Circulation: Genomic and Precision Medicine

RESEARCH LETTER

Cryptic Splice-Altering Variants in *MYBPC3* Are a Prevalent Cause of Hypertrophic Cardiomyopathy

Luis R. Lopes, MD, PhD; Pedro Barbosa, MSc; Mario Torrado, PhD; Ellie Quinn, BSc, MSc; Ana Merino, MD; Juan Pablo Ochoa, MD, PhD; Joanna Jager, MSc; Marta Futema, PhD; Maria Carmo-Fonseca, MD, PhD; Lorenzo Monserrat, MD, PhD; Petros Syrris, PhD; Perry M. Elliott, MBBS, MD

The yield of genetic testing in hypertrophic cardiomyopathy (HCM) is only 40%, even in patients with family histories of the disease. This may be caused by a high prevalence of nongenetic phenocopies or complex genetic mechanisms but may also reflect the inability of conventional diagnostic sequencing to detect intronic variants distant from the essential donor/acceptor dinucleotides with the potential to disrupt splicing (also known as cryptic splice mutations). The availability of whole genomes from the genome aggregation database (gnomAD; https://gnomad.broadinstitute.org) and novel prediction tools has improved the capacity for analyzing and interpreting deep intronic sequences.

In this study, we performed large-scale unbiased screening of intronic variants in *MYBPC3* in 1644 unrelated and consecutive patients with HCM (49.5±15.6 years old at diagnosis, 1103 [67.1%] male, 1000 white [60.8%], 156 Asian [9.5%], and 75 black [4.6%]). All patients gave written informed consent and the study was approved by the regional ethics committee (15/LO/0549). Eight hundred seventy-four probands were screened with next-generation sequencing of the whole-genomic region of 41 genes, and 770 probands were screened using whole-exome sequencing including a ≈100 bp intronic region beyond the intron-exon boundaries. Sequencing, variant calling, filtering, and annotation were as previously described.[1,2] To prioritize intronic variants with an impact on splicing, a deep-learning approach, SpliceAIv1.3 was applied (https://github.com/Illumina/SpliceAI)[3] with a very stringent threshold (≥0.9). To predict branchpoint disruptions, LabBranchoR (https://github.com/jpaggi/labranchor) was run. To evaluate the functional consequences of a novel variant, expression of mRNA was analyzed using blood samples, according to published methods.[4] The data that support the findings of this study are available from the corresponding author upon reasonable request.

Five hundred forty-six patients (33.2%) had coding (or canonical splicing) variants with minor allele frequency $<1\times10^{-4}$ in *MYH7*, *MYBPC3*, *TNNT2*, *TNNI3*, *MYL2*, *MYL3*, *TPM1*, and *ACTC1*. Four variants with a very high (≥0.90) SpliceAI score were detected: 3 located in intron 13 (c.1224-52G>A [0.99; N=18], c.1224-80G>A [0.94; N=3], and c.1224-21A>G [0.90; N=1]) and another, c.906-36G>A (0.91; N=2), in intron 9 (Figure [A]). These 4 variants were present in 24/1644 (2.2% of otherwise mutation-negative patients). All are predicted to cause a cryptic splice site and an expansion to a larger subsequent (micro)exon, in turn leading to a frameshift and stop codon, as previously shown for c.1224-52G>A, c.1224-80G>A, and c.906-36G>A with RNA assays.[5–7]

Eighteen patients had c.1224-52G>A (1.1% compared with minor allele frequency in GnomAD of 2.8×10^{-5} (odds ratio, 197.9 [95% CI, 66.9–585.6]; $P\leq0.0001$). The variants c.1224-80G>A, c.906-36G>A, and c.1224-21A>G were not found in GnomAD. Cosegregation for c.1224-52G>A was demonstrated in 3 families (Figure [B]).

Of the 24 index patients carrying these 4 variants, age of diagnosis was 10 to 72 years, 16 (67%) were male and 6 were Asian (25% versus 9.5% in the overall cohort; $P=0.009$); the remainder were white. Nineteen (79%) had a family history of sudden death and/or HCM. Maximal wall thickness varied between 13 and 30 mm and 6 (25%) had left ventricular outflow tract obstruction. Seven were deemed to be at high sudden death risk

Key Words: cardiomyopathy, hypertrophic ■ cryptic splice sites ■ genetics ■ introns ■ *MYBPC3*

Correspondence to: Luis R. Lopes, MD, PhD, Barts Heart Centre, St. Bartholomew's Hospital, W Smithfield, London EC1A 7BE, United Kingdom. Email luis.lopes1@nhs.net

For Sources of Funding and Disclosures, see page 167.

Circulation: Genomic and Precision Medicine is available at www.ahajournals.org/journal/circgen

Nonstandard Abbreviations and Acronyms

HCM	hypertrophic cardiomyopathy

or already had an implantable cardioverter-defibrillator at baseline.

Finally, a novel variant, c.1898-23A>G with a low SpliceAI score (0.04, GnomAD minor allele frequency 0.000005) segregated with the phenotype in 2 "mutation-negative" families (Figure [C]). These families were enrolled and studied with whole-exome sequencing with the initial aim of novel gene discovery, and this variant was the only suitable candidate found. We have then searched for this variant in the first cohort of 874 patients and found it in an additional proband. As shown in Figure [D], the *MYBPC3* RT-PCR amplification of total RNA showed an additional longer 473 bp band detected in the probands' RNA. Sanger sequencing revealed that these longer bands contained the complete sequence of intron 19. This mis-spliced transcript introduces a premature termination codon in intron 19 and is expected to cause nonsense-mediated mRNA decay, leading to haploinsufficiency. Given the low SpliceAI score for this variant, we sought to understand the mechanism that could explain mis-splicing; we have found a high probability of branchpoint disruption (probability of position being a branchpoint reduced by 60%).

Two recent studies of small cohorts of mutation-negative cases (46[5] and 93[6] probands) described 8 novel cryptic splice-altering variants in *MYBPC3*, with a prevalence of 9% and 6.5%, respectively. In our study, the prevalence was only 2.2% in otherwise mutation-negative patients. The difference may be explained by the very stringent criteria employed and the fact that whole-exome sequencing coverage is limited to ≈100 bp from the exon-intron boundary, meaning that deeper intronic variants might have been missed. None of our candidate variants were described in a previous analysis of *MYBPC3* splice variants at noncanonical splice sites, which used both a splice prediction tool and a mini-gene assay to identify variants that alter *MYBPC3* splicing.[8]

If the variant calling in our cohort was limited to conventional splice-sites, all the families with cryptic splice-altering variants would have been considered mutation-negative. One variant in particular (c.1224-52G>A) was unusually common (1.1%). The cause for the high frequency is uncertain but seems unlikely to be a recent founder effect, considering the heterogeneous geographic origin of the patients. As a comparison, the causal HCM variant considered as the most common to date (*MYBPC3* p.Arg502Trp) has an estimated prevalence of 1.4% to 2.0% (95% CI).[9] Patients with these particular cryptic splicing intronic variants in *MYBPC3*

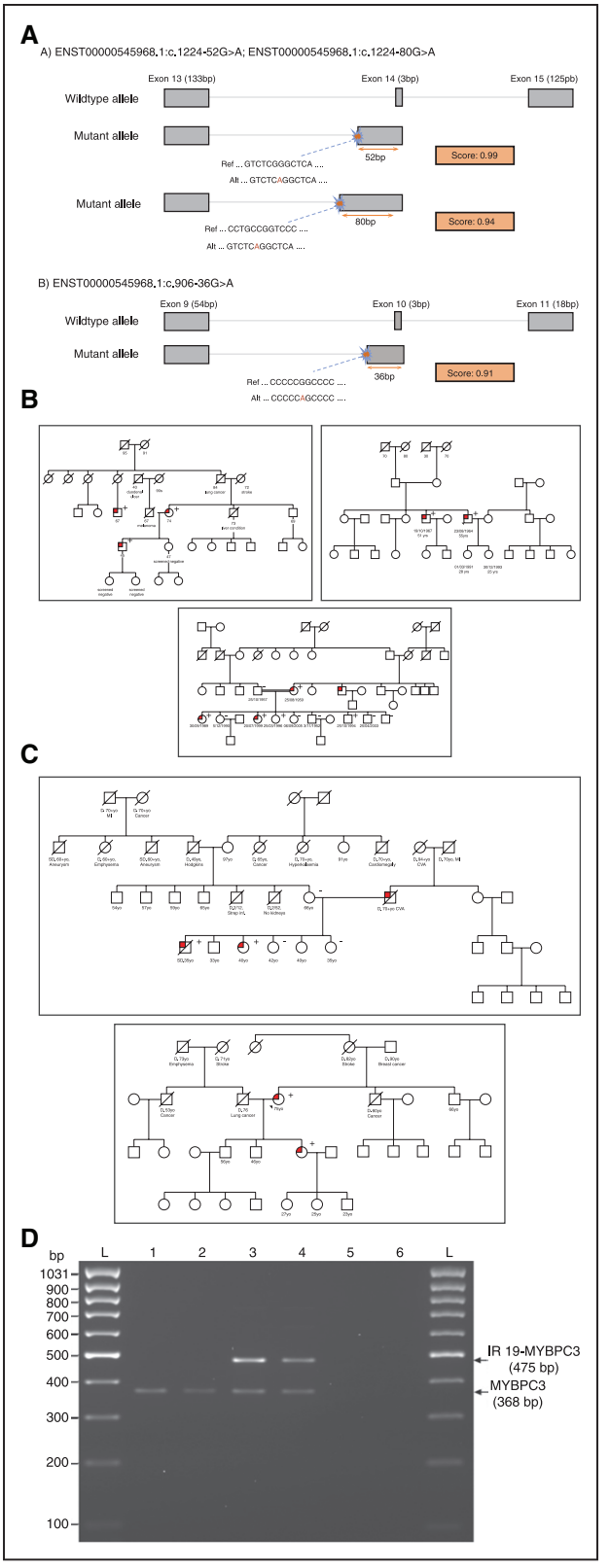

Figure. Bioinformatic analysis, co-segregation studies and RNA assays of cryptic splice-altering variants in *MYBPC3*. **A**, *MYBPC3* deep intronic variants c.1224-52G>A and c.1224-80G>A generate cryptic splice sites within intron 13 with expansion of exon 14. c.906-36G>A generates a cryptic splice site within intron 9 with expansion of exon 10. Score refers to spliceAI score. (*Continued*)

Figure Continued. B, Pedigrees demonstrating co-segregation of the *MYBPC3* c.1224-52G>A variant with HCM phenotype in 3 families. Circles: women; squares: men. +: mutation-positive; —: mutation-negative; colored symbols: affected (hypertrophic cardiomyopathy [HCM]). **C**, Pedigrees demonstrating co-segregation of the *MYBPC3* c.1898-23A>G variant with HCM phenotype in 2 previously "mutation-negative" families. Circles: women; squares: men. +, mutation-positive; —, mutation-negative; colored symbols: affected (HCM). **D**, RT-PCR analysis of the ectopic expression of *MYBPC3* in the blood of controls (lanes 1–2) and index patients carrying the c.1898-23A>G variant (lane 3 and lane 4). Negative controls were loaded in lane 5 (-RT) and 6 (nontemplate). L-100 bp DNA ladder. The lower bands (368 bp, *MYBPC3*) correspond to the normal mRNA, and the longer bands (473 bp, IR19-*MYBPC3*) correspond to a mis-spliced intron 19-retained transcript.

showed a higher prevalence of Asian ethnicity; this finding needs to be replicated in other populations.

Sequencing of deep intronic regions of *MYBPC3* increases the yield of genetic testing and thus improves counseling and evaluation of families with HCM. New phenotype-modifying genetic therapies tailored for splicing altering variants are being tested in animal models; as such, this increase in yield might in the future be translated into tailored therapy.

ARTICLE INFORMATION

Affiliations

Inherited Cardiovascular Disease Unit, St Bartholomew's Hospital, Barts Health NHS Trust (L.R.L., E.Q., P.M.E.). UCL Centre for Heart Muscle Disease, Institute of Cardiovascular Science, University College London, United Kingdom (L.R.L., J.J., M.F., P.S., P.M.E.). European Reference Network for Rare and Low Prevalence Complex Diseases of the Heart (ERN GUARD-HEART; http://guardheart.ern-net.eu/) (L.R.L., P.M.E.). Instituto de Medicina Molecular João Lobo Antunes, Faculdade de Medicina (P.B., M.C.-F.) and LASIGE, Faculdade de Ciências (P.B.), Universidade de Lisboa, Portugal. Institute of Health Sciences, University of A Coruña (M.T.). Cardiology Department, University Hospital of Burgos (A.M.). Cardiology Department, Health in Code, A Coruña, Spain (J.P.O., L.M.)

Sources of Funding

Dr Lopes is a recipient of a Medical Research Council (MRC) Clinical Academic Research Partnership (CARP) award. P. Barbosa is supported by an FCT (Fundacão para a Ciência e Tecnologia, Portugal) Fellowship SFRH/BD/137062/2018. Dr Torrado is funded by a grant (GRC ED431C 2018/38) from the Autonomic Government of Galicia, Spain. J. Jager and M. Futema are funded by the Fondation Leducq Transatlantic Networks of Excellence Program grant (no. 14 CVD03). This work was funded by the British Heart Foundation Program Grant RG/15/8/31480 and the National Institute for Health Research University College London Hospitals Biomedical Research Centre.

Disclosures

Dr Monserrat is shareholder in Health in Code. The other authors report no conflicts.

REFERENCES

1. Lopes LR, Syrris P, Guttmann OP, O'Mahony C, Tang HC, Dalageorgou C, Jenkins S, Hubank M, Monserrat L, McKenna WJ, et al. Novel genotype-phenotype associations demonstrated by high-throughput sequencing in patients with hypertrophic cardiomyopathy. *Heart*. 2015;101:294–301. doi: 10.1136/heartjnl-2014-306387
2. Lopes LR, Futema M, Akhtar MM, Lorenzini M, Pittman A, Syrris P, Elliott PM. Prevalence of TTR variants detected by whole-exome sequencing in hypertrophic cardiomyopathy. *Amyloid*. 2019;26:243–247. doi: 10.1080/13506129.2019.1665996
3. Jaganathan K, Kyriazopoulou Panagiotopoulou S, McRae JF, Darbandi SF, Knowles D, Li YI, Kosmicki JA, Arbelaez J, Cui W, Schwartz GB, et al. Predicting splicing from primary sequence with deep learning. *Cell*. 2019;176:535–548.e24. doi: 10.1016/j.cell.2018.12.015
4. Singer ES, Ingles J, Semsarian C, Bagnall RD. Key value of RNA analysis of MYBPC3 splice-site variants in hypertrophic cardiomyopathy. *Circ Genom Precis Med*. 2019;12:e002368. doi: 10.1161/CIRCGEN.118.002368
5. Bagnall RD, Ingles J, Dinger ME, Cowley MJ, Ross SB, Minoche AE, Lal S, Turner C, Colley A, Rajagopalan S, et al. Whole genome sequencing improves outcomes of genetic testing in patients with hypertrophic cardiomyopathy. *J Am Coll Cardiol*. 2018;72:419–429. doi: 10.1016/j.jacc.2018.04.078
6. Janin A, Chanavat V, Rollat-Farnier PA, Bardel C, Nguyen K, Chevalier P, Eicher JC, Faivre L, Piard J, Albert E, et al. Whole MYBPC3 NGS sequencing as a molecular strategy to improve the efficiency of molecular diagnosis of patients with hypertrophic cardiomyopathy. *Hum Mutat*. 2020;41:465–475. doi: 10.1002/humu.23944
7. Frank-Hansen R, Page SP, Syrris P, McKenna WJ, Christiansen M, Andersen PS. Micro-exons of the cardiac myosin binding protein C gene: flanking introns contain a disproportionately large number of hypertrophic cardiomyopathy mutations. *Eur J Hum Genet*. 2008;16:1062–1069. doi: 10.1038/ejhg.2008.52
8. Ito K, Patel PN, Gorham JM, McDonough B, DePalma SR, Adler EE, Lam L, MacRae CA, Mohiuddin SM, Fatkin D, et al. Identification of pathogenic gene mutations in LMNA and MYBPC3 that alter RNA splicing. *Proc Natl Acad Sci U S A*. 2017;114:7689–7694. doi: 10.1073/pnas.1707741114
9. Whiffin N, Roberts AM, Minikel E, Zappala Z, Walsh R, O'Donnell-Luria AH, Karczewski KJ, Harrison SM, Thomson KL, Sage H, et al. Using high-resolution variant frequencies empowers clinical genome interpretation and enables investigation of genetic architecture. *Am J Hum Genet*. 2019;104:187–190. doi: 10.1016/j.ajhg.2018.11.012