Reference Genomic: Reference sequences for the reporting of clinically relevant sequence variants. *Nucleic Acids Res.* **42**, D873–D878 (2014).
61. D. Rios, W. M. McLaren, Y. Chen, E. Birney, A. Stabenau, P. Flicek, F. Cunningham, A database and API for variation, dense genotyping and resequencing data. *BMC Bioinformatics* **11**, 238 (2010).
62. W. McLaren, B. Pritchard, D. Rios, Y. Chen, P. Flicek, F. Cunningham, Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics* **26**, 2069–2070 (2010).
63. C. Houdayer, V. Caux-Moncoutier, S. Krieger, M. Barrois, F. Bonnet, V. Bourdon, M. Bronner, M. Buisson, F. Coulet, P. Gaildrat, C. Lefol, M. Léone, S. Mazoyer, D. Muller, A. Remenieras, F. Révillion, E. Rouleau, J. Sokolowska, J. P. Vert, R. Lidereau, F. Soubrier, H. Sobol, N. Sevenet, B. Bressac-de Paillerets, A. Hardouin, M. Tosi, O. M. Sinilnikova, D. Stoppa-Lyonnet, Guidelines for splicing analysis in molecular diagnosis derived from a set of 327 combined in silico/in vitro studies on *BRCA1* and *BRCA2* variants. *Hum. Mutat.* **33**, 1228–1238 (2012).
64. G. Yeo, C. B. Burge, Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J. Comput. Biol.* **11**, 377–394 (2004).
65. F. Grothues, G. C. Smith, J. C. Moon, N. G. Bellenger, P. Collins, H. U. Klein, D. J. Pennell, Comparison of interstudy reproducibility of cardiovascular magnetic resonance with two-dimensional echocardiography in normal subjects and in patients with heart failure or left ventricular hypertrophy. *Am. J. Cardiol.* **90**, 29–34 (2002).
66. A. Gulati, T. F. Ismail, A. Jabbour, N. A. Ismail, K. Morarji, A. Ali, S. Raza, J. Khwaja, T. D. Brown, E. Liodakis, A. J. Baksi, R. Shakur, K. Guha, M. Roughton, R. Wage, S. A. Cook, F. Alpendurada, R. G. Assomull, R. H. Mohiaddin, M. R. Cowie, D. J. Pennell, S. K. Prasad, Clinical utility and prognostic value of left atrial volume assessment by cardiovascular magnetic resonance in non-ischaemic dilated cardiomyopathy. *Eur. J. Heart Fail.* **15**, 660–670 (2013).
67. M. R. Taylor, E. Carniel, L. Mestroni, Cardiomyopathy, familial dilated. *Orphanet. J. Rare Dis.* **1**, 27 (2006).

Acknowledgments: We thank all the patients, healthy volunteers, and participants in the FHS, JHS, and WHI for taking part in this research, and our team of research nurses across the hospital sites. **Funding:** The research was supported by the NIHR Biomedical Research Unit in Cardiovascular Disease at Royal Brompton & Harefield NHS Foundation Trust and Imperial College London, NIHR Imperial Biomedical Research Centre, British Heart Foundation UK (SP/10/10/28431, PG/12/27/29489), European Molecular Biology Laboratory, MRC UK, Wellcome Trust UK (087183/Z/08/Z, 092854/Z/10/Z, WT095908), Fondation Leducq, Tanoto Foundation, Goh Foundation, Academy of Medical Sciences, Arthritis Research UK, Heart Research UK, CORDA, National Medical Research Council (NMRC) Singapore, Rosetrees Trust, European Community's Seventh Framework Programme (FP7) [CardioNeT-ITN-289600; 200754 - the GEN2PHEN project], National Human Genome Research Institute (U54 HG003067), NIH (HL080494, 5-T32-GM007748-33), Howard Hughes Medical Institute, and the Australian National Health and Medical Research Council. The FHS was supported by the NHLBI (N01-HC-25195, 6R01-NS 17950), and genotyping services from Affymetrix Inc. (N02-HL-6-4278). The JHS is supported by NHLBI (N01-HC-95170, N01-HC-95171, N01-HC-95172), the National Institute for Minority Health and Health Disparities, and the National Institute of Biomedical Imaging and Bioengineering. The WHI Sequencing Project is supported by NHLBI (HL-102924), NIH, and U.S. Department of Health and Human Services through contracts N01WH22110, 24152, 32100-2, 32105-6, 32108-9, 32111-13, 32115, 32118-32119, 32122, 42107-26, 42129-32, and 44221. This publication reflects only the author's views, and the funders are

not liable for any use that may be made of the information contained herein. **Author contributions:** Data acquisition and primary analyses: A.M.R., J.S.W., D.S.H., S.S., J.B., A.G.B., R.J.B., R.W., S.J., S.W., F.M., L.E.F., S.G., J.A.L.M., F.C., J.F., S.B.G., D.M.A., P.S.M., M.H., A.M.K., C.S.H., N.R.B., D.J.P., D.P.O., T.R.S., A.D.M., T.J.W.D., A.G., E.J.B., M.H.Y., M.R., M.G., J.G.W., C.J.O., S.K.P., P.J.R.B., D.F., N.H., and J.G.S. Study conception and design, data synthesis, statistical analyses, and manuscript preparation: A.M.R., J.S.W., D.S.H., P.J.R.B., J.G.S., C.E.S., and S.A.C. All authors have seen and approved the final manuscript. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All genomic variants presented in the manuscript have been submitted to ClinVar (accession SCV000189630-SCV000189803). RNAseq data are deposited at ArrayExpress (E-MTAB-2466). Data from population cohorts have previously been deposited into dbGaP (accessions phs000007.v18.p7, phs000307.v3.p7, phs000286.v3.p1, phs000498.v1.p1, phs000200.v1.p1). Transcript annotations, including PSI values, are available at http://cardiodb.org/titin.

Submitted 21 July 2014
Accepted 25 November 2014
Published 14 January 2015
10.1126/scitranslmed.3010134

Citation: A. M. Roberts, J. S. Ware, D. S. Herman, S. Schafer, J. Baksi, A. G. Bick, R. J. Buchan, R. Walsh, S. John, S. Wilkinson, F. Mazzarotto, L. E. Felkin, S. Gong, J. A. L. MacArthur, F. Cunningham, J. Flannick, S. B. Gabriel, D. M. Altshuler, P. S. Macdonald, M. Heinig, A. M. Keogh, C. S. Hayward, N. R. Banner, D. J. Pennell, D. P. O'Regan, T. R. San, A. de Marvao, T. J. W. Dawes, A. Gulati, E. J. Birks, M. H. Yacoub, M. Radke, M. Gotthardt, J. G. Wilson, C. J. O'Donnell, S. K. Prasad, P. J. R. Barton, D. Fatkin, N. Hubner, J. G. Seidman, C. E. Seidman, S. A. Cook, Integrated allelic, transcriptional, and phenomic dissection of the cardiac effects of titin truncations in health and disease. *Sci. Transl. Med.* **7**, 270ra6 (2015).

Science Translational Medicine

Integrated allelic, transcriptional, and phenomic dissection of the cardiac effects of titin truncations in health and disease

Angharad M. Roberts, James S. Ware, Daniel S. Herman, Sebastian Schafer, John Baksi, Alexander G. Bick, Rachel J. Buchan, Roddy Walsh, Shibu John, Samuel Wilkinson, Francesco Mazzarotto, Leanne E. Felkin, Sungsam Gong, Jacqueline A. L. MacArthur, Fiona Cunningham, Jason Flannick, Stacey B. Gabriel, David M. Altshuler, Peter S. Macdonald, Matthias Heinig, Anne M. Keogh, Christopher S. Hayward, Nicholas R. Banner, Dudley J. Pennell, Declan P. O'Regan, Tan Ru San, Antonio de Marvao, Timothy J. W. Dawes, Ankur Gulati, Emma J. Birks, Magdi H. Yacoub, Michael Radke, Michael Gotthardt, James G. Wilson, Christopher J. O'Donnell, Sanjay K. Prasad, Paul J. R. Barton, Diane Fatkin, Norbert Hubner, Jonathan G. Seidman, Christine E. Seidman and Stuart A. Cook

Sci Transl Med **7**, 270ra6270ra6.
DOI: 10.1126/scitranslmed.3010134

What Happens When Titins Are Trimmed?

The most common form of inherited heart failure, dilated cardiomyopathy, can be caused by mutations in a mammoth heart protein, appropriately called titin. Now, Roberts *et al.* sort out which titin mutations cause disease and why some people can carry certain titin mutations but remain perfectly healthy. In an exhaustive survey of more than 5200 people, with and without cardiomyopathy, the authors sequenced the titin gene and measured its corresponding RNA and protein levels. The alterations in titin were truncating mutations, which cause short nonfunctional versions of the RNA or protein. These defects produced cardiomyopathy when they occurred closer to the protein's carboxyl terminus and in exons that were abundantly transcribed. The titin-truncating mutations that occur in the general population tended not to have these characteristics and were usually benign. This new detailed understanding of the molecular basis of dilated cardiomyopathy penetrance will promote better disease management and accelerate rational patient stratification.