Peer-Review Record

Reassessment of genes associated with dilated and hypertrophic cardiomyopathy in a Chinese Han population

J Cardiovasc Aging 2023;3:12. http://dx.doi.org/10.20517/jca.2022.44

by Yang Sun, Hong Wang, Man Huang, Ke Li, Xiuli Song, Lei Xiao, Jiaqi Dai, Linlin Wang, Yanghui Chen, Dongyang Wu, Ting Yu, Rui Li, Fei Ma, Zongzhe Li, Peng Chen, Hong Wang, Yan Wang, Zhongsheng Sun, Li Jin, Dao Wu Wang, Guangzhi Chen, Dao Wen Wang

Received: 24 Dec 2022 **| First Decision:** 16 Jan 2023 **| Revised:** 30 Jan 2023 **| Accepted:** 31 Jan 2023 **| Published:** 10 Feb 2023

Academic Editor: Ali J. Marian

Reviewer 1: Anonymous

Reviewer 2: Anonymous

Round 1

Reviewer 1 Report

The authors report whole exome sequencing (WES) findings in 1,059 patients with DCM, 1,175 patients with HCM, and compare the findings to those in 514 control individuals. They have analyzed the variants in ~ 500 genes. They report that the TTN and MYBPC3 contained the most burdensome variants in DCM and HCM groups. They identify 35 genes associated with each of the DCM or HCM phenotype and show association of the variants in MUC16, KMT2C, and FBN1 with DCM and those in KMT2C, RYR2, and SCN5A with HCM. The KTM2C gene variants were identified in 7.2% and 7.4% of DCM and HCM cases. The authors provide the genetic landscape of Chinese patients with DCM and HCM and provide a website (www.cardioexome.cn) to access the data.

The strength of the data is the large sample size of the study. However, the main weakness of the data is an inadequate data analysis. Consequently, no firm conclusions could be made. The authors are encouraged to re-analyze their data to exclude spurious results and expand the findings that are authentic. As presented the risk of spurious findings is too high and the conclusions could not be supported. Examples are discussed below.

Whole exome sequencing (next generation sequencing) has a considerable error rate, depending on the instrument it could be as high as 1%. It is unclear how the sequencing errors were identified and corrected and how the variants were validated.

The authors should expand on the analysis of the TTN variants. They have provided the data inadequately. Considering that the not all TTN truncating variants are pathogenic, the authors should try to further stratify the truncating variants and assess their populations frequencies among the three study groups. This is crucial for the TTN gene. The authors might wish to portrait the distributions of the TTN truncating variants across the TTN gene/protein in each study population to visually illustrates the data (as in gnomAD).

Those genes that have a statistically significant distributions of the truncating and missense variants should be identified in Figure 1, panels A and B.

It is also unclear how the missense variants were annotated for pathogenicity. It appears that all missense variants were included in the data presented in Figure 1B.

The regression data in Figure 2, as presented are insufficient to assess enrichment of the variants in each gene in the cases. A p value-based analysis is inadequate and is expected to lead to false association. A more robust statistical approach is needed, which would be expected to change the overall findings. The reviewer suspect that the apparent associations of KMT2C variants with HCM and DCM are spurious.

Figure 3, as presented, is uninformative. It is unclear what the X axis represent.

Diagnoses of DCM and HCM: It is unclear to this reviewer whether the diagnoses of DCM and HCM adheres those described in references 5, 6 and 7. Would one be able to differentiate the athlete heart with a dilated left ventricle and an LV ejection of 48% from a true DCM, based on the criteria described in the paper? Ditto for HCM, did you not include LVEF as a diagnostic criterion?

Please provide the data range or 95% CI of the data shown in table 1. This will help the reader to assess the robustness of the clinical diagnosis.

It is also unusual that only 26.7% and 28.9% of the patients with DCM and HCM, respectively were male (as opposed to the anticipated ~ 50%). What would be the reason for this? Some sort of ascertainment bias?

Please edit the manuscript for clarity and accuracy of scientific presentation. Please kindly note that casual errors reduce the confidence on robustness of your data. Examples:

Page 5, line 12. Please define “DP”.

Page 5, lines 12 and 13. The sentence is unclear. Please revise.

Page 5, line 14. What is ANNVAR? Is it a typo? Do you mean ANNOVAR ?

ExAC dataset is now include in the gnomAD and is not a separate database any longer.

The Y axis label in Figure 1, panels A and B states “proportion”. The denominators differ among the three groups, making the figure is unclear. Is it the percentage?

Figure S1. Please define the abbreviations. What is “OWN”? Do you mean the present study? The figure legend states “population stratification of the study cohort”, which is fine but then the labels lists AFR, AMR, and EUR. These populations are not part of the study cohort. Clarity is needed.

Author Response

1. Whole exome sequencing (next generation sequencing) has a considerable error rate, depending on the instrument it could be as high as 1%. It is unclear how the sequencing errors were identified and corrected and how the variants were validated.

Thank you very much for your suggestion. To avoid the error rate as much as possible, we performed our whole exome sequencing in a deep coverage, with the mean read depth of >100×. Then our quality control were conducted according to the best practice of GATK. To further exclude the false positive variants, variants with read depth of < 20 were defined as missing and then removed when the missing rate was greater than 20% across the whole cohort. We have described the quality control in Methods section. (See page 5, lines 3-13).

The whole exome sequencing (WES) data were processed in accordance with the best practice of The Genome Analysis Toolkit.8 Adapters and low-quality reads were trimmed by Trimmomatic9, and the filtered reads were aligned to the human reference genome GRCH37 (hg19) using the Burrows-Wheeler Alignment Tool. After marking duplicates, sorting bam and recalibration, HaplotypeCaller called variants of each sample, followed by consolidation across multiple samples with the GenomicsDBImport and GenotypeGVCFs. Finally, the variant call set was recalibrated by the VariantRecalibrator and ApplyRecalibration. (Supplementary materials)
Bioinformatics
Variants with read depth of < 20 were defined as missing and then removed when the missing rate was greater than 20% across the whole cohort.

2. The authors should expand on the analysis of the TTN variants. They have provided the data inadequately. Considering that the not all TTN truncating variants are pathogenic, the authors should try to further stratify the truncating variants and assess their populations frequencies among the three study groups. This is crucial for the TTN gene. The authors might wish to portrait the distributions of the TTN truncating variants across the TTN gene/protein in each study population to visually illustrates the data (as in gnomAD).

Thank you very much for your suggestion. To provide the data adequately, we further stratified the TTN truncating variants and assessed their population frequencies among the three study groups. (see page 7, lines 19-21 and page 8, lines 1-4)

As the TTNtv variants involved exons that were spliced in >90% in the transcripts are considered more likely to be pathogenic, we further stratified the truncating variants and assessed their populations frequencies among the three study group. In total, there are 111 variants with percentage of spliced in > 90%, accounting for 120 DCM patients, one control and zero HCM patients, where 85 variants in A band, 15 variants in I band, six variants in M band, three in Z disk and two near Z disk (Figure 2).

Moreover, to visually illustrate the data, we added a new figure (Figure 2) to portrait the distributions of the TTN truncating variants across the TTN protein in each study population.

3. Those genes that have a statistically significant distributions of the truncating and missense variants should be identified in Figure 1, panels A and B.

Thank you very much for your valuable suggestions. We have identified the genes that have a statistically significant distributions of the truncating and missense variants in Figure 1, panels A and B.

4. It is also unclear how the missense variants were annotated for pathogenicity. It appears that all missense variants were included in the data presented in Figure 1B.
The regression data in Figure 2, as presented are insufficient to assess enrichment of the variants in each gene in the cases. A p value-based analysis is inadequate and is expected to lead to false association. A more robust statistical approach is needed, which would be expected to change the overall findings. The reviewer suspect that the apparent associations of KMT2C variants with HCM and DCM are spurious. Figure 3, as presented, is uninformative. It is unclear what the X axis represent.

Thank you very much for your suggestions. Indeed, we included all rare missense variants in the data presented in Figure 1B, aiming to show a profile of rare missense across the populations. Furthermore, to sufficiently assess enrichment of the variants in each gene in the cases, we filtered for deleterious variants, defined as loss-of function variants and missense variants predicted to be deleterious by REVEL and M-CAP. And then we performed gene based association test using rare deleterious variants for comparison of DCM with Control and HCM with Control. To avoid false positive results as much as possible, we adopted Benjamini-Hochberg to adjust the P values. As you expected, after the pathogenicity filtration, *KMT2C* variants were not significantly enriched in cases and therefore, we moved Figure 3 into supplementary materials as Figure S3 and revised the corresponding text in methods, results and discussion sections. Finally, we have added the description of X axis in figure legends of the new Figure S3.

See Methods section, page 5, lines 14 - 18:
Rare variants were defined as those with a minor allele frequency of < 0.001 in East Asian databases from ExAC, The 1000 Genomes Project, and gnomAD. Deleterious variants were defined as loss-of function variants (nonsense variants, frameshift variants, and canonical splicing-disrupting variants) and missense variants predicted to be deleterious by REVEL and M-CAP

See Methods section, page 6, lines 12 - 15:
The optimal sequence kernel association test (SKAT-O) was used to assess associations of candidate genes and risks of DCM and HCM. Moreover, rare deleterious variants were filtered for further SKAT-O test. Considering the multiple comparison, we adopted Benjamini-Hochberg to adjust the p values.

See Results section, page 10, lines 9 - 19:
Considering the multiple comparison may lead to false positive results, we adjusted the p values using Benjamini-Hochberg. After adjustment, only the associations of MYH7 (adjusted p = 0.006) and MYBPC3 (adjusted p = 0.006) with HCM remained significantly.

Furthermore, rare deleterious variants defined as loss-of function and missense variants predicted to be pathogenic by REVEL and M-CAP were filtered. Then we performed another SKAT-O test based on these rare deleterious variants and identified 11 and five genes significantly enriched in DCM and HCM, respectively (Table S5 and Table S6). After adjustment for multiple comparison, the associations of TTN (adjusted p = 0.003) with DCM and MYH7 (adjusted p = 0.003) and MYBPC3 (adjusted p = 0.003) with HCM remained significantly.

See Discution section, page 12, lines 16 - 19:
However, after filtration for pathogenicity of rare variants, *KMT2C* was not significantly associated with DCM or HCM, which indicates that the single variant in *KMT2C* might be not pathogenic enough to cause DCM or HCM and more studies required to explore the contribution of *KMT2C* in cardiomyopathy.

5. Diagnoses of DCM and HCM: It is unclear to this reviewer whether the diagnoses of DCM and HCM adheres those described in references 5, 6 and 7. Would one be able to differentiate the athlete heart with a dilated left ventricle and an LV ejection of 48% from a true DCM, based on the criteria described in the paper? Ditto for HCM, did you not include LVEF as a diagnostic criterion?

We are very sorry for the unclear description of diagnoses of DCM and HCM. For the diagnoses of DCM, both dilation of ventricular and systolic dysfunction were included while for the diagnoses of HCM, LVEF was not included. The dilation of ventricular was adjusted for age and body surface area to exclude athlete heart.

6. Please provide the data range or 95% CI of the data shown in table 1. This will help the reader to assess the robustness of the clinical diagnosis.

Thank you very much for your suggestions. We have added the 95% CI of the data in table 1.

It is also unusual that only 26.7% and 28.9% of the patients with DCM and HCM, respectively were male (as opposed to the anticipated ~ 50%). What would be the reason for this? Some sort of ascertainment bias?

We are truly sorry for the mistakes. We have revised the proportion of male patients. In this study, 73.3% and 71.1% of the patients with DCM and HCM were male. We think that it might be because of the Chinese Han background, as in the study by Lei Song, *et al.,* the proportion of male HCM patients is 64.6%, similar to our study.(PMID:33361398)

7. Please edit the manuscript for clarity and accuracy of scientific presentation. Please kindly note that casual errors reduce the confidence on robustness of your data. Examples: Page 5, line 12. Please define "DP". Page 5, lines 12 and 13. The sentence is unclear. Please revise. Page 5, line 14. What is ANNVAR? Is it a typo? Do you mean ANNOVAR ? ExAC dataset is now include in the gnomAD and is not a separate database any longer. The Y axis label in Figure 1, panels A and B states "proportion". The denominators differ among the three groups, making the figure is unclear. Is it the percentage? Figure S1. Please define the abbreviations. What is "OWN"? Do you mean the present study? The figure legend states "population stratification of the study cohort", which is fine but then the labels lists AFR, AMR, and EUR. These populations are not part of the study cohort. Clarity is needed.

We are truly sorry for the mistakes and unclear description. We have revised the corresponding text.

Reviewer 2 Report

The authors report the results of a WES screening in a large Han population with DCM and HCM. The analysis was performed in 1059 DCM patients, 1175 HCM patients and 514 controls. The authors selected 500 candidate genes including definitive genes as defined by ClinGen and genes reported as potential disease associated genes. The results show significant overlap with previous population reports (mainly Caucasian), with some interesting variation, such as the lack of SCN5A variant carriers. The authors report other putative novel genes. Finally, they developed an open-access website (www.cardioexome.cn) to make the data available to the research community.

This paper is well written and adds to our knowledge of DCM genetics, specifically in a subgroup which has probably not been extensively investigated (Han).

I have only minor comments:
- I wonder if the authors can comment about the clinical relevance of the novel genes identified
- also, which strategy should be suggested to strengthen the evidence of their causal role.

Author Response

I have only minor comments: - I wonder if the authors can comment about the clinical relevance of the novel genes identified - also, which strategy should be suggested to strengthen the evidence of their causal role.

Thank you very much for your positive comments. As you suggested, we have added comment about the clinical relevance of the novel genes identified and discussed the future strategy to strengthen the evidence of their causal role. See page 13, lines 14 – 21:

After pathogenicity filtration, *FBN1* is still significantly associated with DCM and *SCN5A*, *RYR2* remains significant enrichment in HCM patients. The novel genes identified in the current study added more evidence for genetic diagnosis of DCM and HCM, especially the data deposited in our website, which may contribute to assessment of clinical pathogenicity in clinical practice. However, to confirm the causal role of these associated genes, more efforts are required, of which molecular dynamics analysis helps to filter for key variants and construction of animal model and linkage analysis are the key strategies to confirm it.