A new locus-specific database (LSDB) for mutations in the TGFBR2 gene: UMD-TGFBR2.
Melissa Yana Frederic, Dalil Hamroun, Laurence Faivre, Catherine Boileau, Guillaume Jondeau, Mireille Claustres, Christophe Béroud, Gwenaëlle Collod-Béroud

To cite this version:

HAL Id: inserm-00343980
https://inserm.hal.science/inserm-00343980v2
Submitted on 20 Dec 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
A New Locus-Specific Database (LSDB) for Mutations in the TGFBR2 Gene: UMD-TGFBR2

Melissa Yana Frederic,1,2 Dalil Hamroun,3 Laurence Faivre,4 Catherine Boileau,5–7 Guillaume Jondeau,6,8 Mireille Claustres,1–3 Christophe Béroud,1–3 and Gwenäelle Collod-Béroud1,2*

1INSERM, U 827, Montpellier, F-34000, France; 2Université Montpellier1, Unité Fonctionnelle de Recherche (UFR) de Médecine, Laboratoire de Génétique Moléculaire, Montpellier, F-34000, France; 3Centre Universitaire Hospitalier (CHU) Montpellier, Hôpital Arnaud de Villeneuve, Laboratoire de Génétique Moléculaire, Montpellier, F-34000, France; 4Centre Universitaire Hospitalier (CHU) Dijon, Centre de Génétique, Dijon, F-21000, France; 5INSERM, U 781, Paris, F-75015, France; 6Université Versailles Saint Quentin-en-Yvelines, Unité Fonctionnelle de Recherche (UFR) Paris-Ile-de-France-Ouest (P.I.F.O.), Garches, F-92000, France; 7Assistance Publique-Hôpitaux de Paris (AP-HP), Hôpital Ambroise Paré, Laboratoire de Biochimie, d’Hormonologie et de Génétique moléculaire, Boulogne, F-92000, France; 8Assistance Publique-Hôpitaux de Paris (AP-HP), Hôpital Bichat, Consultation Marfan, Paris, F-75000, France

The implication of mutations in the TGFBR2 gene, known to be involved in cancers, in Marfan syndrome (MFS) and later in Loeys-Dietz syndrome (LDS) and Familial Thoracic Aortic Aneurysms and Dissections (TAAD2) gives a new example of the complexity of one gene involved in multiple diseases. To date, known TGFBR2 mutations are not disease-specific and many mutations have to be accumulated before genotype-phenotype relationships emerge. To facilitate mutational analysis of the TGFBR2 gene, a locus-specific database has been set up with the Universal Mutation Database (UMD) software. The version of the computerized database contains 85 entries. A total of 12 mutations are reported to be involved in MFS, six in incomplete MFS, 30 in LDS type I, 10 in LDS type II, seven in TAAD2, and 20 in various cancers. The database is accessible online at http://www.umd.be

KEY WORDS: TGFBR2; TGF beta signalopathies; Marfan syndrome; Loeys Dietz syndrome; locus-specific database

INTRODUCTION

Many diseases are caused by an inherited variation in just one gene. Among more than 7,000 human rare diseases inherited in a Mendelian fashion (one gene-one disease), nearly 4,000 genes associated with a given phenotype have now been mapped. These molecular genetic analyses identify a more complex scenario in which the same gene can be involved in disorders thought to be clinically distinct. The TGFBR2 gene (MIM# 190182) belongs to this more complex category. Transforming growth factor beta receptor type II (TβR-II) is a transmembrane protein serine/threonine kinase well known as having a tumor suppressor role and inhibiting cell proliferation. Receptor activation occurs upon binding of TGFβ to TβR-II, which then recruits and phosphorylates TβR-I, which propagates the signal to downstream targets, Smads (Fig. 1). Smad complexes then translocate to the nucleus, where together with a coactivator or a corepressor they regulate transcription of target genes [Attisano et al., 1996; Souchelnytskyi et al., 1996; Ullrich and Schlessinger, 1990; Wieser et al., 1995; Wrana et al., 1994a, 1994b]. Perturbations of the TGFβ/Smad signaling system are the cause of various forms of human cancers as well as developmental disorders [Massague, 1998]. With somatic mutations and loss of expression of genes for various components of the TGFβ/SMAD signaling pathway, growth inhibition is lost, resulting in unregulated cell growth with tumor formation. A large number of pancreatic and colorectal cancers have mutations in some component of the pathway. Mutations in the gene encoding TβR-II, TGFBR2, usually reported in colon and gastric cancer lead to truncated, inactive TGFBR2 product by mutations in a polyadenosine tract (BatII region) [Markowitz et al., 1995]. In most of these cases, both TGFBR2 alleles present with a mutation in the BatII sequence. In some cases, however, a different mutation inactivates the second allele such as loss of heterozygosity, addition of a GT dinucleotide to a (GT)$_3$ sequence in the kinase domain coding region, or missense mutations predicted to inactivate this kinase domain. TGFBR2 then shares the two-hit inactivation mechanism of other tumor suppressor genes.

In 2004, we demonstrated the implication of six TGFBR2 mutations in Marfan syndrome (MFS) [Mizuguchi et al., 2004]. MFS, the founding member of heritable disorders of connective tissue,
is a dominantly inherited condition characterized by tall stature and skeletal deformities, dislocation of the lens, and propensity to aortic dissection. The syndrome is characterized by considerable variation in the clinical presentation between families and also within the same family. The leading cause of premature death is progressive dilation of the aortic root and ascending aorta, causing aortic incompetence and dissection. The average life expectancy has risen significantly since 1972, which is partly due to the benefit arising from aortic follow-up with echocardiography, aortic surgery, and medical therapy with beta blockers. The first mutations involved in classical MFS phenotypes have been found in the FBN1 gene encoding fibrillin-1, the major component of the extracellular matrix [Dietz et al., 1991]. FBN1 mutations have now been identified in a spectrum of phenotypically-related connective tissue disorders, termed type I fibrillinopathies, including severe neonatal Marfan syndrome [Kainulainen et al., 1994], autosomal dominantly inherited ectopia lentis [Kainulainen et al., 1994], isolated skeletal features (or Marfanoid skeletal syndrome) [Ades et al., 2002; Hayward et al., 1994; Milewicz et al., 1995], “MASS phenotype” (MASS is an acronym that designates the involvement of the mitral valve, aorta, skeleton, and skin) [Dietz et al., 1993], familial or isolated forms of aortic aneurysms [Milewicz et al., 1996], and autosomal dominant Weill-Marchesani syndrome [Favier et al., 2003]. Heterozygous mutations in TGFBR2, a putative tumor-suppressor gene involved in several malignancies, are also associated with MFS [Mizuguchi et al., 2004]. The implication of TGFBR2 mutations, not only in MFS [Boileau et al., 2005] but also in other autosomal dominant Mendelian disorders such as Loeys-Dietz syndrome (LDDAA) [Loeys et al., 2005] and Familial Thoracic Aortic Aneurysms and Dissections (TAAD2) [Pannu et al., 2005], revealed all the complexity of the TGF-β signal transduction network. Mutations in the TGFBR2 gene in MFS2, LDS, and TAAD2 define a new group of Marfan syndrome-related connective tissue disorders: TGFβ signalopathies.

To date, known TGFBR2 mutations are not disease-specific and many mutations have to be accumulated before genotype-phenotype relationships emerge. To facilitate mutational analysis of the TGFBR2 gene, and as for FBN1 [Boileau et al., 1996; Boileau et al., 1997, 1998, 2003], we have created a human TGFBR2 mutation database, Universal Mutation Database (UMD)-TGFBR2. The information regarding these mutations is standardized to facilitate their mutational analysis and, when a greater number of mutations has been collected, identify structure-function and phenotype-genotype relationships. This database gives access to a software package that provides specific routines and optimized multicriteria research and sorting tools [Beroud et al., 2000, 2005].

**THE TGFBR2 GENE AND TBRII**

TGFBR2 gene maps to 3p22 [Matthew et al., 1994]. The TβRII protein is encoded by 567 codons in seven exons (Fig. 2A) [Takahashi et al., 1996]. The translated product of the TGFBR2 gene is composed of several subdomains (Fig. 2B). Starting from the N-terminus, there is a signal peptide (1–22) [Lin et al., 1992], a relatively short cysteine-rich extracellular domain (51–142), serine/threonine protein kinase catalytic domains (codons 246–544), and the C-terminus. The extracellular region is N-glycosylated [Wells et al., 1997] and contains 12 cysteines probably involved in the general folding of this region. Three of these cysteines form a characteristic cluster near the transmembrane sequence [Chidli et al., 1993; Massague, 1992]. The transmembrane region and the cytoplasmic juxtamembrane region have no singular structural features. The cytoplasmic region contains kinase domains conforming to the canonical sequence of a serine/threonine protein kinase domain (Hanks et al., 1988). TβRII is regulated intracellularly by autophosphorylation on at least three serine residues (Ser213 in the juxtamembrane region and Ser409 et Ser416 in the 1-loop region of the kinase domain [Luo and Lodish, 1997]). The TβRII exists in an alternative form arising from the presence of a 25-amino acid insert following the signal sequence [Hirai and Fujita, 1996; Suzuki et al., 1994].

![TGF-β signaling](image-url)

**FIGURE 1.** TGF-β signaling. Binding of dimeric TGF-β to type II receptor (TβR-II) in concert with type I receptor (TβR-I) leads to the formation of a receptor complex and the phosphorylation of TβR-I. Thus activated, TβR-I subsequently phosphorylates a receptor-regulated SMAD (R-SMAD), allowing this protein to associate with coactivator (CO-SMAD) and move into the nucleus. In the nucleus, the SMAD complex associates with a DNA-binding partner and this complex binds to specific enhancers in target genes, activating transcription. Smurfs (Smad-ubiquitin regulatory factor) bind to the activated TGF-β receptor complex via I-Smads. Smurf then ubiquitinates the receptor, causing its rapid proteasomal degradation. The CO-SMAD/R-SMAD can interact with a great number of transcriptional coactivators/corepressors to positively or negatively regulate effector genes, so that the interpretation of a signal depends on the cell-type and cross-talk with other signaling pathways.
THE UMD-TGFBR2 DATABASE

The database lists known point mutations, deletions or insertions, and splice mutations in the TGFBR2 gene. Its purpose is to collect in a standardized accessible and summary form the molecular and the clinical data on the causative mutations of Marfan syndrome, and TGFβ signalopathies and cancers. (A TGFBR1 mutation database is currently under construction.)

The present version of the database contains 85 entries corresponding to mutations either published or only reported in meeting proceedings (Supplementary Table S1A and B, available online at http://www.interscience.wiley.com/jpages/1059-7794/suppmat). As in previous editions of the UMD [Beroud et al., 2000, 2005], mutation names are given according to nomenclature guidelines (www.hgvs.org) and numbered with respect to the TGFBR2 gene cDNA sequence (+1 = A of ATG) obtained from GenBank database (GenBank database accession number BC040499; Homo sapiens transforming growth factor, beta receptor II (70/80 kDa), mRNA (cDNA clone IMAGE:5287742). For each mutation, information is provided at several levels: 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 when data are available). In addition, we have annotated the TGFBR2 sequence for highly conserved domains (HCD). These data list for a given position the known arguments such as: transmembrane region, ATP-binding region, proton-acceptor region, posttranslational modification of a residue as N-linked glycosylation or serine phosphorylation (SWISS-PROT accession number P37173) and very conserved amino acids of unknown function (comparison of the HBG054502 gene family).

ANALYSIS OF THE DATABASE

The great majority of the mutations are spread throughout the serine/threonine kinase domains (Fig. 2C). We have included repeat observations of the same mutation except for the mutations located in the polyadenosine tract (BatII region) (c.383delA, c.382_383delAA and c.382dup) corresponding to the hotspot observed in cancers [Markowitz et al., 1995] that are represented once. A total of 36 recurrent mutations have been reported corresponding to 11 mutational events: p.R356P (x2), p.N384S (x2), p.D446N (x2), p.S449F (x2), p.R460C (x3), p.R460H (x6), p.C461Y (x2), p.E526Q (x2), p.R528C (x3), p.R528H (x8), and p.R537C (x4). In these recurrent mutational events, mutations are not disease-specific as the same mutation can be either involved in different genetic diseases, or in cancer and genetic disorder. For example, p.R537C is reported in

![Figure 2. Mutations in the TGFBR2 gene.](https://example.com/figure2.png)

A: Genomic structures of the TGFBR2 gene. The seven exons are represented (squares). Respective position in AA are given. B: Domain organization of TBR2. Position of the different modules are given in AA. C: Reported mutations.
colon cancer [Lu et al., 1995], in MFS [Mizuguchi et al., 2004], and in LDS type I [Loeys et al., 2006]. Today, two recurrent mutations are only involved in LDS type I: p.C461Y and p.R528C [Loeys et al., 2006]. When looking at reported transmissions, 20 somatic mutations are involved in cancers and 17 de novo mutations vs. 16 familial cases are observed.

For 28 mutations, detailed data are not available. Mutations are mainly missense with 78 reported cases among which 46 concern highly conserved amino acids. Two nonsense mutations, p.R495X [Loeys et al., 2006] and p.R497X [Singh et al., 2006], are located in the same region, the cytoplasmic serine/threonine kinase domain DXI. Four splice-site mutations are described. Three concern the canonical splice site sequences (c.95–2A>G, c.1397–2A>G, and c.1397–1G>A) and one is a synonymous amino acid substitution in the last nucleotide of exon 6 resulting in abnormal splicing [Mizuguchi et al., 2004]. For each of the four mutations, insertion or deletion sequences have been experimentally characterized (see Table 1).

The p.T315M variation, which was reported initially in familial nonpolyposis colorectal cancer [Lu et al., 1998], was observed in two patients with MFS and in nine of 497 controls by Ki et al. [2005]. Ki et al. [2005] presumed then that this variation is a rare polymorphism in the Korean population (estimated allele frequency of 0.01). This variation is however considered as probably pathogenic based on chemical properties (Blosum62 [Henikoff and Henikoff, 1992], and biochemical value [Yu, 2001]). In the absence of conclusive data, this variation has been classified as polymorphism. In the same way, sequence variations, c.944C>T [Lu et al., 1998], c.1159G>A [Lucke et al., 2001; Matyas et al., 2006], and c.1159G>C [Matyas et al., 2006], for which no clear argument for their pathogenicity has been reported are considered as polymorphism until more data are collected (see Supplementary Table S2).

The p.T315M variation, which was reported initially in familial nonpolyposis colorectal cancer [Lu et al., 1998], was observed in two patients with MFS and in nine of 497 controls by Ki et al. [2005]. Ki et al. [2005] presumed then that this variation is a rare polymorphism in the Korean population (estimated allele frequency of 0.01). This variation is however considered as probably pathogenic based on chemical properties (Blosum62 [Henikoff and Henikoff, 1992], and biochemical value [Yu, 2001]). In the absence of conclusive data, this variation has been classified as polymorphism. In the same way, sequence variations, c.944C>T [Lu et al., 1998], c.1159G>A [Lucke et al., 2001; Matyas et al., 2006], and c.1159G>C [Matyas et al., 2006], for which no clear argument for their pathogenicity has been reported are considered as polymorphism until more data are collected (see Supplementary Table S2).

The diagnosis of Shprintzen-Goldberg syndrome reported by Kosaki et al. [2006] in a Korean patient is debatable (c.IVS5–2A>G or c.1397–2A>G mutation). This patient fulfills the preliminary guidelines suggested for the diagnosis of Shprintzen-Goldberg syndrome. However, Robinson et al. [2006] argued that the diagnosis of Loeys-Dietz syndrome would also be appropriate for this individual, especially in light of the presence of bifid uvula and the sigmoid configuration of the brachcephalic, left common carotid, and left subclavian arteries. This patient is then reported as having LDS type I.

Further studies will be now necessary to characterize a large number of germline mutations in this gene and determine if the spectrum of clinical features associated with TGFBR2 gene mutations is as large as that of FBN1 gene mutations.

### DATABASE UPDATE

The current database and subsequent updated versions are available at www.umd.be.

Notification of omissions and errors in the current version as well as specific phenotypic data would be gratefully received by the corresponding author (g.colloq-beroud@montp.inserm.fr). The software package is available on a collaborative basis. The software will be expanded as the database grows and according to the requirements of its users. New functions could be implemented. Users of the database must cite this article.

<table>
<thead>
<tr>
<th>Splice site type</th>
<th>Reference</th>
<th>Mutation Reference</th>
<th>Observed Consequences</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>c.IVS1^2A G( c.95^2A G)</td>
<td>Loeys et al. [2005]</td>
<td>Experimentally demonstrated (skipping of nucleotides 95–112)</td>
</tr>
<tr>
<td></td>
<td>c.IVS5^2A G( c.1397^2A G)</td>
<td>Kosaki et al. [2006]</td>
<td>Experimentally demonstrated (inclusion of 30 nt)</td>
</tr>
<tr>
<td></td>
<td>c.IVS5^1G A( c.1397^1G A)</td>
<td>Loeys et al. [2006]</td>
<td>Experimentally demonstrated (inclusion of 30 nt)</td>
</tr>
<tr>
<td></td>
<td>c.1524G A( p.Gin508Gin)</td>
<td>Mizuguchi et al. [2004]</td>
<td>Experimentally demonstrated (inclusion of 23 nt)</td>
</tr>
</tbody>
</table>

*Consensus values for each potential donor or acceptor splice site are calculated for wild type (WT) and mutant sequence according to Shapiro and Senapathy [1987] and Senapathy et al. [1990] (1 000 = strong splice site; 0 = not a splice site).

Observed consequences describes the abnormalities found at the mRNA level.
ACKNOWLEDGMENTS

This work was supported by grants from GIS-Institut des Maladies Rares 2004 and ANR-05PCOD-14-02 (both to C.B. and G.J.). M.Y.F. is supported by a grant from Association Française contre les Myopathies (AFM).

REFERENCES


