



OPEN ACCESS

ORIGINAL ARTICLE

# Targeted sequence capture and GS-FLX Titanium sequencing of 23 hypertrophic and dilated cardiomyopathy genes: implementation into diagnostics

Olaf R F Mook,<sup>1</sup> Martin A Haagmans,<sup>1</sup> Jean-François Soucy,<sup>2</sup> Judith B A van de Meerakker,<sup>1</sup> Frank Baas,<sup>3</sup> Marja E Jakobs,<sup>3</sup> Nynke Hofman,<sup>1</sup> Imke Christiaans,<sup>1</sup> Ronald H Lekanne Deprez,<sup>1</sup> Marcel M A M Mannens<sup>1</sup>

► Additional material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/jmedgenet-2012-101231>).

<sup>1</sup>Department of Clinical Genetics, Academic Medical Center, Amsterdam, The Netherlands

<sup>2</sup>Medical Genetics Division, CHU Sainte-Justine, Montréal, Canada

<sup>3</sup>Department of Genome Analysis, Academic Medical Center, Amsterdam, The Netherlands

## Correspondence to

Dr O R F Mook, Department of Clinical Genetics, Academic Medical Center, Meibergdreef 9, Amsterdam 1105 AZ, The Netherlands; [o.r.mook@amc.uva.nl](mailto:o.r.mook@amc.uva.nl)

MAH, J-FS and JBAM contributed equally.

Received 10 September 2012

Revised 24 May 2013

Accepted 27 May 2013

Published Online First

19 June 2013

## ABSTRACT

**Background** Genetic evaluation of cardiomyopathies poses a challenge. Multiple genes are involved but no clear genotype–phenotype correlations have been found so far. In the past, genetic evaluation for hypertrophic (HCM) and dilated (DCM) cardiomyopathies was performed by sequential screening of a very limited number of genes. Recent developments in sequencing have increased the throughput, enabling simultaneous screening of multiple genes for multiple patients in a single sequencing run.

**Objective** Development and implementation of a next generation sequencing (NGS) based genetic test as replacement for Sanger sequencing.

**Methods and Results** In order to increase the number of genes that can be screened in a shorter time period, we enriched all exons of 23 of the most relevant HCM and DCM related genes using on-array multiplexed sequence capture followed by massively parallel pyrosequencing on the GS-FLX Titanium. After optimisation of array based sequence capture it was feasible to reliably detect a large panel of known and unknown variants in HCM and DCM patients, whereby the unknown variants could be confirmed by Sanger sequencing.

**Conclusions** The rate of detection of (pathogenic) variants in both HCM and DCM patients was increased due to a larger number of genes studied. Array based target enrichment followed by NGS showed the same accuracy as Sanger sequencing. Therefore, NGS is ready for implementation in a diagnostic setting.

## INTRODUCTION

Genetic cardiomyopathies represent a group of diseases that are caused by alterations in over 50 genes lacking clear genotype–phenotype correlations.<sup>1–3</sup> As sequential testing of all those genes is both time consuming and expensive, a better testing strategy should therefore be cheaper and have shorter turnaround times. Over time, different approaches have been developed. In our laboratory, cardiomyopathy genes were tested using denaturing high performance liquid chromatography (dHPLC) followed by Sanger sequencing and Sanger sequencing without dHPLC as a pre-screen. Array based sequencing is a technique developed in the last few years that results in higher throughput, diminished costs, and shorter turnaround times.<sup>4</sup> However, this technology is less

sensitive to insertion/deletion detection, is less amenable to panel modification (by adding new genes, for example), and has a very limited sequencing capacity in comparison to Illumina GAI sequencing, for example.<sup>5</sup> Next generation sequencing (NGS) is an attractive approach for high throughput screening of a large number of genes in a time and cost effective manner. At the moment the most popular choices are the Mi/HiSeq (Illumina), SOLiD (Life Technologies) and GS-FLX Titanium (Roche). Gowrisankar *et al*<sup>5</sup> published their results on using the Illumina GAI for the sequencing of pooled concatenated PCR products from five patients with dilated cardiomyopathy (DCM) who were previously analysed on their sequencing array. They concluded that NGS showed an analytical sensitivity that outperformed array based sequencing, mainly because of better indel detection, but it had a higher turnaround time in their hands. Meder *et al*<sup>6</sup> used the SOLiD 3 to test for mutations in 47 cardiomyopathy related and candidate genes in 10 patients diagnosed with either hypertrophic cardiomyopathy (HCM) or DCM. They concluded that an array based enrichment combined with SOLiD sequencing of 47 genes can be performed with high accuracy (but with known difficulties in detection of indels using short reads<sup>7</sup>) and in a cost effective manner.

For our study, we selected GS-FLX Titanium because of significantly longer reads, short run times and easy data analysis, making the transition to use in a diagnostic setting easier without the need for complex bioinformatics. Furthermore, the long reads of GS-FLX Titanium allow for more accurate mapping (pseudogene discrimination), detection of small indels, and can compensate for small untargeted regions of capture arrays.

In the present study we show that array based sequence capture can be optimised for balanced exon coverage. Optimised sequence capture in combination with GS-FLX Titanium sequencing allows for accurate detection of variants in 23 genes involved in HCM and DCM.

## METHODS

### Subjects and clinical evaluation

This study comprises an initial pilot study followed by additional experiments in which a total of 63



Open Access  
Scan to access more  
free content

**To cite:** Mook ORF, Haagmans MA, Soucy J-F, *et al.* *J Med Genet* 2013;**50**:614–626.

patients were included. All patients have a confirmed diagnosis of HCM or DCM according to international criteria and were identified at or referred to the Department of Clinical Genetics, Academic Medical Center (AMC), Amsterdam, for screening of cardiomyopathy related genes offered at the Department of DNA diagnostics, AMC, the Netherlands.

For an initial pilot we included five HCM patients with a confirmed pathogenic mutation in either the *MYBPC3* or *MYH7* gene. In an additional experiment we included nine probands who were diagnosed with HCM but had no pathogenic mutation in eight HCM genes (*MYBPC3*, *MYH7*, *MYL2*, *MYL3*, *TNNI3*, *TNNT2*, *TPM1*, and *GLA*) screened in our laboratory by Sanger sequencing. Furthermore, we also included 19 probands diagnosed with DCM. However, DCM patients were not routinely screened for all these genes, in contrast to HCM patients. Finally, we included 30 cardiomyopathy index patients who were registered for routine DNA diagnostics for all HCM genes. These patients were used to validate the entire procedure with the intention to implement the procedure for diagnostics. Informed consent was obtained from all the patients.

### Array design

We designed four custom 385K oligonucleotide microarrays according the Nimblegen Rebal algorithm to select unique probes (Roche NimbleGen). In the first design we targeted all exons, including 100 bp of the flanking upstream and downstream intronic sequence, of 18 genes with known involvement in HCM and/or DCM. The design comprises 236 exons, targeting 92.5 kb (table 1). In the second design we rebalanced the capture probes by the addition of replicate probes at regions with low coverage in order to obtain a more evenly distributed exon coverage. For regions with a coverage between 75% and 61% of the mean coverage we added one additional probe, between 60% and 51% we added two additional probes, between 50% and

41% we added three additional probes, between 40% and 31% we added four additional probes, between 30% and 21% we added five additional probes, and for  $\leq 20\%$  we added six additional probes. In the third design we fine-tuned the balancing and added five more genes, making the total number of genes 23, containing 292 exons and targeting 117.5 kb. These 23 genes were selected because they were already used in diagnostics supplemented with candidate genes that were selected based on published evidence, with a focus on mutation detection rate. In the fourth design we rebalanced the last five genes that were added (table 1).

A schematic overview of the different versions of the arrays used, the hybridisation protocol used, the array design used for a patient group and its improved performance as represented in coverage statistics is provided in online supplementary figure S1.

### Sample preparation

DNA was isolated from peripheral blood leucocytes using an automated DNA isolator (Gentra). Then 0.5–5  $\mu\text{g}$  DNA was fragmented according to the manufacturer's instructions (Covaris). DNA quality was assessed by running the samples on a DNA7500 chip on the Bioanalyzer (Agilent). Each sample was bar coded by ligation of GSMID-adaptors or RL-multiplex identifiers (MID) adapters using standard Roche protocols. Libraries were amplified by linker mediated PCR to obtain sufficient amounts of starting material for the sequence capture; 3  $\mu\text{g}$  of amplified library was loaded onto the array according to manufacturer's instructions ([http://www.nimblegen.com/products/lit/SeqCap\\_UserGuide\\_Titanium\\_Delivery\\_v1p1.pdf](http://www.nimblegen.com/products/lit/SeqCap_UserGuide_Titanium_Delivery_v1p1.pdf)). For on-array multiplexed samples, 5–10 amplified libraries were mixed equimolarly and then 3  $\mu\text{g}$  of the mixture was loaded onto the array. For multiplex experiments an equimolar pool of bar code specific blocker oligo's was added in the same concentration as used for non-multiplexed hybridisations.

**Table 1** Genes present on the capture array

Genes on V1/V2	Genes on V3	Chromosome	Exons	NM_number	Location/function
	<i>ACTC1</i>	15q14	6	NM_005159.4	Sarcomere
<i>CSRP3</i>	<i>CSRP3</i>	11p15.1	5	NM_003476.3	Z disk
<i>DES</i>	<i>DES</i>	2q35	9	NM_001927.3	Intermediate filament
<i>EMD</i>	<i>EMD</i>	Xq28	6	NM_000117.2	Nuclear lamina
<i>GLA</i>	<i>GLA</i>	Xq22.1	7	NM_000169.2	Lysosome function
<i>LAMP2</i>	<i>LAMP2</i>	Xq24	11	NM_002294.2 NM_013995.2 NM_001122606.1	Lysosome function
	<i>LDB3</i>	10q23.2	13	NM_007078.2 NM_001080116.1	Z disk
<i>LMNA</i>	<i>LMNA</i>	1q22	12	NM_170707.2 ENST00000368294 ENST00000368297	Nuclear lamina
<i>MYBPC3</i>	<i>MYBPC3</i>	11p11.2	34	NM_000256.3	Sarcomere
<i>MYH7</i>	<i>MYH7</i>	14q11.2	37	NM_000257.2	Sarcomere
<i>MYL2</i>	<i>MYL2</i>	12q24.11	7	NM_000432.3	Sarcomere
<i>MYL3</i>	<i>MYL3</i>	3p21.31	7	NM_000258.2	Sarcomere
	<i>PLN</i>	6q22.31	1	NM_002667.3	Cardiac Ca <sup>++</sup> ATPase inhibitor
<i>PRKAG2</i>	<i>PRKAG2</i>	7q36.1	16	NM_016203.3	AMP activated protein kinase
<i>SCN5A</i>	<i>SCN5A</i>	3p22.2	28	NM_198056.2 NM_001099404.1 NM_001160160.1	Sodium ion channel
	<i>SGCD</i>	5q33.2	8	NM_172244 NM_000337	Dystrophin associated protein complex
<i>TAZ</i>	<i>TAZ</i>	Xq28	11	NM_000116.3	Mitochondrial cardiolipin transacylase
<i>TCAP</i>	<i>TCAP</i>	17q12	2	NM_003673.2	Z disk
<i>TNNC1</i>	<i>TNNC1</i>	3p21.1	6	NM_003280.1	Sarcomere
<i>TNNI3</i>	<i>TNNI3</i>	19q13.42	8	NM_000363.4	Sarcomere
<i>TNNT2</i>	<i>TNNT2</i>	1q32.1	15	NM_000364.2 NM_001001430.1	Sarcomere
<i>TPM1</i>	<i>TPM1</i>	15q22.2	9	NM_000366.5 NM_001018005.1 NM_001018020.1	Sarcomere
	<i>VCL</i>	10q22.2	21	NM_014000.2	Z disk

### Target enrichment and sequencing

Hybridisation, post-hybridisation washes and elution of the enriched sample was performed according to the manufacturer's instructions ([http://www.nimblegen.com/products/lit/SeqCap\\_UserGuide\\_Titanium\\_Delivery\\_v1p1.pdf](http://www.nimblegen.com/products/lit/SeqCap_UserGuide_Titanium_Delivery_v1p1.pdf)). To increase target enrichment, the enriched samples were hybridised a second time on the same array used the first time. This second hybridisation reduces background sequences (off-target sequences) that are still present after the first hybridisation. It is the reduction of background that results in higher on-target percentages. To verify successful hybridisation capture, we performed quantitative PCR (qPCR) on four control loci before and after array enrichment. The relative fold enrichment was calculated using the values of delta crossing point (CP) (ie, the difference between average CP of non-captured and average CP of captured samples) according  $E^{\delta\text{-CP}}$  where E is the efficiency of the qPCR assay for a particular amplicon.

The enriched library was diluted, annealed to capture beads, and clonally amplified by emulsion PCR. After emulsion PCR, beads with clonal amplicons were enriched and deposited on a picotitre plate and sequenced on the GS-FLX Titanium.

### Mapping, variant detection and classification

The obtained sequence reads were mapped against the human reference genome (hg19) with the Roche GS Reference Mapper (V2.6) using the default software settings. Output was restricted to the targeted regions as defined by the sequence capture design. Coverage statistics were extracted from the mapping output files using custom scripts. Variants were automatically detected during mapping and annotated with known gene (refSeq genes from <http://genome.ucsc.edu/>) and single nucleotide polymorphism (SNP) information (dbSNP130 from <http://genome.ucsc.edu/>). They were denoted as high quality differences (HCDiffs) when the variation was present in at least three non-duplicate reads that included at least one forward and one reverse read, or when it was seen in at least five reads with quality scores over 20. Variants that did not meet the above criteria were collected in the AllDiff files. Occasionally, known variants end up in the AllDiff table but not in the HCDiff table. A custom script was written for the comparison of the HCDiff and AllDiff variants and selection of variants with a variant percentage  $\geq 20\%$ .

The script also identifies, at the single base resolution, regions with a coverage lower than  $16\times$ . These regions are additionally analysed by Sanger sequencing. This threshold has been reported by Hoischen *et al*<sup>8</sup> to be sufficient for diagnostic testing. We also calculated the minimal number of reads needed statistically. For the statistics we have used the following criteria:

1. for a heterozygous variant the allele frequency is 50%
2. a variant is reported when the variant percentage is  $\geq 20\%$ .

All individual chances that a variant is missed at a given coverage is calculated in R using:

$$x\text{---seq}(300) \text{ and } c\text{---pbinom}((x*0.2),x,0.5).$$

With these criteria we calculated a 99% sensitivity (comparable to a Phred quality score of 20) at  $16\times$ . The frequency of each individual coverage given a mean coverage  $\pm$ SD is calculated in R using `y <- pnorm(x, mean coverage, SD)` with `x <- seq(300)`.

Then the chance that a variant is missed in an experiment with mean coverage  $\pm$ SD is calculated as: sum of (chance variant missed at given coverage)  $\times$  (frequency at given coverage). Since we use a  $16\times$  threshold, the chance that a variant is missed is calculated for the  $16\times$  to  $300\times$  coverage interval. At a coverage

of  $100\pm 35\times$  this chance is  $4.03E-5$  (0.004%). An example of the calculations is given in online supplementary figure 2.

The following criteria were used to classify variations/mutations. We use a list of mutation specific features based on in silico analysis using the mutation interpretation software AlaMut (V1.5). A score is given depending on the outcome of a prediction test for each feature (ie, Grantham distance). Then, depending on the total score and the availability of the variant in at least 300 ethnically matched control alleles (data obtained from the literature and/or the internet, eg, 1000 Genomes: <http://browser.1000genomes.org/index.html>; Exome Variant Server: <http://evs.gs.washington.edu/EVS>, or from own control alleles), we classified them as: not pathogenic; as a variant of unknown clinical significance; VUS1, unlikely to be pathogenic; VUS2, uncertain; or VUS3, likely to be pathogenic. Family information (co-segregation), phenotypic features and/or functional analysis are needed to classify a variant as (putatively) pathogenic. The protocol used for the classification of variants has recently been published by van Spaendonck-Zwarts *et al*<sup>9</sup> All variants of interest identified in our HCM and DCM group were confirmed by Sanger sequencing.

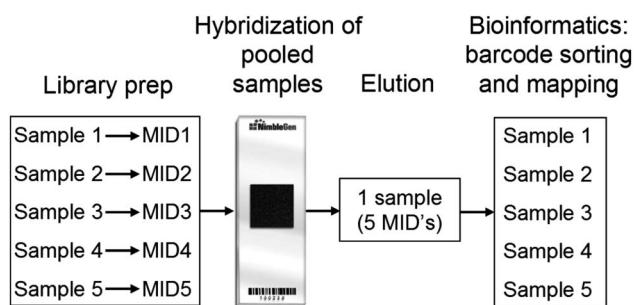
## RESULTS

### Target enrichment and on-target percentage

We evaluated and optimised the ability of Roche NimbleGen 385K Custom Sequence Capture Arrays to enrich patient samples for every exon in a panel of 23 genes. The degree of enrichment for the captured samples was estimated by qPCR analysis of four control loci. The average relative enrichment for the four control loci was  $>100$ -fold for all samples sequenced. The on-target percentages, calculated as number of bases present on the targeted region divided by the total number of mapped bases, varied between 12.8% and 33.1%, using the standard sequence capture protocol. When calculating the ratio of mean base coverage of the 92.5 kb targeted region divided by the mean base coverage outside the targeted region ( $\sim 3$  Gb), the enrichment factor is still 4720-fold and 16 046-fold respectively. A mean on-target percentage using the standard protocol was on average  $\sim 20\%$ . Although this is the equivalent of an enrichment factor of  $\sim 8000$ -fold, it is still not high enough to sequence an acceptable number of patients at an acceptable base coverage on the GS-FLX Titanium. Therefore we introduced a second round of hybridisation of the same sample on the previously used capture array. This procedure increased the on-target percentage to 75–92% (an 80% on-target represents an enrichment factor of 129 729-fold). Base coverage was reproducible both within and between the experiments with 99.8% of the targeted sequence covered.

### On-array multiplexing

The use of capture arrays gives a limitation in the number of samples that can be processed in a given time. Moreover, the costs of one capture array per sample are a significant part of the total costs of the entire procedure. Therefore, we used unique MIDs (bar coding) which are introduced during library preparation. After amplification and quantification of the individual libraries, equimolar mixtures of 5–10 patients were hybridised to one capture array according to the protocol. From that moment on the 5–10 patients were treated as one sample. After sequencing, the patients were separated based on their barcode and the variants were determined using the gsMAPPER (Roche) (figure 1). In general, 100K reads per patient within each multiplex experiment were sequenced. The difference in the number of reads per patient within a multiplexed sample usually was



**Figure 1** A schematic representation of an on-array multiplexed sequence capture experiment. After individual introduction of a unique bar code (multiplex identifiers) to each sample during the library preparation, all samples are mixed equimolarly and hybridised to one capture array. From that time on, the five multiplexed samples are treated as one sample during hybridisation, washes and elution as well as all post-capture steps including linker-mediated PCR, emulsion PCR and sequencing.

below twofold (75K to 150K reads per patient). Comparable number of reads per patient indicates that on-array multiplexing is feasible. Only occasionally a sample has less than half of the reads that was aimed for.

### Proof of principle experiment

In a first proof of principle experiment we evaluated a custom design capture array (V1) to enrich patient samples for every exon of 18 genes with known involvement in cardiomyopathy (table 1). In this experiment we included five HCM patients with different known pathogenic mutations (two substitutions, a 2 bp deletion, and a 1 bp and 26 bp insertion,) in either the *MYBPC3* or *MYH7* gene. All five pathogenic variants (*MYH7* c.1436A>C, *MYH7* c.1207C>T, *MYBPC3* c.2373\_2374dup, *MYBPC3* c.676\_701dup26, and *MYBPC3* c.2864\_2865del) were detected with sequencing on the GS-FLX Titanium. The whole coding region including splice junctions were Sanger sequenced for at least the gene where a pathogenic mutation was identified in the above mentioned HCM patients. Therefore, direct comparison of all the variants (pathogenic and non-pathogenic) found with Sanger sequencing and GS-FLX Titanium sequencing was possible. This showed that 57 out of 57 (100%) variants were detected with GS-FLX Titanium sequencing. Fifty-three out of 57 variants were present in the HCDiff list whereas four out of 57 were present in the AllDiff list. Evaluation of the variants present in the AllDiff list showed that three out of four variants were variants at the same position (*MYBPC3* c.1093-24C>T) and were not in the HCDiff list due to low coverage at the flank of the targeted region. The fourth variant (*MYBPC3* c.2373dup) was present in a sufficient number of reads that were of good quality, so there was no clear reason why it was not present in the HCDiff list. Therefore, in all the following analysis also the variants present in the AllDiffs with a variant percentage >20% were extracted from the list.

### Optimisation of exon coverage and addition of more genes

Exon coverage was determined for all 18 genes and all exons except exon 5 of *PRKAG2* were enriched by the capture array. Coverage per exon varied significantly between genes but also within genes (figure 2A). The mean base coverage for the targeted bases was  $38.3 \pm 14.8$  (mean  $\pm$  SD) after a single hybridisation. Exons with coverage significantly below the mean coverage were rebalanced on the next design (V2) by adding more probes for

poorly covered regions (both more identical and probes from the same region but at a different position were added). The number of probes added per target region was dependent on the coverage. After enrichment with this rebalanced capture array and sequencing the mean coverage per exon was calculated again. The mean base coverage for the targeted bases after a single hybridisation was  $41.6 \pm 12.9$  (mean  $\pm$  SD). Lower SD at a higher mean coverage implies a more even coverage when using the V2 array. In fact, exons that had a low coverage in the V1 array were more evenly distributed after enrichment with the V2 array (figure 2B). The missing exon from the V1 array (exon 5 of *PRKAG2*) was targeted in the V2 with seven times the standard amount of probes, but still this particular exon was not present in the reads. Evaluation of the characteristics of this region showed a high GC content of 76%. It is likely that during pre-capture and post-capture amplification there is poor amplification of this region and therefore it is not represented in the post-capture library. Since the on-target percentages were low we switched to a protocol that used a double hybridisation. After the second hybridisation the mean base coverage for the targeted bases was  $62.8 \pm 42.3$  (mean  $\pm$  SD). Introduction of a second hybridisation increased the on-target percentage drastically, however, the SD of the mean coverage also increased. In the next version of the capture array (V3), we performed an additional balancing step and added the coding exons of five more genes, resulting in targeting a total of ~300 exons of 23 genes. The mean base coverage for the targeted bases was  $97.6 \pm 44.7$  (mean  $\pm$  SD). The percentage of bases covered at least once was 99.8% (only missed 90 bases of exon 5 *PRKAG2*: chr7:151329155-151329224). In the final version of the capture array (V4) we have rebalanced the final five genes. The mean base coverage for the targeted bases was  $103.7 \pm 35.9$  (mean  $\pm$  SD). With the last design there was an average of 1.64 fragment per patient that had one or more bases at  $<16\times$  coverage and therefore needed to be analysed with a Sanger reaction. This represents an average of 0.36% of the analysed bases.

When comparing double hybridisation on the V2 array with the V4 array (mean  $\pm$  SD  $62.8 \pm 42.3$  vs  $103.7 \pm 35.9$ ) it clearly shows improved performance due to design optimisation.

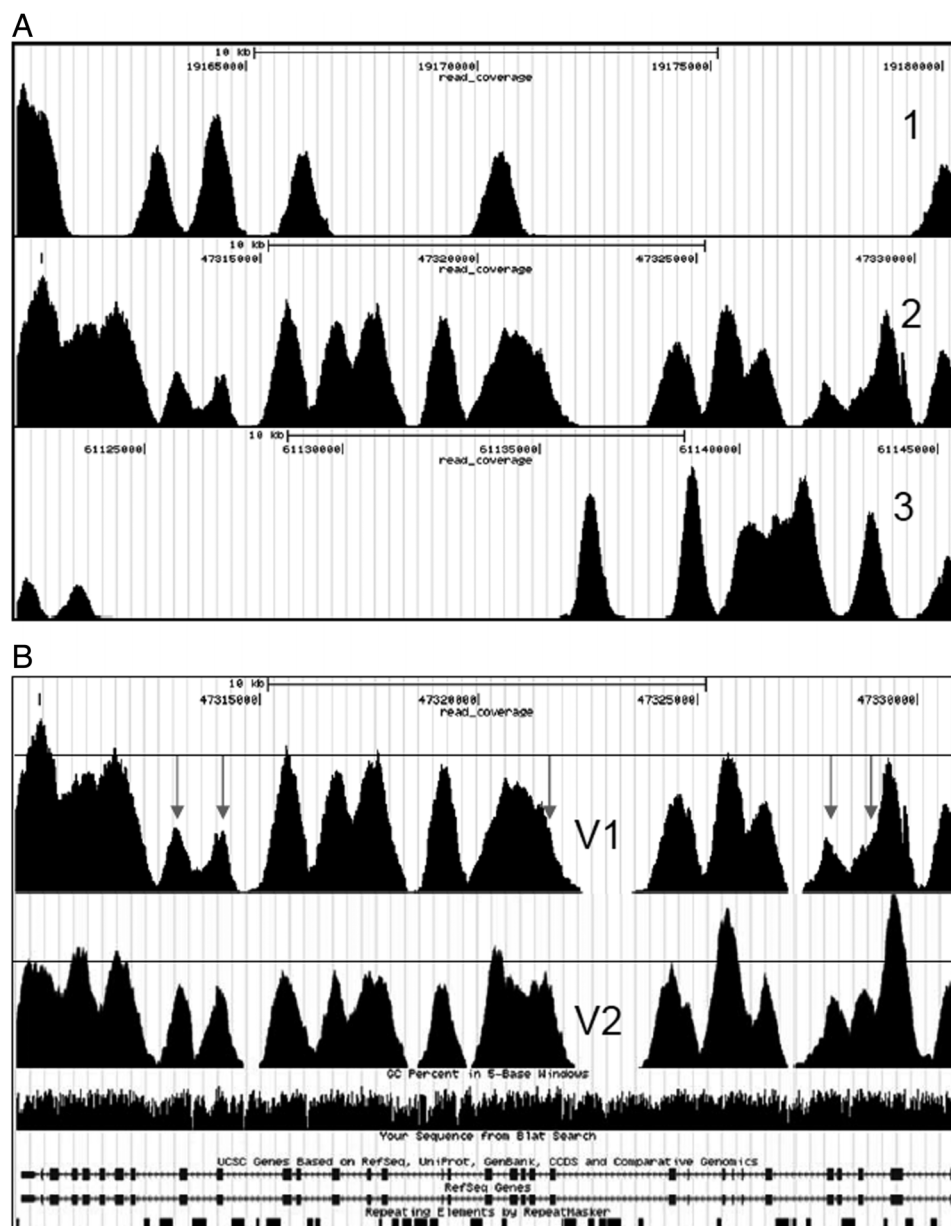
### Screening of DCM patients already seen in diagnostics but no mutation found

Seventeen patients diagnosed with DCM and two patients with a DCM family history were selected for screening of the genes present on the V2 and V3 arrays. In 10 out of 19 patients we found variants with different probability of pathogenicity (for more details see table 2). Two variants, the *PLN* c.40\_42del and the *MYH7* c.415G>T, were both observed twice in unrelated patients, with one patient having two variants (*MYH7* c.415G>T + *LAMP2* c.661G>A). All the other variants were present solitary. All the variants present in *TNNI3*, *EMD*, *SCN5A*, *MYH7*, and *LAMP2* were not observed in all the previously tested probands for these genes. The variant present in *PLN* has been reported as pathogenic<sup>10 11</sup> and is a well-known founder mutation in the Netherlands. One of the two variants in *SCN5A* (NM\_198056.2 : c.4859C>T) has been reported previously.<sup>12-16</sup> The variant in *LDB3* has not been found in 200 HCM/DCM/non-compaction cardiomyopathy patients screened in Erasmus Medical Center, Rotterdam (personal communication).

### Comparison of variants found with Sanger sequencing and GS-FLX Titanium sequencing in HCM patients

In order to make a side by side comparison between variants that were found with Sanger sequencing and variants that were found by GS-FLX Titanium sequencing, we have selected nine

**Figure 2** Representative example of a gene with even exon coverage (1), intermediate exon coverage (2), and uneven exon coverage (3) before rebalancing of the capture array (A) and exon coverage distribution before (V1) and after (V2) rebalancing of the capture array (B). Note that poorly covered exons or exon flanks (arrows) are far below the average coverage (horizontal line) before rebalancing. After rebalancing exon coverage is far more uniform.



patients diagnosed with HCM but with no pathogenic mutation identified. In these patients, all the coding exons of the genes offered for DNA diagnostics (*MYBPC3*, *MYH7*, *MYL2*, *MYL3*, *TNNI3*, *TNNT2*, *TPM1*, and *GLA* genes) were already analysed with dHPLC and/or Sanger sequencing. Within the targeted region of the V3 capture array, 258 variants were detected with dHPLC and/or Sanger sequencing. Of these 258 variants, three were variants of unknown clinical significance (*TNNT2*, c.877C>T,<sup>20 21</sup> *TPM1*, c.755A>G and *MYBPC3*, c.3392T>C<sup>22</sup>), while the others were non-pathogenic coding and non-coding variants. These variants served as positive control in the direct comparison of Sanger sequencing versus GS-FLX Titanium Roche sequencing. Evaluation of the variants present in the HCDiff and AllDiff lists showed a total of 262 variants. From the 258 variants that were potentially detectable, one variant (*MYL2* c.353+46dup) was missed in two patients by the GS-FLX Titanium. This means six additional variants (one heterozygote and five homozygotes) were found with GS-FLX Titanium sequencing, but not with dHPLC. All additional variants were confirmed with Sanger sequencing. This

results in a detection rate of 101.6%, compared to the dHPLC and/or Sanger sequencing. If one excludes the missed variants by the dHPLC, 258 variants were potentially detectable. From these, 256 were found in either HCDiff or AllDiff, while one variant (*MYL2*, c.353+46dup) was missing in two patients resulting in a corrected detection rate of 99.2%. It should be noted that all the other 48 known non-coding indels of different lengths were detected correctly.

On top of the known variants and putative mutations we also found two additional amino acid changes from which one also introduced a potential splice site loss in genes that were not analysed before. These variants are VCL: NM\_014000.2: c.1543G>A (amino acid change and loss of splice site), LDB3: NM\_007078.2: c.566C>T. We classified them as VUS2 (see table 3).

#### Side by side comparison of GS-FLX Titanium sequencing and Sanger sequencing in 30 index patients

Finally, the fully optimised capture array (V4) was used for a second direct comparison in 30 HCM index patients by

**Table 2** Clinical parameters and variants found in DCM patients

DNA nr.	Gender	Age at diagnosis (years)	EF (%)	Gene	g.Var	Reference sequence	c.Var	p.Var	PolyPhen2 HumVar	PolyPhen2 HumDiv	Grantham dist	SIFT	AGVGD	Freq. in EVS (variant alleles/wt alleles)	Classification*	Additional information
Patient 1	F	64	?	<i>TNNI3</i>	55665408 A>G	NM_000363.4	539A>G	p. Asp180Gly	Probably damaging	Possibly damaging	94	0.04	C0	Not found	VUS2	3× DCM index patients, <sup>17 18</sup> 1 × LVNC. <sup>19</sup> Mother of our index with clinical suspicion is not a carrier
Patient 2	M	51	17													
Patient 3	M	44	?	<i>EMD</i>	153609246 C>T	NM_000117.2	454C>T	p. Arg152Cys	Benign	Benign	180	0.00	C55	Not found	VUS3	No
Patient 4	F	50	13	–	–		–	–	–	–		–	–			
Patient 5	M	40	40	–	–		–	–	–	–		–	–			
Patient 6	F	29	37	<i>PLN</i>	118880124_118880126del	NM_002667.3	40_42del	p.Arg14del						Not looked at	Pathogenic	Founder in the Netherlands; mouse model+DCM family <sup>10</sup> ; large DCM family. <sup>11</sup>
Patient 7†	F	22	51	<i>SCN5A</i>	38595955 T>C	NM_198056.2	4628T>C	p. Val1543Ala	Benign	Benign	64	0.00	C65	Not found	VUS2	No
Patient 8	F	55	20	–	–		–	–	–	–		–	–			
Patient 9	F	28	35	–	–		–	–	–	–		–	–			
Patient 10	F	27	30	<i>MYH7</i>	23901935 G>T	NM_000257.2	415G>T	p.Val139Leu	Probably damaging	Probably damaging	32	0.00	C0	Not found	VUS3	Found in 3 other index patients‡
	F	–	–	<i>LAMP2</i>	119581776 G>A	NM_002294.2	661G>A	p.Gly221Arg	Probably damaging	Probably damaging	125	0.00	C0	21/10542	VUS2	No
Patient 11	M	56	?	–	–		–	–	–	–		–	–			
Patient 12	F	52	38	<i>MYH7</i>	23901935 G>T	NM_000257.2	415G>T	p.Val139Leu	Probably damaging	Probably damaging	32	0.00	C0	Not found	VUS3	Found in 3 other index patients‡
Patient 13†	M	40	47	–	–		–	–	–	–		–	–			
Patient 14	M	35	21	<i>SCN5A</i>	38593004 C>T	NM_198056.2	4859C>T	p. Thr1620Met	Probably damaging	Probably damaging	81	0.00	C65	Not found	VUS3	Large IVF family (monoallelic with rare polymorphism). <sup>13</sup> Functional test not conclusive. <sup>12–16</sup>
Patient 15	F	41	25	–	–		–	–	–	–		–	–			
Patient 16	F	73	?	<i>LDB3</i>	88451727 A>G	NM_007078.2	764A>G	p.Lys255Arg	Benign	Benign	26	0.00	C25	Not found	VUS1	Not
Patient 17	V	25	41	<i>PLN</i>	118880124_118880126del		40_42del	p.Arg14del						Not looked at	Pathogenic	Founder in the Netherlands; mouse model+DCM family <sup>10</sup> ; large DCM family. <sup>11</sup>
Patient 18	M	49	37	–	–		–	–	–	–		–	–			
Patient 19§	V	?	?	<i>MYH7</i>	23896054 T>C	NM_000257.2	1976T>C	p. Met659Thr	Probably damaging	Probably damaging	81	0.00	C65	Not found	Pathogenic	Large family (co-segregation in 10 affected members)‡

\*For further details see Methods section.

†Patient does not have observed dilatation yet but has a strong family history with DCM.

‡Identified in cardiomyopathy patients analysed during routine diagnostics in our laboratory.

§Patient also has hypertrophy so DCM could be a result of HCM.

DCM, dilated cardiomyopathy; EF, ejection fraction; HCM, hypertrophic cardiomyopathies; IVF, idiopathic ventricular fibrillation; LVNC, left ventricular non-compaction; n.a. not available; ?, decreased EF, no exact measurement performed.

comparing the seven genes (*MYBPC3*, *MYH7*, *TNNT2*, *TNNI3*, *TPM1*, *MYL2*, and *MYL3*) sequenced with both Sanger sequencing and GS-FLX Titanium sequencing. From the 319 variants present, a total of 317 variants were detected with Sanger sequencing and 318 with GS-FLX Titanium sequencing. One variant was not detected with GS-FLX Titanium sequencing (*MYL2*: NM\_000432.3: c.353+20delG) and two variants were not detected with Sanger sequencing (*MYH7*: NM\_000257.2 c.1231G>A and *MYH7*: NM\_000257.2: c.2967T>C). Re-examination of these two variants by Sanger sequencing confirmed both variants. Re-evaluation of the original traces also showed the two variants but they were not picked up due to high background. Also, here it must be noted that the other 51 known indels were detected with GS-FLX Titanium sequencing. This represents a sensitivity of 317/319 (99.4%) and 318/319 (99.7%) for Sanger sequencing and GS-FLX Titanium sequencing, respectively.

In the 30 patients we found  $26.7 \pm 5.8$  (mean  $\pm$  SD) variants per patient. From these, 17 variants ( $0.57 \pm 0.77$ , mean  $\pm$  SD) have varying probabilities of pathogenicity. These include variants in genes routinely screened with Sanger sequencing—that is, *MYH7*: NM\_000257.2: c.1178C>T,<sup>26–28</sup> c.1231G>A,<sup>29–32</sup> c.2002C>A, c.5135G>A, c.1987C>T, c.121G>A and *MYBPC3*: NM\_000256.3: c.2827C>T,<sup>33</sup> c.3004C>T (twice), c.2864\_2865del, *MYL2*: NM\_000432.3: c.37G>A<sup>34–37</sup> and *TPM1*: NM\_000366.5: c.618A>G, but also in genes that are part of the additional genes in the NG cardiomyopathy panel—that is, *SCN5A*: NM\_198056.2: c.3157G>A<sup>38, 39</sup>, *PRKAG2*: NM\_016203.3: c.253C>T, *TCAP*: NM\_003673.2: c.37\_39del,<sup>40, 41</sup> *CSRP3*: NM\_003476.3: c.10T>C<sup>42, 43</sup>, *LMNA*: NM\_170707.2: c.1804G>A,<sup>44, 45</sup> *DES*: NM\_001927.3: c.935A>C (table 4). Ten patients had a single variant, two patients had two variants each (*MYH7*: NM\_000257.2: c.1178C>T combined with *SCN5A*: NM\_198056.2: c.3157G>A, and *MYBPC3*: NM\_000256.3: c.3004C>T combined with *LMNA*: NM\_170707.2: c.1804G>A), and one patient had three variants (*MYH7*: NM\_000257.2: c.121G>A, *MYBPC3*: NM\_000256.3: c.3004C>T, and *DES*: NM\_001927.3: c.935A>C). All variants were confirmed with Sanger sequencing.

## DISCUSSION

In the present study we show the feasibility of array based target enrichment of 23 HCM and DCM genes combined with NGS for diagnostic mutation screening in HCM/DCM patients. The use of DNA bar codes enabled pre-capture multiplexing, a step essential for increasing throughput and reducing the costs of analysis.<sup>46</sup> This approach not only enables the simultaneous and comprehensive investigation of multiple target genes but also facilitates the analysis of multiple samples in parallel. We have developed an optimised capture array that is at least as sensitive as the current standard, Sanger sequencing. For HCM, the current capture array analyses three times more genes as compared with the current Sanger sequencing panel within the same turn-around time. Screening of more genes results in a higher diagnostic yield. For DCM patients, but not for HCM patients, we have developed a similar assay for *TTN*. This enables us to screen *TTN* on top of the 23 reported genes only for DCM patients.

We and others have shown that NGS technologies are on the verge of being broadly used in clinical laboratories. It is very likely that these new technologies will replace traditional (Sanger sequencing based and array based) sequencing tests for genetically heterogeneous disorders like HCM/DCM.

Targeted exome massively parallel sequencing has great potential for both research and clinical use. However, sequencing of the entire exome for diseases with 'limited' numbers of genes to be investigated is not practically feasible yet due to relatively high costs, variable depth of exon coverage, the extent of data analysis, and data storage. To overcome these problems, sequence capture based target enrichment for a limited number of genes followed by NGS can be an approach that is logistically and financially feasible.

Our first pilot study with five HCM patients with known pathogenic mutations showed that array based target enrichment and NGS could easily detect different types of known mutations (substitutions, insertions and deletions) and numerous non-pathogenic variants already detected with Sanger sequencing. All 57 variants detected with Sanger sequencing were also found with GS-FLX Titanium sequencing, including a coding 26 bp insertion in the *MYBPC3* gene (figure 3). Because we detected 100% of the variants present in five HCM patients we proceeded with testing of additional patients. Nevertheless, in this pilot study we have observed that exon coverage varies significantly within one sample (determined by array design) resulting in lower confidentiality of particular variants (eg, *MYBPC3*: NM\_000256.3: c.1093-24C>T). A balanced representation of all targeted exons would reduce the average coverage needed to detect variants with high confidentiality, consequently lowering the false negative rate. Therefore, we have designed arrays with a more balanced coverage as has previously been proposed by others.<sup>47, 48</sup> The rebalanced capture array has been used to analyse nine HCM patients and 19 DCM patients. The rebalanced design showed that 99.80% of the targeted coding bases were covered at least once and 99.64% at least 16 $\times$ . We and others<sup>8</sup> have calculated that at 15–16 $\times$  coverage a 99% sensitivity is obtained. This means that for an experiment with a mean coverage of  $100 \pm 35 \times$  the statistical chance that a variant is missed in a patient is 0.004% (for calculation see supplementary figure S2).

The bases with low coverage were exon 5 from *PRKAG2* (NM\_016203.3) in all patients and exon 2 from *LAMP2* (NM\_002294.2) in about half of the patients. This is likely due to a high GC content, a phenomenon observed before.<sup>49</sup> In the current test these exons are analysed by standard Sanger sequencing.

From the nine HCM patients the HCM gene panel had been analysed with Sanger sequencing. In these patients 99.2% of the present variants were detected with GS-FLX Titanium sequencing. The two undetected non-coding variants (identical in two individuals) were both times a single nucleotide insertion present in a region with multiple homopolymer stretches of four to six nucleotides. Variant detection and in particular indels in homopolymer stretches is a known problem with pyrosequencing due to incorrect base calling in these regions.<sup>50</sup> In our datasets it has become clear that homopolymer stretches up to 5-mers are generally called properly, while 6-mers and more in general result in improper base calling. Screening all coding regions for  $\geq 6$ -mers showed 17 exons with  $\geq 6$ -mers. For diagnostic purposes, these exons are analysed by Sanger sequencing until better base calling algorithms are available.

Apart from the variants present in the already used gene panel, numerous additional variants were found in the genes that were not Sanger sequenced previously. Two variants (indicated in bold), which both result in amino acid changes (one amino acid change also predicts a splice site loss) (table 3), are of possible clinical significance. The variant in *LDB3* has been reported before<sup>23–25</sup> while the *VCL* variant is novel.

**Table 3** Clinical parameters and variants found in HCM patients

	Gender	Age at diagnosis (years)	MLVWT (mm)	SCD in family	Gene	Var g.	Reference sequence	Var c.	Var p.	PolyPhen2 HumVar	PolyPhen2 HumDiv	Grantham dist	SIFT	AGVGD	Freq in EVS (variant alleles/wt alleles)	Classification*	Additional information	
Patient 20	M	?	?	?	<b>TPM1</b>	63356265 A>G	NM_001018020.1	775A>G	p. Lys259Glu	Not reliable	Not reliable	56	0.00	C55	4/10 754	VUS2	No (not present in transcript NM_000366.5)	
Patient 21	M	65	17	No	–	–	–	–	–	–	–	–	–	–	–	–	–	
Patient 22	F	58	16	Yes	<b>VCL</b>	75854219 G>A	NM_014000.2	1543G>A	p. Gly515Ser	Probably damaging	Probably damaging	56	0.00	C55	Not found	VUS2	No (splicing could be affected)	
Patient 23	M	62	15	No	–	–	–	–	–	–	–	–	–	–	–	–	–	
Patient 24	M	79	17	No	–	–	–	–	–	–	–	–	–	–	–	–	–	
Patient 25	M	44	20	?	<b>LDB3</b>	88441437 C>T	NM_007078.2	566C>T	p. Ser189Leu	Benign	Benign	145	0.1	C0	2/10 806	VUS3	DCM/LVNC family (called S196L), co-segregation in 4 affected members. <sup>23</sup> Functionally tested. <sup>24 25</sup>	
					<b>MYBPC3</b>	47354463 T>C	NM_000256.3	3392T>C	p. Ile1131Thr	Benign	Possibly damaging	89	0.03	C55	7/12 331	VUS1	3 other DCM/NCCM index patients† (one of them together with the pathogenic Dutch founder mutation c.2373dup in MYBPC3 <sup>22</sup> ; in another family no co-segregation in an affected uncle)	
Patient 26	M	56	23	No	–	–	–	–	–	–	–	–	–	–	–	–	–	
Patient 27	M	50	19	No	–	–	–	–	–	–	–	–	–	–	–	–	–	
Patient 28	M	67	28	?	<b>TNNT2</b>	201328349 C>T	NM_000364.2	877C>T	p. Arg293Cys	Probably damaging	Probably damaging	180	0.00	C0	Not found	VUS3	3 HCM index patients (called R286C). <sup>20 21</sup> 3 HCM index patients†	

Variants in the genes highlighted in bold are variants that were found on top of the variants that were already detected by Sanger sequencing. The variant present in VCL (patient 22) introduced a loss of a splice site on top of the amino acid change.

\*For further details see Methods section.

†Identified in cardiomyopathy patients analysed during routine diagnostics in our laboratory.

DCM, dilated cardiomyopathy; HCM, hypertrophic cardiomyopathy; LVNC, left ventricular non-compaction; MLVWT, maximal left ventricular wall thickness; n.a., not available; NCCM, non-compaction cardiomyopathy; SCD, sudden cardiac death; ?, unknown or no exact measurement performed.



**Table 4** Clinical parameters and variants found in index patients

	Gender	Age at diagnosis (years)	MLVWT (mm)	SCD in family	Gene	Var g.	Reference sequence	Var c.	Var p.	PolyPhen2 HumVar	PolyPhen2 HumDiv	Grantham dist	SIFT	AGVGD	Freq in EVS (variant alleles/wt alleles)	Classification*	Additional information
Patient 29	M	24	17	?	<i>SCN5A</i>	38622493 G>A	NM_198056.2	3157G>A	Glu1053Lys	Probably damaging	Possibly damaging	56	0.00	C55	Not found	Pathogenic	3 BrS index patients (one had a second <i>SCN5A</i> mutation: R1583C). <sup>38</sup> Functional test: no binding to Ankrin-G <sup>39</sup>
					<i>MYH7</i>	23898517 C>T	NM_000257.2	1178C>T	Ala393Val	Possibly damaging	Possibly damaging	64	0.00	C0	Not found	VUS2	No
Patient 30	F	69	19	No	–	–	–	–	–	–	–	–	–	–	–	–	–
Patient 31†	M	65	11	No	<i>MYL2</i>	111356964G>A	NM_000432.3	37G>A	Ala13Thr	Possibly damaging	Benign	58	0.08	C0	6/13 000	VUS1	3 families and an index: sometimes together with N1327K in <i>MYH7</i> , not always good co-segregation. <sup>34–36</sup> Functional test: not conclusive <sup>37</sup>
Patient 32	M	?	?	?	<i>PRKAG2</i>	151478451C>T	NM_016203.3	253C>T	Pro85Ser	Possibly damaging	Benign	74	0.12	C0	Not found	VUS2	No
Patient 33	M	68	20	No	–	–	–	–	–	–	–	–	–	–	–	–	–
Patient 34	F	55	16	No	–	–	–	–	–	–	–	–	–	–	–	–	–
Patient 35	M	16	18	Yes	<i>MYBPC3</i>	47356671 C>T	NM_000256.3	2827C>T	Arg943X	–	–	–	–	–	Not found	Pathogenic	Dutch founder mutation. <sup>33</sup>
Patient 36‡	F	68	14	Yes	–	–	–	–	–	–	–	–	–	–	–	–	–
Patient 37	M	33	19	No	–	–	–	–	–	–	–	–	–	–	–	–	–
Patient 38	F	58	15	No	<i>MYH7</i>	23898464 G>A	NM_000257.2	1231G>A	Val411Ile	Possibly damaging	Probably damaging	29	0.00	C0	Not found	VUS3	4 HCM index patients. <sup>29–32</sup>
Patient 39	M	31	?	Yes	<i>MYH7</i>	23896028 C>A	NM_000257.2	2002C>A	His668Asn	Probably damaging	Probably damaging	68	0.00	C65	Not found	VUS3	Mutation was not found back in both parents (father is probably affected)
Patient 40	M	46	15	No	–	–	–	–	–	–	–	–	–	–	–	–	–
Patient 41§	M	62	?	Yes	–	–	–	–	–	–	–	–	–	–	–	–	–
Patient 42	M	69	31	No	<i>TCAP</i>	37821649_37821651 del	NM_003673.3	37_39del	Glu13del	–	–	–	–	–	Not looked at	VUS1	8 index patients (LOVD); 2 HCM index patients. <sup>41</sup> In 0.4–0.5% of the control alleles <sup>40</sup>
Patient 43	M	42	22	No	–	–	–	–	–	–	–	–	–	–	–	–	–
Patient 44	F	74	21	No	<i>MYH7</i>	23884860 G>A	NM_000257.2	5135G>A	Arg1712Gln	Possibly damaging	Probably damaging	43	0.00	C0	Not found	VUS1	9 Dutch HCM index cases (one family no co-segregation in affected daughter)

Continued

Table 4 Continued

	Gender	Age at diagnosis (years)	MLVWT (mm)	SCD in family	Gene	Var g.	Reference sequence	Var c.	Var p.	PolyPhen2 HumVar	PolyPhen2 HumDiv	Grantham dist	SIFT	AGVGD	Freq in EVS (variant alleles/wt alleles)	Classification*	Additional information
Patient 45	M	39	15	No	–	–	–	–	–	–	–	–	–	–	–	–	–
Patient 46	F	19	25	No	–	–	–	–	–	–	–	–	–	–	–	–	–
Patient 47	F	71	?	Yes	<i>CSRP3</i>	19213986 T>C	NM_003476.3	10T>C	Trp4Arg	Possibly damaging	Possibly damaging	101	0.31	C0	42/10 716	VUS2	1 HCM index patient¶; 1 DCM index patient¶; 0.5% DCM index patients, 0.6% HCM index patients and 0.5% of control patients. <sup>43</sup> Functional test: Knock in mice with phenotype <sup>42</sup>
Patient 48	F	56	15	Yes	<i>LMNA</i>	156108384G>A	NM_170707.2	1804G>A	Gly602Ser	Benign	Benign	56	1.00	C0	12/12 980	VUS2	1 HCM index¶; 3× diabetes (+ ovary syndrome). <sup>45</sup> 1× EDMD2. <sup>44</sup> Functional test not conclusive <sup>45</sup>
					<i>MYBPC3</i>	47355294 C>T	NM_000256.3	3004C>T	Arg1002Trp	Probably damaging	Probably damaging	101	0.00	C65	20/12 316	VUS2	2 HCM index patients¶
Patient 49	M	43	28	No	<i>MYBPC3</i>	47356633_47356634del	NM_000256.3	2864_2865del	p. Pro955ArgfsX95	–	–	–	–	–	n.a	Pathogenic	Dutch founder mutation. <sup>33</sup>
Patient 50	M	50	18	No	–	–	–	–	–	–	–	–	–	–	–	–	–
Patient 51	M	60	20	Yes	–	–	–	–	–	–	–	–	–	–	–	–	–
Patient 52	F	38	26	No	<i>MYH7</i>	23896043C>T	NM_000257.2	1987C>T	Arg663Cys	Probably damaging	Possibly damaging	180	0.00	C55	Not found	Pathogenic	5 HCM index patients; 3 HCM index patients <sup>26–28</sup>
Patient 53	F	32	20	Yes	<i>DES</i>	220285587A>C	NM_001927.3	935A>C	Asp312Ala	Probably damaging	Probably damaging	126	0.00	C0	14/12 993	VUS2	No
					<i>MYBPC3</i>	47355294C>T	NM_000256.3	3004C>T	Arg1002Trp	Probably damaging	Probably damaging	101	0.00	C65	20/12 316	VUS2	2 HCM index patients¶
					<i>MYH7</i>	23902821G>A	NM_000257.2	121G>A	Asp41Asn	Benign	Possibly damaging	23	0.13	C0	Not found	VUS2	No
Patient 54**	M	52	12	Yes	–	–	–	–	–	–	–	–	–	–	–	–	–
Patient 55	M	45	17	Yes	–	–	–	–	–	–	–	–	–	–	–	–	–
Patient 56††	F	49	12	Yes	–	–	–	–	–	–	–	–	–	–	–	–	–
Patient 57	M	55	17	No	–	–	–	–	–	–	–	–	–	–	–	–	–
Patient 58‡‡	M	32	13	Yes	–	–	–	–	–	–	–	–	–	–	–	–	–

\*For further details see Methods section.

†Also fibrosis consistent with a mild form of HCM.

‡Son also has MLVWT 14 mm.

§Non-compaction cardiomyopathy.

¶Identified in cardiomyopathy patients analysed during routine diagnostics in our laboratory.

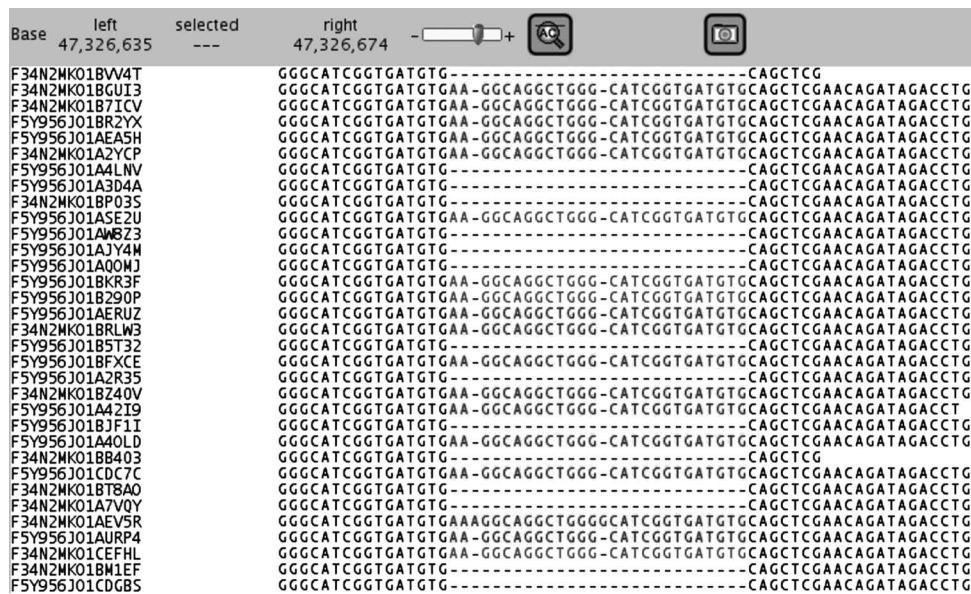
\*\*Myocardial crypts on MRI consistent with an early form of HCM.

††HCM in sister in whom DNA diagnostics is not possible.

‡‡Also SAM and mild outflow tract obstruction.

BrS, Brugada syndrome; DCM, dilated cardiomyopathy; EDMD, Emery-Dreifuss muscular dystrophy;

HCM, hypertrophic cardiomyopathy; LOVD, Leiden Open Variation Database; MLVWT, maximal left ventricular wall thickness; SAM, systolic anterior motion of the mitral valve; SCD, sudden cardiac death; ?, unknown or no exact measurement performed.



**Figure 3** Mapping of a coding 26 bp duplication in a hypertrophic cardiomyopathy patient. In contrast with short read platforms, a heterozygous insertion of 26 bp is easily mapped and reported.

In the DCM panel we detected 11 potentially pathogenic mutations in 10 individuals. The genes involved are *MYH7* (three mutations), *TNNI3* (one mutation), *PLN* (two identical mutations), *LAMP2* (one mutation), *LDB3* (one mutation), *EMD* (one mutation), and *SCN5A* (two mutations) (for details see table 2). The variant in *TNNI3*, *PLN*, and *SCN5A* have already been reported in the literature,<sup>10 11 17–19</sup> and all other variants are novel. Identification of additional variants in both HCM and DCM patients shows an increase in diagnostic yield upon extension of the gene panel sequenced.

Finally, we evaluated 30 index patients in parallel with Sanger sequencing. In these experiments we especially focused on the diagnostic value of this approach. In 10 patients a single potentially pathogenic variant was found, in two patients two potentially pathogenic variants were found, and in one patient three potentially pathogenic variants were identified. From the 17 variants 11 were in the genes regularly screened with Sanger sequencing, and six were found in genes that were additionally screened due to GS-FLX Titanium sequencing (table 4). This shows that the clinical sensitivity can be increased by the addition of extra genes. In this experiment we were able to compare directly the sensitivity of Sanger sequencing and GS-FLX Titanium sequencing. Sensitivity of GS-FLX Titanium sequencing was at least as good as that of Sanger sequencing (99.7% vs 99.4%, respectively), showing that GS-FLX Titanium sequencing can replace Sanger sequencing in a diagnostic setting.

Small insertions and deletions (1–50 bp) represent the second most frequent class of variation in the human genome after SNPs.<sup>51</sup> Throughout all the experiments described here we have particularly focused on the detection of indels. Two non-coding variants (*MYL2* c.353+20 and c.353+46) were not reported in the GS-FLX Titanium variant list. Both variants are present in a region with multiple homopolymer stretches of four to six nucleotides, likely the underlying reason for improper variant calling. Nevertheless, the other 106 indels present in this study were called properly. In fact, pathogenic mutations caused by an insertion of a G nucleotide, deletion of CT nucleotides, deletion of AGA nucleotides, as well as an insertion of 26 bases were called correctly. Furthermore, multiple known non-coding indels like insertion of a C nucleotide, deletion of a

C nucleotide, insertion of AC nucleotides, deletion of CTTCT nucleotides, insertion of ATTTT nucleotides, insertion of ATTTTGTTTT nucleotides, and insertion of ACAG nucleotides were all detected in these patients.

In conclusion, we have shown that on-array multiplexed sequence capture in combination with GS-FLX Titanium sequencing is suitable for reliable variant detection (sensitivity of 99.7%) and will increase clinical sensitivity in cardiomyopathy patients. To date, NGS is used as a research tool with high confidence and ease. In the present paper we demonstrate that due to continuing improvements in throughput, accuracy, cost and ease of data analysis, it has become feasible to apply NGS in a diagnostic setting.

**Contributors** All authors included on the paper fulfil the criteria of authorship, as they contributed to the study conception and design, acquisition of data, or analysis and interpretation of data, drafting the article or revising it for intellectual content, and approving the final version submitted. In addition, ORFM designed the study and experimental set-up, performed the experiments, analysed data, and wrote the paper. RHLD and MMAMM designed the study, analysed data, and wrote the paper. FB designed the experimental set-up. J-FS analysed data and contributed to drafting the manuscript. JVM contributed to the study design and drafting of the manuscript. NH and IC contributed to identification and selection of patients. MAH performed the experiments and developed the bioinformatical analysis. MEJ performed the experiments.

**Competing interests** None.

**Patient consent** Obtained.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Open Access** This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 3.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/3.0/>

**REFERENCES**

- 1 Kimura A. Molecular basis of hereditary cardiomyopathy: abnormalities in calcium sensitivity, stretch response, stress response and beyond. *J Hum Genet* 2010;55:81–90.
- 2 Luk A, Ahn E, Soor GS, Butany J. Dilated cardiomyopathy: a review. *J Clin Pathol* 2009;62:219–25.

- 3 Soor GS, Luk A, Ahn E, Abraham JR, Woo A, Ralph-Edwards A, Butany J. Hypertrophic cardiomyopathy: current understanding and treatment objectives. *J Clin Pathol* 2009;62:226–35.
- 4 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.
- 5 Gowrisankar S, Lerner-Ellis JP, Cox S, White ET, Manion M, LeVan K, Liu J, Farwell LM, Iartchouk O, Rehm HL, Funke BH. Evaluation of second-generation sequencing of 19 dilated cardiomyopathy genes for clinical applications. *J Mol Diagn* 2010;12:818–27.
- 6 Meder B, Haas J, Keller A, Heid C, Just S, Borries A, Boisguerin V, Scharfenberger-Schmeer M, Stahler P, Beier M, Weichenhan D, Strom TM, Pfeufer A, Korn B, Katus HA, Rottbauer W. Targeted next-generation sequencing for the molecular genetic diagnostics of cardiomyopathies. *Circ Cardiovasc Genet* 2011;4:110–122.
- 7 Mullaney JM, Mills RE, Pittard WS, Devine SE. Small insertions and deletions (INDELs) in human genomes. *Hum Mol Genet* 2010;19:R131–6.
- 8 Hoischen A, Gillissen C, Arts P, Wieskamp N, van d V, Vermeer S, Steehouwer M, de VP, Meijer R, Seiquerios J, Knoers NV, Buckley MF, Scheffer H, Veltman JA. Massively parallel sequencing of ataxia genes after array-based enrichment. *Hum Mutat* 2010;31:494–9.
- 9 van Spaendonck-Zwarts KY, van Rijsingen IA, van den Berg MP, Lekanne Deprez RH, Post JG, van Mil AM, Asselbergs FW, Christiaans I, van Langen IM, Wilde AA, de Boer RA, Jongbloed JD, Pinto YM, van Tintelen JP. Genetic analysis in 418 index patients with idiopathic dilated cardiomyopathy: overview of 10 years' experience 1. *Eur J Heart Fail* 2013;15(6):628–36.
- 10 Dewitt MM, MacLeod HM, Soliven B, McNally EM. Phospholamban R14 deletion results in late-onset, mild, hereditary dilated cardiomyopathy. *J Am Coll Cardiol* 2006;48:1396–8.
- 11 Haghighi K, Kolokathis F, Gramolini AO, Waggoner JR, Pater L, Lynch RA, Fan GC, Tsiapras D, Parekh RR, Dorn GW, MacLennan DH, Kremastinos DT, Kranias EG. A mutation in the human phospholamban gene, deleting arginine 14, results in lethal, hereditary cardiomyopathy. *Proc Natl Acad Sci USA* 2006;103:1388–93.
- 12 Baroudi G, Carbonneau E, Pouliot V, Chahine M. SCN5A mutation (T1620M) causing Brugada syndrome exhibits different phenotypes when expressed in *Xenopus* oocytes and mammalian cells. *FEBS Lett* 2000;467:12–16.
- 13 Chen Q, Kirsch GE, Zhang D, Brugada R, Brugada J, Brugada P, Potenza D, Moya A, Borggrefe M, Breithardt G, Ortiz-Lopez R, Wang Z, Antzelevitch C, O'Brien RE, Schulze-Bahr E, Keating MT, Towbin JA, Wang Q. Genetic basis and molecular mechanism for idiopathic ventricular fibrillation. *Nature* 1998;392:293–6.
- 14 Dumaine R, Towbin JA, Brugada P, Vatta M, Nesterenko DV, Nesterenko VV, Brugada J, Brugada R, Antzelevitch C. Ionic mechanisms responsible for the electrocardiographic phenotype of the Brugada syndrome are temperature dependent. *Circ Res* 1999;85:803–9.
- 15 Makita N, Shirai N, Wang DW, Sasaki K, George AL Jr, Kanno M, Kitabatake A. Cardiac Na(+) channel dysfunction in Brugada syndrome is aggravated by beta (1)-subunit. *Circulation* 2000;101:54–60.
- 16 Makita N, Mochizuki N, Tsutsui H. Absence of a trafficking defect in R1232W/T1620M, a double SCN5A mutant responsible for Brugada syndrome. *Circ J* 2008;72:1018–19.
- 17 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;3:155–61.
- 18 Rampersaud E, Siegfried JD, Norton N, Li D, Martin E, Hershberger RE. Rare variant mutations identified in pediatric patients with dilated cardiomyopathy. *Prog Pediatr Cardiol* 2011;31:39–47.
- 19 Hoedemaekers YM, Caliskan K, Michels M, Frohn-Mulder I, van der Smagt JJ, Pfeifferkorn JE, Wessels MW, ten Cate FJ, Sijbrands EJ, Dooijes D, Majoor-Krakauer DF. The importance of genetic counseling, DNA diagnostics, and cardiologic family screening in left ventricular noncompaction cardiomyopathy. *Circ Cardiovasc Genet* 2010;3:232–9.
- 20 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.
- 21 Milliou A, Anastasakis A, D'Cruz LG, Theopistou A, Rigopoulos A, Rizos I, Stamatiopoulos S, Toutouzas P, Stefanadis C. Low prevalence of cardiac troponin T mutations in a Greek hypertrophic cardiomyopathy cohort. *Heart* 2005;91:966–7.
- 22 Alders M, Jongbloed R, Deelen W, van den Wijngaard A, Doevendans P, Ten CF, Regitz-Zagrosek V, Vosberg HP, van Li, Wilde A, Dooijes D, Mannens M. The 2373insG mutation in the MYBPC3 gene is a founder mutation, which accounts for nearly one-fourth of the HCM cases in the Netherlands. *Eur Heart J* 2003;24:1848–53.
- 23 Vatta M, Mohapatra B, Jimenez S, Sanchez X, Faulkner G, Perles Z, Sinagra G, Lin JH, Vu TM, Zhou Q, Bowles KR, Di LA, Schimmenti L, Fox M, Chrisco MA, Murphy RT, McKenna W, Elliott P, Bowles NE, Chen J, Valle G, Towbin JA. Mutations in Cypher/ZASP in patients with dilated cardiomyopathy and left ventricular non-compaction. *J Am Coll Cardiol* 2003;42:2014–27.
- 24 Arimura T, Inagaki N, Hayashi T, Shichi D, Sato A, Hinohara K, Vatta M, Towbin JA, Chikamori T, Yamashina A, Kimura A. Impaired binding of ZASP/Cypher with phosphoglucosylase 1 is associated with dilated cardiomyopathy. *Cardiovasc Res* 2009;83:80–8.
- 25 Li Z, Ai T, Samani K, Xi Y, Tzeng HP, Xie M, Wu S, Ge S, Taylor MD, Dong JW, Cheng J, Ackerman MJ, Kimura A, Sinagra G, Brunelli L, Faulkner G, Vatta M. A ZASP missense mutation, S196L, leads to cytoskeletal and electrical abnormalities in a mouse model of cardiomyopathy. *Circ Arrhythm Electrophysiol* 2010;3:646–56.
- 26 Olivetto 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. Myofibrillar protein gene mutation screening and outcome of patients with hypertrophic cardiomyopathy. *Mayo Clin Proc* 2008;83:630–8.
- 27 Song L, Zou Y, Wang J, Wang Z, Zhen Y, Lou K, Zhang Q, Wang X, Wang H, Li J, Hui R. Mutations profile in Chinese patients with hypertrophic cardiomyopathy. *Clin Chim Acta* 2005;351:209–16.
- 28 Van Driest SL, Jaeger MA, Ommen SR, Will ML, Gersh BJ, Tajik AJ, Ackerman MJ. Comprehensive analysis of the beta-myosin heavy chain gene in 389 unrelated patients with hypertrophic cardiomyopathy. *J Am Coll Cardiol* 2004;44:602–10.
- 29 Yu B, Sawyer NA, Caramins M, Yuan ZG, Saunderson RB, Pamphlett R, Richmond DR, Jeremy RW, Trent RJ. Denaturing high performance liquid chromatography: high throughput mutation screening in familial hypertrophic cardiomyopathy and SNP genotyping in motor neurone disease. *J Clin Pathol* 2005;58:479–85.
- 30 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.
- 31 Millat G, Bouvagnet P, Chevalier P, Dauphin C, Jouk PS, Da CA, Prieur F, Bresson JL, Faivre L, Eicher JC, Chassaing N, Crehalet H, Porcher R, Rodriguez-Lafresse 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.
- 32 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.
- 33 Christiaans I, Nannenber EA, Dooijes D, Jongbloed RJ, Michels M, Postema PG, Majoor-Krakauer D, van den Wijngaard A, Mannens MM, van Tintelen JP, van Langen IM, Wilde AA. Founder mutations in hypertrophic cardiomyopathy patients in the Netherlands. *Neth Heart J* 2010;18:248–54.
- 34 Hougs L, Havndrup O, Bundgaard H, Kober L, Vuust J, Larsen LA, Christiansen M, Andersen PS. One third of Danish hypertrophic cardiomyopathy patients with MYH7 mutations have mutations [corrected] in MYH7 rod region. *Eur J Hum Genet* 2005;13:161–5.
- 35 Klaassen S, Probst S, Oechslin E, Gerull B, Krings G, Schuler P, Greutmann M, Hurlimann D, Yegitbasi M, Pons L, Gramlich M, Drenckhahn JD, Heuser A, Berger F, Jenni R, Thierfelder L. Mutations in sarcomere protein genes in left ventricular noncompaction. *Circulation* 2008;117:2893–901.
- 36 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.
- 37 Szczesna-Cordary D, Guzman G, Ng SS, Zhao J. Familial hypertrophic cardiomyopathy-linked alterations in Ca<sup>2+</sup> binding of human cardiac myosin regulatory light chain affect cardiac muscle contraction. *J Biol Chem* 2004;279:3535–42.
- 38 Kapplinger JD, Tester DJ, Alders M, Benito B, Berthet M, Brugada J, Brugada P, Fressart Y, Guerschicoff A, Harris-Kerr C, Kamakura S, Kyndt F, Koopmann TT, Miyamoto Y, Pfeiffer R, Pollevick GD, Probst V, Zumhagen S, Vatta M, Towbin JA, Shimizu W, Schulze-Bahr E, Antzelevitch C, Salisbury BA, Guicheney P, Wilde AA, Brugada R, Schott JJ, Ackerman MJ. An international compendium of mutations in the SCN5A-encoded cardiac sodium channel in patients referred for Brugada syndrome genetic testing. *Heart Rhythm* 2010;7:33–46.
- 39 Mohler PJ, Rivolta I, Napolitano C, LeMaillet G, Lambert S, Priori SG, Bennett V. Nav1.5 E1053K mutation causing Brugada syndrome blocks binding to ankyrin-G and expression of Nav1.5 on the surface of cardiomyocytes. *Proc Natl Acad Sci USA* 2004;101:17533–8.
- 40 Marziliano N, Pilotto A, Grasso M, Pasotti M, Arbustini E. Deletion of Glu at codon 13 of the TCAP gene encoding the titin-cap-telethonin is a rare polymorphism in a large Italian population. *Mol Genet Metab* 2006;89:286–7.
- 41 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.
- 42 Knoll R, Kostin S, Klede S, Savvatis K, Klinge L, Stehle I, Gunkel S, Kotter S, Babicz K, Sohns M, Miocic S, Didie M, Knoll G, Zimmermann WH, Thelen P, Bickeboller H, Maier LS, Schaper W, Schaper J, Kraft T, Tschope C, Linke WA, Chien KR. A common MLP (muscle LIM protein) variant is associated with cardiomyopathy. *Circ Res* 2010;106:695–704.
- 43 Geier C, Gehmlich K, Ehler E, Hassfeld S, Perrot A, Hayess K, Cardim N, Wenzel K, Erdmann B, Krackhardt F, Posch MG, Osterziel KJ, Bublak A, Nagele H, Scheffold T, Dietz R, Chien KR, Spuler S, Furst DO, Nurnberg P, Ozcelik C. Beyond the sarcomere: CSRP3 mutations cause hypertrophic cardiomyopathy. *Hum Mol Genet* 2008;17:2753–65.
- 44 Scharner J, Brown CA, Bower M, Iannaccone ST, Khatri IA, Escolar D, Gordon E, Felice K, Crowe CA, Grosman C, Meriggioli MN, Asamoah A, Gordon O, Gnocchi VF, Ellis JA, Mendell JR, Zammit PS. Novel LMNA mutations in patients with Emery-Dreifuss muscular dystrophy and functional characterization of four LMNA mutations. *Hum Mutat* 2011;32:152–67.
- 45 Young J, Morbois-Trabut L, Couzinet B, Lascols O, Dion E, Bereziat V, Feve B, Richard I, Capeau J, Chanson P, Vigouroux C. Type A insulin resistance syndrome revealing a novel lamin A mutation. *Diabetes* 2005;54:1873–8.
- 46 Harakalova M, Mokry M, Hrdlickova B, Renkens I, Duran K, van Roekel H, Lansu N, van Roosmalen M, de Bruijn E, Nijman IJ, Kloosterman WP, Cuppen E. Multiplexed array-based and in-solution genomic enrichment for flexible and cost-effective targeted next-generation sequencing. *Nat Protoc* 2011;6:1870–86.
- 47 Almomani R, van der Heijden J, Ariyurek Y, Lai Y, Bakker E, van GM, Breuning MH, den Dunnen JT. Experiences with array-based sequence capture; toward clinical applications. *Eur J Hum Genet* 2011;19:50–5.
- 48 Raca G, Jackson C, Warman B, Bair T, Schimmenti LA. Next generation sequencing in research and diagnostics of ocular birth defects. *Mol Genet Metab* 2010;100:184–92.
- 49 Shen P, Wang W, Krishnakumar S, Palm C, Chi AK, Enns GM, Davis RW, Speed TP, Mindrinos MN, Scharfe C. High-quality DNA sequence capture of 524 disease candidate genes. *Proc Natl Acad Sci USA* 2011;108:6549–54.
- 50 Luo C, Tsementzi D, Kyrpides N, Read T, Konstantinidis KT. Direct comparisons of Illumina vs. Roche 454 sequencing technologies on the same microbial community DNA sample. *PLoS ONE* 2012;7:e30087.
- 51 Bansal V, Libiger O. A probabilistic method for the detection and genotyping of small indels from population-scale sequence data. *Bioinformatics* 2011;27:2047–53.