The Egyptian Journal of Medical Human Genetics 18 (2017) 381–387

HOSTED BY

Contents lists available at ScienceDirect

The Egyptian Journal of Medical Human Genetics

journal homepage: www.sciencedirect.com

Original article

A comparative study of mutation screening of sarcomeric genes (*MYBPC3*, *MYH7*, *TNNT2*) using single gene approach *versus* targeted gene panel next generation sequencing in a cohort of HCM patients in Egypt

Heba Sh. Kassem [a,c], Roddy Walsh [b,d], Paul J. Barton [b,d], Besra S. Abdelghany [a], Remon S. Azer [a], Rachel Buchan [b,d], Shibu John [b,d], Ahmed Elguindy [a], Sarah Moharem-ElGamal [a,e], Hala M. Badran [a,f], Hoda Shehata [a,c], Stuart A. Cook [b,g,h,i], Magdi H. Yacoub [a,b,*]

[a] *MagdiYacoub Heart Foundation, Aswan Heart Centre, Egypt*
[b] *National Heart and Lung Institute, Imperial College London, London SW3 6LY, UK*
[c] *Alexandria Faculty of Medicine, Egypt*
[d] *NIHR Cardiovascular Biomedical Research Unit, Royal Brompton and Harefield NHS Foundation Trust and Imperial College London, London SW3 6NP, UK*
[e] *National Heart Institute, Cairo, Egypt*
[f] *Menofia Faculty of Medicine, Egypt*
[g] *MRC Clinical Sciences Centre, Imperial College London, London W12 0NN, UK*
[h] *National Heart Research Institute Singapore, National Heart Centre Singapore, Singapore 168752, Singapore*
[i] *Division of Cardiovascular & Metabolic Disorders, Duke-National University of Singapore, Singapore 169857, Singapore*

ARTICLE INFO

Article history:
Received 4 April 2017
Accepted 24 May 2017
Available online 8 June 2017

Keywords:
Genetics
Hypertrophic cardiomyopathy
Egypt
Sanger
Next generation sequencing

ABSTRACT

Background: NGS enables simultaneous sequencing of large numbers of associated genes in genetic heterogeneous disorders, in a more rapid and cost-effective manner than traditional technologies. However there have been limited direct comparisons between NGS and more established technologies to assess the sensitivity and false negative rates of this new approach. The scope of the present manuscript is to compare variants detected in *MYBPC3*, *MYH7* and *TNNT2* genes using the stepwise dHPLC/Sanger *versus* targeted NGS.

Methods: In this study, we have analysed a group of 150 samples of patients from the Bibliotheca Alexandrina-Aswan Heart Centre National HCM program. The genetic testing was simultaneously undertaken by high throughput denaturing high-performance liquid chromatography (dHPLC) followed by Sanger based sequencing and targeted next generation deep sequencing using panel of inherited cardiac genes (ICC). The panel included over 100 genes including the 3 sarcomeric genes. Analysis of the sequencing data of the 3 genes was undertaken in a double blinded strategy.

Results: NGS analysis detected all pathogenic and likely pathogenic variants identified by dHPLC (50 in total, some samples had double hits). There was a 0% false negative rate for NGS based analysis. Nineteen variants were missed by dHPLC and detected by NGS, thus increasing the diagnostic yield in this co- analysed cohort from 22.0% (33/150) to 31.3% (47/150).

Of interest to note that the mutation spectrum in this Egyptian HCM population revealed a high rate of homozygosity in *MYBPC3* and *MYH7* genes in comparison to other population studies (6/150, 4%). None of the homozygous samples were detected by dHPLC analysis.

Conclusion: NGS provides a useful and rapid tool to allow panoramic screening of several genes simultaneously with a high sensitivity rate amongst genes of known etiologic role allowing high throughput analysis of HCM patients and relevant control series in a less characterised population.

1. Introduction

Hypertrophic cardiomyopathy (HCM) is a genetic cardiac disease characterised by left ventricular hypertrophy and myofibrillar disarray [1]. HCM is typically inherited in an autosomal dominant manner and, with an estimated worldwide reported prevalence of

Abbreviations: NGS, next generation sequencing; HCM, hypertrophic cardiomyopathy; MYBPC3, myosin binding protein C; MYH7, myosin heavy chain 7; TNNT2, cardiac troponin T2; MYF, Magdi Yacoub Heart Foundation; BA, bibliotheca alexandrina.

Peer review under responsibility of Ain Shams University.

* Corresponding author at: Aswan Heart Center, Kasr El Haggar Street, Aswan, Egypt.
E-mail address: m.yacoub@imperial.ac.uk (M.H. Yacoub).

http://dx.doi.org/10.1016/j.ejmhg.2017.05.002
1110-8630/

1:500, it is the most common inherited cardiac disease and a leading cause of premature sudden cardiac death (SCD) in the young [2].

Mutations in almost 50 genes have been implicated in HCM, including genes that encode for sarcomeric, Z-disc, calcium handling and mitochondrial proteins [3], although the evidence for many of these associations is doubtful [4]. Until recently however, genetic testing for HCM has focused on mutations in myosin heavy chain 7 (*MYH7*), myosin-binding protein C3 (*MYBPC3*) and cardiac troponin T (*TNNT2*), which account for between 35% and 60% of cases [5]. Traditionally such studies have been performed by single gene analysis, utilising either polymerase chain reaction (PCR) or denaturing high performance liquid chromatography (dHPLC) followed by direct Sanger sequencing. Using these approaches, HCM mutation profiles from many different population groups have been published over the last decade [6–18], including an Egyptian cohort of 192 HCM patients [19].

The recent advent of next generation sequencing (NGS) technologies has the power to revolutionise the genetic diagnosis of heterogeneous diseases, such as HCM. NGS can be used for either targeted gene panel or whole genome/exome sequencing [20], and enables rapid and cost-effective analysis of the three core HCM genes, as well as more comprehensive screening of all genes associated with HCM. However, until now there have been limited studies to assess the sensitivity of NGS and to compare its ability to detect variants with more established and traditional methods. In the present study, we have compared variant detection in the three commonly involved sarcomeric genes: *MYH7*, *MYBPC3* and *TNNT2* in a cohort of 150 Egyptian HCM patients using dHPLC/Sanger based analysis followed by NGS targeted deep sequencing assay.

2. Patients and methods

2.1. HCM patient population

The present study comprised 150 unrelated HCM index patients recruited from the different geographic regions of Egypt. Patients were diagnosed according to standard clinical evaluation by 2D echocardiography at Aswan Heart Centre and satellite HCM clinics as part of National BA HCM National Program. Demographic and clinical data of this cohort was described in our earlier paper [19]. The samples analysed were included based on availability of sufficient DNA for analysis by both assays.

The present study has been carried out in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki) and was approved by Aswan Heart Centre Research Ethics Committee. Research subjects provided informed consent for participation in the study.

Genomic DNA was extracted from 4 to 6 ml of blood samples donated from index HCM patients and control subjects using Promega wizard genomic DNA purification Kit. Primers used for amplification of the coding exons and flanking intronic sequences of the three candidate sarcomeric genes: MYH7 (NM_000257.2, MYBPC3 (NM_000256.3), and TNNT2 (NM_001001430.1) were previously described [9,21,22].

New primers set used for amplifying exon 5 were designed in the present study to bypass the intronic deletion polymorphism (c.506-12delC) commonly encountered among Egyptians (present at a frequency of 23%, as reported in our earlier study [19]) to enable screening of the full span of exon 5 (MyBPC3 exon 5 new primers: F primer TTTGCAGGTGGCAGCAT; R primer GTCCCCT CTCTCCGTGTCTCC).

2.2. dHPLC/Sanger analysis

Initial screening for mutations in the three candidate sarcomeric genes: *MYH7* (NM_000257.2, 38 coding exons, 3rd to 40th exons), *MYBPC3* (NM_000256.3, 34 exons, 1st to 34th), and *TNNT2* (NM_001001430.1, 16 coding exons, 2nd to 17th) was undertaken using heteroduplex analysis by dHPLC using WAVE™, Transgenomics, DNA Fragment Analysis System. The conditions for dHPLC were developed on the basis of amplicon-specific melting profiles predicted by the NAVIGATOR software (Transgenomics, San Jose, California, USA).

Samples showing a variant profile (different from the wild type pattern) in any of the amplicons were subjected to bidirectional sequencing using automated dye terminator cycle- capillary electrophoresis (ABI 3500 Applied Biosystems, Foster City, California, USA) to determine the nature of the sequence change, as shown in Fig 1.

2.3. NGS targeted resequencing

The samples were re-sequenced blindly using a custom capture assay (Agilent SureSelect), with an Illumina HiSeq2500 system which targets commonly involved sarcomeric genes (*MYBPC3*, *MYH7*, *TNNT2*) implicated in hypertrophic cardiomyopathy. Reads were demultiplexed (allowing zero mismatches) with HiSeq Control software and quality were checked in FastQC v.0.10.1. Low quality (<20) bases were trimmed using PrinSeq v0.20.4 [23] and the reads were aligned to hg19 reference using BWA v0.7.5 [24].

Marking duplicate reads, local realignment around indels and base quality score recalibration process were done in Picard v1.109 and GATK v2.8-1[25]. Alignment summary metrics and callability and coverage reports were calculated using Picard, Samtools v0.1.18 [26], Bedtools v2.11.2 [27] and in house perl scripts.

A subset file (ontarget) was created, based on reads mapping quality >8. This "ontarget" file was used to make consistent variant calls with GATK HaplotypeCallerandUnifiedGenotyper. Bases covered by at least 10 reads with a mapping quality ≥10 and base quality ≥20 were denoted as "callable", i.e. adequately covered for variant calling with recommended GATK parameters [28]. Variants were annotated using the Ensembl API v75_37 [29] and Human Gene Mutation Database (HGMD) Professional version 2013.4 [30].

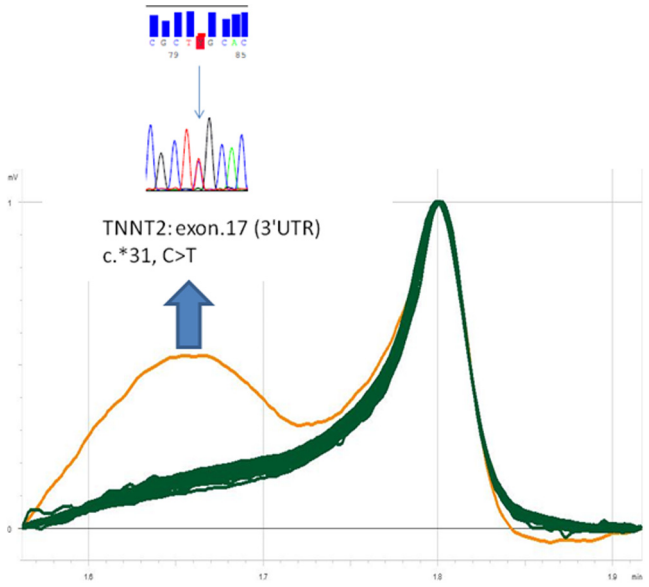

Fig. 1. Electropherogram of dHPLC screening of TNNT2 exon 17, sample showing a variant profile (in orange) is subjected to Sanger based sequencing analysis to identify the nature of the sequence change.

3. Results

The current cohort were co-analysed by two approaches of complimentary dHPLC/Sanger and by targeted resequencing NGS techniques. Both analyses were undertaken independently and blindly.

Twenty-four likely pathogenic variants were detected in *MYH7*, *MYBPC3* and *TNNT2* by dHPLC/Sanger, in 35 patients (Table 1). Two patients had double hits, rendering a diagnostic yield using dHPLC/Sanger analysis of 22% (33/150) in this cohort.

The 150 samples were then analysed on an NGS targeted panel for deep sequencing of the three sarcomeric genes, with the average percentage of coding sequence considered callable as follows: *MYH7*: 97.5%, *MYBPC3*: 99.8%, *TNNT2*: 99.9%. All variants identified by dHPLC/Sanger were also detected using NGS, representing a 0% false negative rate for this technique in this study. In addition, a further 21variants were not identified by dHPLC/Sanger and were detected only by NGS (Table 1).

All variants detected only by NGS were subsequently confirmed by Sanger based sequencing using capillary electrophoresis. NGS assay showed a 0% false positive rate for these three genes. In total, fourteen samples were classified as having putative pathogenic variant over the number of patients detected by dHPLC/Sanger. Hence, NGS contributed towards a higher positive diagnostic yield of 31.3% (47/150 patients) in comparison to the yield of dHPLC (22%%, 33/150) in the same co-analysed cohort (Fig 2).

3.1. Discrepancies between dHPLC and NGS

The 19 individual variants that were detected only by NGS and missed by dHPLC were investigated to determine why the initial screening had failed to detect them.

The dHPLC profile was similar to the wild type profile for 12 variants. Fig 3 shows examples of samples with variants not picked by dHPLC as a different profile from the wild. This is a known limitation in sensitivity of dHPLC technique, with a reported sensitivity of 95%[31].

Detailed analysis of samples not detected by dHPLC showed six variants were present in a homozygous state (c.1227-2A>G, c.1321G>A, c.2458C>G, c.2618C>A, c.3676C>T in *MYBPC3* and c.1064C>T in *MYH7*). Homozygosity is not likely to be detectable by dHPLC as this method relies on detection of mismatched heteroduplexes which is only possible to be detected in DNA of heterozygous individuals (Fig 3B) and is considered a limitation in the dHPLC assay.

Additional issues included the initial primer design for *MYBPC3* exon 5 covered a region enclosing the commonly encountered SNP c.506-12delC (rs11570050), which occurs at a frequency of 0.23 in Egyptians, [19]. The presence of the allele with the deletion provided a technical limitation in reading exon 5 in full. In the patient sample, the variant of interest with relevance to pathogenecity, c.649A > G, occurred in cis with the c.506delC SNP. Sanger sequencing failed to identify the variant with the earlier described primer set [21]. However following detection of the variant by NGS, it was subsequently confirmed by Sanger sequencing with redesigned primers (described above in Materials and methods section).

3.2. Classification of MYH7, MYBPC3 and TNNT2 variants

In total, 50 variants were detected in 61 patient samples by NGS, 35 were scored as putative pathogenic, 12 were scored as variants of uncertain significance (VUS) and 3 were scored as likely benign. The variants detected in this study were assessed for their likely pathogenic effect based on variant type, previous reports linking them to HCM, and variant frequencies reported in web based publically accessed databases [32]. In addition to testing for segregation in affected families whenever possible as illustrated in pedigrees of Fig 4.

Truncating variants (nonsense, frameshift or essential splice site) or missense variants not present at a frequency >0.01% in ExAC [33] and absent from 200 population matched controls were considered as putatively pathogenic. Any variant present at a frequency of >0.1% in ExAC and not previously linked to disease was considered to be likely benign (Table 1). All others were classified as variants of unknown significance (VUS), which included two variants, *MYBPC3*:c.148A>G in sample PM48/PA45 since it did not co-segregate in affected family members of PM48, and also *MYBPC3*:c.3800G>A a second hit in sample PA4 which was non co-segregating in two affected family members in addition to the segregating pathogenic mutation p.R719Q detected in all affected family members (Fig 4).

MYBPC3 contributed for almost half of the variants detected in the present cohort. The likely pathogenic variants detected included 10 missense and 5 truncating variants including 2 non sense mutations, an essential splice site variant c.1227-2A>G (in a homozygous state) and two frame shift mutations (A179QfsX59, D506TfsX7). The latter two variants were each detected in 3 Egyptian HCM patients in addition to a nonsense mutation, p.W1098X, was detected in 4 unrelated index patients.

There were 18 putative pathogenic variants detected in *MYH7* in 23 patients, and PA8 had double hits. There were 16 missense and two truncating variants, one is an essential splice variant c.5791-1G>A and another is frame shift mutation, p.S1924AfsX9. The latter mutation was detected in 5 patients in the current cohort (5/150, 3%) and was absent in population matched controls and other population databases. Most of the missense variants were detected in the head motor region of *MYH7* (14/16, 87.5%) which is a recognised mutation hotspot [33]. In *TNNT2*, 2 likely pathogenic missense variants were detected in 3 patients (Table 1).

Of the 35 distinct pathogenic variants, 21 have previously been linked to cardiomyopathy and 14 are novel and described only in the Egyptian HCM patients Several patients had more than a single variant, in addition to the six homozygous patients (highlighted by an asterisk in Table 1). Three patients were double heterozygotes. One patient (PA26) had missense variants in *MYBPC3* (c.2311G>A, p.V771M) and also in *MYH7* (c.799C>G, p.L267V), the second (PM43) had a frameshift variant in *MYH7* (c.5769delG) and a missense variant in *MYBPC3* (c.2470G>A, p.D824N), the third (PA8) had 2 hits in *MYH7* (c.925G>A, p.D309N & c.5769delG) These double hits provides a possible explanation for the phenotype heterogeneity commonly described in HCM.

4. Discussion

Next generation sequencing is an immensely powerful approach to genetic analysis with the potential to transform diagnosis of genetically heterogeneous diseases like HCM. Targeted resequencing of a panel of known cardiomyopathy genes allows for a comprehensive genetic screening in a relatively quick and inexpensive manner. NGS approaches are now being used in diagnostic laboratories, though most continue to validate any findings with targeted mutation Sanger sequencing. However the sensitivity of NGS in a diagnostic setting, and an evaluation of the false negative rate of this technology, remains difficult to assess as traditional sequencing approaches and NGS are rarely run in parallel.

In this study, we have compared the performance of NGS with a well characterised and clinically established sequencing method (dHPLC/Sanger) in detecting variants in the three major HCM genes in a cohort of Egyptian HCM patients. NGS detected all 50variants

Table 1
Details of the variants detected in this study by both dHPLC/Sanger and NGS and by NGS alone in the three genes analysed.

Gene	Variant nomenclature at coding DNA level	Variant nomenclature at protein level	Variant type	Num. Samples (IDs)		ExAC count (freq.)	Classification of pathogenicity	Novelty of putative pathogenic variants
				Both	NGS			
MYBPC3	c.355G>A	p.E119K	Missense	1 (PM64)		0	Putative pathogenic	No
MYBPC3	c.416C>G	p.S139X	Nonsense	1 (PA15)		0	Putative pathogenic	Yes (this cohort only)
MYBPC3	c.534_541del	p.A179QfsX59	Frameshift	3 (PA25, PA46, PA52)		0	Putative pathogenic	Yes (this cohort only)
MYBPC3	c.772G>A	p.E258K	Missense	1 (PA59)		3 (0.00003903)	Putative pathogenic	No
MYBPC3	c.956A>C	p.E319A	Missense	1 (PA33)		6 (0.00007891)	Putative pathogenic	No
MYBPC3	c.1227-2A>G		Essential Splice Site		1* (P90)	0	Putative pathogenic	No
MYBPC3	c.1516delG	p.D506TfsX7	Frameshift	3 (P63, P70, P62)		0	Putative pathogenic	Yes (this cohort only)
MYBPC3	c.2308G>A	p.D770N	Missense		1 (P88)	1 (0.00000830)	Putative pathogenic	No
MYBPC3	c.2311G>A	p.V771M	Missense	1 (PA26)		1 (0.00005661)	Putative pathogenic	No
MYBPC3	c.2458C>G	p.R820G	Missense		1* (PA20)	0	Putative pathogenic	No
MYBPC3	c.2470G>A	p.D824N	Missense	1 (PM43)		1 (0.00000829)	Putative pathogenic	No
MYBPC3	c.2977C>T	p.L993F	Missense	1 (PA16)		0	Putative pathogenic	Yes (this cohort only)
MYBPC3	c.3293G>A	p.W1098X	Nonsense	4 (PA6, PA42, PA56, PA93)		0	Putative pathogenic	Yes
MYBPC3	c.3412C>T	p.R1138C	Missense	1 (PU11)		3 (0.00003719)	Putative pathogenic	No
MYBPC3	c.3676C>T	p.R1226C	Missense		1* (PM34)	7 (0.00005805)	Putative pathogenic	No
MYH7	c.632C>T	p.P211L	Missense		1 (PA27)	3 (0.00002471)	Putative pathogenic	No
MYH7	c.665A>G	p.Q222R	Missense		1 (P3)	0	Putative pathogenic	No
MYH7	c.746G>A	p.R249Q	Missense	1 (PA55)		0	Putative pathogenic	No
MYH7	c.799C>G	p.L267V	Missense	1 (PA26)		0	Putative pathogenic	Yes (this cohort only)
MYH7	c.925G>A	p.D309N	Missense	1 (PA8)	1 (PA40)	3 (0.00002486)	Putative pathogenic	No
MYH7	c.1064C>T	p.A355V	Missense		1* (P80)	0	Putative pathogenic	No
MYH7	c.1182C>A	p.D394E	Missense	1 (PA54)		0	Putative pathogenic	Yes (this cohort only)
MYH7	c.1816G>A	p.V606M	Missense		1 (P35)	0	Putative pathogenic	Yes
MYH7	c.2147G>C	p.G716A	Missense	1 (PA30)		0	Putative pathogenic	Yes (this cohort only)
MYH7	c.2155C>T	p.R719W	Missense	1 (PM60)		0	Putative pathogenic	No
MYH7	c.2156G>A	p.R719Q	Missense	2 (PA4, PM68)		0	Putative pathogenic	No
MYH7	c.2389G>A	p.A797T	Missense	1 (P16)		4 (0.00003296)	Putative pathogenic	No
MYH7	c.2609G>A	p.R870H	Missense		1 (P77)	0	Putative pathogenic	No
MYH7	c.2779G>A	p.E927K	Missense	1 (PA12)		0	Putative pathogenic	No
MYH7	c.4145G>A	p.R1382Q	Missense	1 (PM65)		0	Putative pathogenic	No
MYH7	c.4258C>T	p.R1420W	Missense		1 (PM24)	1 (0.00000824)	Putative pathogenic	No
MYH7	c.5769delG	p.S1924AfsX9	Frameshift	3 (PA24, PM3, PM43)	2 (PA11, PA8)	0	Putative pathogenic	Yes (this cohort only)
MYH7	c.5791-1G>A		Essential Splice Site		1 (P33)	0	Putative pathogenic	No
TNNT2	c.221T>C	p.L74S	Missense	2 (PM1, PM12)		0	Putative pathogenic	No
TNNT2	c.857G>A	p.R286H	Missense	1 (P102)		7 (0.00007882)	Putative pathogenic	No
MYBPC3	c.148A>G	p.S50G	Missense	1 (PM48)	1 (PA45)	3 (0.00003887)	VUS[1]	NA
MYBPC3	c.530G>A	p.R177H	Missense	1 (P62)		92 (0.00094933)	VUS	NA
MYBPC3	c.1321G>A	p.E441K	Missense	4 (PA38, PA123, P12, P101)	1* (P65)	18 (0.00016221)	VUS	NA
MYBPC3	c.1458-17C>G		Splice Region	1 (P17)		0	VUS	NA
MYBPC3	c.1458-7C>A		Splice Region	1 (P89)		0	VUS	NA
MYBPC3	c.1564G>A	p.A522T	Missense		1 (PA26)	47 (0.00039070)	VUS	NA
MYBPC3	c.2149-8C>T		Splice Region	1 (P100)		2 (0.00002204)	VUS	NA
MYBPC3	c.2618C>A	p.P873H	Missense		1* (P20)	13 (0.00020272)	VUS	NA

Table 1 (*continued*)

Gene	Variant nomenclature at coding DNA level	Variant nomenclature at protein level	Variant type	Num. Samples (IDs)		ExAC count (freq.)	Classification of pathogenicity	Novelty of putative pathogenic variants
				Both	NGS			
MYBPC3	c.3800G>A	p.R1267H	Missense		1 (PA4)	5 (0.00004239)	VUS[1]	NA
MYH7	c.4520-3C>T		Splice Region	1 (PA1)		11 (0.00009061)	VUS[2]	NA
TNNT2	c.690-4G>T		Splice Region	1 (P32)		19 (0.00015650)	VUS	NA
TNNT2	c.832C>T	p.R278C	Missense	1 (PM68)		40 (0.00042906)	VUS	NA
MYBPC3	c.472G>A	p.V158M	Missense	2 (P14, P16)		2348 (0.09043291)	Likely benign	NA
MYBPC3	c.649A>G	p.S217G	Missense		2 (PM46, PA70)	248 (0.00226757)	Likely benign	NA
MYBPC3	c.2149-5C>T		Splice Region	1 (P87)		118 (0.00127947)	Likely benign	NA

* – homozygous. 1 – variant does not segregate with disease in affected families. 2 – present at a frequency of 0.005 in Egyptian population control samples.

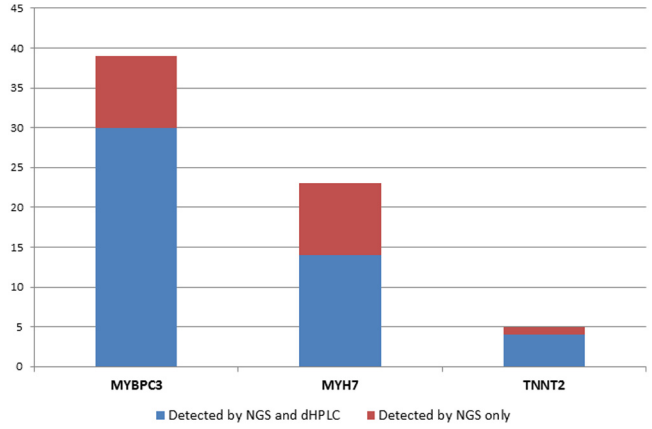

Fig. 2. Number of all variants (pathogenic, VUS and benign) detected in each gene by both dHPLC/Sanger and targeted NGS, or by targeted NGS only and confirmed by Sanger.

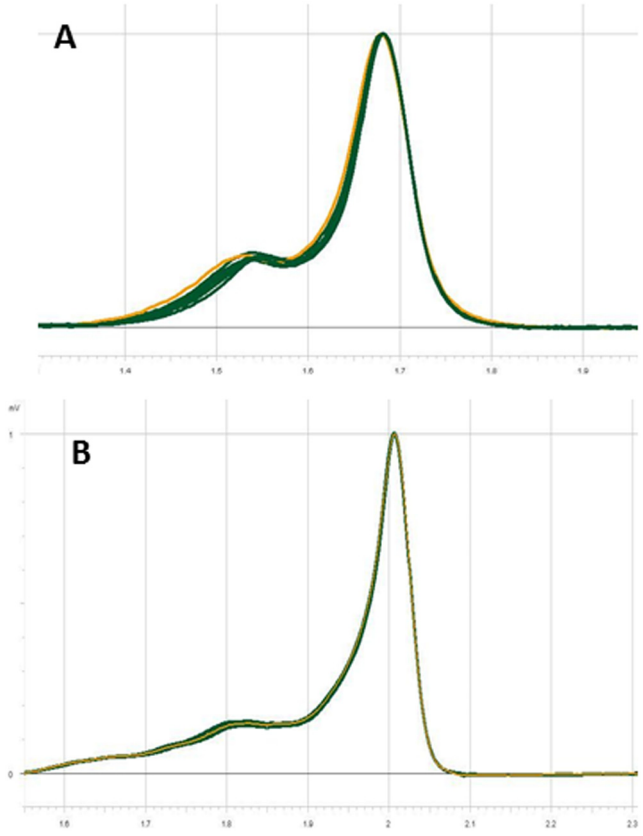

Fig. 3. dHPLC WAVE™ Analysis showing electropherogram of wild profile samples in GREEN and missed sample for variant detection in ORANGE. [A] *MYH7* exon 8: P3 (c.665A>G; p.Q222R), heterozygous. [B] *MYBPC3* exon 15: P65 (c.1321G>A; p.E441K, homozygous).

identified by dHPLC, giving it a zero false negative rate in this study and indicating that this technology is sufficiently sensitive for use in diagnostic laboratories.

In addition to detecting all of the variants identified by dHPLC/Sanger, targeted NGS assay found an additional 19 variants in this cohort, increasing the genetic diagnostic yield for this cohort from 22% (33/150 patients) in comparison to 31.3% (47/150). This has highlighted the limitations in the sensitivity of dHPLC, both in detecting variants with a profile too similar to wild-type and in its intrinsic inability to detect homozygous variants. All variants identified solely by NGS were subsequently verified by Sanger sequencing, giving an overall false positive rate of 0% for NGS in the three genes analysed.

The issue of homozygous mutations is likely to be of particular importance in certain disease conditions and population groups. The consanguinity rate among Egyptians has been reported to be between 30% and 40% [34],indicating that homozygous variants are likely to be more prevalent among Egyptians than in other populations. Indeed in this study, six homozygous patients were identified (five in *MYBPC3*, one in *MYH7*). While there have been some reports of homozygous mutations linked to HCM in these three genes[6,13,18], the proportion of homozygosity in the Egyptian cohort (4%, 6/150) is significantly higher than in the reported studies. Accordingly, the mutation screening using dHPLC which detect variants through formation of heterduplexes could detect variants in heterzygosity including double and compound heterzgyote samples. However, such assay would have a limitation in detecting variants in homozygous status and would be of lower sensitivity in such populations with expected higher rate of homozgosity due to high prevalence of consanguinity. Utilising sequencing technologies, like NGS, that are able to detect these variants in comparison to primary mutation screening by dHPLC is critical for accurate genetic diagnosis in such highly consanguineous populations. The effect of homozygosity and mutation dosage on HCM

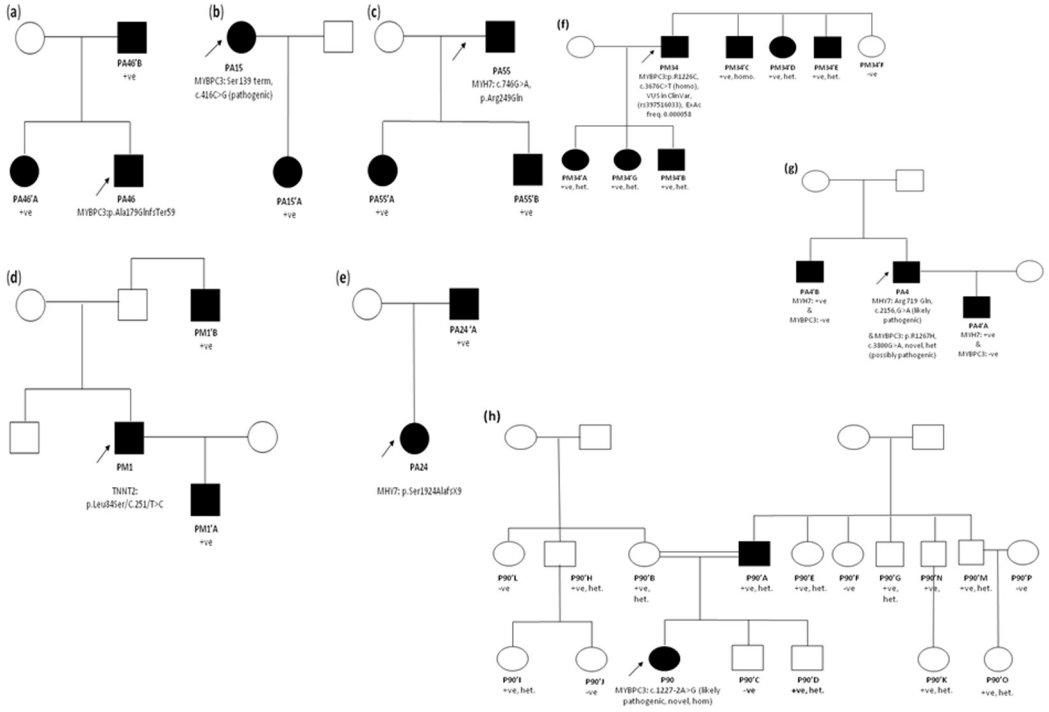

Fig. 4. Pedigrees of HCM patients showing segregation of sarcomeric variants with HCM phenotypte.

clinical phenotype is currently under investigation and will be reported in a future publication.

In a comparison with other population cohorts: there were some variants that were detected only in Egyptian HCM patients such as A179QfsX59, D506TfsX7 in *MYBPC3* and S1924AfsX9 in *MYH7*. Those variants were absent in all other population databases, and were absent in 200 population matched controls as reported in our earlier study [19]. These variants are anticipated as possibly pathogenic by frame shift nature. The fact that they were detected in several index patients may suggest founder effect, however further extended family and haplotype studies are needed to study this observation.

The frequency of involvement of the 3 sarcomeric genes observed in HCM patients in the present study is similar to that observed in Caucasian cohorts reported in Europe and the U.S. In our earlier report of 192 HCM patients, *MYBPC3* was found as the most prevalent HCM gene in respect to variant detection among Egyptian HCM patients [19]. Indeed, the variant detection frequencies in Egyptian patients is quite similar to most reported Caucasian populations [3,6–10,12–14] and Tunisian population[35]. In contrast, *MYH7* gene was the most prevalent HCM gene in published studies from East Asia [11,15–17] and South America [18]. In a comparison with other population cohorts, the pattern of the 3 sarcomeric genes involvement observed in the present study is similar to that observed in Caucasian cohorts reported in Europe and the U.S., in contrary to the reported commoner involvement of *MYH7* observed in Asian population.

However, for interest of population comparisons, this may be limited due to the fact that reported data had also included variants of uncertain significance and hence may not reflect a true difference among different populations. It should also be noted that yield for each gene in these studies is likely to be influenced by several factors such as the sensitivity of the sequencing technologies used, differences in the classification of pathogenic variants and the proportion of familial HCM cases in each cohort.

This study has several limitations which include the relatively small number of non-consecutive HCM samples and the fact that the pathogenic potential of the variants were determined by comparison to the ExAC database which included different ethnicity but no data as yet from the Egyptians. This is influential in determining the accurate positive yield in the 3 sarcomeric genes (*MYBPC3*, *MYH7*, *TNNT2*) and may result in downgrading of several previously described pathogenic variants to variants of uncertain significance (VUS), causing transient discrepancies among different population reports. It also highlights the urgency for including large numbers of population relevant control series in ExAC database to enable its relevance to different population studies. Furthermore, the proper interpretation of VUS necessitates undertaking functional assays and cosegregation studies in extended families to determine if those variants do indeed bear any pathogenic role. However, this limitation should not affect the validity of the comparison between the use of NGS and Sanger, in the same population from a technical perspective.

In conclusion, we have demonstrated that NGS technology can be used for accurate genetic diagnosis of HCM, significantly outperforming the previously used dHPLC high throughput screening technique with no false positives or false negatives detected in this study in relevance to confirmation by Sanger sequencing.

As we have shown, an NGS approach is particularly sensitive and advantageous for defining the specific genetic architecture of the studied cohort. This comparative study and the earlier study by Chin et al. [20] support use of NGS targeted genes panel deep sequencing in a clinical diagnostic setting for genetically heterogeneous single gene disorders, such as inherited cardiomyopathies, however Sanger sequencing remain as indispensable tool for confirmation of variant detection and family members screening.

Conflict of interest

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

Acknowledgements

This work is supported by Magdi Yacoub Heart Foundation and a grant from Egyptian Ministry of Scientific Research-Science and Technology Development STDF Grant Number 5309. Also by the NIHR Cardiovascular Biomedical Research Unit at Royal Brompton & Harefield NHS Foundation Trust and Imperial College London. We thank the Bibliotheca Alexandrina for support of the BA MYF Genetics Lab.

References

[1] Wigle ED. Novel insights into the clinical manifestations and treatment of hypertrophic cardiomyopathy. Curr Opin Cardiol 1995;10:299–305.
[2] Maron BJ, Gardin JM, Flack JM, Gidding SS, Kurosaki TT, Bild DE. Prevalence of hypertrophic cardiomyopathy in a general population of young adults. Echocardiographic analysis of 4111 subjects in the CARDIA Study. Coronary Artery Risk Development in (Young) Adults. Circulation 1995;92:785–9.
[3] Lopes LR, Zekavati A, Syrris P, Hubank M, Giambartolomei C, Dalageorgou C, et al. Genetic complexity in hypertrophic cardiomyopathy revealed by high-throughput sequencing. J Med Genet 2013;50:228–39.
[4] Walsh R, Buchan R, Wilk A, John S, Felkin LE, Thomson KL, et al. Defining the genetic architecture of hypertrophic cardiomyopathy: re-evaluating the role of non-sarcomeric genes. Eur Heart J 2017. doi: http://dx.doi.org/10.1093/eurheartj/ehw603.
[5] 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.
[6] Richard P, Charron P, Carrier L, Ledeuil C, Cheav T, Pichereau C, et al. Hypertrophic cardiomyopathy: distribution of disease genes, spectrum of mutations, and implications for a molecular diagnosis strategy. Circulation 2003;107:2227–32.
[7] Erdmann J, Daehmlow S, Wischke S, Senyuva M, Werner U, Raible J, et al. Mutation spectrum in a large cohort of unrelated consecutive patients with hypertrophic cardiomyopathy. Clin Genet 2003;64:339–49.
[8] Van Driest SL, Ellsworth EG, Ommen SR, Tajik AJ, Gersh BJ, Ackerman MJ. Prevalence and spectrum of thin filament mutations in an outpatient referral population with hypertrophic cardiomyopathy. Circulation 2003;108:445–51.
[9] Van Driest SL, Jaeger MA, Ommen SR, Will ML, Gersh BJ, Tajik AJ, et al. Comprehensive analysis of the beta-myosin heavy chain gene in 389 unrelated patients with hypertrophic cardiomyopathy. J Am Coll Cardiol 2004;44:602–10.
[10] Van Driest SL, Vasile VC, Ommen SR, Will ML, Tajik AJ, Gersh BJ, et al. Myosin binding protein C mutations and compound heterozygosity in hypertrophic cardiomyopathy. J Am Coll Cardiol 2004;44:1903–10.
[11] Song L, Zou Y, Wang J, Wang Zh, Zhen Y, Lou K, et al. Mutations profile in Chinese patients with hypertrophic cardiomyopathy. Clin Chim Acta 2005;351:209–16.
[12] Olivotto I, Girolami F, Ackerman MJ, Nistri S, Bos JM, Zachara E, et al. Myofilament protein gene mutation screening and outcome of patients with hypertrophic cardiomyopathy. Mayo Clin Proc 2008;83:630–8.
[13] Garcia-Castro M, Coto E, Reguero JR, Berrazueta JR, Alvarez V, Alonso B, et al. Mutations in Sarcomeric Genes *MYH7*, *MYBPC3*, *TNNT2*, *TNNI3*, and *TPM1* in Patients With Hypertrophic Cardiomyopathy. Rev Esp Cardiol (Engl Ed) 2009;62:48–56.
[14] Millat G, Bouvagnet P, Chevalier Ph, Dauphin C, Jouk PS, Da Costa A, et al. Prevalence and spectrum of mutations in a cohort of 192 unrelated patients with hypertrophic cardiomyopathy. Eur J Med Genet 2010;53:261–7.
[15] Otsuka H, Arimura T, Abe T, Kawai H, Aizawa Y, Kubo T, et al. Prevalence and distribution of sarcomeric gene mutations in Japanese patients with familial hypertrophic cardiomyopathy. Circ J 2012;76:453–61.
[16] Liu W, Liu W, Hu D, Zhu T, Ma Z, Yang J, et al. Mutation spectrum in a large cohort of unrelated Chinese patients with hypertrophic cardiomyopathy. Am J Cardiol 2013;112:585–9.
[17] Zou Y, Wang J, Liu X, Wang Y, Chen Y, Sun K, et al. Multiple gene mutations, not the type of mutation, are the modifier of left ventricle hypertrophy in patients with hypertrophic cardiomyopathy. Mol Biol Rep 2013;40:3969–76.
[18] Marsiglia JD, Credidio FL, de Oliveira TG, Reis RF, Antunes Mde O, de Araujo AQ, et al. Screening of *MYH7*, *MYBPC3*, and *TNNT2* genes in Brazilian patients with hypertrophic cardiomyopathy. Am Heart J 2013;166:775–82.
[19] Kassem HSh, Azer RS, Ayad MS, Moharem-Elgamal S, Magdy G, Elguindy A, et al. Early results of sarcomeric gene screening from the Egyptian National BA-HCM Program. J Cardiovasc Transl Res 2013;6:65–80.
[20] Chin EL, da Silva C, Hegde M. Assessment of clinical analytical sensitivity and specificity of next-generation sequencing for detection of simple and complex mutations. BMC Genet 2013;19:6–14.
[21] Carrier L, Bonne G, Bahrend E, Yu B, Richard P, Niel F, et al. Organization and sequence of human cardiac myosin binding protein C gene (MYBPC3) and identification of mutations predicted to produce truncated proteins in familial hypertrophic cardiomyopathy. Circ Res 1997;80:427–34.
[22] Farza H, Townsend PJ, Carrier L, Barton PJ, Mesnard L, Bahrend E, et al. Genomic organisation, alternative splicing and polymorphisms of the human cardiac troponin T gene. J Mol Cell Cardiol 1998;30:1247–53.
[23] Schmieder R, Edwards R. Quality control and preprocessing of metagenomic datasets. Bioinformatics 2011;27:863–4.
[24] Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 2009;25:1754–60.
[25] McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 2010;20:1297–303.
[26] Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics 2009;25:2078–9.
[27] Quinlan AR, Hall IM. Hall, BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 2010;26:841–2.
[28] Pua CJ, Bhalshankar J, Miao K, Walsh R, John S, Lim SQ, et al. Development of a comprehensive sequencing assay for inherited cardiac condition genes. J Cardiovasc Transl Res 2016;9:3–11. doi: http://dx.doi.org/10.1007/s12265-016-9673-5.
[29] McLaren W, Pritchard B, Rios D, Chen Y, Flicek P, Cunningham F. Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. Bioinformatics 2010;26:2069–70.
[30] Stenson PD, Mort M, Ball EV, Shaw K, Phillips AD, Cooper DN. The Human Gene Mutation Database: building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine. Hum Genet 2014;133:1–9.
[31] Girolami F, Olivotto I, Passerini I, Zachara E, Nistri S, Re F, et al. 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 2006;7:601–7.
[32] Kassem HSh, Girolami F, Sanoudou D. Molecular genetics made simple. Glob Cardiol Sci Pract 2012;1:1–5. doi: http://dx.doi.org/10.5339/gcsp.2012.6.
[33] Walsh R, Thomson KL, Ware JS, Funke BH, Woodley J, McGuire KJ, et al. Reassessment of Mendelian gene pathogenicity using 7,855 cardiomyopathy cases and 60,706 reference samples. Genet Med 2017;19:192–203.
[34] Temtamy S, Aglan M. Consanguinity and genetic disorders in Egypt. Middle East J Med Genet 2012;1:12–7.
[35] Jaafar N, Go'mez J, Kammoun I, Zairi I, Ben Amara W, Kachboura S, et al. Spectrum of mutations in hypertrophic cardiomyopathy genes among tunisian patients. Genet Test Mol Biomark 2016;11:674–9.