SUPPLEMENTAL MATERIAL

METHODS

Study Design

The Cardiovascular Health Improvement Project (CHIP) is a biorepository with a historical collection of genotype and phenotype data, family history, DNA, and aortic tissue from participants with thoracic aortic disease. Thoracic aortic disease was defined as any pathology of the thoracic aorta, including aneurysm, dissection/intramural hematoma, and rupture of the aorta. Between August 2013 and December 2015, 1752 participants were enrolled in the CHIP biorepository, and of those, 265 cases had a diagnosis of thoracic aortic dissection including type A or type B aortic dissection or thoracic aortic rupture with or without aortic aneurysm. Age-, sex, and ancestry-matched controls (n=265) were identified as previously described from the Michigan Genomics Initiative (MGI), which is a surgical-based biobank.[1] In brief, we matched thoracic aortic dissection cases from CHIP to MGI controls of the same sex, age range (-5, +10) at time of enrollment, and minimum Euclidean distance as calculated from the first two principal components of genotype data indicative of genetic ancestry. Principal components were obtained by principal component analysis (PCA) in PLINK 1.9[2] on 58,563 genotyped variants with > 0.05 minor allele frequency. For 83 CHIP samples without genotypes from a customized Illumina HumanCoreExome v12.1 bead array, we used self-reported ancestry instead of principal component-based ancestry to identify controls with similar genetic ancestry. In the event that insufficient DNA was available for the best matched control, we moved sequentially through the top 10 best matched controls. All study procedures were approved by the Institutional Review Board (HUM00052866 and HUM00094409) and all subjects provided informed consent.

Clinical Characteristics

The electronic medical record was systematically reviewed for all thoracic aortic dissection cases (hereon referred to as cases). Specifically, the electronic medical record was used to verify demographics, clinical diagnoses, family history, surgical history, clinical genetic testing results, medications, comorbidities, and systemic features. Patients were excluded (n=18) during electronic medical record review if a traumatic aortic dissection (n=9, accident or illicit drug use) or abdominal aortic rupture (n=9, etiology is typically atherosclerotic in nature) was identified. All cases with a clinical diagnosis of Marfan syndrome or a research-level pathogenic variant identified in *FBN1* were reviewed using the Revised Ghent Nosology (Supplementary Tables 5 and 6).[3] The clinical characteristics were reviewed in conjunction with the clinical genetic testing results (when available), and compared to the whole exome sequencing results.

All cases completed a family history questionnaire with a trained research assistant at the time of enrollment to CHIP. The family history questionnaire asked participants to recall whether any first or second degree relatives had pathology to the thoracic aorta, including aneurysm, dissection/intramural hematoma, or rupture of thoracic aorta. For this manuscript, we focused on first degree relatives, and thoracic aortic disease was collapsed into a single categorical variable with "yes" equaling positive and "no" equaling negative family history. This process was repeated for each first degree relative (mother, father, siblings, and children). Clinical characteristics for the cases (pathogenic carriers versus non-pathogenic carriers) are presented as median and inter-quartiles for continuous data and n (%) for categorical data.

Whole Exome Sequencing

DNA samples from whole blood for cases and controls (n=530) were prepared for whole exome sequencing as outlined by the Northwest Genomics Center (NWGC, University of Washington).

528 samples were approved for sequencing with sufficient DNA quality. DNA libraries underwent exome capture using Roche/Nimblegen SeqCap EZ v2.0 (~36.5 MB target). NWGC's sequencing pipeline is a combined suite of Illumina software and other industry standard software packages (i.e., Genome Analysis ToolKit [GATK], Picard, BWA-MEM, SAMTools, and in-house custom scripts) and consisted of base calling, alignment, local realignment, duplicate removal, quality recalibration, data merging, variant detection, genotyping and annotation. Variant detection and genotyping were performed using the HaplotypeCaller tool from GATK[4] and hard filtering was performed (GATK v3.4). Exome completion was defined as having > 90% of the exome target at > 8X coverage and >80% of the exome target at > 20X coverage. Exome completion and several metrics including capture efficiency, raw error rates, and sample contamination validation were used for standard quality control assessment. A total of 521 samples, 260 cases and 261 controls, and 323,867 variants (single nucleotide polymorphisms and insertion/deletions) passed standard quality control and were released to researchers.

Additional sample and variant filtering

Bi-allelic sites were extracted and lower coverage genotypes with depth (DP) < 5 were masked out. All samples met the quality control threshold of an individual level call rate > 0.9. Poor quality sites with site-level call rate < 0.9 were excluded. Variants significantly deviating from HWE with p-value < $1x10^{-6}$ were also removed. KING[5] was used to identify five sample pairs as duplicates, and the sample with the lowest call rate was excluded, leaving 258 cases and 258 controls. Concordance with Exome+GWAS array genotypes was > 0.999 across all minor allele frequencies. The final analysis set was comprised of 240 cases and 258 controls and 299,195 variants. We opted to keep all cases and controls that passed quality control procedures, rather than reduce the sample size by only including complete pairs.

Annotation of variants with clinical implications

We focused on the following genes which confer a dominantly inherited risk for thoracic aortic dissection and with definitive and strong evidence of association of hereditary thoracic aortic aneurysm and dissection: *ACTA2, COL3A1, FBN1, MYH11, SMAD3, TGFB2, TGFBR1, TGFBR2, MYLK, LOX,* and *PRKG1*.[6,7] A total of 248 variants in these genes were annotated using dbNSFPv3.5a. and reviewed by a single researcher blinded to case or control status of the sample in which the variant was identified. Variants were then annotated as pathogenic, variants of unknown significance (VUS), or benign. Protein isoforms that are major isoforms expressed in smooth muscle cells or used in previous publications were used to predict amino acid changes (Supplementary Table 7). To define pathogenic variants, we annotated variants based on the ACMG-AMP standards and guidelines.[8] Additionally, established rules[7] were used to classify rare variants as pathogenic or disease-causing. Rare variants were annotated as variants of unknown significance if lacking proof of pathogenicity. Variants were considered benign if they are nonsynonymous mutations with MAF $\geq$ 0.005 in ExAC Non-Finnish Europeans[9] or in a nonrelevant isoform, are synonymous mutations, or occurred > ±2 bp from intron/exon boundaries.

Molecular Inversion Probe Sequencing

Molecular Inversion Probe Sequencing (MIPS) was performed as a technical replicate of cases and controls that were whole exome sequenced and found to carry a pathogenic variant. MIPS was first performed on DNA from the same extraction used for whole exome sequencing. An additional round of MIPS was performed from a second DNA isolation to serve as a sample replicate. A custom targeted sequencing panel was designed for 116 genes using single molecule molecular inversion probes or smMIPS.[10] Coding exon coordinates were retrieved from the UCSC Genome Browser "knownGene" table (build GRCh37/hg19) and padded by 5 bp in each direction to include

splice sites. Probes were designed and prepared as previously described.[11] For each sample, approximately 9 ng of purified smMIPS probes were combined with 250 ng genomic DNA. The captured material was amplified by PCR using barcoded primers. The resulting PCR products were pooled for one lane of paired-end 150 bp sequencing on an Illumina HiSeq 4000 instrument at the University of Michigan Sequencing Core.

Reads were aligned to the human genome reference (build GRCh37/hg19) using bwa mem[12] and a custom pipeline (available at https://github.com/kitzmanlab/mimips) was used to remove smMIPS probe arm sequences and remove reads with duplicated molecular tags. Variant calling of MIPS sequencing results for both single nucleotide variants and insertions/deletions was performed using the GotCloud[13] pipeline. An iterative filtering process was performed after variant calling to remove variants with a depth < 10, then samples with call rates < 0.6, followed by variants with a call rate < 0.8, and finally samples with call rates < 0.9.

Statistical analysis for burden of variants in cases and controls

Univariate comparisons between pathogenic and non-pathogenic carrier clinical characteristics were performed using Chi-square with Yates' continuity correction or Fisher's exact test when any expected cell counts were < 5 for categorical data, and Wilcoxon rank sum tests were used for continuous data. Multivariable logistic regression was used to identify associations between risk factors and pathogenic variant carriers.

To test for association between carriers of a given variant class and case/control status we used Fisher's exact test when any expected cell counts were < 5 and Chi-square test with Yates' continuity correction otherwise. This was done using the statistical programming language R version 3.5.1. We identified first degree relatives using KING2[5] and whole exome sequencing variant calls. For the two first degree relative pairs we found in the cases, we retained the first

sample acquired (proband) for the analysis resulting in 238 cases. A sample carrying at least one of a variant class was considered a carrier. We performed burden tests for association with case status across the 11 genes for all pathogenic variants (N=24) and VUS (N=86). We first excluded carriers of pathogenic variants before testing for association with case status for carriers of VUS (N_{cases}=213, $N_{controls}$=258). A Bonferroni threshold of 0.003 was used to account for 17 independent tests, which are assumed to be independent.

Data Visualization

Annotated Fibrillin 1 protein domains from Pfam 31.0[14] and a modified version of GenVisR 1.14.1[15] were used for data visualization. Variants falling in mutation splice sites are not included in this protein-level visualization.

REFERENCES

1. Yang B, et al. Protein-altering and regulatory genetic variants near GATA4 implicated in bicuspid aortic valve. *Nat Commun*. 2017;8:15481.

2. Chang CC, et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*. 2015;4:7.

3. Loeys BL, et al. The revised Ghent nosology for the Marfan syndrome. *J Med Genet*. 2010;47:476-485.

4. Van der Auwera GA, et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics*. 2013;43:11 10 11-33.

5. Manichaikul A, et al. Robust relationship inference in genome-wide association studies. *Bioinformatics*. 2010;26:2867-2873.

6. Brownstein AJ, et al. Genes Associated with Thoracic Aortic Aneurysm and Dissection: An Update and Clinical Implications. *Aorta (Stamford)*. 2017;5:11-20.

7. Guo DC, et al. Heritable Thoracic Aortic Disease Genes in Sporadic Aortic Dissection. *J Am Coll Cardiol*. 2017;70:2728-2730.

8. Richards S, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med*. 2015;17:405-424.

9. Lek M, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature*. 2016;536:285-291.

10. Hiatt JB, et al. Single molecule molecular inversion probes for targeted, high-accuracy detection of low-frequency variation. *Genome Res*. 2013;23:843-854.

11. Yoon JK, et al. microDuMIP: target-enrichment technique for microarray-based duplex molecular inversion probes. *Nucleic Acids Res*. 2015;43:e28.

12. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *ArXiv13033997 Q-Bio* 2013;

13. Jun G, et al. An efficient and scalable analysis framework for variant extraction and refinement from population-scale DNA sequence data. *Genome Res*. 2015;25:918-925.

14. Finn RD, et al. Pfam: the protein families database. *Nucleic Acids Res*. 2014;42:D222-230.

15. Skidmore ZL, et al. GenVisR: Genomic Visualizations in R. *Bioinformatics*. 2016;32:3012-3014.

Supplementary Note

Concordance between research-level whole exome sequencing and clinical genetic testing

20 (20/240) aortic dissection cases had previous clinical genetic testing in their medical record. For 13 patients our findings agreed with clinical genetic testing (pathogenic=5, no findings=5, VUS=3). The remaining 7 cases had discrepancies between the clinical genetic testing and research-level WES and variant annotation. For one patient, we identified a VUS in *MYH11,* which was not one of the 6 genes clinically evaluated, and for another patient, clinical genetic testing identified a VUS in 2 genes (*CBS, COL5A1*) which were not identified in the 11 HTAAD genes that we annotated. (Supplementary Table 3).

In another patient, functional annotations and the protein domain affected were sufficient evidence for classification as VUS in both *MYLK* and *COL3A1* which were clinically evaluated in 2016 but considered benign. For 1 patient, clinical genetic testing found double heterozygous genotypes for 2 VUS variants in *COL5A1* and *CBS* which were not annotated in our research-level genetic testing.

3 patients were found to have likely pathogenic or possibly causative variants by clinical genetic testing which we annotated as VUS. Finally, one patient had a 2012 clinical genetic testing result of pathogenic which we annotated as a VUS due to lack of evidence for pathogenicity.

Pathogenic variants in commonly used databases

Of the 17 of our 24 pathogenic variants present in ClinVar as of September 30, 2018, 12 are pathogenic, 8 are listed as pathogenic, 5 as likely pathogenic, 1 as conflicting interpretation of pathogenicity, and 3 as VUS. 1 of the VUS variants is for a non-aortic phenotype—Wolff-Parkinson-White pattern. 15 of the 24 variants have an rsID in dbSNP v151 for the hg19 chromosomal position, but only 11 of those have reference and alternate alleles corresponding to

the variation catalogued in our cohort. For example, dbSNP lists rs113935744 as having reference allele T and alternate allele A whereas our sample was a carrier for alternate allele C.

5,344 loss of function and missense variants from the 11 genes of interest were obtained from gnomAD v2.1. 930 of those variants are listed in ClinVar with the same reference and alternate alleles as gnomAD. 16 of those are Pathogenic or Pathogenic/Likely pathogenic. By summing the allele counts across variants, we estimate pathogenic variants in these genes have a background prevalence of 9.396×10^{-6} (30 occurrences in 3,193,956 alleles). Of the 24 pathogenic variants, only 2 are catalogued (rs779512296 and rs761857514) in gnomAD. rs779512296 has an allele frequency of 0.00002891 in the gnomAD Latino population and 0.000008801 in the non-Finnish European population. rs761857514 has an allele frequency of 0.00003267 in the South Asian population.

In this effort we used pathogenicity filtering criteria tailored to our phenotype of interest. As previously shown, using a typical pathogenicity filter (predicted deleterious by at least two of Polyphen2, SIFT, and MutationTaster; 0.5% maximum allele frequency across European Americans and African Americans in the Exome Variant Server; and 5% maximum allele frequency in 1000G) there is a high background prevalence of protein-altering variants in a population[4]. For example, default filtering on GeneVetter (genevetter.kidneyomics.org) identifies 322 of 2,535 (12.7%) 1000 Genomes samples as pathogenic variant carriers, which is a higher background prevalence than we might expect for TAAD. Using the same filter in our cohort, we identify 48 cases (19.3%) and 22 controls (8.5%) as carriers for a pathogenic variant.

Comparisons of cases versus controls

After quality control, we had 240 cases and 258 controls rather than 265 age, sex, and ancestry matched pairs remaining. We confirmed that the distribution of age, sex, and ethnicity was similar after the attrition of matched cases/controls during quality control (Supplementary Figures 1-3). These samples were used to

test for association between disease and pathogenic/VUS variant carrier status (Supplementary Table 1 and 4). To ensure these comparisons were robust to slightly unbalanced case/control matching, we performed logistic regression using age, sex, and carrier status as predictors of case/control status to replicate the analysis in Supplementary Table 4. The Wald test p-value for affect of VUS on case status adjusted for age/sex is 0.06, similar to the Chi-square p-value of 0.07. For pathogenic variants we had 0 controls as carriers so we used Firth's bias-Reduced penalized-likelihood logistic regression as implemented in the R package logistf. The p-value from the profile penalized log likelihood is 1.5e-8, similar to the Chi-square test p-value of 2.8e-7.

	Pathogenic (N=24)							VUS (N=86)						
Gene	**Cases* (N=237)**	**Controls (N=258)**	**Fisher Exact Test p-value**	**Odds Ratio Estimate**	**Odds Ratio 95% Confidence Interval**	**Chi-square test p-value (Yates' continuity correction)**	**Chi-square test statistic (Yates' continuity correction)**	**Cases* (N=213)**	**Controls (N=258)**	**Fisher Exact Test p-value**	**Odds Ratio Estimate**	**Odds Ratio 95% Confidence Interval**	**Chi-square test p-value (Yates' continuity correction)**	**Chi-square test statistic (Yates' continuity correction)**
ACTA2	NA	NA						1	0	0.452	Inf	0.031, Inf		
COL3A1	1	0	0.48	Inf	0.028, Inf			9	5				0.237	1.4
FBN1	18	0				2.05e-5	18.14	12	15				1	1.18e-29
LOX	1	0	0.48	Inf	0.028, Inf			NA	NA					
MYH11	NA	NA						18	13				0.194	1.69
MYLK	NA	NA						5	4	0.738	1.525	0.323, 7.789		
PRKG1	2	0	0.23	Inf	0.204, Inf			3	3	1	1.213	0.161, 9.16		
SMAD3	2	0	0.23	Inf	0.204, Inf			4	0	0.041	Inf	0.805, Inf		
TGFB2	NA	NA						5	3	0.477	2.040	0.392, 13.290		
TGFBR1	NA	NA						3	3	1	1.214	0.161, 9.16		
TGFBR2	1	0	0.48	Inf	0.028, Inf			7	9				1	7e-31

Supplementary Table 1. Association between variants of a given class and case/control status per each of the 11 HTAAD genes. A sample from each of the two related pairs in the cases was removed while the first ascertained sample was retained. When testing the VUS class of variants, only cases without a pathogenic variant were considered. Accounting for multiple testing using a Bonferroni threshold of 0.003, the only significant association identified is for pathogenic variants in *FBN1*.

Chr	Pos	Variant type	Ref	Alt	Sample (NHLBI_ID)	Sample (GWAS/MIPS ID)	WES (GT:AD:DP:GQ:PL)	MIPS_v1 Variant call (GT:DP:GQ:PL)	MIPS_v1 Quality	MIPS_v2 Variant call (GT:DP:GQ:PL for SNPs, GT:PL:DP:AD:GQ for indels)	MIPS_v2 Quality
15	48707913	SNP	T	C	16554	58432	0/1:28,17:45:99:488,0,896	0/1:267:99:255,0,255	Failed individual level call rate filter	0/1:676:99:255,0,255	Pass
15	48713849	SNP	G	C	19082	113392	0/1:36,35:71:99:952,0,1142	NA	Sample not sequenced	0/1:222:255:255,0,255	sample filtered out due to high missingness in first pass, variant filtered by SVM filter
15	48714232	SNP	C	A	11353	57411	0/1:43,33:76:99:931,0,1329	0/1:1165:99:255,0,255	Pass	0/1:1050:99:255,0,255	Pass
15	48719947	Indel	TGAAGCAGTACCCTTCCC	T	17339	57403	0/1:26,12:38:99:427,0,4465	NA	Indel calling not performed	0/1:1189:.:583,586,20:43177,0,38932	Pass
15	48722967	SNP	A	G	15731	58466	0/1:8,6:14:99:175,0,237	0/1:1065:99:255,0,255	Pass	0/1:742:99:255,0,255	Pass
15	48725107	SNP	C	T	12040	57445	0/1:19,16:35:99:427,0,631	0/1:607:99:255,0,255	Failed individual level call rate filter	0/1:1564:99:255,0,255	Pass
15	48730109	SNP	G	A	11487	57396	0/1:12,7:19:99:216,0,401	0/1:86:99:255,0,255	Pass	0/1:162:99:255,0,255	Pass
15	48744873	SNP	C	T	16426	113380	0/1:13,13:26:99:318,0,361	NA	Sample not sequenced	0/1:242:99:255,0,255	Pass
15	48760660	SNP	A	G	17258	113401	0/1:28,35:63:99:975,0,807	NA	Sample not sequenced	0/1:154:255:255,0,255	sample filtered out in second pass due to missingness rate, variant passes filter
15	48764793	SNP	A	G	15339	57412	0/1:22,24:46:99:724,0,693	0/1:3287:99:255,0,255	Pass	0/1:8893:99:255,0,255	Pass
15	48773870	Indel	C	CT	12144	57419	0/1:24,29:53:99:741,0,561	NA	Indel calling not performed	0/1:1748:.:836,905,7:24198,0,21469	Pass
15	48782066	SNP	C	A	11080	58472	0/1:32,20:52:99:533,0,1034	0/1:881:99:255,0,255	Failed individual level call rate filter	0/1:2433:99:255,0,255	Pass
15	48786401	SNP	C	G	16641	57386	0/1:49,52:101:99:1347,0,1369	0/1:104:99:255,0,255	Pass	0/1:218:99:255,0,255	Pass
15	48802264	Indel	G	GT	11970	57402	0/1:27,38:65:99:1212,0,802	NA	Indel calling not performed	0/1:2213,0,2047:162:79,83,0:.	sample filtered out due to high missingness in first pass
15	48802366	SNP	T	C	15837	57597	0/1:12,16:28:99:460,0,355	0/1:372:99:255,0,255	Pass	0/1:240:99:255,0,255	Pass
15	48808561	SNP	T	C	13555	113354	0/1:19,16:35:99:525,0,561	NA	Sample not sequenced	0/1:415:99:255,0,255	Pass
15	48812913	SNP	G	A	16930	57832	0/1:27,26:53:99:773,0,755	0/1:427:99:255,0,255	Pass	0/1:753:99:255,0,255	Pass
15	48888576	SNP	C	T	17920	57617	0/1:21,16:37:99:480,0,699	NA	No coverage in sequencing bam file	0/1:525:99:255,0,255	Pass
15	67457370	Indel	TGAA	T	10317	57605	0/1:17,19:36:99:727,0,647	NA	Indel calling not performed	0/1:180:.:95,85,0:2617,0,3122	Pass
15	67462935	Indel	TA	T	16115	58000	0/1:25,41:66:99:1336,0,736	NA	Indel calling not performed	0/1:89:.:49,39,1:1005,0,1307	Pass

15	67462935	Indel	TA	T	13332	58351	0/1:40,27:67:99:810,0,1268	NA	Indel calling not performed	0/1:39:.:16,23,0:632,0,396	Pass
2	189858169	SNP	G	A	15202	57577	0/1:46,36:82:99:1077,0,1385	0/1:363:99:255,0,255	Pass	0/1:446:99:255,0,255	Pass
3	30732950	SNP	G	A	17845	57458	0/1:16,23:39:99:667,0,499	0/1:234:99:255,0,255	Pass	0/1:621:99:255,0,255	Pass
10	53227579	SNP	G	A	19825	57370	0/1:42,47:89:99:1590,0,1192	0/1:39:99:255,0,255	Failed individual level call rate filter	0/1:301:99:255,0,255	Pass
10	53227579	SNP	G	A	10712	57607	0/1:91,58:149:99:1689,0,2706	0/1:2002:99:255,0,255	Pass	0/1:168:255:255,0,255	sample filtered out in second pass due to high missingness
5	121412592	Indel	CCAGA	C	14301	57653	0/1:44,39:83:99:1506,0,2525	NA	Indel calling not performed	0/1:1650:.:757,879,14:31292,0,25368	Pass

Supplementary Table 2. Confirmation of WES variant calls with Molecular Inversion Probe Sequencing (MIPS). Two rounds of MIPS were performed to confirm the pathogenic variant calls in all 26 patients. In round 1, 22 of the 26 samples were sequenced. In round 2, all samples were sequenced.

CLIA			Research			
CLIA year	**Clinical Genetic Results**	**Classification**	**Variant**	**Classification**	**Gene**	**Rationale for discrepancy**
2015	Heterozygous for the p.R192Q pathogenic mutation in the PRKG1 gene	Pathogenic	10:53227579	Pathogenic	PRKG1	Concordant
2010	Mutation: FBN1 Exon 22 Nucleotide: c.2728G>C Amino Acid:Asp910His	Pathogenic	15:48786401	Pathogenic	FBN1	Concordant
NA	Genetically confirmed MFS	Pathogenic	15:48782066	Pathogenic	FBN1	Concordant
			2:189856434	VUS	COL3A1	Concordant
NA	clinical genetic testing, no variant identified	No findings				
2014	Panel was negative for everything, COL3A1 TGFBR1 TGFBR2, ACTA2, SMAD3, TGFB2 tested	No findings	16:15820794	VUS	MYH11	Not tested in CLIA panel
2012	SMAD3 genetic mutation	Pathogenic	15:67462935	Pathogenic	SMAD3	Concordant
2012	VUS from TGFBR2	VUS	3:30713866	VUS	TGFBR2	Concordant
2016	No genetic mutations discovered, 22 gene panel including COL3A1 and MYLK	No findings	3:123337545	VUS	MYLK	MYLK p.T1814I is absent in the ExAC and gnomAD database. T1814 alteration is not reported before so it is unclear whether alter this amino acid lead to TAD. Multiple functional prediction programs suggest that this variant is damaging.
			2:189863424	VUS	COL3A1	In triple helical region but didn't alter critical Glycine
NA	6 gene vascular aneurysm panel and fibrillin 1 sequencing were negative	No findings				
NA	SMAD3 mutation related to Loeys-Dietz syndrome	Pathogenic	15:67462935	Pathogenic	SMAD3	Concordant
2017	Patient was negative for panel	No findings				
2014	SMAD 3 likely pathogenic variant	Likely pathogenic	16:15844048	VUS	MYH11	MYH11 p.K1256del is not found in the ExAC and gnomAD database. Deletion of this amino acid is not reported before so it is unclear whether deletion of this amino acid lead to

						TAD. Couple of single amino acid deletion flanking K1256 are found in the gnomAD and ExAC datebases. In the gnomAD v2.1 control database, there are 6 K1263del alleles and 2K1231del alleles.
			15:67482824	VUS	SMAD3	SMAD3 p.V410 is found in the ExAC with low MAF (5.53E-04). Some functional prediction programs suggest damaging and other suggest benign.
2014	SMAD3 gene mutation in exon 9, c.1228G>T, p.Val410Phe	Likely pathogenic	15:67482824	VUS	SMAD3	SMAD3 p.V410 is found in the ExAC with low MAF (5.53E-04). Some functional prediction programs suggest damaging and other suggest benign.
NA	Only was tested for Marfan and was found to be negative	No findings				
2012	Possibly causative SMAD3 mut (c.331T>A)	Possibly causative	15:67457357	VUS	SMAD3	No evidence for pathogenicity
2016	VUS in COL3A1 p. V5291	VUS	2:189860493	VUS	COL3A1	Concordant
2016	Heterozygous for the p.R369C pathogenic mutation in the CBS gene. Heterozygous for the p.P435A (c.1303C>G) VUS in the COL5A1 gene	VUS	21:44480591	NA	CBS	Not one of 11 HTAAD genes
			9:137623480	NA	COL5A1	Not one of 11 HTAAD genes
2012	FBNI exon 32 Nucleotide: c. 4057G>A Amino: Gly1353Arg	Likely pathogenic	15:48766755	VUS	FBN1	Reported in patients, no evidence for pathogenicity. Located in EGF-like 22 calcium binding domain and is not a critical amino acid for the domain.
2013	TGFBR1 Exon 5 Nuc: c.949C>T AA: His317Tyr	Likely pathogenic	9:101904961	VUS	TGFBR1	No evidence for pathogenicity
2013	No mutations found	No findings				

Supplementary Table 3. Concordance between research-level and clinical genetic testing in 20 patients with CLIA-certified genetic testing results.

Variant class (# of variants in class)		Cases	Controls	Chi-square test p-value (Yates' continuity correction)	Chi-square test statistics (Yates' continuity correction)
		n=238	**n =258**		
pathogenic (24)	Non-carrier	213	258	2.79e-7	26.39
	Carrier	25	0		
		n=213	**n=258**		
VUS (86)	Non-carrier	155	207	0.072	3.25
	Carrier	58	51		

Supplementary Table 4. Association between variants of a given class and case/control status across all 11 genes. A sample from each of the two related pairs in the cases was removed while the first ascertained sample was retained. When testing the VUS class of variants, only cases without a pathogenic variant were considered.

Basis for Diagnosis of Marfan Syndrome		
	Non-Pathogenic (n=6)	Pathogenic (n=18)
Family history absent or unknown (n)		
AD + EL		1
AD + *FBN1* mutation	1	4
EL + *FBN1* mutation		
AD + Systemic score ≥ 7		
AD + EL + *FBN1* mutation		
AD + EL + Systemic score ≥ 7	1	1
AD + Systemic score ≥ 7 + *FBN1* mutation		
Family history present (n)		
AD	3	10
EL		1
Systemic score ≥ 7	1	1
AD + EL		
AD + Systemic score ≥ 7		
EL + Systemic score ≥ 7		
AD + EL + Systemic score ≥ 7		

Abbreviations: AD=aortic dissection; EL=ectopia lentis

Supplementary Table 5. Diagnosis of Marfan Syndrome based on the Revised Ghent Nosology. According to the Revised Ghent Nosology, a positive family history is based on a diagnosis of Marfan Syndrome among a first degree family member.

Comparison of Phenotypic Features in Patients with and without Pathogenic Variants in *FBN1*

Features	Non-Pathogenic (n=6)	Pathogenic (n=18)
Ectopia lentis	1	2
Systemic score	5 (3, 8)	3 (2, 6)
Arachnodactyly	5	11
Pectus carinatum	1	4
Pectus excavatum	1	3
Dural ectasia	3	4
Reduced US/LS + increased arm/height + no severe scoliosis	0	2
Scoliosis	2	5
Kyphosis	3	0
Plain pes planus	0	3
Skin striae	1	2
Myopia	2	5
Mitral valve prolapse	1	5

Values are median (interquartile range) or n (%).

Abbreviations: US/LS=upper segment/lower segment ratio.

Supplementary Table 6. Comparison of phenotypic features in patients with and without pathogenic variants in *FBN1*. If a systemic feature is not listed above then it did not occur in any of the cases.

Gene	NCBI ID
ACTA2	NM_001141945.1
COL3A1	NM_000090.3
FBN1	NM_000138.4
LOX	NM_002317.5
MYH11	NM_002474.2
MYLK	NM_053025.3
PRKG1	NM_001098512.3
SMAD3	NM_005902.3
TGFB2	NM_003238.3
TGFBR1	NM_004612.2
TGFBR2	NM_003242.5

Supplementary Table 7: mRNA-seq isoforms used to identify the predicted amino acid change. Typically, this is a major isoform expressed in smooth muscle cells. For some proteins, previous publication's isoform was chosen. NM indicates manually annotated and reviewed mRNAs.

SUPPLEMENTARY FIGURE LEGENDS

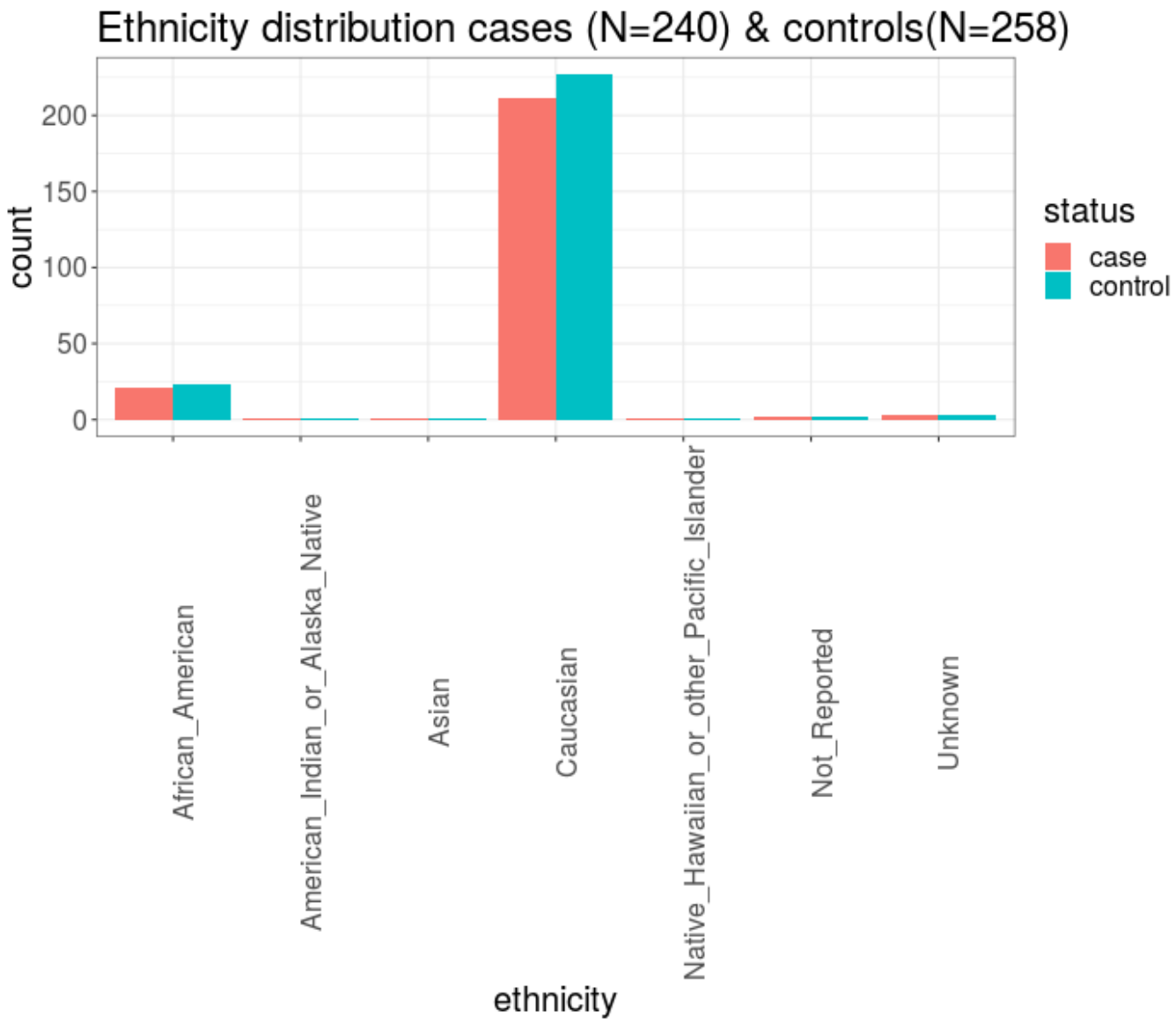
Supplementary Figure 1. Age Distribution Between Cases (n=240) and Controls (n=258). The For distribution of age was similar after the attrition of matched cases/controls during quality control.

Supplementary Figure 2. Ethnicity Distribution Between Cases (n=240) and Controls (n=258). The For distribution of ethnicity was similar after the attrition of matched cases/controls during quality control.

Supplementary Figure 3. Sex Distribution Between Cases (n=240) and Controls (n=258). The For distribution of sex was similar after the attrition of matched cases/controls during quality control.

Supplementary Figure 1.

Supplementary Figure 2.



Supplementary Figure 3.