Deletions of specific exons of *FHOD3* detected by next-generation-sequencing are associated with hypertrophic cardiomyopathy

Short running title: *FHOD3* deletions associated with HCM

Authors: Juan P. Ochoa[1,2], Luis R. Lopes[3,4,5], Marlene Pérez-Barbeito[2], Laura Cazón[2], María M. de la Torre-Carpente[6], Natalia Sonicheva-Paterson[2], David de Uña-Iglesias[1,2], Ellen Quinn[4], Svetlana S. Kuzmina-Krutetskaya[7], Jose A. Garrote[8], Perry M. Elliott[3,4,5], Lorenzo Monserrat[2].

Author's affiliations:
1. Cardiovascular Research Group. Biomediacal Research Institute (INIBIC), Complexo Hospitalario Universitario de A Coruña (CHUAC), Sergas. A Coruña University (UDC).
2. Health in Code S.L. Scientific Department. A Coruña (Spain).
3. Institute of Cardiovascular Science, University College London
4. Barts Heart Centre, St. Bartholomew's Hospital, Barts Health NHS Trust. London (UK)
5. European Reference Network for Rare and Low Prevalence Complex Diseases of the Heart (ERN GUARD-HEART; http://guardheart.ern-net.eu/).
6. Cardiology. Hospital Universitario Rio Hortega. Valladolid (Spain).
7. Cardiovascular teaching department of North-Western State Medical University named after I.I. Mechnikov. St. Petersburg (Russia)
8. Molecular Genetics Laboratory. Service of Laboratory Medicine. Hospital Universitario Rio Hortega. Valladolid (Spain).

Funding sources: LRL is funded by a Medical Research Council (MRC).

Conflict of Interest Statement: Juan P. Ochoa, Marlene Perez-Barbeito, Natalia Sonicheva-Paterson, and Laura Cazón-Varela are employees of Health in Code SL. Lorenzo Monserrat is a stakeholder and CEO of Health in Code SL. All of the remaining authors have no conflicts of interest to declare.

Data Availability Statement: The data that support the findings of this study are available on request from the corresponding author, JPO. The data are not publicly available due to information that could compromise the privacy of research participants.

Address for Correspondence
Dr. Juan Pablo Ochoa
Health in Code SL, As Xubias s/n
Edificio O Fortín, Hospital Marítimo de Oza (15006), A Coruña, Spain.
Telephone: +34 881 600 003
Fax: +34 981 167 093
E-mail: juanpablo.ochoa@healthincode.com
Twitter: @OchoaJP

ABSTRACT:

Despite new strategies, such as evaluating deep intronic variants and new genes in whole-genome-sequencing studies, the diagnostic yield of genetic testing in hypertrophic cardiomyopathy (HCM) is still around 50%. *FHOD3* has emerged as a novel disease-causing gene for this phenotype, but the relevance and clinical implication of copy-number-variations (CNVs) have not been determined. In this study, CNVs were evaluated using a comparative depth-of-coverage strategy by NGS in 5,493 hypertrophic cardiomyopathy probands and 2,973 disease-controls. We detected three symmetrical deletions in *FHOD3* that involved exons 15 and 16 in three HCM families (no CNVs were detected in the control group). These exons are part of the diaphanous inhibitory domain of FHOD3 protein, considered a cluster of mutations for HCM. The clinical characteristics of the affected carriers were consistent with those reported in *FHOD3* in previous studies.

This study highlights the importance of performing CNV analysis systematically in NGS genetic testing panels for HCM, and reinforce the relevance of the *FHOD3* gene in the disease.

KEYWORDS: Cardiomyopathies; Cardiomyopathy, Hypertrophic; Formins; FHOD3 protein, human; Next-Generation Sequencing; DNA Copy Number Variations; Genetic Testing

INTRODUCTION

Hypertrophic cardiomyopathy (HCM) is the most frequent inherited genetic heart disease, affecting one in 250-500 individuals of the general population. Massive parallel sequencing has contributed to the understanding of the disease, allowing the sequencing of a large number of genes in a short time at a sustainable cost. However, the diagnostic yield of genetic studies in HCM is around 50%[1]. New strategies to increase the yield of genetic testing include the sequencing of deep intronic variants, screening for copy number variations (CNVs) in disease-causing genes, and sequencing of a broader range of candidate genes. In any case, a high proportion of these currently genotype-negative patients might have a non-Mendelian form of the disease as it is suggested by the lower family history and a more benign disease course in such cases.

Mutations in the Formin Homology 2 Domain Containing 3 gene (*FHOD3*) have been recently identified as a possible cause for HCM[2]. Formins are proteins that are present in almost all cells and tissues due to their role in the polymerization of actin. FHOD3 – the protein encoded by *FHOD3* gene- is expressed almost exclusively in the heart muscle, participating in the development and maintenance of the thin filament of the sarcomere[3]. Most disease-causing variants reported to date are non-truncating (non-synonymous variants and small in-frame deletions). More recently, some variants affecting splicing were identified cosegregating with the phenotype in a few families[4]. However, the relevance and clinical implication of CNVs in this gene have not been evaluated.

METHODS

PATIENTS: From February 2014, *FHOD3* was evaluated by next-generation sequencing in 15,225 consecutive unrelated probands referred to our center for genetic

diagnosis. Samples were received mainly from centers in Spain, the United Kingdom, Denmark, the United States, Germany, and Argentina. The clinical phenotypes were established by each center prior to the genetic studies. 6,539 probands had a diagnosis or clinical suspicion of HCM. An additional 3,688 index cases with no evidence of structural cardiac disease (mainly channelopathies and aortic diseases) were used as an internal control population. The predominant ethnicity was European (more than 90% of the probands). Variant frequencies in the general population were extracted from the gnomAD database version r2.0 (http://gnomad.broadinstitute.org). The study protocol was approved by the Research Ethics Committee of A Coruña-Ferrol (registry code 2015/576).

DNA SEQUENCING AND CNV ANALYSIS:

Coding exons and intronic boundaries of 213 genes related to inherited cardiovascular diseases and sudden cardiac death were captured using a custom probe library (SureSelect Target Enrichment Kit for Illumina paired-end multiplexed sequencing method) and sequenced using the HiSeq 1500 platform (Illumina, San Diego, California, USA) following Illumina protocols. The complete list of genes is included in Table S1. The read depth (number of times that a base was sequenced by independent reads) for every nucleotide of genes of interest (including *FHOD3*) was greater than 30x (mean 250-400x). Exons that did not fulfill this standard were complementarily sequenced using the Sanger method. Sequence data analysis was done using a proprietary bioinformatics pipeline that includes sample demultiplexing as well as all the steps necessary to obtain a report of annotated variants together with their coverage and corresponding quality parameters. CNVs were analyzed using a comparative depth-of-coverage strategy[5] and were confirmed by an alternative method, including NGS strategies (amplicon sequencing analysis) and conventional methods (PCR

amplification and Sanger sequencing with specifically designed primers; Table S2). For Sanger sequencing, after cleanup, the amplified fragments were sequenced with a BigDye Terminator v3.1 Cycle Sequencing Kit and Applied Biosystems 3730 DNA Analyzer (Applied Biosystems) according to the manufacturer's instructions. Chromatograms were analyzed using Chromas Lite and the UCSC (University of California, Santa Cruz) Genome Browser tool. The identified variants were reported according to the HGVS (Human Genome Variant Society) nomenclature. The pathogenicity of identified variants was established according to the recommendations of the American College of Medical Genetics and Genomics (ACMG).

RESULTS:

From the 15,225 consecutive probands, CNVs were analyzed in 12,781 (84%). The other 2,444 samples were not suitable for CNV analysis because they did not pass the minimal bioinformatics quality filtering; most of these samples were derived from paraffin-embedded tissue blocks or saliva samples in which the quality of the DNA was not optimal.

CNVs were detected in *FHOD3* in three patients with a diagnosis of HCM (3 out of 5,493 HCM probands in which CNVs were suitable for analysis; 0.05%) and none of the 2,973 controls. They consisted of *FHOD3* deletions (at chromosome band 18q12.2) including one or two exons and part of flanking intronic sequences (Table S3). The affected exons 15 and 16 comprise residues p.Arg612-Phe657 and p.Ser657-Arg674, respectively. Since these deletions involve symmetric exons, the loss of genetic material would result in in-frame variants that do not change the reading frame of the protein. None of the probands carried additional genetic variants that could explain the phenotype. All the deletions were confirmed and the preliminary breakpoints

determined using additional NGS strategies (amplicon sequencing analysis); the precise breakpoints of these large deletions were defined at base-pair resolution by PCR amplification and Sanger sequencing as well (Fig. 1). No deletions involving exclusively exon 15 or 16 are reported in subjects considered controls from the gnomAD database.

DETAILED DESCRIPTION OF THE FAMILIES AND CHARACTERISTICS OF THE DELETIONS:

FAMILY A (Fig. 2A): The proband was a 51-year-old Spanish male with hypertension and paroxysmal atrial fibrillation who has been diagnosed with HCM at age 36, with no family history of HCM or sudden cardiac death (SCD). His last echocardiogram showed asymmetrical septal hypertrophy (maximum left ventricular wall thickness [MLVWT] 21 mm), moderate dilatation of the left atrium (50 mm in the parasternal long-axis plane) and no left ventricular (LV) outflow tract obstruction. His ejection fraction (EF) was normal (71%). His 26-year-old son, a carrier of the *FHOD3* deletion, and his 20-year-old daughter, a non-carrier, were evaluated and considered as clinically unaffected with normal echocardiogram and ECG.

The identified CNV in the patient consisted of a 2,678-bp deletion encompassing the nucleotide sequence from c.1,836-1,312 of intron 14 to c.1,970+1,231 of intron 15 of the *FHOD3* gene, with a 27-bp insertion at the junction (Fig 1A). The proximal breakpoint was localized adjacent to a 1,374bp LINE termed L1ME3C, and the distal breakpoint was embedded in a 304bp SINE termed AluSz (Fig. S1). On manual alignment, the inserted 27-bp fragment was identical to the reference sequence at chr18:g.34,268,898-34,268,924 position, which lies 6,134 bp downstream from the insertion breakpoint. Non-homologous DNA end-joining (NHEJ), the predominant DNA repair mechanism in mammals, does not require any microhomology between the

two DNA ends. Since we did not find extensive microhomology at the junctions, we proposed NHEJ as responsible in this case.

FAMILY B (Fig. 2B): The proband was a 61-year-old male from the United Kingdom with a family history of HCM who had been diagnosed with the disease at age 51 years. He showed an apical distribution of LV hypertrophy (MLVWT 13 mm) with repolarization abnormalities on his ECG. An implantable cardio-defibrillator (ICD) was implanted at age 53 years for primary prevention (his major risk factors were non-sustained ventricular tachycardia (NSVT) on Holter monitoring and a family history of SCD). His 18-year-old son was also clinically affected, but he did not consent to genetic testing. Two of the proband's sisters and one brother had a diagnosis of HCM; the brother had died suddenly in his mid-forties and was found to have HCM at autopsy. He was an obligate carrier because his affected son was genotyped and harbored the same deletion as the family proband.

An 11kb (11,259-bp) deletion encompassing the nucleotide sequence from c.1,836-1,527 of intron 14 to c.2,022-2,042 of intron 16 of *FHOD3* was identified (Fig. 1B). The proximal breakpoint is embedded in a 1,374-bp LINE termed L1ME3C, and the distal junction–flanking sequence contains no interspersed repetitive elements within >1Kb surrounding the breakpoint. No homology or inserted sequence was observed at the breakpoint junction, suggesting again an NHEJ mechanism.

FAMILY C (Fig. 2C): The proband of this Russian family was a 58-year-old male, diagnosed in his late thirties with an asymmetrical non-obstructive septal hypertrophy (MLVWT 31 mm). His son was considered to be probably affected as, at age 16, he had an MLVWT of 12 mm and hypertrabeculation of the LV apex, lateral wall, and septum.

Both father and son harbored the same deletion; one clinically unaffected daughter was a non-carrier.

A deletion of 11kb (10,503-bp) encompassing the nucleotide sequence from c.1,836-2,543 of intron 14 to c.2,021+2,217 of intron 16 of *FHOD3* was identified (Fig. 1C). The proximal breakpoint is embedded in a 1,374bp LINE termed L1ME3C; the distal junction–flanking sequence contains no interspersed repetitive elements within >1Kb surrounding the breakpoint and also had a 105-bp insertion at the junction derived from an L1MEf L1 family LINE within the human genome. L1MEf retrotransposition was unlikely. The complexity of the CNV also precluded an NHEJ, making a replication-based mechanism the most plausible explanation for disease expression. To overcome the replication fork error, the replication machinery would have restarted DNA synthesis via a different replication fork located in L1ME3C before switching back to the original replication fork resulting in the deletion of two exons (from 15 to 16) and partial insertion of L1MEf. We observed regions of microhomology at the junctions, which allows this switching to occur.

DISCUSSION:

FHOD3 is a new HCM associated gene that may explain around 2% of cases. In this article, we report the first three families with CNVs in *FHOD3* associated with an HCM phenotype, and propose new disease mechanisms for the observed genetic variants.

The three deletions described affected symmetrical exons, predicting the loss of 45 and 62 amino acids of a relevant domain of the gene (Fig. 3). One of the limitations of our study is that RNA sequencing of this splice-variants was not performed to confirm this mechanism at the transcript level. In any case, since they involve symmetric exons 15 and 16 with no change in the reading frame of the protein sequence, they should not be considered loss-of-function variants *per se*.

The clinical characteristics of the families described in our study are consistent with those reported in the largest study of *FHOD3*, in which pathogenic variants were non-synonymous or small in-frame deletions (loss of one single amino acid). On the other hand, the pathogenicity of loss-of-function variants in *FHOD3* is still unclear. They are reported with low frequencies in public databases of controls such as gnomAD but are constrained with an observed/expected ratio of 0.25. Truncating nonsense and frameshift variants were not enriched in cohorts of cardiomyopathy patients[1] and we were unable to prove cosegregation with the disease in our previous study[2]. More information is needed to determine the clinical significance of these variants.

Only a few studies in the literature have systematically evaluated CNVs in HCM genetic studies. A few deletions in sarcomeric genes (*MYBPC3, TNNC1, ACTC1, MYL2*, and *MYL3*) that might explain the phenotype have been reported[6, 7]. We have no evidence of similar deletions described before in *FHOD3*. A few CNVs are reported in DECIPHER and ClinVar, but deletions were much larger involving the whole *FHOD3* gene and other adjacent genes and were related to more complex phenotypes.

In our cases, the three deletions affect a part of the diaphanous inhibitory domain -DID- of FHOD3 and more specifically the coil-coiled domain involving amino acids 623-655. This supercoiled alpha-helix is considered a clustering site of mutations associated with HCM. An attractive hypothesis is that the loss of the residues encompassed in our deletions could lead to a predominance of FHOD3 protein in an activated state, as it was described for mutations in the paralogue FHOD1[8] . A similar mechanism could occur with the splicing variants previously described by Semsarian et al.[4], both affecting the same essential splice donor nucleotide of exon 12 and predicting the skipping of this symmetrical exon. They would lead to an in-frame loose of 120 amino acids, from p.Ser429 to p.Ser549, also considered part of the DID domain of FHOD3 and where the

pathogenic variants p.Ser527del and p.Tyr528Cys were reported (Fig. 3). These two variants together represent 0.5% of HCM cases in our cohort (19 and 15 probands respectively, from 6788 consecutive HCM patients sequenced by NGS with panels that included the gene), reinforcing the relevance of this exon.

Using current ACMG criteria, splice-site variants that predict the skipping of exon 12 and 15-16, or deletions involving exclusively these exons should be considered likely pathogenic. We cannot extrapolate this information to larger deletions, or deletions of other exons, especially when they predict a change in the reading frame: these variants should be considered of uncertain significance and more information obtained (i.e. evidence of segregation in families or functional studies) for correct categorization.

In conclusion, our study highlights the importance of performing CNV analysis systematically in NGS genetic testing panels for HCM, and reinforce the relevance of the *FHOD3* gene in the disease.

REFERENCES

1. Walsh R, Mazzarotto F, Whiffin N, et al. Quantitative approaches to variant classification increase the yield and precision of genetic testing in Mendelian diseases: the case of hypertrophic cardiomyopathy. *Genome Med*. 2019;11(1):5. doi:10.1186/s13073-019-0616-z
2. Ochoa JP, Sabater-Molina M, García-Pinilla JM, et al. Formin Homology 2 Domain Containing 3 (FHOD3) Is a Genetic Basis for Hypertrophic Cardiomyopathy. *J Am Coll Cardiol*. 2018;72(20):2457-2467. doi:10.1016/j.jacc.2018.10.001
3. Ushijima T, Fujimoto N, Matsuyama S, et al. The actin-organizing formin protein Fhod3 is required for postnatal development and functional maintenance of the adult heart in mice. *J Biol Chem*. 2018;293(1):148-162. doi:10.1074/jbc.M117.813931
4. Semsarian C, Ingles J, Bagnall RD. Revisiting Genome Sequencing Data in Light of Novel Disease Gene Associations. *J Am Coll Cardiol*. 2019;73(11):1365-1366. doi:10.1016/j.jacc.2018.12.060
5. de Uña D. System and method to detect structural genetic variants. *OEPM (Oficina Española de Patentes y Marcas)*. Madrid, Spain. 2019:P 201731242.
6. Mademont-Soler I, Mates J, Yotti R, et al. Additional value of screening for minor genes and copy number variants in hypertrophic cardiomyopathy. *PLoS One*. 2017;12(8):e0181465. doi:10.1371/journal.pone.0181465
7. Lopes LR, Murphy C, Syrris P, et al. Use of high-throughput targeted exome-sequencing to screen for copy number variation in hypertrophic cardiomyopathy. *Eur J Med Genet*. 2015;58(11):611-616. doi:10.1016/j.ejmg.2015.10.001
8. Schulte A, Stolp B, Schönichen A, et al. The Human Formin FHOD1 Contains a Bipartite Structure of FH3 and GTPase-Binding Domains Required for

Activation. *Structure*. 2008;16(9):1313-1323. doi:10.1016/j.str.2008.06.008

LEGENDS:

FIGURE 1: The three deletions detected in this study. In the top of the panels, a graph showing a comparative depth-of-coverage is shown: x-axis corresponds to the determined genomic region and the y-axis the depth-coverage (number of readings). The red line represents each patient coverage, which is compared to the green line, representing the median coverage of control samples sequenced in the same run. In red it can be observed a 50% decrease in the coverage of the fragments corresponding to the affected exons, indicating the heterozygous deletion in the explored regions. In the bottom, Sanger sequencing confirmation of the deletions with their precise breakpoints is shown using the Chromas lite program. Panel A also shows the deletion confirmed by amplicon-based NGS using the Integrative Genomics Viewer (IGV) graphic.

FIGURE 2: Pedigrees of the three HCM families carrying CNVs in *FHOD3* identified in the present study.

FIGURE 3: Diagram of the FHOD3 protein, its relevant domains, and their concordance with the exonic regions of the gene. In blue, all the non-truncating disease-causing variants reported to date in the literature. In the bottom of the figure, it can be observed the location of the deletions described in this paper, and the splicing variants described by Semsarian and colleagues.