

Open

# Reassessment of Mendelian gene pathogenicity using 7,855 cardiomyopathy cases and 60,706 reference samples

Roddy Walsh, BSc, MSc<sup>1,2</sup>, Kate L. Thomson, BSc, FRCPPath<sup>3,4</sup>, James S. Ware, PhD, MRCP<sup>1,2,5</sup>, Birgit H. Funke, PhD, FACMG<sup>6,7</sup>, Jessica Woodley, BSc<sup>3</sup>, Karen J. McGuire, BSc<sup>3</sup>, Francesco Mazzarotto, BSc, MSc<sup>1,2</sup>, Edward Blair, BMSc, MRCP<sup>8</sup>, Anneke Seller, PhD<sup>3</sup>, Jenny C. Taylor, PhD<sup>9,10</sup>, Eric V. Minikel, MS<sup>11-14</sup>, Exome Aggregation Consortium<sup>14</sup>, Daniel G. MacArthur, PhD<sup>11,12,14,15</sup>, Martin Farrall, FRCPPath<sup>4,10</sup>, Stuart A. Cook, PhD, MRCPath<sup>2,5,16,17</sup> and Hugh Watkins, MD, PhD<sup>4,10</sup>

**Purpose:** The accurate interpretation of variation in Mendelian disease genes has lagged behind data generation as sequencing has become increasingly accessible. Ongoing large sequencing efforts present huge interpretive challenges, but they also provide an invaluable opportunity to characterize the spectrum and importance of rare variation.

**Methods:** We analyzed sequence data from 7,855 clinical cardiomyopathy cases and 60,706 Exome Aggregation Consortium (ExAC) reference samples to obtain a better understanding of genetic variation in a representative autosomal dominant disorder.

**Results:** We found that in some genes previously reported as important causes of a given cardiomyopathy, rare variation is not clinically

informative because there is an unacceptably high likelihood of false-positive interpretation. By contrast, in other genes, we find that diagnostic laboratories may be overly conservative when assessing variant pathogenicity.

**Conclusions:** We outline improved analytical approaches that evaluate which genes and variant classes are interpretable and propose that these will increase the clinical utility of testing across a range of Mendelian diseases.

*Genet Med* advance online publication 17 August 2016

**Key Words:** clinical genetics; Exome Aggregation Consortium; inherited cardiomyopathy; Mendelian genetics; variation interpretation

## INTRODUCTION

Interpretation of rare genetic variation, whether in a clinical diagnostic or research setting, has not kept pace with the accelerating data generation using high-throughput DNA sequencing. Increasingly extensive gene panels, as well as whole-exome and genome sequencing, are used to interrogate the growing number of genes implicated in Mendelian diseases.<sup>1</sup> However, such panels only modestly increase the number of high-confidence diagnostic results while identifying ever larger numbers of variants of uncertain significance<sup>2,3</sup>; these inconclusive results not only reduce the clinical utility of testing but also can lead to misinterpretation and misdiagnosis.

Central to the challenge of rare variant interpretation is the paradox that individually rare variants are now seen to

be collectively common. Although it is accepted that a common variant can be excluded as a cause of a rare and penetrant Mendelian disease, the community has been slower to recognize that many rare variants identified in Mendelian disease genes are innocent bystanders and some “rare” variants are not rare at all. Recent population sequencing efforts have raised awareness of these issues (e.g., the 1000 Genomes Project,<sup>4</sup> the Exome Sequencing Project (<http://evs.gs.washington.edu/EVS>)), but the full extent is now revealed in the Exome Aggregation Consortium (ExAC) data set (<http://exac.broadinstitute.org>), in which the average exome contains 7.6 rare non-synonymous variants (minor allele frequency (MAF) <0.1%) in well-characterized dominant disease genes, with the majority being very rare or “private.”<sup>5</sup> Clearly, only a small minority can actually cause a penetrant Mendelian disease.<sup>6</sup>

The first two authors and the last two authors contributed equally to this work.

<sup>1</sup>NIHR Royal Brompton Cardiovascular Biomedical Research Unit, Royal Brompton Hospital and Imperial College London, London, UK; <sup>2</sup>National Heart and Lung Institute, Imperial College London, UK; <sup>3</sup>Oxford Medical Genetics Laboratory, Oxford University Hospitals NHS Foundation Trust, The Churchill Hospital, Oxford, UK; <sup>4</sup>Radcliffe Department of Medicine, University of Oxford, Oxford, UK; <sup>5</sup>MRC Clinical Sciences Centre, Imperial College London, UK; <sup>6</sup>Laboratory for Molecular Medicine, Partners HealthCare Personalized Medicine, Cambridge, Massachusetts, USA; <sup>7</sup>Department of Pathology, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, USA; <sup>8</sup>Department of Clinical Genetics, Oxford University Hospitals NHS Foundation Trust, The Churchill Hospital, Oxford, UK; <sup>9</sup>Oxford NIHR Biomedical Research Centre, Oxford, UK; <sup>10</sup>The Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK; <sup>11</sup>Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts, USA; <sup>12</sup>Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA; <sup>13</sup>Program in Biological and Biomedical Sciences, Harvard Medical School, Boston, Massachusetts, USA; <sup>14</sup>Exome Aggregation Consortium (ExAC), Cambridge, Massachusetts, USA; <sup>15</sup>Department of Medicine, Harvard Medical School, Boston, Massachusetts, USA; <sup>16</sup>National Heart Centre Singapore, Singapore; <sup>17</sup>Duke-National University of Singapore, Singapore. Correspondence: Hugh Watkins ([hugh.watkins@rdm.ox.ac.uk](mailto:hugh.watkins@rdm.ox.ac.uk)) or Stuart A. Cook ([stuart.cook@imperial.ac.uk](mailto:stuart.cook@imperial.ac.uk)). Members of the Exome Aggregation Consortium are listed in the **Supplementary Information** online.

Submitted 22 March 2016; accepted 10 May 2016; advance online publication 17 August 2016. doi:10.1038/gim.2016.90

The challenges of variant interpretation in Mendelian disorders are particularly well illustrated by inherited cardiomyopathies: hypertrophic cardiomyopathy (HCM), dilated cardiomyopathy (DCM), and arrhythmogenic right ventricular cardiomyopathy (ARVC). These largely autosomal dominant disorders are relatively common, genetically heterogeneous, and medically important;<sup>7</sup> consequently, cardiomyopathy genes feature prominently in the American College of Medical Genetics and Genomics list of proposed genes to be routinely analyzed in all exome or genome sequencing.<sup>8</sup> Although clinical genetic testing in cardiomyopathy has been available for more than a decade, the number of genes reported as disease-causing has increased dramatically in recent years, often without robust evidence.

Here, we leveraged two substantial resources to better understand and interpret rare variation in cardiomyopathy genes. We compared sequence data from 7,855 individuals who had a clinical diagnosis of cardiomyopathy with 60,706 reference samples from the ExAC consortium, the first data set powered to assess variant alleles present in the population at a range of 1:1,000–100,000 that might previously have been considered pathogenic yet may in fact be too common to cause penetrant Mendelian disease.

Through these analyses, we aimed to define the genes, regions of genes, and/or classes of variants that can be reliably interpreted in a clinical setting and in doing so enhance variant interpretation and increase clinical diagnostic yields.

## MATERIALS AND METHODS

### Clinical and control cohorts

Data from 3,267 individuals with a clinical diagnosis of HCM, 559 with DCM, and 361 with ARVC were obtained from the Oxford Medical Genetics Laboratory for up to 16 genes for HCM, 28 genes for DCM, and 8 genes for ARVC (Supplementary Table S1a online). Data from Partners Laboratory of Molecular Medicine (LMM) were downloaded from the supplemental files of previous publications and included data from up to 18 genes sequenced in 632–2,912 HCM patients<sup>9</sup> and up to 46 genes sequenced in 121–756 DCM patients<sup>3</sup> (Supplementary Table S1b online). Because there were no significant differences in the proportion of cases of rare variants in each gene between the two laboratories (Fisher's exact test), the data were combined (Supplementary Table S3a,b online).

Data were downloaded from ExAC (<http://exac.broadinstitute.org>; version 0.3, January 2015). Only genes with a high proportion of coding regions covered to a median sequence depth of  $>30\times$  and only high-quality (PASS filter) variants were included in our analyses. In addition, we adjusted the total number of ExAC samples per gene based on the mean coverage at the variant sites of interest.

Please refer to Supplementary Note S1 online for more detailed information on each component cohort.

### Defining an allele frequency threshold for rare variation

The single most common confirmed pathogenic variant in both clinical cohorts was *MYBPC3* c.1504C>T (p.Arg502Trp), which

was found in 104/6,179 HCM cases (1.7%; 95 CI: 1.4–2.0%); this variant was only observed three times in ExAC (MAF  $2.5 \times 10^{-5}$ ). We therefore applied a MAF threshold of  $1 \times 10^{-4}$  as a conservative upper bound because variants more frequent than this in the general population would not be expected to be highly penetrant pathogenic mutations (see Supplementary Note S2 online). This MAF does not exclude the possibility of more common deleterious founder variants in specific populations where the genetic architecture of cardiomyopathy is not well defined.

### Calculation of rare variant frequency in cardiomyopathy and ExAC cohorts

For each gene, the frequency of rare variants (MAF  $<1 \times 10^{-4}$ ) in ExAC was calculated by dividing the sum of the adjusted allele count by the mean of the total adjusted alleles. The frequency of rare variation in the cardiomyopathy cohorts was calculated by dividing the sum of rare variants identified in cardiomyopathy cases by the total number of patients analyzed for each gene. Only likely protein-altering variants in designated canonical transcripts (Supplementary Table S2 online) were analyzed: missense, in-frame insertions/deletions, frameshift, nonsense, and variants affecting the splice donor and acceptor regions (first and last two bases of each intron). Analyses were performed for all protein-altering variants and separately for variants predicted to be nontruncating (missense and in-frame insertions and deletions) and truncating (frameshift, nonsense, splice donor/acceptor). To ensure that population-specific variants did not have a confounding effect on this analysis, we compared the results seen in the LMM DCM cohort for all samples and for Caucasians only (see Supplementary Note S4 online).

In total, after data from both clinical laboratories were combined and poorly covered genes in ExAC were excluded, 20 genes sequenced in 632–6,179 HCM patients, 46 genes sequenced in 121–1,315 DCM patients, and 8 genes sequenced in 93–361 ARVC patients were analyzed. See Supplementary Table S2 online for full details of cohort sizes for each gene.

### Comparison of variation between cardiomyopathy and ExAC cohorts

For each gene, the frequency of rare variation in the clinical cohort was compared with that in ExAC. Case excess was defined by subtracting the proportion of individuals in ExAC with a filtered variant from the proportion in the clinical cohort. We made the simplifying assumption that the frequency of rare benign variants was equivalent in cases and ExAC, and that the frequency of pathogenic variants in ExAC is sufficiently low so as not to affect this comparison. A Fisher's exact test was performed to test the significance of observed excess in cases.

For each gene and variant class, we calculated two related metrics: the odds ratio (OR) with 95% confidence intervals and the etiological fraction (EF)<sup>10–12</sup> (calculated as: (OR-1)/OR  $\times 100$ ; for further information on EF please refer to Supplementary Note S3 online). These metrics were calculated

for all protein-altering variants and separately for predicted nontruncating and truncating subsets.

All statistical tests used in these analyses are two-sided, unless otherwise stated, and analysis was undertaken using Stata statistical analysis software (StataCorp. 2007. Stata Statistical Software: Release 10. College Station, TX: StataCorp LP.).

#### Distribution of missense variants in *MYH7*

To identify putative hotspots of pathogenic missense mutations in *MYH7*, distinct rare missense variants from the HCM, DCM, and ExAC cohorts were mapped along the protein sequence. Nonrandom mutation cluster (NMC),<sup>13</sup> implemented in the iPAC Bioconductor R package, was used to identify clusters of variants in each cohort (R source code of NMC algorithm: <https://bioconductor.org/packages/release/bioc/html/iPAC.html>).

#### Analysis of research cardiomyopathy cohorts

Research cardiomyopathy cohorts are defined as published studies from research laboratories where patient samples were subjected to sequencing across panels of cardiac genes. The HCM research cohort<sup>14</sup> sequenced 874 patients across 35 genes (12 associated primarily with HCM, 7 with DCM, 7 with ARVC, and 9 with arrhythmias, as stated by Lopes *et al.*). The DCM research cohort<sup>15–17</sup> comprised 312–324 patients sequenced for 12 confirmed and putative DCM genes (not including *TTN*). Putative pathogenic variants in these studies are not identified by clinical-grade classification but rather by criteria such as variant type, population frequency, and in silico algorithm prediction—the details of the criteria used in each study are described in the published articles.

Rare variant frequencies and case excess were calculated for each gene as described for the clinical cohorts. The number of variants reported to be putatively pathogenic for each gene in these studies was compared with the number predicted to be pathogenic based on the case excess observed in these cohorts.

#### HGMD cardiomyopathy mutations in ExAC

Variants in the Human Genome Mutation Database (HGMD; professional version 2015.1) associated with HCM, DCM, or ARVC (“disease-causing mutations” with a HGMD tag of DM and DM?) were identified based on manual curation of the HGMD disease terms. The total allele frequency and count from ExAC were extracted for each variant. Polymorphisms (ExAC MAF >1×10<sup>-2</sup>) were removed from the analysis. The number of HGMD variants present in ExAC was calculated at any frequency and with MAF >1×10<sup>-4</sup>. The total number of ExAC alleles and the total number of ExAC individuals with HGMD-associated cardiomyopathy variants were also calculated for each disease. Additionally, the ExAC frequencies of HGMD cardiomyopathy variants previously observed only once in the Exome Sequencing Project were analyzed to assess how the enhanced resolution of ExAC can clarify previously uninterpretable variants.

## RESULTS

### Comparison of rare variation between cardiomyopathy cohorts and ExAC

Variants identified by sequencing of putative cardiomyopathy genes in cases ( $n = 7,855$ ) were collated by disease and gene (**Supplementary Table S1a,b** online). We compared the burden of rare protein-altering variants (ExAC MAF <1×10<sup>-4</sup>) detected in 20 HCM genes, 48 DCM genes, and 8 ARVC genes in HCM, DCM, and ARVC cases, respectively, with the burden observed in ExAC. Predicted truncating and nontruncating variants were analyzed separately (**Table 1** and **Supplementary Table S4a–c** online).

As expected,<sup>18–20</sup> rare variation in the two major HCM genes accounted for the majority of variation in HCM cases (*MYBPC3*, 19.0% of cases; *MYH7*, 14.2%). Rare variants were less numerous in other well-characterized HCM genes (*TNNI3*, *TNNT2*, *TPM1*, *MYL2*, *MYL3*, *ACTC1*, *PLN*) and phenocopy genes (*GLA*, *LAMP2*, *PRKAG2*) ( $\leq 2\%$  cases per gene). For each of these genes there is a significant ( $P < 0.05$  after Bonferroni correction) excess of variation in cases as compared with ExAC, thus confirming their association with disease (**Figure 1** and **Table 1** and **Supplementary Table S5a** online). However, for several more recently reported HCM genes (*TNNC1*<sup>21</sup>, *MYOZ2*<sup>22</sup>, *ACTN2*<sup>23</sup>, *ANKRD1*<sup>24</sup>) there was no significant excess of rare genetic variation in these HCM cases.

DCM is highly genetically heterogeneous, with up to ~60 implicated genes.<sup>20,25,26</sup> In the clinical cohorts, truncating variants in *TTN* were most common (14.6%), in accordance with our findings in large research cohorts.<sup>27,28</sup> The prevalence of rare variants in other well-characterized DCM genes was modest (*MYH7*, 5.3%; *LMNA*, 4.4%; *TNNT2*, 2.9%; and *TPM1*, 1.9%) but significantly enriched compared with ExAC (**Figure 1** and **Table 1** and **Supplementary Table S5b** online). However, with the exception of truncating variants in *DSP* (2.8%), there was limited burden and modest or no significant excess variation in the remaining 40 genes tested. In ARVC, the five major genes each showed significant excess in cases (**Table 1** and **Supplementary Table S5c** online).

Overall, the yield of pathogenic (P) and likely pathogenic (LP) variants was 32% for HCM, 13% for DCM (but note that *TTN* was only sequenced in one-third of samples), and 36% for ARVC. Of note, in the genes robustly supported by an excess of pathogenic and likely pathogenic variants, even variants of uncertain significance were seen in excess over ExAC, suggesting that clinical laboratories may be overly conservative (**Figure 1**).

### Interpretation of variation by gene and variant class

Many variants in confirmed disease genes can be interpreted with confidence based on cumulative experience (e.g., multiple occurrences of segregation in families, de novo mutations, founder variants) and/or functional insights (e.g., null alleles in haploinsufficient genes). However, our ability to evaluate the pathogenicity of novel variants depends on the signal-to-noise ratio. For each gene and variant class, we calculated two related

**Table 1** Summary of the study results, including the number of cardiomyopathy cases and ExAC reference samples analyzed, case excess, OR, and EF for nontruncating and truncating variants

Gene	Cases	ExAC samples	Nontruncating variants			Truncating variants		
			Case excess	OR	EF	Case excess	OR	EF
Hypertrophic cardiomyopathy								
<i>ACTC1</i>	4185	60198	<b>0.00464</b>	<b>8.59 (5.06–14.5)</b>	<b>0.88 (0.79–0.93)</b>	-0.00004	2.88 (0.14–59.9)	–
<i>ACTN2</i>	1439	60388	-0.00451	0.58 (0.30–1.12)	–	0.00059	7.00 (0.84–58.1)	0.86 (0.00–0.98)
<i>ANKRD1</i>	807	60300	0.00068	1.23 (0.39–3.84)	0.19 (0.00–0.73)	0.00072	2.41 (0.33–17.6)	0.59 (0.00–0.93)
<i>CSRP3</i>	2167	60647	0.00081	1.28 (0.63–2.60)	0.22 (0.00–0.61)	0.00102	3.82 (1.14–12.7)	0.74 (0.00–0.92)
<i>FHL1</i>	1535	60268	<b>0.00724</b>	<b>6.95 (3.85–12.5)</b>	<b>0.86 (0.72–0.92)</b>	0.00065	–	–
<i>GLA</i>	3700	59404	<b>0.00630</b>	<b>7.39 (4.68–11.6)</b>	<b>0.86 (0.78–0.92)</b>	0.00000	–	–
<i>LAMP2</i>	3290	59987	0.00045	1.23 (0.60–2.51)	0.19 (0.00–0.60)	0.00395	–	–
<i>MYBPC3</i>	6179	47155	<b>0.07988</b>	<b>5.70 (5.12–6.34)</b>	<b>0.82 (0.80–0.84)</b>	<b>0.09074</b>	<b>118.7 (86.0–163.8)</b>	<b>0.99 (0.99–0.99)</b>
<i>MYH7</i>	6112	60471	<b>0.12786</b>	<b>12.0 (10.8–13.3)</b>	<b>0.92 (0.91–0.92)</b>	0.00034	1.71 (0.66–4.41)	0.42 (0.00–0.78)
<i>MYL2</i>	4185	60516	<b>0.00873</b>	<b>6.74 (4.69–9.70)</b>	<b>0.85 (0.78–0.90)</b>	0.00048	3.10 (0.89–10.7)	0.68 (0.00–0.91)
<i>MYL3</i>	4185	60601	<b>0.00706</b>	<b>5.00 (3.43–7.27)</b>	<b>0.80 (0.70–0.86)</b>	-0.00004	2.90 (0.14–60.3)	–
<i>MYOZ2</i>	632	60436	-0.00078	0.67 (0.09–4.78)	–	-0.00020	3.82 (0.23–64.6)	–
<i>NEXN</i>	632	59277	0.00416	1.61 (0.76–3.41)	0.38 (0.00–0.70)	0.00094	2.47 (0.34–18.0)	0.60 (0.00–0.93)
<i>PLN</i>	2167	60475	0.00096	3.35 (1.01–11.1)	0.70 (0.00–0.91)	<b>0.00179</b>	<b>27.9 (6.99–111.8)</b>	<b>0.96 (0.81–0.99)</b>
<i>PRKAG2</i>	3973	57747	<b>0.00487</b>	<b>1.95 (1.40–2.71)</b>	<b>0.49 (0.26–0.63)</b>	0.00038	4.15 (0.86–20.0)	0.76 (0.00–0.95)
<i>TNNC1</i>	632	59156	-0.00054	1.40 (0.09–22.8)	–	-0.00004	18.7 (0.90–390.0)	–
<i>TNNI3</i>	6047	53856	<b>0.01947</b>	<b>11.6 (8.93–15.0)</b>	<b>0.91 (0.89–0.93)</b>	0.00057	2.43 (0.99–6.00)	0.59 (0.00–0.84)
<i>TNNT2</i>	6103	56924	<b>0.01556</b>	<b>8.46 (6.52–10.9)</b>	<b>0.88 (0.84–0.91)</b>	<b>0.00150</b>	<b>6.04 (2.83–12.9)</b>	<b>0.83 (0.61–0.93)</b>
<i>TPM1</i>	4447	58727	<b>0.01400</b>	<b>18.0 (12.4–26.1)</b>	<b>0.94 (0.92–0.96)</b>	-0.00004	2.64 (0.13–55.0)	–
<i>TTR</i>	632	60635	0.00180	2.34 (0.58–9.55)	0.57 (0.00–0.89)	-0.00004	13.6 (0.71–265.4)	–
Dilated cardiomyopathy								
<i>ABCC9</i>	590	60453	0.00386	1.49 (0.70–3.15)	0.33 (0.00–0.68)	0.00095	2.33 (0.32–16.9)	0.57 (0.00–0.94)
<i>ACTC1</i>	1103	60198	0.00301	5.92 (2.11–16.6)	0.83 (0.53–0.94)	0.00087	27.3 (2.47–301.4)	0.96 (0.60–1.00)
<i>ACTN2</i>	895	60388	-0.00182	0.83 (0.41–1.67)	–	-0.00010	5.19 (0.29–92.1)	–
<i>ANKRD1</i>	426	60300	-0.00304	0.38 (0.02–6.17)	–	-0.00052	2.24 (0.14–36.7)	–
<i>CASQ2</i>	121	59954	0.00364	1.80 (0.25–12.9)	0.44 (0.00–0.92)	0.00780	18.5 (2.49–137.2)	0.95 (0.60–0.99)
<i>CAV3</i>	121	60262	-0.00124	3.28 (0.20–53.2)	–	-0.00014	29.1 (1.67–508.2)	–
<i>CRYAB</i>	425	50492	0.00273	2.38 (0.59–9.69)	0.58 (0.00–0.90)	-0.00010	10.7 (0.60–195.3)	–
<i>CSRP3</i>	945	60647	0.00029	1.10 (0.35–3.45)	0.09 (0.00–0.71)	0.00070	2.92 (0.39–21.6)	0.66 (0.00–0.95)
<i>DES</i>	894	51128	0.00347	1.79 (0.84–3.80)	0.44 (0.00–0.74)	0.00188	7.16 (1.64–31.2)	0.86 (0.39–0.97)
<i>DSC2</i>	427	60064	-0.00470	0.50 (0.12–2.00)	–	-0.00026	4.26 (0.25–71.0)	–
<i>DSG2</i>	427	59667	-0.00523	0.57 (0.18–1.78)	–	-0.00072	1.60 (0.10–26.0)	–
<i>DSP</i>	427	59620	0.01606	1.55 (0.99–2.43)	0.35 (0.00–0.59)	<b>0.02740</b>	<b>41.0 (21.4–78.4)</b>	<b>0.98 (0.95–0.99)</b>
<i>DTNA</i>	121	60219	0.00400	1.94 (0.27–13.9)	0.49 (0.00–0.93)	-0.00010	38.1 (2.14–680.4)	–
<i>EMD</i>	590	59063	-0.00061	1.02 (0.14–7.34)	–	0.00000	–	–
<i>FHL1</i>	355	60268	-0.00170	1.14 (0.07–18.3)	–	0.00000	–	–
<i>FHL2</i>	425	59228	-0.00065	0.78 (0.11–5.60)	–	-0.00010	10.7 (0.60–190.3)	–
<i>GLA</i>	541	59404	-0.00138	0.92 (0.06–14.9)	–	0.00000	–	–

Genes with a significant excess in cases over ExAC reference samples are shown in bold type. See **Supplementary Tables 1–14** for full details.

OR and EF with 95% confidence intervals.

EF, etiological fraction; ExAC, Exome Aggregation Consortium; OR, odds ratio.

**Table 1** Continued on next page

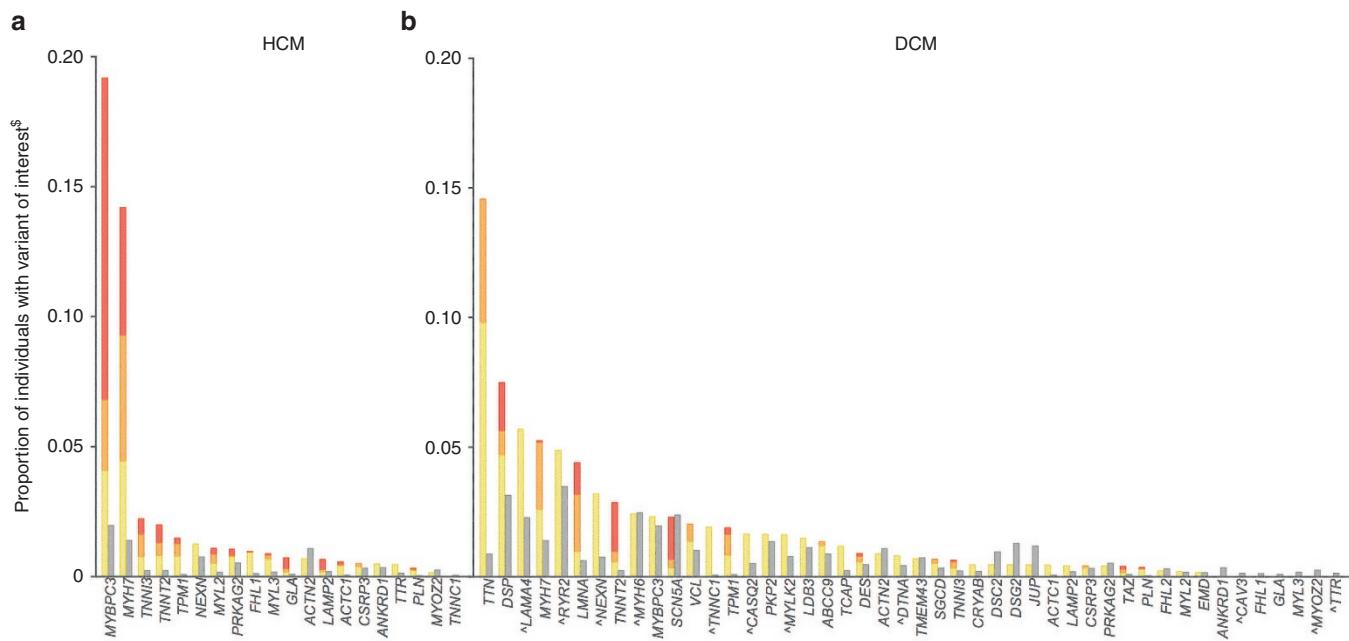
Table 1 Continued

Gene	Cases	ExAC samples	Nontruncating variants			Truncating variants		
			Case excess	OR	EF	Case excess	OR	EF
<i>JUP</i>	425	55896	-0.00709	0.40 (0.10–1.59)	–	-0.00018	6.25 (0.37–106.9)	–
<i>LAMA4</i>	121	59741	0.02779	2.34 (1.03–5.33)	0.57 (0.03–0.81)	-0.00102	4.06 (0.25–66.0)	–
<i>LAMP2</i>	532	59987	0.00102	1.90 (0.47–7.70)	0.47 (0.00–0.87)	0.00000	–	–
<i>LDB3</i>	740	58166	0.00412	1.39 (0.76–2.54)	0.28 (0.00–0.61)	-0.00048	1.33 (0.08–21.8)	–
<i>LMNA</i>	1044	51266	<b>0.02264</b>	<b>4.83 (3.31–7.06)</b>	<b>0.79 (0.70–0.86)</b>	<b>0.01519</b>	<b>99.7 (42.5–233.5)</b>	<b>0.99 (0.98–1.00)</b>
<i>MYBPC3</i>	1161	47155	0.00614	1.33 (0.92–1.94)	0.25 (0.00–0.48)	-0.00086	0.50 (0.03–8.15)	–
<i>MYH6</i>	121	60328	0.00143	1.06 (0.34–3.34)	0.06 (0.00–0.70)	-0.00128	3.24 (0.20–52.5)	–
<i>MYH7</i>	1315	60471	<b>0.03821</b>	<b>3.99 (3.09–5.14)</b>	<b>0.75 (0.68–0.81)</b>	0.00028	1.59 (0.22–11.6)	0.37 (0.00–0.91)
<i>MYL2</i>	543	60516	0.00030	1.20 (0.17–8.62)	0.17 (0.00–0.88)	-0.00024	3.84 (0.23–64.4)	–
<i>MYL3</i>	543	60601	-0.00178	0.51 (0.03–8.26)	–	-0.00004	22.3 (1.07–465.0)	–
<i>MYLK2</i>	121	57226	0.00885	2.17 (0.53–8.80)	0.54 (0.00–0.89)	-0.00014	27.7 (1.59–482.6)	–
<i>MYOZ2</i>	121	60436	-0.00236	1.73 (0.11–27.9)	–	-0.00020	19.8 (1.17–337.8)	–
<i>NEXN</i>	156	59277	0.01872	3.78 (1.39–10.2)	0.74 (0.28–0.90)	0.00577	10.0 (1.37–73.7)	0.90 (0.27–0.99)
<i>PKP2</i>	427	58406	0.00359	1.28 (0.60–2.71)	0.22 (0.00–0.63)	-0.00076	1.57 (0.10–25.5)	–
<i>PLN</i>	1095	60475	0.00232	6.64 (2.00–22.0)	0.85 (0.50–0.95)	0.00085	13.8 (1.54–123.7)	0.93 (0.35–0.99)
<i>PRKAG2</i>	546	57747	-0.00154	0.70 (0.17–2.84)	–	-0.00012	7.04 (0.40–123.4)	–
<i>RYR2</i>	121	53692	0.01589	1.49 (0.66–3.40)	0.33 (0.00–0.71)	-0.00114	3.71 (0.23–60.3)	–
<i>SCN5A</i>	304	56420	-0.01002	0.56 (0.21–1.51)	–	<b>0.00923</b>	<b>16.5 (5.05–54.1)</b>	<b>0.94 (0.80–0.98)</b>
<i>SGCD</i>	590	50725	0.00194	1.61 (0.51–5.04)	0.38 (0.00–0.80)	0.00147	8.61 (1.10–67.3)	0.88 (0.09–0.99)
<i>TAZ</i>	740	51321	0.00144	2.96 (0.72–12.1)	0.66 (0.00–0.92)	0.00133	69.4 (4.34–1111.3)	0.99 (0.77–1.00)
<i>TCAP</i>	590	55233	<b>0.00968</b>	<b>5.51 (2.56–11.8)</b>	<b>0.82 (0.61–0.92)</b>	-0.00024	3.22 (0.19–54.1)	–
<i>TMEM43</i>	427	60220	0.00013	1.02 (0.33–3.19)	0.02 (0.00–0.69)	-0.00040	2.87 (0.17–47.3)	–
<i>TNNC1</i>	156	59156	<b>0.01869</b>	<b>35.1 (10.6–115.7)</b>	<b>0.97 (0.91–0.99)</b>	-0.00004	75.6 (3.61–1581.1)	–
<i>TNNI3</i>	1239	53856	0.00460	3.46 (1.68–7.12)	0.71 (0.40–0.86)	-0.00042	0.97 (0.06–15.9)	–
<i>TNNT2</i>	1254	56924	<b>0.02657</b>	<b>13.8 (9.52–20.2)</b>	<b>0.93 (0.89–0.95)</b>	-0.00030	1.30 (0.08–21.5)	–
<i>TPM1</i>	1111	58727	<b>0.01626</b>	<b>20.8 (12.2–35.5)</b>	<b>0.95 (0.92–0.97)</b>	0.00176	52.9 (7.45–376.2)	0.98 (0.87–1.00)
<i>TTN</i>	460	58613	–	–	–	<b>0.13689</b>	<b>19.7 (15.0–26.0)</b>	<b>0.95 (0.93–0.96)</b>
<i>TTR</i>	121	60635	-0.00136	3.02 (0.19–48.9)	–	-0.00004	71.2 (3.66–1387.6)	–
<i>VCL</i>	590	59393	0.00541	1.56 (0.80–3.02)	0.36 (0.00–0.67)	<b>0.00646</b>	<b>21.3 (7.23–62.8)</b>	<b>0.95 (0.86–0.98)</b>
Arrhythmogenic right ventricular cardiomyopathy								
<i>DES</i>	93	51128	0.02790	7.54 (2.37–24.0)	0.87 (0.34–0.96)	-0.00036	16.5 (0.99–278.1)	–
<i>DSC2</i>	351	60064	0.01056	2.15 (1.01–4.57)	0.53 (0.00–0.78)	<b>0.00544</b>	<b>21.5 (4.93–93.8)</b>	<b>0.95 (0.58–0.99)</b>
<i>DSG2</i>	354	59667	<b>0.02164</b>	<b>2.83 (1.58–5.05)</b>	<b>0.65 (0.31–0.80)</b>	<b>0.01340</b>	<b>19.8 (7.82–50.4)</b>	<b>0.95 (0.84–0.98)</b>
<i>DSP</i>	352	59620	<b>0.03172</b>	<b>2.10 (1.36–3.24)</b>	<b>0.52 (0.23–0.69)</b>	<b>0.05896</b>	<b>89.9 (52.7–153.6)</b>	<b>0.99 (0.98–0.99)</b>
<i>JUP</i>	94	55896	<b>0.07331</b>	<b>7.80 (3.76–16.1)</b>	<b>0.87 (0.69–0.94)</b>	-0.00018	28.1 (1.64–484.0)	–
<i>LMNA</i>	93	51266	-0.00608	0.87 (0.05–14.0)	–	-0.00014	32.2 (1.85–562.7)	–
<i>PKP2</i>	361	58406	0.00382	1.30 (0.58–2.92)	0.23 (0.00–0.65)	<b>0.26240</b>	<b>484.7 (331.5–708.7)</b>	<b>1.00 (1.00–1.00)</b>
<i>TMEM43</i>	94	60220	-0.00690	0.76 (0.05–12.2)	–	-0.00040	13.0 (0.78–215.3)	–

Genes with a significant excess in cases over ExAC reference samples are shown in bold type. See **Supplementary Tables 1–14** for full details.

OR and EF with 95% confidence intervals.

EF, etiological fraction; ExAC, Exome Aggregation Consortium; OR, odds ratio.



**Figure 1 Proportion of individuals with rare variants in hypertrophic cardiomyopathy (HCM) and dilated cardiomyopathy (DCM) in combined clinical cohorts (data from Oxford Medical Genetics Laboratory and Laboratory of Molecular Medicine) compared with Exome Aggregation Consortium (ExAC) data (gray columns). Variants counted were single-nucleotide changes or small insertion/deletion variants detected in the coding region  $\pm 2$  bp, with an ExAC minor allele frequency of  $<1 \times 10^{-4}$  (see Methods). Clinical cohorts: HCM,  $n = 632$  to 6,179 and DCM,  $n = 121$  to 1,315 (see Table 1, Supplementary Table S5a,b online). Information on reported pathogenicity class (red = pathogenic (P), orange = likely pathogenic (LP), yellow = variant of uncertain significance (VUS)) is overlaid. See Supplementary Tables S1, S4a,b online for full details.  $\wedge$  = genes analyzed in fewer than 200 cases. ExAC:  $n =$  mean of total adjusted allele count for rare variant carriers. For HCM genes,  $n$  ranged from 47,153 to 60,647; for DCM genes,  $n$  was 42,697 to 60,647 (see Supplementary Table S5a,b online). CTF1 and RBM20 were removed from analysis due to poor coverage in ExAC.**

metrics: the odds ratio (OR) (ratio of odds of cardiomyopathy comparing rare variant carriers with noncarriers) and the etiological fraction (EF), which is a commonly used measure in epidemiology<sup>10–12</sup> that estimates the proportion of cases in which the exposure (in this case, a rare variant in a gene) was causal (see Supplementary Note S3 online).

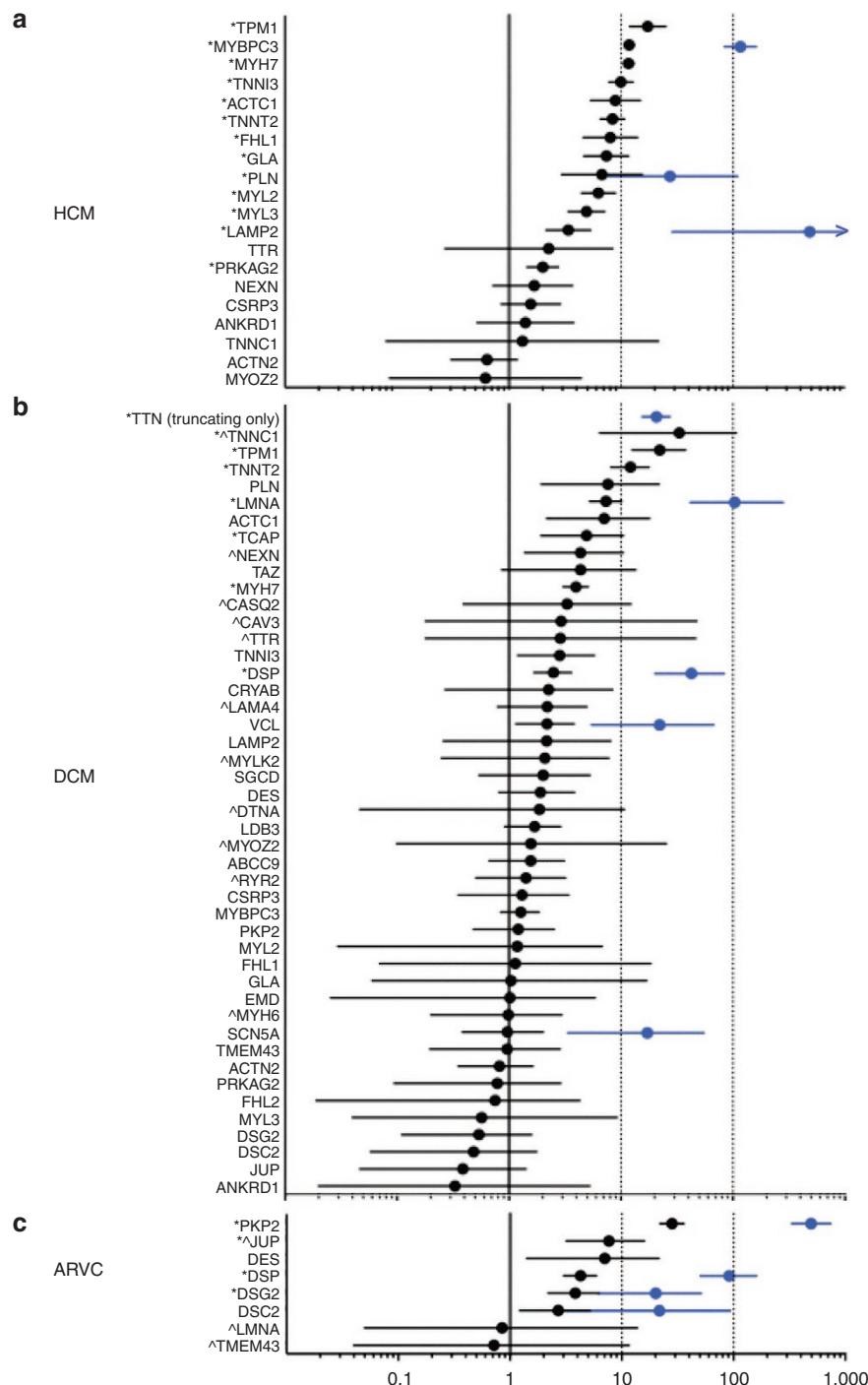
These analyses reaffirm high ORs and EFs in key cardiomyopathy genes and also highlight a number of previously reported cardiomyopathy genes that show limited disease association when compared with a very large number of reference samples (Figure 2 and Table 1 and Supplementary Table S5a–c online). As expected, many genes have divergent results for truncating as compared with nontruncating variants; for example, *MYH7* has an OR of 1 (0.5–4.5) for truncating variants versus an OR of 12 (10.9–13.3) for nontruncating variants in HCM cases.

This observation confirms the widely accepted view that missense alleles of *MYH7* act as dominant negatives in HCM whereas truncating variants are not pathogenic. In genes whose truncating alleles are disease-causing, ORs are typically higher owing to the lower rate of truncating variants in the population. As expected, truncating variation in *MYBPC3* associates strongly with HCM (the result of haploinsufficiency<sup>29</sup>), but neither truncating nor nontruncating variants in *MYBPC3* show a significant association with DCM (OR = 1.3 (0.8–1.8); EF = 0.21 (0–0.46)), which is a finding that fits with mechanistic insights but challenges some widely held viewpoints.<sup>15,30</sup> Among the

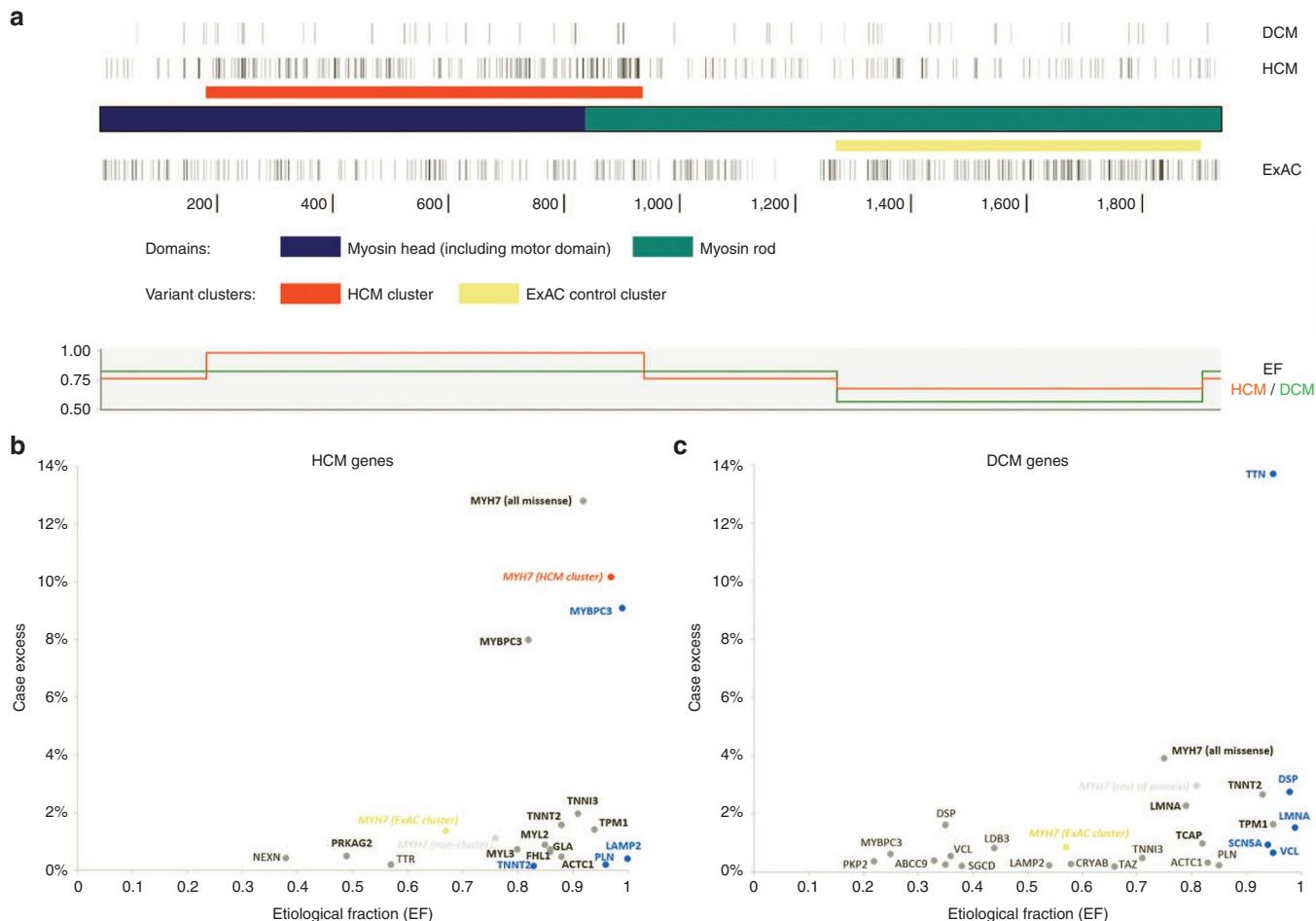
ARVC genes, truncating alleles are informative for four major genes (and particularly common for *PKP2* and *DSP*), whereas nontruncating variants in these genes are difficult to interpret reliably (Figure 2 and Table 1 and Supplementary Tables S4c and S5c online).

#### Using protein domain knowledge to improve variant interpretation

At the gene level, ORs for nontruncating variants are typically modest; in the absence of prior clinical experience or functional data, interpretation is often uncertain. This may be improved by considering protein topology because pathogenic variants often cluster in specific regions in cases.<sup>31,32</sup> We evaluated the distribution of rare missense alleles in *MYH7*, which encodes a protein with well-characterized functional and structural domains to determine whether systematic analysis of variant distribution refines interpretation. Nonrandom mutation cluster analysis<sup>13</sup> revealed a significant cluster ( $P < 3 \times 10^{-15}$ ; false discovery rate  $q < 5 \times 10^{-13}$ ) between residues 181 and 937 in HCM cases, whereas in ExAC variants were depleted in this region and instead clustered between residues 1,271 and 1,903 ( $P < 3 \times 10^{-8}$ ; false discovery rate  $q < 4 \times 10^{-5}$ ) (Figure 3a). These data more precisely define the boundaries of mutation-enriched and depleted zones that can be used to generate more discriminating EFs; for example, for rare variants in HCM patients,



**Figure 2** Odds ratios (OR) with 95% confidence interval (CI) for each gene tested in the hypertrophic cardiomyopathy (HCM) ( $n = 632$  to  $6,179$ ), dilated cardiomyopathy (DCM) ( $n = 121$  to  $131$ ), and arrhythmic right ventricular cardiomyopathy (ARVC) ( $n = 93$  to  $361$ ) clinical cohorts compared with Exome Aggregation Consortium (ExAC) reference samples ( $n = \text{mean of total adjusted allele count for rare variant carriers}$  for HCM genes,  $n = 47,153$  to  $60,647$ ; for DCM genes,  $n = 42,697$  to  $60,647$ ; and for ARVC genes,  $n = 51,126$  to  $60,218$ ). See **Supplementary Table S5a–c** online for data used to generate this plot. Data have been plotted (log<sub>10</sub> scale) for all protein-altering variants (black) and truncating variants (blue). For truncating variants, OR with 95% CI have been plotted for genes where a statistically significant difference was observed for this variant type on FET. \*Statistically significant Fisher's exact test (FET) ( $P = 0.05$  with Bonferroni correction; HCM  $P \leq 0.0025$ ; DCM  $P \leq 0.001$ ; and ARVC  $P \leq 0.006$ ). ^Genes analyzed in fewer than 200 cases. CTF1 and RBM20 were removed from analysis owing to poor coverage in ExAC.



**Figure 3 Variant interpretation.** (a) Distribution of rare (Exome Aggregation Consortium (ExAC)) minor allele frequency  $<1 \times 10^{-4}$ ) MYH7 missense variants in hypertrophic cardiomyopathy (HCM) ( $n = 864$ ) and dilated cardiomyopathy (DCM) ( $n = 69$ ) clinical cohorts and ExAC ( $n = 816$ ), with the key myosin protein regions highlighted. Using nonrandom mutation cluster analysis (NMC), (refer to **Supplementary Methods** online), a significant variant cluster ( $P < 3 \times 10^{-15}$ , false discovery rate (FDR),  $q < 5 \times 10^{-13}$ ) was identified between residues 181 and 937 (involving the motor domain, lever arm, and part of the rod) in HCM cases, and depletion in this region and a significant cluster ( $P < 3 \times 10^{-8}$ , FDR  $q < 4 \times 10^{-5}$ ) was identified between residues 1,271 and 1,903 (in the part of the rod that forms the filament backbone) in ExAC samples. The etiological fraction (EF) for a rare MYH7 missense variant identified in a HCM proband ranges from 0.97 in the HCM cluster to 0.67 in the control cluster. Vertical gray bars depict the positions of variants in cohorts; grayscale shows variant density where variants are coincident. An overview of the genetic landscape of HCM (b) and DCM (c) for truncating (blue) and nontruncating (gray) variants, as well as MYH7 missense variants in the clusters identified in (a) (orange, disease cluster; yellow, ExAC control cluster). The case excess (y-axis) is the frequency of rare variation in disease cohorts over and above the frequency in ExAC and indicates the relative importance of the gene and variant class to the genetic etiology of each cardiomyopathy. The etiological fraction (EF) (x-axis) is an estimate of the proportion of affected carriers where the variant caused the disease; it is a measure of the interpretability of variants of this class (see **Supplementary Table S4a,b** online for full details). This measure is an average of all variants of a given class; some of which will be pathogenic but others will be benign.

EFs range from 0.97 in the HCM cluster to 0.67 in the control cluster (Figure 3a,b).

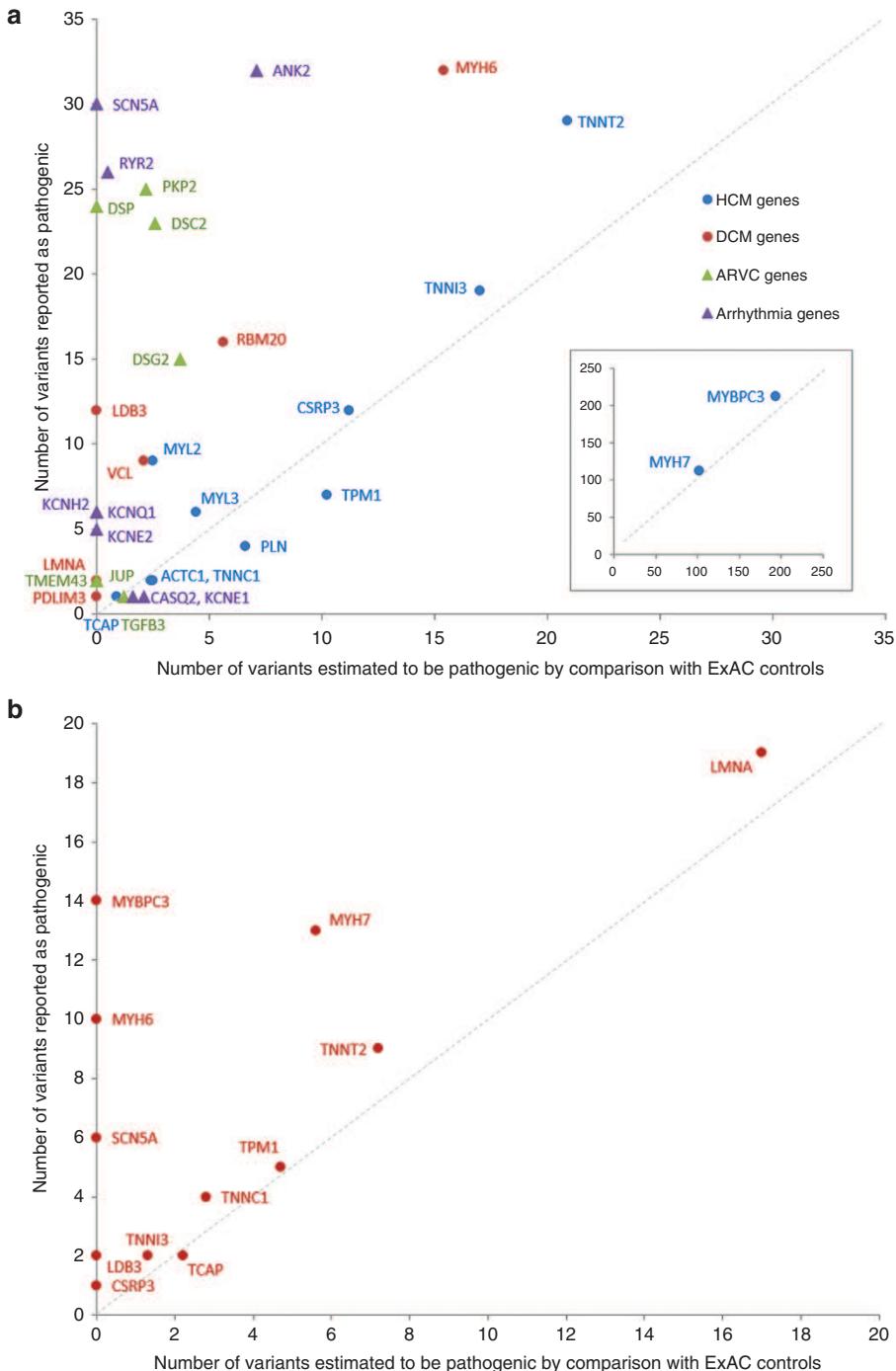
### Application of findings

To facilitate the application of these findings for research and clinical use, we provide an overview of the genetic landscape of cardiomyopathy as represented by patient referrals received by UK and US clinical testing laboratories. This shows the relative importance of cardiomyopathy genes within these patient populations (measured as a “case excess”) and their interpretability (expressed as the etiological fraction) (Figure 3b,c). Furthermore, we have created a Web resource, Atlas of Cardiac Genetic Variation (<http://cardiobd.org/ACGV>), to aid those

assessing the relevance of specific genes and classes of variants to cardiomyopathies.

### Reassessing extended gene panel cardiomyopathy research studies

Several research studies<sup>14,15,17,30</sup> using extended gene panels have reported genetic overlap between diverse cardiac diseases that appear at odds with known disease mechanisms. We surmise that many such studies have not adequately accounted for background genetic variation, have relied on variant data from incompletely annotated disease-centered databases, and have not used segregation. Here, using ExAC data, we present a reanalysis of two representative research studies.<sup>14–17</sup>



**Figure 4 Comparison of the number of variants reported as putatively pathogenic in research studies.** Hypertrophic cardiomyopathy (HCM) (a) and dilated cardiomyopathy (DCM) (b) in research studies (using generic analysis criteria such as variant class, missense-effect predictions, and variant population frequency in the Exome Sequencing Project) with those predicted as pathogenic by the excess of variation in cases over Exome Aggregation Consortium (ExAC) in each gene. For the HCM study<sup>14</sup> (a), genes are colored according to the cardiac disease for which they are primarily associated, as defined by Lopes *et al*.<sup>14</sup> Although there is good concordance between the research findings and the ExAC predictions for established HCM genes, for genes primarily associated with DCM, ARVC, and arrhythmias, the variation in cases is similar to that in ExAC. In the DCM study<sup>15–17</sup> (b), variation burden in MYBPC3, SCN5A, and MYH6 is similar between the published research cases and ExAC, suggesting that most variants in these genes are unlikely to cause DCM.

In the research HCM cases, the excess variation in the known HCM genes (e.g., *MYBPC3*, *MYH7*) is substantial. By contrast, the measured variation in DCM, ARVC, and ion channel genes in the HCM patients—although also substantial—is similar to

that seen in ExAC with little, if any, excess burden (Figure 4a and Supplementary Table S6 online). This suggests that the majority of these variants, although individually rare, are benign bystanders, and that any overlap between the disorders has been overestimated.

In the DCM research studies, some genes proposed on the basis of these and other recent studies as among the most common causes of DCM (e.g., *MYBPC3*, *MYH6*, and *SCN5A*<sup>15,17,30</sup>) in fact showed no excess variation (Figure 4b and Supplementary Table S7 online).

#### Reassessment of variants previously reported as pathogenic

We examined the ExAC frequency of variants previously reported to cause HCM, DCM, or ARVC, as catalogued by HGMD (Supplementary Tables S8–S10 online). A substantial number of purported disease-causing variants for HCM (25.2%; 322/1,280), DCM (29.2%; 222/759), and ARVC (34.6%, 167/483) were observed in ExAC. Although presence in ExAC does not preclude pathogenicity, a significant number are present at a frequency incompatible with causation of penetrant cardiomyopathy (6.5% of HCM, 11.9% of DCM, and 13.5% of ARVC variants are present at MAF  $>1 \times 10^{-4}$ ; Supplementary Tables S11 and S12 online). Of HGMD variants that could not be excluded as disease-causing using the Exome Sequencing Project (the largest control data set prior to ExAC) due to an allele count of one, 75% can now be discounted by ExAC refinement. In total, 11.7, 19.6, and 20.1% of individuals in ExAC have reported HCM, DCM, and ARVC variants, respectively; this is far in excess of disease prevalence. Hence, variant prioritization based on HGMD status alone is not advised for cardiomyopathy genes, a fact that is increasingly apparent with larger control data sets.

## DISCUSSION

We present an analysis of data from 7,855 individuals referred for clinical diagnostic testing for inherited cardiomyopathies, along with 60,706 ExAC reference samples. These data exemplify the many challenges of variant interpretation in genetically heterogeneous disorders. We propose that in the absence of a large matched case-control series, the approaches described here, using data from large patient cohorts and broader reference data sets such as ExAC, may be applied to a range of multigenic, multiallelic diseases.

We show that the pathogenicity of disease genes originally identified through family linkage is resoundingly validated, for example, the majority of sarcomere genes in HCM. However, genes implicated in cardiomyopathy through candidate gene studies, including genes on panel tests in routine clinical use, are often not convincingly associated with disease. For example, *MYBPC3*, *MYH6*, and *SCN5A* have all been reported to be major contributors to DCM<sup>15,17,30</sup> but show little or no excess burden despite adequate numbers and power; instead, we see that these are in fact genes that have the highest background variation.

We also show that it is crucial not only to distinguish variant classes but also to assess these in light of known disease mechanisms for each gene and disorder. For example, cardiomyopathy-causing variant in most myofilament proteins incorporate into the sarcomere and act as dominant negatives

(HCM mutations are activating, whereas DCM mutations decrease myofibrillar function).<sup>33</sup> Hence, protein-truncating variants that do not incorporate would not be expected to cause these conditions, and this is borne out in our data. By contrast, *MYBPC3* truncation alleles cause HCM through haploinsufficiency, making it unlikely that they could also cause DCM, which we confirm with our findings.

We summarize our analyses of cardiomyopathy genes in two measures, capturing the contribution of each gene to a disease (case excess) and our ability to interpret variation in each gene (etiological fraction (EF)). EF can be interpreted as the proportion of affected carriers in which the variant caused the disease (i.e., the proportion of true positives). EF is based on pooled rare variant frequency data, so it summarizes the average risk across many variants in a gene (some of which will be pathogenic but others will be benign) and will be particularly useful for selecting panels of genes that are informative for discrete phenotypes. Of critical importance, the probability that a novel variant is pathogenic depends on the clinical status of the individual carrying the variant and will be considerably lower in individuals with a remote/unrelated clinical diagnosis. This will be even more problematic with incidental findings during exome or genome sequencing with major implications for the recommendations to return apparently actionable findings.<sup>8</sup>

Although detailed phenotyping of the clinical cases was not available, we are confident that the clinical diagnoses are robust because the current clinical practice is to test only individuals with a confirmed diagnosis.<sup>19,20</sup> The proportion of cases with inherited cardiomyopathy is unknown because evidence of familial disease is not a requirement for testing. The clinical sensitivity (proportion of patients with a pathogenic variant) in our case series was lower than that of previous surveys, which may reflect more restricted testing of stringently selected cases, typically from multiply affected families with severe disease.<sup>34,35</sup> However, the cohorts studied are representative of those encountered by clinical diagnostic laboratories, rather than a highly selected subset.

Despite high levels of confidence in interpreting many well-characterized variants (which may give ORs in the hundreds), diagnostic laboratories are understandably cautious when interpreting a variant that has not been seen before. Our analyses demonstrate that for many genes even variants currently reported with variants of uncertain significance show a several-fold excess over the background in ExAC (Figure 1). More refined interpretation of variants in validated genes, for example, leveraging domain information, could increase the diagnostic yield of genetic testing and is likely to lead to much more substantial gains than the expansion of gene panels.

In contrast to the conservative strategy of clinical laboratories, research studies often report large “yields.” Some may not adequately control for the background rate of rare variation or may include genes for other conditions; as a result, genes nominated as important contributors to disease have little if any excess variation in cases. Testing of broad gene panels and overly inclusive interpretation of variants may lead to erroneous

conclusions about pleiotropic effects of genetic variation<sup>14,30</sup> and overestimates of double/compound mutations<sup>36</sup> and the population prevalence of the disease.<sup>37</sup>

Despite the absence of demonstrable excess of rare variation in a gene, specific variants identified in family studies may still be disease-causing. However, if such variants are a small minority of rare variants in cases, then testing will yield more false positives than true positives. For some of the genes that show no excess (Figure 1), the original reports did not include any variant with robust evidence of segregation (i.e., LOD >3), and the possibility exists that the reported disease association is entirely spurious. An argument is often made that variants could be contributing as modifiers.<sup>38,39</sup> This remains possible; however, in the absence of any significant overrepresentation in cases, the more parsimonious interpretation is that they are phenotypically silent. We have not tested more common variants (MAF >1 × 10<sup>-4</sup>) that could be mechanistically informative but are likely to have smaller effects,<sup>40</sup> and we have not evaluated individual-level data to assess the impact of coinheritance of variants, which are limitations of the analyses.

A further limitation of this study is that the case and ExAC data were not generated using a single sequencing method; none of the methods used was expected to have 100% sensitivity for variant detection. Although these technical limitations could have marginal effects on estimates of rare variant frequency and OR/EF values, we do not expect them to alter the key conclusions of this study. Because ethnicity data were not available for the Oxford Medical Genetics Laboratory or LMM HCM patients, we were unable to confirm the extent to which the cohorts used in this study were matched by race. However, by studying the aggregate burden of multiple very rare variants, we expect that any confounding effects by individual population-specific variants in cases or ExAC will be limited. Supporting this assumption, an analysis of the LMM DCM cohort comparing findings from all populations with the Caucasian-only subsets revealed that the conclusions are robust (Supplementary Table S13 online).

In conclusion, we have demonstrated that new opportunities for large-scale comparison of rare variation in Mendelian disease genes between patient cohorts and the wider population can identify the genes, regions of genes, and/or classes of variants that can be reliably interpreted in a clinical setting. For validated disease genes, there is clear potential to increase the yield of correctly interpreted, actionable variants. At the same time, problems must be avoided by recognizing that many implicated genes, as well as a significant proportion of variants, may not be robust. As clinical genetic testing moves to ever-larger gene panels and whole-exome and genome sequencing, an understanding of gene and variant pathogenicity will be increasingly important to deliver reliable interpretation.

## SUPPLEMENTARY MATERIAL

Supplementary material is linked to the online version of the paper at <http://www.nature.com/gim>

## ACKNOWLEDGMENTS

The research was supported by the NIHR Oxford Biomedical Research Centre, the NIHR Biomedical Research Unit in Cardiovascular Disease at the Royal Brompton & Harefield NHS Foundation Trust and Imperial College London, the Wellcome Trust, Fondation Leducq, the Tanoto Foundation, the SingHealth Duke-NUS Institute of Precision Medicine (PRISM), the Medical Research Council, the Academy of Medical Sciences, the British Heart Foundation, Arthritis Research UK, and the National Medical Research Council (NMRC) Singapore. K.L.T. is the recipient of a National Institute for Health Research (NIHR) doctoral fellowship (NIHR-HCS-D13-04-006). M.F. and H.C.W. acknowledge support from a Wellcome Trust core award (090532/Z/09/Z) and the BHF Centre of Research Excellence in Oxford (RE/13/1/30181). ExAC was partially supported by U54DK105566 (to D.G.M.) from the National Institutes of Diabetes and Digestive and Kidney Diseases of the National Institutes of Health. We thank the staff at the Oxford Medical Genetics Laboratories (OMGL), Oxford University Hospitals NHS Foundation Trust, for generating and interpreting data used in these analyses.

## DATA AVAILABILITY

Web resource Atlas of Cardiac Genetic Variation (<http://cardiob.org/ACGV>)

## DISCLOSURE

Professor Stuart Cook occasionally consults for Illumina Inc. The other authors declare no conflict of interest.

## REFERENCES

1. Xue Y, Ankala A, Wilcox WR, Hegde MR. Solving the molecular diagnostic testing conundrum for Mendelian disorders in the era of next-generation sequencing: single-gene, gene panel, or exome/genome sequencing. *Genet Med* 2015;17:444–451.
2. Woorderchak-Donahue W, VanSant-Webb C, Tvardik T, et al. Clinical utility of a next generation sequencing panel assay for Marfan and Marfan-like syndromes featuring aortopathy. *Am J Med Genet Part A* 2015;167:1747–1757.
3. Pugh TJ, Kelly MA, Gowrisankar S, et al. The landscape of genetic variation in dilated cardiomyopathy as surveyed by clinical DNA sequencing. *Genet Med* 2014;16:601–608.
4. An integrated map of genetic variation from 1,092 human genomes. *Nature* 2012;491:56–65.
5. Exome Aggregation Consortium; Lek M, Karczewski K, Minikel E, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature*, e-pub ahead of print 17 August, 2016. <http://biomedrxiv.org/content/early/2015/01/30/030338>.
6. Blekhman R, Man O, Herrmann L, et al. Natural selection on genes that underlie human disease susceptibility. *Curr Biol* 2008;18:883–889.
7. Watkins H, Ashrafian H, Redwood C. Inherited cardiomyopathies. *N Engl J Med* 2011;364:1643–1656.
8. Green RC, Berg JS, Grody WW, et al.; American College of Medical Genetics and Genomics. ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. *Genet Med* 2013;15:565–574.
9. Alfares AA, Kelly MA, McDermott G, et al. Results of clinical genetic testing of 2,912 probands with hypertrophic cardiomyopathy: expanded panels offer limited additional sensitivity. *Genet Med*. 2015;17:880–888.
10. Cole P, MacMahon B. Attributable risk percent in case-control studies. *Br J Prev Soc Med* 1971;25:242–244.
11. Robins JM, Greenland S. Estimability and estimation of excess and etiologic fractions. *Stat Med* 1989;8:845–859.
12. Greenland S, Robins JM. Conceptual problems in the definition and interpretation of attributable fractions. *Am J Epidemiol* 1988;128:1185–1197.
13. Ye J, Pavlicek A, Lunney EA, Rejto PA, Teng CH. Statistical method on nonrandom clustering with application to somatic mutations in cancer. *BMC Bioinformatics* 2010;11:11.

14. Lopes LR, Syrris P, Guttmann OP, et al. Novel genotype-phenotype associations demonstrated by high-throughput sequencing in patients with hypertrophic cardiomyopathy. *Heart* 2015;101:294–301.
15. Hershberger RE, Norton N, Morales A, Li D, Siegfried JD, Gonzalez-Quintana J. Coding sequence rare variants identified in MYBPC3, MYH6, TPM1, TNNC1, and TNNI3 from 312 patients with familial or idiopathic dilated cardiomyopathy. *Circ Cardiovasc Genet* 2010;3:155–161.
16. Parks SB, Kushner JD, Nauman D, et al. Lamin A/C mutation analysis in a cohort of 324 unrelated patients with idiopathic or familial dilated cardiomyopathy. *Am Heart J* 2008;156:161–169.
17. Hershberger RE, Parks SB, Kushner JD, et al. Coding sequence mutations identified in MYH7, TNNT2, SCNSA, CSRP3, LBD3, and TCAP from 313 patients with familial or idiopathic dilated cardiomyopathy. *Clin Transl Sci* 2008;1:21–26.
18. Ho CY, Charron P, Richard P, Girolami F, Van Spaendonck-Zwarts KY, Pinto Y. Genetic advances in sarcomeric cardiomyopathies: state of the art. *Cardiovasc Res* 2015;105:397–408.
19. Elliott PM, Anastasakis a, Borger M a, et al. 2014 ESC Guidelines on diagnosis and management of hypertrophic cardiomyopathy: The Task Force for the Diagnosis and Management of Hypertrophic Cardiomyopathy of the European Society of Cardiology (ESC). *Eur Heart J* 2014;35:2733–2779.
20. Ackerman MJ, Priori SG, Willems S, et al. 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). *Hear Rhythm* 2011;8:1308–1339.
21. Hoffmann B, Schmidt-Traub H, Perrot A, Osterziel KJ, Gessner R. First mutation in cardiac troponin C, L29Q, in a patient with hypertrophic cardiomyopathy. *Hum Mutat* 2001;17:524.
22. Osio A, Tan L, Chen SN, et al. Myozenin 2 is a novel gene for human hypertrophic cardiomyopathy. *Circ Res* 2007;100:766–768.
23. Chiu C, Bagnall RD, Ingles J, et al. Mutations in alpha-actinin-2 cause hypertrophic cardiomyopathy: a genome-wide analysis. *J Am Coll Cardiol* 2010;55:1127–1135.
24. Arimura T, Bos JM, Sato A, et al. Cardiac ankyrin repeat protein gene (ANKRD1) mutations in hypertrophic cardiomyopathy. *J Am Coll Cardiol* 2009;54:334–342.
25. Hershberger RE, Siegfried JD. Update 2011: clinical and genetic issues in familial dilated cardiomyopathy. *J Am Coll Cardiol* 2011;57:1641–1649.
26. McNally EM, Golbus JR, Puckelwartz MJ. Genetic mutations and mechanisms in dilated cardiomyopathy. *J Clin Invest* 2013;123:19–26.
27. Herman DS, Lam L, Taylor MR, et al. Truncations of titin causing dilated cardiomyopathy. *N Engl J Med* 2012;366:619–628.
28. Roberts AM, Ware JS, Herman DS, et al. Integrated allelic, transcriptional, and phenomic dissection of the cardiac effects of titin truncations in health and disease. *Sci Transl Med* 2015;7:270ra6.
29. Marston S, Copeland O, Gehmlich K, Schlossarek S, Carrier L. How do MYBPC3 mutations cause hypertrophic cardiomyopathy? *J Muscle Res Cell Motil* 2012;33:75–80.
30. Haas J, Frese KS, Peil B, et al. Atlas of the clinical genetics of human dilated cardiomyopathy. *Eur Heart J* 2015;36:1123–135a.
31. Turner TN, Douville C, Kim D, et al. Proteins linked to autosomal dominant and autosomal recessive disorders harbor characteristic rare missense mutation distribution patterns. *Hum Mol Genet* 2015;24:5995–6002.
32. Kapplinger JD, Landstrom AP, Bos JM, Salisbury BA, Callis TE, Ackerman MJ. Distinguishing hypertrophic cardiomyopathy-associated mutations from background genetic noise. *J Cardiovasc Transl Res* 2014;7:347–361.
33. Robinson P, Mirza M, Knott A, et al. Alterations in thin filament regulation induced by a human cardiac troponin T mutant that causes dilated cardiomyopathy are distinct from those induced by troponin T mutants that cause hypertrophic cardiomyopathy. *J Biol Chem* 2002;277:40710–40716.
34. Richard P, Charron P, Carrier L, et al. Hypertrophic cardiomyopathy: distribution of disease genes, spectrum of mutations, and implications for a molecular diagnosis strategy. *Circulation* 2003;107:2227–2232.
35. Ingles J, Sarina T, Yeates L, et al. Clinical predictors of genetic testing outcomes in hypertrophic cardiomyopathy. *Genet Med* 2013;15:972–977.
36. Xu T, Yang Z, Vatta M, et al.; Multidisciplinary Study of Right Ventricular Dysplasia Investigators. Compound and digenic heterozygosity contributes to arrhythmogenic right ventricular cardiomyopathy. *J Am Coll Cardiol* 2010;55:587–597.
37. Semsarian C, Ingles J, Maron MS, Maron BJ. New perspectives on the prevalence of hypertrophic cardiomyopathy. *J Am Coll Cardiol* 2015;65:1249–1254.
38. Golbus JR, Puckelwartz MJ, Dellefave-Castillo L, et al. Targeted analysis of whole genome sequence data to diagnose genetic cardiomyopathy. *Circ Cardiovasc Genet* 2014;7:751–759.
39. Andreasen C, Nielsen JB, Refsgaard L, et al. New population-based exome data are questioning the pathogenicity of previously cardiomyopathy-associated genetic variants. *Eur J Hum Genet* 2013;21:918–928.
40. Bezzina CR, Barc J, Mizusawa Y, et al. Common variants at SCN5A-SCN10A and HEY2 are associated with Brugada syndrome, a rare disease with high risk of sudden cardiac death. *Nat Genet* 2013;45:1044–1049.



This work is licensed under a Creative Commons Attribution 4.0 International License.

The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>