SCIENCE CHINA
Life Sciences

• RESEARCH PAPER •

January 2017 Vol.60 No.1: 57–65
doi: 10.1007/s11427-016-0333-3

Variants of genes encoding collagens and matrix metalloproteinase system increased the risk of aortic dissection

Zongzhe Li[1], Chengming Zhou[1], Lun Tan[1], Peng Chen[1], Yanyan Cao[1], Chenze Li[1], Xianqing Li[1], Jiangtao Yan[1], Hesong Zeng[1], Dao-Wu Wang[2*] & Dao-Wen Wang[1**]

[1]*Division of Cardiology, Departments of Internal Medicine and Genetic Diagnosis Center, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430030, China;*
[2]*Department of Cardiology, the First Affiliated Hospital, Nanjing Medical University, Nanjing 210029, China*

Received October 9, 2016; accepted November 1, 2016; published online December 13, 2016

Aortic dissection (AD) is a devastating, heterogeneous condition of aorta. The homeostasis between collagens and matrix metalloproteases (MMPs)/tissue inhibitors of MMPs (TIMPs) system in the extracellular matrix plays an important role for structure and functions of aorta. However, our knowledge on association between variants of genes in this system and pathogenesis of AD is very limited. We analyzed all yet known coding human genes of collagens (45 genes), MMPs/TIMPs (27 genes) in 702 sporadic AD patients and in 163 matched healthy controls, by using massively targeted next-generation and Sanger sequencing. To define the pathogenesis of potential disease-causing candidate genes, we performed transcriptome sequencing and pedigree co-segregation analysis in some genes and generated *Col5a2* knockout rats. We identified 257 pathogenic or likely pathogenic variants which involved 88.89% (64/72) genes in collagens-MMPs/TIMPs system and accounted for 31.05% (218/702) sporadic AD patients. In them, 84.86% patients (185/218) carried one variant, 12.84% two variants and 2.30% more than two variants. Importantly, we identified 52 novel probably pathogenic loss-of-function (LOF) variants (20 nonsense, 16 frameshift, 14 splice sites, one stop-loss, one initiation codon) in 11.06% (50/452) AD patients, which were absent in 163 controls (P=2.5×10^{-5}). Transcriptome sequencing revealed that identified variants induced dyshomeostasis in expression of collagens-TIMPs/MMPs systems. The *Col5a2*$^{-/-}$ rats manifested growth retardation and aortic dysplasia. Our study provides a first comprehensive map of genetic alterations in collagens-MMPs/TIMPs system in sporadic AD patients and suggests that variants of these genes contribute largely to AD pathogenesis.

aortic dissection, collagen, matrix metalloproteinase, next-generation sequencing, genetic diagnosis

Citation: Li, Z., Zhou, C., Tan, L., Chen, P., Cao, Y., Li, C., Li, X., Yan, J., Zeng, H., Wang, D.W., and Wang, D.W. (2017). Variants of genes encoding collagens and matrix metalloproteinase system increased the risk of aortic dissec. Sci China Life Sci 60, 57–65. doi: 10.1007/s11427-016-0333-3

INTRODUCTION

Aortic dissection (AD) is a disease of life-threatening emergency associated with high mortality and is difficult to be predicted (Gao et al., 2016; Nienaber and Eagle, 2003; Suzuki et al., 2009). AD usually occurs in non-syndromic and sporadic form (~80%) or in association with hereditary genetic disorders, such as familial thoracic aortic dissection (fTAAD), Marfan syndrome (MFS), the vascular type of Ehlers-Danlos syndrome (EDS), Loeys-Dietz syndrome (LDS), and arterial tortuosity syndrome (ATS) (Goldfinger et al., 2014; Guo et al., 2011; Hoffjan, 2012; Khau Van Kien et al., 2004; Ritelli et al., 2009). Although previous genetic studies in familial

*Corresponding author (email: dwwang@tjh.tjmu.edu.cn)
**Corresponding author (email: david37212@hotmail.com)

AD cases have identified several genes (Table S1 in Supporting Information) responsible for this disease, large fraction of clinical cases, especially sporadic cases, still lack genetic targets (Campens et al., 2015; Guo et al., 2011).

Except for a genome-wide association study (GWAS), no previous studies have systematically investigated the genetic basis of sporadic spontaneous AD or the relationship between genotype and phenotype heterogeneity, especially for collagens and matrix metalloproteases (MMPs)/tissue inhibitors of MMPs (TIMPs) (LeMaire et al., 2011). However, GWASs were performed using only known genetic markers involved in both cases and healthy controls, the ability to identify novel pathogenesis is limited compared with direct sequencing (Marian and Belmont, 2011; Salomon et al., 2016). It is clear that the imbalance of the homeostasis between collagens and MMPs/TIMPs system is a key responsible for extracellular matrix (ECM) degeneration and structure and functions of aorta, and therefore may contribute to the pathogenesis of AD (Barbour et al., 2007; Theruvath et al., 2012; Tsamis et al., 2013). However, to date, except for gene *COL3A1*, no pathogenic genes in collagens and MMP/TIMPs system have been demonstrated in AD patients (Sakai et al., 2012; Smith et al., 2011).

To systematically study the pathogenic basis of genes in collagens and MMP/TIMPs system in sporadic AD, we sequenced all 72 related genes in the system in 702 patients with sporadic AD and in 163 matched healthy controls. According to the guideline (Richards et al., 2015), we identified 257 pathogenic or likely pathogenic variants in 31.05% (218/702) patients, including 52 novel probably pathogenic loss-of-function (LOF) variants.

RESULTS

Targeted resequencing output and quality

We obtained sequence data for 2,496 amplicons of 2,279 exons from 72 targeted genes for 702 sporadic AD patients and 165 controls. On average, we gained 4,756,177 mapped reads per sample with a mean base coverage depth of 716-fold. In summary, 99.70% of the target amplicons were covered at least once, 98.62% were covered at least 20 times, and 96.06% were covered at least 100 times in coding regions.

Identification of pathogenic or likely pathogenic variants

Using stringent criteria (Richards et al., 2015), in all we identified two pathogenic and 255 likely pathogenic variants in the 72 genes in 31.05% (218/702) of the AD patients, and most of these genes harbored more than one likely pathogenic mutation (Figure 1 and Table S2 in Supporting Information).

Importantly, we identified 52 probably pathogenic LOF variants in collagens or MMPs/TIMPs genes in the 702 AD cases. We did not observe any pathogenic or probably pathogenic LOF variants in these genes in the 165 controls (P=2.5×10^{-5}). These LOF variants include 20 nonsense, 16 frameshift, 14 splice-sites, one stop-loss, and one initiation-codon-break (Table 1). In addition, none of these LOF variants carriers had potential pathogenic variants in the 12 known AD pathogenic genes (Table S1 in Supporting Information).

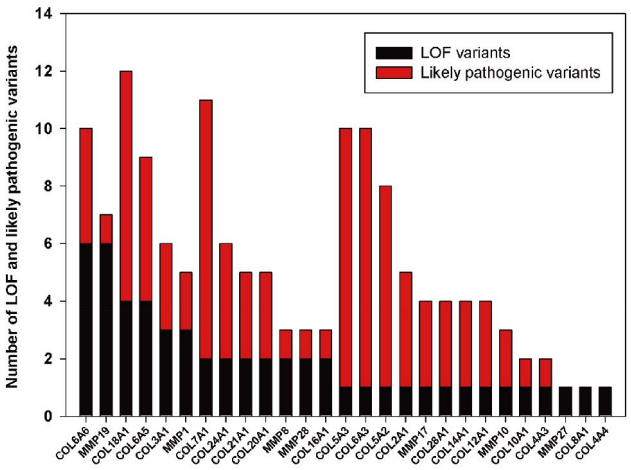

Figure 1 The LOF and likely pathogenic variants identified in 702 AD cases. This panel shows the 54 novel identified LOF and likely pathogenic variants identified in 702 AD patients of collagens and MMPs/TIMPs genes. The LOF variants are shown in black and likely pathogenic variants are shown in red.

Copy number variation (CNV) analysis

To evaluate the copy number across the targeted genes and to identify potential large heterozygous or homozygous deletion, the copy number of all sequenced regions of the 702 cases and 163 controls using Ion reporter 5.0. CNVs were analyzed and then further validated by real-time PCR. However, we did not find any large deletion in all participants.

Correlations between phenotype and genotype

It is clear that sporadic AD presents high genetic heterogeneity, as at least 64 genes are implied to responsible for its pathogenesis in our study. The genotype-phenotype relationship is complex in this disease. We firstly found genes that were previously thought to be associated with specific syndromes (*COL4A3*, *COL4A4*, *COL4A5*: Alport syndrome, *COL6A3*, *COL6A6*, *COL6A5*: Bethlem myopathy, *COL18A1*: Knobloch syndrome, *COL14A1*: punctate palmoplantar keratoderma, *MMP19*: Cavitary optic disc anomalies, and *COL7A1*: Epidermolysis bullosa dystrophica) (Theocharis et al., 2016) also contribute to non-syndromic sporadic AD pathogenesis. We did not identify related symptoms in our AD patients after carefully examination. Furthermore, the same pathogenic or likely pathogenic variants were never shared by more than four patients in our study, and most of the variants were unique

Table 1 LOF and pathogenic variants in collagens and MMPs/TIMPs coding genes for 702 AD patients[a)]

Count*	Gene name	Variant function	Protein change	Coding change	Novelty
1	*COL10A1*	Initiation codon	p.Met1Thr	c.2T>C	Novel
1	*COL12A1*	Nonsense	p.Arg3035Ter	c.9103C>T	Novel
1	*COL14A1*	Nonsense	p.Arg1473Ter	c.4417C>T	Novel
1	*COL16A1*	Frameshift deletion	p.Gln992fs	c.2974_2974delC	Novel
1	*COL16A1*	Splicesite_5	–	chr1:32155995_32155996delCT	Novel
1	*COL18A1*	Frameshift insertion	p.Pro1286fs	c.3852_3853insCA	Novel
1	*COL18A1*	Frameshift insertion	p.Asp1296fs	c.3885_3886insC	Novel
1	*COL18A1*	Nonsense	p.Glu378Ter	c.1132G>T	Novel
1	*COL18A1*	Splicesite_5	–	chr21:46912599A>G	Novel
2	*COL20A1*	Frameshift insertion	p.Leu613fs	c.1835_1836insCGG	Novel
1	*COL21A1*	Frameshift insertion	p.Tyr392fs	c.1173_1174insA	Novel
1	*COL21A1*	Splicesite_3	–	chr6:55929375C>T	Novel
1	*COL24A1*	Frameshift deletion	p.Glu450fs	c.1347_1347delA	Novel
1	*COL24A1*	Nonsense	p.Arg1206Ter	c.3616C>T	rs187705574
1	*COL28A1*	Frameshift deletion	p.Ser1020fs	c.3059_3059delC	Novel
1	*COL2A1*	Splicesite_5	–	chr12:48373349_48373350delTG	Novel
2	*COL3A1*	Missense	p.Ala1259Val	c.3776C>T	CM122891**
1	*COL3A1*	Splicesite_3	–	chr2:189868871T>A	Novel
1	*COL4A3*	Frameshift Block substitution	p.Pro553fs	c.1657_1664delCCAGGTGAinsT	Novel
1	*COL4A4*	Stoploss	p.Ter1691Tyr	c.5073G>C	Novel
1	*COL5A2*	Splicesite_3	–	chr2:189922044A>T	Novel
1	*COL5A3*	Splicesite_5	–	chr19:10099854C>G	Novel
1	*COL6A3*	Nonsense	p.Gln1760Ter	c.5278C>T	Novel
1	*COL6A5*	Frameshift deletion	p.Arg1384fs	c.4146_4147delAG	Novel
1	*COL6A5*	Frameshift insertion	p.Ala568fs	c.1699_1700insG	Novel
1	*COL6A5*	Splicesite_3	–	chr3:130188326G>T	Novel
1	*COL6A5*	Splicesite_5	–	chr3:130098259A>G	Novel
4	*COL6A6*	Nonsense	p.Arg1480Ter	c.4438A>T	rs140413590
2	*COL6A6*	Splicesite_3	–	chr3:130318654_130318657de-lAGTA	Novel
1	*COL7A1*	Frameshift insertion	p.Gly2233fs	c.6696_6697insC	Novel
1	*COL7A1*	Nonsense	p.Arg1343Ter	c.4027C>T	Novel
1	*COL8A1*	Nonsense	|p.Arg225Ter	c.673C>T	Novel
1	*MMP1*	Frameshift deletion	p.Asp408fs	c.1222_1222delG	Novel
1	*MMP1*	Splicesite_3	–	chr11:102663334A>T	Novel
1	*MMP1*	Splicesite_5	–	chr11:102661539T>C	Novel
1	*MMP10*	Frameshift deletion	p.Asp193fs	c.576_580delAGATA	Novel
1	*MMP17*	Frameshift deletion	p.Tyr512fs	c.1534_1534delT	Novel
2	*MMP19*	Nonsense	p.Arg408Ter	c.1222C>T	Novel
4	*MMP19*	Nonsense	p.Gln5Ter	c.13C>T	rs17844787
1	*MMP27*	Nonsense	p.Gln330Ter	c.988C>T	Novel
1	*MMP28*	Nonsense	p.Arg391Ter	c.1173C>T	Novel
1	*MMP28*	Nonsense	p.Gln49Ter	c.145C>T	Novel
1	*MMP8*	Frameshift deletion	p.Lys39fs	c.117_117delG	Novel
1	*MMP8*	Splicesite_5	–	chr11:102584189C>T	Novel

a) *, indicates count of the variants in AD patients; **, CM122891 is the pathogenic number in HGMD database.

to single patient. These results indicate that there is no "hotspot" in this disease.

In all explained cases, 12.84% (28/218) harbored two variants, 1.83% (4/218) harbored three variants and one harbored four variants. We investigated the patients who had multiple variants and found that they suffered from AD in earlier onset age (Figure 2A), and lower percentage of DeBaKey Type 3 AD (mild type of AD) (Figure 2B).

To evaluate the effect of these AD-associated genes in different systems, we compared the average AD onset age between collagens and MMPs/TIMPs genes (Figure 2C), as well as percentage of DeBaKey Type 3 AD (Figure 2D). Results showed that there is no significant difference of AD onset age or severity between collagens and MMPs/TIMPs genes. However, if patients carried them in both systems, they were observed to suffer from worse prognosis (Figure 2C and D).

Pedigree co-segregation analysis

We performed pedigree co-segregation analysis of a 45-yeay-old-onset AD patient (AD337, II-1), who carried only a nonsense variant of *COL14A1* (Figure 3A). The father (I-1) died

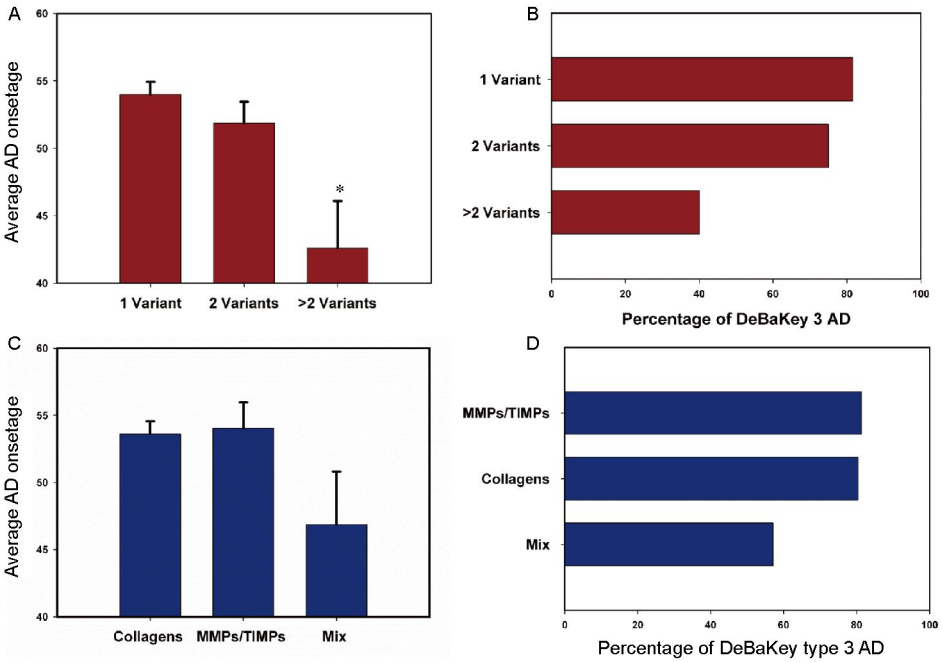

Figure 2 Correlations between phenotype and genotype. A, Average onsetage of AD patients carried different number of pathogenic or likely pathogenic variants. The error bars indicate standard error (SE). *, $P<0.05$. B, Percentage of DeBaKey type 3 AD cases in different number of pathogenic or likely pathogenic variants. C, Average onsetage of AD patients carried pathogenic or likely pathogenic variants in different genes. The error bars indicate standard error (SE). D, Percentage of DeBaKey type 3 AD cases in different genes.

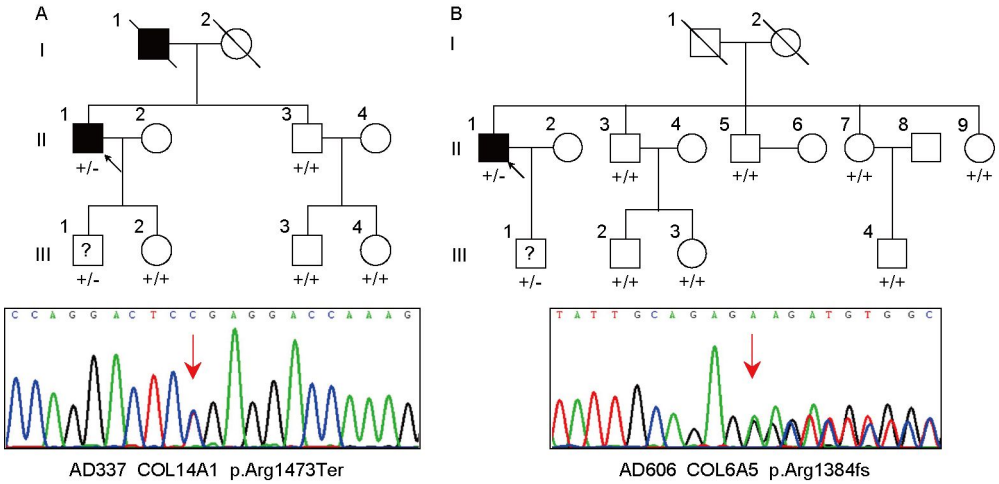

Figure 3 Pedigree co-segregation analysis. A, The pedigree co-segregation analysis of a COL14A1 nonsense variant carried AD patient (AD337). B, The pedigree co-segregation analysis of a COL6A5 frameshift deletion variant carried AD patient (AD606).

because of AD at the same age of the proband. The proband and his son (III-1), 20-year-old, carried the same variant in *COL14A1* (p.Arg1473Ter). All other their relatives are of wild type. Maybe his son is too young to develop AD. All pathogenic and likely pathogenic variants we identified in sporadic AD cases are heterozygous, implying that there might be *de novo* variants. In order to validate this assumption, we sequenced all healthy relatives of AD patient AD606 with a *COL6A5* frameshift variant, and the results supported our hypothesis (Figure 3B).

Col5a2 knockout rats display thoracic aortic dysplasia

To evaluate the pathogenesis of *COL5A2* gene defect in AD, we generated four *Col5a2* knockout rats (*Col5a2*$^{-/-}$) by inducing a large deletion containing the exon 1 of *Col5a2*. The four *Col5a2*$^{-/-}$ rats showed growth retardation. Histological analysis revealed thoracic aortic dysplasia at nine weeks. In the lesion region, the media was demonstrated loss of smooth muscle cells. The elastin fibers were loss, disarray and fragmentation in the same region. In the rest part of the loose hyperplasia media there was increased collagen deposition versus control. The adventitia is loose from the media and showed hyperplasia (Figure 4). These changes are similar to pathology of human aorta dissection.

The mRNA-expression analysis

Firstly, we detected all selected genes except for MMP7 by transcriptome sequencing in healthy control aortas. Analysis of mRNA-expression revealed that *TIMP1*, *COL6A2*, *COL1A1*, *COL1A2*, *TIMP3*, *COL3A1*, *COL4A2* and *COL4A1* were most aorta-enriched in healthy control aortas (Figure S1 in Supporting Information).

By mRNA-expression analysis, we found that seven of nine variants (one LOF variant and eight missense variants) result in significant downregulation of their corresponding mRNA in aorta wall of AD patients while one splice variant and other two missense variants result in significant upregulation of their mRNAs (Figure S2A and B in Supporting Information). Interestingly, these three upregulated genes are all most enriched genes in normal aorta, which indicate that the upregulation of these genes may largely affect the ECM structure of aorta wall.

We investigated the expressed mRNA profiles of one patient who carried FBN1 frameshift pathogenic variant and six patients who carried collagens or MMPs/TIMPs likely pathogenic variants compared with healthy controls. Importantly, we found that TIMP3 and TIMP1 were downregulated in FBN1-affected AD profile, but TIMP1 was significantly unregulated in collagen-related AD profiles. This indicates that the pathogenic mechanism of collagen-related AD is different from classic FBN1-affected AD.

Relationship between common SNP and collagen or MMPs/TIMPs related AD pathogenesis

To investigate the common SNPs' contribution of collagen and MMPs/TIMPs related AD pathogenesis, we compared

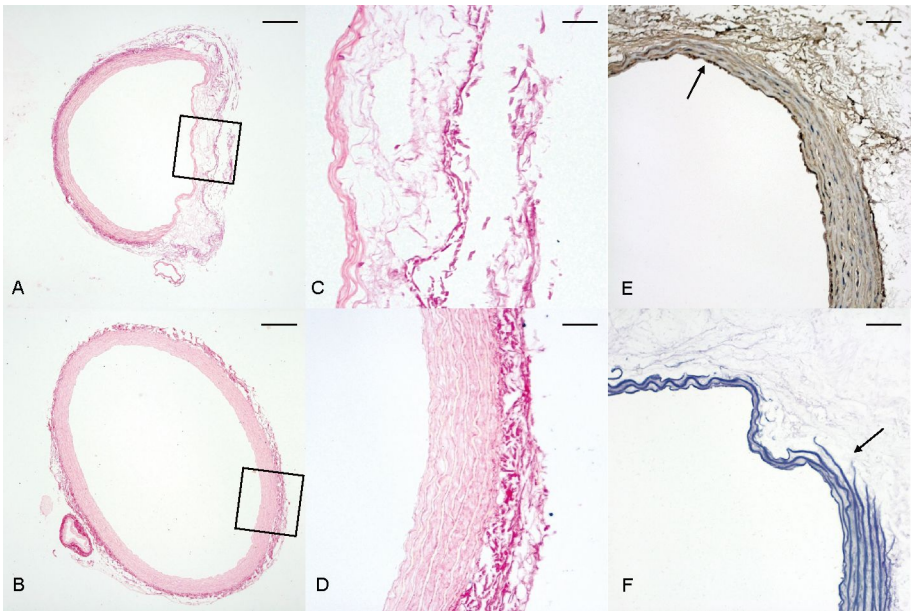

Figure 4 The histological staining of a thoracic aortic lesion region of *Col5a2*$^{-/-}$ rat and control at nine weeks. A, Sirius red staining of the thoracic aortic lesion region of *Col5a2*$^{-/-}$ rat. Bar=200 μm. B, Sirius red staining of the thoracic aortic of a wild type rat. Bar=200 μm. C, Higher magnification of the lesion region of the arterial wall of a *Col5a2*$^{-/-}$ rat showing less elastin fibers in the loose hyperplasia media. Bar=50 μm. D, Higher magnification of the control artery. Bar=50 μm. E, The thoracic aortic lesion region of a *Col5a2*$^{-/-}$ rat was stained with mouse anti-α-smooth muscle actin. Arrow indicates loss of smooth muscle cells in the lesion region of the media. Bar=50 μm. F, Van Gieson staining of the thoracic aortic lesion region of a *Col5a2*$^{-/-}$ rat. Arrow indicates loss, disarray and fragmentation of the elastin fibers in the lesion region of the media. Bar=50 μm.

the minor allele frequency of all SNPs located in exons in the 72 genes between AD cases and controls. However, there is no significantly statistical difference, which implies that rare variants might be the key points of AD pathogenesis.

DISCUSSION

To our knowledge, this is a first comprehensive targeted resequencing study specifically for genetic alterations in genes of collagens and MMPs/TIMPs system in sporadic AD patients to date. We identified a series of LOF and likely pathogenic variants in selected new genes in 702 AD patients, indicating important new causes of the pathogenesis of this disease (Figure 1 and Tables 1 and S3 in Supporting Information). Besides, pedigrees co-segregation analysis and mRNA-expression analysis further support these genes are responsible for AD pathogenesis. Finally, we demonstrate that *COL5A2* is a novel disease causing gene of sporadic AD by *Col5a2*$^{-/-}$ knock-out rat models.

The *COL5A1*, *COL5A2 and COL5A3* genes encode type V fibrillar collagen (Connizzo et al., 2015). Pathogenic variants in these genes were previously deemed to cause EDS classic type (Connizzo et al., 2015; Ritelli et al., 2013). Classic type EDS are rare autosomal dominant connective tissue disorders that are characterized by skin hyperextensibility, abnormal wound healing/atrophic scars, joint hypermobility and should exclude aortic defects (Ritelli et al., 2013). However, our study provided evidences that these genes also take part in AD pathogenesis. Similarly, *COL3A1*, *COL24A1* and *COL2A1* are also fibrillar-coding collagens, and are widely and abundantly expressed in aorta walls (Theocharis et al., 2016). Mutations of these genes and consequential structure of abnormal collagens will directly reduce the fastness of aorta.

Collagen type VI (*COL6A6*, *COL6A5*, *COL6A3*) are identified most LOF variants in our study, and the relationship with AD pathogenesis was supported by mRNA-analysis as well as pedigree co-segregation analysis. Previous study associated collagen type VI with Bethlem myopathy (Fitzgerald et al., 2013). However, our LOF variants affected AD patients did not have related symptoms or family history after carefully physical examination and inquiry. Collagen type VI and type XXVIII (*COL28A1*) belongs to Beaded-filament-forming collagens, which interact with various ECM proteins, and they are thus important in ECM protein binding and co-localization (Theocharis et al., 2016).

Collagen type XVIII (*COL18A1*) is a multiplexin, involving in production of endostatin which potently inhibits endothelial cell proliferation (Theocharis et al., 2016). mRNA-analysis found that *COL18A1* was enriched in healthy human aorta, and a splice variant in our AD patient resulted in nine times elevation in *COL18A1* mRNA in aorta compared with the controls (Figure 5A and C).

In addition, Collagen type VII (*COL7A1*) is an anchoring fibril; type X (*COL10A1*), VIII (*COL8A1*) and IV (*COL4A3*, *COL4A4*) are networking-forming collagens; type XII (*COL12A1*), XIV (*COL14A1*), XVI (*COL16A1*), XX (*COL20A1*), XXI (*COL21A1*) are FACITs (fibril-associated collagens with interrupted triple helices) (Theocharis et al., 2016). They anchor or connect with various ECM elements, and provide aorta capable and flexible. The LOF variants in these genes may break the inner connection of ECM, thereupon interrupt the homeostasis of ECM.

The MMPs/TIMPs have robust associations with AD in association studies (Song et al., 2013; Wang et al., 2014; Zhang et al., 2009), indicating that functional variants within MMPs/TIMPs genes may lead to degrading of various components of the aortic ECM or result in aorta inflammation, therefore, playing an important role in AD pathogenesis (Theruvath et al., 2012). However, none of the MMPs coding genes have been regarded as AD-causing gene in the past. Our data identified LOF (four frameshift deletion; five nonsense; three splice site) and likely pathogenic variants in MMPs in AD patients for the first time.

Although our study identify a series of LOF and like pathogenic variants in sporadic AD patients, missing heritability still remains, which indicates that targeted sequencing is not enough to find full spectrum of sporadic AD. More comprehensive strategies such as whole exome sequencing or whole genome sequencing should be considered in future studies.

In conclusion, this is a first study on targeted resequencing of the genes in all collagens and MMPs/TIMPs system in large scale AD patients to date, and provide a comprehensive map of genetic alternations in this system. We identified *COL5A2* as novel sporadic AD pathogenic genes. We also highlight collagen type VI (*COL6A6*, *COL6A5*, *COL6A3*), *COL18A1*, and *MMP19* genes in AD pathogenesis. Our data also indicate that next-generation targeted resequencing may be a powerful and effective method for the genetic diagnosis of AD, which can be used for risk stratification among patients' first-degree relatives.

MATERIALS AND METHODS

Study design and eligibility

In this study, we recruited two cohorts of Han Chinese descent, including 702 sporadic AD patients (cases) and 163 matched unrelated healthy volunteers (controls). Recruitment of the participants was conducted at the Tongji Hospital. The cases were selected from patients with AD as defined by computed tomography (CT) or magnetic resonance imaging. The controls were randomly recruited from healthy individuals undergoing routine health examinations. All controls were determined to be free of AD by medical history inquiry and definitive imaging examinations including ultrasonography and CT. All enrolled participants underwent a review of medical history, physical examinations, several

laboratory assessments of AD risk factors and a consultation to determine their ethnic origin.

This study was approved by the Institutional Review Board of Tongji Hospital. Written informed consent was obtained from all participants. All the experiments were conducted according to the principles expressed in the Declaration of Helsinki.

Gene selection

To systemically study genetic defects of genes in collagens and MMPs/TIMPs system in pathogenic process in AD, we collected all thus far known human genes coding collagens and MMPs/TIMPs for resequencing, including 45 collagen coding genes, 23 MMPs coding genes and four TIMPs coding genes (Table S3 in Supporting Information).

Targeted semiconductor sequencing

Genomic DNA was extracted from blood samples of all participants and was diluted to 5 ng μL^{-1}. We performed next-generation semiconductor sequencing using the Ion torrent platform. Briefly, adapter-ligated libraries were constructed using the Ion AmpliSeq™ Library Kit 2.0 and custom-ordered multiplex PCR primer pools (Life Technologies, USA) following the manufacturer's protocol. After purification, libraries were quantitated using a Qubit 2.0 fluorometer (Invitrogen, USA) and were combined at equal molar ratios for sequencing as previously described (Li et al., 2014).

Bioinformatic analysis

The next-generation sequencing data were initially processed using the Ion Torrent platform-specific software Torrent Suite v5.0. The sequences were aligned to the hg19/GRCh37 human reference genome to analyze the coverage and call variants. Then, all variants were annotated using Ion reporter 5.0. We removed common variants in the 1000-Genomes-Project database, the UCSC common SNP database, the Exome Aggregation Consortium (ExAC) database, and the Exome Sequencing Project (ESP). To predict the possible impact of detected non-synonymous variants in the exons on the encoded proteins, all the missense substitutions were scored and predicted using SIFT and PolyPhen-2. The degrees of conservation across multiple species of the non-synonymous variants were estimated by the Grantham and phyloP scores.

Criteria for pathogenic and likely pathogenic variants

We divided the AD-associated variants into "pathogenic" and "likely pathogenic" according to the guideline (Richards et al., 2015). The following criteria were used to determine that a variant was pathogenic: it should be validated by Sanger sequencing; and it should have been reported as causal variants in the NCBI ClinVar database or HGMD database.

The following criteria were used to determine that a variant was likely pathogenic: it should be found in one or more cases and be absent in all healthy controls; it should be validated by Sanger sequencing; it should not be found as common SNP in the 1000-Genomes-Project database, the UCSC common SNP database, the ExAC database, and the ESP; it should be an exonic non-synonymous variant or canonical±2 splice sites; if it is a missense substitution, it should be predicted conservative by Grantham or phyloP score; and both of the SIFT and Polyphen-2 bioinformatic predictions should be damaging.

Sanger sequencing validation

PCR amplification was optimized in accordance with the manual for Taq™ Hot Start version (TaKaRa, Japan). To exclude any possible false-positive errors, Sanger sequencing was performed for all pathogenic or likely pathogenic variants using the Big Dye v.1.1 terminator cycle sequencing kit and an Applied Biosystems 3500xl capillary sequencer (Applied Biosystems, CA). The technically uncovered 3,972 bp regions of the 72 selected genes were sequenced directly by Sanger sequencing.

Statistical analysis

Fisher exact test was used to assess the difference in variant frequencies between cases and controls. The statistical analysis was performed with the use of SPSS software, version 20.0 (SPSS). A two-sided *P* value of less than 0.05 was considered to indicate statistical significance.

Gene expression analysis

We performed transcriptome sequencing and real-time PCR to evaluate the tissue expression of some identified novel genes that may be responsible for AD pathogenesis in aorta tissue from seven AD patients who underwent aortic surgery. Total RNA was extracted from frozen human aorta tissue using a TRIzol Reagent (Invitrogen). For real-time PCR, total RNA was reverse-transcribed to cDNA using M-MLV First Strand Kit (Invitrogen), and amplified using Power SYBR Green PCR Master Mix (Invitrogen) on Step One Plus real time PCR system (Applied Biosystems). We performed comprehensive RNA sequencing of aorta samples of AD patients, as well as healthy controls using Ion AmpliSeq™ Transcriptome Human Gene Expression Research Panel (Life Technologies) following the manufacturer's protocol (https://ion-community.thermofisher.com/docs/DOC-8980), and analysis the data by using Ion reporter 5.0.

Preparation of *Col5a2* knockout rats

SD rats were bred in standard cages in an Assessment and Accreditation of Laboratory Animal Care-accredited SPF animal facility. All animal protocols were approved by the Animal Care and Use Committees of the Institute of Laboratory Animal Science of Chinese Academy of Medical Sciences (ILAS-GC-2015-002).

The paired synthesized oligonucleotides for sgRNAs were annealed and cloned into the pUSD-sgRNA expression vector (#51132, Addgene, USA) before *in vitro* transcription and microinjection. The Cas9 expression plasmid (#44758, Addgene) was linearized with *Age* I and used as the template for *in vitro* transcription using the T7 Ultra Kit (AM1345, Ambion, USA). sgRNA expression plasmids were linearized with *Dra* I and used as templates for *in vitro* transcription using the MEGAshortscript Kit (AM1354, Ambion). Transcribed Cas9 mRNA and sgRNA were both purified by using the MEGAclear Kit (AM1908, Ambion). SD rats were purchased from Beijing Vital River Laboratories animal center are housed in standard cages and maintained on a 12-h light/dark cycle with food and water. The microinjection of fertilized rat eggs was performed in Zhang's lab, Institute of Laboratory Animal Science, Chinese Academy of Medical Sciences. New pups genomic DNA was extracted from the tails of 7-day-old rats for sequencing. Rats with wanted mutations were prepared for consequence experiment.

Compliance and ethics *The author(s) declare that they have no conflict of interest.*

Acknowledgements *We thank Drs. Mingjia Ma and Zeming Fang for their efforts in recruiting patients for this study. We especially acknowledge all the participants in this study. This work was supported by the National Natural Science Foundation of China (91439203), and National Key Basic Research Program of China (2012CB518004, 2012CB517801).*

Barbour, J.R., Spinale, F.G., and Ikonomidis, J.S. (2007). Proteinase systems and thoracic aortic aneurysm progression. J Surgical Res 139, 292–307.

Campens, L., Callewaert, B., Muiño Mosquera, L., Renard, M., Symoens, S., De Paepe, A., Coucke, P., and De Backer, J. (2015). Gene panel sequencing in heritable thoracic aortic disorders and related entities—results of comprehensive testing in a cohort of 264 patients. Orphanet J Rare Dis 10, 9.

Connizzo, B.K., Freedman, B.R., Fried, J.H., Sun, M., Birk, D.E., and Soslowsky, L.J. (2015). Regulatory role of collagen V in establishing mechanical properties of tendons and ligaments is tissue dependent. J Orthop Res 33, 882–888.

Fitzgerald, J., Holden, P., and Hansen, U. (2013). The expanded collagen VI family: new chains and new questions. Connect Tissue Res 54, 345–350.

Gao, Y., Wu, W., Yu, C., Zhong, F., Li, G., Kong, W., and Zheng, J. (2016). A disintegrin and metalloproteinase with thrombospondin motif 1 (ADAMTS1) expression increases in acute aortic dissection. Sci China Life Sci 59, 59–67.

Goldfinger, J.Z., Halperin, J.L., Marin, M.L., Stewart, A.S., Eagle, K.A., and Fuster, V. (2014). Thoracic aortic aneurysm and dissection. J Am Coll Cardiol 64, 1725–1739.

Guo, D.C., Regalado, E.S., Minn, C., Tran-Fadulu, V., Coney, J., Cao, J., Wang, M., Yu, R.K., Estrera, A.L., Safi, H.J., Shete, S.S., and Milewicz, D.M. (2011). Familial thoracic aortic aneurysms and dissections: identification of a novel locus for stable aneurysms with a low risk for progression to aortic dissection. Circ Cardiovasc Genet 4, 36–42.

Hoffjan, S. (2012). Genetic dissection of marfan syndrome and related connective tissue disorders: an update 2012. Mol Syndromol 3, 47–58.

Khau Van Kien, P., Wolf, J.E., Mathieu, F., Zhu, L., Salve, N., Lalande, A., Bonnet, C., Lesca, G., Plauchu, H., Dellinger, A., Nivelon-Chevallier, A., Brunotte, F., and Jeunemaitre, X. (2004). Familial thoracic aortic aneurysm/dissection with patent ductus arteriosus: genetic arguments for a particular pathophysiological entity. Eur J Hum Genet 12, 173–180.

LeMaire, S.A., McDonald, M.L.N., Guo, D.C., Russell, L., Miller, C.C., Johnson, R.J., Bekheirnia, M.R., Franco, L.M., Nguyen, M., Pyeritz, R.E., Bavaria, J.E., Devereux, R., Maslen, C., Holmes, K.W., Eagle, K., Body, S.C., Seidman, C., Seidman, J.G., Isselbacher, E.M., Bray, M., Coselli, J.S., Estrera, A.L., Safi, H.J., Belmont, J.W., Leal, S.M., and Milewicz, D.M. (2011). Genome-wide association study identifies a susceptibility locus for thoracic aortic aneurysms and aortic dissections spanning FBN1 at 15q21.1. Nat Genet 43, 996–1000.

Li, Z., Huang, J., Zhao, J., Chen, C., Wang, H., Ding, H., Wang, D.W., and Wang, D.W. (2014). Rapid molecular genetic diagnosis of hypertrophic cardiomyopathy by semiconductor sequencing. J Transl Med 12, 173.

Marian, A.J., and Belmont, J. (2011). Strategic approaches to unraveling genetic causes of cardiovascular diseases. Circul Res 108, 1252–1269.

Nienaber, C.A., and Eagle, K.A. (2003). Aortic dissection: new frontiers in diagnosis and management: part I: from etiology to diagnostic strategies. Circulation 108, 628–635.

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E., Spector, E., Voelkerding, K., Rehm, H.L., and Rehm, H.L. (2015). 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 17, 405–423.

Ritelli, M., Dordoni, C., Venturini, M., Chiarelli, N., Quinzani, S., Traversa, M., Zoppi, N., Vascellaro, A., Wischmeijer, A., Manfredini, E., Garavelli, L., Calzavara-Pinton, P., and Colombi, M. (2013). Clinical and molecular characterization of 40 patients with classic Ehlers-Danlos syndrome: identification of 18 *COL5A1* and 2 *COL5A2* novel mutations. Orphanet J Rare Dis 8, 58.

Ritelli, M., Drera, B., Vicchio, M., Puppini, G., Biban, P., Pilati, M., Prioli, M.A., Barlati, S., and Colombi, M. (2009). Arterial tortuosity syndrome in two Italian paediatric patients. Orphanet J Rare Dis 4, 20.

Sakai, H., Suzuki, S., Mizuguchi, T., Imoto, K., Yamashita, Y., Doi, H., Kikuchi, M., Tsurusaki, Y., Saitsu, H., Miyake, N., Masuda, M., and Matsumoto, N. (2012). Rapid detection of gene mutations responsible for non-syndromic aortic aneurysm and dissection using two different methods: resequencing microarray technology and next-generation sequencing. Hum Genet 131, 591–599.

Salomon, M.P., Li, W.L.S., Edlund, C.K., Morrison, J., Fortini, B.K., Win, A.K., Conti, D.V., Thomas, D.C., Duggan, D., Buchanan, D.D., Jenkins, M.A., Hopper, J.L., Gallinger, S., Le Marchand, L., Newcomb, P.A., Casey, G., and Marjoram, P. (2016). GWASeq: targeted re-sequencing follow up to GWAS. BMC Genomics 17, 176.

Smith, L.B., Hadoke, P.W.F., Dyer, E., Denvir, M.A., Brownstein, D., Miller, E., Nelson, N., Wells, S., Cheeseman, M., and Greenfield, A. (2011). Haploinsufficiency of the murine Col3a1 locus causes aortic dissection: a novel model of the vascular type of Ehlers-Danlos syndrome. Cardiovasc Res 90, 182–190.

Song, Y., Xie, Y., Liu, F., Zhao, C., Yu, R., Ban, S., Ye, Q., Wen, J., Wan, H., Li, X., Ma, R., and Meng, Z. (2013). Expression of matrix metalloproteinase-12 in aortic dissection. BMC Cardiovasc Disord 13, 34.

Suzuki, T., Distante, A., Zizza, A., Trimarchi, S., Villani, M., Salerno Uriarte, J.A., De Luca Tupputi Schinosa, L., Renzulli, A., Sabino, F., Nowak, R., Birkhahn, R., Hollander, J.E., Counselman, F., Vijayendran, R., Bossone, E., Eagle, K., and Eagle, K. (2009). Diagnosis of acute aortic dissection by D-dimer: the international registry of acute aortic dissection substudy on biomarkers (IRAD-Bio) experience. Circulation 119, 2702–2707.

Theocharis, A.D., Skandalis, S.S., Gialeli, C., and Karamanos, N.K. (2016). Extracellular matrix structure. Adv Drug Deliver Rev 97, 4–27.

Theruvath, T.P., Jones, J.A., and Ikonomidis, J.S. (2012). Matrix metalloproteinases and descending aortic aneurysms: parity, disparity, and switch. J Cardiac Surgery 27, 81–90.

Tsamis, A., Krawiec, J.T., and Vorp, D.A. (2013). Elastin and collagen fibre microstructure of the human aorta in ageing and disease: a review. J R

Soc Interface 10, 20121004–20121004.
Wang, X., Liu, O., Qin, Y., Zhang, H., and Lv, Y. (2014). Association of the polymorphisms of MMP-9 and TIMP-3 genes with thoracic aortic dissection in Chinese Han population. Acta Pharmacol Sin 35, 351–355.
Zhang, X., Shen, Y.H., and LeMaire, S.A. (2009). Thoracic aortic dissection: are matrix metalloproteinases involved? Vascular 17, 147–157.

SUPPORTING INFORMATION

Figure S1 Normalized expression level of top 15 genes.

Figure S2 The mRNA differentially expression level of 12 novel identified LOF or likely pathogenic variants in AD cases.

Table S1 Known pathogenic genes for AD

Table S2 Likely pathogenic variants in collagens and MMPs/TIMPs coding genes for 702 AD patients

Table S3 Collagens and MMPs/TIMPs coding genes

The supporting information is available online at life.scichina.com and www.springerlink.com. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.