ORIGINAL ARTICLE

Genetic complexity in hypertrophic cardiomyopathy revealed by high-throughput sequencing

Luis R Lopes,[1] Anna Zekavati,[2] Petros Syrris,[1] Mike Hubank,[2] Claudia Giambartolomei,[3] Chrysoula Dalageorgou,[1] Sharon Jenkins,[1] William McKenna,[1] Uk10k Consortium,[4] Vincent Plagnol,[3] Perry M Elliott[1]

ABSTRACT

Background Clinical interpretation of the large number of rare variants identified by high throughput sequencing (HTS) technologies is challenging. The aim of this study was to explore the clinical implications of a HTS strategy for patients with hypertrophic cardiomyopathy (HCM) using a targeted HTS methodology and workflow developed for patients with a range of inherited cardiovascular diseases. By comparing the sequencing results with published findings and with sequence data from a large-scale exome sequencing screen of UK individuals, we sought to quantify the strength of the evidence supporting causality for detected candidate variants.

Methods and results 223 unrelated patients with HCM (46±15 years at diagnosis, 74% males) were studied. In order to analyse coding, intronic and regulatory regions of 41 cardiovascular genes, we used solution-based sequence capture followed by massive parallel resequencing on Illumina GAIIx. Average read-depth in the 2.1 Mb target region was 120. Rare (frequency<0.5%) non-synonymous, loss-of-function and splice-site variants were defined as candidates. Excluding titin, we identified 152 distinct candidate variants in sarcomeric or associated genes (89 novel) in 143 patients (64%). Four sarcomeric genes (*MYH7*, *MYBPC3*, *TNNI3*, *TNNT2*) showed an excess of rare single non-synonymous single-nucleotide polymorphisms (nsSNPs) in cases compared to controls. The estimated probability that a nsSNP in these genes is pathogenic varied between 57% and near certainty depending on the location. We detected an additional 94 candidate variants (73 novel) in desmosomal, and ion-channel genes in 96 patients (43%).

Conclusions This study provides the first large-scale quantitative analysis of the prevalence of sarcomere protein gene variants in patients with HCM using HTS technology. Inclusion of other genes implicated in inherited cardiac disease identifies a large number of non-synonymous rare variants of unknown clinical significance.

INTRODUCTION

Hypertrophic cardiomyopathy (HCM), defined as left ventricular hypertrophy in the absence of abnormal loading conditions, occurs in approximately one in every 500 adults and can cause sudden cardiac death at all ages, and progressive deterioration in left ventricular function.[1–6] In 50–60% of adolescents and adults with the disease, HCM is inherited as an autosomal dominant trait caused by mutations in cardiac sarcomere protein genes. Mutations in genes encoding Z-disc or calcium-handling proteins account for less than 1% of cases, and a further 5% of patients have metabolic disorders, neuromuscular disease, chromosome abnormalities and genetic malformation syndromes.[7–17] The disease is characterised by a highly heterogeneous phenotype, a highly variable intra- and interfamily expressivity and incomplete penetrance. This genotype-phenotype plasticity is largely unexplained.

Although current clinical guidelines recommend routine genetic testing in patients with HCM,[18 19] its use in everyday clinical practice has been limited by the cost and complexity of conventional sequencing technologies. Advances in high throughput sequencing technology (HTS) have the potential to solve this problem by analysing substantially larger genomic regions at a lower cost than conventional capillary Sanger sequencing,[20] but they may also pose new challenges. In particular, the potential to identify a large number of rare variants that are also found in the general population, and which have little or no effect on disease phenotypes, could make attribution of causality to candidate variants using conventional methods, such as cosegregation analysis in large pedigrees, impractical.

The aim of this study was to explore the clinical implications of a HTS strategy for patients with HCM using a targeted HTS methodology and workflow developed for patients with a range of inherited cardiovascular diseases. By comparing the sequencing results with published findings, and with sequence data from a large-scale exome sequencing screen of 1287 Caucasian UK individuals (UK10K project, http://www.uk10k.org), we quantify the strength of the evidence supporting causality for detected candidate variants.

METHODS

Patients and clinical evaluation

The study cohort comprised unrelated consecutively evaluated patients with HCM referred to an inherited cardiovascular disease unit at The Heart Hospital, University College London (UCL), London, UK. Patients underwent 12-lead ECG, echocardiography, symptom-limited upright exercise testing with simultaneous respiratory gas analysis and ambulatory ECG monitoring. HCM was diagnosed in probands when the maximum left ventricular wall thickness (MLVWT) on two-dimensional

echocardiography measured 13 mm or more in at least one myocardial segment, or when MLVWT exceeded two SDs corrected for age, size and gender in the absence of other diseases that could explain the hypertrophy. In individuals with unequivocal disease in a first-degree relative, a diagnosis was made using extended familial criteria for HCM.[21]

Targeted gene enrichment and sequencing for case samples

The project was approved by the UCL/UCLH Joint Research Ethics Committee. All patients provided written informed consent and received genetic counselling prior to venesection. Blood samples were collected at routine clinic visits, and DNA was isolated from peripheral blood lymphocytes.

The study was designed to screen 2.1 Mb of genomic DNA sequence per patient, covering coding, intronic and selected regulatory regions of 20 genes known to be associated with HCM and dilated cardiomyopathy (DCM), 17 genes implicated in other inherited cardiomyopathies and arrhythmia syndromes, and a further four candidate genes (table 1).

A web-based design tool, eArray (Agilent Technologies, Santa Clara, California, USA) was employed to design an initial SureSelect (Agilent) capture library of oligonucleotides (RNA bait groups) based on the target gene sequences using the following parameters: library size 1×55 K; length 120; tiling 1×. Control samples from patients with HCM who were known to carry disease-causing sequence variants, previously detected with conventional Sanger sequencing, were used in pilot studies to validate the method. The library was used to capture the target regions from eight patients which were then sequenced (single end) on an Illumina GAIIx platform with 35 base read length. Single-end sequencing of control samples with known HCM-related variants identified regions of low coverage possibly associated with suboptimal sample processing steps or low capture efficiency. The pilot study was used to optimise the protocol and to redesign the capture library, introducing double density cover to regions of low coverage, and increase cover at the 5′ regulatory ends of genes. The following steps were adopted: single-end adapters were replaced with paired-end adapters; sequencing read length was increased from 35 bp to 75 bp, and the capture RNA bait library was redesigned with eArray to enrich regions with low coverage. The new design included an additional 965 RNA baits at 2× tiling for all <30 read depth regions, 2× tiling for 20 genes associated with HCM and DCM and redesigned sequence regions were extended 5–10 Kb upstream of all genes. As a result, 23 637 120 bp RNA baits were redesigned to target a total of 2.1 Mb of genomic DNA. In order to increase efficiency and reduce costs, we adopted a 75 bp paired-end multiplexed sequencing method which allowed us to pool 12 samples into a single lane of an Illumina GAIIX flow cell and, thus, taking into account internal controls, sequence a total of 84 samples in a single instrument run. The multiplex sequencing protocol was also tested using control samples. In early sequencing test runs a total of 21 samples from HCM patients were sequenced as part of developing and optimising the method (data not shown).

For phase two, sample preparation was carried out as recommended by Agilent and initially based on the SureSelect Target Enrichment for Illumina paired-end multiplexed sequencing method. Genomic DNA shearing (3 μg) per patient was performed on a Covaris E220 instrument in 96-well plates. Fragmented DNA was end-repaired and 'A' base addition was performed using the NEBNext DNA Sample Prep Master mix Set 1 (New England BioLabs). Ligation of indexing-specific paired-end adapters to DNA samples was performed using the

Table 1 Name of the targeted genes, Ensembl accession number, chromosomal position and size

Gene	Ensembl number	Location Chromosome: base range	Number (bp)
MYBPC3	ENSG00000134571	chr11:47352958–47374253	21295
MYH7	ENSG00000092054	chr14:23881948–23904870	22922
TNNI3	ENSG00000129991	chr19:55663137–55669100	5963
TNNT2	ENSG00000118194	chr1:201328143–201346805	18662
TPM1	ENSG00000140416	chr15:63334838–63364111	29273
MYL2	ENSG00000111245	chr12:111348626–111358404	9778
MYL3	ENSG00000160808	chr3:46899357–46904973	5616
ACTC1	ENSG00000159251	chr15:35080297–35087927	7630
TNNC1	ENSG00000114854	chr3:52485108–52488057	2949
MYH6	ENSG00000197616	chr14:23851199–23877482	26283
TTN	ENSG00000155657	chr2:179390720–179672150	281430
PDLIM3	ENSG00000154553	chr4:186422852–186456712	33860
CSRP3	ENSG00000129170	chr11:19203578–19223589	20011
DES	ENSG00000175084	chr2:220283099–220291459	8360
LMNA	ENSG00000160789	chr1:156084461–156109878	25417
LDB3	ENSG00000122367	chr10:88428426–88495822	67396
VCL	ENSG00000035403	chr10:75757872–75879912	122040
TCAP	ENST00000309889	chr17:37821599–37822806	1207
PLN	ENSG00000198523	chr6:118869442–118881586	12144
RBM20	ENSG00000203867	chr10:112404155–112599227	195072
JUP	ENSG00000173801	chr17:39910859–39942964	32105
DSP	ENSG00000096696	chr6:7541870–7586946	45076
PKP2	ENSG00000057294	chr12:32943682–33049780	106098
DSG2	ENSG00000046604	chr18:29078027–29128813	50786
DSC2	ENSG00000134755	chr18:28645944–28682388	36444
RYR2	ENSG00000198626	chr1:237205702–237997288	791586
TMEM43	ENST00000306077	chr3:14166440–14185180	18740
TGF-β3	ENST00000238682	chr14:76424442–76448092	23650
KCNQ1	ENSG00000053918	chr11:2466221–2870339	404118
KCNH2	ENSG00000055118	chr7:150642050–150675014	32964
SCN5A	ENSG00000183873	chr3:38589554–38691164	101610
KCNE1	ENSG00000180509	chr21:35818989–35828063	9074
KCNE2	ENSG00000159197	chr21:35736323–35743440	7117
ANK2	ENST00000394537	chr4:113970785–114304894	334109
CASQ2	ENSG00000118729	chr1:116242628–116311426	68798
CAV3	ENSG00000182533	chr3:8775496–8788450	12954
KCNJ2	ENSG00000123700	chr17:68165676–68176181	10505
PLEC	ENSG00000178209	chr8:144989321–145025044	35723
GJA1	ENST00000282561	chr6:121756745–121770872	14127
PKP4	ENSG00000144283	chr2:159313476–159537938	224462
PNN	ENSG00000100941	chr14:39644387–39652421	8036
Total			3285390

Ensembl: Feb. 2009 (GRCh37/hg19). The total size of genomic sequence for 41 loci was approximately 3 Mb, which was reduced to 2.1 Mb of capture sequence following the exclusion of repetitive DNA regions in the custom RNA bait library design with eArray (Agilent).

DES, desmin; PLN, phospholamban; VCL, vinculin; MYBPC3: myosin binding protein C, cardiac; MYH7: myosin, heavy chain 7, cardiac muscle, beta; TNNI3: troponin I type 3 (cardiac); TNNT2: troponin T type 2 (cardiac); TPM1: tropomyosin 1 (alpha); MYL2: myosin, light chain 2, regulatory, cardiac, slow; MYL3: myosin, light chain 3, alkali; ventricular, skeletal, slow; ACTC1: actin, alpha, cardiac muscle 1; TNNC1: troponin C type 1 (slow); MYH6: myosin, heavy chain 6, cardiac muscle, alpha; TTN: titin; PDLIM3: PDZ and LIM domain 3; CSRP3: cysteine and glycine-rich protein 3 (cardiac LIM protein); DES: desmin; LMNA: lamin A/C; LDB3: LIM domain binding 3; VCL: vinculin; TCAP: titin-cap; PLN: phospholamban; RBM20: RNA binding motif protein 20; JUP: junction plakoglobin; DSP: desmoplakin; PKP2: plakophilin 2; DSG2: desmoglein 2; DSC2: desmocollin 2; RYR2: ryanodine receptor 2 (cardiac); TMEM43: transmembrane protein 43; THFbeta3: transforming growth factor, beta 3; KCNQ1: potassium voltage-gated channel, KQT-like subfamily, member 1; KCNH2: potassium voltage-gated channel, subfamily H (eag-related), member 2; SCN5A: sodium channel, voltage-gated, type V, alpha subunit; KCNE1: potassium voltage-gated channel, Isk-related family, member 1; KCNE2: potassium voltage-gated channel, Isk-related family, member 2; ANK2: ankyrin 2, neuronal; CASQ2: calsequestrin 2 (cardiac muscle); CAV3: caveolin 3; KCNJ2: potassium inwardly-rectifying channel, subfamily J, member 2; PLEC: plectin; GJA1: gap junction protein, alpha 1, 43kDa; PKP4: plakophilin 4; PNN: pinin, desmosome associated protein.

Illumina Paired-End Genomic DNA Sample Prep Kit, and the subsequent amplification of the adapter-ligated library was carried out with Herculase II Fusion DNA Polymerase (Agilent). Hybridisation of amplified libraries to the SureSelect biotinylated RNA library (baits) was performed at 650°C for 24 h on a GeneAmp PCR System 9700 (AppliedBiosystems). Addition of index tags to the library preparation was achieved by PCR using the Illumina Multiplex Sample Preparation Oligonucleotide Kit and Herculase II Fusion DNA Polymerase (Agilent). Following the introduction of the SureSelectXT Target Enrichment protocol, all the above steps were performed with Agilent reagents as recommended by the manufacturer. Clean-up of DNA samples was performed according to the protocol using Agencourt AMPure XP beads and Dynal MyOne Streptavidin T1 magnetic beads (Invitrogen, hybridisation step). Quality of DNA samples throughout the protocol was assessed with Agilent 2100 Bioanalyzer DNA assays. All steps were performed manually using individual PCR tubes or 96-well plates except the incubation of the hybrid-capture/bead solution (hybridisation step) on 96-well plates, which was carried out on the Bravo Automated Liquid Handling Platform (Agilent) using an Agilent instrument protocol with small modifications. Samples were subjected to standard Illumina protocols for cluster generation and sequencing. Paired-end multiplexed sequencing was performed on an Illumina GAIIX with 12 samples tagged with different index sequences (Illumina) combined in each lane.

Potentially pathogenic variants and variants with a read depth below 15 were confirmed by conventional dideoxy sequencing using BigDye Terminator V.3.1 sequencing chemistry (AppliedBiosystems) on a 3130 × l capillary sequencer (AppliedBiosystems). We also sequenced, using conventional dideoxy sequencing, those exons with an average read depth below 15. All *MYH6* variants were also confirmed using dideoxy sequencing, due to the high homology between *MYH7* and *MYH6* at DNA level, which could potentially generate false positive results.

Sequencing of UK10K control samples (http://www.uk10k.org)

DNA (1–3 µg) was sheared to 100–400 bp using a Covaris E210 or LE220 (Covaris, Woburn, Massachusetts, USA). Sheared DNA was subjected to Illumina paired-end DNA library preparation and enriched for target sequences (Agilent Technologies; Human All Exon 50 Mb ELID S02972011) according to the manufacturer's recommendations (Agilent Technologies; SureSelectXT Automated Target Enrichment for Illumina Paired-End Multiplexed Sequencing). Enriched libraries were sequenced using the HiSeq platform (Illumina) as paired-end 75 base reads according to the manufacturer's protocol.

Bioinformatic sequence data analysis

For the HCM samples, paired-end reads were aligned using the Novoalign software V.2.7.19 on the human reference genome build hg19 using quality score calibration, soft clipping and Illumina adapter trimming. Following the exclusion of PCR duplicate reads (Picard MarkDuplicate tool), insertion-deletions (indels) and single-nucleotide polymorphisms (SNPs) were called using the software SAMtools (V.0.1.18, using single sample calling).[22] Variants (SNPs/indels) were filtered on the basis of the Phred scaled genotype quality score (minimum value of 30, as computed by SAMtools). For the UK10K samples, alignment was performed using Bowtie and the calling algorithm merged the output of SAMtools (V.0.1.17, single sample calling) and GATK Unified Genotyper (V.1.3-21). All samples were annotated using Annovar.[23] We defined our set of candidate variants for further analysis based on frequency and function. The frequency filter

used the allele frequency estimates from the 1000 genomes project database,[24] and we used a 0.5% cut-off (based on the November 2010 and May 2011 releases). For the functional filter, exonic non-synonymous, loss-of-function and splice-site variants located in one of the 41 targeted genes were included in this candidate set. After filtering, variants present in the dbSNP build 135 database[25] were identified, but not excluded from the analysis. We also identified variants that were previously published in the literature as disease causing mutations. Prediction of in silico pathogenicity for novel missense variants was performed using Polyphen2 and SIFT prediction software.[26 27] A variant was predicted to be pathogenic if it was classified as 'damaging' by SIFT and 'possibly or 'probably damaging' by Polyphen2.

Analysis of UK10K control samples

Sequencing results were compared with a set of 1287 UK controls with exome sequence data generated by the UK10K project (http://www.uk10k.org). These samples are the subset of UK10K exomes for which ethics enabled their use as control samples. None of the UK10K control samples was recruited on the basis of a cardiac phenotype. To limit the technical difficulties associated with comparing sets of variants in controls and cases generated using different protocols and analysed with the same tools but in different laboratories, we restricted our comparison to non-synonymous SNPs (nsSNPs), hence excluding indels and larger copy number variants. We retrieved the data from the UK10K project and annotated nsSNPs as candidates using the same protocol and thresholds that were applied to the set of HCM cases. To avoid biases associated with variable coverage between cases and controls, we only considered in this case-control comparison the exons that were sequenced with a read depth of 10 or more in both the UK10K dataset and our HCM case collection.

Statistical assessment of the case control comparison of candidate nsSNPs

This analysis was restricted to the 180 Caucasian HCM cases, which could be matched to the 1287 UK10K control samples. All computations were performed using the statistical software R. Frequencies of candidate nsSNPs were compared between cases and controls. We then used the case-control data to infer the proportion of HCM cases explained by rare nsSNPs variants in each gene. We used a profile likelihood approach to estimate this parameter of interest (see online supplementary material—additional statistical methods).

For genes showing a significant excess of rare nsSNPs, point estimates for the probability that a rare nsSNP is causal for HCM were estimated using the formula: (proportion of carriers of rare nsSNPs in cases—proportion of carriers of rare nsSNPs in controls)/proportion of carriers of rare nsSNPs in cases (see online supplementary material—additional statistical methods).

RESULTS

Study population

Two-hundred-and-twenty-three unrelated patients with HCM were studied. The mean age at initial evaluation was 46 ±15 years (5–76); 165 (74%) were men. Mean MLVWT was 19.5±4.6 mm. Table 2 summarises the demographic and clinical characteristics of the patients.

Summary of sequence data

The median value of the per-sample average read depth in the 2.1 Mb target region across the samples was 120. Only four out 223 samples had an average read depth lower than 40, with a

Table 2 Demographic, clinical and echocardiographic characterisation of the patients

	Frequency (percentage) or mean±standard deviation	Range
Age at initial evaluation (years)	46±15	5–76
Male	165 (74)	
Ethnicity—Caucasian	180 (80.7)	
Family history HCM	82 (37)	
NYHA class III or IV	26 (12)	
Maximal left ventricular wall thickness (mm)	19.5±4.6	11–35
Left atrial diameter (mm)	45±7	31–63
Left ventricular end-diastolic diameter (mm)	45.4±6.1	29–68
Left ventricular end-systolic diameter (mm)	27.3±6.1	12–50
Fractional shortening (%)	40.3±8.9	16–69

HCM: hypertrophic cardiomyopathy; NYHA: New York Heart Association.

minimum of 20. Combining all samples and taking the median value across all samples, 91.2% of the target region was covered to a depth of 10 or more, and 85.2% to a depth of 20 or more (figure 1). Variants were filtered on the basis of the SAMtools Phred scale quality score greater than 20. We successfully validated, by Sanger sequencing, all 50 variants with a sufficient quality score but a read depth lower than 15, except for one frameshift deletion in *KCNH2*. This indicates that the approach has a very low false positive rate even with a limited read depth. We also screened the following genomic regions using conventional dideoxy sequencing, because of low average read depth across the samples: *MYBPC3* exons 5 and 13; *TNNI3* exon 1; *KCNH2* exons 1 and 13. Only one false negative was found, c.459delC, p.P153fsX5 in *MYBPC3*. This indicates that the approach has a very low rate of false negatives. Additionally, all the *MYH6* variants detected with HTS were confirmed by Sanger sequencing.

Genotyping results
Initial genotype calling generated 21 939 exonic and splice-site calls distinct from the reference sequence, that corresponded to a total of 1758 distinct variants present in the 223 patients. After exclusion of synonymous substitutions, we found 9180 exonic and splice-site calls (994 distinct variants).

After filtering (as described in the methods), we selected 480 distinct rare non-synonymous exonic or splice-site variants (641 calls) as candidates for further analysis. In total, 209 patients

(93.7%) carried at least one variant in the target genes, 177 when excluding *TTN* (79.4%). One-hundred-and-sixty-one (72.2%) patients carried multiple variants, 98 when excluding *TTN* (43.9%).

Variants in sarcomeric, Z-disc and calcium-handling genes
One hundred-and-two distinct rare variants in eight sarcomeric protein genes (*MYH7*, *MYBPC3*, *TNNT2*, *TNNI3*, *MYL2*, *MYL3*, *ACTC1* and *TPM1*) were identified in 110 (49%) patients. Fifty-nine (58%) of these sarcomere protein gene variants were previously published as pathogenic mutations.[8 11 15 16 28–63] Nineteen (19%) were novel missense variants predicted in silico to be pathogenic, and 19 (19%) were novel potential loss-of-function variants. In total, 97% of these sarcomere variants present in 106 patients (48% of our cohort) were considered strong disease-causing candidates.

The distribution of sarcomere variants, including *MYH6*, is shown in figure 2. Twenty-five patients (11%) carried multiple candidate variants in sarcomere protein genes. Twelve (63%) of the 19 double and triple heterozygotes carried a cardiac myosin binding protein C (*MYBPC3*) variant as one of the candidate sarcomere variants. Two patients were compound heterozygotes for *MYBPC3*, one patient was homozygous for a *MYBPC3* variant, one patient was a compound heterozygote for β-myosin heavy chain (*MYH7*) and two patients were compound heterozygotes for α-myosin heavy chain (*MYH6*). Eighteen rare variants in *MYH6*, one previously published as a disease-causing mutation in a family with congenital heart disease[64] and eight predicted in silico to be pathogenic, were present in 19 patients (8.5%).

Expanding the analysis to a panel of 19 sarcomeric and related genes, previously associated with HCM or DCM (sarcomeric, Z-disc and calcium-handling genes, excluding the highly variable titin) resulted in the detection of 152 distinct rare variants in 143 patients (64%) (figure 3). The number of variants found in each gene is summarised in table 3. From those 152 variants, 63 (41%), present in 79 patients (35%), have been previously published as pathogenic mutations. Thirty-seven (24%) were novel missense variants predicted in silico to be pathogenic, and 23 (15%) were novel nonsense, frameshift indels or splice-site variants predicted to cause loss-of-function (see online supplementary table S1 and table 3). In total, a majority (75%) of these variants are strong candidates for pathogenicity, and were present in 131 patients (59% of our cohort).

Thirty-two distinct candidate variants in Z-disc and calcium-handling genes were detected. Three were previously published as disease-causing mutations: one cysteine and glycine-rich protein 3 (*CSRP3*) published pathogenic mutation[55] (in DCM); one telethonin (*TCAP*) published pathogenic mutation[52] (in

Figure 1 Box-and-whisker plot, showing the read-depths across the targeted genes.
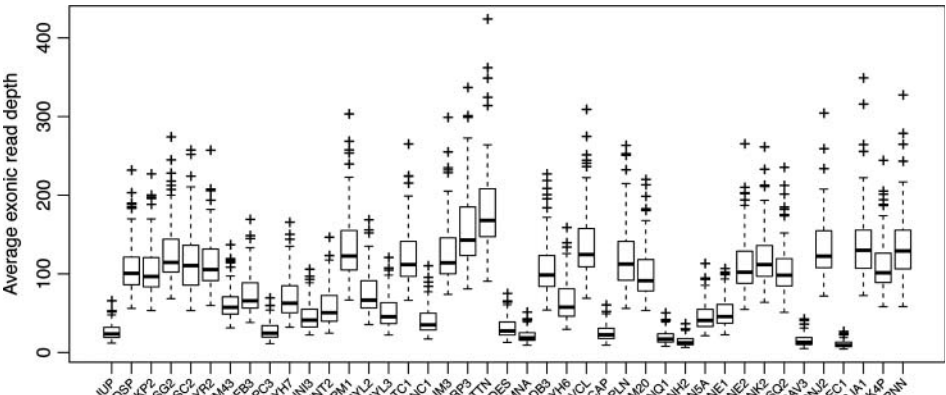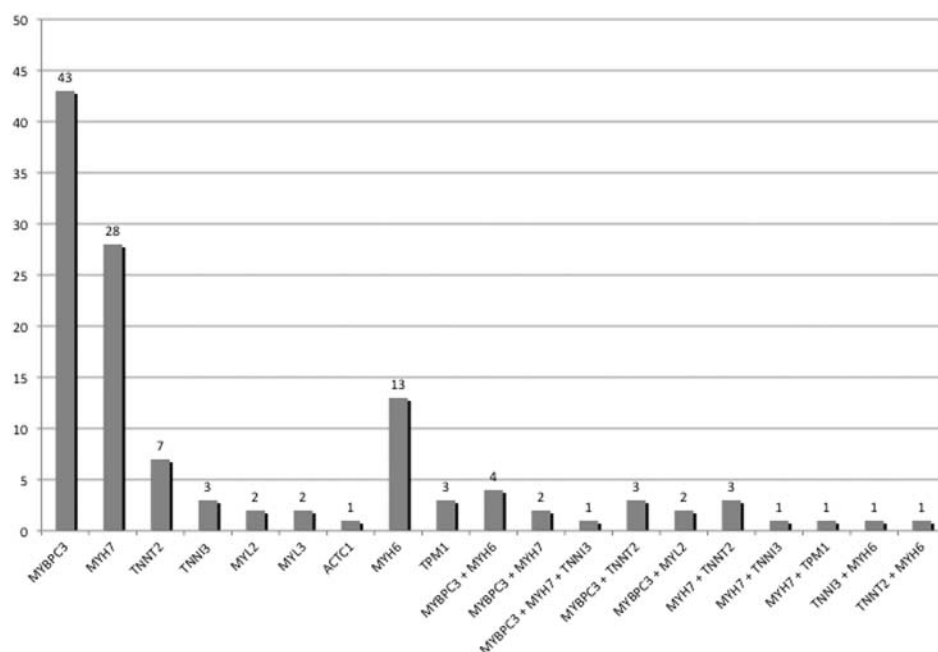
Figure 2 Number of patients with variants in each of the sarcomeric genes.

HCM); and one RNA-binding protein 20 (*RBM20*) pathogenic mutation[54] (in DCM). Two additional *CSRP3* variants were predicted to affect a canonical splice-site, probably causing loss-of-function, and 10 novel missense variants were predicted in silico to be pathogenic: one phospholamban (*PLN*), one *CSRP3*, one lamin (*LMNA*), four *RBM20*, one vinculin (*VCL*), one desmin (*DES*) and one LIM-domain binding protein 3 (*LDB3*).

Table 4 shows the distribution of patients according to the strength of the evidence supporting causality for the detected candidate variants.

Statistical analysis of nsSNPs case-control data

For each gene, we combined nsSNPs to test for an overall enrichment in HCM cases compared with the general population. For our control set, we used the UK10K exome sequence dataset (Methods). To avoid technical artefacts associated with indel calling, and to properly match cases and controls, we restricted this analysis to nsSNPs and the 180 HCM Caucasian HCM cases. Data for the eight sarcomere genes most commonly implicated in HCM are summarised in table 5 and figure 4. A complete table of 19 sarcomere and associated genes is provided
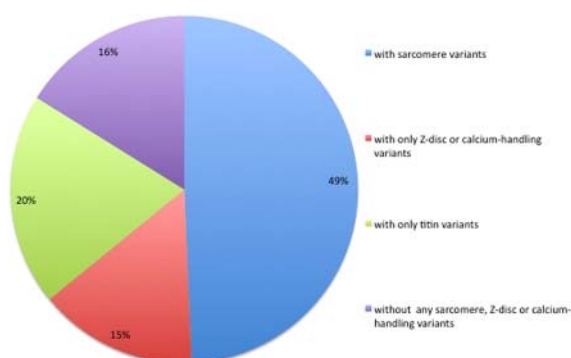

Figure 3 Percentage of patients with rare variants in hypertrophic cardiomyopathy/dilated cardiomyopathy associated genes. This figure is only reproduced in colour in the online version.

in online supplementary table S2. Four out of the 20 cardiomyopathy genes (*MYH7*, *MYBPC3*, *TNNI3*, *TNNT2*) showed an excess of rare nsSNP in cases compared with controls (two-tailed Fisher exact $p < 0.05$), which is consistent with the established causal role of rare nsSNPs in these genes. We used these case-control data to extrapolate the proportion of HCM cases in the general population explained by variants in each of these genes (Methods and table 5). Rare nsSNPs in *MYH7* explained the largest fraction, with between 9.6 and 20.7% of HCM cases (95% CI). We note that *MYBPC3* harbours a significant number of loss-of-function indels that are excluded in this analysis, therefore, underestimating the contribution of *MYBPC3*.

Assuming that an excess of candidate variants in HCM cases reflects their disease-causing potential, these data can be used to estimate the probability that a candidate nsSNP found in a HCM case is disease causing (Methods and table 5). We found that these estimates are largely dependent of the genetic variability for each gene in the general population. As an example, nsSNPs in *MYH7* explain between 9.6% and 20.7% of HCM cases, and a rare nsSNP in *MYH7* found in a HCM case is estimated to have 86% probability to be causal for HCM (table 5). Higher estimates are obtained for rare nsSNP in the genes *TNNT2* and *TNNI3*, even though the contribution of these genes to HCM cases is lower. This is a consequence of the much reduced presence of nsSNPs in these genes in the control population (0.3% and 0%, respectively) compared with *MYH7* (2.5%)—table 5.

Recurrent variants in HCM cases

We then investigated whether nsSNP variants were found in multiple UK HCM cases, suggesting potentially common genetic causes of HCM. We found three rare nsSNPs for which the single SNP case-control p value was less than 0.05, in *MYPC3*, *MYH7* and *TNNT2* (table 6). All were previously published as disease-causing mutations.

Titin

Two-hundred-and-nineteen rare titin variants were identified in 142 probands (63.6%). Two-hundred-and-nine were novel missense variants. One of the variants is a missense mutation

Table 3 Number of distinct rare variants in sarcomeric, Z-disc and calcium-handling genes

Genes	Total number of distinct rare variants in each gene	Published	Novel Missssense variants predicted to be pathogenic in silico	Nonsense, frameshift or splice-site variants predicted to cause loss of function
Sarcomere genes				
MYBPC3	46	23	2	17
MYH7	31	21	9	
TNNI3	5	4	1	
TNNT2	9	8		1
MYL2	3	2	1	
MYL3	2		2	
ACTC1	1		1	
TPM1	5	1	3	1
MYH6	19	1	8	2
TTN	219	1		9
TNNC1	0			
Subtotal	339	61	27	30
Z disc/related, calcium handling				
PLN	4		1	
CSRP3	3	1	1	1
TCAP	2	1		
LDB3	2		1	
VCL	6		1	
PDLIM3	2			
DES	3		1	1
LMNA	1		1	
RBM20	9	1	4	
Subtotal	32	3	10	2
Total	371	64	37	32

previously described as pathogenic in HCM (R8500H).[65] Nine other variants were predicted to cause loss of function in 10 patients: two are frameshift insertions potentially leading to a truncated protein (one in a patient that carried one TNNT2 variant of unknown significance, the other in a patient with a MYBPC3 splice-site variant); another is a nonsense variant that probably leads to the synthesis of a truncated protein, and six more are splice-site variants predicted to cause exon skipping (one in a patient that also carried a frameshift small deletion in MYBPC3, and another in a patient who carried a published MYH7 mutation; the other four patients did not carry any other sarcomere or related variant) (see online supplementary table S1). Thirty patients (13%) had titin candidate variants in isolation. Twenty-two patients (10%) had titin variants only in association with desmosomal gene candidate variants or ion channel disease-associated gene variants, but not other sarcomere (or related) variants. This means that 171 patients (77% of the cohort) carried a TTN candidate variant in association with sarcomere, Z-disc or calcium-handling gene variants. By contrast, in the 11 patients (5%) carrying potentially truncating variants or the published mutation, six patients carried the TTN variant isolated or only associated with ion-channel/desmosome genes (55%), and five patients in association with sarcomere variants (45%).

Table 4 Level of evidence for the pathogenicity of the distinct variants. ('Others'—novel missense variants not predicted to be pathogenic in silico)

	Number of patients (%)
Sarcomere genes (MYH7, MYBPC3, TNNT2, TNNI3, MYL2, MYL3, ACTC1, TPM1)	110 (49.3)
Published mutation	71 (31.8)
Loss of function/in silico predicted to be damaging	35 (15.7)
Others	4 (1.8)
Other sarcomere and sarcomere-associated genes (titin excluded): MYH6, VCL, CSRP3, DES, LDB3, LMNA, RBM20, TCAP, PLN, PDLIM3	33 (14.8)
Published mutation	4 (1.8)
Loss of function/in silico predicted to be damaging	18 (8.1)
Others	11 (4.9)
Total	143 (64.1)

Table 5 Rare nsSNPs frequency comparison between our sequencing results and a set of 1287 UK controls with exome sequence data generated by the UK10K project (http://www.uk10k.org) for the eight sarcomere genes most commonly associated with HCM

Gene	Rare nsSNPs—Frequency in controls	Rare nsSNPs—Frequency in patients	p Value	95% CI for the proportion of cases explained	Probability that a nsSNP is causal
MYH7	0.025	0.172	3.86E-13	(0.096 to 0.207)	0.856
TNNT2	0.003	0.044	8.41E-05	(0.014 to 0.071)	0.930
TNNI3	0.000	0.017	0.002	(0.003 to 0.042)	1.000
MYBPC3	0.045	0.106	0.007	(0.014 to 0.104)	0.570
MYL2	0.007	0.022	0.065	(0 to 0.044)	NA
MYL3	0.004	0.011	0.208	(0 to 0.0301)	NA
ACTC1	0.0008	0.006	0.230	(0 to 0.0167)	NA
TPM1	0.000	0.006	0.123	(0 to 0.0167)	NA

To avoid technical artefacts associated with indel calling, and to properly match cases and controls, we restricted this analysis to nsSNPs and the 180 HCM UK Caucasian HCM cases. The columns show: the proportion of the 1287 UK controls with exome sequence data generated by the UK10K project (http://www.uk10k.org) and 180 HCM Caucasian cases that carry rare nsSNPs in these genes (rare defined by frequency less than 0.5% in the 1000 Genomes dataset), a Fisher exact test p value to quantify the case-control difference, the 95% CI for the estimated proportion of HCM cases explained by rare nsSNPs variants in each gene (see online supplementary statistical methods), and (in the rightmost column) the estimated probability that a rare nsSNP found in a HCM case in each gene is disease causing (see online supplementary statistical methods).
HCM, hypertrophic cardiomyopathy; nsSNP, non-synonymous single nucleotide polymorphism.

Variants in genes associated with arrhythmogenic cardiomyopathy and ion-channel disease

Ninety-four distinct candidate variants in genes implicated in arrhythmogenic cardiomyopathy (44) and ion-channel disease (50) were present in 96 patients (43%). Twenty-one (24%) of these variants were previously published[66–81] (including 13 published variants of unknown significance in desmosomal genes, one disease-causing mutation in *PKP2*, and seven disease-causing mutations in 12 patients—5%—in ion-channel disease genes; table 7 and see online supplementary table S1). A further 20 variants (23%) are novel missense variants predicted in silico to be pathogenic, and three (3%) are potential loss-of-function variants. In total, aproximately half these variants are predicted to have a biological effect. Approximately two-thirds (57 patients) of these 96 patients also carried variants in sarcomere or related genes.

DISCUSSION

This study provides the first large-scale quantitative analysis of the prevalence of sarcomere protein gene variants in patients with HCM using HTS technology. Our HTS protocol achieved adequate coverage of the targeted genomic DNA, identifying

likely pathogenic sarcomeric variants in 49% of patients. We report, for the first time, the prevalence of all types of variants (including missense) in *TTN* and find a higher than expected number of novel rare variants in *MYH6*, although the total number in both genes was similar to that found in normal controls. Inclusion of other genes implicated in inherited cardiac disease resulted in the identification of a large number of non-synonymous rare variants. While the overall frequency of these variants was similar to the control population, published data and in silico prediction tools suggest that some of these have the potential to modify the disease phenotype.

Determining pathogenicity of sequence variants

Even when using conventional sequencing technology, the genetic heterogeneity of HCM and the high frequency of novel variants with uncertain effects on gene function present considerable challenges for clinical interpretation. Ideally, novel variants should be subjected to functional studies, but these are costly, time consuming, and often impractical in the clinical setting. Similarly, cosegregation analysis within families can be helpful, but is uninformative in small pedigrees and often difficult to orchestrate.

Figure 4 Rare nsSNPs frequency comparison between our sequencing results and a set of 1287 UK controls with exome sequence data generated by the UK10K project (http://www.uk10k.org) for the eight sarcomere genes most commonly associated with hypertrophic cardiomyopathy. nsSNP: non-synonymous single nucleotide polymorphism. The frequency of *MYH7, MYBPC3, TNNT2* and *TNNI3* candidate nsSNPs is significantly higher in our cohort, as also shown in table 5. This figure is only reproduced in colour in the online version.
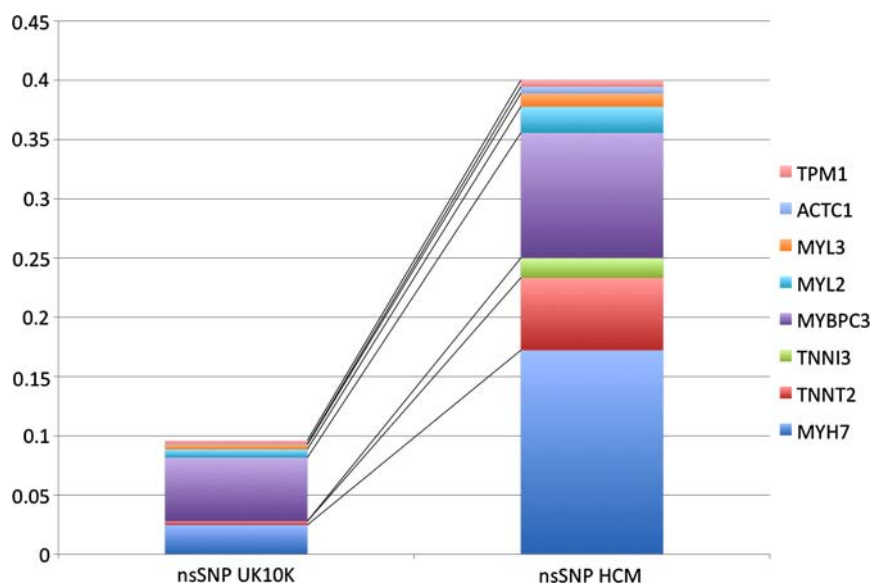
Table 6 Candidate variants present in multiple HCM cases for which the single nsSNP case-control p value between HCM cases and UK10K controls is p<0.05

Gene	Amino acid change	Calls (n)	Frequency in patients	UK10K MAF	p Value	dbSNP135	Published as disease-causing
MYBPC3	NM_000256:c.G1484A:p.R495Q	3	0.008333333	0	0.001833766	Not present	Yes
TNNT2	NM_001001432:c.C296T:p.A104V	3	0.008333333	0	0.001833766	Not present	Yes
MYH7	NM_000257:c.G1988A:p.R663H	3	0.008333333	0	0.001833766	Not present	Yes

HCM, hypertrophic cardiomyopathy; MAF, minimum allele frequency; nsSNPs, non-synonymous single nucleotide polymorphism.

The recent availability of sequence datasets for large cohorts makes it possible to statistically compare the distribution of rare variants between controls and cases. While this statistical approach cannot on its own identify a variant as causal, it provides insights into the genetic architecture of HCM. In our study, we applied a threshold of 0.5% based on the 1000 genomes project dataset that eliminated 90% of the calls and halved the number of distinct variants, while retaining the most likely causative ones. A recently published study[82] analysed the presence and frequency of DCM-associated variants in the *NHLBI* National Heart, Lung and Blood Institute exome sequencing project database and in the dbSNP build 131. The authors proposed a preliminary allele frequency cut-off of 0.04%. If applied to our dataset, this would exclude some clinically important variants, including five disease-causing published sarcomere variants and five published ion-channel disease-causing variants, suggesting that this threshold is too stringent.

Furthermore, we compared our sequencing results with a set of high-depth exomes generated by the UK10K sequencing project (http://www.uk10k.org) and investigated whether we could identify an excess of rare variants in cases that would be consistent with a pathogenic role. Four sarcomeric genes (*MYH7*, *MYBPC3*, *TNNI3*, *TNNT2*) showed a significant excess of rare nsSNPs compared with controls. Assuming a simple dominant model, we estimated that rare nsSNPs in these four genes explained between 12.7% and 53.2% of HCM cases (table 5). Additionally, we proposed a statistical approach that estimates the probability that a nsSNP candidate variant is causal (table 5).

Comparison with previous studies

Studies that have used conventional genetic sequencing techniques to screen patients with HCM suggest that approximately 50–60% of individuals carry mutations in one of eight cardiac sarcomere protein genes.[8 83] This is also the approximate yield described in our centre, using conventional sequencing methodology.[84 85] In this study, the distribution of variants among individual sarcomere protein genes and genes encoding Z-disc and calcium-handling proteins was similar to that reported previously, with the exception of *MYH6* and titin, in which a higher frequency of variants was found.

Table 7 Number of distinct rare variants in genes associated with arrhythmogenic cardiomyopathy and ion channel disease

Genes	Total number of distinct rare variants in each gene	Published	Novel — Misssense variants predicted to be pathogenic in silico	Nonsense, frameshift or splice-site variants predicted to cause loss of function
Genes associated with arrhythmogenic cardiomyopathy				
DSC2	8	5		1
DSG2	9	4		
PKP2	6	3	1	
DSP	16	2	8	1
JUP	3		1	
TMEM43	2			1
TGF-β3	0			
Subtotal	44	14	10	3
Genes associated with ion-channel disease				
KCNQ1	4	1	1	
KCNH2	2			
SCN5A	9	1	5	1
KCNE1	2			
KCNE2	0			
KCNJ2	0			
ANK2	18	3	2	1
CASQ2	2			
CAV3	1	1		
RYR2	12	1	2	4
Subtotal	50	7	10	6
TOTAL	94	21	20	9

In humans, cardiac myosin heavy chain exists as two isoforms, α and β, that are encoded by the tandemly arranged genes *MYH6* and *MYH7*, respectively, situated on chromosome 14. As β-myosin heavy chain is the predominantly expressed isoform in human heart, most studies in patients with HCM have not screened the *MYH6* gene, but evidence that mutations in *MYH6* can cause HCM comes from a few case reports.[86] *MYH6* has also been recently implicated in familial atrial septal defects[64] and sick sinus syndrome.[87] However, in this study, the frequency of rare *MYH6* variants in patients was similar to the control exome sequencing project population, questioning the importance of this gene in HCM. Further functional studies and family evaluation are currently being performed to determine the pathogenicity of the identified SNPs.

Several recent studies have focussed attention on the role of the largest protein in mammals, titin, in heart muscle disease.[88–90] Titin is found in skeletal and cardiac muscle, where it forms an elastic filament bound at the N-terminus to the Z-disc and at the C-terminus to myosin and myosin-binding protein C. The inextensible A-band region of the filament consists of regular patterns of immunoglobulin-like and fibronectin repeats, whereas the I band region is composed of multiple extensible segments (or 'spring' elements) including PEVK, N2A (skeletal and cardiac muscle) and N2B (cardiac only). Titin is encoded by a single gene on chromosome 2 that undergoes complex differential splicing to produce isoforms with variable elastic properties. Titin has a major role in determining the mechanical properties of the heart through its effects on passive tension during myocardial stretch and restoring forces during early ventricular filling, and appears to be an important biomechanical sensor and organisational element within the sarcomere.[88] Nevertheless, titin has been difficult to sequence and study due to its size, large number of isoforms and unsolved tertiary structure. In this study, we identified a large number of novel titin variants in two-thirds of the probands, the majority occurring in association with variants in other genes. Nine of the titin variants, present in 10 patients (4% of the cohort) are predicted to cause loss-of-function, which is more than the proportion of potentially truncating mutations recently reported in a subcohort of patients with HCM.[89] Importantly, the majority of patients carrying truncating and the published mutation do not show any associated rare sarcomere or associated variant, which is strikingly different from the observed when considering all the titin variants. The significance of the large number of SNPs is more difficult to assess. All the individual variants present in our cohort occurred with a frequency less that 0.5% in the 1000 genomes project,[24] suggesting that a proportion of them is, at the very least, modulators of the phenotype. However, the overall frequency of variants in the HCM cohort was actually lower than that seen in the control exome population. This latter finding is difficult to interpret because the annotation of titin variants is made extremely complex by the large number of possible isoforms/transcripts which are not accounted for in existing databases. Further work on understanding the role of titin in HCM is necessary.

Clinical significance of non-sarcomeric variants

Heterozygous mutations in genes encoding desmosomal proteins have been identified in up to 70% of patients with non-syndromic autosomal dominant forms of arrhythmogenic right ventricular cardiomyopathy,[91] and latterly in up to 15% of patients with DCM.[92] In this study, we identified a large number of desmosomal candidate variants, most of which were classified as variants of unknown significance. As with *TTN*, the majority occurred in patients who had at least one sarcomere protein (or related) gene variant, making it difficult to determine their pathogenic role. The same was true for the many variants detected in ion channel genes. Nonetheless, we speculate that the previously published pathogenic mutations in *RYR2*, *ANK2*, *CAV3* and *SCN5A* may be potential phenotype modifiers in HCM, and we are now clinically re-evaluating patients with these variants.

Clinical implications

The targeted HTS protocol used in this study produced similar results to a conventional Sanger sequencing protocol focussed on a small number of sarcomere genes, but also identified large numbers of rare non-synonymous sequence variants in non-sarcomeric genes. These additional data are important as they suggest a possible role for hitherto unsuspected disease-modifying genetic variants in the disease and highlight the challenge that increasing use of HTS will pose for variant interpretation and genetic counselling in everyday clinical practice.

Standard approaches to variant interpretation will increasingly need to be complemented by other strategies and the novel quantitative methods presented in this study provide one way of determining the probability that a variant is disease causing. Additional tools that integrate genetic data with high throughput functional analyses and more sophisticated in silico prediction models coupled with improved clinical phenotyping will also be required.

CONCLUSIONS

A targeted HTS strategy in HCM identifies a large number of nsSNPs in sarcomeric and non-sarcomeric genes. Four sarcomeric genes (*MYH7*, *MYBPC3*, *TNNI3*, *TNNT2*) showed an excess of rare single non-synonymous SNPs (nsSNPs) in cases compared with controls. The frequency of non-sarcomeric variants was similar to the control population, but the clinical significance of individual variants requires further study.

REFERENCES

1 Elliott P, McKenna WJ. Hypertrophic cardiomyopathy. *Lancet* 2004;363:1881–91.
2 Maron BJ, McKenna WJ, Danielson GK, Kappenberger LJ, Kuhn HJ, Seidman CE, Shah PM, Spencer WH III, Spirito P, Ten Cate FJ, Wigle ED. American college of cardiology/european society of cardiology clinical expert consensus document on hypertrophic cardiomyopathy. A report of the american college of cardiology foundation task force on clinical expert consensus documents and the european society of cardiology committee for practice guidelines. *J Am Coll Cardiol* 2003;42:1687–713.
3 Gersh BJ, Maron BJ, Bonow RO, Dearani JA, Fifer MA, Link MS, Naidu SS, Nishimura RA, Ommen SR, Rakowski H, Seidman CE, Towbin JA, Udelson JE, Yancy CW. 2011 accf/aha guideline for the diagnosis and treatment of hypertrophic cardiomyopathy: a report of the american college of cardiology foundation/american heart association task force on practice guidelines. Developed in collaboration with the american association for thoracic surgery, american society of echocardiography, american society of nuclear cardiology, heart failure society of america, heart rhythm society, society for cardiovascular angiography and interventions, and society of thoracic surgeons. *J Am Coll Cardiol* 2011;58:e212–60.
4 Elliott PM, Poloniecki J, Dickie S, Sharma S, Monserrat L, Varnava A, Mahon NG, McKenna WJ. Sudden death in hypertrophic cardiomyopathy: identification of high risk patients. *J Am Coll Cardiol* 2000;36:2212–18.
5 Elliott PM, Gimeno Blanes JR, Mahon NG, Poloniecki JD, McKenna WJ. Relation between severity of left-ventricular hypertrophy and prognosis in patients with hypertrophic cardiomyopathy. *Lancet* 2001;357:420–4.
6 Elliott PM, Gimeno JR, Thaman R, Shah J, Ward D, Dickie S, Tome Esteban MT, McKenna WJ. Historical trends in reported survival rates in patients with hypertrophic cardiomyopathy. *Heart* 2006;92:785–91.
7 Marian AJ, Roberts R. The molecular genetic basis for hypertrophic cardiomyopathy. *J Mol Cell Cardiol* 2001;33:655–70.
8 Morita H, Rehm HL, Menesses A, McDonough B, Roberts AE, Kucherlapati R, Towbin JA, Seidman JG, Seidman CE. Shared genetic causes of cardiac hypertrophy in children and adults. *N Engl J Med* 2008;358:1899–908.
9 Van Driest SL, Ommen SR, Tajik AJ, Gersh BJ, Ackerman MJ. Yield of genetic testing in hypertrophic cardiomyopathy. *Mayo Clin Proc* 2005;80:739–44.
10 Millat G, Bouvagnet P, Chevalier P, Dauphin C, Jouk PS, Da Costa A, Prieur F, Bresson JL, Faivre L, Eicher JC, Chassaing N, Crehalet H, Porcher R, Rodriguez-Lafrasse C, Rousson R. Prevalence and spectrum of mutations in a cohort of 192 unrelated patients with hypertrophic cardiomyopathy. *Eur J Med Genet* 2010;53:261–7.
11 Richard P, Charron P, Carrier L, Ledeuil C, Cheav T, Pichereau C, Benaiche A, Isnard R, Dubourg O, Burban M, Gueffet JP, Millaire A, Desnos M, Schwartz K, Hainque B, Komajda M. Hypertrophic cardiomyopathy: distribution of disease genes, spectrum of mutations, and implications for a molecular diagnosis strategy. *Circulation* 2003;107:2227–32.
12 Geisterfer-Lowrance AA, Kass S, Tanigawa G, Vosberg HP, McKenna W, Seidman CE, Seidman JG. A molecular basis for familial hypertrophic cardiomyopathy: a beta cardiac myosin heavy chain gene missense mutation. *Cell* 1990;62:999–1006.
13 Watkins H, Conner D, Thierfelder L, Jarcho JA, MacRae C, McKenna WJ, Maron BJ, Seidman JG, Seidman CE. Mutations in the cardiac myosin binding protein-c gene on chromosome 11 cause familial hypertrophic cardiomyopathy. *Nat Genet* 1995;11:434–7.
14 Thierfelder L, Watkins H, MacRae C, Lamas R, McKenna W, Vosberg HP, Seidman JG, Seidman CE. Alpha-tropomyosin and cardiac troponin t mutations cause familial hypertrophic cardiomyopathy: a disease of the sarcomere. *Cell* 1994;77:701–12.
15 Poetter K, Jiang H, Hassanzadeh S, Master SR, Chang A, Dalakas MC, Rayment I, Sellers JR, Fananapazir L, Epstein ND. Mutations in either the essential or regulatory light chains of myosin are associated with a rare myopathy in human heart and skeletal muscle. *Nat Genet* 1996;13:63–9.
16 Kimura A, Harada H, Park JE, Nishi H, Satoh M, Takahashi M, Hiroi S, Sasaoka T, Ohbuchi N, Nakamura T, Koyanagi T, Hwang TH, Choo JA, Chung KS, Hasegawa A, Nagai R, Okazaki O, Nakamura H, Matsuzaki M, Sakamoto T, Toshima H, Koga Y, Imaizumi T, Sasazuki T. Mutations in the cardiac troponin i gene associated with hypertrophic cardiomyopathy. *Nat Genet* 1997;16:379–82.
17 Keren A, Syrris P, McKenna WJ. Hypertrophic cardiomyopathy: the genetic determinants of clinical disease expression. *Nat Clin Pract Cardiovasc Med* 2008;5:158–68.
18 Charron P, Arad M, Arbustini E, Basso C, Bilinska Z, Elliott P, Helio T, Keren A, McKenna WJ, Monserrat L, Pankuweit S, Perrot A, Rapezzi C, Ristic A, Seggewiss H, van Langen I, Tavazzi L. Genetic counselling and testing in cardiomyopathies: a position statement of the european society of cardiology working group on myocardial and pericardial diseases. *Eur Heart J* 2010;31:2715–26.
19 Hershberger RE, Lindenfeld J, Mestroni L, Seidman CE, Taylor MR, Towbin JA. Genetic evaluation of cardiomyopathy—a heart failure society of america practice guideline. *J Card Fail* 2009;15:83–97.
20 Metzker ML. Sequencing technologies—the next generation. *Nat Rev Genet* 2010;11:31–46.
21 McKenna WJ, Spirito P, Desnos M, Dubourg O, Komajda M. Experience from clinical genetics in hypertrophic cardiomyopathy: proposal for new diagnostic criteria in adult members of affected families. *Heart* 1997;77:130–2.
22 Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. The sequence alignment/map format and samtools. *Bioinformatics* 2009;25:2078–9.
23 Wang K, Li M, Hakonarson H. Annovar: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 2010;38:e164.
24 1000 Genomes Project Consortium A map of human genome variation from population-scale sequencing. *Nature* 2010;467:1061–73.
25 Smigielski EM, Sirotkin K, Ward M, Sherry ST. Dbsnp: a database of single nucleotide polymorphisms. *Nucleic Acids Res* 2000;28:352–5.
26 Ng PC, Henikoff S. Sift: predicting amino acid changes that affect protein function. *Nucleic Acids Res* 2003;31:3812–14.
27 Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. A method and server for predicting damaging missense mutations. *Nat Methods* 2010;7:248–9.
28 Brito D, Madeira H. Malignant mutations in hypertrophic cardiomyopathy: fact or fancy? *Rev Port Cardiol* 2005;24:1137–46.
29 Maron BJ, Maron MS, Semsarian C. Double or compound sarcomere mutations in hypertrophic cardiomyopathy: a potential link to sudden death in the absence of conventional risk factors. *Heart Rhythm* 2012;9:57–63.
30 Niimura H, Patton KK, McKenna WJ, Soults J, Maron BJ, Seidman JG, Seidman CE. Sarcomere protein gene mutations in hypertrophic cardiomyopathy of the elderly. *Circulation* 2002;105:446–51.
31 Morner S, Richard P, Kazzam E, Hellman U, Hainque B, Schwartz K, Waldenstrom A. Identification of the genotypes causing hypertrophic cardiomyopathy in northern sweden. *J Mol Cell Cardiol* 2003;35:841–9.
32 Van Driest SL, Vasile VC, Ommen SR, Will ML, Tajik AJ, Gersh BJ, Ackerman MJ. Myosin binding protein c mutations and compound heterozygosity in hypertrophic cardiomyopathy. *J Am Coll Cardiol* 2004;44:1903–10.
33 Girolami F, Olivotto I, Passerini I, Zachara E, Nistri S, Re F, Fantini S, Baldini K, Torricelli F, Cecchi F. A molecular screening strategy based on beta-myosin heavy chain, cardiac myosin binding protein c and troponin t genes in italian patients with hypertrophic cardiomyopathy. *J Cardiovasc Med (Hagerstown)* 2006;7:601–7.
34 Niimura H, Bachinski LL, Sangwatanaroj S, Watkins H, Chudley AE, McKenna W, Kristinsson A, Roberts R, Sole M, Maron BJ, Seidman JG, Seidman CE. Mutations in the gene for cardiac myosin-binding protein c and late-onset familial hypertrophic cardiomyopathy. *N Engl J Med* 1998;338:1248–57.
35 Zimmerman RS, Cox S, Lakdawala NK, Cirino A, Mancini-DiNardo D, Clark E, Leon A, Duffy E, White E, Baxter S, Alaamery M, Farwell L, Weiss S, Seidman CE, Seidman JG, Ho CY, Rehm HL, Funke BH. A novel custom resequencing array for dilated cardiomyopathy. *Genet Med* 2010;12:268–78.
36 Millat G, Chanavat V, Crehalet H, Rousson R. Development of a high resolution melting method for the detection of genetic variations in hypertrophic cardiomyopathy. *Clin Chim Acta* 2010;411:1983–91.
37 Nanni L, Pieroni M, Chimenti C, Simionati B, Zimbello R, Maseri A, Frustaci A, Lanfranchi G. Hypertrophic cardiomyopathy: two homozygous cases with "Typical" Hypertrophic cardiomyopathy and three new mutations in cases with progression to dilated cardiomyopathy. *Biochem Biophys Res Commun* 2003;309:391–8.
38 Konno T, Shimizu M, Ino H, Matsuyama T, Yamaguchi M, Terai H, Hayashi K, Mabuchi T, Kiyama M, Sakata K, Hayashi T, Inoue M, Kaneda T, Mabuchi H. A novel missense mutation in the myosin binding protein-c gene is responsible for hypertrophic cardiomyopathy with left ventricular dysfunction and dilation in elderly patients. *J Am Coll Cardiol* 2003;41:781–6.
39 Watkins H, McKenna WJ, Thierfelder L, Suk HJ, Anan R, O'Donoghue A, Spirito P, Matsumori A, Moravec CS, Seidman JG, Seidman C. Mutations in the genes for cardiac troponin T and alpha-tropomyosin in hypertrophic cardiomyopathy. *N Engl J Med* 1995;332:1058–64.
40 Moolman JC, Corfield VA, Posen B, Ngumbela K, Seidman C, Brink PA, Watkins H. Sudden death due to troponin t mutations. *J Am Coll Cardiol* 1997;29:549–55.
41 Yamauchi-Takihara K, Nakajima-Taniguchi C, Matsui H, Fujio Y, Kunisada K, Nagata S, Kishimoto T. Clinical implications of hypertrophic cardiomyopathy associated with mutations in the alpha-tropomyosin gene. *Heart* 1996;76:63–5.
42 Morita H, Larson MG, Barr SC, Vasan RS, O'Donnell CJ, Hirschhorn JN, Levy D, Corey D, Seidman CE, Seidman JG, Benjamin EJ. Single-gene mutations and increased left ventricular wall thickness in the community: the framingham heart study. *Circulation* 2006;113:2697–705.
43 Alpert NR, Mohiddin SA, Tripodi D, Jacobson-Hatzell J, Vaughn-Whitley K, Brosseau C, Warshaw DM, Fananapazir L. Molecular and phenotypic effects of

heterozygous, homozygous, and compound heterozygote myosin heavy-chain mutations. *Am J Physiol Heart Circ Physiol* 2005;288:H1097–102.

44 Watkins H, Rosenzweig A, Hwang DS, Levi T, McKenna W, Seidman CE, Seidman JG. Characteristics and prognostic implications of myosin missense mutations in familial hypertrophic cardiomyopathy. *N Engl J Med* 1992;326:1108–14.

45 Gruver EJ, Fatkin D, Dodds GA, Kisslo J, Maron BJ, Seidman JG, Seidman CE. Familial hypertrophic cardiomyopathy and atrial fibrillation caused by arg663his beta-cardiac myosin heavy chain mutation. *Am J Cardiol* 1999;83:13H–18H.

46 Kaski JP, Syrris P, Esteban MT, Jenkins S, Pantazis A, Deanfield JE, McKenna WJ, Elliott PM. Prevalence of sarcomere protein gene mutations in preadolescent children with hypertrophic cardiomyopathy. *Circ Cardiovasc Genet* 2009;2:436–41.

47 Fananapazir L, Dalakas MC, Cyran F, Cohn G, Epstein ND. Missense mutations in the beta-myosin heavy-chain gene cause central core disease in hypertrophic cardiomyopathy. *Proc Natl Acad Sci USA* 1993;90:3993–7.

48 Woo A, Rakowski H, Liew JC, Zhao MS, Liew CC, Parker TG, Zeller M, Wigle ED, Sole MJ. Mutations of the beta myosin heavy chain gene in hypertrophic cardiomyopathy: critical functional sites determine prognosis. *Heart* 2003;89:1179–85.

49 Erdmann J, Daehmlow S, Wischke S, Senyuva M, Werner U, Raible J, Tanis N, Dyachenko S, Hummel M, Hetzer R, Regitz-Zagrosek V. Mutation spectrum in a large cohort of unrelated consecutive patients with hypertrophic cardiomyopathy. *Clin Genet* 2003;64:339–49.

50 Consevage MW, Salada GC, Baylen BG, Ladda RL, Rogan PK. A new missense mutation, arg719gln, in the beta-cardiac heavy chain myosin gene of patients with familial hypertrophic cardiomyopathy. *Hum Mol Genet* 1994;3:1025–6.

51 Mohiddin SA, Begley DA, McLam E, Cardoso JP, Winkler JB, Sellers JR, Fananapazir L. Utility of genetic screening in hypertrophic cardiomyopathy: prevalence and significance of novel and double (homozygous and heterozygous) beta-myosin mutations. *Genet Test* 2003;7:21–7.

52 Bos JM, Poley RN, Ny M, Tester DJ, Xu X, Vatta M, Towbin JA, Gersh BJ, Ommen SR, Ackerman MJ. Genotype-phenotype relationships involving hypertrophic cardiomyopathy-associated mutations in titin, muscle lim protein, and telethonin. *Mol Genet Metab* 2006;88:78–85.

53 Mogensen J, Kubo T, Duque M, Uribe W, Shaw A, Murphy R, Gimeno JR, Elliott P, McKenna WJ. Idiopathic restrictive cardiomyopathy is part of the clinical expression of cardiac troponin i mutations. *J Clin Invest* 2003;111:209–16.

54 Refaat MM, Lubitz SA, Makino S, Islam Z, Frangiskakis JM, Mehdi H, Gutmann R, Zhang ML, Bloom HL, MacRae CA, Dudley SC, Shalaby AA, Weiss R, McNamara DM, London B, Ellinor PT. Genetic variation in the alternative splicing regulator rbm20 is associated with dilated cardiomyopathy. *Heart Rhythm* 2012;9:390–6.

55 Knoll R, Hoshijima M, Hoffman HM, Person V, Lorenzen-Schmidt I, Bang ML, Hayashi T, Shiga N, Yasukawa H, Schaper W, McKenna W, Yokoyama M, Schork NJ, Omens JH, McCulloch AD, Kimura A, Gregorio CC, Poller W, Schaper J, Schultheiss HP, Chien KR. The cardiac mechanical stretch sensor machinery involves a z disc complex that is defective in a subset of human dilated cardiomyopathy. *Cell* 2002;111:943–55.

56 Charron P, Dubourg O, Desnos M, Bennaceur M, Carrier L, Camproux AC, Isnard R, Hagege A, Langlard JM, Bonne G, Richard P, Hainque B, Bouhour JB, Schwartz K, Komajda M. Clinical features and prognostic implications of familial hypertrophic cardiomyopathy related to the cardiac myosin-binding protein c gene. *Circulation* 1998;97:2230–6.

57 Lakdawala NK, Thune JJ, Maron BJ, Cirino AL, Havndrup O, Bundgaard H, Christiansen M, Carlsen CM, Dorval JF, Kwong RY, Colan SD, Kober LV, Ho CY. Electrocardiographic features of sarcomere mutation carriers with and without clinically overt hypertrophic cardiomyopathy. *Am J Cardiol* 2011;108:1606–13.

58 Ingles J, Doolan A, Chiu C, Seidman J, Seidman C, Semsarian C. Compound and double mutations in patients with hypertrophic cardiomyopathy: implications for genetic testing and counselling. *J Med Genet* 2005;42:e59.

59 Laredo R, Monserrat L, Hermida-Prieto M, Fernandez X, Rodriguez I, Cazon L, Alvarino I, Dumont C, Pinon P, Peteiro J, Bouzas B, Castro-Beiras A. [beta-myosin heavy-chain gene mutations in patients with hypertrophic cardiomyopathy]. *Rev Esp Cardiol* 2006;59:1008–18.

60 Ohsuzu F, Katsushika S, Akanuma M, Nakamura H, Harada H, Satoh M, Hiroi S, Kimura A. Hypertrophic obstructive cardiomyopathy due to a novel t-to-a transition at codon 624 in the beta-myosin heavy chain (beta-mhc) gene possibly related to the sudden death. *Int J Cardiol* 1997;62:203–9.

61 Olivotto I, Girolami F, Ackerman MJ, Nistri S, Bos JM, Zachara E, Ommen SR, Theis JL, Vaubel RA, Re F, Armentano C, Poggesi C, Torricelli F, Cecchi F. Myofilament protein gene mutation screening and outcome of patients with hypertrophic cardiomyopathy. *Mayo Clin Proc* 2008;83:630–8.

62 Waldmuller S, Muller M, Rackebrandt K, Binner P, Poths S, Bonin M, Scheffold T. Array-based resequencing assay for mutations causing hypertrophic cardiomyopathy. *Clin Chem* 2008;54:682–7.

63 Gimeno JR, Monserrat L, Perez-Sanchez I, Marin F, Caballero L, Hermida-Prieto M, Castro A, Valdes M. Hypertrophic cardiomyopathy. A study of the troponin-t gene in 127 spanish families. *Rev Esp Cardiol* 2009;62:1473–7.

64 Posch MG, Waldmuller S, Muller M, Scheffold T, Fournier D, Andrade-Navarro MA, De Geeter B, Guillaumont S, Dauphin C, Yousseff D, Schmitt KR, Perrot A, Berger F,

65 Hetzer R, Bouvagnet P, Ozcelik C. Cardiac alpha-myosin (myh6) is the predominant sarcomeric disease gene for familial atrial septal defects. *PLoS One* 2011;6:e28872.

65 Arimura T, Bos JM, Sato A, Kubo T, Okamoto H, Nishi H, Harada H, Koga Y, Moulik M, Doi YL, Towbin JA, Ackerman MJ, Kimura A. Cardiac ankyrin repeat protein gene (ankrd1) mutations in hypertrophic cardiomyopathy. *J Am Coll Cardiol* 2009;54:334–42.

66 Quarta G, Muir A, Pantazis A, Syrris P, Gehmlich K, Garcia-Pavia P, Ward D, Sen-Chowdhry S, Elliott PM, McKenna WJ. Familial evaluation in arrhythmogenic right ventricular cardiomyopathy: impact of genetics and revised task force criteria. *Circulation* 2011;123:2701–9.

67 Klauke B, Kossmann S, Gaertner A, Brand K, Stork I, Brodehl A, Dieding M, Walhorn V, Anselmetti D, Gerdes D, Bohms B, Schulz U, Zu Knyphausen E, Vorgerd M, Gummert J, Milting H. De novo desmin-mutation n116s is associated with arrhythmogenic right ventricular cardiomyopathy. *Hum Mol Genet* 2010;19:4595–607.

68 Fressart V, Duthoit G, Donal E, Probst V, Deharo JC, Chevalier P, Klug D, Dubourg O, Delacretaz E, Cosnay P, Scanu P, Extramiana F, Keller D, Hidden-Lucet F, Simon F, Bessirard V, Roux-Buisson N, Hebert JL, Azarine A, Casset-Senon D, Rouzet F, Lecarpentier Y, Fontaine G, Coirault C, Frank R, Hainque B, Charron P. Desmosomal gene analysis in arrhythmogenic right ventricular dysplasia/cardiomyopathy: spectrum of mutations and clinical impact in practice. *Europace* 2010;12:861–8.

69 Basso C, Czarnowska E, Della Barbera M, Bauce B, Beffagna G, Wlodarska EK, Pilichou K, Ramondo A, Lorenzon A, Wozniek O, Corrado D, Daliento L, Danieli GA, Valente M, Nava A, Thiene G, Rampazzo A. Ultrastructural evidence of intercalated disc remodelling in arrhythmogenic right ventricular cardiomyopathy: an electron microscopy investigation on endomyocardial biopsies. *Eur Heart J* 2006;27:1847–54.

70 van der Zwaag PA, Jongbloed JD, van den Berg MP, van der Smagt JJ, Jongbloed R, Bikker H, Hofstra RM, van Tintelen JP. A genetic variants database for arrhythmogenic right ventricular dysplasia/cardiomyopathy. *Hum Mutat* 2009;30:1278–83.

71 Medeiros-Domingo A, Bhuiyan ZA, Tester DJ, Hofman N, Bikker H, van Tintelen JP, Mannens MM, Wilde AA, Ackerman MJ. The ryr2-encoded ryanodine receptor/ calcium release channel in patients diagnosed previously with either catecholaminergic polymorphic ventricular tachycardia or genotype negative, exercise-induced long qt syndrome: a comprehensive open reading frame mutational analysis. *J Am Coll Cardiol* 2009;54:2065–74.

72 Mohler PJ, Le Scouarnec S, Denjoy I, Lowe JS, Guicheney P, Caron L, Driskell IM, Schott JJ, Norris K, Leenhardt A, Kim RB, Escande D, Roden DM. Defining the cellular phenotype of "Ankyrin-b syndrome" Variants: human ank2 variants associated with clinical phenotypes display a spectrum of activities in cardiomyocytes. *Circulation* 2007;115:432–41.

73 Mohler PJ, Splawski I, Napolitano C, Bottelli G, Sharpe L, Timothy K, Priori SG, Keating MT, Bennett V. A cardiac arrhythmia syndrome caused by loss of ankyrin-b function. *Proc Natl Acad Sci USA* 2004;101:9137–42.

74 Wattanasirichaigoon D, Vesely MR, Duggal P, Levine JC, Blume ED, Wolff GS, Edwards SB, Beggs AH. Sodium channel abnormalities are infrequent in patients with long qt syndrome: identification of two novel scn5a mutations. *Am J Med Genet* 1999;86:470–6.

75 Sherman J, Tester DJ, Ackerman MJ. Targeted mutational analysis of ankyrin-b in 541 consecutive, unrelated patients referred for long qt syndrome genetic testing and 200 healthy subjects. *Heart Rhythm* 2005;2:1218–23.

76 Cronk LB, Ye B, Kaku T, Tester DJ, Vatta M, Makielski JC, Ackerman MJ. Novel mechanism for sudden infant death syndrome: persistent late sodium current secondary to mutations in caveolin-3. *Heart Rhythm* 2007;4:161–6.

77 Garcia-Pavia P, Syrris P, Salas C, Evans A, Mirelis JG, Cobo-Marcos M, Vilches C, Bornstein B, Segovia J, Alonso-Pulpon L, Elliott PM. Desmosomal protein gene mutations in patients with idiopathic dilated cardiomyopathy undergoing cardiac transplantation: a clinicopathological study. *Heart* 2011;97:1744–52.

78 den Haan AD, Tan BY, Zikusoka MN, Llado LI, Jain R, Daly A, Tichnell C, James C, Amat-Alarcon N, Abraham T, Russell SD, Bluemke DA, Calkins H, Dalal D, Judge DP. Comprehensive desmosome mutation analysis in north americans with arrhythmogenic right ventricular dysplasia/cardiomyopathy. *Circ Cardiovasc Genet* 2009;2:428–35.

79 Kapplinger JD, Landstrom AP, Salisbury BA, Callis TE, Pollevick GD, Tester DJ, Cox MG, Bhuiyan Z, Bikker H, Wiesfeld AC, Hauer RN, van Tintelen JP, Jongbloed JD, Calkins H, Judge DP, Wilde AA, Ackerman MJ. Distinguishing arrhythmogenic right ventricular cardiomyopathy/dysplasia-associated mutations from background genetic noise. *J Am Coll Cardiol* 2011;57:2317–27.

80 Syrris P, Ward D, Asimaki A, Evans A, Sen-Chowdhry S, Hughes SE, McKenna WJ. Desmoglein-2 mutations in arrhythmogenic right ventricular cardiomyopathy: a genotype-phenotype characterization of familial disease. *Eur Heart J* 2007;28:581–8.

81 Millat G, Chevalier P, Restier-Miron L, Da Costa A, Bouvagnet P, Kugener B, Fayol L, Gonzalez Armengod C, Oddou B, Chanavat V, Froidefond E, Perraudin R, Rousson R, Rodriguez-Lafrasse C. Spectrum of pathogenic mutations and associated

polymorphisms in a cohort of 44 unrelated patients with long qt syndrome. *Clin Genet* 2006;70:214–27.

82 Norton N, Robertson PD, Rieder MJ, Zuchner S, Rampersaud E, Martin E, Li D, Nickerson DA, Hershberger RE. Evaluating pathogenicity of rare variants from dilated cardiomyopathy in the exome era. *Circ Cardiovasc Genet* 2012;5:167–74.

83 Bos JM, Towbin JA, Ackerman MJ. Diagnostic, prognostic, and therapeutic implications of genetic testing for hypertrophic cardiomyopathy. *J Am Coll Cardiol* 2009;54:201–11.

84 Page SP, Kounas S, Syrris P, Christiansen M, Frank-Hansen R, Anderson PS, Elliott PM, McKenna WJ. Cardiac myosin binding protein-c mutations in families 1396 with hypertrophic cardiomyopathy: disease expression in relation to age, gender, 1397 and long term outcome. *Circ Cardiovasc Genet* 2012. 5(2):156–66

85 Pasquale F, Syrris P, Kaski JP, Mogensen J, McKenna WJ, Elliott P. Long-term outcomes in hypertrophic cardiomyopathy caused by mutations in the cardiac troponin t gene. *Circ Cardiovasc Genet* 2012;5:10–17.

86 Carniel E, Taylor MR, Sinagra G, Di Lenarda A, Ku L, Fain PR, Boucek MM, Cavanaugh J, Miocic S, Slavov D, Graw SL, Feiger J, Zhu XZ, Dao D, Ferguson DA, Bristow MR, Mestroni L. Alpha-myosin heavy chain: a sarcomeric gene associated with dilated and hypertrophic phenotypes of cardiomyopathy. *Circulation* 2005;112:54–9.

87 Holm H, Gudbjartsson DF, Sulem P, Masson G, Helgadottir HT, Zanon C, Magnusson OT, Helgason A, Saemundsdottir J, Gylfason A, Stefansdottir H, Gretarsdottir S, Matthiasson SE, Thorgeirsson GM, Jonasdottir A, Sigurdsson A, Stefansson H, Werge T, Rafnar T, Kiemeney LA, Parvez B, Muhammad R, Roden DM, Darbar D, Thorleifsson G, Walters GB, Kong A, Thorsteinsdottir U, Arnar DO, Stefansson K. A rare variant in myh6 is associated with high risk of sick sinus syndrome. *Nat Genet* 2011;43:316–20.

88 LeWinter MM, Granzier H. Cardiac titin: a multifunctional giant. *Circulation* 2010;121:2137–45.

89 Herman DS, Lam L, Taylor MR, Wang L, Teekakirikul P, Christodoulou D, Conner L, DePalma SR, McDonough B, Sparks E, Teodorescu DL, Cirino AL, Banner NR, Pennell DJ, Graw S, Merlo M, Di Lenarda A, Sinagra G, Bos JM, Ackerman MJ, Mitchell RN, Murry CE, Lakdawala NK, Ho CY, Barton PJ, Cook SA, Mestroni L, Seidman JG, Seidman CE. Truncations of titin causing dilated cardiomyopathy. *N Engl J Med* 2012;366:619–28.

90 Gautel M. The sarcomeric cytoskeleton: who picks up the strain? *Curr Opin Cell Biol* 2011;23:39–46.

91 Sen-Chowdhry S, Morgan RD, Chambers JC, McKenna WJ. Arrhythmogenic cardiomyopathy: Etiology, diagnosis, and treatment. *Annu Rev Med* 2010;61:233–53.

92 Elliott P, O'Mahony C, Syrris P, Evans A, Rivera Sorensen C, Sheppard MN, Carr-White G, Pantazis A, McKenna WJ. Prevalence of desmosomal protein gene mutations in patients with dilated cardiomyopathy. *Circ Cardiovasc Genet* 2010;3:314–22.