Chapter 8: The Marfan Mutation Database

Gwenaëlle Collod-Beroud, Catherine Boileau

The Marfan Mutation Database

Gwenaëlle Collod-Béroud[1] and Catherine Boileau[2-3]

1- Laboratoire de Génétique Moléculaire, Institut Universitaire de Recherche Clinique, 641 avenue du Doyen Gaston Giraud, 34093 Montpellier Cedex 5, France. gwenaelle.beroud@igh.cnrs.fr

2- INSERM U383, Clinique Maurice Lamy, Hôpital Necker-Enfants Malades, 149 rue de Sèvres, 75743 Paris, cedex 15, France. boileau@necker.fr

3- Laboratoire de Biochimie, d'Hormonologie et de Génétique Moléculaire, Hôpital Ambroise Paré, AP-HP et Université Paris-Ife de France Ouest, 9 avenue Charles de Gaulle, 92104 Boulogne cedex, France. catherine.boileau@apr.ap-hop-paris.fr

Mutations in the fibrillin gene (*FBN1*) were described at first in Marfan syndrome (MFS) patients. Subsequently, the gene was shown to harbor mutations related to a spectrum of conditions clinically related to MFS, the "type-1 fibrillinopathies". In an effort to standardize the information regarding these mutations, to facilitate their mutational analysis and the identification of structure/function and phenotype/genotype relationships, we created, in 1995, a mutation database named "UMD-*FBN1*". The database follows the guidelines on mutation databases of the Hugo Mutation Database Initiative and gives access to a software package that provides specific routines and optimized multicriteria research and sorting tools. Recently, we have also developed a *FBN1* polymorphism database in order to facilitate diagnosis.

Mutations in the *FBN1* gene are associated with a wide range of phenotypes that show considerable variation in tissue distribution, timing of onset, and severity. The severe end of this broad spectrum of phenotypes is neonatal MFS. Conditions at the mild end include the MASS syndrome (Mitral valve prolapse, Aortic dilatation, and Skin and Skeletal manifestations) (OMIM#604308), mitral valve prolapse syndrome (OMIM#157700), isolated skeletal features, familial ascending aortic aneurysm and dissection (OMIM#132900), and ectopia lentis (OMIM#129600) with relatively mild skeletal features. Many of these conditions show significant overlap with MFS and are quite common in the general population. Finally, two syndromes

[dominant Weill-Marchesani[1] and Shprintzen-Goldberg OMIM#182212] have also been associated with mutations in the *FBN1* gene.

In an effort to standardize the information regarding *FBN1* mutations, we created a computerized database in 1995 using the UMD (Universal Mutation Database) software. This generic software was devised by Cariello et al. (ref 2) to create locus-specific databases (LSDBs) with the 4th Dimension® package from ACI[2,3]. The software includes an optimized structure to assist and secure data entry and to allow the input of clinical information. The software package contains routines for the analysis of the database. The use of the 4th Dimension ® (4D) SGDB gives access to optimized multicriteria research and sorting tools to select records from any field. The software has been successfully adapted to many different genes either involved in cancer (*p53*[2,4], *APC*[5,6], *VHL*[7,8], and *WT1*[9]) or in genetic diseases (*LDLR*[10-12], *VLCAD* (very long-chain acyl-CoA dehydrogenase) (unpublished) and *MCAD* (medium-chain acyl-CoA dehydrogenase) (unpublished)). Routines were specifically developed for the *FBN1* database and are described below[3,13-15].

DATABASE:

The *FBN1* database contains mutations either published or only reported in meetings proceedings, or contributed by the co-authors of published papers[13-16]. Recently, the database was modified to follow the guidelines on mutation databases of the Hugo Mutation Database Initiative including the new nomenclature[17]. Mutation names are numbered with respect to the *FBN1* gene cDNA sequence obtained from the Genbank database (Genbank database accession number L13923 ;

Complete coding sequence of HUM-FIBRILLIN Homo sapiens fibrillin mRNA). Intron-exon boundaries are as defined by Pereira et al. (ref. 18) and module organization is as in SWISS-PROT (accession number P35555).

The mutation records of the database include point mutations, large and small deletions, insertions, and splice mutations in the *FBN1* gene. The database cannot accommodate complex mutations or two mutations on the same allele. These are entered as two different records linked by the same sample ID. Each record contains the molecular and the clinical data for a given mutation in a standardized, easily accessible and summary form. Several levels of information are provided: at the gene level (exon and codon number, wild type and mutant codon, mutational event, mutation name), at the protein level (wild type and mutant amino acid, affected domain) and at the clinical level (absence or presence of skeletal, ocular, cardiovascular, central nervous system and other various manifestations). Data on fibrillin protein biosynthesis classification groups have been added when available as described in Aoyama et al. (ref. 19). In 2003, the database was updated to 563 mutations, of which over 400 are new entries.

ROUTINES

Initially, 6 routines were specifically developped[13]. The database now contains 13 routines that are described in the order in which they were implemented[14,15]:

1- "**Position**" studies the distribution of mutations at the nucleotide level to identify preferential mutation sites.

2- "**Statistical evaluation of mutational events**" is comparable to "Position" but also indicates the mutational event. The result can either be displayed as a table or in a graphic representation.

3- "**Frequency of mutations**" allows one to study the relative distribution of mutations at all sites and to sort them according to their frequency. A graphic representation is also available and displays a cumulative chart of mutation distributions.

4- "**Stat exons**" studies the distribution of mutations in the different exons. It enables detection of a statistically significant difference between observed and expected mutations.

5- "**Protein**" studies the distribution of mutational events in the various protein domains and repeated motifs (NH2 unique region, EGF-like motifs, cb EGF-like motifs, 8-cysteine motifs, hybrid motifs, proline rich region and COOH unique region).

6- "**Insertions and deletion analysis**" searches for repeated sequences flanking the mutation and possibly involved in the mutational mechanism.

7- "**Restriction enzyme**" appears on the first page of the mutation file. If the mutation modifies a restriction site, the program shows a restriction map displaying the new or abolished site and the enzyme of interest.

8- "**Amino acid type search**" studies the mutations with respect to phylogenetic conservation. In effect, the fibrillin gene has been identified and sequenced in two mammalian species [complete coding sequence of mouse fibrillin (*Fbn-1*) mRNA and complete coding sequence of bovine fibrillin (*BovFib*) mRNA, Genbank database

accession numbers L29454 and L28748 respectively]. The identity at the amino level between the human and bovine sequences is 97.8% and between the human and the mouse sequences of 96.2%. Therefore, the routine lists the mutations affecting non-conserved amino acids in the bovine, in the mouse, or in both sequences.

9- "**Binary comparison**" compares two mutation groups, each group being defined by distinct research criteria chosen from the database (molecular, clinical, age of onset, sex...). The result can be displayed as either of several graphic representations (by amino acids, by exon, or by protein domain) of the distribution of the sorted mutations. Furthermore the sorted mutations can appear by group or in a detailed format (insertion, deletion, missense, nonsense).

10- "**Amino acid changes**" lists for each of the 20 amino acids the observed substitutions throughout the protein.

11- "**Base modification**" lists the observed mutations with respect to their position within the codon for each of the 4 bases.

12- "**CpG**" studies the distribution of mutations occuring at CpG sites throughout the coding sequence. The result is displayed in a graphic representation.

13- "**Distribution of mutations**" lists the proportion of each of the mutational events observed in a selected group of mutation records.

Finally, several routines can be simultaneously studied and multicriteria searches can be performed.

MUTATION ANALYSES:

To date, 563 *FBN1* mutations have been identified and reported. Since the software cannot accommodate complex mutational events in a given individual, several mutations are not included in the current version of the database (3901_3904del; 3908_3909del (ref. 20); 1642del3ins20pb and 1888delAAinsC (ref. 21) and 1882_1884delinsAAA (ref. 22). The double mutant 3212T>G;3219A>G (I1071S;E1073D) reported by Wang et al. [23] is reported in two different records linked by the same sample ID, as well as for double mutant 3797A>T;5746T>A (Y1266F;C1916S) found in a French proband (unpublished data). Other mutations are spread throughout almost the entire gene without obvious predilection for any given region. So far, mutations were thought to be private and generally non recurrent. However, the recent database update shows that approximately 12 % of mutations are recurrent (56 recurrent mutations representing 156 events). Most of these mutations have been found in 2 to 3 probands. However, 8 are found in more than 4 subjects (table 1). Interestingly, almost all these mutational events affect a CpG, a mutation hot spot, and could truly be recurrent rather than associated with an unique haplotype. The case of mutation I2585T is baffling since it does not affect a CpG, it is absent in over 300 control chromosomes and has been found only in Marfan patients. Although these subjects were identified in different countries, the existence of a founder effect cannot be ruled out and further molecular and functional data need to be collected to investigate this finding.

Information on transmission is available for 398 mutations. Among these, there is a surprising number of *de novo* mutations compared to transmitted mutations (188 *de novo* vs. 210 familial cases). This finding could imply that sporadic cases are far more frequent than the 25% usually reported. However, the observed figure may

reflect a bias related to the fact that clinical diagnosis of MFS is often difficult and that the yield in mutation identification is low in the incomplete forms of the syndrome. Thus, molecular studies may be biased in favor of the study of probands with a complete and severe form of the disease. In effect, screening for FBN1 mutations has remained laborious, expensive, and limited to the 65 exons and their intronic flanking sequences. Thus, the mutations located within the large non coding sequence as well as in the promoter and the 5′ and 3′ regulatory regions go undetected. Finally, genetic heterogeneity could also partly explain the low number of mutations reported for familial forms of MFS.

The 492 mutations located within the coding sequence are generally distributed in all exons. However, when comparing the number of mutational events expected in a given exon to the number of mutations identified in the given exon, statistically significant differences appear in a few exons: there is under-representation of mutations in exons 45 (cbEGF-like#27) and 57 (8 cystein#7) while over-representation is observed in exons 13 (cbEGF-like#4), 26 (cbEGF-like#12), 27 (cbEGF-like#13), 28 (cbEGF-like#14) and 43 (cbEGF-like#25) (table 2). Over-representation in exons 26 to 28 can be explained by the fact that almost all the mutations identified in neonatal cases of MFS are located within this area (exons 24-32). Furthermore, mutations in this region are more likely to be associated with a severe clinical phenotype[24]. Severity probably leads to a bias in the types of patients selected for mutation detection.

The fibrillin gene has been identified and sequenced in vertebrates species: bovine (L28748), mouse (L29454), rat (AF135059), dog (partial cDNA, AF29080), pig (AF073800) and invertebrates: medusa (partial cDNA, L39930). The identity at the

amino acid level between mammals is so high (for example 97.8% human-bovine, 98% human-rat, and 96.2% human-mouse) that very often phylogenic conservation should be observed at the amino acid position affected by a given missense mutation. In fact, all reported mutations in the *FBN1* gene affect a conserved amino acid with respect to the bovine sequence. It is interesting to note that only 6 mutational events affect amino acids that are not conserved between mouse and man: three deletions (4179_4187del (Lesley Adès, Katherine Holman, personal communication 2000), 4177_4177delG and 7965_7977del (ref. 21), a duplication (6409_6411dup (Collod-Beroud, in preparation), a nonsense mutation (6339T>G) (ref. 25), for which the causality is obvious, and a missense mutation (3382G>A corresponding to V1128I) (ref. 26). In the medusa (*Podocoryne carnea*) the primary amino acid sequence (>40% sequence identity with mammalians), the highly repetitive multidomain structure, as well as the beaded microfibril appearance of fibrillin are highly conserved in man[27]. Fibrillin, as well as collagen, is thought to be a very early invention of metazoans in evolution[28]. Reber-Müller et al[27] suggest that with the invention of fibrillin, resilience and elasticity might have been added to the characteristics of ECM, thus providing the biomechanical basis for the development of a free-swimming medusa life stage.

Large rearrangements: In the *FBN1* gene, no major deletions have been reported except for the exons 60-62 genomic deletion[29] and two other multi-exon deletions[30]. Deletions of contiguous EGF-like domains have different effects depending on their location within the fibrillin-1 molecule. Deletion of the three contiguous cbEGF-like domains encoded by exons 44-46 resulted in a severe phenotype with onset in infancy and a rapidly progressing clinical course[30]. In frame

deletion of exons 42-43 was characterized in a patient presenting with bilateral ectopia lentis and Marfanoid skeletal features[30] and finally, deletion of the cbEGF-like domains encoded by exons 60-62 in the C-terminal domain of fibrillin-1 results in a much less severe phenotype characterized by a moderate Marfanoid habitus[29]. In the patient presenting with inframe deletion of exons 42-43, two sets of identical pentamers (cagta and ggaaa) were identified near the breakpoints in intron 41 and 43. In the deletion of exons 44-46, the exchange occurred within an identical pentamer (atttt). None of these sequences are known to predispose to genomic rearrangements. For exons 60-62 genomic deletion, data are not available.

Presently, the Human Gene Mutation Database (HGMD at www.hgmd.org) which is the largest general mutation database, contains 33 252 mutations in 1338 genes (public dataset numbers available online). Of these mutations, 1827 (5.49%) are gross deletions, 274 (0.86%) gross insertions and duplications, 57 (0.17%) are repeat variations, and 340 (1.0%) are complex rearrangements. Thus, although generally less frequent than point mutations, major rearrangements represent a mutational mechanism found in many disease genes. Therefore, it is surprising that so few have been reported in the *FBN1* gene. However, it is unclear if this mutation type has always been searched since Southern blotting is a long and time consuming technique that is no longer performed in many diagnostic laboratories.

Small insertion/deletion mutations (<20nt), duplications: Among the 80 small insertion/deletion mutations, 72 create a premature termination codon (PTC). These account for 12.9 % of the total mutations (small insertions: 22 cases and small deletions: 50 cases). They act as dominant negatives but display a highly variable

clinical phenotype, from severe to mild. The severity of the phenotype is may be related to the quantitative expression of the mutant allele and to the percentage of truncated proteins incorporated in the microfibrils[20,31]. In the 55 small deletions reported (table 3), 18 single base pair deletions can be the result of a mechanism of slipped mispairing, 4 small deletions (5791_5793delGTT, 8525_8529delTTAAC, 755_762del and 3603_3668del) are flanked by direct repeats and 3 mutations are deletions of a repeated sequence (635_636delCA, 3355_3358delAGAG and 4920_4923delTGAA). For the 30 other mutations, the mechanisms have yet to be determined by the search, among others, for the presence of quasi-palindromic sequences, inverted repeats or symmetric elements which facilitate the formation of secondary-structure intermediates[32]. Twenty-seven insertions have been reported so far (table 4). Six are insertions within runs of identical bases and can be explained by slippage mispairings at the replication fork. Eleven single base pair insertions correspond to the duplication of an existing base. Seven insertions are small duplications of existing sequences. For three mutations, the mechanisms, similar to those described above, have yet to be determined.

Splice mutations : The pre-mRNA splicing machinery recognizes exons and joins them together to form mRNAs with intact translational reading frames. Splicing requires canonical sequences at the intron/exon boundary. Three categories of mutations can be identified. The first one corresponds to mutations in canonical sequences and represents 60/73 splice mutations found the *FBN1* gene. They cause abnormal splicing patterns by the use of the nearest and strongest consensus splice site. The second category (10 mutations) corresponds to mutations not located in

canonical sequences (table 5). These last few years, different studies have indicated that distinct sequence elements that are distant from the splice sites are also needed for normal splicing. These elements can affect splice-site recognition during constitutive splicing and also play important roles in directing alternative splicing[33] [Cooper, 1997]. They can be auxiliary splicing elements (ASEs) required for cell-specific modulation of alternative splicing within introns that flank alternative exons, or exonic splicing enhancers (ESEs) within both coding and noncoding exons that direct specific recognition of splice sites during constitutive and alternative splicing. Two exonic mutations, a non sense mutation 6339T>G (Y2113X) (ref. 34) and a silent exonic mutation 6354C>T (I2118I) (ref. 35) have been reported as inducing in-frame skipping of the entire exon 51 and demonstrate the existence of an ESE sequence in exon 51 (refs. 36,37). Thirteen other mutations could belong to this category with mutations up to 53 bp away from the canonical sequence. In most cases, cDNA analysis is not available and abnormal splicing has not been demonstrated. Therefore, causality is still uncertain. Finally, the third category of mutations is provided by single base pair changes that introduce novel splice sites that substitute for the wild-type sites. A single recurrent mutation possibly creating a potential donor splice site has been reported but an abnormal splicing pattern has not been demonstrated (3294C>T). In the majority of splice site mutations, exon skipping results in an in-frame mRNA and produces a mutant fibrillin-1 missing a whole domain. The mutant allele produces abnormal monomers that considerably interfere with the assembly of fibrillin molecules in the microfibrils network. In a small number of patients (9 cases), the skipping of an exon causes a frameshift, a premature termination codon (PTC) and reduced mutant RNA levels through nonsense-

mediated decay of the mutant transcript[38]. Furthermore, MFS patients have been reported in whom the donor splice site mutation results in the incorporation of intronic sequence (IVS46+1G>A, IVS27+1G>A) into the transcript or in the use of a cryptic splice site inducing partial exon deletion (IVS18+2T>C, IVS37+5G>T).

Nonsense/missense: Nonsense (61 cases) and missense (337 cases) mutations represent 10.9% and 60.3% of mutations, respectively. Among missense mutations, more than three quarters (263/337) are located in calcium binding modules. These mutations either create (20/263, 7.6%) or substitute (129/263, 49%) cysteine residues potentially implicated in disulfide bonding. Pulse-chase studies on fibrillin-1 secretion from MFS patient fibroblasts have shown that these mutations often result in a delay in secretion/intracellular retention of profibrillin[19,39-41]. As three disulfide bonds are required to maintain the native cbEGF-like module fold, suppression or addition of cysteine residues would result in cbEGF-like module misfolding, which impairs trafficking[42-44]. The majority of the remaining mutations in these modules affects residues of the calcium consensus sequence and results in reduced calcium affinity, which may in turn destabilize the interface between two consecutive cbEGF-like modules. Calcium binding would rigidify the interdomain region between two cbEGF-like modules and allow multiple tandem cbEGF-like modules to take on a rigid, rod-like conformation[45-47]. Increased protease susceptibility due to reduced calcium affinity is a mechanism also reported for missense mutations. This pathological mechanism emphasizes the importance of calcium binding for the structural integrity of fibrillin-1. Mutations which do not belong to one of these subclasses may likely be involved in protein-protein interactions. Other modules are

carriers of one quarter of missense mutations and pathological mechanisms have yet to be clearly demonstrated.

The global molecular analysis of *FBN1* mutations reveals 2 classes of mutations. The first one, which represents more than one third of the mutations (38.6 %), contains mutations predicted to result in shortened fibrillin-1 molecules: 61 nonsense mutations, 71 splicing errors, 23 insertions and duplications, 51 deletions and 10 inframe deletions. They act as dominant negatives but display a highly variable clinical phenotype, of which severity is directly related to the quantitative expression of the mutant allele and to the percentage of truncated or shortened proteins incorporated in the microfibrils[20,31]. The second one represents less than two thirds (60.3%) of the mutations and contains missense mutations, mostly located in cbEGF-like modules (78%). They can be subclassified in (a) mutations creating or substituting cysteine residues potentially implicated in disulfide bonding and consequently in the correct folding of the monomer and (b) amino acids implicated in calcium binding and subsequently in interdomain linkage, rigidification of monomer and in protease susceptibility.

FBN1 POLYMORPHISM DATABASE:

Recently, we created an independent database for *FBN1* polymorphisms. Its goal is to make available a complete set of *FBN1* gene variations (mutations and polymorphisms) so that causative mutations may be quickly identified. The polymorphism database contains molecular as well as population data: size and

ethnicity of all the populations in which a given variation was identified, the number of tested chromosomes, and the frequency of each allele. In the future, patients in which these polymorphisms have been found should be added. This information can be helpful to determine if certain *FBN1* genotypes are associated with more severe phenotypes. These data should provide tools to start to interpret the phenotypic variability associated with a mutation in different probands or in the same family.

CONCLUSION

Elucidating the molecular basis of MFS and related fibrillinopathies is the major goal of the teams working on this subject[16,48]. The extreme clinical variability, the difficulties associated with clinical diagnosis and the low detection rate of mutations in this large gene all conspire to negatively impact on progress. At present it is not possible to predict the phenotype for a given *FBN1* mutation. On the one hand, mutations affecting different positions within a given module may be associated with quite different phenotypes. On the other hand, mutations affecting an analogous residue within two different modules may also be associated with differing phenotypes. Therefore, it is apparent that neither the location of the affected structural module in the protein nor the position of the altered residue is, in itself, sufficient to predict potential genotype-phenotype correlations[50]. The high degree of intrafamilial variability suggests that environmental and perhaps stochastic or epigenetic factors are important for the phenotypic expression of disease. The effects of unknown modifier (enhancing or protecting) genes on the clinical expression as well as conjugation of different alleles of the multiple fibrillin-interacting proteins are likely to constitute the foundation of an enhanced susceptibility for disease severity. All these hypotheses are starting points for future research.

Références :

1. Faivre L, Gorlin R, Wirtz M, Godfrey M, Dagoneau N, Lemerrer M, Collod-Béroud G, Boileau C, Munnich A, Cormier-Daire V. In-frame fibrillin-1 gene deletion in autosomal dominant Weill-Marchesani syndrome. *J Med Genet*. 2003;40:34-36.

2. Cariello N, Cui L, Béroud C and Soussi T. Database and software for the analysis of mutations in the human p53 gene. *Cancer Res*. 1994;54: 4454–4460.

3. Béroud C, Collod-Béroud G, Boileau C, Soussi T and Junien C. UMD (Universal mutation database): a generic software to build and analyze locus specific databases. *Hum Mutat*. 2000;15:86-94.

4. Béroud C and Soussi T. p53 gene mutation: software and data-base. *Nucl Acids Res*. 1998b;26:200–204.

5. Béroud C and Soussi T. APC gene: database of germline and somatic mutations in human tumors and cell lines. *Nucl Acids Res*. 1996a;24:121–124.

6. Laurent-Puig P, Béroud C and Soussi T. APC gene: database o germline and somatic mutations in human tumors and cel lines. *Nucl Acids Res*. 1998;26:269–270.

7. Béroud C, Joly D, Gallou C, Staroz F, Orfanelli M and Junien C. Software and database for the analysis of mutations in the VHL gene. *Nucl Acids Res*. 1998a;26:256–258.

8. Gallou C, Joly D, Méjean A, Starosz F, Martin N, Tarlet G, Orfanelli M, Bouvier R, Droz D, Chrétien Y, Maréchal J, Richard S, Junien C and Béroud C. Mutations of the VHL gene is sporadic renal cell carcinoma definition of a risk fac-tor for VHL patients to develop an RCC. *Hum Mutat*. 1999;13:464–475.

9. Jeanpierre C, Béroud C, Niaudet P and Junien C. Software anddatabase for the analysis of mutations in the human WT gene. *Nucl Acids Res*. 1998;26:271–274.

10. Varret M, Rabes J, Collod-Béroud G, Junien C, Boileau C and Béroud C. Software and database for the analysis of mutations in the human LDL receptor gene. *Nucl Acids Res*. 1997;25:172-80.

11. Varret M, Rabes J, Thiart R, Kotze M, Baron H, Cenarro A, Descamps O, Ebhardt M, Hondelijn J, Kostner G, Miyake Y, Pocovi M, Schmidt H, Schuster H, Stuhrmann M, Yamamura T, Junien C, Béroud C and Boileau C. LDLR Database, 2nd ed. New additions to the database and the soft-ware, and results of the first molecular analysis. *Nucl Acids Res*. 1998;26:248–252.

12. Villeger L, Abifadel M, Allard D, Rabes J, Thiart R, Kotze M, Béroud C, Junien C, Boileau C and Varret M. The UMD-LDLR database: Additions to the software and 490 new entries to the database. *Hum Mutat*. 2002;20:81-87.

13. Collod G, Béroud C, Soussi T, Junien C and Boileau C. Software and database for the analysis of mutations in the human FBN1 gene. *Nucl Acids Res*. 1996;24(1):137-140.

14. Collod-Béroud G, Béroud C, Ades L, Black C, Boxer M, Brock DJ, Godfrey M, Hayward C, Karttunen L, Milewicz D, Peltonen L, Richards R, Wang M, Junien C and Boileau C. Marfan Database (second edition): software and database for the analysis of mutations in the human FBN1 gene. *Nucl Acids Res*. 1997;25(1): 147-50.

15. Collod-Béroud G, Béroud C, Ades L, Black C, Boxer M, Brock D, Holman K, de Paepe A, Francke U, Grau U, Hayward C, Klein H, Liu W, Nuytinck L, Peltonen L, Alvarez Perez A, Rantamaki T, Junien C and Boileau C. Marfan Database (third edition): new mutations and new routines for the software. *Nucl Acids Res*. 1998;26(1):229-3.

16. Collod-Béroud G, S Le Bourdelles, L Ades, L Ala-Kokko, P Booms , M Boxer, A Child, P Comeglio, A de Paepe, JC. Hyland, K Holman, I Kaitila, G Matyas, L Nuytinck, L Peltonen, T Rantamaki, P Robinson, B Steinman, C Junien, C Béroud, C Boileau. New update of the UMD-*FBN1* mutation database and creation of a FBN1 polymorphism database. *Hum Mutat*. 2003, in press

17. den Dunnen J and Antonarakis S. Nomenclature for the description of human sequence variations. *Hum Genet*. 2001;109:121-4.

18. Pereira L, D'Alessio M, Ramirez F, Lynch JR, Sykes B, Pangilinan T and Bonadio J. Genomic organization of the sequence coding for fibrillin, the defective gene product in Marfan Syndrome. *Hum Mol Genet*. 1993;2:961-968.

19. Aoyama T, Francke U, Dietz H and Furthmayr H. Quantitative differences in biosynthesis and extracellular deposition of fibrillin in cultured fibroblasts distinguish five groups of Marfan syndrome patients and suggest distinct pathogenetic mechanisms. *J Clin Invest*. 1994;94:130-137.

20. Nijbroek G, Sood S, McIntosh I, Francomano C, Bull E, Pereira L, Ramirez F, Pyeritz R and Dietz H. Fifteen novel FBN1 mutations causing Marfan syndrome detected by heteroduplex analysis of genomic amplicons. *Am J Hum Genet*. 1995;57:8-21.

21. Schrijver I, Liu W, Odom R, Brenn T, Oefner P, Furthmayr H and Francke U. Premature termination mutations in FBN1: Distinct effets on differential allelic

expression and on protein and clinical phenotypes. *Am J Hum Genet*. 2002;71:223-237.

22. Rommel K, Karck M, Haverich A, Schmidtke J and Arslan-Kirchner M.. Mutation screening of the fibrillin-1 (FBN1) gene in 76 unrelated patients with Marfan syndrome or Marfanoid features leads to the identification of 11 novel and three previously reported mutations. *Hum Mutat*. 2002;20:406-407.

23. Wang M, Kishnani P, Decker-Phillips M, Kahler SG, Chen YT and Godfrey M. Double mutant fibrillin-1 (FBN1) allele in a patient with neonatal Marfan syndrome. *J Med Genet*. 1996;33(9):760-3.

24. Tiecke F, Robinson P, Booms P, Katzke S, Hagemeier C, Neumann L, Hinkel G, Brenner R, Hövels-Gürich H, Fuchs J, Skovby F and Rosenberg T. Classic, early-onset severe and neonatal Marfan syndrome: Twelve mutations and genotype-phenotype correlations in FBN1 exons 24-40. *Eur J Hum Genet*. 2001;9:13-21.

25. Dietz H, Valle D, Francomano C, Kendzior RJ, Pyeritz R and Cutting R. The skipping of constitutive exons in vivo induced by nonsense mutations. *Science*. 1993b;259:680-683.

26. Loeys B, Nuytinck L, Delvaux I, De Bie S and De Paepe A. Genotype and phenotype analysis of 171 patients referred for molecular study of the fibrillin-1 gene FBN1 because of suspected Marfan syndrome. *Arch Intern Med*. 2001;161(20):2447-54.

27. Reber-Müller S, Spissinger T, Schuchert P, Spring J and Schmid V. An extracellular matrix protein of jellyfish homologous to mammalian fibrillins forms different fibrils depending on the life stage of the animal. *Dev Biol*. 1995;169:662-672.

28. Doolittle R. Reconstructing history with amino acid sequences. *Protein Sci*. 1992;1:191-200.

29. Kainulainen K, Sakai L, Child A, Pope F, Puhakka L, Ryhanen L, Palotie A, Kaitila I and Peltonen L. Two mutations in Marfan syndrome resulting in truncated fibrillin polypeptides. *Proc Natl Acad Sci USA*. 1992;89:5917-5921.

30. Liu W, Schrijver I, Brenn T, Furthmayr H and Francke U. Multi-exon deletions of the FBN1 gene in Marfan syndrome. *BMC Med Genet*. 2001;2(1): 1.

31. Karttunen L, Ukkonen T, Kainulainen K, Syvänen A-C and Peltonen L. Two novel fibrillin-1 mutations resulting in premature termination codons but in different mutant transcript levels and clinical phenotypes. *Hum Mutat*. 1998;supplement 1:S34-S37.

32. Krawczak M and Cooper D. Gene deletions causing human genetic disease: mechanisms of mutagenesis and the role of the local DNA sequence environment. *Hum Genet*. 1991;86:425-441.

33. Cooper T and Mattox W. Gene regulation'97. The regulation of splice-site selection, and its role in Human disease. *Am J Hum Genet*. 1997;61:259-266.

34. Dietz H, McIntosh I, Sakai L, Corson G, Chalberg S, Pyeritz R and Francomano C. Four novel FBN1 mutations: significance for mutant transcript level and EGF-like domain calcium binding in the pathogenesis of Marfan syndrome. *Genomics*. 1993a;17:468-475.

35. Liu W, Qian C and Francke U. Silent mutation induces exon skipping of fibrillin-1 gene in Marfan syndrome. *Nat Genet*. 1997;16(4):328-9.

36. Dietz H and Kendzior RJ. Maintenance of an open reading frame as an additional level of scrutiny during splice site selection. *Nat Genet*. 1994;8(2):183-8.

37. Caputi M, Kendzior RJ and Beemon K. A nonsense mutation in the fibrillin-1 gene of a Marfan syndrome patient induces NMD and disrupts an exonic splicing enhancer. *Genes Dev*. 2002;16(14):1754-9.

38. Frischmeyer PA and Dietz HC. Nonsense-mediated mRNA decay in health and disease. *Hum Mol Genet*. 1999;8(10): 1893-1900.

39. Aoyama T, Tynan K, Dietz HC, Francke U and Furthmayr H. Missense Mutations Impair Intracellular Processing of Fibrillin and Microfibril Assembly in Marfan Syndrome. *Hum Mol Genet*. 1993;2(12):2135-2140.

40. Halliday D, Hutchinson S, Kettle S, Firth H, Wordsworth P and PA. H. Molecular analysis of eight mutations in FBN1. *Hum Genet*. 1999;105:587-597.

41. Schrijver I, Liu W, Brenn T, Furthmayr H and Francke U. Cysteine substitutions in epidermal growth factor-like domains of fibrillin-1: distinct effects on biochemical and clinical phenotypes. *Am J Hum Genet*. 1999;65:1007-20.

42. Johnson A and Haigh N. The ER translocon and retrotranslocation: is the shift into reverse manual or automatic? *Cell*. 2000;102:709-712.

43. Lippincott-Schwartz J, Roberts T and Hirshberg K. Secretory protein trafficking and organelle dynamics in living cells. *Ann Rev Cell Dev Biol*. 2000;16:557-589.

44. Lord J, Davey J, Frigerio L and Roberts L. Endoplasmic reticulum-associated protein degradation. *Cell Dev Biol*. 2000;11:159-164.

45. Downing A, Knott V, Werner J, Cardy C, Campbell I and Handford P. Solution structure of a pair of calcium-binding epidermal growth factor-like domains: implications for the Marfan syndrome and other genetic disorders. *Cell*. 1996;85:597-605.

46. Knott V, Downing AK, Cardy CM and Handford P. Calcium binding properties of an epidermal growth factor-like domain pair from human fibrillin-1. *J Mol Biol*. 1996;255(1): 22-7.

47. Cardy CM and Handford PA. Metal ion dependency of microfibrils supports a rod-like conformation for fibrillin-1 calcium-binding epidermal growth factor-like domains. *J Mol Biol*. 1998;276(5):855-60.

48. Collod-Béroud G, Boileau C. Marfan syndrome in the third millenium. *Eur J Hum Genet*., 2003;10: 673-681.

49. Robinson P, Booms P, Katzke S, Ladewig M, Neumann L, Palz M, Pregla R, Tiecke F and Rosenberg T. 2002. Mutations of FBN1 and genotype-phenotype correlations in Marfan syndrome and related fibrillinopathies. Hum Mutat 20: 153-61.

50. Palz M, Tiecke F, Booms P, Goldne rB, Rosenberg T, Fuchs J, Skovby F, Schumache rH, Kaufmann U, von Kodolitsch Y, Nienaber C, Leitner C, Katzke S, Vetter B, Hagemeier C and Robinson P. Clustering of mutations associated with mild Marfan-like phenotypes in the 3' region of FBN1 suggests a potential genotype-phenotype correlation. *Am J Med Genet*. 2000;91:212-21.

Table 1: Recurrent mutations found in the *FBN1* gene.

Events	Position (AA)	WT codon	Mutant codon	Mutation name	Number of records
184C>T	62	CGT	TGT	R62C	5
IVS2+1G>A	83	CCC	spl+1	IVS2+1G>A	5
364C>T	122	CGC	TGC	R122C	4
1633C>T	545	CGC	TGC	R545C	4
3037G>A	1013	GGA	AGA	G1013R	6
4930C>T	1644	CGA	TGA	R1644X	4
IVS46+5G>A	1930	GAT	spl-2	IVS46+5G>A	8
7754T>C	2585	ATT	ACT	I2585T	6

Table 2: FBN1 exons displaying an abnormal number of mutations

Exon	Domain encoded	Expected mutations	Observed mutations	*p*
1	NH2-ter	**8.0**	1	>0.02
7	cb EGF-like # 1	**6.0**	**1**	>0.05
13	cb EGF-like # 4	7.2	**17**	>0.001
26	cb EGF-like # 12	7.6	**18**	>0.001
27	cb EGF-like # 13	6.9	**16**	>0.001
28	cb EGF-like # 14	7	**13**	>0.05
43	cb EGF-like # 25	7.1	**13**	>0.05
45	cb EGF-like # 27	**7.5**	2	>0.05
57	TGF β BP-like # 7	**11.8**	5	>0.05
65	NH2-ter	20	**7**	>0.01

The ‘‘Stat exons’’ routine studies the distribution of exonic mutations and enables detection of a statistically significant difference between observed and expected mutations. The algorithm takes into account the mutability of each base from an exon. The **mutability is defined as follows:** for each base, the significance of a mutation is defined by its ability to produce a new amino acid. In these conditions, the specific position of the base within the codon has a major incidence. If any substitution result in a new amino acid, its individual mutability is 3. If only two substitutions result in a new amino acid, mutability is 2... The exon’s mutability is defined by the addition of all mutabilities for each base. The expected value is calculated by the formula:

(exon mutability / by the CDNA mutability)*observed mutations

The *p* value is calculated using the usual Chi square formula.

Table 3: Small deletions identified in the *FBN1* gene involving repeated sequences.

Mutation name	File #	Exon	Mutation position	Codon	WT codon	Mutant codon	Event	Repeat type	WT AA	Mutant AA
7965_7977del	594	63	7963	2655	GCG	del13c	Stop at 2677	Deletion flanked by direct repeats	Ala	Fr.
5791_5793delGTT	467	47	5791	1931	GTT	del3a	In frame del	Deletion flanked by direct repeats	Val	InF
8525_8529delTTAAC	328	65	8524	2842	CTT	del5b	Stop at 2848	Deletion flanked by direct repeats	Leu	Fr.
755_762del	225	7	754	252	GCC	del8b	Stop at 261	Deletion flanked by direct repeats	Ala	Fr.
4178_4178delA	586	33	4177	1393	GAA	del1b	Stop at 1412	Deletion in a repetition of A	Glu	Fr.
3767_3767delA	585	30	3766	1256	AAT	del1b	Stop at 1275	Deletion in a repetition of A	Asn	Fr.
7577_7577delA	435	61	7576	2526	AAT	del1b	Stop at 2681	Deletion in a repetition of A	Asn	Fr.
4704_4704delA	208	37	4702	1568	AAA	del1c	Stop at 1580	Deletion in a repetition of A	Lys	Fr.
7291_7291delA	207	58	7291	2431	ACT	del1a	Stop at 2437	Deletion in a repetition of A	Thr	Fr.
3192_3192delA	82	25	3190	1064	GAA	del1c	Stop at 1087	Deletion in a repetition of A	Glu	Fr.
1836_1836delA	102	14	1834	612	AAA	del1c	Stop at 624	Deletion in a repetition of A	Lys	Fr.
3238_3238delC	500	26	3238	1080	CTC	del1a	Stop at 1087	Deletion in a repetition of C	Leu	Fr.
526_526delC	492	5	526	176	CAG	del1a	Stop at 189	Deletion in a repetition of C	Gln	Fr.
4485_4485delC	286	36	4483	1495	ACC	del1c	Stop at 1519	Deletion in a repetition of C	Thr	Fr.
2399_2399delC	230	19	2398	800	CCT	del1b	Stop at 802	Deletion in a repetition of C	Pro	Fr.
5311_5311delC	205	43	5311	1771	CGG	del1a	Stop at 1892	Deletion in a repetition of C	Arg	Fr.
4020_4020delC	52	32	4018	1340	ACC	del1c	Stop at 1412	Deletion in a repetition of C	Thr	Fr.
124_124delG	436	1	124	42	GCC	del1a	Stop at 107	Deletion in a repetition of G	Ala	Fr.
6996_6996delT	522	56	6994	2332	CTT	del1c	Stop at 2397	Deletion in a repetition of T	Leu	Fr.
4356_4356delT	495	35	4354	1452	CTT	del1c	Stop at 1474	Deletion in a repetition of T	Leu	Fr.
6018_6018delT	206	48	6016	2006	CTT	del1c	Stop at 2058	Deletion in a repetition of T	Leu	Fr.
1604_1604delT	51	13	1603	535	TTA	del1b	Stop at 578	Deletion in a repetition of T	Leu	Fr.
635_636delCA	354	6	634	212	ACA	del2b	Stop at 221	Deletion of a repeated sequence	Thr	Fr.
4920_4923delTGAA	193	39	4918	1640	AAT	del4c	Stop at 1648	Deletion of a repeated sequence	Asn	Fr.
3355_3358delAGAG	583	27	3355	1119	AGA	del4a	Stop at 1160	Deletion of a repeated sequence	Arg	Fr.
6497_6616del	555	52-53	6496	2166	GAT	del120b	In frame del	gDNA defect not found yet	Asp	InF

Reference sequence on which the +1 nt residue is based is L13923.
File # : indicates the file record in the database **; Mutant codon:** "**del**" followed by the number of deleted bases then by the position of the deletion **"a":** first base of the codon deleted, **"b":** second base of the codon deleted, **"c":** third base of the codon deleted. **InF**: in frame; **Fr**: Frameshift;

Table 4: Insertions and duplications reported in the *FBN1* gene and involving repeated sequences.

Mutation name	File #	Exon	Mutation position	Codon	WT codon	Mutant codon	Event	Type	Mutation type	WT AA	Mutant AA
4699_4721dup	519	37	4720	1574	TGT	ins23c	Stop at 1588	Fr.	Duplication	Cys	Fr.
5470_5484dup	518	44	5485	1829	GGC	ins15a	In frame ins	Fr.	Duplication	Gly	InF
6409_6411dup	513	52	6412	2138	AAA	ins3a	In frame ins	Fr.	Duplication	Lys	InF
3228_3232dup	554	26	3232	1078	CCT	ins5b	Stop at 1089	Fr.	Duplication	Pro	Fr.
5470_5484dup	481	44	5485	1829	GGC	ins15a	In frame ins	Fr.	Duplication	Gly	InF
5236_5240dup	248	42	5239	1747	CTC	ins5c	Stop at 1894	Fr.	Duplication	Leu	Fr.
5134_5137dup	9	41	5137	1713	AAC	ins4b	Stop at 1735	Fr.	Duplication	Asn	Fr.
4703_4704dup	423	37	4705	1569	GCC	ins2a	Stop at 1581	Fr.	Duplication = Insertion after an A repeat	Ala	Fr.
3444_3444dup	237	27	3445	1149	AAC	ins1a	Stop at 1158	Fr.	Duplication = Insertion after a C repeat, Consensus sequence for N-glycosylation	Asn	Fr.
6285_6285dup	595	50	6285	2096	TGC	Ins1a	Stop at 2104		Duplication = Insertion after an A repeat	Cys	Fr.
5065_5065dup	424	40-41	5065	1689	GAT	ins1b	Stop at 1702	Fr.	Duplication = Insertion after a G repeat	Asp	Fr.
7818_7818dup	575	62-63	7819	2607	GAT	ins1a	Stop at 2607	Fr.	Duplication = Insertion after a T repeat	Asp	Fr.
3525_3525dup	584	28	3526	1176	GGG	ins1a	Stop at 1192	Fr.	One base duplication	Gly	Fr.
3525_3525dup	584	28	3526	1176	GGG	ins1a	Stop at 1192	Fr.	One base duplication	Gly	Fr.
7116_7116dup	553	57	7117	2373	CCC	ins1a	Stop at 2376	Fr.	One base duplication	Pro	Fr.
623_623dup	469	6	622	208	CTC	ins1c	Stop at 222	Fr.	One base duplication	Leu	Fr.
2586_2586dup	434	21	2587	863	GAG	ins1a	Stop at 863	Fr.	One base duplication	Glu	Fr.
5499_5499dup	433	44	5500	1834	GAC	ins1a	Stop at 1834	Fr.	One base duplication	Asp	Fr.
959_959dup	432	8	958	320	TAC	ins1c	Stop at 320	Fr.	One base duplication	Tyr	Fr.
1378_1378dup	306	11	1378	460	TGT	ins1b	Stop at 475	Fr.	One base duplication	Cys	Fr.
7790_7790dup	298	62	7789	2597	CTC	ins1c	Stop at 2607	Fr.	One base duplication	Leu	Fr.
2570_2570dup	297	21	2569	857	GTC	ins1c	Stop at 859	Fr.	One base duplication	Val	Fr.
6185_6185dup	254	50	6184	2062	TAT	ins1c	Stop at 2062	Fr.	One base duplication	Tyr	Fr.

Reference sequence on which the +1 nt residue is based is L13923.
File # : indicates the file record in the database **; Mutant codon**: "**ins**" followed by the number of inserted bases then by the position of the insertion **"a"**: insertion before the first base of the codon, **"b"**: insertion before the second base of the codon, **"c"**: insertion before the third base of the codon. **InF**: in frame; **Fr**: Frameshift.

Table 5: Unusual splice mutations

Mutation name	File #	Exon	Mutation position	Codon	WT codon	Mutant codon	Event	Probable effect on cDNA	WT AA	Mutant AA
I. Mutations not located in canonical donor or acceptor splice site sequence										
IVS28+15del3	216	28-29	3589	1197	GAC	spl+15	del3	ND	Asp	Spl.
IVS56+17delG	479	56-57	6997	2333	GAC	spl+17	delG	ND	Asp	Spl.
IVS30+28C>A	217	30-31	3838	1280	GAT	spl+28	C->A	ND	Asp	Spl.
IVS12+21G>A	189	12-13	1588	530	GAC	spl+21	G->A	ND	Asp	Spl.
IVS52+50C>T	155	52-53	6496	2166	GAT	spl+50	C->T	ND	Asp	Spl.
IVS20+53T>A	213	20-21	2539	847	GAA	spl+53	T->A	ND	Glu	Spl.
IVS2-7T>G	314	2-3	247	83	CCC	spl-7	T->G	ND	Pro	Spl.
IVS35-8G>A	440	35-36	4459	1487	GAT	spl-8	G->A	ND	Asp	Spl.
6339T>G	57	51	6339	2113	TAT	TAG	T->G	Nonsense mutation inducing skipping of exon 51, in frame	Tyr	Stop
6354C>T	145	51	6354	2118	ATC	ATT	C->T	Silent mutation inducing skipping of exon 51, in frame	Ile	Ile
II. Mutations creating potential donor or acceptor splice site										
3294C>T	215	26	3294	1098	GAC	GAT	C->T	ND	Asp	Asp
3294C>T	359	26	3294	1098	GAC	GAT	C->T	ND	Asp	Asp
3294C>T	329	26	3294	1098	GAC	GAT	C->T	ND	Asp	Asp

Reference sequence on which the +1 nt residue is based is L13923.
File # : indicates the file record in the database**. ND**: Skipping of exon not demonstrated by cDNA analysis. **Spl**: splice; **PTC**: Premature Termination Codon.