

Difficulties in diagnosing Marfan syndrome using current *FBN1* databases

Kristian A. Groth, MD^{1,2}, Mette Gaustadnes, MSc, PhD², Kasper Thorsen, PhD, MSc², John R. Østergaard, MD³, Uffe Birk Jensen, MD⁴, Claus H. Gravholt, PhD, DMSc^{2,5} and Niels H. Andersen, PhD, DMSc¹

Purpose: The diagnostic criteria of Marfan syndrome (MFS) highlight the importance of a *FBN1* mutation test in diagnosing MFS. As genetic sequencing becomes better, cheaper, and more accessible, the expected increase in the number of genetic tests will become evident, resulting in numerous genetic variants that need to be evaluated for disease-causing effects based on database information. The aim of this study was to evaluate genetic variants in four databases and review the relevant literature.

Methods: We assessed background data on 23 common variants registered in ESP6500 and classified as causing MFS in the Human Gene Mutation Database (HGMD). We evaluated data in four variant databases (HGMD, UMD-*FBN1*, ClinVar, and UniProt) according to the diagnostic criteria for MFS and compared the results with the classification of each variant in the four databases.

Results: None of the 23 variants was clearly associated with MFS, even though all classifications in the databases stated otherwise.

Conclusion: A genetic diagnosis of MFS cannot reliably be based on current variant databases because they contain incorrectly interpreted conclusions on variants. Variants must be evaluated by time-consuming review of the background material in the databases and by combining these data with expert knowledge on MFS. This is a major problem because we expect even more genetic test results in the near future as a result of the reduced cost and process time for next-generation sequencing.

Genet Med advance online publication 26 March 2015

Key Words: ClinVar; Exome Sequencing Project; Human Gene Mutation Database; next-generation sequencing; variant databases

Patients with classic Marfan syndrome (MFS) tend to be tall, with long limbs and fingers, as described by Marfan¹ in 1896. The syndrome's description has since evolved considerably, however, and a variety of clinical manifestations are now well established. The genetic background, that is, mutations in the fibrillin-1 gene (*FBN1*),² was discovered much later but is now a cornerstone in the diagnosis of MFS.

The definition of MFS is based on several sets of criteria, the latest being the Ghent II nosology.³ It is considered to be an autosomal-dominant disorder associated with a mutation in *FBN1* and phenotypical manifestations of MFS. The Ghent II criteria highlight the importance of a *FBN1* mutation test; for this reason, the focus on genetic testing has increased considerably when either diagnosing or excluding MFS.⁴

With the introduction of next-generation sequencing, the sequencing price and the process time have been reduced considerably, and the amount of genetic sequencing data has been multiplied, resulting in an enormous amount of data that need to be evaluated. Moreover, access to genetic testing in the clinical setting is becoming more common and widespread, and the number of patients who are genetically tested is rapidly increasing.

The vast majority of genetic variants found are benign and represents a part of our genetic variation, but some—maybe

just one in a given patient—may be pathogenic and the cause of a given disease. So even though sequencing has become easier and more accessible, the evaluation of these sequencing data has not evolved with the same speed as the technical evolution of next-generation sequencing. This is a considerable problem given that a patient must receive the correct genetic diagnosis.

The majority of variants are single-nucleotide variants. In reality, the genotype–phenotype correlation is essential to determining the pathogenicity of *FBN1* in MFS.

A range of tools for evaluating variants has been developed, but a majority of these tools are not exact and can be used only for guidance when evaluating genetic variants. In addition, a number of databases exist with data collected for published and nonpublished variants. The databases are often incorrect and have many incorrect interpretations of published data.^{5,6}

Yang et al.⁶ recently presented an evaluation of common variants in the National Heart, Lung, and Blood Institute GO Exome Sequencing Project (ESP) classified as “disease causing” in the widely used Human Gene Mutation Database (HGMD).⁷ Yang et al. expected to find a maximum of two patients with disease-causing *FBN1* variants in the ESP database but found 100 individuals with 23 different variants, indicating a misinterpretation of variants in the HGMD database. The aim of this

¹Department of Cardiology, Aarhus University Hospital, Aarhus, Denmark; ²Department of Molecular Medicine, Aarhus University Hospital, Aarhus, Denmark; ³Centre for Rare Diseases, Department of Paediatrics, Aarhus University Hospital, Aarhus, Denmark; ⁴Department of Clinical Genetics, Aarhus University Hospital, Aarhus, Denmark; ⁵Department of Endocrinology and Internal Medicine, Aarhus University Hospital, Aarhus, Denmark. Correspondence: Kristian A. Groth (kristian.groth@ki.au.dk)

Submitted 20 November 2014; accepted 9 February 2015; advance online publication 26 March 2015. doi:10.1038/gim.2015.32

study was to evaluate the quality of variant databases as they relate to these 23 likely benign variants.

MATERIALS AND METODS

Yang *et al.*⁶ identified 23 *FBN1* variants (**Supplementary Table S1** online) in ESP that were classified as disease-causing in HGMD. We searched the HGMD professional database (**Supplementary Table S1** online), the UMD-*FBN1* database⁸ (**Supplementary Table S2** online), the ClinVar database⁹ (**Supplementary Table S3** online), and the UniProt database¹⁰ (**Supplementary Table S4** online) for the 23 variants. In each database we identified reference material in as much detail as possible (**Supplementary Tables S1–S4** online). Published peer-reviewed articles were all identified via PubMed searches, and all material was accessible. The UMD-*FBN1* database also contained data classified as “personal communication,” which is not published in the literature and therefore accessible only via the UMD-*FBN1* database homepage (www.umd.be/FBN1/).

Each variant was evaluated according to published information (**Supplementary Table S5** online). The accessible data were evaluated according to the Ghent II nosology, and each variant was evaluated as “not MFS,” “maybe MFS” and “inconclusive” (**Table 1**). “Not MFS” indicates that the variant probably does not cause MFS based on a majority of reported phenotypes that

do not fulfill the Ghent II nosology. “Maybe MF” indicates that the variant could cause MFS, but there is not full documentation for a genotype–phenotype association. “Inconclusive” indicates that the data were insufficient to evaluate the variant’s effect on the phenotype. The term “MFS” also was intended to be used to describe variants that cause an evident MFS phenotype, but none of the variants had a clear genotype–phenotype association with an MFS phenotype fulfilling the Ghent II nosology. The manual evaluation of the variants then was compared with the database conclusions (**Table 2**).

RESULTS

Only the HGMD database contained all variants. The UMD-*FBN1* database contained all but one variant, ClinVar contained eight variants, and UniProt contained five variants (**Table 2**). The majority of references were overlapping in all databases. None of the databases did cover all references, and all four databases had unique references that were not recorded in the three other databases.

As expected, the HGMD database classified all 23 variants as “disease-causing mutations” and associated all the variants with MFS. The UMD-*FBN1* database had records for 22 variants, all classified as “mutation,” but one variant was in a single sub-record classified as a “polymorphism.” This specific variant was

Table 1 Summary of articles reviewed and the diagnostic conclusion as assessment for the documented phenotype

Number	Variant	Patients with MFS (n)	Patient with unknown MFS status (n)	Patients without MFS (n)	Diagnostic conclusion ^a
1	c.59A>G	0	0	1	Inconclusive
2	c.1027G>A	0	1	0	Inconclusive
3	c.1345G>A	(1)	0	0	Maybe MFS
4	c.2056G>A	0	0	1	Inconclusive
5	c.2927G>A	(1)	1	0	Maybe MFS
6	c.3058A>G	0	0	3	Not MFS
7	c.3422C>T	0	2	0	Inconclusive
8	c.3509G>A	0	8	12	Not MFS
9	c.3797A>T	0	1	0	Inconclusive
10	c.3845A>G	0	3	0	Inconclusive
11	c.4270C>G	3	4	5	Maybe MFS
12	c.6055G>A	0	0	2	Not MFS
13	c.6700G>A	0	3	2	Not MFS
14	c.7241G>A	0	2	0	Inconclusive
15	c.7379A>G	0	3	0	Inconclusive
16	c.7660C>T	0	1	0	Inconclusive
17	c.7661G>A	1	0	1	Inconclusive
18	c.7702G>A	0	2	0	Inconclusive
19	c.7846A>G	0	2	0	Inconclusive
20	c.7852G>A	0	2	3	Not MFS
21	c.8081G>A	0	1	0	Inconclusive
22	c.8176C>T	0	11	15	Not MFS
23	c.8494A>G	0	2	0	Inconclusive

^aThe conclusions are based on published data and according to the Ghent II criteria. “Not MFS” variants probably do not cause MFS based on the fact that the majority of reported phenotypes with this variant do not fulfill the Ghent II criteria. “Maybe MFS” indicates that the variant could result in the MFS phenotype but the full documentation on genotype–phenotype association is not provided. “Inconclusive” indicates that assessable data on the variant is inconclusive to assess the variant effect on phenotype. MFS, Marfan syndrome.

Table 2 Summary of conclusions in databases and conclusions of the manual evaluation of background material in this study

Number	Variant	ESP count	Database				This study's conclusion
			HGMD	UMD- <i>FBN1</i>	ClinVar	UniProt	
1	c.59A>G; p.Y20C	3	DCM: MFS	Mutation: incomplete MFS	NR	MFS	Inconclusive
2	c.1027G>A; p.G343R	2	DCM: MFS	Mutation	NR	NR	Inconclusive
3	c.1345G>A; p.V449I	2	DCM: MFS	Mutation: incomplete MFS	NR	NR	Maybe MFS
4	c.2056G>A; p.A686T	1	DCM: MFS	Mutation	NR	NR	Inconclusive
5	c.2927G>A; p.R976H	2	DCM: MFS	Mutation: classic MFS	Uncertain significance: all highly penetrant	NR	Maybe MFS
6	c.3058A>G; p.T1020A	3	DCM: MFS	(2)Mutation/(1)polymorphism: classic MFS/Lujan-Fryns syndrome/isolated skeletal features	Pathogenic/likely pathogenic: MFS	NR	Not MFS
7	c.3422C>T; p.P1141L	14	DCM: MFS	Mutation: classic MFS/isolated skeletal features	Uncertain significance: all highly penetrant	NR	Inconclusive
8	c.3509G>A; p.R1170H	25	DCM: MFS	Mutation: classic MFS/isolated skeletal features/incomplete MFS	Pathogenic/likely pathogenic: MFS, subdiagnostic variant of	MFS	Not MFS
9	c.3797A>T; p.Y1266F	4	DCM: MFS	Mutation	NR	NR	Inconclusive
10	c.3845A>G; p.N1282S	3	DCM: MFS	Mutation: MFS	NR	NR	Inconclusive
11	c.4270C>G; p.P1424A	4	DCM: MFS	Mutation: incompletes MFS/classic MFS/Shprintzen-Goldberg syndrome/unknown marfanoid syndrome	Pathogenic/likely pathogenic: MFS	MFS	Maybe MFS
12	c.6055G>A; p.E2019K	1	DCM: MFS	Mutation: probable MFS	NR	NR	Not MFS
13	c.6700G>A; p.V2234M	8	DCM: MFS	Mutation: incomplete MFS	Conflicting data from submitters: MFS, all highly penetrant	NR	Not MFS
14	c.7241G>A; p.R2414Q	1	DCM: MFS	Mutation	NR	NR	Inconclusive
15	c.7379A>G; p.K2460R	2	DCM: MFS	Mutation: MFS	NR	NR	Inconclusive
16	c.7660C>T; p.R2554W	1	DCM: MFS	Mutation: incomplete MFS	NR	NR	Inconclusive
17	c.7661G>A; p.R2554Q	1	DCM: MFS	Mutation: classic MFS	NR	NR	Inconclusive
18	c.7702G>A; p.V2568M	1	DCM: MFS	Mutation	NR	NR	Inconclusive
19	c.7846A>G; p.I2616V	4	DCM: MFS	Mutation	NR	NR	Inconclusive
20	c.7852G>A; p.G2618R	2	DCM: MFS	Mutation: incomplete MFS	Pathogenic/likely pathogenic: MFS	MFS	Not MFS
21	c.8081G>A; p.R2694Q	1	DCM: MFS	Mutation	NR	NR	Inconclusive
22	c.8176C>T; p.R2726W	14	DCM: MFS	Mutation: isolated skeletal features/classic MFS/incomplete MFS/MASS/marfanoid syndrome/Lujan-Fryns syndrome/AAA	Conflicting data from submitters: MFS	MFS	Not MFS
23	c.8494A>G; p.S2832G	1	DCM: MFS	mutation	NR	NR	Inconclusive

ESP individuals is the number of alleles registered in ESP6500 with the specific variant. HGMD, UMD-*FBN1*, and ClinVar are stated as classification conclusion and disease association.

DCM, disease-causing mutation; ESP, Exome Sequencing Project; HGMD, Human Gene Mutation Database; MFS, Marfan syndrome; NR, not registered.

still classified as a mutation in the overall database, and only by analyzing the specific input record of the patient with X-linked Lujan-Fryns syndrome was it clear that the record had been classified as a polymorphism. A polymorphism is historically defined as a variant more common than 1% in the background population, but the term is seldom used in the modern literature. In some records in the UMD-*FBNI* database, the variant was associated with a variety of syndromes and characteristic phenotypes. In many cases the same variant was associated with more than one syndrome/phenotype. Even syndromes not associated with *FBNI*-like Lujan-Fryns syndrome were found in the UMD-*FBNI* database.

ClinVar had records of eight variants, of which four were classified under the term “clinical significance” as “pathogenic/likely pathogenic”; two were classified as being of “uncertain significance” and two were classified as “conflicting data from submitters.” The six variants classified as either “pathogenic/likely pathogenic” or “conflicting data from submitters” were connected with MFS, whereas the “uncertain significance” variants were labeled as “all highly penetrant.”

The UniProt database had records of five variants, which all were associated with MFS.

Manual evaluation of the database references did not find evidence of any of the 23 variants being associated with MFS. Of the 23 variants, only 3 variants were classified as “maybe MFS,” indicating that the variant could result in the MFS phenotype but none of the identified references provided full documentation of a genotype–phenotype association. We classified 14 variants as “inconclusive,” indicating that the accessible literature on the variant precluded any definite conclusions concerning genotype–phenotype relations. Six variants were classified as “not MFS” because they most likely do not cause MFS based on the fact that the majority of the reported patients with this variant do not fulfill the Ghent II criteria.

DISCUSSION

The evaluation of 23 variants shows that the databases do contain misleading information on variants and their genotype–phenotype associations. Thus, clearly, the databases cannot be used as a direct source for diagnostics, but only as a tool for seeking additional information about the specific variant. An MFS diagnosis is based on a rather complex set of diagnostic criteria, which has changed over time. Interpretation of *FBNI* variants for defining MFS is rather difficult, and the analyst must have expert knowledge about MFS phenotypes and the diagnostic criteria. Descriptions such as “classic MFS” or “fulfilling the MFS criteria” are not precisely defined and not useful for classifying a given phenotype. MFS can be considered as a diagnosis about which knowledge is constantly evolving, and the diagnostic criteria therefore have to change over time.^{3,11,12} Describing specific phenotypical characteristics in the databases, such as aortic dilatation, ectopia lentis, or scoliosis, is necessary because they do not change with new diagnostic criteria, and reevaluating the MFS diagnosis if the nosology changes would be possible.

MFS is a rare disease, but the exact incidence is not fully known. Different estimates have been reported: from 17.2 per 100,000 (ref. 13) to 4.6 per 100,000 (ref. 14), but 10 in 100,000 is widely quoted.¹⁵ Precisely calculating the expected incidents of disease-causing *FBNI* mutations in the ESP cohort is difficult. ESP contains 6,503 samples from a cohort with heart, lung, and blood disorders. One small subcohort of 29 subjects, notated as “thoracic aortic aneurysms leading to acute aortic dissections,” is noteworthy because aortic dissection is highly associated with MFS. The ESP phenotype data are not publicly available, and verifying in which cohort each variant is detected is not possible. In the general ESP cohort one would expect 0.3 to 1.1 patients with MFS, but this could be supplemented with up to 29 extra patients because of the aortic dissection cohort. An expected maximum of around 30 patients compared with the actual 100 patients indicates an overrepresentation of benign variants among the 23 variants found in ESP and HGMD.

The introduction of next-generation sequencing into clinical diagnostics obviously will provide an increased amount of genotype versus phenotype information, but much of this information will never be reported in the databases because the main source of data in the databases is published peer-reviewed articles. Currently, publication of new or reconfirmed variants is not prioritized by most scientific journals. For this reason, publication bias is inevitable. Because the diagnostic strength of each variant is correlated with the amount of data collected about genotype versus phenotype, the future of MFS diagnostics is dependent on data collection from genetic testing and input directly from laboratories, not via published, peer-reviewed articles. Variant databases need to accept data on an individual level, but providing phenotype information as well is crucial.

The UMD-*FBNI* database does have the possibility for “personal communication” on variants, but such information relies on personal communication and in many cases does not contain specific information about the phenotype. The ClinVar database does provide data from a few laboratories, but these also do not provide phenotype data.

In a broad perspective, the *FBNI* databases do not seem to be ready to incorporate the benefit of the high output of data from next-generation sequencing. The many daily analyses are not incorporated in the databases, and the databases are not ready to “learn” from new data input. Only the ClinVar database provides information on “conflicting data from submitters.” The UMD-*FBNI* database registers all variants as mutations even though the variants may be associated with a variety of (up to seven different) phenotypes. It is necessary for variant databases to be able both to handle many data inputs with different and even conflicting data and to present these data in a way that provides the user with a clear understanding of the currently accessible information on the specific variant.

The evaluation of these 23 highly selected and likely benign variants in the *FBNI* gene shows that not all data in the databases are correctly classified. That references with no other

connection to MFS other than mentioning a variant in the *FBN1* gene can be found in the UMD-*FBN1* (ref. 16), HGMD,¹⁷ and ClinVar¹⁸ databases is worrying. Even a reference based on a publication about incidental findings of *FBN1* variants is, for some reason, suddenly associated with MFS in the HGMD database.¹⁷ UniProt contains only a minority of references compared with HGMD and UMD-*FBN1*, and this might be the reason why this database does not have references to any of these articles. Among the five UniProt records labeled “MFS,” three do not have MFS, one is inconclusive, and one might have MFS.

Yang *et al.*⁶ concluded that the genotype prevalence of MFS was 1:65 but question the causality of some of these variants and suggest “that these variants may not be the monogenic cause of MFS.” We think that their study shows that some researchers tend to use the databases rather uncritically.

Conclusion

The genetic diagnosis of MFS cannot be made reliably using only variant databases; it must be made through time-consuming evaluation of the background material in the databases and by combining these data with expert knowledge on MFS. Because the databases do not provide a reliable interpretation of variants, there is a substantial possibility of misdiagnosing MFS.

SUPPLEMENTARY MATERIAL

Supplementary material is linked to the online version of the paper at <http://www.nature.com/gim>

ACKNOWLEDGMENTS

The authors thank the National Heart, Lung, and Blood Institute GO Exome Sequencing Project and its ongoing studies, which produced and provided exome variant calls for comparison: the Lung GO Sequencing Project (HL-102923), the WHI Sequencing Project (HL-102924), the Broad GO Sequencing Project (HL-102925), the Seattle GO Sequencing Project (HL-102926), and the Heart GO Sequencing Project (HL-103010).

DISCLOSURE

The authors declare no conflict of interest.

REFERENCES

1. Marfan AB. Un cas de deformation des quatre membre plus prononcee aux extremitre caracterisee par l'allongment des os avec un certain degre d'aminicissement. *Bull Mem Soc Med Hop Paris* 1896;13:220–221.
2. Dietz HC, Cutting GR, Pyeritz RE, *et al.* Marfan syndrome caused by a recurrent de novo missense mutation in the fibrillin gene. *Nature* 1991;352:337–339.
3. Loeys BL, Dietz HC, Braverman AC, *et al.* The revised Ghent nosology for the Marfan syndrome. *J Med Genet* 2010;47:476–485.
4. Sheikhzadeh S, Kade C, Keyser B, *et al.* Analysis of phenotype and genotype information for the diagnosis of Marfan syndrome. *Clin Genet* 2012;82:240–247.
5. Bell CJ, Dinwiddie DL, Miller NA, *et al.* Carrier testing for severe childhood recessive diseases by next-generation sequencing. *Sci Transl Med* 2011;3:65ra4.
6. Yang RQ, Jabbari J, Cheng XS, *et al.* New population-based exome data question the pathogenicity of some genetic variants previously associated with Marfan syndrome. *BMC Genet* 2014;15:74.
7. Stenson PD, Ball EV, Mort M, *et al.* Human Gene Mutation Database (HGMD): 2003 update. *Hum Mutat* 2003;21:577–581.
8. Colod-Bérout G, Le Bourdelles S, Ades L, *et al.* Update of the UMD-FBN1 mutation database and creation of an FBN1 polymorphism database. *Hum Mutat* 2003;22:199–208.
9. Landrum MJ, Lee JM, Riley GR, *et al.* ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res* 2014;42(Database issue):D980–D985.
10. UniProt Consortium. Activities at the universal protein resource (UniProt). *Nucleic Acids Res* 2014;42:D191–D198.
11. De Paepe A, Devereux RB, Dietz HC, Hennekam RC, Pyeritz RE. Revised diagnostic criteria for the Marfan syndrome. *Am J Med Genet* 1996;62:417–426.
12. Beighton P, de Paepe A, Danks D, *et al.* International Nosology of Heritable Disorders of Connective Tissue, Berlin, 1986. *Am J Med Genet* 1988;29:581–594.
13. Sun QB, Zhang KZ, Cheng TO, *et al.* Marfan syndrome in China: a collective review of 564 cases among 98 families. *Am Heart J* 1990;120:934–948.
14. Fuchs J. Marfan syndrome and other systemic disorders with congenital ectopia lentis. A Danish national survey. *Acta Paediatr* 1997;86:947–952.
15. Gray JR, Bridges AB, Faed MJ, *et al.* Ascertainment and severity of Marfan syndrome in a Scottish population. *J Med Genet* 1994;31:51–54.
16. Robinson DO, Lin F, Lyon M, *et al.* Systematic screening of FBN1 gene unclassified missense variants for splice abnormalities. *Clin Genet* 2012;82:223–231.
17. Dorschner MO, Amendola LM, Turner EH, *et al.*; National Heart, Lung, and Blood Institute Grand Opportunity Exome Sequencing Project. Actionable, pathogenic incidental findings in 1,000 participants' exomes. *Am J Hum Genet* 2013;93:631–640.
18. Yuan B, Thomas JP, von Kodolitsch Y, Pyeritz RE. Comparison of heteroduplex analysis, direct sequencing, and enzyme mismatch cleavage for detecting mutations in a large gene, FBN1. *Hum Mutat* 1999;14:440–446.