Genome-wide Association Study of Thoracic Aortic Aneurysm and Dissection in the Million Veteran Program

Derek Klarin (dklarin@stanford.edu)
Stanford https://orcid.org/0000-0002-4636-5780
Poornima Devineni
Massachusetts Veterans Epidemiology and Research Information Center, Veterans Affairs Boston Healthcare System
Anoop Sendamarai
Center for Data and Computational Sciences, VA Boston Healthcare System, Boston, MA 02130, USA
https://orcid.org/0000-0002-0476-5428
Anthony Angueira
Institute for Diabetes, Obesity & Metabolism, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA
Sarah Graham
University of Michigan - Department of Internal Medicine Cardiology https://orcid.org/0000-0003-1271-2489
Ying Shen
Baylor College of Medicine
Michael Levin
University of Pennsylvania https://orcid.org/0000-0002-9937-9932
James Pirruccello
Massachusetts General Hospital https://orcid.org/0000-0001-6088-4037
Ida Surakka
University of Michigan
Purushotham Karnam
Center for Data and Computational Sciences, VA Boston Healthcare System, Boston, MA
Tanmoy Roychowdhury
Department of Internal Medicine, Division of Cardiology, University of Michigan, Ann Arbor, MI
Yanming Li
Baylor College of Medicine https://orcid.org/0000-0002-8213-9166
Minxian Wang
Broad Institute https://orcid.org/0000-0002-3753-508X
Krishna Aragam
Massachusetts General Hospital
Kaavya Paruchuri
Massachusetts General Hospital
Gabrielle Shakt
Division of Cardiovascular Medicine, Department of Medicine, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA
Noah Tsao
University of Pennsylvania Perelman School of Medicine

Renae Judy
Department of Surgery, Perelman School of Medicine, University of Pennsylvania, Philadelphia,USA
Ha My Vy
Icahn School of Medicine at Mount Sinai
Shefali Verma
University of Pennsylvania https://orcid.org/0000-0001-5216-4670
Daniel Rader
Perelman School of Medicine at the University of Pennsylvania https://orcid.org/0000-0002-9245-9876
Ron Do
Icahn School of Medicine at Mount Sinai https://orcid.org/0000-0002-3144-3627
Joseph Bavaria
Division of Cardiovascular Surgery, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA
Girish Nadkarni
Icahn School of Medicine at Mount Sinai https://orcid.org/0000-0001-6319-4314
Marylyn Ritchie
University of Pennsylvania, Philadelphia, Pennsylvania, USA. https://orcid.org/0000-0002-1208-1720
Dongchuan Guo
University of Texas Houston
Patrick Ellinor
The Broad Institute of MIT and Harvard https://orcid.org/0000-0002-2067-0533
Scott LeMaire
Texas Heart Institute
Dianna Milewicz
McGovern Medical School, University of Texas Health Science Center at Houston, TX 77030 https://orcid.org/0000-0002-7806-0068
Cristen Willer
University of Michigan https://orcid.org/0000-0001-5645-4966
Pradeep Natarajan
Massachusetts General Hospital https://orcid.org/0000-0001-8402-7435
Phil Tsao
Stanford University School of Medicine https://orcid.org/0000-0001-7274-9318
Saiju Pyarajan
Massachusetts Veterans Epidemiology and Research Information Center, Veterans Affairs Boston Healthcare System
Scott Damrauer
University of Pennsylvania

Article

Keywords:

Posted Date: April 8th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1507463/v1

Version of Record: A version of this preprint was published at Nature Genetics on June 12th, 2023. See the published version at https://doi.org/10.1038/s41588-023-01420-z.

Abstract

Current understanding of the genetic contribution to thoracic aortic aneurysms and dissections (TAAD) has largely been informed through studies focusing on rare, Mendelian forms of disease. In the current analysis we performed the largest genome-wide association study (GWAS) of TAAD to date, testing ~25 million DNA sequence variants in 8,626 participants with (7,050 European, 1,266 African, and 310 Hispanic ancestry individuals) and 453,043 participants without TAAD in the Million Veteran Program. The results were replicated in an independent sample of 4,459 individuals with and 512,463 without TAAD from 6 cohorts. We identified 21 TAAD loci, 17 of which have not been previously reported. We leverage phenome-wide scanning and causal inference methods to provide evidence that TAAD is a non-atherosclerotic aortic disorder distinct form other forms of vascular disease; we subsequently use Bayesian techniques and single cell/nucleus RNA sequencing data to identify causal risk genes and cell types. Finally, we generate a TAAD polygenic risk score and demonstrate that those at the top 5% of polygenic risk are over 4-fold more likely to experience an incident, fatal TAAD event. Our results demonstrate that the genetic architecture of TAAD mirrors that of other complex traits and is not solely inherited through protein altering variants of large effect size. These data provide new mechanistic insights into TAAD risk and improve our understanding of thoracic aortic disease.

Introduction

Thoracic aortic aneurysms and dissections (TAAD) encompass a spectrum of aortic pathology affecting the aortic root, ascending aorta, aortic arch, and descending thoracic aorta. Thoracic aortic aneurysms (TAA), a dilation of the proximal aorta, are known to progressively enlarge over time ultimately leading to rupture and death if not surgically repaired. In addition, dissections of the ascending (Stanford Type A) or descending thoracic (Stanford Type B) are life-threatening condition requiring emergency treatment, often including surgical repair, and are associated with high short and long-term mortality risk[1,2]. Despite the lethality of these conditions, the genetic determinants of TAAD remain largely unknown, with published genome-wide associations studies (GWAS) having revealed only 4 loci reaching genome-wide significance[3-5]. As a result, most of what is understood about the genetics of TAAD has been derived from studies examining rare, pathogenic variants resulting in heritable aortopathy (so called, hereditary TAAD or "HTAAD")[6].

The Million Veteran Program (MVP) is a genomic and precision medicine cohort established in 2011 by the Department of Veterans Affairs (VA) Healthcare System to study how genes affect health and disease. We recently demonstrated that a VA Healthcare System-based biobank can aid in the genetic discovery of aortic disease[7], and allows for the elucidation of causal biology and mechanisms. Leveraging the MVP resource, we sought to: 1) perform a genetic discovery analysis for TAAD across multiple ancestries; 2) explore the spectrum of phenotypic consequences associated with TAAD risk variants, 3) examine the genetic relationship between TAAD and its epidemiologic risk factors, 4) map causal variants and genes for disease, 5) identify causal tissues and cell types, and 6) construct and test a polygenic risk score (PRS) for TAAD (**Figure 1**).

Results

Common Variants Associated with TAAD

We designed a two-phased Genome-wide Association Study (GWAS). The initial MVP discovery analysis was comprised of 8,626 individuals (7,050 European, 1,266 African, and 310 Hispanic ancestry participants) with TAAD and 453,043 disease-free individuals from the same ancestral groups (**Supplementary Fig. 1)**; their baseline

characteristics are presented in **Supplementary Table 1.** Participants with TAAD were more likely to be older, male, prescribed statin therapy, and former smokers.

Through genotype imputation, we obtained 25.4 million, 40.3 million, and 34.9 million DNA sequence variants for analysis in participants of European, African, and Hispanic ancestry, respectively (**Supplementary Table 1**). Following multi-ancestry meta-analysis in the discovery phase, a total of 1,465 variants at 25 loci met a genome-wide significance threshold ($P < 5x10^{-8}$, **Supplementary Fig. 2 and 3**). We replicated the known *FBN1*[3], *ULK4*[4], and *LRP1*[4] loci at genome-wide significance, and the recently identified *TCF7L2*[5] locus with $P < 5x10^{-5}$ (**Supplementary Table 2**). Notably, in MVP we found no evidence of association for 3 variants previously reported in an analysis of 435 thoracic aortic aneurysm cases that lacked independent replication[8], suggesting these rare variant associations may be false positive findings (**Supplementary Table 3**).

Of the 1,465 variants reaching genome-wide significance in MVP, 1,461 were also available for independent testing in external datasets (4,459 individuals with TAAD, 512,463 individuals without TAAD across 6 cohorts of predominantly European ancestry) and were taken forward for replication. Following replication, 21 loci continued to exceed genome-wide significance ($P < 5x10^{-8}$), with the 4 known and 17 novel loci demonstrating a directionally consistent replication $P < 0.05$ (**Table 1, Supplementary Tables 4 and 5**). The *FBN1* variant rs1818275 was the top association result (17.4% frequency for the C allele; OR =1.35; 95%CI: 1.41-1.30; $P = 2.5x10^{-47}$). All of the 5 signals that did not replicate failed to meet the pre-specified $P < 0.05$ for independent replication (**Supplementary Table 6**).

Of the 21 TAAD loci, 16 were directionally consistent across European, African, and Hispanic ancestries in MVP, 13 demonstrated at least nominal significance in African ancestry individuals and 6 in Hispanic ancestry participants (**Supplementary Table 7**).

Anatomic Distribution and Phenotypic Consequences of TAAD Risk Variants

We next explored whether the identified TAAD risk variants were associated with a specific anatomic distribution of aortic disease. We tested the 21 TAAD risk variants for association with ascending/arch, or descending thoracic aortic repair using Current Procedural Terminology (CPT) codes, and for association with aortic dissection (both Stanford Type A and B) based on diagnosis code data. We found that certain genetic variants demonstrated an association with a distinct type of aortic pathology (**Figure 2, Supplementary Table 8**). For example, the *ADAMTS8* locus was more significantly associated with descending thoracic aortic repairs and aortic dissection, while the *FBN1* locus was associated with disease across all 3 phenotypes. In total, 9 of the 21 TAAD loci were associated with 2 or more specific subtypes of aortic disease.

Understanding the full spectrum of phenotypic consequences of a given DNA sequence variant can help identify the mechanism by which a variant or gene leads to disease. Termed a phenome-wide association study (PheWAS), this approach examines the association of a risk variant across a range of phenotypes[9,10]. Using data from the Integrative Epidemiology Unit (IEU) open GWAS project[11,12], we performed a PheWAS of the 21 TAAD lead risk variants across a range of over 2,000 conditions, disease phenotypes, and metabolites. We found that several of the newly identified DNA sequence variants correlated with a range of anthropometric traits such as height and conditions including asthma and migraine at genome-wide significance (**Supplementary Table 9).** Notably, for 7 of the 21 variants the TAAD risk allele was associated with *increased* diastolic blood pressure (DBP). In contrast, for all 9 variants that demonstrated an association with systolic blood pressure (SBP), the TAAD risk allele was associated with *decreased* SBP. In a sensitivity analysis, for the 7 variants demonstrating a PheWAS association with increased DBP, we re-tested

their association with TAAD accounting for DBP using summary-based methods[13]. While some abrogation in association signal was observed, in each case the TAAD association P value remained genome-wide significant suggesting that blood pressure was not the primary mediator for the observed genetic association (**Supplementary Table 10**). In total, we identified 326 statistically significant ($P < 5.0x10^{-8}$) PheWAS associations across the 21 genetic variants.

Causal Epidemiologic Risk Factors for TAAD

In observational studies smoking, hyperlipidemia, hypertension, and standing height have been suggested as independent risk factors for TAAD[14,15]. We performed Mendelian randomization analyses using genetic instruments for a lifetime smoking index[16], lipids (triglycerides, HDL/LDL cholesterol)[17], blood pressure (SBP and DBP)[18], and height[19] (**Supplementary Table 11**). Consistent with the epidemiologic literature, we observed that a 1 standard deviation genetic increase in lifetime smoking index, diastolic blood pressure, and height were associated with increased risk of TAAD [OR 1.42 for smoking ~20 cigarettes a day for 15 years and stopping 17 years ago; 95% CI 1.14-1.77; P = 0.002; OR 1.88 per 10 mmHg increase in DBP; 95% CI 1.62-2.18; $P = 4.0 \times 10^{-17}$; OR 1.23 per 7.6cm increase in height; 95% CI 1.15-1.33; $P = 3.0 \times 10^{-8}$, two-sided Bonferroni P < 0.007 (**Figure 3)**]. Our results remained robust to multiple sensitivity analyses, including the weighted median[20] as well as MR-PRESSO[21] and MR-Egger[22] tests for evidence of horizontal pleiotropy (**Supplementary Table 12**). We did not detect a significant association between a genetic 10 mmHg increase in SBP and TAAD risk or an association between any lipid fraction and TAAD.

Identification of Candidate Causal TAAD Risk Genes

We next sought to identify causal TAAD risk genes and variants. Prior human genetic evidence strongly suggests that *FBN1*[23], *ELN*[24], and *LRP1*[25,26] are the causal genes at 3 of the identified loci, and a recent report provided evidence that *TCF7L2* is the likely causal gene acting at the locus[5] (**Supplementary Table 13**). For the remaining loci, we examined the genetic literature and possible causal variants that result in a protein altering consequence (missense mutations) in high LD with the lead variant ($R^2 > 0.8$). In addition, we hypothesized that TAAD risk variants may be acting by inducing expression changes locally in the proximal aortic wall and performed a fine-mapping TWAS[27] and colocalization analyses using aortic expression quantitative trait locus (eQTL) data from the Genotype Tissue Expression Project (GTEx)[28]. When combining the above strategies, we identified 7 additional putative causal genes: *THSD4*, *COL6A3*, *CDH13*, *NOC3L*, *SCAI*, *PRDM6*, and *ADAMTS8* (**Supplementary Tables 13,14**). At 5 of the putative causal genes, we were also able to fine-map the locus to five or fewer causal variants (**Supplementary Table 15**).

Notably, we observed that decreased *THSD4* expression was associated with an increased risk of TAAD. The protein product of *THSD4*, ADAMTSL6, is a known microfibril-associated protein that promotes fibrillin-1 matrix assembly[29]. In recent report, rare deleterious *THSD4* variants segregated in families with a history of thoracic aortic aneurysms, and *Thsd4*$^{+/-}$mice were found to have progressive thoracic aortic dilation[30]. Taken together, these results suggest that common *THSD4* variants may cause TAAD through diminished gene expression in the thoracic aorta.

Gene Expression Analyses Reveal TAAD-Relevant Cell Types

We next sought to identify the critical tissues and cell types for TAAD risk variants. At the genome-wide level, we first used stratified LD score regression[31] to identify TAAD relevant tissues and cell types. We combined publicly available expression data from GTEx[28] and an aggregation of microarray gene expression data sets comprising 37,427 samples in human, mouse, and rat[31] (previously referred to as the "Franke lab dataset") to evaluate for significant enrichment of a specific tissues or cell types with genetic TAAD risk association signals across 205 different tissues/cell types. Not surprisingly, the aorta demonstrated the strongest enrichment ($P_{enrichment} = 6.2x10^{-6}$, **Supplementary Table 16**). Interestingly, we also observed an enrichment for uterine myometrium, chondrocytes, and osteoblasts (Bonferroni $P_{enrichment} < 0.002$). Human myometrium is primarily composed of smooth muscle cells, consistent with the well-recognized critical role of vascular smooth muscle cells (VSMCs) in TAAD pathogenesis. Much like fibroblasts, chondrocytes produce and maintain collagen and proteoglycans, and in cell culture osteoblasts are nearly indistinguishable from fibroblasts[32] – providing evidence for a key role of fibroblasts in TAAD development.

We next examined whether a putative causal TAAD risk gene set demonstrated a significant cell type enrichment. We generated a gene set of 21 putative causal TAAD risk genes: the 11 candidate causal genes identified from our GWAS, as well as 10 additional previously identified definitive or strong HTAAD genes ("Category A")[6]. Of note, *FBN1* was present both in our GWAS gene set and the HTAAD gene set. We then tested whether the 21 putative causal TAAD risk genes were enriched in specific cell types identified from publicly available single nuclear RNA expression data (snRNA) from ascending and descending thoracic aorta[33]. Using the fast gene set enrichment software, we observed that the causal TAAD gene set was significantly enriched in VSMC (P = 0.0078), and a suggestive enrichment was also observed with fibroblasts (P = 0.02), though this was no longer significant after Bonferroni correction (P < 0.01 = 0.05/5 cell types; **Supplementary Table 17)**. These findings are consistent with our results from the stratified LD score regression analysis highlighting VSMCs and fibroblasts as causal TAAD cell types.

Finally, we sought to identify relevant cell types for the individual putative causal TAAD risk genes leveraging scRNA sequencing data from normal and aneurysmal ascending aortic aneurysm tissue[34] and snRNA sequencing data from normal ascending and descending thoracic aorta[33]. For the 9 genes we hypothesized promote TAAD risk through changes in aortic gene expression ("candidate causal genes through changes in expression" in **Supplementary Table 14**), we first qualitatively assessed for a prioritized cell type through: 1) the percentage of cells expressing the gene in a given cell type cluster, and 2) the magnitude of average gene expression in each cell type (**Figure 4a,b,c**). We then tested the differential expression of each gene among putative cell type clusters between ascending aortic aneurysm and normal aortic tissue samples using *Seurat*[35] (**Figure 4d,e**). Integrating this evidence, we prioritized a series of causal cell types for each gene (**Figure 4f**). For example, *COL6A3* was prioritized to be acting in fibroblasts, consistent with a reported involvement in smooth muscle cell-elastin contact within the aortic wall[36]. *CDH13* (encoding cadherin-13) was prioritized to be acting in fibroblasts, mesenchymal-stromal cells, and endothelial cells. Cadherin-13 signaling has been shown to be protective for endothelial cells in the setting of oxidative stress[37], and its reported role in angiogenesis[38] suggests it may play a role in aortic remodeling during aneurysmal degeneration across multiple cell types. In total, we prioritized at least one candidate causal cell type for 8 of the 9 genes thought to be acting through changes in gene expression in the thoracic aorta.

Polygenic Risk Score Generation for TAAD

Lastly, we sought to examine the contribution of polygenic inheritance on TAAD risk. We generated TAAD PRS including 1,189,073 variants from the MVP discovery GWAS summary statistics (8,626 multi-ancestry TAAD

cases, 453,043 controls) and a linkage disequilibrium panel from 1000 Genomes[39] whole genome sequencing data. To increase the number of independent variants included in our score, we used the PRS-CSx software which uses Bayesian methods to generate posterior genetic variants effect sizes under coupled continuous shrinkage priors[40]. We first validated the PRS using prevalent data from the MassGeneral Brigham Biobank (775 cases, 24,518 controls of European ancestry). We observed that the TAAD PRS was strongly associated with prevalent TAAD, with a 1 standard deviation increase in PRS associated with a 57% increased risk of disease (OR_{PRS} = 1.57, 95% CI =1.46-1.69, P_{PRS} =4.6x10^{-32}). Individuals with a PRS in the 95th percentile or higher were 2.67 times more likely to be diagnosed with TAAD (OR = 2.67, 95% CI =2.11-3.39, P_{PRS} =6.3x10^{-16}, **Figure 5a**). After this initial validation step, we then tested this PRS in two additional cohorts.

We first assessed the performance of the PRS in European-ancestry individuals in the Cardiovascular Health Improvement Project(CHIP)/Michigan Genomics Initiative(MGI) cohort (3,743 cases and 51,898 controls). We again observed an effect estimate greater than 2.0 for the top 5% PRS (OR = 2.28, 95% CI =2.03-2.56, P_{PRS} =1.8x10^{-47}, **Fig 4a**). We next restricted the data to those individuals with targeted or exome sequencing available (1,842 cases and 1,887 controls), and compared the increase in AUC afforded from the PRS and a set of rare TAAD risk variants that were manually curated as “pathogenic or likely pathogenic” for HTAAD according to ACMG best practices[41]. While the effect estimate of these pathogenic variants was substantially larger than that observed for the top 5% PRS ($OR_{pathogenic}$ = 11.1; P = 1.4 x10^{-10}), we noted that when modeling TAAD risk, the addition of the PRS improved the AUC value in a similar amount as the presence of a pathogenic TAAD risk variant (**Supplementary Table 18**). In a sensitivity analysis, we noted similar results when considering rare, deleterious variants as defined by missense variants with a REVEL score[42] > 0.5 or a LOFTEE[43] high confidence predicted loss of function variant within 1 of the 11 HTAAD genes.

We then sought to examine whether the TAAD PRS was associated with an increase in *incident* TAAD, and incident TAAD-related mortality. Using the UK Biobank, we first tested the PRS with all incident TAAD events, and then tested the PRS with incident TAAD events listed as a primary or secondary cause of death in the electronic health record using Cox proportional hazards models. We observed that those in the top 5% of the PRS were 2.5-fold more likely to experience an incident TAAD event during a median of 11.2 years of follow up, and demonstrated greater than 4-fold higher risk of TAAD-related mortality (**Figure 5b**).

Discussion

In the current study, we identified 17 novel TAAD loci and localized the anatomic distribution of these TAAD risk variants. We examined the phenotypic consequences of TAAD lead risk variants with PheWAS, and through Mendelian randomization we demonstrate that elevated diastolic blood pressure, taller standing height, and smoking are likely causal epidemiologic risk factors for TAAD. Leveraging bulk and single cell/single nucleus RNA sequencing data, we identified causal tissues and cell types for TAAD. Lastly, we developed a genome-wide polygenic risk score for TAAD that identifies a subset of the population at substantially greater risk for TAAD.

These findings permit several conclusions. First, we provide substantial evidence that the genetic architecture of TAAD mirrors that of other complex traits. Current understanding of the pathophysiology of TAAD has largely been informed through the investigation and identification of rare deleterious variants within what are now termed HTAAD genes. These identified genetic variants, generally missense or nonsense mutations, substantially alter a gene’s protein product and subsequently disrupt critical functions in VSMC contraction, extracellular matrix stabilization, or TGF-b signaling[6]. While prior GWAS have identified 4 TAAD risk loci[3-5], given the relatively rare incidence of TAAD in the

population it remained unclear whether common or rare variants were the primary driver of TAAD heritability. In the current study, we increase the number of TAAD risk loci by a factor of 5 and identify putative causal risk genes that likely affect disease through changes in gene expression akin to other common complex traits. These findings offer new potential targets for therapeutic intervention as well as firmly establishes TAAD as a complex trait.

Second, despite its morphologic similarities with infrarenal abdominal aortic aneurysm (AAA), our results support the notion that TAAD is a distinct disorder from the rest of the atherosclerotic cardiovascular disease spectrum. Early observational studies of TAAD often studied abdominal and thoracic aortic aneurysms together[15], and suggested common risk factors for both diseases included hypertension, smoking, and hyperlipidemia. Here, through causal inference methods we provide genetic support for a causal role of smoking and hypertension on TAAD risk. However, unlike for AAA[44,45] and other cardiovascular diseases[46,47], circulating lipoproteins do not appear to play a substantial role in TAAD development. The loci revealed in our genetic discovery analysis highlight the role of extracellular matrix integrity in TAAD, and the tissue and cell type enrichment analyses underscore the importance of VSMCs in TAAD risk. Prior lineage mapping studies suggest that the differing embryologic origin of VSMCs in the thoracic aorta from the rest of the arterial tree may play a role in susceptibility to atherogenic conditions[48]. In light of these findings, we hypothesize that therapies focusing on restoring extracellular matrix stability, rather than atherosclerotic risk factor modification, will be more likely to provide a substantial impact on TAAD prevention.

Third, our data provide evidence for the clinical utility of TAAD polygenic risk prediction. Although recent literature focusing on PRS application has demonstrated an ability to risk stratify the population at large[49], critics have highlighted that this research has focused on diseases in which genetic testing is seldom indicated[50], unlikely to significantly alter clinical outcomes[51], or that may not be justified within current healthcare cost structures[7]. However, unlike other diseases, testing for heritable causes of TAAD is already performed - typically through exome sequencing in those with familial or early-onset TAAD syndromes to identify pathogenic variants[52]. Here, we demonstrate that individuals within the upper tail of the polygenic risk score distribution are at substantially greater risk to experience TAAD or TAAD-related mortality. Furthermore, we observed similar gains in the calculated AUC statistic when adding pathogenic or likely pathogenic TAAD risk variants or polygenic risk scores (TAAD PRS) to risk prediction models, suggesting that there is an additional benefit for PRS testing to include common variation beyond current targeted or exome sequencing panels. Taken together, our data suggest that extending current genetic panels to include testing for polygenic TAAD risk could significantly increase the yield of current genetic testing and may be warranted.

Our study should be interpreted in the context of its limitations. First, our TAAD phenotype is based on EHR diagnosis and procedural code data, and may result in misclassification of case status. However, such misclassification should, on average, reduce statistical power for discovery and bias results toward the null. Second, the Veterans Affairs Healthcare System population is overwhelmingly male and our ability to detect sex-specific genetic associations in discovery was limited. Third, power to detect differential expression associations and identify causal TAAD cell types in sc/snRNA-sequencing data may be limited by sequencing depth, sample size, or tissue processing techniques. Fourth, while we observe a significant Mendelian randomization result for the height exposure and TAAD, we cannot rule out the possibility that this association is driven by the pleiotropy of height-associated variants[53], rather than a true causal association. Finally, a number of the TAAD risk loci demonstrate genome-wide significant associations with increased DBP. Our Mendelian randomization results support a causal role for DBP in TAAD susceptibility, however the mt-COJO analysis suggests that DBP is not the sole driver of TAAD risk at these regions of the genome. Disentangling the effects of hypertension on proximal aortic dilation/dissection - a pathologic process which alters human blood pressure homeostasis - is likely to require model systems to completely elucidate the complex mechanisms at work.

In conclusion, our data provide new mechanistic insights into TAAD risk and demonstrate that its genetic architecture is akin to that of other complex traits. We identify causal risk factors, cell types, and genes that may be used to inform clinical care.

Methods

Study Populations

We conducted a discovery genetic association analysis using DNA samples and phenotypic data from the Million Veteran Program (MVP, **Supplementary Fig. 1**). In MVP, individuals aged 18 to over 100 years have been recruited from 63 Veterans Affairs (VA) Medical Centers across the United States. After quality control, we identified 7,050 participants of European, 1,266 of African, and 310 of Hispanic ancestry with TAAD and 453,043 controls free of clinical evidence of disease. For variants with meeting genome-wide significance ($P < 5x10^{-8}$) in MVP, we sought replication of our findings with data from a meta-analysis of 6 external datasets comprising 4,459 TAAD cases and 512,463 controls (**Supplementary Table 19**). Additional details of MVP and replication genetic data and quality control are available in the Online Data Supplement.

TAAD Phenotype Definitions

From the participants passing quality control in MVP, individuals were defined as having TAAD based on possessing at least two of the ICD-9/10 codes/CPT codes outlined in **Supplementary Table 20** in their electronic health record (EHR) on separate dates, and possessing zero codes suggesting a possible history of biscupid aortic valvular disease (**Supplementary Table 21).** Individuals were defined as not having TAAD if they had zero diagnosis/procedure codes suggesting a diagnosis of TAAD (**Supplementary Table 22**) and their EHR reflected two or more separate encounters in the VA Healthcare System in each of the two years prior to enrollment in MVP. For the repair phenotypes, individuals were identified as having an ascending or aortic arch repair, descending thoracic aortic repair, or dissection based on the procedure or diagnosis codes outlined in **Supplementary Table 20**. In the replication cohorts, TAAD definitions are described in **Supplementary Table 19.** MVP received ethical/study protocol approval by the VA Central Institutional Review Board, the analysis in UK Biobank was approved by a local Institutional Review Board at Partners Healthcare (protocol 2013P001840), and informed consent was obtained for all participants. Additional information regarding experimental design and participants are provided in the Life Sciences Reporting Summary.

PheWAS of TAAD Risk Variants

The IEU open GWAS project[11,12] is a publicly available online repository of a wide array of summary statistics from previously published GWAS and the UK Biobank. For TAAD lead variants identified in our GWAS analysis, we queried the phenotypes available in the IEU open GWAS project to perform a PheWAS (Phenome-wide Association Study[10]) across a range of over 2,000 conditions, disease phenotypes, and metabolites. Details of the sensitivity analysis re-examining the DBP-associated variants with TAAD accounting for blood pressure using mt-COJO summary-based methods[13] are described in the supplementary appendix.

TAAD Risk Factor Mendelian Randomization Analyses

Mendelian randomization (MR) analyses for smoking (through a lifetime smoking index), lipid levels (triglycerides, HDL and LDL cholesterol), blood pressure (systolic blood pressure [SBP] and diastolic blood pressure [DBP]), and height exposures were performed with TAAD as the outcome. Given that some of the above exposure summary statistics included UK Biobank data, the TAAD outcome summary data included all of the studies in our analysis except UK Biobank, encompassing 12,422 TAAD cases and 578,768 controls. Genetic instruments were selected as DNA sequence variants that associated with the exposure at genome-wide significance ($P<5x10^{-8}$) with an $R^2<0.001$. All clumping was performed using the TwoSampleMR package of R[54]. Genetic instruments were constructed for the lifetime smoking index (462,690 participants)[16], lipid levels (up to 188,577 participants)[17], blood pressure (up to 757,601 participants)[18], and height (253,288 individuals)[19] using publicly available summary statistics (**Supplementary Table 11)**. Inverse-variance weighted MR was used for the primary analysis, with weighted-median[20] MR performed as sensitivity analysis, allowing for up to 50% of the weight of each instrument to be drawn from invalid instruments while controlling type-I error. MR-Egger[22] analysis was performed to evaluate for horizontal pleiotropy, as was the MR-PRESSO[21] test which consists of three parts: 1) the global test for horizontal pleiotropy, 2) the outlier corrected causal estimate which corrects for the detected horizontal pleiotropy, and 3) the distortion test which tests if the causal estimate is significantly different after outlier adjustment.

Causal TAAD Gene and Variant Identification

We prioritized candidate causal genes at each of the identified TAAD risk loci by aggregating evidence from i) prior genetic, clinical, or functional studies, ii) the closest gene to the lead risk variant, iii) genes with protein-altering variants in high linkage disequilibrium ($R^2 > 0.8$) with the lead TAAD risk variant, iv) cis-eQTLs from the Genotype-Tissue Expression Project (GTEx) dataset in aortic tissue[28] with association $P < 5x10^{-6}$, v) results from FOCUS[27] v0.5, a fine-mapping technique to identify causal genes in a transcriptome-wide association study[55] (TWAS) using bulk RNA-seq data from post-mortem aortic tissue (387 individuals from GTEx) and TAAD meta-analysis (discovery and replication) summary statistics, and v) results of a colocalization analysis from our TAAD GWAS meta-analysis and expression quantitative trait locus (eQTL) data from GTEx bulk RNA-seq data in aortic tissue using the *coloc* R package[56]. Further methodologic details of these analyses and an analysis identifying putative causal TAAD risk variants are described in the supplementary appendix.

Stratified LD Score Regression Analysis

As an initial enrichment analysis, we partitioned the heritability of TAAD using stratified LD-score regression[31]. Stratified LD-score regression leverages GWAS summary to estimate the heritability explained by each functional classification while accounting for LD structure and other annotations. For this analysis, we combined the TAAD meta-analysis summary statistics and a previously published set of 205 cell type annotations from GTEx[28] and the previously defined “Franke lab dataset[31].” For this analysis, we approximated the LD structure from Europeans within the 1000 Genomes[39] reference panel, and set a Bonferroni corrected $P < 0.00024$ (0.05/205 annotations) for statistical significance.

Enrichment Analysis with Human Thoracic Aorta snRNA-seq Data

We generated a gene set of 21 putative causal TAAD risk genes: the 11 candidate causal genes identified from our GWAS, as well as 10 additional previously identified definitive or strong HTAAD genes ("Category A")[6], and overlapped our gene set with publicly available single nucleus RNA-sequencing data from ascending and descending thoracic aorta specimens[33]. We downloaded single nucleus expression data and existing tSNE cluster annotations for each cell from the Broad Institute single cell portal (https://singlecell.broadinstitute.org/single_cell), and combined the clusters into 5 overarching cell categorizations - VSMCs, fibroblasts, endothelial cells, leukocytes, and other cells (**Supplementary Table 17**). We then calculated average expression for each gene across all cells in that cell type. For each cell type, we calculated the enrichment P-value for our list of causal TAAD risk genes using the 'fgsea' R package v1.20.0, which looks for overrepresentation of our gene list in ranked genes for each cell type, as implemented in R 4.1. A Bonferroni two-sided P <0.01 (0.05/5 cell types) was used to declare statistical significance.

Cell Type Prioritization in scRNA/snRNA-seq Data from Thoracic Aorta

For 9 genes that we hypothesized influence TAAD risk through alterations in gene expression ("candidate causal genes through changes in expression" in **Supplementary Table 14**), we prioritized causal cell types using scRNA/snRNA-seq data generated from thoracic aorta specimens. Previously published single cell RNA sequencing from control (n=3) and ascending thoracic aortic aneurysm cases (n=8) were reanalyzed using *Seurat* (v4)[34,57]. Briefly, dimensionality reduction was previously performed using tSNE and identification of cluster-defining genes was performed using the FindAllMarkers function. Feature plots (order=TRUE, min.cutoff='q1', max.cutoff='q95', raster=TRUE, pt.size=2.5), violin plots (split.plot=TRUE,split.by='stim'), dot plots and heat maps were generated using *Seurat* (v4). Within each cluster, differential gene expression between clusters based on case/control status was performed with FindAllMarkers (group.by = 'stim'), and a Bonferroni P < 0.01 was set for statistical significance. Previously published single nucleus RNA sequencing from thoracic aorta was downloaded from the Broad Single Cell Portal[33]. The cells were processed in R Studio with *Seurat* (v4) according to the pipeline above, with metadata of the original clustering added. Briefly, cells were filtered based on the following parameters: 250 < nFeatureRNA < 2500; nCount_RNA >500; and percent.mt <0.5%. Variable features were scaled using ScaleData for percent.mt. Dimensionality reduction with UMAP was performed and dotplots were generated with group.by ='Category' corresponding to original clusters from Pirruccello et al[33].

TAAD Polygenic Risk Score Generation

A weighted polygenic risk score (PRS) represents an individual's risk of a given disease conferred by the sum of the effects of many common DNA sequence variants. A weight is assigned to each genetic variant based on its strength of association with disease risk (β). Individuals are then additively scored in a weighted fashion based on the number of risk alleles they carry for each variant in the PRS.

To generate our scores, we used summary statistics from the MVP multi-ancestry discovery GWAS (8,626 TAAD cases, 453,043 controls) and a LD panel from 1000 Genomes[39] whole genome sequencing data. To increase the number of independent variants included in our score, we used the PRS-CSx v1.0 software which uses Bayesian methods to generate posterior genetic variants effect sizes under coupled continuous shrinkage priors[40]. The latest iteration of the software allows for the integration of summary statistics across multiple populations to improve polygenic predictions across ancestries. The European, African, and Hispanic ancestry summary statistics were input with European (EUR), African (AFR), and Ad Mixed American (AMR) reference panels from 1000 Genomes, and default

software parameters were used including the use of HapMap[58] imputed variants for PRS generation and allowing PRS-CSx to generate the global shrinkage parameter phi through a Bayesian approach.

We then tested our scores in 3 separate datasets. For initial validation, we tested the normalized PRS using prevalent data from the MassGeneral Brigham Biobank in 775 individuals with and 24,518 individuals without TAAD of European ancestry. Next, we assessed the performance of the PRS in an updated freeze of the CHIP-MGI cohort (3,743 cases and 51,898 controls, and subsequently compared these results (in terms of effect estimate and AUC) to the effects of a set of previously curated, pathogenic or likely pathogenic TAAD risk variants among individuals with exome sequencing data available (a subset of 1,842 cases and 1,887 controls). In a sensitivity analysis, we also examined the effects of rare missense variants with a REVEL score[42] > 0.5 and high-confidence LOFTEE[43] predicted loss of function variants in 1 of the 11 HTAAD genes. Lastly, we examined whether the TAAD PRS was associated with an increase in risk of incident TAAD and TAAD-related mortality within the UK Biobank among 281 participants with incident TAAD, 22 participants with TAAD-related deaths, and 359,000 participants without TAAD of European ancestry.

Statistical Analysis

In our primary discovery analysis, genotyped and imputed DNA sequence variants in individuals of European, African, and Hispanic ancestry were tested for association with TAAD using logistic mixed models as performed in the REGENIE v2.0 statistical software program[59]. We included in step 1 of REGENIE (i.e. prediction of individual trait values based on the

genetic data) variants that were directly genotyped, had a minor allele frequency >1%,

<10% missingness, Hardy-Weinberg equilibrium test P-value>10^{-15}. The

association model used in step 2 of REGENIE included as covariates age, sex, and 5 principal components of ancestry. Next, associated statistics across of European, African, and Hispanic ancestry MVP participants were meta-analyzed using an inverse variance-weighted fixed effects method as implemented in the METAL software program[60]. We excluded variants with a high amount of heterogeneity (I^2 statistic > 75%) across the three ancestries.

For variants meeting genome-wide significance for TAAD ($P < 5x10^{-8}$) we sought replication of our findings from a combination of 6 external cohorts representing 4,459 individuals with TAAD and 512,463 individuals without TAAD. Details of participant selection, quality control, phenotyping, and statistical analysis are presented in **Supplementary Table 19**.

We defined significant novel TAAD associations as those that were at least nominally significant in replication (P<0.05), were directionally consistent in both cohorts, and had an overall $P < 5x10^{-8}$ (genome-wide significance) in the discovery and replication cohorts combined. Novel loci were defined as being greater than 500,000 base-pairs away from a known TAAD genome-wide associated lead variant. Additionally, linkage disequilibrium information from the 1000 Genomes Project[39] was used to determine independent variants where the association peak extended beyond 500,000 base-pairs. All logistic regression values of P were two-sided.

In the PheWAS analysis, DNA sequence variants were queried in the IEU open GWAS project[11,12], an online resource of association statistics from previously conducted GWAS, and used a genome-wide significant P value threshold (two-

sided $P < 5 \times 10^{-8}$) to declare statistical significance.

In our Mendelian randomization analyses, a random-effects inverse-variance weighted method was used as the main analysis, with sensitivity analyses performed for the statistically significant associations as described above. We set a two-sided $P < 0.007$ (0.05/7 traits) for statistical significance.

In the association analysis of TAAD risk variants with specific aortic repair or dissection, we tested each variant with ascending/aortic arch repair, descending thoracic aortic repair, or aortic dissection through logistic regression adjusting for age, sex, and 5 principal components of ancestry within European MVP participants using the R statistical software program[61] (R version 3.6). Given the known prior association with TAAD we used a nominally significant two-sided $P < 0.05$ to declare significance.

In our PRS analysis, logistic regression models (prevalent cases) were used to estimate odds ratios and 95% confidence intervals for the associations of the continuous PRS (1 standard deviation unit) with TAAD in the MassGeneral Brigham Biobank and CHIP-MGI adjusting for age, sex, 5 principal components. We additionally calculated the prevalence of TAAD for the 5% of individuals with the highest PRS relative to the rest of the population, and generated confidence intervals using R (version 4.1). In CHIP-MGI, we also tested the association of rare, pathogenic variants with TAAD risk using logistic regression and adjusting for age, sex, and principal components. In a sensitivity analysis, we examined the TAAD risk conferred through rare missense variants with a REVEL score[42] > 0.5 and high-confidence LOFTEE[43] predicted loss of function variants in 1 of the 11 HTAAD genes. Following analysis, an AUC statistic was generated for each of these models.

In the UK Biobank, we tested the association of the 5% of individuals with the highest TAAD PRS relative to the rest of the population with incident TAAD events and incident TAAD-related mortality using Cox proportional hazard models adjusting for age, sex, and 5 principal components of ancestry in the White British subset of UK Biobank participants. Prevalent cases were excluded, and individuals were censored upon death, when experiencing the relevant event, or at the end of follow up (a median of 11.1 years). We declared a $P < 0.0125$ for statistical significance (0.05/4 tests – associations for the: 1) the continuous PRS, 2) the top 5% PRS, 3) rare TAAD risk variants, and 4) TAAD-related death). All P values were two-sided.

Declarations

Data Availability

The full summary level association data from the MVP TAAD discovery analysis from this manuscript will be available through dbGAP, accession code phs001672.v2.p1.

Acknowledgements

This publication does not represent the views of the Department of Veteran Affairs or the United States Government. Use of the data was facilitated through UK Biobank Application 7089. The authors acknowledge the Michigan Genomics Initiative participants, Precision Health at the University of Michigan, the University of Michigan Medical School Central Biorepository, and the University of Michigan Advanced Genomics Core for providing data and specimen storage, management, processing, and distribution services, and the Center for Statistical Genetics in the Department of Biostatistics at the School of Public Health for genotype data curation, imputation, and management in support of the research reported in this publication/grant application/presentation. We acknowledge the Penn

Medicine BioBank (PMBB) for providing data and thank the patient-participants of Penn Medicine who consented to participate in this research program. We would also like to thank the Penn Medicine BioBank team and Regeneron Genetics Center for providing genetic variant data for analysis. The PMBB is approved under IRB protocol# 813913 and supported by Perelman School of Medicine at University of Pennsylvania, a gift from the Smilow family, and the National Center for Advancing Translational Sciences of the National Institutes of Health under CTSA award number UL1TR001878.

Sources of Funding

This work was supported by funding from the Department of Veterans Affairs Office of Research and Development, Million Veteran Program Grant MVP000; Department of Veterans), I01-BX003362 (P.S.T.), IK2-CX001780 (S.M.D.). S.A.L., Y.L, Y.H.S. are supported by grants from the American Heart Association (AHA) Vascular Diseases Strategically Focused Research Networks (SFRN) (AHA18SFRN33960114). S.A.L. is supported in part by the Jimmy and Roberta Howell Professorship in Cardiovascular Surgery at Baylor College of Medicine. P.N. is supported by grants from the National Heart, Lung, and Blood Institute (R01HL142711, R01HL148050, R01HL151283, R01HL127564, R01HL151152) and the National Human Genome Research Institute (U01HG011719). K.A. is supported by grants from the National Institutes of Health (1K08HL153937) and the American Heart Association (862032). K.P. is supported by a grant from the National Heart, Lung and Blood Institute (5-T32HL007208-43). A.R.A. is supported by award number F30-DK120062. D.M.M. is supported by the National Heart, Lung, and Blood Institute (RO1HL62594), and the John Ritter and Remembrin' Benjamin Foundations. M.G.L. is supported by the Institute for Translational Medicine and Therapeutics of the Perelman School of Medicine at the University of Pennsylvania and the NIH/NHLBI National Research Service Award postdoctoral fellowship (T32HL007843).

Competing Interests

The spouse of C.J.W. is employed by Regeneron. S.A.L. is a consultant for Terumo Aortic and Cerus, and a principal investigator for clinical studies sponsored by Terumo Aortic and CytoSorbents. P.N. reports grants from Amgen, Apple, AstraZeneca, Boston Scientific, and Novartis, is a consultant for Apple, AstraZeneca, Blackstone Life Sciences, Foresite Labs, geneXwell, Novartis, Roche / Genentech, and TenSixteen Bio, is a scientific co-founder of TenSixteen Bio, and spousal employment at Vertex, all unrelated to the present work. K.A.A. reports consulting for Sarepta Therapeutics and a research collaboration with Novartis. M.D.R. is on the scientific advisory board for Cipherome.

Code Availability

Data was collected using the EasyQC package and REGENIE v2 software program as outlined in the online methods. Additional software used for analysis include the Coloc R package, FOCUS/TWAS v0.5, Eigensoft v6, PRSCSx v1.0, Seurat v4. Clear code for analysis is available at their associated website (see text), and any additional code will be provided at reasonable request.

References

1. Tsai, T.T. *et al.* Long-term survival in patients presenting with type B acute aortic dissection: insights from the International Registry of Acute Aortic Dissection. *Circulation* **114**, 2226-31 (2006).

2. Erbel, R. *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* **35**, 2873-926 (2014).

3. LeMaire, S.A. *et al.* 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 (2011).

4. Guo, D.C. *et al.* Genetic Variants in LRP1 and ULK4 Are Associated with Acute Aortic Dissections. *Am J Hum Genet* **99**, 762-769 (2016).

5. Roychowdhury, T. *et al.* Regulatory variants in TCF7L2 are associated with thoracic aortic aneurysm. *Am J Hum Genet* **108**, 1578-1589 (2021).

6. Renard, M. *et al.* Clinical Validity of Genes for Heritable Thoracic Aortic Aneurysm and Dissection. *J Am Coll Cardiol* **72**, 605-615 (2018).

7. Klarin, D. *et al.* Genetic Architecture of Abdominal Aortic Aneurysm in the Million Veteran Program. *Circulation* (2020).

8. Ashvetiya, T. *et al.* Identification of novel genetic susceptibility loci for thoracic and abdominal aortic aneurysms via genome-wide association study using the UK Biobank Cohort. *PLoS One* **16**, e0247287 (2021).

9. Denny, J.C. *et al.* PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. *Bioinformatics* **26**, 1205-10 (2010).

10. Denny, J.C. *et al.* Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. *Nat Biotechnol* **31**, 1102-10 (2013).

11. Elsworth, B. *et al.* The MRC IEU OpenGWAS data infrastructure. *bioRxiv*, 2020.08.10.244293 (2020).

12. Hemani, G. *et al.* The MR-Base platform supports systematic causal inference across the human phenome. *Elife* **7**(2018).

13. Zhu, Z. *et al.* Causal associations between risk factors and common diseases inferred from GWAS summary data. *Nature Communications* **9**, 224 (2018).

14. Isselbacher, E.M. Thoracic and abdominal aortic aneurysms. *Circulation* **111**, 816-28 (2005).

15. Reed, D., Reed, C., Stemmermann, G. & Hayashi, T. Are aortic aneurysms caused by atherosclerosis? *Circulation* **85**, 205-11 (1992).

16. Wootton, R.E. *et al.* Evidence for causal effects of lifetime smoking on risk for depression and schizophrenia: a Mendelian randomisation study. *Psychol Med*, 1-9 (2019).

17. Global Lipids Genetics Consortium *et al.* Discovery and refinement of loci associated with lipid levels. *Nat Genet* **45**, 1274-83 (2013).

18. Evangelou, E. *et al.* Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits. *Nat Genet* **50**, 1412-1425 (2018).

19. Wood, A.R. *et al.* Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat Genet* **46**, 1173-86 (2014).

20. Burgess, S., Bowden, J., Fall, T., Ingelsson, E. & Thompson, S.G. Sensitivity Analyses for Robust Causal Inference from Mendelian Randomization Analyses with Multiple Genetic Variants. *Epidemiology* **28**, 30-42 (2017).

21. Verbanck, M., Chen, C.Y., Neale, B. & Do, R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nat Genet* **50**, 693-698 (2018).

22. Bowden, J., Davey Smith, G. & Burgess, S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int J Epidemiol* **44**, 512-25 (2015).

23. Marfan, A.B. Un cas de deformation congenitale des quatres membres plus prononcee aux extremities characterisee par lallongement des os avec un certain degre d amincissement. *Bull Mem Soc Med Hop Paris* **13**, 220 (1896).

24. Szabo, Z. *et al.* Aortic aneurysmal disease and cutis laxa caused by defects in the elastin gene. *J Med Genet* **43**, 255-8 (2006).

25. Boucher, P., Gotthardt, M., Li, W.P., Anderson, R.G. & Herz, J. LRP: role in vascular wall integrity and protection from atherosclerosis. *Science* **300**, 329-32 (2003).

26. Davis, F.M. *et al.* Smooth muscle cell deletion of low-density lipoprotein receptor-related protein 1 augments angiotensin II-induced superior mesenteric arterial and ascending aortic aneurysms. *Arterioscler Thromb Vasc Biol* **35**, 155-62 (2015).

27. Mancuso, N. *et al.* Probabilistic fine-mapping of transcriptome-wide association studies. *Nat Genet* **51**, 675-682 (2019).

28. GTEx Consortium. Genetic effects on gene expression across human tissues. *Nature* **550**, 204-213 (2017).

29. Tsutsui, K. *et al.* ADAMTSL-6 is a novel extracellular matrix protein that binds to fibrillin-1 and promotes fibrillin-1 fibril formation. *J Biol Chem* **285**, 4870-82 (2010).

30. Elbitar, S. *et al.* Pathogenic variants in THSD4, encoding the ADAMTS-like 6 protein, predispose to inherited thoracic aortic aneurysm. *Genet Med* **23**, 111-122 (2021).

31. Finucane, H.K. *et al.* Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat Genet* **50**, 621-629 (2018).

32. Ducy, P., Schinke, T. & Karsenty, G. The osteoblast: a sophisticated fibroblast under central surveillance. *Science* **289**, 1501-4 (2000).

33. Pirruccello, J.P. *et al.* Deep learning enables genetic analysis of the human thoracic aorta. *Nat Genet* (2021).

34. Li, Y. *et al.* Single-Cell Transcriptome Analysis Reveals Dynamic Cell Populations and Differential Gene Expression Patterns in Control and Aneurysmal Human Aortic Tissue. *Circulation* **142**, 1374-1388 (2020).

35. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902.e21 (2019).

36. Dingemans, K.P., Teeling, P., Lagendijk, J.H. & Becker, A.E. Extracellular matrix of the human aortic media: an ultrastructural histochemical and immunohistochemical study of the adult aortic media. *Anat Rec* **258**, 1-14 (2000).

37. Joshi, M.B. *et al.* T-cadherin protects endothelial cells from oxidative stress-induced apoptosis. *Faseb j* **19**, 1737-9 (2005).

38. Philippova, M. *et al.* Atypical GPI-anchored T-cadherin stimulates angiogenesis in vitro and in vivo. *Arterioscler Thromb Vasc Biol* **26**, 2222-30 (2006).

39. The 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* **526**, 68-74 (2015).

40. Ruan, Y. *et al.* Improving Polygenic Prediction in Ancestrally Diverse Populations. *medRxiv*, 2020.12.27.20248738 (2021).

41. 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* **17**, 405-24 (2015).

42. Ioannidis, N.M. *et al.* REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. *Am J Hum Genet* **99**, 877-885 (2016).

43. Karczewski, K.J. *et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434-443 (2020).

44. Klarin, D. *et al.* Genetics of blood lipids among ~300,000 multi-ethnic participants of the Million Veteran Program. *Nat Genet* **50**, 1514-1523 (2018).

45. Harrison, S.C. *et al.* Genetic Association of Lipids and Lipid Drug Targets With Abdominal Aortic Aneurysm: A Meta-analysis. *JAMA Cardiol* **3**, 26-33 (2018).

46. Do, R. *et al.* Common variants associated with plasma triglycerides and risk for coronary artery disease. *Nat Genet* **45**, 1345-52 (2013).

47. Levin, M.G. *et al.* Prioritizing the Role of Major Lipoproteins and Subfractions as Risk Factors for Peripheral Artery Disease. *Circulation* **In Press**(2021).

48. Majesky, M.W. Developmental basis of vascular smooth muscle diversity. *Arterioscler Thromb Vasc Biol* **27**, 1248-58 (2007).

49. Khera, A.V. *et al.* Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat Genet* **50**, 1219-1224 (2018).

50. Baker, E. & Escott-Price, V. Polygenic Risk Scores in Alzheimer's Disease: Current Applications and Future Directions. *Front Digit Health* **2**, 14 (2020).

51. Reitter-Pfoertner, S. *et al.* The influence of thrombophilia on the long-term survival of patients with a history of venous thromboembolism. *Thromb Haemost* **109**, 79-84 (2013).

52. Wolford, B.N. *et al.* Clinical Implications of Identifying Pathogenic Variants in Individuals With Thoracic Aortic Dissection. *Circ Genom Precis Med* **12**, e002476 (2019).

53. Marouli, E. *et al.* Rare and low-frequency coding variants alter human adult height. *Nature* (2017).

54. Zheng, J. *et al.* LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics* **33**, 272-279 (2017).

55. Gusev, A. *et al.* Integrative approaches for large-scale transcriptome-wide association studies. *Nat Genet* **48**, 245-52 (2016).

56. Guo, H. *et al.* Integration of disease association and eQTL data using a Bayesian colocalisation approach highlights six candidate causal genes in immune-mediated diseases. *Hum Mol Genet* **24**, 3305-13 (2015).

57. Angueira, A.R. *et al.* Defining the lineage of thermogenic perivascular adipose tissue. *Nat Metab* **3**, 469-484 (2021).

58. Frazer, K.A. *et al.* A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**, 851-61 (2007).

59. Mbatchou, J. *et al.* Computationally efficient whole-genome regression for quantitative and binary traits. *Nat Genet* **53**, 1097-1103 (2021).

60. Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190-1 (2010).

61. R Development Core Team. R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria.* **ISBN 3-900051-07-0, URL http://www.R-project.org.**(2008).

Tables

Table 1 – 17 Novel TAAD risk loci after discovery in MVP and independent replication

Chr:Pos (hg19)	rsid	EA	NEA	EAF	Overall OR	Overall 95% CI	Overall P	Gene/Locus*	Annotation
1:9436538	rs4596926	A	G	0.404	1.10	1.07-1.13	3.94E-11	*(SPSB1)*	regulatory_region_variant
2:164915279	rs13002621	G	A	0.222	1.12	1.08-1.15	2.68E-12	*AC092684.1*	intron_variant
2:19729131	rs1863777	T	C	0.361	1.14	1.11-1.17	1.67E-19	*(OSR1)*	intergenic_variant
2:238227919	rs6759927	G	A	0.301	1.11	1.08-1.14	3.57E-12	*COL6A3*	downstream_gene_variant
4:146814640	rs7666150	C	T	0.504	1.13	1.10-1.16	6.58E-17	*ZNF827*	intron_variant
5:122441482	rs337128	C	A	0.513	1.12	1.09-1.15	2.74E-16	*PRDM6*	intron_variant
5:95566562	rs55745974	A	T	0.658	1.20	1.16-1.23	1.54E-32	*CTD-2337A12.1*	intron_variant
7:35293972	rs336284	A	G	0.485	1.10	1.08-1.14	6.23E-13	*TBX20*	upstream_gene_variant
7:73431693	rs62465578	T	G	0.431	1.18	1.15-1.21	2.94E-31	*(ELN)*	regulatory_region_variant
9:127883905	rs139650453	A	G	0.875	1.15	1.10-1.20	4.98E-10	*SCAI*	intron_variant
10:96061793	rs4394764	G	A	0.810	1.15	1.11-1.19	4.52E-17	*PLCE1*	intron_variant
11:130271647	rs747249	G	A	0.619	1.16	1.13-1.19	5.50E-25	*ADAMTS8*	downstream_gene_variant
13:22861921	rs9316871	A	G	0.783	1.21	1.18-1.25	3.03E-32	*(AL354828.1)*	regulatory_region_variant
13:50810171	rs2765768	G	A	0.307	1.10	1.07-1.13	4.85E-11	*DLEU1*	intron_variant
15:71612514	rs1441358	T	G	0.643	1.12	1.09-1.15	4.50E-16	*THSD4*	intron_variant
16:83045790	rs7500448	G	A	0.241	1.13	1.09-1.16	2.47E-13	*CDH13*	intron_variant
17:2097583	rs1002135	G	T	0.398	1.11	1.09-1.14	5.10E-16	*SMG6*	intron_variant

* Genes for variants that are outside the transcript boundary of a protein-coding gene are shown with nearest candidate gene in parentheses [eg, *(OSR1)*]
Abbreviations: EA, Effect Allele; NEA, Non Effect Allele; EAF, Effect Allele Frequency; OR, Odds Ratio; CI, Confidence Interval

Figures

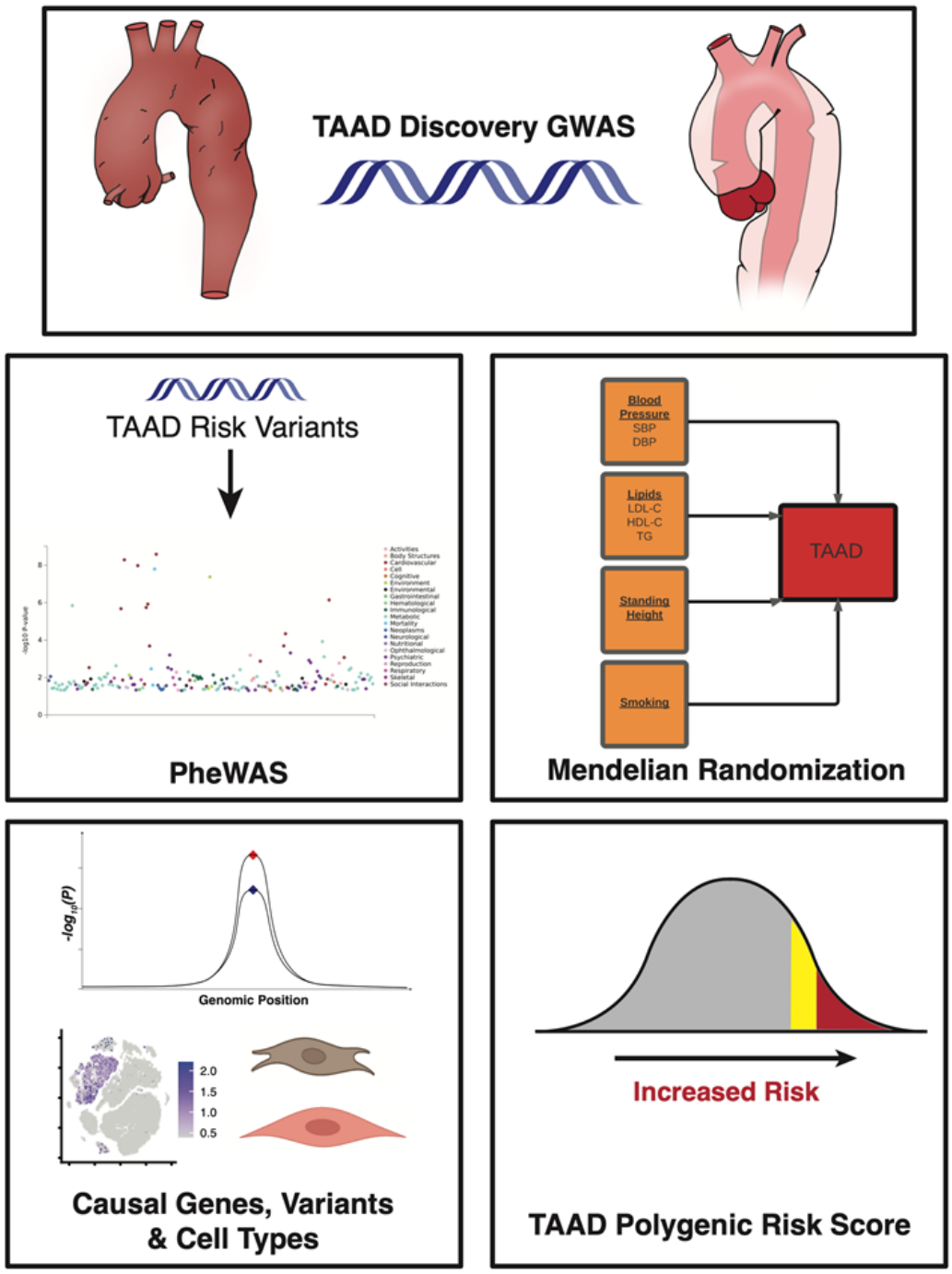

Figure 1

Overall study design. In the current study, we first performed a TAAD discovery GWAS in the Million Veteran Program (MVP), with replication from 6 external datasets. Secondary analyses included: a PheWAS of lead TAAD risk variants, Mendelian randomization analyses with known epidemiologic risk factors for disease, a series of analyses to identify causal genes, variants and cell types for TAAD leveraging colocalization techniques and single cell/single nuclear RNA sequencing data, and the generation/testing of a TAAD polygenic risk score (PRS).

Abbreviations: TAAD, Thoracic Aortic Aneurysm and Dissection; GWAS, Genome-wide Association Study; PheWAS, Phenome-wide Association Study; SBP, Systolic Blood Pressure; DBP, Diastolic Blood Pressure; LDL-C, Low-Density Lipoprotein Cholesterol; HDL-C, High-Density Lipoprotein Cholesterol; TG, Triglycerides

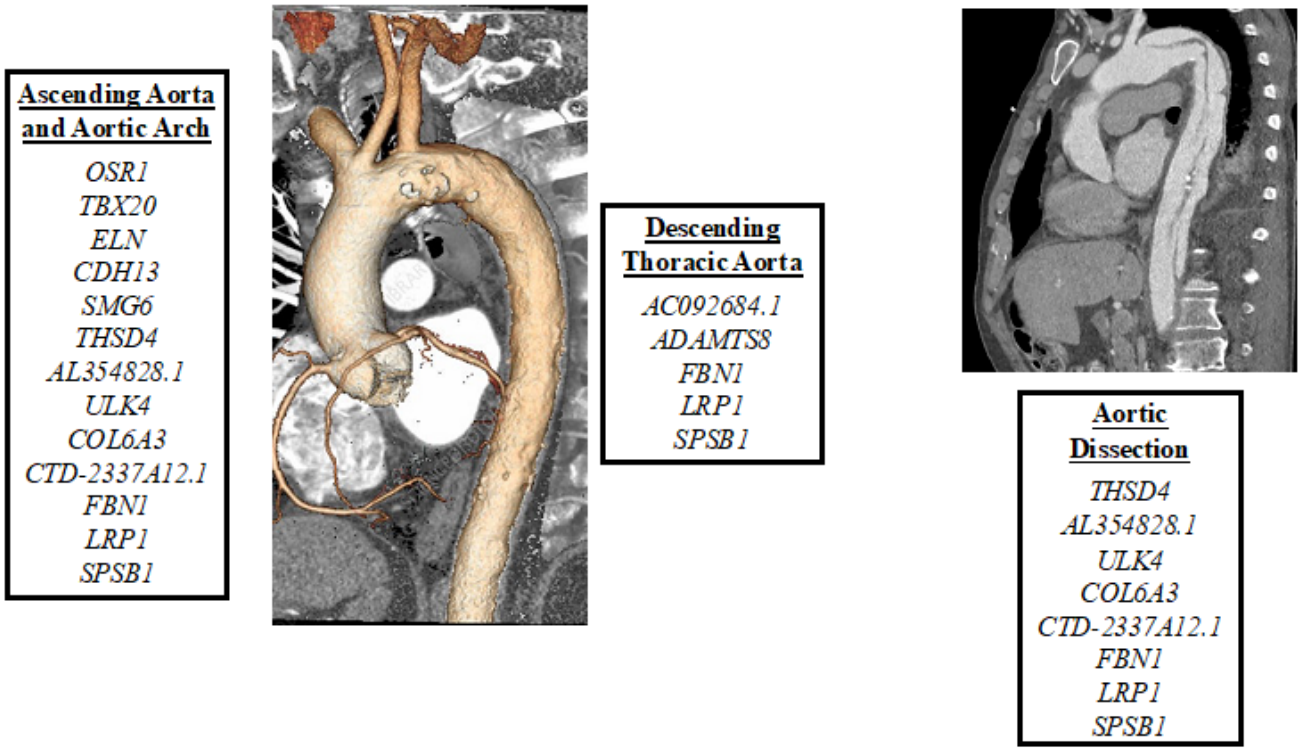

Figure 2

Anatomic distribution of TAAD risk variants in the thoracic aorta. We tested the 21 TAAD lead risk variants with ascending aorta/aortic arch repair (N = 354 cases), descending thoracic aortic repair (N = 214 cases), or aortic dissection (N=524 cases) phenotypes through logistic regression within European MVP participants. Variants were declared to be significantly associated with the aortic repair or dissection phenotype if the two-sided P value of association < 0.05.

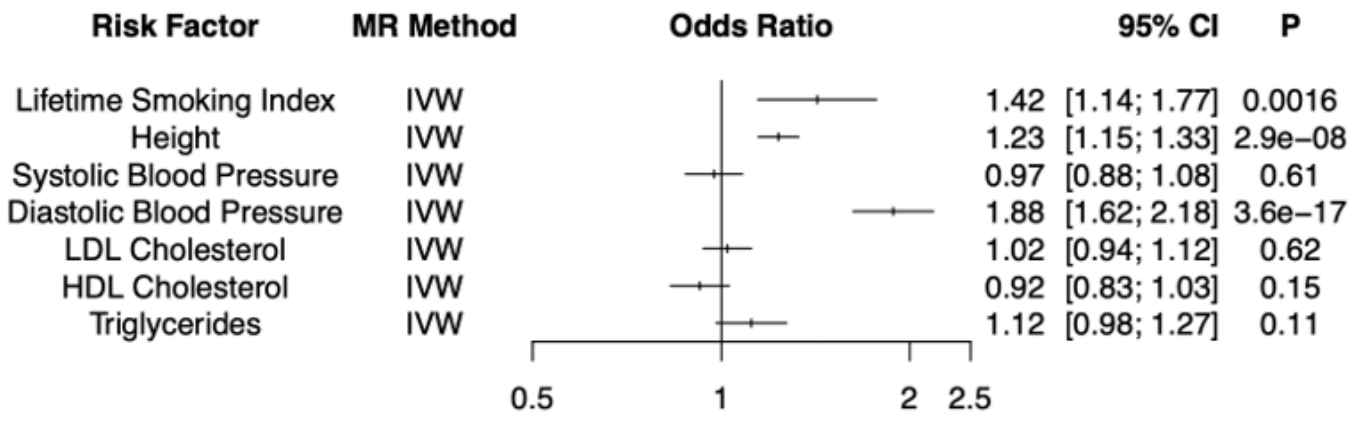

Figure 3

Mendelian randomization analyses of Epidemiologic Risk Factors for TAAD. Logistic regression association results for multiple epidemiologic risk factor exposures with the TAAD outcome in two-sample Mendelian randomization

analyses. The lifetime smoking TAAD odds ratio reflects a per genetic increase in smoking ~20 cigarettes a day for 15 years and stopping 17 years ago. The odds ratio for height reflects a 1 standard deviation genetic increase in standing height (~7.6cm). The systolic and diastolic blood pressure odds ratios correspond to the change in TAAD risk per 10 mmHg increase in the blood pressure trait. The lipids odds ratios reflect the change in TAAD risk per standard deviation genetic increase in lipid fraction. Two-sided values of P are displayed, and we set a two-sided P < 0.007 (0.05/7 traits) for statistical significance.

Abbreviations: IVW, Inverse-Variance Weighted; MR, Mendelian Randomization; CI, Confidence Interval

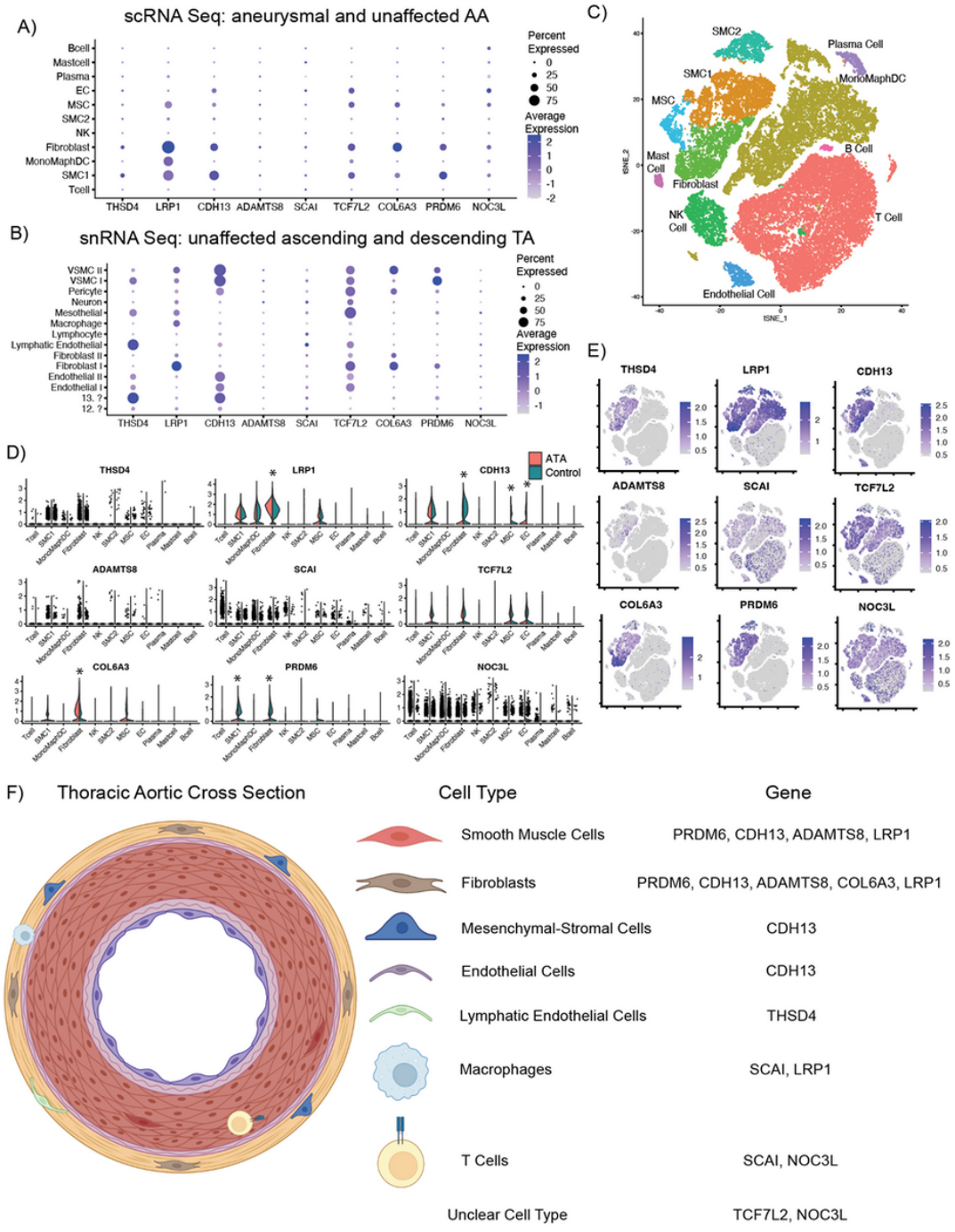

Figure 4

Causal TAAD cell types. Dot plots for each of the 9 candidate causal genes likely affecting TAAD risk based on changes in gene expression in (A) scRNA-seq data from aneurysmal and unaffected ascending aorta[34], and (B) snRNA-seq data from unaffected ascending and descending thoracic aorta[33]. C) TSNE plot of cell type clusters for scRNA-seq data from aneurysmal and unaffected ascending thoracic aorta. D) Violin plots and E) relative expression of each of the 9 candidate causal genes likely affecting TAAD risk based on changes in gene expression for each cell type. * Indicates p value <0.01 after Bonferroni correction for differential expression in aneurysmal ascending thoracic aortic tissue cases versus unaffected tissue[34]. F) Prioritized cell type(s) for each of the 9 candidate causal genes above depicted along a representative thoracic aortic cross section. Abbreviations: AA, Ascending Aorta; TA, Thoracic Aorta; ATA, Ascending Thoracic Aneurysm; scRNA Seq, single cell RNA sequencing data; snRNA Seq, single nucleus RNA sequencing data; VSMC, Vascular Smooth Muscle Cell; SMC, Smooth Muscle Cell; MonoMaphDC, Monocyte/Macrophage/Dendritic Cell; NK, Natural Killer Cell; EC, Endothelial Cell; ?, unclear cell type as referenced in [33]; MSC, Mesothelial Cell or Mesenchymal Stromal Cell

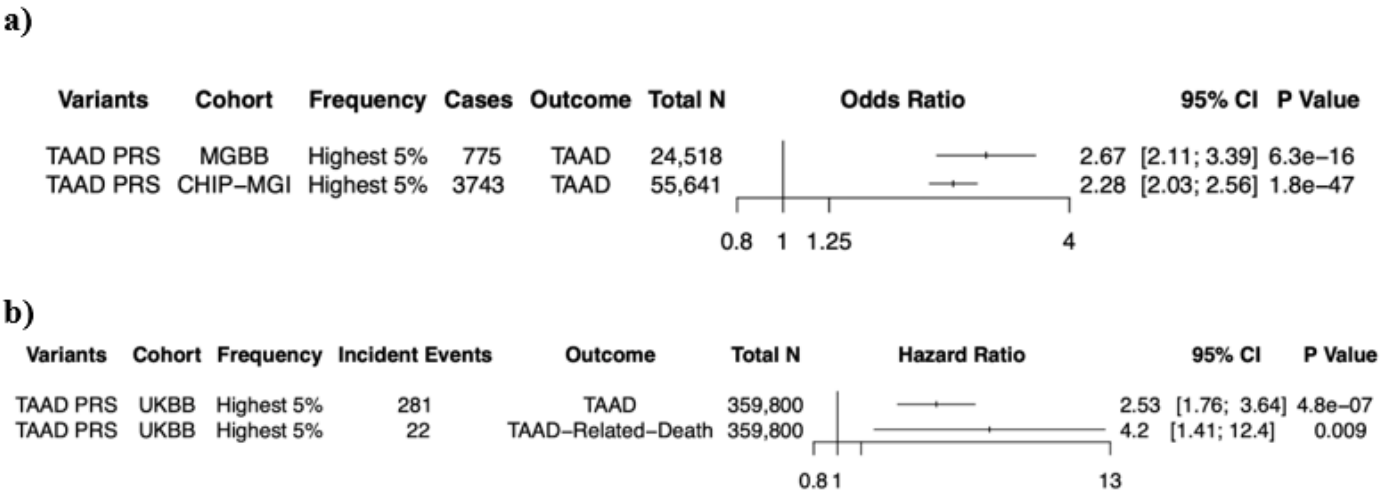

Figure 5

TAAD Polygenic Risk. A) Logistic regression odds ratios and 2-sided P values for the association of the top 5% of the TAAD PRS with prevalent TAAD in the MassGeneral Brigham and CHIP-MGI biobanks. Results were combined in an inverse variance-weighted fixed effects meta-analysis. B) Hazard ratios and 2-sided P values for the association of the top 5% of the TAAD PRS with incident TAAD and TAAD-related-death in the UK Biobank.

Abbreviations: TAAD, Thoracic Aortic Aneurysm and Dissection; PRS, Polygenic Risk Score; MGBB, MassGeneral Brigham Biobank; UKBB, UK Biobank; OR, Odds ratio; HR, Hazard Ratio

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- TAADGWASSupplementaryTables.v1.5.xlsx

- TAADSupplementaryFigShortTablesv1.5.docx