

Research Article

# Marfan syndrome: whole-exome sequencing reveals *de novo* mutations, second gene and genotype–phenotype correlations in the Chinese population

Yuduo Wu<sup>1,2,3</sup>, Hairui Sun<sup>2,3,4</sup>, Jianbin Wang<sup>5</sup>, Xin Wang<sup>2,3,6</sup>, Ming Gong<sup>1,2,3</sup>, Lu Han<sup>2,7</sup>, Yihua He<sup>2,3,6</sup> and  Hongjia Zhang<sup>1,2,3</sup>

<sup>1</sup>Department of Cardiac Surgery, Beijing Anzhen Hospital, Capital Medical University, Beijing, China; <sup>2</sup>Key Laboratory of Medical Engineering for Cardiovascular Disease, Ministry of Education, Beijing, China; <sup>3</sup>Beijing Key Laboratory of Maternal-Fetal Medicine and Fetal Heart Disease, Beijing Anzhen Hospital, Capital Medical University, Beijing, China; <sup>4</sup>School of Biological Science and Medical Engineering, Beihang University, Beijing, China; <sup>5</sup>School of Life Science, Tsinghua University, Beijing, China; <sup>6</sup>Ultrasound Department of Beijing Anzhen Hospital, Beijing, China; <sup>7</sup>Department of Cardiac Surgery, Beijing Chaoyang Hospital, Capital Medical University, Beijing, China

**Correspondence:** Hongjia Zhang (zhanghongjia722@ccmu.edu.cn) or Yihua He (yihuahecho@163.com)



Marfan syndrome (MFS) is a dominant monogenic disease caused by mutations in fibrillin 1 (*FBN1*). Cardiovascular complications are the leading causes of mortality among MFS. In the present study, a whole-exome sequencing of MFS in the Chinese population was conducted to investigate the correlation between *FBN1* gene mutation and MFS. Forty-four low-frequency harmful loci were identified for the *FBN1* gene in HGMD database. In addition, 38 loci were identified in the same database that have not been related to MFS before. A strict filtering and screening protocol revealed two patients of the studied group have double mutations in the *FBN1* gene. The two patients harboring the double mutations expressed a prominent, highly pathological phenotype in the affected family. In addition to the *FBN1* gene, we also found that 27 patients had mutations in the *PKD1* gene, however these patients did not have kidney disease, and 16 of the 27 patients expressed aortic related complications. Genotype-phenotype analysis showed that patients with aortic complications are older in the family, aged between 20 and 40 years.

## Introduction

Marfan's syndrome (MFS) is one of the most common autosomal dominant connective tissue diseases, with an incidence of 2–3/10000, caused by mutations in the fibrillin-1 gene (*FBN1*) [1]. Since 1991, when the mutation of *FBN1* was identified as the pathogenic factor of MFS [2], about 2900 variant sites have been identified in the *FBN1* gene. More than 90% of patients diagnosed with MFS revealed mutations in the *FBN1* gene [3,4]. Clinically MFS patients with *FBN1* mutation(s), express a range of phenotypes from mild to severe disease. Thus, single *FBN1* mutation cannot fully explain the heterogeneity of clinical phenotypes in MFS patients, who can express different pathologies in the eyes, skin, musculoskeletal, cardiovascular, and pulmonary systems. Despite other complications, cardiovascular pathologies remain the major risk factor for death in MFS patients [4].

The differences in the clinical phenotypes of patients with MFS exist even within the same family harboring the same genetic mutations. Moreover, the onset of the disease at different ages may further lead to different clinical phenotypes (REF). So far, no strong genotype–phenotype correlation between *FBN1* variants and MFS has been reported except for neonatal forms of MFS associated with missense variants in exons 25–33 (exons 24–32 according to next-generation sequencing technology) [5,6]. In 2017, the study

Received: 01 October 2020  
Revised: 11 November 2020  
Accepted: 16 November 2020

Accepted Manuscript online:  
17 November 2020  
Version of Record published:  
04 December 2020

**Table 1 Variants identified by WES in 131 samples**

	SNPs	INDELS
Total	573805	98055
Frequency	209343	45239
Function	63465	4888
Exonic function	42213	3973
Deleterious	28197	3482

Sites (>2 bp) that did not affect alternative splicing were removed.

**Table 2 Double mutation sites in the *FBN1* of MFS patients**

Patients	ACMG classification	Priority	POS	Esp			Func	ExonicFunc	AAChange (NM.000138)
				Genomic	6500sv2	GnomAD			
P_7	VUS	H	48725156	-	-	-	exonic	missense SNV	exon55:c.C6646T;p.L2216F
P_7	LikelyPathogenic	H	48780367	-	-	-	exonic	frameshift deletion	exon27:c.3279delT;p.F1093fs
P_33	VUS	H	48719977	-	-	-	splicing	splicing	exon58:c.6998-7C>T
P_33	Pathogenic	H	48757762	-	-	-	splicing	splicing	exon40:c.4942+3-4A>G
P_41	VUS	H	48936888	-	-	-	exonic	missense SNV	exon2:c.G79A;p.A27T
P_41	Pathogenic	H	48738902	-	-	-	splicing	splicing	exon47:c.5788+1-2G>A
P_76	VUS	H	48780631	-	-	-	exonic	missense SNV	exon26:c.A3142G;p.I1048V
P_76	VUS	H	48802333	-	-	-	exonic	missense SNV	exon14:c.G1622A;p.C541Y
P_101	VUS	H	48780631	-	-	-	exonic	missense SNV	exon26:c.A3142G;p.I1048V
P_101	VUS	H	48802333	-	-	-	exonic	missense SNV	exon14:c.G1622A;p.C541Y
P_113	VUS	H	48936954	-	-	0.00001627	exonic	missense SNV	exon2:c.C13T;p.R5C
P_113	Pathogenic	H	48704912	-	-	-	exonic	stopgain	exon65:c.C8080T;p.R2694X
P_123	VUS	H	48936888	-	-	-	exonic	missense SNV	exon2:c.G79A;p.A27T
P_123	VUS	H	48802247	-	-	-	exonic	missense SNV	exon14:c.T1708C;p.C570R
P_136	VUS	H	48808487	-	-	-	exonic	missense SNV	exon11:c.G1220A;p.G407D
P_136	Pathogenic	H	48704912	-	-	-	exonic	stopgain	exon65:c.C8080T;p.R2694X

Priority: H: high; cytoband: The chromosome segment where the mutation sites are located (observed by Giemsa staining); POS: Absolute position of the mutation site on chromosome; Genomic, Esp6500sv2 and gnomAD: The frequency of newly discovered sites in the gnomAD, Genomic, and Esp6500sv2 databases; Func: Note the region of the mutation site; ExonicFunc: SNV or InDel variation types of exons; AAChange: amino acid change. The mutations in HGMD were shown in Red while other mutations shown in Black.

of Linnea and Franken et al. indicated that patients with severe phenotypes die earlier, and the mutation usually occurs in exons 24–32; It was also found that patients with haploinsufficiency-*FBN1* mutation and *FBN1* truncation mutations had a high risk of developing an acute aortic event or requiring surgical intervention [7]. In 2019, our center conducted a study on the relationship between genotypes and phenotypes in MFS patients with aortic dissection and revealed a significantly higher frequency of frameshift and nonsense mutations exist in these patients, whereas missense mutations were at higher frequency in patients with aortic aneurysm [8].

However, the previous studies and findings thus far don't account to some clinical phenotypic differences expressed in MFS patients. Therefore, we investigated whether there are *de novo* gene sites or additional genes contributing to the phenotypic differences among MFS patients or whether there are two mutations in MFS patients.

## Materials and methods

### Sample collection and selection

The cohort consisted of 131 MFS patients, including 82 probands and their family members. Among them, there were 69 males and 62 females. The average age of the patients at the time of enrollment was  $25.00 \pm 14.54$  (range 0.4–64), and the average age of cardiovascular involvement was  $22.64 \pm 14.67$  (range 0–64). The diagnosis of MFS patients was performed in accordance with the 2010 revised Ghent criteria [9]. Demographic and clinical data of patients were collected from medical records of the Ultrasound Department and Department of Cardiac Surgery at Beijing Anzhen Hospital, Capital Medical University, Beijing, China. Informed consent for DNA analysis was obtained from patients in line with local institutional review board requirements at the time of collection.

## Sequencing and data analysis

### DNA extract and sequencing

Genomic DNA extracted from peripheral blood of each patient was fragmented to an average size of 180–280 bp and subjected to DNA library creation using established Illumina paired-end protocols. The Agilent SureSelect Human All ExonV6 Kit (Agilent Technologies, Santa Clara, CA, U.S.A.) was used for exome capture according to the manufacturer's instructions. The Illumina Novaseq 6000 platform (Illumina Inc., San Diego, CA, U.S.A.) was utilized for genomic DNA sequencing in Novogene Bioinformatics Technology Co., Ltd (Beijing, China) to generate 150-bp paired-end reads with a minimum coverage of 10× for ~99% of the genome (mean coverage of 100×).

### Sanger sequencing

After targeted NGS sequencing, the mutation obtained was further verified in the proband and remaining affected family members by Sanger sequencing. PCR amplification was performed using the forward primer (caactcctgtgagctgttc) and reverse primer (acgttgccacagtgtcc). The obtained sequence was compared with the FBN1 reference gene (NM\_000138.4) to identify mutations.

### Data analysis

After sequencing, basecall files conversion and demultiplexing were performed with bcl2fastq software (Illumina). The resulting FASTQ data were submitted to in-house quality control software for removing low quality reads, and then were aligned to the reference human genome (hs37d5) using the Burrows-Wheeler Aligner (bwa) [9], and duplicate reads were marked using sambamba tools [10]. SNP/INDEL calling: Single nucleotide variants (SNVs) and indels were called with SAMtools to generate gVCF [11]. The raw calls of SNVs and INDELS were further filtered with the following inclusion thresholds: (1) read depth > 4; (2) Root-Mean-Square mapping quality of covering reads > 30; (3) the variant quality score > 20.

### Annotation

Annotation was performed using ANNOVAR (2017June8) [12]. Annotations included minor allele frequencies from public control data sets as well as deleteriousness and conservation scores enabling further filtering and assessment of the likely pathogenicity of variants.

### Rare variants filtering

Firstly, Filtering of rare variants was performed as follows: (1) variants with a MAF less than 0.01 in 1000 genomic data (1000g\_all) [13], esp6500siv2\_all [14], gnomAD data (gnomAD\_ALL and gnomAD\_EAS) [15] and in house Novo-Zhonghua exome database from Novogene. (2) Only SNVs occurring in exons or splice sites (splicing junction 10 bp) were further analyzed since we are interested in amino acid changes. (3) Then synonymous SNVs that were not relevant to the amino acid alternation predicted by dbSNV were discarded; The small fragment non-frameshift (<10 bp) indel in the repeat region defined by RepeatMasker were also discarded. (4) Variations were screened according to scores of SIFT [16], Polyphen [17], MutationTaster [18] and CADD [19] software. The potentially deleterious variations were reserved if the score of more than half of these four softwares support pathogenicity of variations [20]. Sites (>2 bp) that did not affect alternative splicing were removed. Secondly, the newly discovered variants with MAF less than 0.001/0.0001 in gnomAD data (gnomAD\_ALL and gnomAD\_EAS) and in-house Novo-Zhonghua exome database from Novogene.

### Classification of alterations

In order to better predict the pathogenicity of the variants, this part refers to the previous research classification of our center [21]. The American College of Medical Genetics and Genomics (ACMG) variant classification recommendations were utilized for all reported variants (Richards et al., 2015). Notably, the following two conditions of missense variants were considered strong evidences of pathogenicity: “Well-established *in vitro* or *in vivo* functional studies supportive of a deleterious effect on the gene or gene product”. (1) A missense variant that created or destroyed a cysteine residue. 2) A missense variant that affected conserved residues in the EGF-like domain consensus sequence (D/N) X (D/N) (E/Q) X<sub>m</sub> (D/N) X<sub>n</sub> (Y/F) (m and n represent variable numbers of residues). Positive result: pathogenic or likely pathogenic variant(s) in a known disease gene associated with the reported phenotype.

### Possible diagnosis

Variant(s) in a known disease gene possibly associated with the reported phenotype. This category includes novel variants, including missense variants or in-frame insertions/deletions in disease genes that overlap with the provided

**Table 3 Data filtering results of double mutation sites in eight MFS patients**

Patients	Whether_freq_0.001	Whether_freq_0.0001	Genomic	Esp6500siv2	GnomAD	ACMG_classification	Priority
P_7	TRUE	FALSE				VUS	H
P_7	TRUE	TRUE				LikelyPathogenic	H
P_33	TRUE	FALSE				VUS	H
P_33	TRUE	TRUE				Pathogenic	H
P_41	TRUE	TRUE				Pathogenic	H
P_41	FALSE	FALSE				VUS	H
P_76	TRUE	FALSE				VUS	H
P_76	TRUE	FALSE				VUS	H
P_101	TRUE	TRUE				VUS	H
P_101	TRUE	TRUE				VUS	H
P_113	TRUE	TRUE				Pathogenic	H
P_113	TRUE	FALSE				VUS	H
P_123	TRUE	TRUE				VUS	H
P_123	FALSE	FALSE				VUS	H
P_136	TRUE	TRUE				Pathogenic	H
P_136	TRUE	TRUE				VUS	H

Whether\_freq\_0.001: Whether the locus is less than 0.001 or has no record in gnomAD data (gnomAD\_ALL and gnomAD\_EAS) and in house Novo-Zhonghua exome database from Novogene; Whether\_freq\_0.0001: Whether the locus is less than 0.0001 or has no record in gnomAD data (gnomAD\_ALL and gnomAD\_EAS) and in house Novo-Zhonghua exome database from Novogene. True means yes and False means no. Genomic, Esp6500siv2 and gnomAD: The frequency of newly discovered sites in the gnomAD, Genomic, and Esp6500siv2 databases;

**Table 4 Age comparison of patients in different aortic phenotype groups**

<i>P</i> value	No aortic phenotype
Width	0.6075
Dissection	0.004868
Aneurysm	0.000214

Width: aortic widening; Aneurysm: aneurysm of aorta; Dissection: aortic dissection.

phenotype for the patients, and a single rare or highly suspicious novel variant of uncertain significance (VUS) known to be *in trans* with a pathogenic/likely pathogenic variant in a gene that explains the reported phenotype.

### Candidate gene

Variant(s) predicted to be deleterious in a gene that has not been previously implicated in MFS or for which the published data supporting an association may not yet be definitive were considered potential novel candidates. Supporting data were based on model organism data, copy number variant data, tolerance of the gene to sequence variation, data regarding tissue or developmental timing of expression, or knowledge of the gene function and pathway analysis.

### Uncertain result

VUS in a known disease gene and a patient phenotype consistent with the reported disease spectrum (e.g. uncertainty is limited to the pathogenicity of the variant due to a lack of parent samples to assess for *de novo* occurrence and determining the phase of variants in recessive disorders). This category also includes recessive conditions that overlap with the phenotype provided for the fetus in which only a single pathogenic/likely pathogenic variant is identified.

### Negative result

No variants in genes associated with the reported phenotype identified.

## Analysis of genotype–phenotype correlations

In order to avoid bias for a certain locus due to the family relationship, we only extracted the proband samples from the families of enrolled MFS patients and analyzed the relationship between the genotype and the phenotype. Since some families have no proband, if there is only one member of the family, this member will be selected as the proband; if

there is more than one member, a patient will be randomly selected as the proband. Finally, a set of 82 samples were obtained. Requirement (1) frequency: reserve variants with MAF less than 0.0001 in gnomAD data (gnomAD\_ALL and gnomAD\_EAS) and in house Novo-Zhonghua exome database from Novogene; (2) reserve variation with functional loss (classical splicing site + - 1 / 2, code shifting, nonsense). At least two software in Sift, polyph2, and mutationastar predict harmful missense variation; classic splicing site + - 3 to 10 mutations, retain two software predicted mutations affecting splicing [22]. At the same time, the relationship between the shared genes and disease phenotype was analyzed for some genes with a large number of shared samples.

## Statistical analysis

The quantitative data and variables are respectively presented as the mean  $\pm$  standard deviation (SD) and frequencies or percentages. Quantitative variables were compared using Student's *t* test. A value of  $P < 0.05$  was considered significant. Statistical analysis was performed by SPSS 22.0.

## Results

### De novo mutation loci in *FBN1* gene

All site information of WES in 131 MFS patients is shown in Table 1 and details of sequencing quality are shown in Table 1. Furthermore, 82 low-frequency deleterious loci were found in the *FBN1* gene (recorded in OMIM database and MFS related pathogenic genes, Table 2). Of the 82 loci, 44 loci have been reported (Supplemental Excel S1), and 38 loci have never been reported before to cause MFS according to the HGMD database (Supplementary Excel S2). In order to further improve the quality of the sequencing data, we set a higher data filtering level for the newly discovered 82 *FBN1* loci in gnomAD and Asia, and Nuohe's local database. The more stringent sequencing criteria yielded fewer new possibly pathogenic mutations (Table 3).

### Two-mutations in *FBN1* gene

We found eight patients (5 males and 3 females) had more than one SNP or INDEL on *FBN1* (Table 2). The age of the patients with MFS related symptoms was under 35 years old. Only P\_123 didn't present cardiovascular complications, but the patient had high myopia and dislocated lens. In order to further ensure the accuracy of two mutations sites, we set more strict filter conditions (set the data filtering standard of 0.001 / 0.0001, as shown in Table 3), and finally determined that only P\_101 (exon26:c.A3142G;p.I1048V/exon14:c.G1622A;p.C541Y) and P\_136 (exon11:c.G1220A;p.G407D/exon65:c.C8080T;p.R2694X) have two mutations in the *FBN1* gene at the same time (According to the incidence rate of 1/10000 in MFS patients, two mutations were confirmed). Both patients presented with widened aorta, and the younger brother of P\_136 was also an MFS patient, but the phenotype was not as obvious as in patient P\_136. MFS onset in the two patients was at a relatively earlier age, P\_101 disease onset at 6 years of age and P\_136 at 7 years.

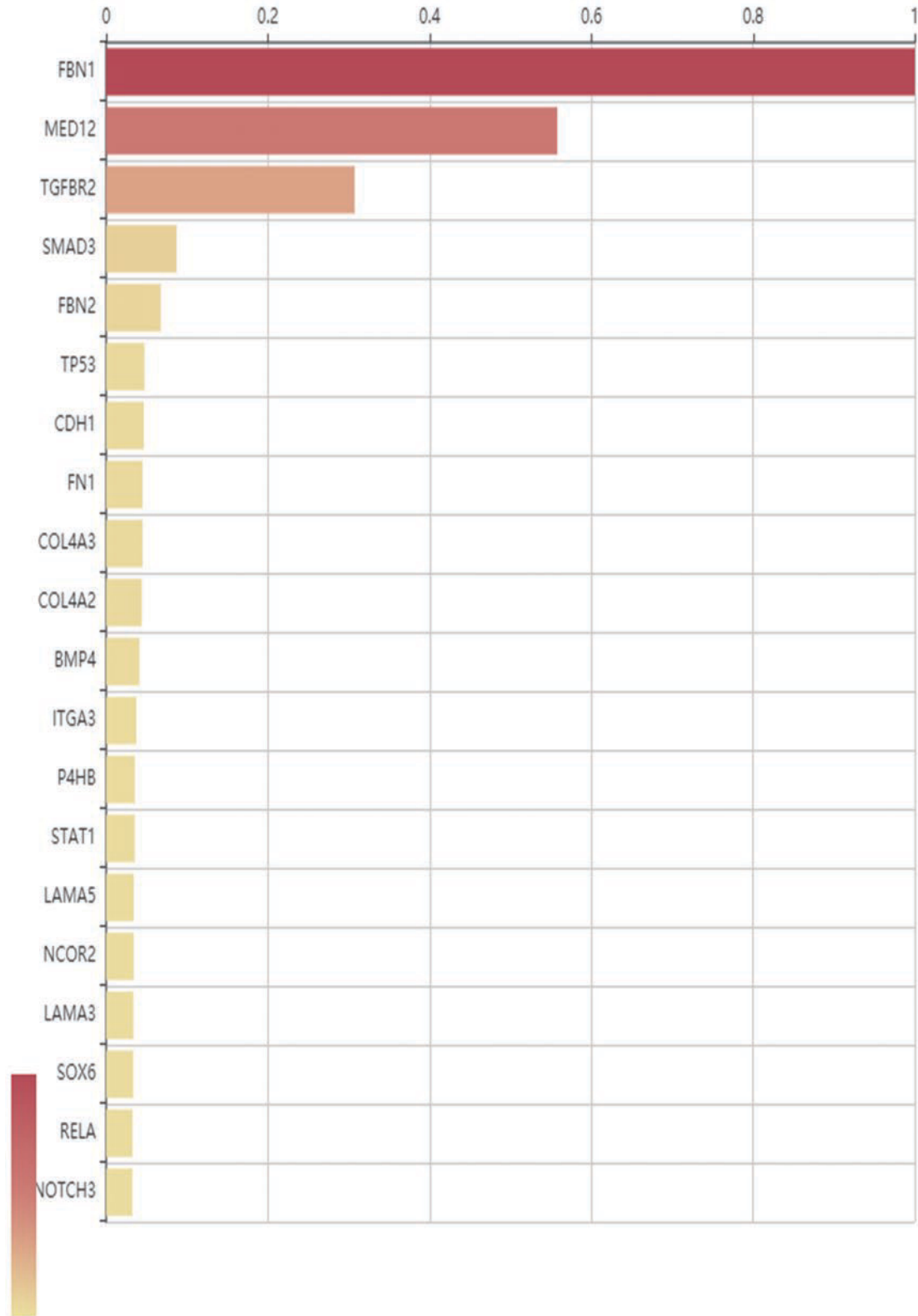
### The second gene mutation

Surprisingly the sequencing results revealed the presence of mutations in genes other than the *FBN1* gene. These genes were *PKD1*, *PKD2*, *FLNA*, *NKX2-5* and *ACVRL1*. Mutations in *PKD1* gene appeared at much higher frequency than others. Twenty-seven patients (13 males and 14 females) have mutations in *PKD1* which is located on chromosome 16 in humans. The average age of disease onset is 27.27 years in these patients. Sixteen of the patients suffered from cardiovascular complications. Additionally we found that 14 of 27 patients have missense mutations in *FBN1*. Although mutations in other genes (*MFAP5*, *SMARTC2*, and *SMAD3*) have been detected but they didn't appear to contribute to MFS according to the low pathogenicity after filtering.

### Analysis of genotype–phenotype correlations in MFS

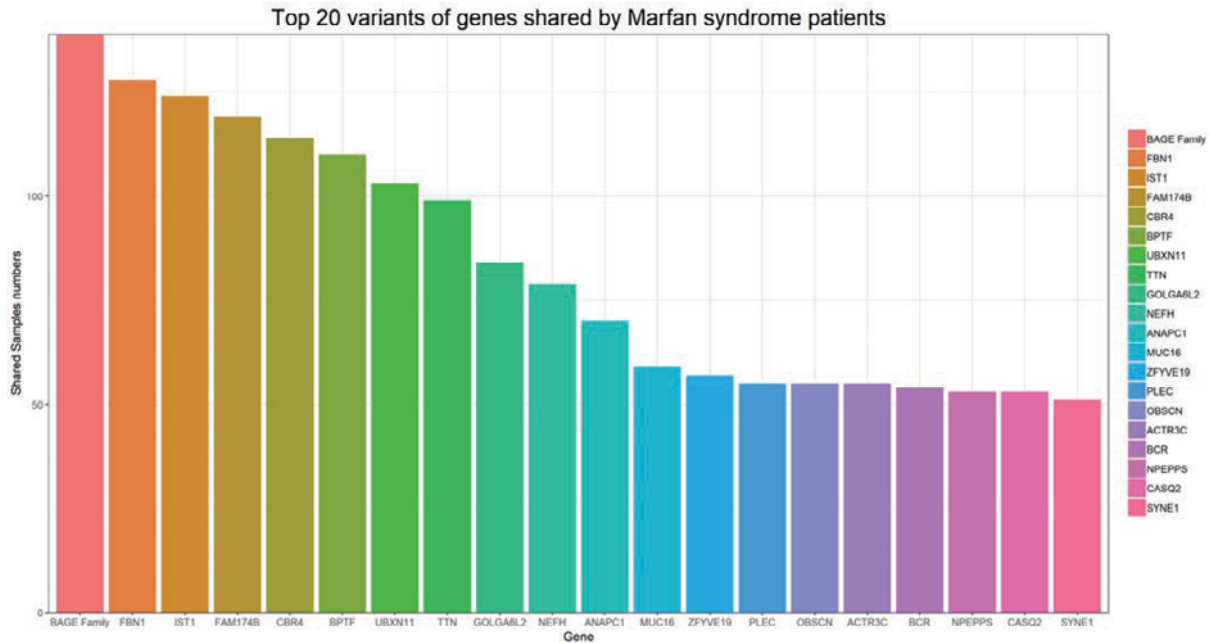
In order to construct a genotype–phenotype correlations map, the sequencing results along with various databases were combined to screen and sort the affected genes. Genes were sorted based on the strength of the correlation between genetic mutations and MFS (Figure 1). The top 10 candidate genes associated with phenotype or disease are: *FBN1*, *MED12*, *TGFBR2*, *SMAD*, *FBN2*, *TP53*, *CDH1*, *FN1*, *COL4A3*, and *COL4A2*.

Data analysis also indicated that there are many genes related to cardiomyopathy, such as *TTN*, *NEFH*, *PLEC*, *CASQ2*, and *SYNE1* (Figure 2). At the same time patients with the aortic dissection and aneurysm were further studied based on the relationship between the age at disease onset and the occurrence of aortic events (divided into four aortic phenotypes of no, width, aneurysm, and dissection). The correlation was mainly observed in patients



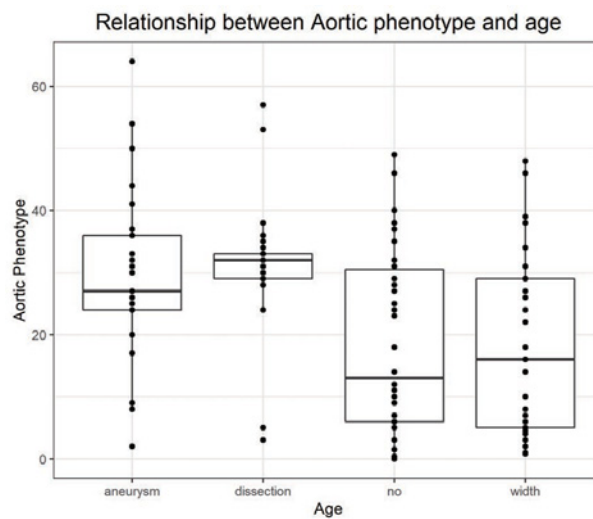
**Figure 1. Correlations between mutated genes and MFS according to sequencing results**

Only the top 20 mutated genes were shown. The maximal relevance score is 1, and the ranking is based on the built-in rules.



**Figure 2. Top 20 variants of common genes mutated in patients with MFS**

The top 10 candidate genes associated with phenotype or disease are: FBN1, MED12, TGFBR2, SMARD, FBN2, TP53, CDH1, FN1, COL4A3, COL4A2. There are many genes related to cardiomyopathy, such as TTN, NEFH, PLEC, CASQ2, SYNE1.



**Figure 3. Correlations between aortic phenotype and age**

According to the patient's aortic phenotype, the patients were divided into four groups. The correlation was mainly observed in patients between 20 and 40 years of age. Whereas the age distribution of the patients with no aortic phenotype or aortic width had a wider range with more patients younger than 20 years of age.

between 20 and 40 years of age. Whereas the age distribution of the patients with no aortic phenotype or aortic width had a wider range with more patients younger than 20 years of age (Figure 3).

Statistical analysis using *t* test revealed a significant difference between the age of patients with no aortic phenotype and those with aortic aneurysm and aortic dissection ( $P < 0.05$ ). Whereas no significant difference was found between the age of patients showing no aortic phenotype and those with only aortic widening ( $P > 0.05$ ) (Table 4).

The patients harboring two-mutations in the *FBN1* gene showed prominent MFS phenotypes compared to other family members. Moreover, the external manifestations of patients are also obvious, but not every patient has aortic complications.

## Discussion

Pathogenic mutations in the *FBN1* gene have been reported to cause MFS pathologies affecting peripheral tissues, skeletal anomalies and aortic complications [1]. Among the 131 Marfan patients enrolled in this study, 38 patients were found in HGMD database to harbor novel mutations in *FBN1* loci. The findings in the present study expand the spectrum of the *FBN1* site mutation information database. The novel loci were not detected in the early stage of genomic testing [5]. It is proved that WES technology is superior to panelgenomic testing for the discovery of new mutation sites in MFS patients, and the coverage is more extensive, but due to the current testing costs, WES technology is subject to certain restrictions [5]. However, with the advancement of science and technology, the current testing costs have dropped to affordable levels. The present study provides additional evidence that WES technology should be widely used in clinical and scientific research for more accuracy and sensitivity.

To this extent, one of our novel findings is the detection of two mutations in the *FBN1* gene that exacerbate the disease pathology in MFS patients. The mutations we report in the present study are: exon26:c.A3142G:p.I1048V and exon14:c.G1622A:p.C541Y from one patient (P\_101), exon11:c.G1220A:p.G407D and exon65:c.C8080T:p.R2694X from a second patient (P\_136). Previous reports indicate that all patients with two mutations in *FBN1* have obvious MFS phenotypes. Nonetheless the incidence is extremely low in the population, and the clinical significance is still unknown [23,24]. In our study, patients with two mutations in *FBN1* manifested external phenotypes and early disease onset, which is consistent with previous reports [6]. Although the specific mechanism through which the two mutations occur in *FBN1* in MFS patients is still not clear, the current reports pointed out that the probability for such mutation to cause severe MFS is very high [24,25,26]. This highlights the importance of evaluating patients with severe MFS, early onset and high penetrance for the presence of two mutations in the *FBN1* gene.

In this group of MFS patients, we also found 27 patients with *PKD1* gene mutations, but the mutations are located on the patient's chromosome 16. Previously, it has been reported that *PKD1* mutation is one of the main genes that cause polycystic kidney disease in patients. The pathogenic genes are all located on chromosome 15 of patients, and the disease is a late-onset disorder [27,28]. In recent years, it has been found that some patients with *PKD1* gene mutation suffers from aortic dissection [29]. Although *PKD1* gene mutation and aortic related complications were found in some patients, polycystic kidney did not appear in these patients. In addition, *PKD1* gene mutations in this group of patients are located on chromosome 16, so far, there is no report on whether the mutation of *PKD1* gene in addition to chromosome 15 will also cause polycystic kidney disease in MFS patients. Moreover, polycystic kidney disease is a late onset disease while MFS patients are all early onset, in which the appearance of other phenotypes still need further follow-up, but we did not find any renal lesions in the existing older MFS patients. Therefore, the detection of this genetic mutation in patients with MFS, whether it alone causes aortic related complications, or whether it works together with *FBN1* gene, needs further research and demonstration.

Although peripheral tissues of Marfan patients are often affected, aortic complications are the main mortality risk factors. Previous studies have shown that MFS patients with severe phenotypes are prone to mutations in the *FBN1* gene in the exon 24-32 region [7]. Early research in our center also showed that patients with *FBN1* frameshift mutations and nonsense mutations are prone to aortic dissection, and missense mutations are prone to aortic aneurysms [8]. Based on the observation that older patients in the same family seemed to have either aortic aneurysm or aortic dissection we further analyzed data considering age as a variable. Patients were grouped in four categories based on the aortic phenotype. Group-1: aortic widening; Group-2: aortic aneurysm; Group-3: aortic dissection and Group-4: No aortic phenotype (no). The correlation between the aortic phenotype and age was significant and obvious as seen in the box plot (Figure 3). This study found that patients with aortic aneurysms and dissections are between 20 and 40 years of age. This finding highlights the importance of being more vigilant for MFS patients who are in this age group to prevent the occurrence of aortic adverse events.

In the present study, only exons from MFS patient's DNA were studied, but related reports showed that MFS can also be caused by mutations in the intron region of the *FBN1* gene, and these mutations can have an impact on the prognosis and phenotype of patients. We are working to obtain a large sample size and multicenter data to analyze gene information of introns in patients with MFS in the near future in order to improve genetic information of patients with MFS. And, for other gene loci in the article, such as *PKD1*, etc., since there are no samples during data analysis, we cannot perform sanger verification for these genes. However, as shown by the data provided, these sites' sequencing



data are good (such as total depth and mutation depth etc.). We believe that these site mutations are very reliable. However, we will study this part in-depth in the future.

## Conclusion

In the present study, through the whole-exome sequencing of MFS patients, we found many new MFS mutation sites and patient double mutation sites, which further verified the pathogenicity of the *FBN1* gene to MFS patients. At the same time, it is believed that *FBN1* gene mutation, which the main factor leading to patients with aortic dissection or aneurysm, and double mutation sites are a crucial factor that aggravates the patient's phenotype. Besides, we also found that the age of MFS patients with related aortic complications in the same family is generally between 20 and 40 years, which is older than other MFS patients in the same family with no or mild aortic phenotype, which can provide a basis for clinicians to have better disease prognosis and prevent the occurrence of aortic adverse events. The clinical difference of MFS patients is vast, which may be caused by modification factors, such as the gene loci found on *PKD1* in this sequencing. We consider these loci may be modified factors, but they do not occur on the disease-causing chromosomes. Due to the lack of samples in this part of the sequencing process, we cannot verify this part of the data. However, for this research, we are only preliminary research, and the relationship between these factors needs further research and discussion in the future.

## Data Availability

Data are available upon request.

## Competing Interests

The authors declare that there are no competing interests associated with the manuscript.

## Funding

This study was supported by National Key R&D Program of China [grant number 2017YFC1308000] and Beijing Municipal Administration of Hospitals' Ascent Plan [grant number DFL20180602].

## Author Contribution

Yuduo Wu designed research and wrote the paper. Hairui Sun and Jianbin Wang performed experiments. Xin Wang, Ming Gong and Lu Han worked on associated data collection; Yihua He and Hongjia Zhang reviewed the manuscript in preparation; All authors have read and approved the final manuscript.

## Ethics Approval and Consent to Participate

This study was approved by the ethics committee of Beijing Anzhen Hospital (Institutional Review Board File 2014019, approval number: 2017YFC1308000) and consisted of the principles outlined in the Declaration of Helsinki.

## Consent for Publication

All authors have agreed to publish this manuscript

## Acknowledgements

We gratefully acknowledge all the patients and their families, and all the authors for their contributions. Meanwhile, thanks for the lab that helped us with this genetic test.

## Abbreviations

ACMG, American College of Medical Genetics and Genomics; *FBN1*, fibrillin 1; MFS, Marfan syndrome; VUS, variant of uncertain significance.

## References

- 1 Judge, D.P. and Dietz, H.C. (2005) Marfan's syndrome. *Lancet* **366**, 1965–1976, [https://doi.org/10.1016/S0140-6736\(05\)67789-6](https://doi.org/10.1016/S0140-6736(05)67789-6)
- 2 Dietz, H.C., Cutting, G.R., Pyeritz, R.E. et al. (1991) Marfan syndrome caused by a recurrent de novo missense mutation in the fibrillin gene. *Nature* **352**, 337–339
- 3 Robinson, P.N., Arteaga-Solis, E., Collod-Béround, G. et al. (2006) The molecular genetics of Marfan syndrome and related disorders. *J. Med. Genet.* **43**, 769–787

- 4 Beatens, M., Van Laer, L., De Leeneer, K. et al. (2011) Applying massive parallel sequencing to molecular diagnosis of Marfan and Loey-Dietz syndromes. *Hum. Mutat.* **32**, 1053–1062
- 5 Yuan, B., Neiar, J., Pehlivan, D. et al. (2019) Clinical exome sequencing reveals locus heterogeneity and phenotypic variability of cohesinopathies. *Genet. Med.* **21**, 663–675
- 6 Cook, J.R., Carta, L., Galatioto, J. and Ramirez, F. (2015) Cardiovascular manifestations in Marfan syndrome and related diseases; multiple genes causing similar phenotypes. *Clin. Genet.* **87**, 11–20
- 7 Franken, R., Teixido-Tura, G., Brion, M. et al. (2017) Relationship between fibrillin-1 genotype and severity of cardiovascular involvement in Marfan syndrome. *Heart* **103**, 1795–1799
- 8 Xu, S., Li, L., Fu, Y. et al. (2020) Increased frequency of FBN1 frameshift and nonsense mutations in Marfan syndrome patients with aortic dissection. *Mol. Genetics Genomic Med.* **8**, e1041
- 9 Li, H. and Durbin, R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754, <https://doi.org/10.1093/bioinformatics/btp324>
- 10 Tarasov, A., Vilella, A.J., Cuppen, E., Nijman, I.J. and Prins, P. (2015) Sambamba: fast processing of NGS alignment formats. *Bioinformatics* **31**, 2032–4, <https://doi.org/10.1093/bioinformatics/btv098>
- 11 Li, H., Handsaker, B., Wysoker, A. et al. (2009) The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079, <https://doi.org/10.1093/bioinformatics/btp352>
- 12 Wang, K., Li, M. and Hakonarson, H. (2010) ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164–e164, <https://doi.org/10.1093/nar/gkq603>
- 13 Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., 1000 Genomes Project Consortium et al. (2015) A global reference for human genetic variation. *Nature* **526**, 68–74
- 14 Exome Variant Server (2017) *NHLB GO Exome Sequencing Project (ESP)*, Seattle, WA, Available at: <http://evs.gs.washington.edu/EVS> (accessed February 21)
- 15 (2017) gnom AD. Available at: <https://doi.org/10.1101/030338> (accessed February 21)
- 16 Kumar, P., Henikoff, S. and Ng, P.C. (2009) Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat. Protoc.* **4**, 1073–1081, <https://doi.org/10.1038/nprot.2009.86>
- 17 Adzhubei, I.A. et al. (2010) A method and server for predicting damaging missense mutations. *Nat. Methods.* **7**, 248–249, <https://doi.org/10.1038/nmeth0410-248>
- 18 Schwarz, J.M., Rodelsperger, C., Schuelke, M. and Seelow, D. (2010) MutationTaster evaluates disease-causing potential of sequence alterations. *Nat. Methods.* **7**, 575–576, <https://doi.org/10.1038/nmeth0810-575>
- 19 Kircher, M. et al. (2014) A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* **46**, 310–315, <https://doi.org/10.1038/ng.2892>
- 20 Muona, M., Berkovic, S.F., Dibbens, L.M., Oliver, K.L., Maljevic, S., Bayly, M.A. et al. A recurrent de novo mutation in KCNC1 causes progressive myoclonus epilepsy. *Nat. Genet.* **47**, 39–46, <https://doi.org/10.1038/ng.3144>
- 21 Shijun, X., Li, L., Yuwei, F. et al. (2020) Increased frequency of FBN1 frameshift and nonsense mutations in Marfan syndrome patients with aortic dissection. *Mol Genet Genomic Med* **8**, e1041
- 22 Aubart, M., Gazal, S., Arnaud, P. et al. (2018) Association of modifiers and other genetic factors explain Marfan syndrome clinical variability. *Eur. J. Hum. Genet.* **26**, 1759–1772, <https://doi.org/10.1038/s41431-018-0164-9>
- 23 Renard, M., Francis, C., Ghosh, R. et al. (2018) Clinical Validity of Genes for Heritable Thoracic Aortic Aneurysm and Dissection. *J. Am. Coll. Cardiol.* **72**, 605–615, <https://doi.org/10.1016/j.jacc.2018.04.089>
- 24 Arnaud, P., Hanna, N., Aubart, M. et al. (2017) Homozygous and compound heterozygous mutations in the FBN1 gene: unexpected findings in molecular diagnosis of Marfan syndrome. *J. Med. Genet* **54**, 100–103, <https://doi.org/10.1136/jmedgenet-2016-103996>
- 25 Van Dijk, F.S., Hamel, B.C., Hillhorst-Hofstee, Y. et al. (2009) Compound-heterozygous Marfan syndrome. *Eur. J. Med. Genet* **52**, 1–5, <https://doi.org/10.1016/j.ejmg.2008.11.004>
- 26 Li, J., Lu, C., We, W. et al. (2019) Application of next-generation sequencing to screen for pathogenic mutations in 123 unrelated Chinese patients with Marfan syndrome or a related disease. *Science China. Life Sci.* **62**, 1630–1637, <https://doi.org/10.1007/s11427-018-9491-8>
- 27 Eisenberger, T., Decker, C., Hiersche, M. et al. (2015) An efficient and comprehensive strategy for genetic diagnostics of polycystic kidney disease. *PLoS ONE* **10**, e0116680, <https://doi.org/10.1371/journal.pone.0116680>
- 28 Pei, Y. (2011) Practical genetics for autosomal dominant polycystic kidney disease. *Nephron Clin. Pract.* **118**, c19–c30, <https://doi.org/10.1159/000320887>
- 29 Garcia-Gonzalez, M.A., Jones, J.G., Allen, S.K. et al. (2007) Evaluating the clinical utility of a molecular genetic test for polycystic kidney disease. *Mol. Genet. Metab.* **92**, 160–167, <https://doi.org/10.1016/j.ymgme.2007.05.004>