

81127_Auto_EditedC.docx

Name of Journal: *World Journal of Clinical Cases*

Manuscript NO: 81127

Manuscript Type: ORIGINAL ARTICLE

Retrospective Study

Study of pathogenic genes in a pedigree with familial dilated cardiomyopathy
pathogenic genes in a pedigree with familial dilated cardiomyopathy

Xin-Ru Zhang, Hang Ren, Fang Yao, Yang Liu, Chun-Li Song

Abstract

BACKGROUND

Dilated cardiomyopathy (DCM) is a genetically heterogeneous cardiac disorder characterized by left ventricular dilation and contractile dysfunction. The substantial genetic heterogeneity evident in patients with DCM contributes to variable disease severity and complicates overall prognosis, which can be very poor.

AIM

To identify pathogenic genes in DCM through pedigree analysis.

METHODS

Our research team identified a patient with DCM in the clinic. Through investigation, we found that the family of this patient has a typical DCM pedigree. High-throughput sequencing technology, next-generation sequencing, was used to sequence the whole exomes of seven samples in the pedigree.

RESULTS

A novel and potentially pathogenic gene mutation-ANK2p.F3067L-was discovered. The mutation was completely consistent with the clinical information for this DCM pedigree.

Sanger sequencing was used to further verify the locus of the mutation in pedigree samples. These results were consistent with those of high-throughput sequencing.

CONCLUSIONS

ANK2p.F3067L is considered a novel and potentially pathogenic gene mutation in DCM.

Key Words: Dilated cardiomyopathy; Gene mutation; Whole exomes sequencing; Sanger sequencing; ANK2p.F3067L; Potentially pathogenic gene

Zhang XR, Ren H, Yao F, Liu Y, Song CL. Study of pathogenic genes in a pedigree with familial dilated cardiomyopathy. *World J Clin Cases* 2023; In press

Core Tip: Our research team identified a typical dilated cardiomyopathy (DCM) pedigree clinically. High-throughput sequencing technology, namely second-generation sequencing, is used to sequence the entire exon group of seven samples in the pedigree. A new potential pathogenic gene mutation ANK2p.F3067L was found in DCM.

INTRODUCTION

Dilated cardiomyopathy (DCM) is a type of genetically heterogeneous cardiomyopathy^[1]. It is the primary cause of heart transplantation and the third most common cause of heart failure, malignant arrhythmia, and sudden death^[2]. DCM has obvious genetic heterogeneity, with nearly 50% of DCM cases caused by genetic factors that are dominant in pathogenesis^[3]. However, research on the pathogenic genes associated with DCM is lacking. With the development of next-generation sequencing technology, it has recently been found that DCM is related to variations in genes encoding sarcomeric, cytoskeletal, nuclear membrane, and desmosomal proteins such as TTN, MYH7, TNNT2, LMNA, and DSP^[4,5]. However, only ~40% of familial DCM patients harbor known hereditary changes in pathogenic genes^[6], while the etiology of the remaining 60% of familial DCM patients remains unclear. Therefore, the identification of pathogenic genes in DCM through pedigree analysis has become an ardent focus of current research.

⁵ We believe that our study makes a significant contribution to the literature and will be of interest to the readership of your journal because we discovered the rs764952487 (ANK2p.F3067L) locus-a novel mutation locus of pathogenic genes-in a DCM family characterized in this study. It would be useful for clinical implications of these results in medicine, especially general medicine.

Clinical presentation

The family with the DCM pedigree resides in Nongan, a county in Jilin Province. The patient's parents were cousins (both of his parents are deceased-his father having succumbed to DCM). There are five brothers and three sisters in his generation of the family pedigree. Among the eight siblings, six were DCM patients (five males and one female). Based on information from the patient, there are more than 30 living family members in his family. At the time of writing this manuscript, we have obtained 19 samples from this pedigree, including ²⁴ II-2, II-3 and his wife, II-4 and his wife, ⁶ II-6, II-7, II-8, III-1, III-2 and his wife, III-3, III-4, IV-1, IV-2, IV-3, IV-4, IV-5, and IV-6. According to

clinical tests on existing samples, the key clinical information confirmed that these patients are afflicted with DCM. The pedigree chart is presented in Figure 1, from which it may be inferred that DCM in the patient's pedigree may be autosomal dominant.

According to their chief complaints, clinical medical records, and electrocardiogram and echocardiogram results, seven patients were confirmed to have DCM in the pedigree. Among these seven patients, two (I-1, II-1) died of DCM. Four samples of surviving patients were collected (II-2, II-3, II-4, and II-7; sample II-5 was not collected). Although II-2 had no obvious clinical symptoms, echocardiography showed enlargement of the left ventricle. The proband II-3 was first diagnosed at the age of 47, presenting with chest tightness and shortness of breath when fatigued. At the age of 57, mild activity after a common viral illness lead to chest tightness, shortness of breath, cough, and phlegm. These symptoms prompted him to visit a doctor for the first time. Echocardiography revealed a left ventricular end-diastolic diameter of 67 mm and left ventricular ejection fraction (LVEF) of 23%. The results indicated that his heart was significantly enlarged and his LVEF reduced. Coronary angiography was performed, ruling out cardiac enlargement caused by coronary heart disease. The clinical diagnosis was DCM. In the past 5 years, the proband was admitted to the hospital more than 10 times due to heart failure. He was equipped with an implantable cardioverter defibrillator pacemaker at the age of 60. In addition, II-4 and II-7 were also diagnosed with DCM. They also underwent coronary angiography, again ruling out cardiac enlargement caused by coronary heart disease. The clinical characteristics of the patients with DCM in the pedigree are shown in Table 1.

METHODS AND MATERIALS

Subjects

The pedigree consisted of 26 family members (13 men and 13 women; Figure 1). They were diagnosed according to the World Health Organization 1995 diagnostic criteria (left ventricle end-diastolic diameter $> 2.7 \text{ cm/m}^2$; fractional shortening $< 25\%$).

Linkage analysis

Seven subjects were included in the linkage study. Genomic DNA was extracted from peripheral blood samples. Linkage analysis of the peripheral blood of seven family members was performed.

Exome sequencing

Five micrograms of DNA from each of the two affected male individuals was used to construct exome libraries using the Agilent SureSelect exome capture system (Agilent Technologies, Santa Clara, CA, United States). The libraries were sequenced on the Illumina Solexa GAIIX platform following the manufacturer's instructions (Illumina, San Diego, CA, United States). Raw image files were processed using the Illumina pipeline (version 1.3.4) for base calling and generating reads. Reads were aligned to the human reference genome (University of California Santa Cruz, UCSC hg19; Santa Cruz, CA, United States) using Burrows-Wheeler aligner (BWA, version 0.7.5a-r405) software^[7]. Single nucleotide polymorphisms (SNPs) and indels (insertions and deletions) were identified using the genome analysis toolkit (GATK, version 4.1.4.1)^[8]. SNPs from regions with read depth > 4 and quality scores > 20 (Q20 scores) were reserved for subsequent analysis. SNPs and indels were annotated using ANNOVAR (version 2017-07-17)^[9]. Based on the SNP database (dbSNP) and 1000 genomes annotation with PolyPhen prediction, any non-synonymous variants not assigned a 'benign' prediction were considered damaging.

Identification of pathogenic genes

Whole-exome sequencing and analysis: To determine the pathogenic mutations of the DCM pedigree, we sequenced the whole exomes of seven samples (II-3, II-4 and his wife, II-7, II-8, III-3, and IV-2) and evaluated the data quality using FastQC (version 0.11.7)^[10]. In addition, we removed adapter sequences and filtered low-quality sequences using Trimmomatic (version 0.39)^[11]. The results of the filtered exome data are shown in Table 2. Then we aligned the paired-end sequences based on the human reference genome hg38

using BWA (version 0.7.5a-r405)^[7]. The results for the comparative analysis of full exon data are presented in Table 3. To evaluate these results, we performed statistical analysis on the coverage of all exon regions. The average exon coverage for each sample was 106×–133× and approximately 95% of the regions reached 20× coverage. The read depth attained during sequencing was deemed sufficient for downstream data processing, with results listed in Table 4.

Identification of potentially pathogenic gene mutations: To identify potentially pathogenic gene mutations, we first identified sequence mutations using GATK software based on sequence alignment results. A total of 790998 sequence variation sites were obtained from seven samples. Subsequently, the sequence variations were annotated using ANNOVAR (version 2017-07-17)^[9]. These annotations refer to a number of public databases, including RefSeq gene, cytoBand, phast ConsElements46way, genomic Super Dups, gwas Catalog, avsnp147, cosmic70, esp6500siv2_all, nci60, exac03, 1000g2015aug_all, 1000g2015aug_afr, 1000g2015aug_amr, 1000g2015 aug_eur, 1000g2015aug_eas, 1000g2015aug_sas, dbSNP, SIFT, Polyphen2, LRT, Mutation Taster, Mutation Assessor, FATHMM, PROVEAN, VEST3, CADD, DANN, fath mm-MKL, Meta SVM, Meta LR, integrated fit Cons, integrated confidence, GERP++_RS, phyloP7way_vertrebat, phyloP20way_mammalian, phastCons7way_vertrebrate, phastCons20way_mammalian, SiPhy_29way_log Odds, and clinvar _20170130. Next, the sequence variations and annotation results were multilayer-filtered, based primarily on the allele frequency of the locus in the population [the locus with mutation allele frequency (MAF) ≥ 0.05 was filtered according to the MAF value of the site in sequence variation databases, such as esp6500siv2_all, ExAC ALL, ExAC_EAS, ExAC_SAS, 1000g2015aug_all, 1000g2015aug_eas, and 1000g2015aug_sas], locus annotation information, locus harmfulness prediction, clinical information, comparison of sample genotype, and comparison with the Online Mendelian Inheritance in Man database.

In addition, we identified CNVs using HMZDelFinder (version 2016)^[12] and identified variations in chromosome structure using Meerkat. Finally, based on an

investigation of the relevant literature and comprehensive analysis of the filtered mutation sites, rs764952487 (ANK2p.F3067L) emerged as a potential new DCM pathogenic gene mutation site.

RESULTS

Candidate variant analysis

⁷ The frequency of the rs764952487 mutation genotype in the population was less than 0.0001 (1/10000) in each public database. The mutation genotype is located in a conserved region of the genome. Its conservation score in the phastConsElements100way database is 499. It is predicted to be a harmful mutation by LRT, Mutation Taster, PROVEAN, MetaSVM, MetaLR, CAP, fathmm-MKL, and other software. This site is located in the 38th exon of the ankyrin-B gene (ANK2) (reference sequence NM_001148)-an arrhythmia pathogenic gene. The codon changes to c.T9199C and the amino acid changes to p.F3067L (phenylalanine mutates to leucine), characterized as non-synonymous mutations (missense mutations). Information on the mutation sites is shown in Table 5.

According to a study on the protein structure of the ANK2 gene, rs764952487 is located on the helix of ZU5-ZU5-UPA-DD tandem, the neighbor of the death domain (DD). Studies have confirmed that the ZU5-ZU5-UPA domain can form a tightly packed structural supermodel, while DD can enter freely. The formation of the ZZU supermodel does not affect the spectral binding of the anchorin. However, it does change the interface mutation of the ZZU domain and further impairs the function of ankyrin-B, which has become the target of many pathogenic mutations.

Sequencing results for the mutation sites of potential pathogenic genes

To further determine the reliability of gene mutation identification, we obtained the sequence alignment results for rs764952487 in seven samples. As shown in Figure 2, mutations were detected in samples from all three afflicted patients (II-3, II-4, and II-7), while mutations were not detected in the other four samples. In all seven samples, the

sequence coverage of the site location was greater than 100×, ensuring the accuracy of the identification of variant sites.

At the same time, we verified the locus of rs764952487 in all 19 samples, including the above seven samples, using Sanger sequencing. The Sanger sequencing results were consistent with the results for whole-exome sequencing. Moreover, the Sanger results for the rs764952487 Locus in all samples were consistent with the clinical information. This further indicated that the rs764952487 Locus may constitute a novel pathogenic gene mutation in DCM. The clinical information and mutation site genotype of the rs764952487 Locus in all samples are shown in Table 4. The Sanger sequencing results for some of the samples are shown in Figure 3. As shown in Figure 3A, the rs764952487 Locus for sample II is a heterozygous mutant. As shown in Figure 3B, the rs764952487 Locus for sample III-1 is wild-type.

Table 6 shows that among the samples collected in the pedigree so far, all DCM patients carry the ANK2p.F3067L locus. No abnormalities were found in family members who did not carry ANK2p.F3067L. Younger participants carrying ANK2p.F3067L had not yet reported clinical symptoms or received abnormal ultrasound results. In summary, ANK2p.F3067L-a novel and potentially pathogenic mutation, was discovered by genetic investigation of this DCM pedigree. Considering the influence of ANK mutations on prolonged QT intervals, all DCM patients in the pedigree were reviewed. The mutation did not cause arrhythmia in patient II-3.

DISCUSSION

Ankyrin-B is a member of the anchor protein family. As a multifunctional membrane adapter, ankyrin-B is an important ion channel protein auxiliary. It is widely expressed in human tissues and highly expressed in cardiomyocytes, playing an important role in a variety of cardiac physiological functions. Ankyrin-B plays a vital role in the localization of ion transporters, ion channel localization, calcium homeostasis regulation, and membrane stability of cardiomyocytes (Figure 4). As shown in Figure 1, Na⁺/K⁺-ATPase, Na⁺/Ca²⁺ exchanger (NCX1), and inositol 1,4,5-trisphosphate receptor in the

sarcoplasmic reticulum and transverse tubules of cardiomyocytes, maintain normal contractile function and signal transduction function in cardiomyocytes^[13]. The targeting and post-translational stability of NCX1 in cardiomyocytes depends on the expression of ankyrin-B, while elevated expression of NCX1 is directly related to myocardial contractile function^[14]. The elevated expression of NCX1 leads to an imbalance in intracellular calcium homeostasis and activates calpains. The activation of calpains partially degrades contractile proteins and inhibits myocardial systolic function, resulting in cardiac enlargement and eventually leading to DCM^[15,16]. Therefore, the expression of ankyrin-B is important for maintaining the normal contractile function of cardiomyocytes.

ANK2 is a gene coding for ankyrin-B. Its size is approximately 560 kb, consisting of 53 exons on human chromosome 4. In the present study, we found that *ANK2* is a common pathogenic gene in arrhythmia syndrome. Its genetic mutation can lead to a broad-spectrum arrhythmia phenotype called "anchor protein-B syndrome".

Although arrhythmia syndrome and hereditary cardiomyopathy are considered to be different genetic diseases, there is a significant phenotypic overlap between them. Clinically, cardiomyopathy is often associated with arrhythmias and abnormal cardiac conduction. Conversely, gene mutations associated with arrhythmia syndrome are also linked to morphological and structural abnormalities in certain types of cardiomyopathy^[17-19]. It has been confirmed that common arrhythmia genes such as *PLN*, *SCN5A*, *KCNQ1*, *KCNH2*, and *KCNE2* are associated with hereditary cardiomyopathy^[20]. Recent studies have found that there are multiple co-pathogenic genes in arrhythmia syndrome and DCM (such as *ABCC9* and *SCN5A*)^[18,21]. A number of researchers have found that variations in the *SCN5A* gene can cause calcium homeostasis imbalance, resulting in myocardial injury and cardiac enlargement. This eventually leads to DCM^[19-23]. *In vitro* studies have shown that most *ANK2* gene mutations can cause abnormal expression and distribution of ankyrin-B and its binding proteins, resulting in an imbalance of calcium homeostasis in cardiomyocytes^[24,25].

1 In recent years, an increasing number of researchers have found that *ANK2* gene mutations not only cause arrhythmia syndrome but are also related to the occurrence of

cardiomyopathy. As early as 2003, Mohler *et al*^[26] found that ANK2+/- myocardial cells exhibited spontaneous contraction capacity decline and intracellular calcium dynamic disorder. In 2015, Lopes *et al*^[20] found that mutations in ANK2 can modify the phenotypic expression of hypertrophic cardiomyopathy. In 2017, Swayne *et al*^[24] reported the case of a severe idiopathic DCM patient (sudden death at the age of 36) and a DCM patient with pulmonary hypertension. Both of these patients carried the ANK2 gene mutation. In the same year, Forleo *et al*^[27] also found ANK2 gene variation in patients with DCM. Although there is a host of ankyrin-B isoforms in the heart, the primary isoform is canonical 220-kD ankyrin-B^[28]. Ankyrin-B-188 and ankyrin-B-212 are two ankyrin-B isoforms present in the heart. Ankyrin-B-188 is expressed in human ventricular cardiomyocytes, which regulates NCX expression, whereas ankyrin-B-212 is expressed in cardiomyocytes and skeletal muscle, is localized to the M-line, and exclusively interacts with obscurin^[29]. A cardiomyocyte-specific ankyrin-B-knockout mouse was designed (as ankyrin-B-null mice die shortly after birth) that developed a phenotype including dramatic structural abnormalities, biventricular dilation, reduced ejection fraction, cardiac fibrosis, premature death, and exercise-induced death^[30].

In summary, our research team discovered the rs764952487 (ANK2p.F3067L) locus-a novel mutation locus of pathogenic genes-in the DCM family characterized in this study. The codon changes to c.T9199C and the amino acid changes to p.F3067L (phenylalanine mutates to leucine). It is a non-synonymous mutation (*i.e.*, missense mutation). The frequency of the rs764952487 mutation genotype in the population of each public database is less than 0.0001 (1/10 000). The mutation genotype is located in a conserved region of the genome. Its conservation score in the phastCons-Elements100way database is 499. It is predicted to be a harmful mutation by LRT, MutationTaster, PROVEAN, MetaSVM, MetaLR, CAP, fathmm-MKL, and other software. Sanger sequencing of the rs764952487 locus in 19 pedigrees was performed by our team to validate the whole-exome sequencing results. The clinical information of this DCM family is consistent with that of the former. Among all samples collected thus far, all DCM patients carried ANK2p.F3067L, while no abnormalities were found in non-

ANK2p.F3067L carriers. Young people carrying ANK2p.F3067L did not report clinical symptoms or receive abnormal ultrasound results. This indicates that ANK2p.F3067L is likely a pathogenic gene mutation in DCM and that its influence on the onset of DCM is age-dependent. Our team will continue to monitor the pedigree, especially the young ANK2p.F3067L carriers who have not yet developed DCM. We will further verify the function of ANK2p.F3067L *in vitro* and *in vivo*, as a novel gene mutation in this pedigree.

CONCLUSION

In summary, ANK2p.F3067L is considered a novel and potentially pathogenic gene mutation in DCM.

17%

SIMILARITY INDEX

PRIMARY SOURCES

1	www.ncbi.nlm.nih.gov Internet	136 words — 4%
2	preview-bmcmedgenet.biomedcentral.com Internet	100 words — 3%
3	www.researchsquare.com Internet	23 words — 1%
4	"Clinical Cardiogenetics", Springer Science and Business Media LLC, 2020 Crossref	22 words — 1%
5	journalofethnicfoods.biomedcentral.com Internet	22 words — 1%
6	www.peterlang.com Internet	21 words — 1%
7	pr.hec.gov.pk Internet	16 words — 1%
8	Serveh Kamrava, Ali Sahimi, Justin Ichida, Muhammad Sahimi. "Machine-learning algorithm for identifying and predicting amyotrophic lateral sclerosis causal mutations", Cold Spring Harbor Laboratory, 2022 Crossref Posted Content	15 words — < 1%

9	Jason R. Cowan, Lorien Salyer, Nathan T. Wright, Daniel D. Kinnamon et al. " Gain-of-Function Variants in Dilated Cardiomyopathy ", Circulation: Genomic and Precision Medicine, 2020 Crossref	14 words — < 1 %
10	coek.info Internet	14 words — < 1 %
11	Zhen Miao, Xiaohong Chen, Honglei Zhou, Ping Liu, Shaoli Fu, Jiajie Yang, Yuhang Gao, Yupeng Ren, Dong Rong. "Interfacing MXene Flakes on a Magnetic Fiber Network as a Stretchable, Flexible, Electromagnetic Shielding Fabric", Nanomaterials Internet	13 words — < 1 %
12	insights.ovid.com Internet	13 words — < 1 %
13	journals.plos.org Internet	12 words — < 1 %
14	Rüdiger Riesch, Moritz Muschick, Dorothea Lindtke, Romain Villoutreix et al. "Transitions between phases of genomic differentiation during stick-insect speciation", Nature Ecology & Evolution, 2017 Crossref	11 words — < 1 %
15	bmcgenomics.biomedcentral.com Internet	11 words — < 1 %
16	www.frontiersin.org Internet	11 words — < 1 %
17	dm5migu4zj3pb.cloudfront.net Internet	9 words — < 1 %

18	link.springer.com Internet	9 words — < 1%
19	patentimages.storage.googleapis.com Internet	9 words — < 1%
20	www.science.org Internet	9 words — < 1%
21	Peng, Hao, Xiaojuan Xu, Lusi Zhang, Xuehong Zhang, Hexiang Peng, Yu Zheng, Sanchuan Luo, Hui Guo, Kun Xia, Jiada Li, Hongliang Yao, and Zhengmao Hu. "GLA variation p.E66Q identified as the genetic etiology of Fabry disease using exome sequencing", Gene, 2016. Crossref	8 words — < 1%
22	irdrjournal.com Internet	8 words — < 1%
23	"Abstracts of ESCVS 2003", Cardiovascular Surgery, 200311 Crossref	6 words — < 1%
24	hdl.handle.net Internet	6 words — < 1%

EXCLUDE QUOTES OFF
EXCLUDE BIBLIOGRAPHY OFF

EXCLUDE SOURCES OFF
EXCLUDE MATCHES OFF