



Association of modifiers and other genetic factors explain Marfan syndrome clinical variability

Melodie Aubart^{1,2} · Steven Gazal^{3,4} · Pauline Arnaud^{1,5} · Louise Benarroch¹ · Marie-Sylvie Gross¹ · Julien Buratti^{6,7} · Anne Boland⁸ · Vincent Meyer⁸ · Habib Zouali⁸ · Nadine Hanna⁵ · Olivier Milleron⁹ · Chantal Stheneur⁹ · Thomas Bourgeron^{6,7,10} · Isabelle Desguerre^{2,11} · Marie-Paule Jacob¹ · Laurent Gouya^{10,12} · Emmanuelle Génin¹³ · Jean-François Deleuze⁸ · Guillaume Jondeau^{1,9,10} · Catherine Boileau^{1,5,10}

Received: 4 September 2017 / Revised: 27 March 2018 / Accepted: 11 April 2018
© European Society of Human Genetics 2018

Abstract

Marfan syndrome (MFS) is a rare autosomal dominant connective tissue disorder related to variants in the *FBNI* gene. Prognosis is related to aortic risk of dissection following aneurysm. MFS clinical variability is notable, for age of onset as well as severity and number of clinical manifestations. To identify genetic modifiers, we combined genome-wide approaches in 1070 clinically well-characterized *FBNI* disease-causing variant carriers: (1) an *FBNI* eQTL analysis in 80 fibroblasts of *FBNI* stop variant carriers, (2) a linkage analysis, (3) a kinship matrix association study in 14 clinically concordant and discordant sib-pairs, (4) a genome-wide association study and (5) a whole exome sequencing in 98 extreme phenotype samples.

Three genetic mechanisms of variability were found. A new genotype/phenotype correlation with an excess of loss-of-cysteine variants ($P = 0.004$) in severely affected subjects. A second pathogenic event in another thoracic aortic aneurysm gene or the *COL4A1* gene (known to be involved in cerebral aneurysm) was found in nine individuals. A polygenic model involving at least nine modifier loci (named gMod-M1-9) was observed through cross-mapping of results. Notably, gMod-M2 which co-localizes with *PRKGI*, in which activating variants have already been described in thoracic aortic aneurysm, and gMod-M3 co-localized with a metalloprotease (proteins of extra-cellular matrix regulation) cluster. Our results represent a major advance in understanding the complex genetic architecture of MFS and provide the first steps toward prediction of clinical evolution.

Introduction

Marfan syndrome (MFS) is a rare (prevalence 1/5000) autosomal dominant connective tissue-disorder. The clinical features of MFS may be present in multiple systems (ocular, cardiovascular, skeletal, skin, lung, dura...). Prognosis is related to the life-threatening complications of thoracic aortic aneurysm rupture and dissection (TAAD) [1]. Private variants in the *FBNI* gene encoding fibrillin-1, an extra-

cellular matrix component, are responsible for most cases of MFS (#154700) [2].

MFS clinical variability is notable, even within families, for age of onset as well as severity and number of clinical manifestations. To date, no strong genotype-phenotype correlation between MFS and *FBNI* variants has been reported except for neonatal forms of MFS associated with missense variants in exons 25–33 (according to the NGs, previously 24–32) [3] and truncating variants with a more severe aortic phenotype [4, 5]. However, although a genotype-phenotype correlation may explain interfamilial variability, it cannot account for intrafamilial variability. For intrafamilial variability, the accepted mechanism of variability is mostly the effect of risk alleles in modifier genes and environmental factors. In MFS, the early-onset variability does not support the hypothesis of environmental factors. Identification of genetic modifiers and their alleles remains a

Electronic supplementary material The online version of this article (<https://doi.org/10.1038/s41431-018-0164-9>) contains supplementary material, which is available to authorized users.

✉ Catherine Boileau
catherine.boileau@inserm.fr

Extended author information available on the last page of the article.

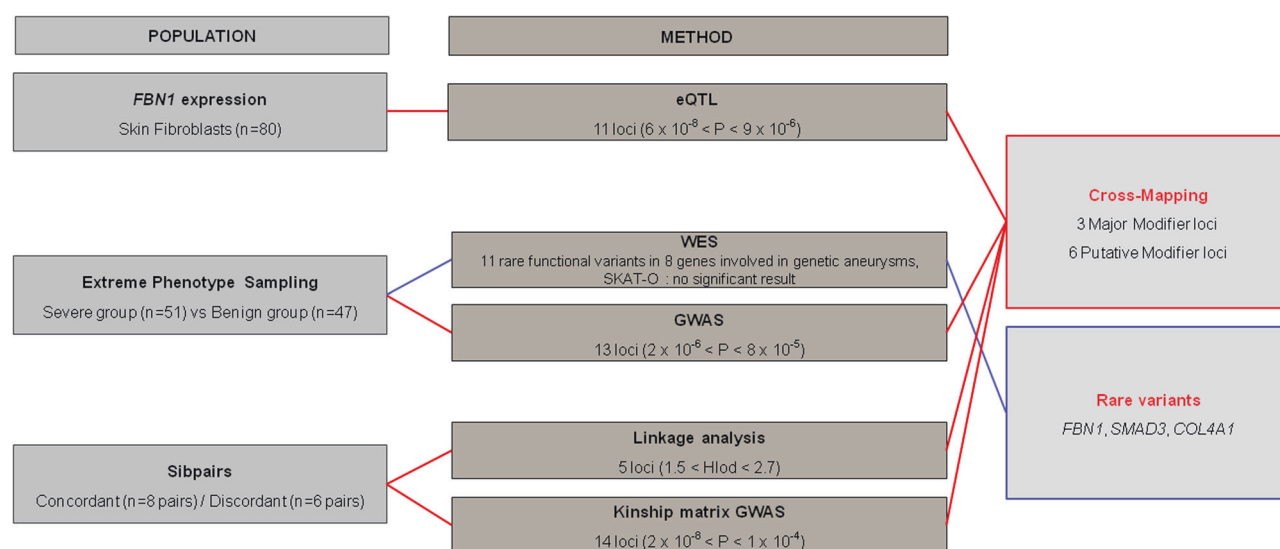


Fig. 1 Study design. Three different samples (skin fibroblasts, EPS and sibpairs) were studied by five different approaches (eQTL, WES, GWAS, linkage analysis, kinship matrix GWAS). Cross mapping of

results from the five approaches revealed modifier loci while WES showed rare coding variants in known TAA genes

challenge as in others rare diseases. Indeed, despite the availability of powerful genome-wide tools, the sample size required to achieve usual statistical significance for discovery and replication stages of a study remains unattainable for classic study designs even through worldwide collaborations in rare diseases. Moreover, the ignorance of the genetic mechanisms underlying phenotype variability (type of disease-initiating variant, gender factor, co-occurrence of rare variants, modifier genes) precludes the a priori design of an optimal research strategy. Building on a collection of MFS samples, we show how the combination of various genome-wide approaches and their joint analysis within a cross-mapping framework can provide interesting insight into mechanisms underlying MFS clinical variability.

Indeed, we demonstrated recently that the level in skin fibroblasts of residual *FBN1* expression in MFS patients with premature-truncating codon variant was significantly associated with some features of the disease [6]. To further identify genetic modifiers, we performed in this work, (1) an expression Quantitative Trait Loci (eQTL) study on skin fibroblasts of MFS patients with premature truncating codon (PTC) variants to identify genetic regulators of *FBN1* expression which could be genetic modifiers of the phenotype, (2) a linkage and association analysis in discordant and concordant MFS because family-based studies enable resistance to the effects of population structure and do not need any prior hypothesis on the frequency of the modifier variants, (3) Whole-Exome sequencing (WES) with a SKAT-O analysis to identify accumulation of different rare coding variants in presumed frequent modifier genes and finally (4) a genome-wide association study (GWAS) to screen for coding or non coding frequent variants in extreme

aortic phenotype of MFS sampling (EPS) which provides enhanced study power (Fig. 1). Results of these analyses were crossed over and enabled a cross-mapping framework of three major and six putative modifier loci, some of which co-localized with excellent candidate genes.

Subjects and methods

Subjects

All subjects originated from the French National Reference Centre for Marfan Syndrome and related disorders and gave their written informed consent for participation in this clinical and genetic study in agreement with the requirements of French regulations. To ensure homogeneity, all retrospective clinical data were reassessed by two physicians (MA and GJ) using the Ghent criteria [7, 8]. Patients were evaluated by geneticists, rheumatologists, cardiologists and ophthalmologists. Systematic slit-lamp examination, cardiac ultrasonography, radiological investigations and molecular analyses were also performed. Clinical features used for clinical severity evaluation were as listed in the Ghent nosology (20) with aortic diameters measured at the level of the sinuses of Valsalva according to Roman [9]. All patients included were adult (>18 years-old) *FBN1* variant-carriers. Missense variants in exons 25–33 were excluded since they are correlated with severe forms of MFS. Patients for the eQTL study had a PTC *FBN1* variant [6].

For EPS, two groups of unrelated patients were considered out of the cohort of 1070 *FBN1* pathogenic variant carriers [557 women (52.1%) and 513 men (47.9%)]: the severe

group (S group) formed by the 5% ($n = 54$) of the collection with the most severe aortic features and the benign group (B group) formed by the 5% ($n = 54$) of the collection with the most benign aortic features, while preserving a sex ratio of 1. The S group was defined as individuals presenting with a dissection or preventive TAA surgery (for a diameter > 50 mm at the Valsalva level [10]) at a young age. The B group was defined as absence of TAA as late as possible in life (no surgery and aortic diameter lower than 2 SDs). Geographic origin of patients should be European and was assessed by a principal component analysis (PCA) reported to 1000 Genomes database [11] (Figure S1).

For the sib-pairs study, individuals were same-sex siblings and at least one had aortic criteria corresponding to the S group and the other aortic criteria either of the S (concordant sib-pairs) or of the B (discordant sib-pairs) group. Clinical and molecular comparisons between groups were performed using a standard chi squared test.

Gene expression

Cell culture conditions and RNA extraction, RT-PCR and Real-time quantitative PCR methods have been described [6] and are in agreement with the minimum information for the publication of quantitative real-time PCR experiments guidelines [12].

Genotyping and quality control

Genotyping of the 80 individuals of the eQTL study was performed using the HumanOmniExpress-24v1.0 array (713,014 SNPs) of Illumina (San Diego, CA, USA). This array is referred as genotyping array 1 in the results section. All SNPs with more than 5% of missing genotypes were removed and SNPs with a MAF > 5% were kept leading to 594,978 SNPs retained. PCA was performed with PLINK and the two first components were used as covariates (Figure S2) [13].

Genotyping of the 102 individuals of the EPS and the 28 individuals of the sib-pairs was performed using the HumanOmniExpressExome-8v1.2 array (964,192 SNPs) of Illumina (San Diego, CA, USA) by the Centre National de Génotypage (CNG, Evry, France). This array is referred as genotyping array 2 in the results section. Using the same criteria as for array 1, 701,570 SNPs were retained. Relatedness among the whole samples was assessed using—genome option of PLINK version 1.90b2b. In the absence of a control group, no Hardy-Weinberg equilibrium test was performed.

Association analyses

A linear regression on the 80 individuals of the eQTL study was conducted using the—linear option of PLINK and the 2

first principal component of PCA as covariables. Fibroblast mRNA levels were quantified as reported [6]. They were combined with genome-wide genotyping data to conduct the eQTL analysis. Single-marker association analysis was performed between the 2 EPS (S and B) using the --assoc option of PLINK under an additive genetic model. To increase the power of the study, the standard linear mixed model implemented in GEMMA software version 0.94.1 [14] was also used for individuals of EPS and of sib-pairs.

To test the association of un-genotyped SNPs, all SNPs of the phase 3 of the 1000 Genomes Project were imputed from their available haplotypes [software IMPUTE2 [15], while pre phasing of data used software SHAPEIT2 [16]]. All SNPs with a low imputation quality (IMPUTE info metric below 0.8) or with a MAF < 0.05 (in the 102 individuals of the EPS) were removed. Association of marker dosage was conducted with SNPtest version 2.5.2 [17] using the two first components of the PCA as covariates. CNV genotypes were inferred using CNVnator [18] merging CNV detected by QuantiSNP version 2 [19] and PennCNV version 1.0.1 [20]. The impact of CNV was then investigated through an association as reported by Gamazon et al. [21]. Finally, genome-wide significance was inferred for each analysis with a Bonferroni correction.

Submission of the data has been done to the GWAS database (NIH) (www.ebi.ac.uk/gwas/).

Linkage analysis

Parametric linkage analyses were performed on genotyped sib-pairs using Merlin version 1.1.2 [22]. PLINK option—inde-pairwise 50 5 0.1 was performed to minimize linkage disequilibrium (LD) between SNPs in the EPS samples. This led to the definition of a set of 42,024 SNPs and their allele frequencies determined in the sample. The genetic model tested was autosomal dominant with complete penetrance or 0.9 of penetrance and no phenocopy. Disease allele frequency was set at 0.01 (and 0.1, 0.001 or 0.00001, similar results, data not shown). To take into account possible genetic heterogeneity, linkage was determined using the heterogeneity Lod (HLod) score implemented in Merlin. The highest likelihood ratio that could be obtained with this sample was approximately HLod 4.2 in a context of genetic homogeneity and if all sib-pairs were informative. Since these conditions were unlikely to be met, only the regions that provided the highest HLod were retained. Fine mapping of linked regions was performed through an association studies on the sibs population. Relatedness between individuals was taken into account using the standard linear mixed model implemented in GEMMA. Kinship matrix was estimated through the—genome option of PLINK.

Whole-exome sequencing (WES)

WES of all cases was performed by the French National Genotyping Centre (CNG). Agilent SureSelect Human All Exon Kits V5 (Agilent Technologies, Santa Clara, CA, USA) were used with PE Illumina HiSeq2000 (Illumina, San Diego, CA, USA) sequencing. Bio-informatics analyses were performed following Genome Analysis Toolkit (GATK) best practices [23, 24]. Variants were annotated, filtered and analyzed using Variant tools [25] and ANNOVAR [26]; allele frequencies were from Exome Aggregation Consortium (ExAC) database. To test the association between rare functional variants and EPS, a SKAT-O (optimal Sequence Kernel Association Test) was used [27] on all functional variants.

The reference sequences used for variant descriptions are: *FBN1* (NG_008805.2, NM_000138.4), *COL4A1* (NG_011544.2, NM_001845.5), *COL3A1* (NG_007404.1, NM_000090.3), *SMAD3* (NG_011990.1, NM_005902.3), *SKI* (NG_013084.1, NM_003036.3), *TGFBR2* (NG_007490.1, NM_003242.5). Exon numbering is according to the NGs.

Results

eQTL of *FBN1* expression in skin fibroblasts

The eQTL analysis was performed to identify only *trans*-eQTL loci. Indeed no haplotype phasing was possible between tested markers and *FBN1* disease-causing variants in isolated patients. eQTL analysis performed in 80 subjects with PTC *FBN1* variants lead to the identification of 16 SNPs with $P < 1 \times 10^{-5}$ distributed among 11 loci (Fig. 2 and Table S1, Figure S3). One SNP on chromosome 11 reached genome-wide significance (P value $< 8 \times 10^{-8}$) with a P -value of 6×10^{-8} . This *trans*-eQTL locus was defined by four SNPs in LD at 11q22.3 (Fig. 2) and is further described below. Interestingly, at the *FBN1* locus (15q21.1) four SNPs, in partial LD, were associated with *FBN1* expression with P -values between 1×10^{-5} and 3×10^{-5} . These SNPs were located 1.5–2 Mb downstream (transcription direction) from the *FBN1* gene.

EPS study

Clinical characteristics and new *FBN1* phenotype/genotype correlation

The S (severe) group comprised 27 males meeting the criteria with ages at surgery between 4 and 25 y.o. and 27 females with ages at surgery from 15 to 33 y.o. The B (benign) group comprised 27 females without surgery

and with aortic diameter < 2 SD with ages between 53 and 79 y.o. and 27 males with ages between 22 and 75 y.o. However, since the youngest of these 27 males were very young, they could possibly no longer meet the diagnostic criteria as they grew older. Therefore, we chose to add additional criterion for men in the B group: Age ≥ 27 y.o. In this way, 21 males with ages between 27 and 75 y.o met the criteria of complete inclusion. After complete genotyping, PCA identified 4 geographic outliers who were excluded thus leading to 51 and 47 subjects for the S and B group, respectively. No difference was observed in MFS non-aortic features of between the two groups (Table S2). Molecular data showed a significant difference of variant types between the two groups (55% PTC and 43% missense variants in the S group, 36 and 64%, respectively, in the B group, $p = 0.05$) (Table 1). Confirming this genotype-phenotype correlation, we found a difference among the missense variants with an excess of loss-of-cysteine variants in the S group (63 vs 20%, $p = 0.004$). Conversely, we observed an excess of cysteine-forming variants in the B group (23 vs 0%, $p = 0.04$). However, we found no difference between the two groups in the exons where the missense variants were located.

Rare coding variants

WES was performed following a preliminary in silico simulation study that provided the following results: Under the hypothesis of the existence of 3 to 10 modifier genes explaining phenotypic variance, a sample of 51 patients would allow observing between 50 to 100% of the modifier genes carrying more than 5 variants among these genes, while only 2 non modifier genes are expected to reach this criteria by chance (Figure S4).

WES in the 51 patients from the S group and 47 patients of the B group identified around 50,000 non-synonymous exonic and splicing variants. No gene was identified with an excess or a lack of variants associated with phenotype severity (Figure S5). Indeed, SKAT-O provided the lowest P -value (1.4×10^{-4}) for *TUBGCP2* (encoding an ubiquitous protein of the gamma-tubulin complex).

In a second step, we performed classic investigation (single-gene tests) on the WES data. No significant variant was observed in the B group. Conversely, in the S group, careful examination of very rare filtered variants identified nine variants of unknown significance (VUS) in nine patients within six genes known to be involved in TAA or other arterial aneurysm (Table 2). Four were considered as non relevant: 2 variants in *COL3A1* (#130050) because (1) they concerned weakly conserved amino acids and the Grantham distances were very low and (2) no familial co-segregation with the severity of the aortic phenotype

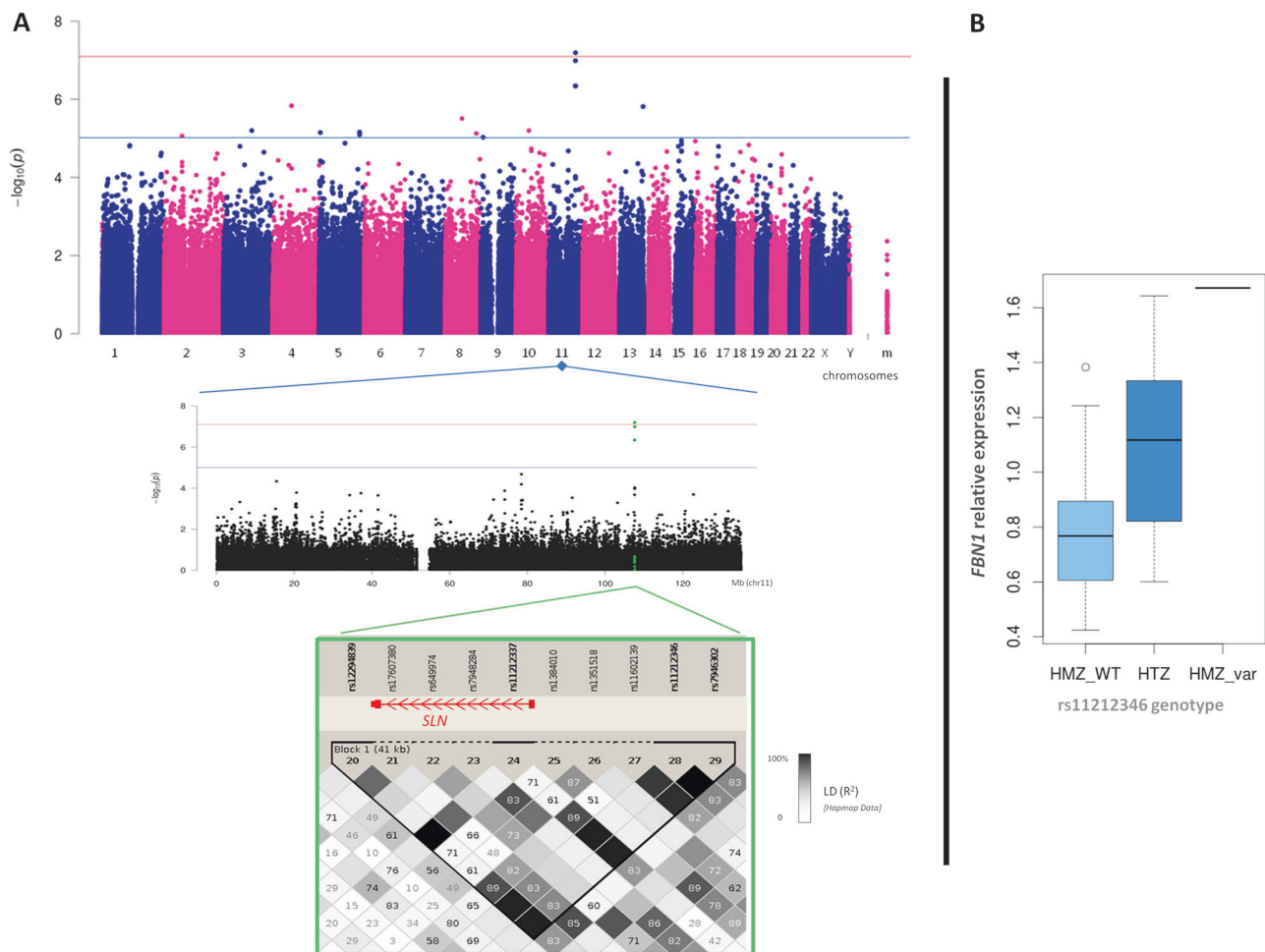


Fig. 2 *FBNI* eQTL Results, focus on chromosome 11. **a** *FBNI* expression was studied in skin fibroblasts of 80 *FBNI* variant-carriers and associated with their genotyping array results in an eQTL study. On the Manhattan Plot, red line represents the significant threshold at 8×10^{-8} , blue line represents P -value 1×10^{-5} . One SNP had significant P -value (rs11212346), with three other SNPs (rs12294839, rs11212330, rs7946302) in linkage disequilibrium (LD). LD is

determined according to R^2 in Haploview (Black squares correspond to $R^2 = 100\%$). **b** Expression of *FBNI* in skin fibroblasts of *FBNI* variant-carriers according to rs11212346 genotype (boxplot: box from first to third quartile shift by the median, extreme lines showing the highest and lowest value, except if they are superior to a default range of 1.5 interquartile)

could be observed; 1 variant in *SKI* (#182212) because it was not located in the Smad-binding domain of *SKI* which is the hot-spot for reported disease-causing variants and because the corresponding codon has a weak coverage in ExAC database (less than 20% of individuals with a coverage > 10) so the rare frequency of the variant cannot be assessed properly and finally 1 variant in *TGFBR2* (#610168) which was considered as benign by Polyphen2 and for which familial segregation showed no cardiovascular features in the variant carriers.

The remaining five variants were considered as relevant: c.1286 G \rightarrow C (p.(Arg429Pro)) in *FBNI*, c.304 G \rightarrow A (p.(Glu102Lys)) in *SMAD3* (#613795), c.1588 C \rightarrow T (p.(Pro530Ser)), c.329 T \rightarrow C (p.(Ile110Thr)) and c.3164 C \rightarrow T (p.(Pro1055Leu)) in *COL4A1* (#180000, #611773, #607595, #175780) (Table 1). The *FBNI* variant is absent

from all databases and family investigation was performed to determine the location of this event *trans* of the disease-causing variant and allowed to diagnose compound heterozygosity. For the *SMAD3* variant, the proband also carried a *de novo FBNI* variant while several family members without this *FBNI* variant also carried the *SMAD3* variant. Some of them had histories of TAA or early-onset osteoarthritis, allowing retrospective diagnosis of aneurysm-osteoarthritis syndrome related to *SMAD3* disease-causing variant in this family in addition to MFS related to a *FBNI* variant in the severe phenotype proband (Fig. 3a). Among the 3 variants in *COL4A1*, none affected a glycine residue as classically reported in the different hereditary angiopathies associated with *COL4A1* variants. The c.3164 C \rightarrow T (p.(Pro1055Leu)) and c.329 T \rightarrow C (p.(Ile110Thr)) in *COL4A1* variant carriers also carried *de novo FBNI* variants and no familial vascular

Table 1 FBN1 variants in the two groups of the EPS study

	S (severe phenotype)		B (benign phenotype)		Chi squared
	<i>n</i>	%	<i>n</i>	%	<i>P</i> -value
Truncating variants (PTC)	28	55	17	36	0.05
Large deletion	3	11% of PTC variants	3	18% of PTC variants	NS
Splice variant	7	25% of PTC variants	2	12% of PTC variants	NS
Nonsense	11	39% of PTC variants	10	59% of PTC variants	NS
Frameshift	7	25% of PTC variants	2	12% of PTC variants	NS
Insertion/deletion without frameshift	1	2	0	0	NS
Missense variants	22	43	30	64	NS
Lost of cystein	14	63% of missense variants	6	20% of missense variants	0.004
Cystein formation	0	0% of missense variants	7	23% of missense variants	0.04
If not cystein, consensus/not consensus aminoacid	4/4	18/18% of missense variants	7/10	23/33% of missense variants	NS
Exons concerned	3 à 63	–	1 à 63	–	
Missense variant in signal peptide	0	0% of missense variants	1	3% of missense variants	NS
Missense variant in hybrid domains	2	9% of missense variants	4	13% of missense variants	NS
Missense variant in TGFB binding domains	3	14% of missense variants	3	10% of missense variants	NS
Missense variant in calcium binding (like) domains	17	77% of missense variants	22	73% of missense variants	NS

The bold values in table 1 correspond to main titles = main lines comparing to others which are subtitles (and sub percentages) of these main lines

features were reported. For the third, and most frequent *COL4A1* variant c.1588 C>T (p.(Pro530Ser)), family investigation showed it was present in a total of four MFS patients, two children with a TAA and two adults who underwent preventive surgery for TAA (at 17 y.o. in the man, 43 y.o. in the woman) (Fig. 3b). Four other family members carried the *FBN1* variant but not the *COL4A1* variant and none of these had TAA.

GWAS in EPS groups

Genotyping with array 2 followed by quality control led to the study of 701,570 SNPs with the Bonferroni-adjusted significance threshold set at $P < 7 \times 10^{-8}$. No SNP reached this value but 22 SNPs tagging 14 different LD blocks reached $P < 10^{-4}$ (Figure S6 and Table S4). Three tagged regions included candidate genes (*COL16A1*, *SMAD6*, and *MEF2C*). No significant results were obtained when association studies were performed with frequent variants typed by WES, or with CNVs, or using imputation with 1000 Genomes SNPs.

Concordant/discordant sib-pairs study

Linkage analysis for severe aortic disease was performed in eight severe concordant and six discordant same-sex sib-pairs. Nine sib-pairs were females (five concordant and four discordant pairs) and five were males (respectively three and two). Assuming an autosomal dominant model with a modifier allele frequency of 0.01 and complete penetrance

of the trait, Hlod_s > 1.5 were found in five chromosomal regions (Table S4). The use of lower allele frequencies or of lower penetrance did not significantly change the results (data not shown). The lead locus had a Hlod of 2.7 (ALPHA = 1) and was located on chromosome 10. This locus comprises 10 genes among which *PRKG1*, a gene encoding a cGMP-dependent protein kinase strongly expressed in the aortic wall. Interestingly, missense variants in this gene are known to be involved in familial TAAD (#615436) [28]. WES was also performed in the sib-pairs and no rare functional variant was found in either of the five genomic regions of interest. Subsequently, an association study was performed in these 28 subjects using array 2 data and a mixed model to account for the kinship matrix. Seventeen SNPs representing 14 loci showed *P*-values between 1.5×10^{-8} and 1×10^{-4} (Table S5). The lowest *P*-value was obtained at 7q21.11 with a SNP located within intron 1 of *MAGI2*, encoding a protein that serves as a scaffold for the assembly of synaptic protein complexes but with GO annotations related to this gene including SMAD binding.

Cross-mapping of the results

Convergence of results between the different studies led to the identification of 3 modifier regions (gMod-M1 to 3) and 6 putative modifier regions (gMod-M4 to 9) (Figs. 4, 5 and Figure S7). gMod-M1 is located at 1p36.12 (Fig. 5a) in a region defined by 3 SNPs in LD with a *P*-value of 4×10^{-5} (GWAS EPS study), a Hlod of 1.7 (sib-pair linkage

Table 2 Results of WES in EPS

Subject	EPS group	cDNA FBN1 variant ID (NM_000138.4)	Protein FBN1 variant ID	Familial transmission of MFS	Gene of the VUS ^a	Disease related to variant in the gene	Transcript ID	cDNA VUS ^a ID	Protein VUS ^a ID	ExAC MAF	UMD LSDB ^b
1	SEVERE	c.4459 + 1 G > C	p.?	sporadic	<i>COL3A1</i>	Vascular Ehler-Danlos syndrome	NM_000090.3	c.3226 C > G	p.(Pro1076Ala)	0%	NA
2		c.2626 T > C	p.(Cys876Arg)	familial	<i>COL3A1</i>	Vascular Ehler-Danlos syndrome	NM_000090.3	c.3818 A > G	p.(Lys1273Arg)	0.00008%	NA
3		c.4615 C > T	p.(Arg1539 ^a)	sporadic	<i>COL4A1</i>	Hereditary angiopathy	NM_001845.5	c.329 T > C	p.(Ile110Thr)	0.02%	NA
4		c.1630 G > C	p.(Gly544Arg)	familial	<i>COL4A1</i>	Hereditary angiopathy	NM_001845.5	c.1588 C > T	p.(Pro530Ser)	0.040%	NA
5		c.4453 T > C	p.(Cys1485Arg)	sporadic	<i>COL4A1</i>	Hereditary angiopathy	NM_001845.5	c.3164 C > T	p.(Pro1055Leu)	0.003%	NA
6		c.1585 C > T	p.(Arg529 ^a)	sporadic	<i>FBN1</i>	Marfan syndrome	NM_000138.4	c.1286 G > C	p.(Arg429Pro)	0%	0
7		c.4817-2 A > C	p.?	sporadic	<i>SKI</i>	Sphrintzen Goldberg syndrome	NM_003036.3	c.1505 C > T	p.(Pro502Leu)	0% (poor coverage)	NA
8		c.6424 T > C	p.(Cys2142Arg)	sporadic	<i>SMAD3</i>	TAAD and osteoarthritis	NM_005902.3	c.304 G > A	p.(Glu102Lys)	0%	0
9		c.(2293 + 1_2294-1)_(2854 + 1_2855-1)del	p.?	sporadic	<i>TGFBR2</i>	TAAD / Marfan syndrome / Loeys-Dietz syndrome	NM_003242.5	c.453 A > C	p.(Glu151Asp)	0%	0
10	BENIGN	c.7583 G > T	p.(Cys2528Phe)	familial	<i>TGFBR2</i>	TAAD	NM_003238.3	c.893 G > A	p.(Arg298Gln)	0.002%	NA
11		c.5339 G > C	p.(Gly1780Ala)	familial	<i>TGFBR1</i>	TAAD /Marfan syndrome/Loeys-Dietz syndrome	NM_004612.3	c.470 G > C	p.(Arg157Pro)	0.020%	0

^aVUS: Variant of unknown significance

^bLSDB: Locus specific database

^cGrantham score: 0–215

^dB benign, D damaging, T Tolerated, NA not available; candidate genes discussed in the results are in red

The reference sequences used for variant descriptions are: FBN1 (NG_008805.2, NM_000138), COL4A1 (NG_011544.2, NM_001845), COL3A1 (NG_007404.1, NM_000090), SMAD3 (NG_011990.1, NM_005902), SKI (NG_013084.1, NM_003036), TGFBR2 (NG_007490.1, NM_003242)

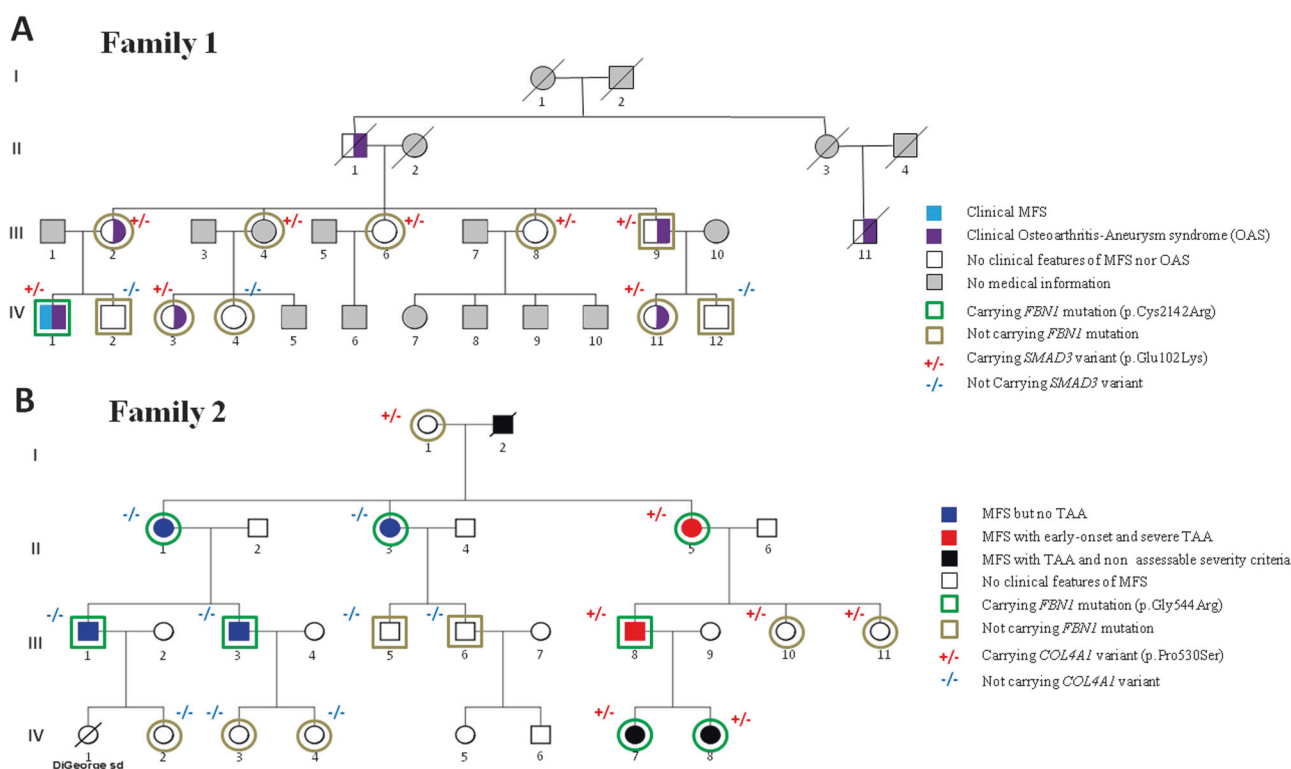


Fig. 3 Examples of segregation of rare modifier variants. **a** Segregation of a *SMAD3* rare variant in a family with a *FBN1* de novo variant-carrier. Individual III-1 is the proband of the family with a clinical diagnosis of MFS (bilateral ectopia lentis with surgery at 13 y.o., severe and early-onset TAA with preventive surgery at 16 y.o., MFS facial features, cleft palate, pectus carinatum, arachnodactyly, dolichostenomelia, hypermobility with Beighton score of 9/9, pes planus, 30° scoliosis, spondylolisthesis, dural ectasia, striae) and osteoarthritis features (early-onset osteophytosis at 30 y.o. and genu valgum surgery at 10 y.o.). His mother, individual III-2, has no ectopia lentis, no TAA but has severe lumbar and shoulder osteo-arthritis and medical history of disc surgery. Sudden death of unknown origin occurred at 55 y.o. in individual II-1. Individual III-4 is an obligate carrier with no clinical information available but her daughter (IV-3) has pectus excavatum, scoliosis less than 20°, hand (fingers and wrist) pain, medical history of Scheuermann disease but no TAA. Individual III-6 has a medical history of surgery for hernia, hypermobility, cleft palate and striae. Individual II-7 has wrist and leg pains without MFS symptoms. Abdominal aortic aneurysm surgery has been performed in individual III-9 at 50 y.o. This individual has also a TAA with an aortic root

measured at 49 mm at Valsalva level by CT-scan and cardiac ultrasound (+3DS), tortuosity of carotid and vertebral arteries, pectus carinatum, spinal osteoarthritis, lumbar scoliosis measured at 22°, toe and ankle surgery. This individual has cardiovascular risk factors (smoking and hypercholesterolemia). His daughter (IV-11) has no TAA but a mitral valve prolapse, a pectus excavatum, a 22° scoliosis and medical history of multiple sprains (knee and ankle). Individual III-11 died suddenly at 20 y.o. without diagnosis. **b** Segregation of a *COL4A1* rare variant in a family carrying a *FBN1* variant. Individual II-5 had preventive aortic surgery at diagnosis of MFS (43 years-old). Individual III-8 had preventive aortic surgery at 17 years-old. Individuals IV-7 (7 y.o.) and IV-8 (4 y.o.) had aortic roots diameters at 3.1 and 2.7 SD, respectively. Individuals II-1, II-3, III-1, III-3 carry the *FBN1* pathogenic variant but not the *COL4A1* variant and have clinical features of MFS but no TAA (aortic root diameter <2 SD). Individuals I-1, III-10 and III-11 (carrying the *COL4A1* variant but not the *FBN1* variant) have no features of MFS and no history of cerebral haemorrhage, but no cerebral investigations have been performed. Individual I-2 had a clinical diagnosis of MFS and died at 44 years-old from internal haemorrhage

analysis) and 7 other SNPs in this region with a P -value of 5×10^{-5} (association-kinship matrix study in sib-pairs). The only candidate gene within the interval was *ECE1*. The second and third regions were mapped both in eQTL and sib-pair studies. gMod-M2 is located at 10q11.2.chromosome 10 (Fig. 5b). At this locus, the highest HLod [2.7 (ALPHA = 1)] was obtained with rs1937662C > A. Interestingly, this SNP was also associated with *FBN1* gene expression in the eQTL study with a P -value of 6×10^{-6} . The region includes the *PRKG1*. gMod-M3 is located at 11q22.3 (Fig. 5c). At this locus, (1) the eQTL study showed a SNP (rs11212346C > T) significantly associated with

FBN1 expression ($P = 6 \times 10^{-8}$), (2) sib-pair linkage analysis led to a HLod of 1.8 (ALPHA = 1) defining a region between 103,500,000 and 108,000,000 and (3). The association-kinship matrix study in sib-pairs revealed 2 SNPs (rs12277780C > A and rs11601883A > G) with a P -values of 9.0×10^{-7} and 2.3×10^{-5} , respectively. The sentinel SNP of the eQTL study is located in a LD-block containing only one expressed gene, *SLN*, which encodes sarcolipin. this small sarcoplasmic protein which is a negative regulator of Ca^{++} ATPase is neither expressed in target tissues nor known to be associated with known physiologically relevant pathways. However, the regional

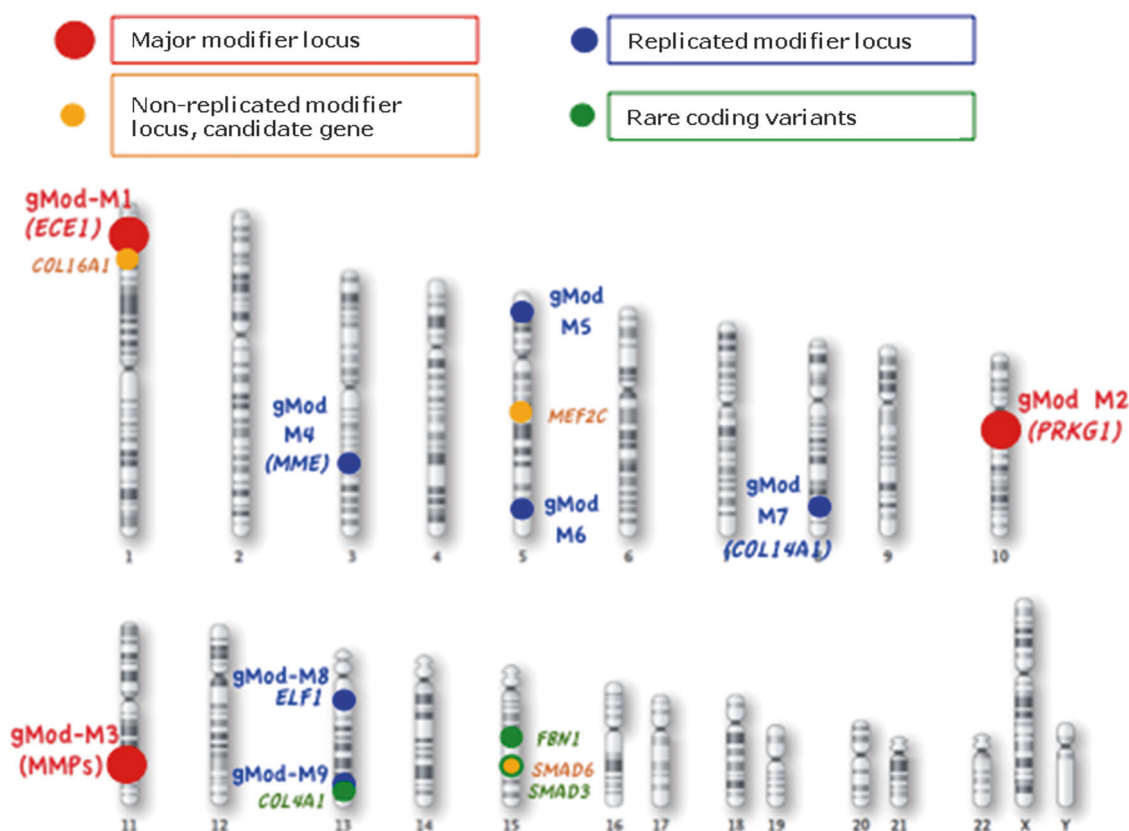


Fig. 4 Graphic representation of crossmapping results. Crossmapped modifier loci are overlaid on a karyotype cartoon. In red are represented the three major modifier loci (crossmapping of three analysis or of two analyses and strong arguments in literature), in blue the six

putative modifier loci (crossmapping in two analyses and no good regional candidate), in yellow excellent candidate genes in a region found in only one analysis, in green genes in which convincing rare coding variants have been found in EPS

linkage peak is close to a cluster of matrix metalloprotease genes (*MMP1*, 3, 7, 8, 10, 12, 13, 20, 27). Finally, 6 putative modifier regions were defined by converging results from only 2 genome-wide studies (gMod-M4 to 9) (Fig S6).

Discussion

Using a large population of *FBN1* variant-carriers with consistent clinical evaluations and extensive follow-up, we combined genome-wide approaches to perform a cross-mapping analysis.

First of all, the EPS groups revealed that truncating variants were more common in the severe group (55 vs 36%, $P=0.05$) as were missense variants with loss of a cysteine residue (63% of the missense variants vs 20%, $P=0.004$). Missense variants creating a cysteine were only present in patients of the benign group (Table 1). The largest previous study of genotype/phenotype correlation found no correlation for aortic phenotype (Faivre et al. 2007, 1013 probands [29]) except for the neonatal form of MFS associated with missense variants in exons 25–33. Patients carrying variants

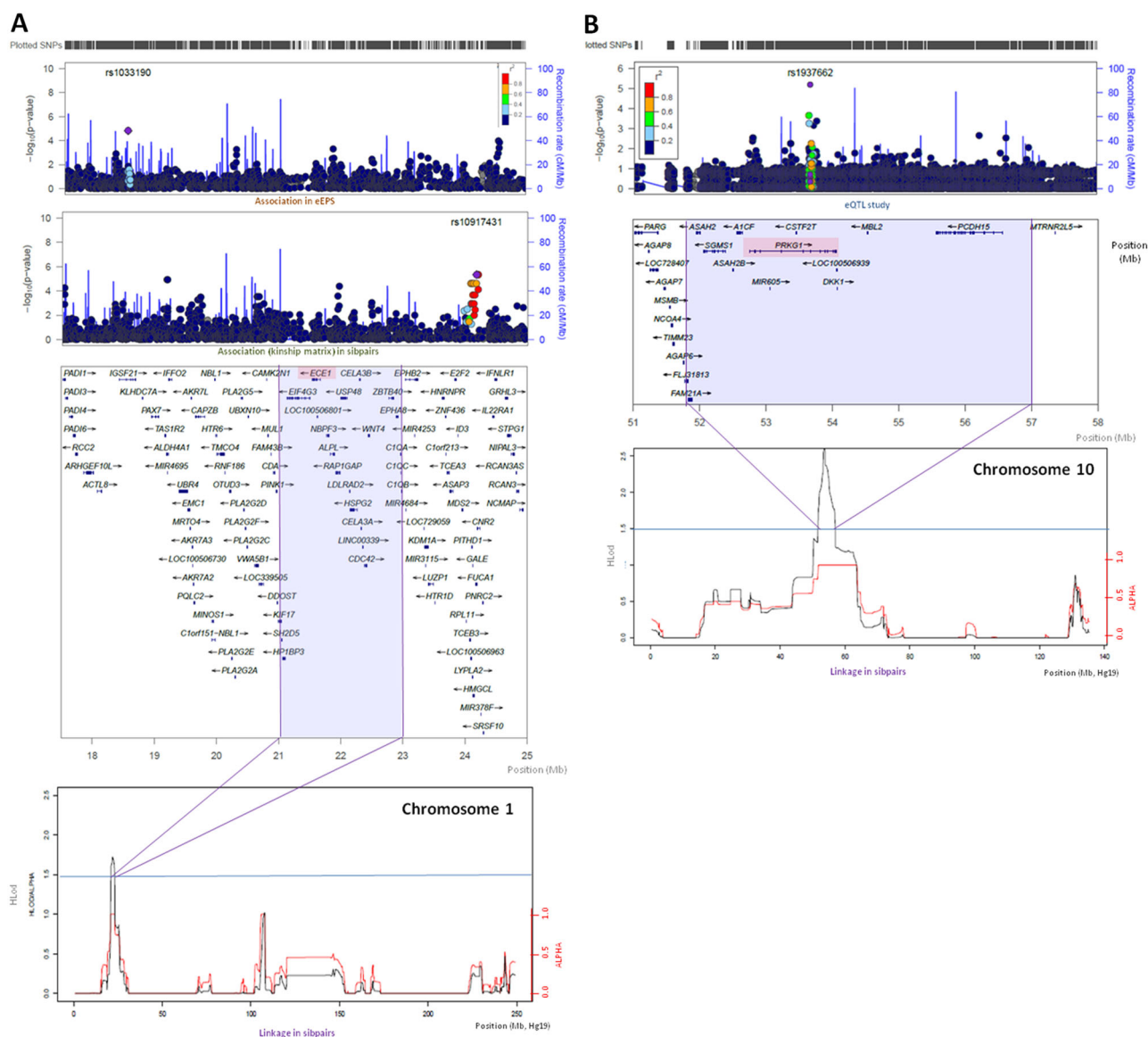
in this region were excluded from our study because the well recognized genotype-phenotype correlation was considered as sufficient to explain phenotype severity in these patients. This would enable us to identify other and less obvious genotype-phenotype correlations that we indeed reported. In this way, more recently, smaller studies [207 and 357 patients in [4] and [5]] provided clues for an association between truncating variants and severe and early-onset aortic dilatation and dissection. Finally, although we cannot exclude that this exclusion could have somewhat biased our results, the number of patients excluded (36/1070) is so small that this is unlikely.

Another genetic mechanism of MFS TAA variability revealed by the EPS study is the possible co-occurrence with the disease-initiating *FBN1* variant of a pathogenic variant in another aneurysm gene. This second event has been found in 5 out of 47 severe patients from the S group and not in patients from the B group. Two cases with severe aortic phenotype carried a second disease event in a known TAA gene: one case was a compound heterozygote with a second variant in the *FBN1* gene [30] and the other case was a double heterozygote carrying a second variant in the *SMAD3* gene. The three other patients carried

a second event in *COL4A1*. This gene encodes a collagen with major expression in the arterial wall including aorta [31]. Furthermore, interactions of type IV collagen with the TGF-beta pathway, central to pathogenic MFS aortic alterations, have been described [32, 33]. Disease-causing variants in the *COL4A1* gene usually affect glycine residues and have been involved in porencephaly or recurrent hemorrhagic stroke and cerebral aneurysm. The three rare modifier variants we report did not affect a glycine and were not associated with cerebral vascular disease in family members but co-segregated with aortic severity as observed for *COL4A1* c.1588 C > T (p.(Pro530Ser)) investigated in a large family (Fig. 3). The existence of rare variants in other potential modifier genes was not supported when combining genome-wide SNP data and WES results in EPS groups,

thus demonstrating that this genetic mechanism of variability is not frequent.

Modifying effects of frequent alleles were investigated through several genome-wide strategies followed by cross-mapping of results. The genome significance threshold was only reached in the eQTL study for the locus on chromosome 11 (gMod-M3). Cross mapping confirmed this locus and revealed two additional modifier loci (gMod-M2 and gMod-M3) (Fig. 5). gMod-M1, located on chromosome 1 in a region containing the *ECE1* gene, encoding the endothelin-converting enzyme from a subfamily of metalloproteases. Endothelin-1 is activated by ECE-1 and has been shown to promote endothelial-to-mesenchymal transition and myofibroblasts resistance to apoptosis and therefore contribute to ECM regulation and fibrotic



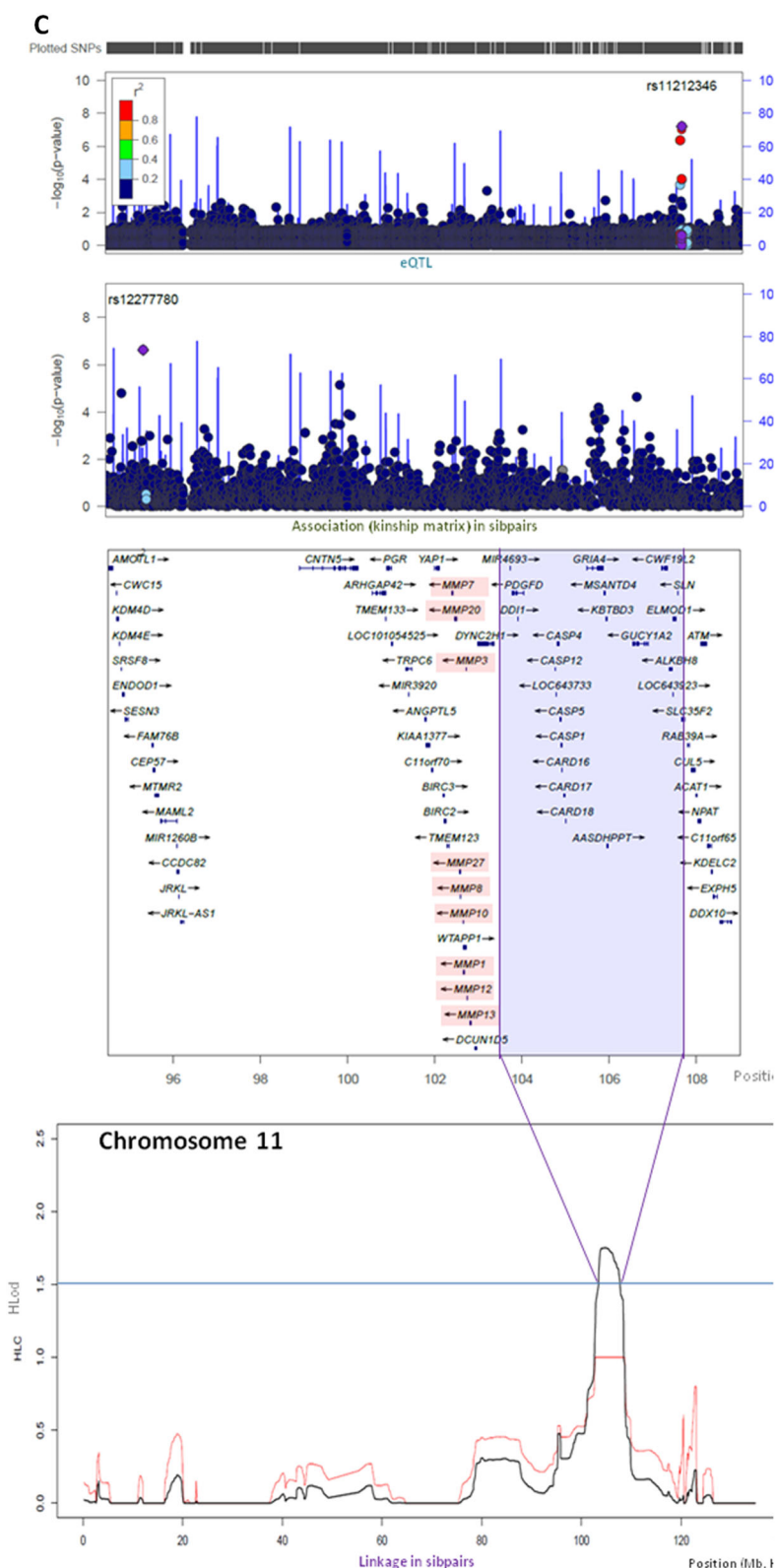


Fig. 5 Modifier regions identified by cross-mapping of genome-wide strategies. **a** gMod-M1; **b** gMod-M2; **c** gMod-M3. For each figure, upper panel shows regional results for eQTL and/or association studies from LocusZoom (<http://locuszoom.sph.umich.edu>). Genome build and LD population are from Hg19/1000Genomes (Nov 2014 EUR). Each point represents the nominal P -value (left y axis). SNPs are colored according to their pairwise correlation (r^2) with the major regional SNP (rs with purple circle). SNPs with missing LD information are shown in grey. Overlaid are recombination rates (right y axis) represented as blue lines. Middle panel provides location of regional genes with respect to upper panel coordinates. The best candidate gene(s) are highlighted. Lower panel shows results of regional linkage analysis in sibpairs: black curve represents hLod-score (left y axis), blue line represents the Hlod 1.5 threshold; red curve represents ALPHA (heterogeneity) (right y axis)

response [34]. Moreover, expression of *ECE1* is high in the aortic wall supporting a role in modification of the aortic phenotype in MFS [31]. gMod-M2, located on chromosome 10 (Fig. 5), contains *PRKG1*, in which gain-of-function variants cause TAA [28]. It encodes a type I cGMP-dependent protein kinase (PKG-1), which is activated upon binding of cGMP and controls SMC relaxation [28]. gModM3 on chromosome 11 defines a region containing the *SLN* gene encoding sarcolipin, a sarcoplasmic protein weakly expressed in skin fibroblasts and in the aortic wall but playing a role in calcium regulation. This locus is statistically the strongest one because it is the result of the cross-mapping strategy but also a significant result of the eQTL study. To identify the mechanisms involved, we tested the correlation between *FBN1* expression and *SLN* expression but found none (data not shown). However, gModM3 is also close to a gene cluster of metalloproteases, enzymes with an important role in ECM regeneration/remodelling. eQTL results suggests that they could be strong candidates for a role in *FBN1* transcription regulation and therefore aortic phenotype.

Identification of genetic modifiers of a disease remains a challenge since a very robust definition of the phenotype need to be obtained. The challenge is even greater in rare diseases. Indeed, despite the availability of powerful genome-wide tools and the lowering of costs, their application in the field of rare disease has been limited by the size of relevant patient samples available. Our results provide evidence that this shortcoming may be overcome by combining methods within the same population, allowing also convergence of observations to be used as a result: this strategy enhances study power (i.e., use of EPS [35]), enables resistance to the effects of population structure (i.e., use of family-based studies) and provides embedded replication (i.e., crossover of results between genome-wide methods). The principal limitations of the cross-mapping strategy, as in every genetic strategy, is the loss of power if there is a great number of underlying genetic factors and the control of confounding factors (gender, age, evolution...). Following this strategy, numerous genetic factors appear to concur to the aortic phenotype variability in MFS: genotype-phenotype correlations, gender effect (severity of aortic phenotype in men), co-occurrence of rare variants in TAA genes and *COL4A1* (responsible for severe aortic phenotype) and detection of loci for frequent modifiers among which 9 were common to at least two methods (Figs. 1, 4) and some include excellent candidate gene such as *PRKG1*, already described as an initiating gene in TAA.

Acknowledgements This work was supported by DHU FIRE (Emergence 2 project for MA), Programme Hospitalier de Recherche Clinique (CRC07032 and P071009 for CS, AOM10108 and CRC15014

for CB), Agence Nationale de la Recherche (NONAGES, ANR-14-CE15-0012-01 for GJ), Fédération de Cardiologie (for GJ), Société Française de Cardiologie (for GJ), Centre National de Génotypage (JFD), Fondation pour la Recherche en Cardiologie.

Funding MA was supported by an INSERM Poste d'accueil grant. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript. We are indebted to the clinicians of the National diagnostic network for the Marfan syndrome and related disorders, patients and families. We thank Elisabeth Tournier-Lasserre and Laurence Olivier-Faivre for their expert advice on rare variants in the *COL4A1* and *SKI* genes, respectively.

Authors contributions Patient recruitment, characterization and data generation: MA, PA, NH, OM, CS, LG, GJ, CB. Genomics and CNV data generation: MA, SG, JB, AB, VM, JFD. Targeted sequencing, Expression studies: MA, PA, LB, MSG, MPJ. Statistical analysis: MA, SG, EG. Steering and manuscript writing committee: MA, SG, PA, LB, MSG, NH, OM, CS, TB, ID, MPJ, LG, EG, JFD, GJ, CB.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Faivre L, Collod-Beroud G, Child A, et al. Contribution of molecular analyses in diagnosing Marfan syndrome and type I fibrillinopathies: an international study of 1009 probands. *J Med Genet.* 2008;45:384–90.
2. Aubart M, Benarroch L, Arnaud P, Collod-Bérout G, Jondeau G, Boileau C. Molecular genetics of the fibrillinopathies. Chichester, UK: John Wiley & Sons, Ltd; 2016. p. 1–13.
3. Faivre L, Masurel-Paulet A, Collod-Bérout G, et al. Clinical and molecular study of 320 children with Marfan syndrome and related type I fibrillinopathies in a series of 1009 probands with pathogenic *FBN1* variants. *Pediatrics.* 2009;123:391–8.
4. Baudhuin LM, Kotzer KE, Lagerstedt SA. Increased frequency of *FBN1* truncating and splicing variants in Marfan syndrome patients with aortic events. *Genet Med J Am Coll Med Genet.* 2015;17:177–87.
5. Franken R, Groenink M, de Waard V, et al. Genotype impacts survival in Marfan syndrome. *Eur Heart J.* 2016;37:3285–90. <https://doi.org/10.1093/eurheartj/ehv739>.
6. Aubart M, Gross M-S, Hanna N, et al. The clinical presentation of Marfan syndrome is modulated by expression of wild-type *FBN1* allele. *Hum Mol Genet.* 2015;24:2764–70.
7. De Paepe A, Devereux RB, Dietz HC, Hennekam RC, Pyeritz RE. Revised diagnostic criteria for the Marfan syndrome. *Am J Med Genet.* 1996;62:417–26.
8. Loeys BL, Dietz HC, Braverman AC, et al. The revised Ghent nosology for the Marfan syndrome. *J Med Genet.* 2010;47:476–85.
9. Roman MJ, Devereux RB, Kramer-Fox R, O'Loughlin J. Two-dimensional echocardiographic aortic root dimensions in normal children and adults. *Am J Cardiol.* 1989;64:507–12.
10. Erbel R, Aboyans V, Boileau C, et al. 2014 ESC Guidelines on the diagnosis and treatment of aortic diseases: Document covering acute and chronic aortic diseases of the thoracic and abdominal aorta of the adult. The Task Force for the Diagnosis and Treatment of Aortic Diseases of the European Society of Cardiology (ESC). *Eur Heart J.* 2014;35:2873–926.

11. An integrated map of structural variation in 2504 human genomes—emss-64772.pdf. <https://www.ncbi-nlm-nih.gov.gate2.inist.fr/pmc/articles/PMC4617611/pdf/emss-64772.pdf>. Accessed 20 Mar 2017.
12. Vandesompele J, De Preter K, Pattyn F, et al. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* 2002;3:RESEARCH0034.
13. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007;81:559–75.
14. Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies. *Nat Genet.* 2012;44:821–4.
15. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* 2009;5:e1000529.
16. Delaneau O, Zagury J-F, Marchini J. Improved whole-chromosome phasing for disease and population genetic studies. *Nat Methods.* 2012;10:5–6.
17. Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet.* 2007;39:906–13.
18. Sanders SJ, Ercan-Sencicek AG, Hus V, et al. Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. *Neuron.* 2011;70:863–85.
19. Colella S, Yau C, Taylor JM, et al. QuantiSNP: an objective Bayes hidden-Markov model to detect and accurately map copy number variation using SNP genotyping data. *Nucleic Acids Res.* 2007;35:2013–25.
20. Wang K, Li M, Hadley D, et al. PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res.* 2007;17:1665–74.
21. Gamazon ER, Cox NJ, Davis LK. Structural architecture of SNP effects on complex traits. *Am J Hum Genet.* 2014;95:477–89.
22. Abecasis GR, Cherny SS, Cookson WO, Cardon LR. Merlin—rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet.* 2002;30:97–101.
23. McKenna A, Hanna M, Banks E, et al. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 2010;20:1297–303.
24. Van der Auwera GA, Carneiro MO, Hartl C, et al. From FastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. In: Bateman A, Pearson WR, Stein LD, Stormo GD, Yates JR, (eds). *Current protocols in bioinformatics*. Hoboken, NJ, USA: John Wiley & Sons, Inc; 2013. p. 11.10.1–11.10.33.
25. San Lucas FA, Wang G, Scheet P, Peng B. Integrated annotation and analysis of genetic variants from next-generation sequencing studies with variant tools. *Bioinformatics.* 2012;28:421–2.
26. Wang K, Li M, Hakonarson H. ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 2010;38:e164–e164.
27. Wang GT, Peng B, Leal SM. Variant association tools for quality control and analysis of large-scale sequence and genotyping array data. *Am J Hum Genet.* 2014;94:770–83.
28. Guo D, Regalado E, Casteel DE, et al. Recurrent gain-of-function mutation in PRKG1 causes thoracic aortic aneurysms and acute aortic dissections. *Am J Hum Genet.* 2013;93:398–404.
29. Faivre L, Collod-Beroud G, Loeys BL, et al. Effect of mutation type and location on clinical outcome in 1013 probands with Marfan syndrome or related phenotypes and FBN1 mutations: an international study. *Am J Hum Genet.* 2007;81:454–66.
30. Arnaud P, Hanna N, Aubart M, et al. Homozygous and compound heterozygous mutations in the FBN1 gene: unexpected findings in molecular diagnosis of Marfan syndrome. *J Med Genet.* 2017;54:100–3.
31. GTEx Consortium. Human genomics. The genotype-tissue expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science.* 2015;348:648–60.
32. Wahl SM, Allen JB, Weeks BS, Wong HL, Klotman PE. Transforming growth factor beta enhances integrin expression and type IV collagenase secretion in human monocytes. *Proc Natl Acad Sci.* 1993;90:4577–81.
33. Bobik A. Transforming growth factor-betas and vascular disorders. *Arterioscler Thromb Vasc Biol.* 2006;26:1712–20.
34. Horowitz JC, Ajayi IO, Kulasekaran P, et al. Survivin expression induced by endothelin-1 promotes myofibroblast resistance to apoptosis. *Int J Biochem Cell Biol.* 2012;44:158–69.
35. Emilsson V, Thorleifsson G, Zhang B, et al. Genetics of gene expression and its effect on disease. *Nature.* 2008;452:423–8.

Affiliations

Melodie Aubart^{1,2} · Steven Gazal^{3,4} · Pauline Arnaud^{1,5} · Louise Benarroch¹ · Marie-Sylvie Gross¹ · Julien Buratti^{6,7} · Anne Boland⁸ · Vincent Meyer⁸ · Habib Zouali⁸ · Nadine Hanna⁵ · Olivier Milleron⁹ · Chantal Steneur⁹ · Thomas Bourgeron^{6,7,10} · Isabelle Desguerre^{2,11} · Marie-Paule Jacob¹ · Laurent Gouya^{10,12} · Emmanuelle Génin¹³ · Jean-François Deleuze⁸ · Guillaume Jondeau^{1,9,10} · Catherine Boileau^{1,5,10}

¹ Laboratory for Vascular Translational Science, INSERM U1148, DHU FIRE, Centre Hospitalo-Universitaire Xavier Bichat (APHP), 46 rue Henri Huchard, Paris 75018, France

² Service de Neuropédiatrie, Hôpital Necker-Enfants-Malades (APHP), 149 rue de Sèvres, Paris 75015, France

³ INSERM, IAME, UMR 1137, Paris 75018, France

⁴ Plateforme de génomique constitutionnelle du GHU Nord, Assistance Publique des Hôpitaux de Paris (APHP), Hôpital Bichat, Paris 75018, France

⁵ Département de Génétique, Centre Hospitalo-Universitaire Xavier Bichat (APHP), 46 rue Henri Huchard, Paris 75018, France

⁶ Institut Pasteur, Human Genetics and Cognitive Functions Unit, Paris 75015, France

⁷ CNRS UMR 3571: Genes, Synapses and Cognition, Institut Pasteur, Paris 75015, France

⁸ Centre National de Génotypage, Institut de Génomique, Evry and Centre d'Etude du Polymorphisme Humain, 2 rue Gaston Crémieux, Paris 91000, France

- ⁹ Centre de Référence pour le Syndrome de Marfan et syndromes apparentés, Service de Cardiologie, Centre Hospitalo-Universitaire Xavier Bichat (APHP), 46 rue Henri Huchard, Paris 75018, France
- ¹⁰ Université Paris 7 Denis Diderot, Paris 75013, France
- ¹¹ Université Paris 5 René Descartes, Paris 75005, France
- ¹² INSERM U1149, Faculté de Médecine site Bichat, 16 rue Henri Huchard, Paris 75018, France
- ¹³ INSERM U1078, CHRU Brest, Université de Bretagne Occidentale, Brest 29200, France