Genome-wide association study reveals novel genetic loci

Roselli, Carolina; Yu, Mengyao; Nauffal, Victor; Georges, Adrien; Yang, Qiong; Love, Katie; Weng, Lu-Chen; Delling, Francesca N; Maurya, Svetlana R; Schrölkamp, Maren

Published in:
European Heart Journal

DOI:
10.1093/eurheartj/ehac049

ESC European Society of Cardiology

European Heart Journal (2022) **00**, 1–14
https://doi.org/10.1093/eurheartj/ehac049

TRANSLATIONAL RESEARCH
Valvular heart disease

Genome-wide association study reveals novel genetic loci: a new polygenic risk score for mitral valve prolapse

Carolina Roselli [1,2], Mengyao Yu[3], Victor Nauffal [1,4], Adrien Georges [3], Qiong Yang [5], Katie Love [6], Lu-Chen Weng [1,6], Francesca N. Delling [7], Svetlana R. Maurya [8], Maren Schrölkamp [8], Jacob Tfelt-Hansen [9,10], Albert Hagège[3,11], Xavier Jeunemaitre[3,11], Stéphanie Debette [12,13], Philippe Amouyel [14], Wyliena Guan [1,6,15], Jochen D. Muehlschlegel [16], Simon C. Body [17], Svati Shah [18,19], Zainab Samad[19,20], Sergiy Kyryachenko [3], Carol Haynes [18], Michiel Rienstra [2], Thierry Le Tourneau [21,22], Vincent Probst[21], Ronan Roussel[23,24], Inez J. Wijdh-Den Hamer [25], Joylene E. Siland [2], Kirk U. Knowlton[26,27], Jean Jacques Schott [21], Robert A. Levine [28], Emelia J. Benjamin [29,30,31], Ramachandran S. Vasan [5,29,32], Benjamin D. Horne [26,33], Joseph B. Muhlestein [26,34], Giovanni Benfari [35], Maurice Enriquez-Sarano [35], Andrea Natale[36], Sanghamitra Mohanty[36], Chintan Trivedi[36], Moore B. Shoemaker[37], Zachary T. Yoneda [37], Quinn S. Wells[38,39,40], Michael T. Baker [38], Eric Farber-Eger [38], Hector I. Michelena[35], Alicia Lundby[8,41], Russell A. Norris [42], Susan A. Slaugenhaupt[43†], Christian Dina [21†], Steven A. Lubitz [1,6,15†], Nabila Bouatia-Naji [3†], Patrick T. Ellinor [1,6,15*†], and David J. Milan[6,44*†]

[1]Cardiovascular Disease Initiative, The Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA; [2]Department of Cardiology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands; [3]Université de Paris, PARCC, Inserm, F-75015 Paris, France; [4]Division of Cardiovascular Medicine, Brigham and Women's Hospital, Boston, MA, USA; [5]School of Public Health, Boston University, Boston, MA, USA; [6]Cardiovascular Research Center, Massachusetts General Hospital, Boston, MA, USA; [7]Division of Cardiology, University of California San Francisco, San Francisco, CA, USA; [8]Department of Biomedical Sciences, Faculty of Health and Medical Sciences, University of Copenhagen, København 2200, Denmark; [9]Department of Cardiology, The Heart Centre, Rigshospitalet, Copenhagen University Hospital, Copenhagen, Denmark; [10]Department of Forensic Medicine, Faculty of Medical Sciences, University of Copenhagen, Copenhagen, Denmark; [11]Assistance Publique–Hôpitaux de Paris, Departments of Cardiology and Genetics, Hôpital Européen Georges Pompidou, 75015 Paris, France; [12]Bordeaux Population Health Research Center, Inserm Center U1219, University of Bordeaux, Bordeaux, France; [13]Department of Neurology, Bordeaux University Hospital, Inserm U1219, Bordeaux, France; [14]Univ. Lille, Inserm, Centre Hosp. Univ Lille, Institut Pasteur de Lille, UMR1167 – RID-AGE- Risk factors and molecular determinants of aging-related diseases, F-59000 Lille, France; [15]Demoulas Center for Cardiac Arrhythmias, Massachusetts General Hospital, Boston, MA, USA; [16]Department of Anesthesiology, Perioperative and Pain Medicine, Brigham and Women's Hospital, Boston, MA, USA; [17]Department of Anesthesiology, Boston University School of Medicine, Boston, MA, USA; [18]Duke Molecular Physiology Institute, Duke University, Durham, NC, USA; [19]Division of Cardiology, Department of Medicine, Duke University School of Medicine, Durham, NC, USA; [20]Department of Medicine, Aga Khan University, Karachi, Pakistan; [21]l'institut du thorax, INSERM, CNRS, Univ Nantes, CHU Nantes, Nantes, France; [22]l'institut du thorax, CHU Nantes, Nantes, France; [23]Cordeliers Research Centre, ImMeDiab Team, INSERM, Université de Paris, Paris, France; [24]Hôpital Bichat-Claude-Bernard, APHP, Department of Diabetology, Paris, France; [25]Department of Cardiothoracic Surgery, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands; [26]Intermountain Medical Center Heart Institute, Salt Lake City, UT, USA; [27]Division of Cardiovascular Medicine, Department of Medicine, University of California San Diego, San Diego, CA, USA; [28]Cardiac Ultrasound Laboratory, Harvard Medical School, Massachusetts General Hospital, Boston, MA, USA; [29]National Heart, Lung, and Blood Institute's and Boston University's, The Framingham Heart Study, Framingham, MA, USA; [30]Section of Cardiovascular Medicine, Boston University School of Medicine, Boston, MA, USA; [31]Department of Epidemiology, Boston University School of Public Health, Boston, MA, USA; [32]School of Medicine, Boston University, Boston, MA, USA; [33]Division of Cardiovascular Medicine, Department of Medicine, Stanford University, Stanford, CA, USA; [34]Cardiology Division, Department of Internal Medicine, University of Utah, Salt Lake City, UT, USA;

* Corresponding authors. Tel: +1 617 459 4688, Email: dmilan@leducq.com (D.J.M.); Tel: +1 617 724 8729, Email: ellinor@mgh.harvard.edu (P.T.E.)
† These authors contributed equally.

[35]Department of Cardiovascular Medicine, Mayo Clinic, Rochester, MN, USA; [36]Texas Cardiac Arrhythmia Institute, St David's Medical Center, Austin, TX, USA; [37]Department of Medicine, Division of Cardiovascular Diseases, Vanderbilt University Medical Center, Nashville, TN, USA; [38]Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA; [39]Department of Pharmacology, Vanderbilt University Medical Center, Nashville, TN, USA; [40]Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, USA; [41]The Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, København 2200, Denmark; [42]Cardiovascular Developmental Biology Center, Department of Regenerative Medicine and Cell Biology, Medical University of South Carolina, Charleston, SC, USA; [43]Center for Genomic Medicine and Department of Neurology, Massachusetts General Hospital Research Institute and Harvard Medical School, Boston, MA, USA; and [44]Leducq Foundation, Boston, MA 02110, USA

Received 2 February 2021; revised 18 August 2021; accepted 1 February 2022

See the editorial comment for this article 'Mitral valve prolapse: will genetics finally solve the puzzle?', by T. Trenkwalder and M. Krane, https://doi.org10.1093/eurheartj/ehac048.

Abstract

Aims Mitral valve prolapse (MVP) is a common valvular heart disease with a prevalence of >2% in the general adult population. Despite this high incidence, there is a limited understanding of the molecular mechanism of this disease, and no medical therapy is available for this disease. We aimed to elucidate the genetic basis of MVP in order to better understand this complex disorder.

Methods and results We performed a meta-analysis of six genome-wide association studies that included 4884 cases and 434 649 controls. We identified 14 loci associated with MVP in our primary analysis and 2 additional loci associated with a subset of the samples that additionally underwent mitral valve surgery. Integration of epigenetic, transcriptional, and proteomic data identified candidate MVP genes including *LMCD1*, *SPTBN1*, *LTBP2*, *TGFB2*, *NMB*, and *ALPK3*. We created a polygenic risk score (PRS) for MVP and showed an improved MVP risk prediction beyond age, sex, and clinical risk factors.

Conclusion We identified 14 genetic loci that are associated with MVP. Multiple analyses identified candidate genes including two transforming growth factor-β signalling molecules and spectrin β. We present the first PRS for MVP that could eventually aid risk stratification of patients for MVP screening in a clinical setting. These findings advance our understanding of this common valvular heart disease and may reveal novel therapeutic targets for intervention.

Key question

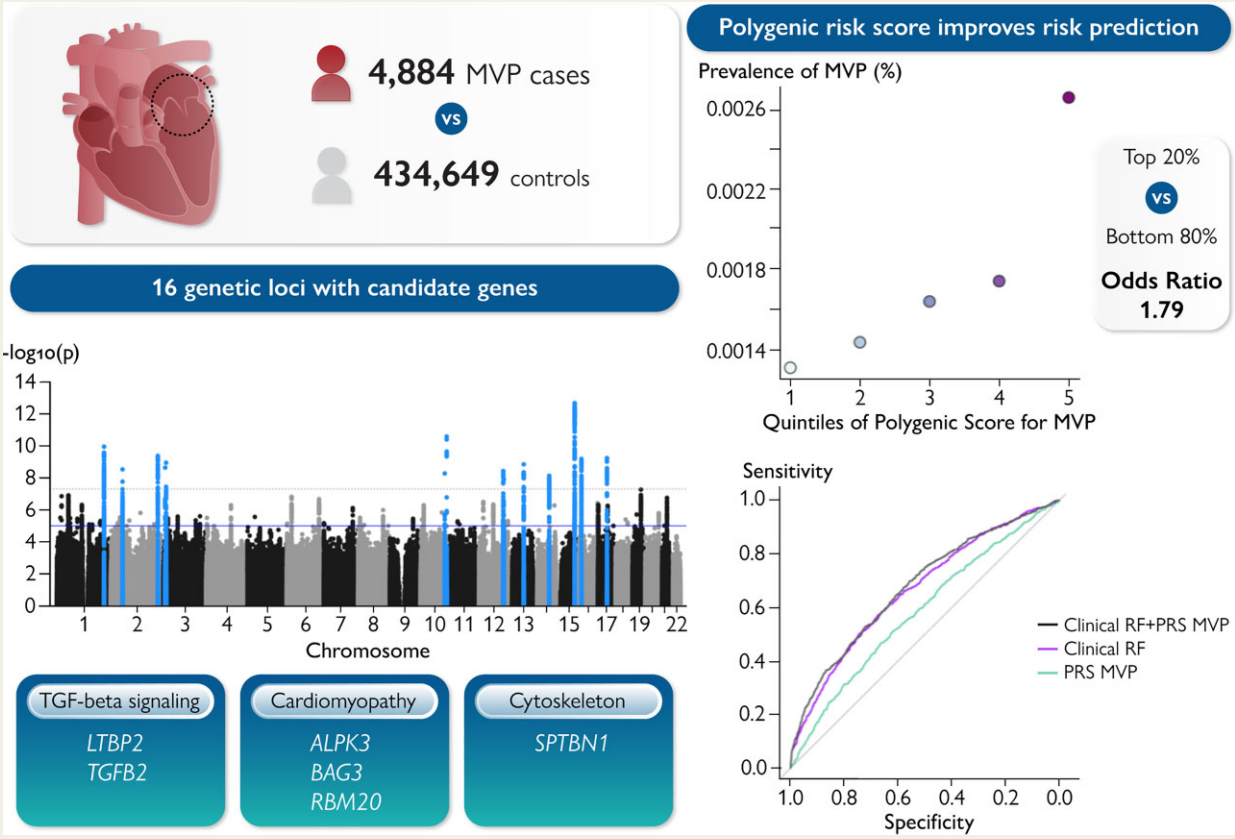
Expand our understanding of the genetic basis for mitral valve prolapse (MVP). Uncover relevant pathways and target genes for MVP pathophysiology. Leverage genetic data for MVP risk prediction.

Key finding

Sixteen genetic loci were significantly associated with MVP, including 13 novel loci. Interesting target genes at these loci included *LTBP2*, *TGFB2*, *ALKP3*, *BAG3*, *RBM20*, and *SPTBN1*. A risk score including clinical factors and a polygenic risk score, performed best at predicting MVP, with an area under the receiver operating characteristics curve of 0.677.

Take-home message

Mitral valve prolapse has a polygenic basis: many genetic variants cumulatively influence pre-disposition for disease. Disease risk may be modulated via changes to transforming growth factor-β signalling, the cytoskeleton, as well as cardiomyopathy pathways. Polygenic risk scores could enhance the MVP risk prediction.



Structured Graphical Abstract This study meta-analysed 4884 mitral valve prolapse (MVP) cases versus 434 649 controls, and discovered 16 genetic loci associated to MVP. Downstream analyses implicated candidate genes involved in TGF-beta signalling, cardiomyopathy and the cytoskeleton. The results from the meta-analysis were used to calculate a polygenic risk score (PRS) to aid prediction of MVP. Adding the PRS to a model with age, sex and clinical risk factors improved MVP risk prediction. Abbreviations, MVP, mitral valve prolapse, p, p-value, PRS, polygenic risk score, RF, risk factors.

Keywords Genome-wide association study • Mitral valve prolapse • Proteomics • RNA-sequencing • Genetic correlation • Polygenic risk score

Translational Perspective

The findings from this genetic analysis may uncover novel causal pathways underlying the pathophysiology of mitral valve prolapse (MVP)—a disease for which no medical therapy currently exists. These insights could lead to the identification of new therapeutic targets, prognostic markers, and modifiable pathways. Additionally, the results from the genome-wide association study may facilitate a step towards personalized medicine in the field of MVP disease. The genetic risk score for MVP may also allow an individualized risk assessment for patients. However, further research is needed to uncover a larger proportion of the underlying genetic basis of MVP.

Introduction

Mitral valve prolapse (MVP) is a common cause of primary mitral valve (MV) insufficiency with a prevalence of ~2.4% in the general adult population.[1] The majority of patients with MVP have a benign course; however, a subset of patients can develop a significant morbidity including severe MV insufficiency, heart failure (HF), endocarditis, atrial fibrillation (AF), and sudden cardiac death.[1] The chronic volume overload of mitral regurgitation can lead over time to dilation of the left ventricle, decline in ejection fraction, and the need for surgical or percutaneous intervention, as described in the latest guidelines from the *European Society of Cardiology* (ESC) on the management of valvular heart disease.[2] The long latent period between initial diagnosis and the need for intervention makes this disease a prime target for medical therapeutic treatment. However, the underlying biological mechanisms and pathways remain elusive, limiting our ability to intervene and alter the natural history of this morbid disease.

Although the vast majority of MVP is sporadic, MVP can less commonly manifest as one component of a multi-system syndromic connective tissue disorder (e.g. Marfan syndrome, Loeys–Dietz syndrome) where transforming growth factor (TGF)-β signalling has been implicated in the pathogenesis. More commonly, MVP presents as a distinct entity without syndromic features where our understanding of the molecular pathogenesis is limited. Familial inheritance and clustering have been demonstrated in several studies and have provided some clues to the molecular causes.[3–6] Two genes in particular, *DCHS1*[7] and *DZIP1*,[8] have been implicated as causal for MVP in inherited cases, but their relevance to sporadic cases remains unclear. A prior genome-wide association study (GWAS) revealed six loci associated with MVP confirming the complex and heterogeneous genetic pathways that are associated with myxomatous degeneration of the mitral valve.[9] Despite the progress made, actionable targets have not yet been identified and our understanding of the genetic underpinnings of MVP remains nascent.

In this study, we sought to identify novel genetic signals for non-syndromic MVP by conducting the largest meta-analysis of GWAS to date. We included >4800 MVP cases and then sought to identify candidate genes at MVP loci by integrating our genetic results with transcriptional and proteomic data from the mitral valve. Finally, we calculated a polygenic risk score (PRS) for MVP based on the meta-analysis results and tested its predictive power in a large-scale biobank.

Methods

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contacts, D.J.M. (dmilan@leducq.com) and P.T.E. (ellinor@mgh.harvard.edu).

Data availability

The summary level result files and the weight files for the MVP PRS generated by this study are available on the Cardiovascular Disease Knowledge Portal at https://cvd.hugeamp.org/. Participant level data for the Framingham Heart Study (FHS) are available at dbGaP.

Study participants

We leveraged six international cohorts/biobanks, including the FHS, the Mass General Brigham (MGB) Biobank, the UK Biobank, the Broad Cardiovascular Disease Initiative (Broad CVDi), the MVP-FRANCE/Three-City Study (3C), and the MVP-NANTES/D.E.S.I.R., to conduct the largest meta-analysis of genetic association studies for MVP. The study design and rationale for the contributing cohorts have been previously published[9–12] and study descriptions are available in Supplementary material online, *Note*. Case/control distribution across the contributing cohorts was as follows: the Broad CVDi (1862/9803), the MVP-FRANCE/3C (1007/1469), the MVP-NANTES/D.E.S.I.R. (478/1283), the UK Biobank (691/392 538), the MGB Biobank (677/23 981), and the FHS (169/5 575) (see Supplementary material online, *Table S1*). In brief, MVP was defined according to contemporary echocardiographic society guidelines as leaflet displacement ≥2 mm beyond the mitral annulus in a long-axis view at end-systole involving one or both MV leaflets when raw imaging data were available for review or via electronic medical record search for MVP diagnosis using ICD and procedure codes or natural language processing (see Supplementary material online, *Note* and *Table S2*). This study complied with the Declaration of Helsinki. The local ethics committees of each study have approved the research protocol and informed consent was obtained from all the study participants.

Genotyping and imputation

Genotyping arrays, calling algorithms, and quality control steps for each study are listed in Supplementary material online, *Table S3*. Variants were filtered prior to the imputation based on quality control measures, including call rate, Hardy–Weinberg equilibrium, minor allele frequency, and additional study-specific criteria (if applicable). Samples were removed due to low call rate and excess heterozygosity. Most studies were imputed with the TOPMed freeze5 imputation panel on the Michigan imputation server[13] using the Eagle v2.4[14] for phasing and Minimac4 (https://genome.sph.umich.edu/wiki/Minimac4) for imputation. The imputation procedure for the UK Biobank is described elsewhere.[12]

Analysis of open chromatin regions in mitral valve and cardiac tissue

Sample acquisition and processing of the mitral valves from mice were described previously.[15] A complete list of Encyclopedia of DNA Elements (ENCODE) data sets used to compare dermal and cardiac fibroblasts samples is available in Supplementary material online, *Table S4*. A brief description of the sample processing can be found in Supplementary material online, *Methods*.

Proteome measurements of mitral valves from humans

Mitral valve non-diseased tissue from five individuals were collected at the Department of Forensic Medicine, University of Copenhagen, and stored at −80°C. Donor information is listed in Supplementary material online, *Table S5*. Further details on tissue processing off-line peptide fractionation, Liquid chromatography-tandem mass spectrometry (LC-MS/MS) measurements, and analytical processing can be found in Supplementary material online, *Methods*.

Genome-wide association studies

The study-specific GWAS results for the Broad CVDi study, the MGB Biobank, and the UK Biobank were calculated using a generalized mixed model framework and the analytical tool SAIGE[16] to account for case–control imbalance. The results for the cohorts MVP-NANTES/D.E.S.I.R. and MVP-FRANCE/3C were calculated in a logistic regression using the SNPTEST.[17] Framingham used the R-package GWAF.[18] All studies included principal components as covariates to correct for ancestry effects, as well as additional covariates including sex, age, or genotyping array. Details on the statistical analysis of the primary GWAS results per study are listed in Supplementary material online, *Table S3*. All study-specific results underwent quality control checks including alignment of allele frequencies to the TOPMed freeze5 reference, inspection of the Manhattan plot, QQ plot, and PZ plot, and appropriate genomic control. We restricted the analysis to variants included in the TOPMed freeze5 reference panel according to the available site files (https://bravo.sph.umich.edu/freeze5/hg38/download). All result files were successfully annotated with GRCh37 and GRCh38 base pair coordinates using the command line LiftOver tool (https://genome.ucsc.edu/cgi-bin/hgLiftOver). Variants were included if imputation quality was >0.3, minor allele frequency × *n* cases × imputation quality was >10, and the discrepancy of the reported study allele frequency to the allele frequency in the TOPMed freeze5 reference panel was <|0.3|.

Meta-analysis

Meta-analyses were run using the METAL v2018-08-28.[19] We applied an inverse variance weighted meta-analysis with the classical approach of using effect sizes and standard errors. A study-specific genomic control correction was applied. Allele frequencies and imputation quality were calculated as an average across the studies. Variants present in just one study were excluded. Base pair locations are reported in GRCh38 throughout this work. The *P*-value cut-off of 5×10^{-8} was applied for a genome-wide significance. Genetic loci at a genome-wide significance have a minimum distance of 500 kb between each other and have also been confirmed as independent through the conditional and joint (COJO) analysis (see description below). The lead variant at each locus is defined as the variant with the lowest *P*-value. Heterogeneity between the studies was evaluated with the *I*-squared (I^2) statistic and significance was determined using the Bonferroni correction for multiple testing of the lead variants (0.05/14 for MVP). The linkage disequilibrium (LD) region is defined as the region spanning all variants with $R^2 \geq 0.6$ to the lead variant. The LD is calculated from the samples of the Broad CVDi study (see description below). The overlapping genes with the LD region are selected based on the longest transcript from all collapsed isoforms based on the gene reference from the Ensemble v84.

Sensitivity analyses

We performed two sensitivity analyses. First, we restricted all MVP cases to individuals that had also undergone MV surgery and recalculated the primary GWAS for each participating study (except the FHS, where the MV surgery phenotype was not available). Case/control distribution for the MV surgery analysis was as follows: the Broad CVDi (359/9803), the MVP-FRANCE/3C (675/1465), the MVP-NANTES/D.E.S.I.R. (474/1283), the UK Biobank (261/392 538), and the MGB Biobank (134/23 827). We then meta-analysed these results using the METAL v2018-08-28 as previously described for the primary analysis.

In a second sensitivity analysis, we excluded individuals with reported case status for either prevalent atrial fibrillation or prevalent heart failure (noAFHF) from the primary set. We again recalculated the summary level results for each cohort (except the MVP-FRANCE, where neither AF nor HF case status was available). Within the Broad CVDi study, a subset of participants had missing status for AF or HF and were excluded. Case/control distribution for the MVP-noAFHF analysis was as follows: the Broad CVDi (814/7981), FHS (167/5485), the MVP-NANTES/D.E.S.I.R. (195/1283), the UK Biobank (544/384 546), and the MGB Biobank (322/18 834). The meta-analysis was performed as previously described for the primary analysis.

Conditional and joint analysis

The COJO analysis on the summary level results for the MVP and MV surgery meta-analyses was calculated with the GCTA-COJO version 1.93.2 beta.[20] As LD reference, we utilized the TOPMed-imputed genotypes of the 11 665 European ancestry samples from the Broad CVDi study that were included in the primary MVP GWAS analysis. Imputed allelic dosages of the LD reference were converted to hard calls in the PLINK v2.0[21] (www.cog-genomics.org/plink/2.0/) with a hard-call-threshold of 0.2. The COJO analysis was performed with the *P*-value cut-off of $P < 5 \times 10^{-8}$. We utilized the LD pair tool to calculate a pairwise LD for all top hits from the MVP and MV surgery GWAS, based on the European ancestry samples from the TOPMed LD reference (https://ldlink.nci.nih.gov).

Linkage disequilibrium score regression and heritability analyses

The LD score regression (LDSC) intercept, single nucleotide polymorphism (SNP) heritability on the liability scale, and partitioned heritability were calculated with the LDSC v1.0.1 (https://github.com/bulik/ldsc). Further details on the analysis can be found in Supplementary material online, *Methods*.

Associations to other phenotypes

The Phenoscanner (http://www.phenoscanner.medschl.cam.ac.uk/) was utilized to look up the sentinel variants at the genome-wide significant loci for the MVP and the MV surgery GWAS results. We included proxy variants with $R^2 > 0.5$ based on the European ancestry LD in the lookup. Reported associations were filtered for *P*-value $< 5 \times 10^{-8}$ and similar traits were reported as a group (e.g. low-density cholesterol, HDL, and total cholesterol are reported under lipid traits). We also investigated whether the identified sentinel variants were located within a previously reported MVP locus[9] (≤500 kb distance between the sentinel variants) and are reporting the LD between those sentinel variants based on the European ancestry samples from 1000 genomes.[22]

Genetic correlation of mitral valve prolapse with other cardiac phenotypes

The estimates and tests were performed using the LDSC method.[23] The results shown correspond to the following GWAS analyses: AF,[10] HF,[24] non-ischaemic cardiomyopathy[25] (NICM), heart rate,[26] PR interval,[27] and cardiac magnetic resonance-derived left ventricular measurements.[28]

Polygenic risk score for mitral valve prolapse

A meta-analysis with the METAL v2018-08-28[19] was performed excluding the summary level results from the UK Biobank. A total of 4193 MVP cases and 42 111 controls were included across the remaining five summary level results. We used these results to derive a PRS with the polygenic risk score-continuous shrinkage (PRS-CS)[29] method using the European samples from the 1000 Genomes Project phase 3[22] as provided by the PRS-CS. We set the global shrinkage parameter (PARAM_PHI) to 1×10^{-2} as recommended for disease GWAS for polygenic traits with limited samples size. All other parameters were left at the default setting. We only included variants available in our test set, the UK Biobank, and a total of 1 097 364 variants were included in the PRS. The PRS was calculated with the PRS-CS-derived weights, using the PLINK v2.00a3LM for the 691 MVP cases and 392 538 in the UK Biobank based on allelic dosages and the flag *scoresum*. The PRS was inverse-normal transformed to have the odds ratios on the standard deviation scale. A logistic regression was performed in the R v3.6.0 to test the association of the quantitative PRS with MVP. We also tested the performance of the models including age and sex, as well as PRS, age, and sex. Two additional models included clinical risk factors [hypertension, all-cause HF, myocardial infarction and diabetes (Type 1 and Type 2)] at baseline, and clinical risk factors plus PRS. Each model included the additional covariates: genotyping array and the first five principal components for genetic ancestry. The package *pROC*[30] in R was used to calculate the area under the receiver operating characteristics curve (AUC) and its 95% confidence interval (CI). We also associated the dichotomized PRS (highest quintile vs. bottom 80%) with MVP in a logistic regression model with the covariates age, sex, genotyping array, and the first five principal components for genetic ancestry, as well as clinical risk factors. The definitions of the clinical risk factors in the UK Biobank can be found in Supplementary material online, *Methods*.

Results

Meta-analysis of mitral valve prolapse

We conducted a meta-analysis, sensitivity analyses, and a number of data-driven downstream methods, including intersecting our results with epigenetic and proteomics data from the MV (*Figure 1*). We conducted the primary meta-analysis across cohorts and biobanks of European ancestry and identified 14 loci associated with MVP at genome-wide significance (*P*-value $<5 \times 10^{-8}$, *Figure 2* and *Table 1*, see Supplementary material online, *Figure S1A–N*). The 14 loci were separated by >500 kb and were independently associated with MVP by the COJO analysis with a stepwise procedure (see Supplementary material online, *Table S6*). The conditional analysis did not reveal independent secondary signals with a genome-wide significance at the loci. Of the 14 identified loci, 12 were novel and 2 had previously been reported to be associated with MVP.[9] These two loci reside at 2q35 (lead variant: rs13399995, *P*-value $= 4.2 \times 10^{-10}$, locus: *TNP1/DIRC3/TNS1*) and 3p25.3 (lead variant: rs165177, *P*-value $= 2.3 \times 10^{-9}$, locus: *LMCD1*). All of the lead variants at the MVP loci were common, with minor allele frequencies ranging from 4 to 50% and well imputed, with an average imputation quality between 0.898 and 0.997. There was no significant heterogeneity observed across the studies (*Table 1*). The genomic inflation factor (λ_{GC}) was 1.097 and the LDSC intercept was 1.0222.

The overall SNP heritability on the liability scale was observed as 0.2234. We calculated the partitioned heritability based on the LDSC baseline model for functional annotations of the genome across 50 categories. We observed significant positive enrichment across 12 of the 50 functional elements, which largely fell into the categories of super enhancers, transcription factor binding site, conserved regions, and regions of active transcription. At repressed regions ('Repressed + 500 pb', *P*-value $= 1.81 \times 10^{-6}$) less heritability than the average was harboured (see Supplementary material online, *Table S7* and *Figure S2*).

Sensitivity analyses

First, since a prior meta-analysis was heavily enriched for cases that had undergone MV surgery, we performed a sensitivity analysis including only the subset of cases ($n = 1903$) that had a diagnosis for MVP and underwent valve surgery. This sensitivity analysis was performed in an effort to examine a sub-population with a more severe MVP phenotype. We observed that the direction of association of the MVP risk alleles at the 14 sentinel variants was concordant between the primary and sensitivity analysis (see Supplementary material online, *Figure S3*). The MV surgery GWAS revealed seven genome-wide significant loci (*P*-value $<5 \times 10^{-8}$) all independently associated with the outcome (*Table 2* and Supplementary material online, *Table S6*) and with concordant risk compared with the primary MVP GWAS results (see Supplementary material online, *Figure S3*). Five of the seven loci overlapped with the primary MVP GWAS results. Two additional loci detected in the MV surgery analysis were located at 8p23.1 (lead variant: rs12676417, *P*-value 1.6×10^{-8}, locus: *PINX1*) and 17p13.3 (lead variant: rs112258894, *P*-value 1.6×10^{-11}, locus: *SMG6*) (*Table 2* and Supplementary material online, *Figure S1O and P*). Most of the sentinel variants identified for the MV surgery analysis had larger effect sizes than those identified in the overall MVP GWAS (see Supplementary material online, *Figure S3*). In sum, we detected a total of 13 novel loci in our work.

Second, we performed a sensitivity analysis that was limited to cases ($n = 2042$) and controls ($n = 418\,129$) without prevalent AF or HF (MVP-noAFHF). The correlation of the *Z*-scores was $R^2 = 0.69$ overall, $R^2 = 0.91$ for variants with a nominal significance (*P*-values $<5 \times 10^{-2}$) and $R^2 = 0.99$ for variants with a genome-wide significance (*P*-value $<5 \times 10^{-8}$) compared with the primary analysis (see Supplementary material online, *Figure S4* and *Table S8*). For all 14 tophits from the primary analysis, we observe the same direction of effect in this sensitivity analysis. The effects of 11 of the 14 tophits show a modest attenuation in the MVP-noAFHF sensitivity analysis (see Supplementary material online, *Figure S5*).

Linking genes at mitral valve prolapse loci to protein expression in human mitral valve and RNA-sequencing data from murine mitral valve tissue

We interrogated bulk RNA-sequencing (RNA-Seq) data of cardiac ($n = 3$) and MV tissue ($n = 2$) in normal mice[8] and found detectable levels of RNA in either cardiac or MV tissue for 20 genes at MVP loci (see Supplementary material online, *Table S9*). In MV tissue, the five most highly expressed genes were *Sec11A*, *Cand2*, *Sptbn1*, *Mzt1*, and *Lmcd1* (see Supplementary material online, *Figure S6*). The top five expressed genes in cardiac tissue were *Cand2*, *Alpk3*, *Rbm20*, *Tgfb2*, and *Atxn2* (see Supplementary material online, *Figure S7*).

We measured protein levels in human MV tissue from five donors without known valve disease and free of heart disease (see Supplementary material online, *Table S5*). The five genes with the highest detected protein levels for genes at MVP GWAS loci in human MV were LTBP2, SPTBN1, LMCD1, BAG3, and SEC11A (see Supplementary material online, *Figure S8* and *Table S10*). High levels of SPTBN1 and LMCD1 were consistent with the RNA-Seq results from the mouse MV tissue, as is the verification of protein expression in human MV of SEC11A and CAND2. We also confirmed protein expression of ATXN2, TGFB2, RBM20, and ALPK3 in the MV.

Genetic correlation analysis with other cardiovascular conditions and previously reported genome-wide association study

Genetic correlation testing, using the LDSC, for MVP with related diseases, electrocardiographic traits, and cardiac magnetic resonance image traits, revealed a positive genetic correlation with AF (*P*-value $= 1.35 \times 10^{-2}$) and longer PR interval (*P*-value $= 8.29 \times 10^{-4}$). Furthermore, larger ventricular volumes (left ventricular end-diastolic volume, *P*-value $= 3.6 \times 10^{-12}$; left ventricular end-systolic volume, *P*-value $= 2.72 \times 10^{-13}$), larger stroke volume (*P*-value $= 3.49 \times 10^{-4}$), all indexed by body surface area, were positively correlated with MVP. We observed a trend towards a positive correlation between MVP and NICM (*P*-value $= 0.0593$) and a negative genetic correlation with left ventricular ejection fraction (*P*-value $= 2.62 \times 10^{-6}$, *Figure 3*).

We found that sentinel variants and their LD-proxies ($R^2 > 0.5$), at 2 of 14 loci were reportedly associated with other traits. Those

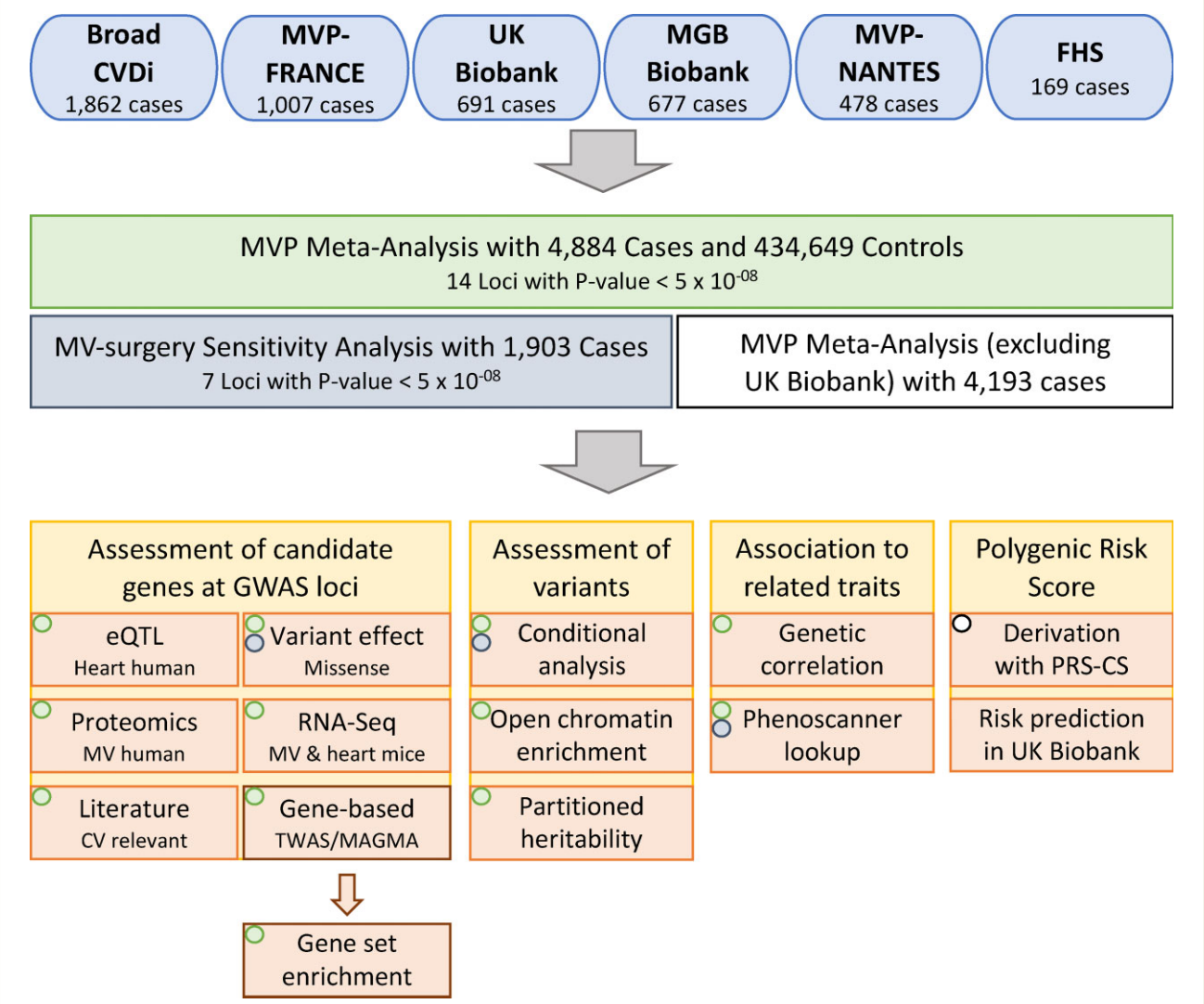


Figure 1 Overview of the study sample and performed analyses. The upper part of the figure shows the six included biobanks, registries, and studies that provided summary level results for primary and secondary meta-analyses and their respective case counts for mitral valve prolapse. The middle section shows the two main analyses of this work with total number of cases and controls that were included, and the total number of loci that were found in each. The primary meta-analysis was based on the summary statistics for the mitral valve prolapse phenotype. The secondary sensitivity analysis was based on the subset of patients who underwent mitral valve surgery in addition to their mitral valve prolapse diagnosis. Additionally, a meta-analysis was calculated without the UK Biobank for the purpose of creating a polygenic risk score. The lower part of the figure summarizes the downstream work that was performed. Analyses to evaluate the candidate genes at the genome-wide association study loci included eQTL lookups in heart tissue from the genotype-tissue expression, assessment of the variant effects regarding missense variation, generation, and interrogation of a protein expression data set of mitral valve tissue in humans and interrogation of RNA-Seq data set from mitral valve and heart tissue from mice, literature evaluation of candidate genes, and gene-based test based on the summary level results and including gene-set enrichment analyses as a follow-up. Furthermore, we assessed variants in a conditional and joint analysis to establish independent signals and their overlap with open chromatin regions from mitral valve and heart tissue, primarily based on the ATAC-Seq data. We also evaluated the overlap with functional elements in the genome in a partitioned heritability analysis. We checked whether mitral valve prolapse had a shared genetic architecture with related traits via genetic correlation analyses. We assessed the association of the tophits to other genome-wide association study traits with the Phenoscanner tool. Finally, we derived a polygenic risk score with the tool PRS-CS and tested the score in the UK Biobank. For each downstream analyses, the circles indicate whether the analysis was performed for the primary results (green circle), for the mitral valve surgery results (grey circle), or for the meta-analysis without the UK Biobank (white circle with black outline). ATAC-Seq, assay for transposase-accessible chromatin using sequencing; CV, cardiovascular; eQTL, expression quantitative trait locus; GWAS, genome-wide association study; MAGMA, Multi-marker Analysis of GenoMic Annotation; MV, mitral valve; MVP, mitral valve prolapse; PRS-CS, polygenic risk score-continuous shrinkage; RNA-Seq, RNA sequencing; TWAS, transcriptome-wide association study.

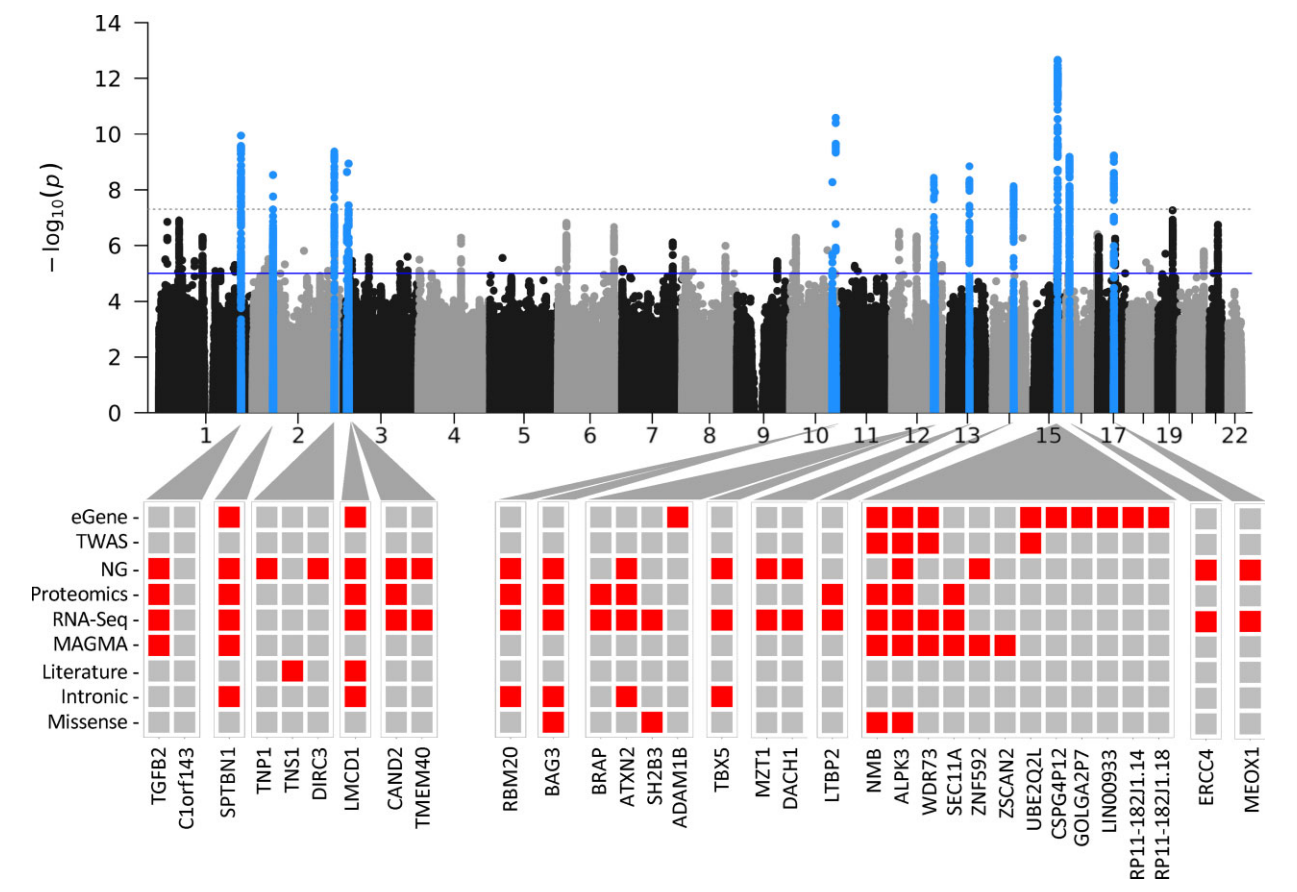


Figure 2 The Manhattan plot and evaluation of genes at each locus of mitral valve prolapse meta-analysis. The Manhattan plot of the meta-analysis results for mitral valve prolapse. Highlighted in blue are the 14 loci with a genome-wide significance (*P*-value $<5 \times 10^{-8}$). The grey dotted line indicates the genome-wide significance cut-off while the blue solid line indicates the subthreshold cut-off (*P*-value $<5 \times 10^{-6}$). Below the Manhattan plot, the candidate genes for each locus are listed and the lines of evidence for each gene are summarized. A gene at the GWAS locus was chosen as a candidate gene if the gene (i) overlapped with a linkage disequilibrium window ($R^2 \geq 0.6$) around the sentinel variant, (ii) had an eQTL to a sentinel variant, or (iii) had a genome-wide significant missense mutation. A red quadrant indicates that the evidence is present. eGene, the sentinel variant has an eQTL for the gene in cardiac atrial or ventricular tissue; TWAS, significant in the transcriptome-wide association study in either atrial or ventricular tissue; NG, nearest gene(s) to the sentinel variant; Proteomics, protein levels in human mitral valve were detected; RNA-Seq, detectable expression levels in mitral valve or heart tissue in RNA-Seq data from mice; MAGMA, significantly associated in the multi-marker analysis; Literature, gene has been implicated for mitral valve prolapse in previously published work; Intronic, the sentinel variant fell within the intron of the gene; Missense, a significantly associated variant was missense for the gene. eQTL, expression quantitative trait locus; eGene, expression quantitative trait locus gene; GWAS, genome-wide association study; LD, linkage disequilibrium; MAGMA, Multi-marker Analysis of GenoMic Annotation; MVP, mitral valve prolapse; NG, nearest gene(s); RNA-Seq, RNA sequencing; TWAS, transcriptome-wide association study.

included autism[31] and schizophrenia[32] at rs35828350 (*ZNF592*, *ALPK3*) and a number of traits at rs10774625 (*ATXN2*) including hypertension,[33] coronary artery disease,[34] and diabetes[35] (see Supplementary material online, *Table S11*).

Candidate causal genes at mitral valve prolapse genome-wide association study loci

At each of the MVP GWAS loci, we considered genes as candidates for causal genes, if they either overlapped with an LD window ($R^2 \geq 0.6$) around the sentinel variant, had an expression quantitative trait locus (eQTL) or had a genome-wide significant missense mutation. We found a total of 33 genes, with most loci having one or two candidates (see Supplementary material online, *Table S12*). *Figure 2* lower panel shows each gene per locus and the summarized evidence for the gene. The evidence indicates (i) an eGene (the sentinel variant has an eQTL for the gene in cardiac atrial or ventricular tissue, see Supplementary material online, *Table S13*), (ii) significant in TWAS (see Supplementary material online, *Table S14*), (iii) the nearest gene, (iv) protein expression in human MV tissue (see Supplementary material online, *Table S10*), (v) RNA expression in murine MV or heart tissue (see Supplementary material online, *Table S9*), (vi) significant in the MAGMA (see Supplementary material online, *Table S14* and *Figure S9*), (vii) associated with MVP in the literature (see Supplementary material online, *Table S15*),

Table 1 Fourteen loci for mitral valve prolapse in the European ancestry meta-analysis

Rsid	Chr	bp (GRCh38)	Locus	RA	NRA	RAF	OR	*P*-value	I^2_{HET}	*P*-value$_{HET}$	Rsq
rs12406058	1	218 495 537	*TGFB2/LYPLAL1*	A	G	0.04	1.52	1.1×10^{-10}	22.1	0.27	0.974
rs12713274	2	54 621 591	*SPTBN1*	A	G	0.76	1.18	3.0×10^{-9}	1.9	0.40	0.997
rs13399995	2	217 032 818	*TNP1/DIRC3/**TNS1***	C	T	0.36	1.17	4.2×10^{-10}	63.7	0.02	0.982
rs165177	3	8 561 557	***LMCD1***	T	C	0.27	1.17	2.3×10^{-9}	14.5	0.32	0.966
rs34871776	3	12 773 870	*TMEM40/CAND2*	T	C	0.48	1.15	1.1×10^{-9}	0	0.54	0.96
rs12573386	10	110 816 801	*RBM20*	C	G	0.63	1.16	5.3×10^{-9}	0	0.88	0.898
rs17099139	10	119 659 975	*BAG3*	C	G	0.73	1.19	2.6×10^{-11}	32.2	0.19	0.988
rs10774625	12	111 472 415	*ATXN2*	G	A	0.49	1.15	3.8×10^{-9}	30.9	0.20	0.993
rs2555005	12	114 402 286	*TBX5*	A	C	0.50	1.15	1.2×10^{-8}	0	0.46	0.918
rs57355895	13	72 228 232	*DACH1/MZT1*	C	CT	0.12	1.24	1.4×10^{-9}	45.5	0.10	0.987
rs11852134	14	74 635 279	*LTBP2/AREL1*	A	G	0.51	1.14	7.5×10^{-9}	8.4	0.36	0.984
rs35828350	15	84 812 610	*ZNF592/ALPK3*	A	G	0.28	1.21	2.2×10^{-13}	45	0.11	0.993
rs13334552	16	13 982 112	*ERCC4/MKL2*	A	G	0.26	1.18	6.6×10^{-10}	0	0.60	0.944
rs55674920	17	43 640 059	*ETV4/MEOX1*	G	A	0.84	1.22	5.9×10^{-10}	0	0.77	0.987

Sentinel variants and association results at the 14 genome-wide significant loci of the MVP meta-analysis. Genomic positions are listed based on the genomic build GRCh38. Locus is labelled as the nearest genes, plus genes in bold that have been implicated for MVP in a previous study.[8] The heterogeneity statistics (I^2_{HET}, *P*-value$_{HET}$) reflect the test for heterogeneity across studies in our results. The Rsq reflects the mean imputation quality across studies.
Chr, chromosome; GrCh38, genome reference consortium human build 38; HET, heterogeneity; I^2, *I*-squared; MVP, mitral valve prolapse; NRA, non-risk allele; OR, odds ratio; RA, risk allele; RAF, risk allele frequency; Rsq, r^2 imputation quality metric.

(viii) sentinel variant is intronic, or (ix) there was a coding missense variant within the significantly associated variants at the locus (see Supplementary material online, *Table S16*). The integration of this information identifies a number of strong candidate genes.

For example, the intronic regions of *LMCD1* overlapped with the MVP GWAS signal at 3p25.3 (*Figure 4A*). The sentinel variant rs165177 also overlapped with epigenetic markers, including DNase-Seq sites from embryonic and adult heart (see Supplementary material online, *Figure S10A* and *Table S17*). *LMCD1* has an eQTL in heart showing that the rs165177-T risk allele for MVP associated with decreased expression of *LMCD1*. LMCD1 was observed to be a highly abundant protein human MV tissue and *Lmcd1* was shown to be expressed in MV and cardiac tissue from RNA-Seq data from mice.

Similarly, the sentinel variant at 2p16.2 overlapped with an intronic region of *SPTBN1* (*Figure 4B*). A closer look across all genes at the locus revealed that *SPTBN1* had the lowest *P*-value in the gene-based analysis MAGMA and TWAS in atrial tissue, despite not reaching experiment-wide significance in the TWAS analysis (see Supplementary material online, *Table S14*). *SPTBN1* had an eQTL

Table 2 Seven loci for a subset of patients with mitral valve prolapse who underwent mitral valve surgery

Rsid	Chr	bp (GRCh38)	Locus	RA	NRA	RAF	OR	*P*-value	*P*-value$_{HET}$
rs34909633	2	217 029 812	*TNP1/DIRC3/**TNS1***	C	G	0.36	1.24	1.8×10^{-8}	0.95
rs165177	3	8 561 557	***LMCD1***	T	C	0.27	1.29	3.2×10^{-9}	0.29
rs12676417	8	10 811 124	*PINX1*	G	A	0.60	1.24	1.6×10^{-8}	0.34
rs1946299	12	114 399 956	*TBX5*	G	A	0.40	1.24	3.4×10^{-8}	0.95
rs888414	14	74 638 202	*LTBP2/AREL1*	G	A	0.61	1.26	8.7×10^{-10}	0.06
rs35828350	15	84 812 610	*ZNF592/ALPK3*	A	G	0.29	1.26	2.0×10^{-8}	0.07
rs112258894	17	2 254 523	***SMG6***	T	C	0.35	1.30	1.6×10^{-11}	0.40

Sentinel variants and association results at the seven genome-wide significant loci for the subset of patients with MVP who underwent mitral valve surgery. Genomic positions are listed based on the genomic build GRCh38. Locus is labelled as the nearest genes, plus genes in bold that have been implicated for MVP in a previous study.[8]
Chr, chromosome; GRCh38, genome reference consortium human build 38; HET, heterogeneity; MVP, mitral valve prolapse; NRA, non-risk allele; OR, odds ratio; RA, risk allele; RAF, risk allele frequency.

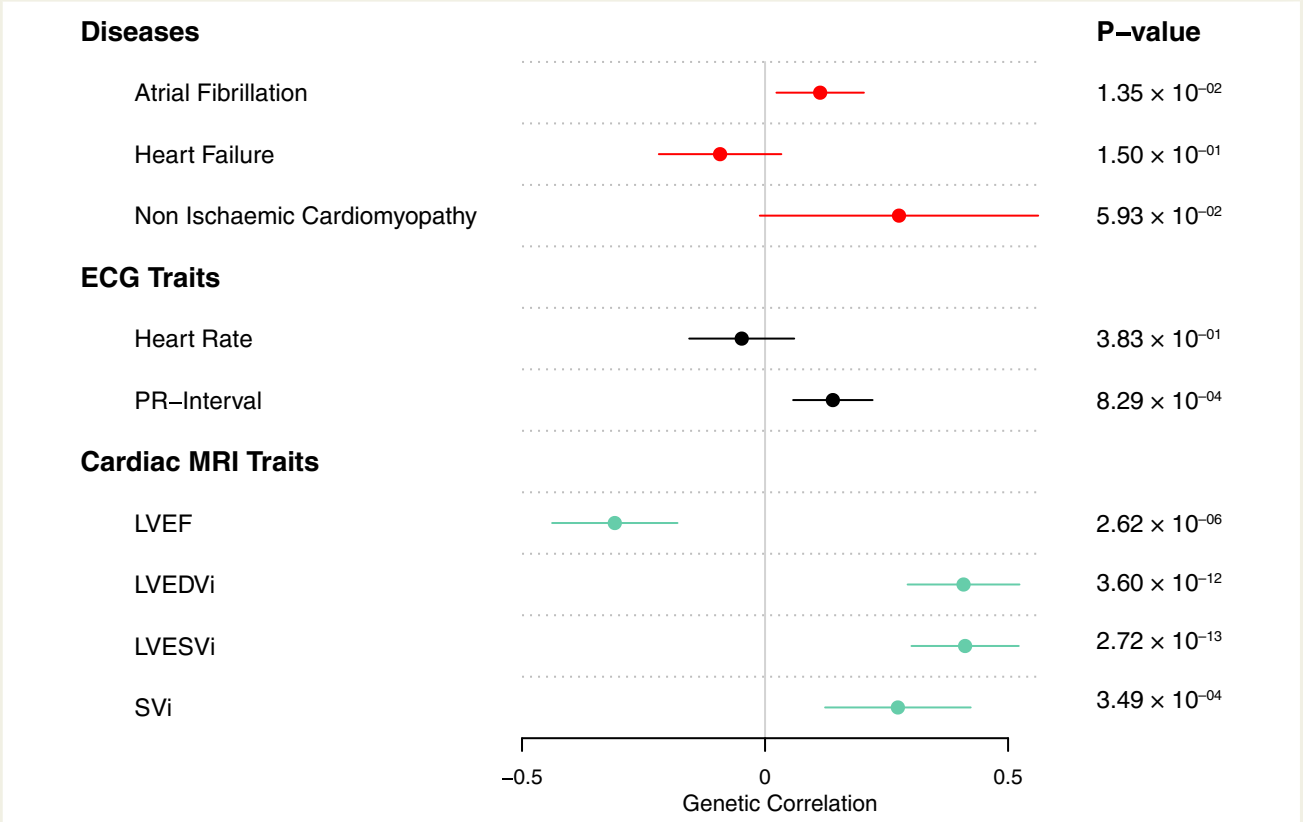


Figure 3 Genetic correlation of mitral valve prolapse with other cardiac phenotypes. Genetic correlation between mitral valve prolapse and cardiovascular phenotypes, including diseases, electrocardiographic traits, and cardiac magnetic resonance imaging traits. The genetic correlation was assessed with the linkage disequilibrium score regression method. ECG, electrocardiogram; LVEF, left ventricular ejection fraction; LVEDVi, left ventricular end-diastolic volume indexed to body surface area; LVESVi, left ventricular end-systolic volume indexed to body surface area; SVi, stroke volume indexed to body surface area; MRI, magnetic resonance imaging; MVP, mitral valve prolapse.

for the sentinel variant rs12713274 in atrial tissue, showing that the rs12713274-A risk allele for MVP correlated with decreased expression for *SPTBN1*. Comparable to the results at the LMCD1 locus, we measured high protein levels of SPTBN1 in human MV tissue, as well as expression of the *Sptbn1* gene in the RNA-Seq experiments for MV tissues of mice.

Other genes emerged as the strongest candidate at a locus but with less concordant evidence, such as *LTBP2* at 14q24.3, with high levels of protein expression and RNA-Seq confirmed gene expression in MV as the main indicators pointing towards a causal gene but no established connection from variant to gene. Similarly, *TGFB2* emerged as the strongest candidate at 1q41 as it the nearest gene, with protein expression in mitral valves and gene expression in murine MV tissue, significantly associated in the multi-marker analysis, and well known to be involved in valve pathogenesis.[36,37]

Other loci are more complex, and the evidence points towards more than one candidate genes per locus, such as *NMB* and *ALPK3* at 15q25.3, both with missense mutations, eQTLs, and significant TWAS results. Additionally, *Nmb* is expressed at very low levels in MV tissue from mice whereas there was stronger support for ALPK3 expression at the protein level in MV. *ALPK3* has prior evidence linking it to non-sarcomeric hypertrophic cardiomyopathy (HCM)[38] (see Supplementary material online, *Table S15*).

Polygenic risk score for mitral valve prolapse

We performed a meta-analysis excluding the UK Biobank and derived a PRS with the PRS-CS method.[29] The PRS-CS method uses summary level statistics as well as an external LD reference, in this case European samples from 1000 Genomes, to calculate a PRS based on roughly 1.1M SNPs. We then applied and tested our score in the UK Biobank. A higher PRS was significantly associated with an increased risk for MVP, as tested in a logistic model [odds ratio (OR) per SD = 1.33, *P*-value = 5.86×10^{-12}]. We evaluated five different models and their power to predict MVP via the AUC. A model including the PRS as covariate had moderate predictive power (AUC = 0.5838, 95% CI 0.5622–0.6053). A model including age and sex had significantly higher (*P*-value = 1.31×10^{-6}) predictive power than the PRS alone (AUC = 0.6477, 95% CI = 0.6275–0.6678). The model including PRS, age, and sex outperformed the model with age and sex with an AUC of 0.6614 (95% CI = 0.6410–0.6819, *P*-value = 0.001677) (see Supplementary material online, *Table S18* and *Figure 5A*). We next included clinical risk factors hypertension, HF, diabetes, and myocardial infarction at baseline into the prediction model. The model including PRS and clinical risk factors had a significantly higher (*P*-value = 0.003467) predictive power

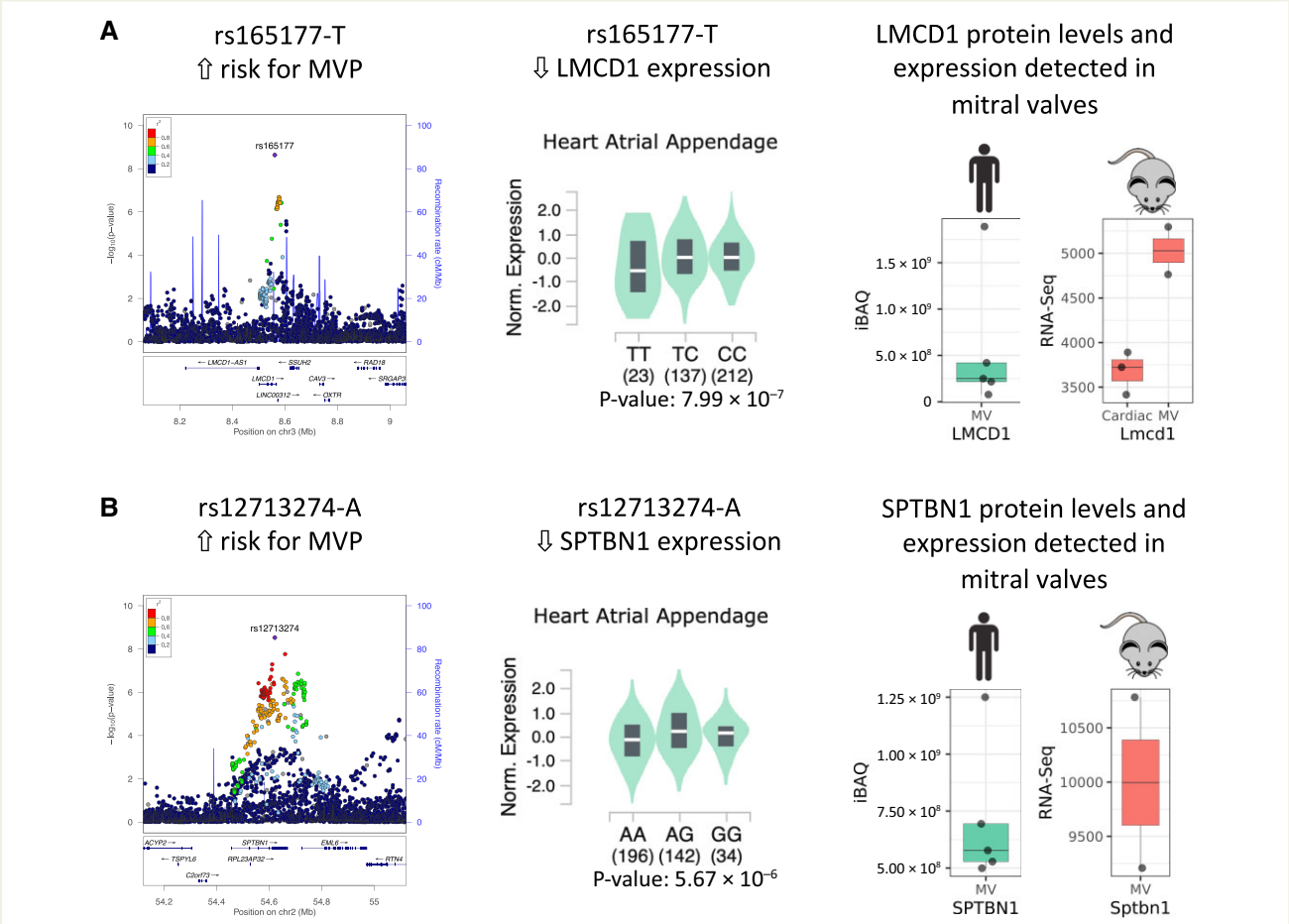


Figure 4 Locus summary for 3p25.3 and the candidate gene *LMCD1* and 2p16.2 and the candidate gene *SPTBN1*. (*A*) Summary of locus at 3p25.3. Left: regional plot showing the linkage disequilibrium pattern for the sentinel variant rs165177 based on the European ancestry linkage disequilibrium observed from 1000 genomes. The T-allele is associated with an increased risk for mitral valve prolapse. Middle: Expression quantitative trait locus for rs165177 in atrial tissue from the genotype-tissue expression. The T-allele is associated with decreased expression of the *LMCD1* gene. Right: Protein level of the candidate gene LMCD1 from human is shown in blue and RNA-Seq expression of *Lmcd1* from mouse tissue, cardiac and mitral valve shown in red. (*B*) Summary of evidence for locus at 2p16.2. Left: regional plot showing linkage disequilibrium pattern for the sentinel variant rs12713274 based on the European ancestry linkage disequilibrium from 1000 genomes. The A-allele is associated with increased risk for mitral valve prolapse. Middle: expression quantitative trait locus for rs12713274 in atrial tissue from the genotype-tissue expression. The A-allele is associated with a decreased expression of the *SPTBN1* gene. Right: Protein levels of the candidate gene SPTBN1 from human are shown in blue and RNA-Seq expression from mouse tissue for *Sptbn1*, mitral valve shown in red. GTEx, genotype-tissue expression; iBAQ, intensity-based absolute quantification; LD, linkage disequilibrium; MV, mitral valve; MVP, mitral valve prolapse; RNA-Seq, RNA sequencing; SNP, single nucleotide polymorphism.

(AUC = 0.6772, 95% CI = 0.6563–0.6982) than the model with clinical risk factors alone (AUC = 0.6649, 95% CI = 0.644–0.6858) (see Supplementary material online, *Table S18* and *Figure 5B*). Individuals in the highest quintile of the PRS had a 1.79-fold increased risk for MVP compared with the bottom 80% of individuals (*P*-value = 3.04×10^{-12}) (*Figure 5C* and *D*). This result was still significant but attenuated when including clinical risk factors into the model (OR = 1.20, *P*-value = 2.85×10^{-11}).

Discussion

We analysed 4884 cases and 434 649 controls and identified 13 novel genetic loci for MVP. While the genetic loci are important in generating a genetic risk score, insights into the pathobiology of MVP require identification of the causal gene(s) at each of these genetic loci. Therefore, we integrated transcriptional and protein expression data from the MV in order to identify a number of candidate genes at these MVP loci. Interesting candidate genes for MVP include *LTBP2*, *TGFB2*, *LMCD1*, and *SPTBN1* (*Structured Graphical Abstract*).

One of the most significant findings in our study is the genetic evidence for TGF-β signalling in MVP. Both *LTBP2* and *TGFB2* are directly involved in TGF-β signalling. The field has long been focused on a central role for this signalling pathway based on the presence of myxomatous mitral valves as part of the Marfan syndrome, a disease of excess TGF-β signalling. In addition, studies of valves from patients

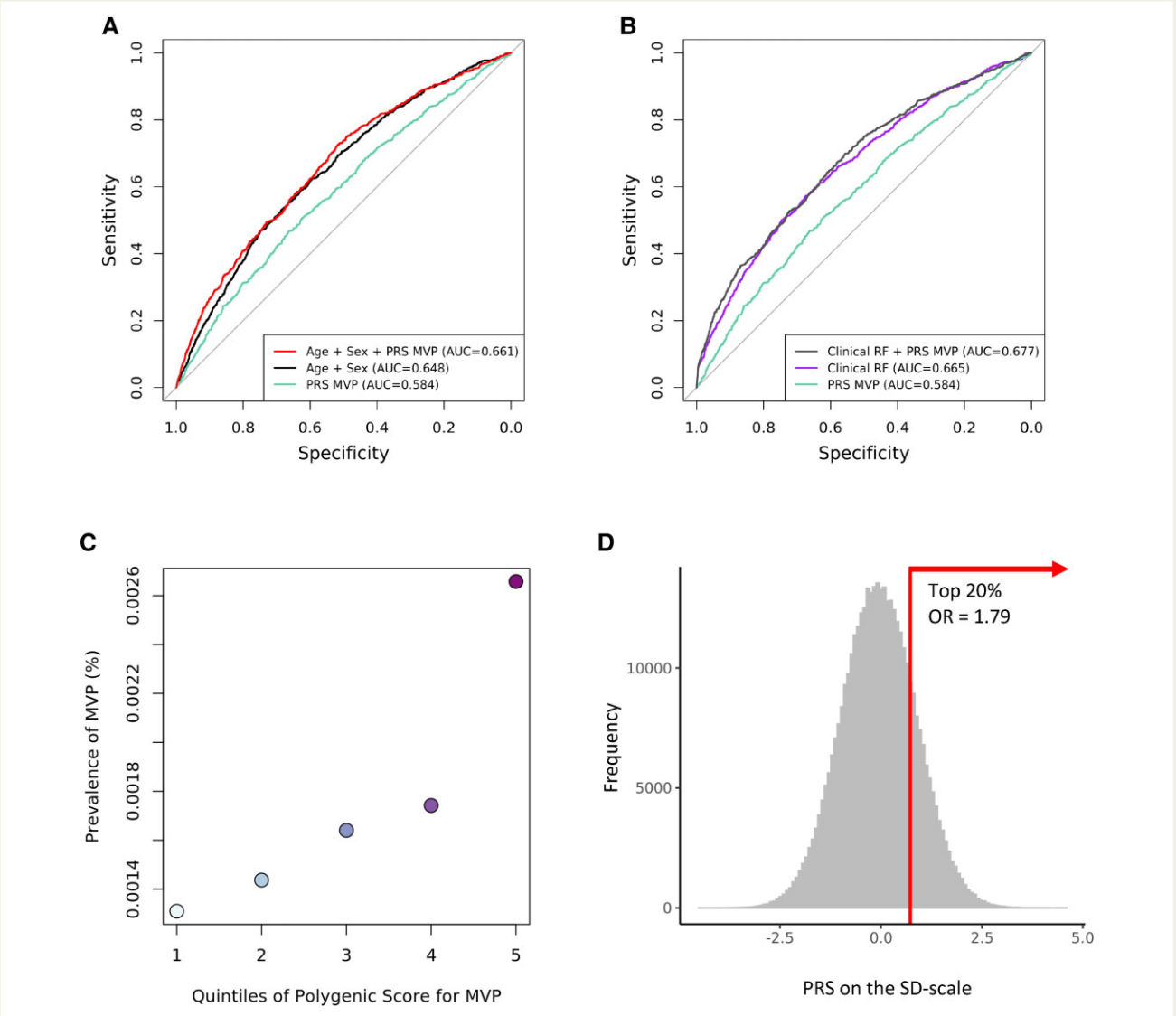


Figure 5 Performance of polygenic risk prediction of mitral valve prolapse in the UK Biobank. (*A*) Receiver operating characteristic curves of prediction models including age, sex, and polygenic risk score. (*B*) Receiver operating characteristic curves of prediction models including clinical risk factors [hypertension, all-cause heart failure, myocardial infarction, and diabetes (Type 1 and Type 2)] for mitral valve prolapse and polygenic risk score. (*C*) Quintiles of polygenic risk are based on a polygenic risk score derived on summary statistics without the UK Biobank, and using the European ancestry samples from the 1000 genomes reference set for linkage disequilibrium. The score was applied to unrelated European ancestry individuals from the UK Biobank ($n = 393\,229$). The individuals were grouped into quintiles based on their polygenic risk score and for each group, the prevalence of mitral valve prolapse was calculated. (*D*) Distribution of the polygenic risk score in the UK Biobank, marked with a red line is the top 20% cut-off for the score. The top 20% presented with an odds ratio of 1.79 compared with the bottom 80%. LD, linkage disequilibrium; MVP, mitral valve prolapse; OR, odds ratio; PRS, polygenic risk score; RF, risk factors; ROC, receiver operating characteristic.

with MVP have also implied a role for TGF-β excess. Thus, the lack of TGF-β genes in the first MVP GWAS was somewhat surprising. In this study, the finding two genes directly involved in TGF-β signalling implies a role for TGF-β in isolated MVP pathogenesis—just as in syndromic MVP. The latent TGF-β b protein 2 or *LTBP2* encodes an extracellular matrix protein that associates with fibrillin-1 containing microfibrils and is involved in regulation of TGF-β signalling. *LTBP2* has also been associated with connective tissue disorders, including isolated ectopia lentis and Weill–Marchesani syndrome.[39] Interestingly, a yet unpublished study investigating a large family of Ashkenazi-Jews with MVP found a single C to T nucleotide substitution in *LTBP2* to segregate with MVP (https://heartvalvesociety.org/meeting/abstracts/2016/P8.cgi). *TGFB2* encodes one of the TGF isoforms. Mutations in this gene can cause familial thoracic aneurysms[40] and an associated high incidence of MVP. These findings directly implicate the TGF-β pathway, in MVP, and may present an opportunity for therapeutic intervention.

The *LMCD1* gene encodes a transcriptional cofactor of *GATA6* that inhibits DNA binding for lung and cardiac specific promoters.[41] LMCD1 has been shown to activate calcineurin/NFAT, a known

signalling pathway in heart valve development.[42,43] In the current work, we found that the sentinel variant at this locus is located within a region of active chromatin and is predicted to disrupt a CREB1 binding site. Furthermore, an increased risk of MVP is linked to reduced *LMCD1* expression, a finding consistent with the observation that knockdown of *lmcd1* of in zebrafish increases atrioventricular valve regurgitation.[9] While calcineurin inhibitors are well known as therapeutic agents, including cyclosporin A and FK-506, activators of this pathway are not in clinical use and may be more challenging to develop for therapeutic purposes.

Another interesting gene resides at a novel locus for MVP is *SPTBN1*, which encodes β2-spectrin, a scaffold protein that connects the actin cytoskeleton to the plasma membrane. A study in mice suggested that β2-spectrin deficiency leads to dysregulation of the cytoskeleton and defective heart development.[44] These findings are aligned with the directionality of our result showing that MVP risk is associated with a decreased expression of *SPTBN1* in the human heart. This may imply a role of β2-spectrin during heart development in humans, where a reduction of SPTBN1 expression may potentially disrupt the cytoskeleton in human MV development.

Mutations in *ALPK3* have been associated with non-sarcomeric HCM and some forms of dilated cardiomyopathy[38] (see Supplementary material online, *Table S15*). Abnormalities in the MV are prevalent in HCM and considered a primary phenotypic expression of the pathology associated with HCM in addition to pathologic myocyte hypertrophy.[45] Elongated MV leaflets as well as myxomatous degeneration of the MV have been reported in HCM.[46] Identifying *ALPK3* as a gene associated with MVP in this study potentially highlights the overlap between myopathic and valvular pathogenic pathways. The gene-set enrichment analysis for the MVP meta-analysis results revealed the KEGG pathway for HCM and implies overlap of the genetic underpinnings of both diseases (see Supplementary material online, *Table S19*). Furthermore, two other loci spanning genes associated with dilated cardiomyopathy including *BAG3* and *RBM20* were found to be associated with MVP further highlighting the link between MVP and cardiomyopathy.[47,48] We additionally observed a strong genetic correlation with the MRI-LV measurement of left ventricular end-systolic volume. This association may be due to chronic mitral regurgitation from MVP causing left ventricular dilation directly, or could be due to a common genetic underpinning of dilated cardiomyopathy and MVP.

Finally, we have shown that the results from our meta-analysis can be utilized to construct the first polygenic risk score for MVP. With available low-cost genotyping chips, PRS could become a building block for individualized risk prediction. While the relatively small absolute risk even in the top quintile of our MVP PRS (1.79-fold increased risk) is not likely to be clinically actionable, these observations may provide a building block for future individualized risk prediction. Our findings demonstrate the potential for an application of our GWAS results in clinical genetic risk stratification that could enhance targeted screening for MVP in the future.

Limitations

Our analysis was conducted in individuals of the European ancestry and we are most likely not capturing the full scope of the human genetic variation associated with MVP. While most studies excluded cases of HCM or congenital heart diseases within the MVP case set, there may be a small overlap of these diagnoses and MVP, predominantly in the biobanks, where echocardiogram data were not available. We interrogated RNA-Seq expression patterns from cardiac and MV tissue from mice. While we expect some overlap with human expression, these results are not necessarily transferable to other species. The eQTL analyses were based on available resources for human cardiac tissue, namely atrial and ventricular tissue but not mitral tissue from diseased and non-diseased individuals, which would be the preferred source for the interrogation of our sentinel variants. While this study presents the largest GWAS for MVP to date, the size of our case set is still limited and we are likely only capturing a subset of the genetic bases for MVP. The exact mechanisms linking our reported genetic loci and candidate genes to MVP remain to be determined. While the application of PRS in clinical risk prediction is promising, the portability of a PRS into individuals with a different ancestry than the derivation cohort is limited, and comes with a reduced predictive power. Therefore, a PRS from a European ancestry analysis alone could further exacerbate existing race-based health disparities in the clinic.

Conclusions

We have identified 16 genetic loci for MVP, a condition that affects ~2.4% of adults. Our findings expand our understanding of the molecular basis for MVP and provide a series of new potential therapeutic targets for this common condition. Finally, we developed a PRS that has the potential to stratify patients by their genetic risk profile for MVP.

Supplementary material

Supplementary material is available at *European Heart Journal* online.

Acknowledgements

Acknowledgements are listed for each study in Supplementary material online, *Note*.

Conflict of interest: A.N. has consulted for Abbott, Biosense Webster, Baylis, Boston Scientific, Biotronik, and Medtronic. R.R. had grants or contracts with Sanofi, Novo Nordisk, Merck, Eli Lilly, Astra-Zeneca, Diabnext, Abbott, and Medtronic. E.J.B. has participated on the advisory board for the NIH/NHLBI CARDIA OSMB. M.R. has received grants from the Dutch Heart Foundation: RAVE V, RED-CVD, CVON-AI, DECISION studies, a grant from the SJM/Abbott: VIP-HF study (to his institution), and a grant from the Medtronic: Cryoballoon AF registry/biobank study (to his institution). P.T.E. receives sponsored research support from Bayer AG and IBM Health, and he has served on advisory boards or consulted for Bayer AG, Quest Diagnostics, MyoKardia, and Novartis. S.A.L. receives sponsored research support from the Bristol Myers Squibb/Pfizer, Bayer AG, and Boehringer Ingelheim, IBM Health and Fitbit, and has consulted for Bristol Myers Squibb/Pfizer, Blackstone Life Sciences, and Bayer AG. C.R. is supported by a grant from Bayer AG to the Broad Institute. A.H. has served on advisory boards or consulted for Amicus, Bristol Myers Squibb, Gilead, Myokardia, Sanofi Genzyme. F.N.D. has consulted for Zogenix. All other authors report no conflicts of interest.

References

1. Freed LA, Levy D, Levine RA, Larson MG, Evans JC, Fuller DL, *et al.* Prevalence and clinical outcome of mitral-valve prolapse. *N Engl J Med* 1999;**341**:1–7.
2. Vahanian A, Beyersdorf F, Praz F, Milojevic M, Baldus S, Bauersachs J, *et al.* 2021 ESC/EACTS Guidelines for the management of valvular heart disease: developed by the task force for the management of valvular heart disease of the European Society of Cardiology (ESC) and the European Association for Cardio-Thoracic Surgery (EACTS). *Eur Heart J* 2021;**43**:561–632.
3. Freed LA, Acierno JS, Dai D, Leyne M, Marshall JE, Nesta F, *et al.* A locus for autosomal dominant mitral valve prolapse on chromosome 11p15.4. *Am J Hum Genet* 2003;**72**:1551–1559.
4. Disse S, Abergel E, Berrebi A, Houot AM, Le Heuzey JY, Diebold B, *et al.* Mapping of a first locus for autosomal dominant myxomatous mitral-valve prolapse to chromosome 16p11.2-p12.1. *Am J Hum Genet* 1999;**65**:1242–1251.
5. Kyndt F, Schott JJ, Trochu JN, Baranger F, Herbert O, Scott V, *et al.* Mapping of X-linked myxomatous valvular dystrophy to chromosome Xq28. *Am J Hum Genet* 1998;**62**:627–632.
6. Nesta F, Leyne M, Yosefy C, Simpson C, Dai D, Marshall JE, *et al.* New locus for autosomal dominant mitral valve prolapse on chromosome 13. *Circulation* 2005;**112**:2022–2030.
7. Durst R, Sauls K, Peal DS, deVlaming A, Toomer K, Leyne M, *et al.* Mutations in DCHS1 cause mitral valve prolapse. *Nature* 2015;**525**:109–113.
8. Toomer KA, Yu M, Fulmer D, Guo L, Moore KS, Moore R, *et al.* Primary cilia defects causing mitral valve prolapse. *Sci Transl Med* 2019;**11**:eaax0290.
9. Dina C, Bouatia-Naji N, Tucker N, Delling FN, Toomer K, Durst R, *et al.* Genetic association analyses highlight biological pathways underlying mitral valve prolapse. *Nat Genet* 2015;**47**:1206–1211.
10. Roselli C, Chaffin MD, Weng LC, Aeschbacher S, Ahlberg G, Albert CM, *et al.* Multi-ethnic genome-wide association study for atrial fibrillation. *Nat Genet* 2018;**50**:1225–1233.
11. 3C Study Group. Vascular factors and risk of dementia: design of the Three-City Study and baseline characteristics of the study population. *Neuroepidemiology* 2003;**22**:316–325.
12. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* 2018;**562**:203–209.
13. Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A, *et al.* Next-generation genotype imputation service and methods. *Nat Genet* 2016;**48**:1284–1287.
14. Loh PR, Danecek P, Palamara PF, Fuchsberger C, Reshef YA, Finucane HK, *et al.* Reference-based phasing using the Haplotype Reference Consortium panel. *Nat Genet* 2016;**48**:1443–1448.
15. Kyryachenko S, Georges A, Yu M, Berrandou T, Guo L, Bruneval P, *et al.* Chromatin accessibility of human mitral valves and functional assessment of MVP risk loci. *Circ Res* 2021;**128**:e84–e101.
16. Zhou W, Nielsen JB, Fritsche LG, Dey R, Gabrielsen ME, Wolford BN, *et al.* Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nat Genet* 2018;**50**:1335–1341.
17. Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet* 2007;**39**:906–913.
18. Chen MH, Yang Q. GWAF: an R package for genome-wide association analyses with family data. *Bioinformatics* 2010;**26**:580–581.
19. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 2010;**26**:2190–2191.
20. Yang J, Ferreira T, Morris AP, Medland SE, Genetic Investigation of ANthropometric Traits (GIANT) Consortium, DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium, *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat Genet* 2012;**44**:369–375.
21. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 2015;**4**:7.
22. 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, *et al.* A global reference for human genetic variation. *Nature* 2015;**526**:68–74.
23. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh PR, *et al.* An atlas of genetic correlations across human diseases and traits. *Nat Genet* 2015;**47**:1236–1241.
24. Shah S, Henry A, Roselli C, Lin H, Sveinbjörnsson G, Fatemifar G, *et al.* Genome-wide association and Mendelian randomisation analysis provide insights into the pathogenesis of heart failure. *Nat Commun* 2020;**11:**163.
25. Aragam KG, Chaffin M, Levinson RT, McDermott G, Choi SH, Shoemaker MB, *et al.* Phenotypic refinement of heart failure in a national biobank facilitates genetic discovery. *Circulation* 2019;**139**:489–501.
26. den Hoed M, Eijgelsheim M, Esko T, Brundel BJJM, Peal DS, Evans DM, *et al.* Identification of heart rate-associated loci and their effects on cardiac conduction and rhythm disorders. *Nat Genet* 2013;**45**:621–631.
27. Ntalla I, Weng LC, Cartwright JH, Hall AW, Sveinbjornsson G, Tucker NR, *et al.* Multi-ancestry GWAS of the electrocardiographic PR interval identifies 202 loci underlying cardiac conduction. *Nat Commun* 2020;**11**:2542.
28. Pirruccello JP, Bick A, Wang M, Chaffin M, Friedman S, Yao J, *et al.* Analysis of cardiac magnetic resonance imaging in 36,000 individuals yields genetic insights into dilated cardiomyopathy. *Nat Commun* 2020;**11**:2254.
29. Ge T, Chen CY, Ni Y, Feng YCA, Smoller JW. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nat Commun* 2019;**10**:1776.
30. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, *et al.* pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* 2011;**12**:77.
31. Autism Spectrum Disorders Working Group of The Psychiatric Genomics Consortium. Meta-analysis of GWAS of over 16,000 individuals with autism spectrum disorder highlights a novel locus at 10q24.32 and a significant overlap with schizophrenia. *Mol Autism* 2017;**8**:21.
32. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* 2014;**511**:421–427.
33. Surendran P, Drenos F, Young R, Warren H, Cook JP, Manning AK, *et al.* Trans-ancestry meta-analyses identify rare and common variants associated with blood pressure and hypertension. *Nat Genet* 2016;**48**:1151–1161.
34. Deloukas P, Kanoni S, Willenborg C, Farrall M, Assimes TL, Thompson JR, *et al.* Large-scale association analysis identifies new risk loci for coronary artery disease. *Nat Genet* 2013;**45**:25–33.
35. Barrett JC, Clayton DG, Concannon P, Akolkar B, Cooper JD, Erlich HA, *et al.* Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. *Nat Genet* 2009;**41**:703–707.
36. Rizzo S, Basso C, Lazzarini E, Celeghin R, Paolin A, Gerosa G, *et al.* TGF-beta1 pathway activation and adherens junction molecular pattern in nonsyndromic mitral valve prolapse. *Cardiovasc Pathol* 2015;**24**:359–367.
37. Ng CM, Cheng A, Myers LA, Martinez-Murillo F, Jie C, Bedja D, *et al.* TGF-β–dependent pathogenesis of mitral valve prolapse in a mouse model of Marfan syndrome. *J Clin Invest* 2004;**114**:1586–1592.
38. Herkert JC, Verhagen JMA, Yotti R, Haghighi A, Phelan DG, James PA, *et al.* Expanding the clinical and genetic spectrum of ALPK3 variants: phenotypes identified in pediatric cardiomyopathy patients and adults with heterozygous variants. *Am Heart J* 2020;**225**:108–119.
39. Haji-Seyed-Javadi R, Jelodari-Mamaghani S, Paylakhi SH, Yazdani S, Nilforushan N, Fan JB, *et al.* LTBP2 mutations cause Weill-Marchesani and Weill-Marchesani-like syndrome and affect disruptions in the extracellular matrix. *Hum Mutat* 2012;**33**:1182–1187.
40. Boileau C, Guo DC, Hanna N, Regalado ES, Detaint D, Gong L, *et al.* TGFB2 mutations cause familial thoracic aortic aneurysms and dissections associated with mild systemic features of Marfan syndrome. *Nat Genet* 2012;**44**:916–921.
41. Rath N, Wang Z, Lu MM, Morrisey EE. LMCD1/Dyxin is a novel transcriptional cofactor that restricts GATA6 function by inhibiting DNA binding. *Mol Cell Biol* 2005;**25**:8864–8873.
42. Chang CP, Neilson JR, Bayle JH, Gestwicki JE, Kuo A, Stankunas K, *et al.* A field of myocardial-endocardial NFAT signaling underlies heart valve morphogenesis. *Cell* 2004;**118**:649–663.
43. Bian ZY, Huang H, Jiang H, Shen DF, Yan L, Zhu LH, *et al.* LIM and cysteine-rich domains 1 regulates cardiac hypertrophy by targeting calcineurin/nuclear factor of activated T cells signaling. *Hypertension* 2010;**55**:257–263.
44. Lim JA, Baek HJ, Jang MS, Choi EK, Lee YM, Lee SJ, *et al.* Loss of β2-spectrin prevents cardiomyocyte differentiation and heart development. *Cardiovasc Res* 2014;**101**:39–47.
45. Maron MS, Olivotto I, Harrigan C, Appelbaum E, Gibson CM, Lesser JR, *et al.* Mitral valve abnormalities identified by cardiovascular magnetic resonance represent a primary phenotypic expression of hypertrophic cardiomyopathy. *Circulation* 2011;**124**:40–47.
46. Levine RA, Hagége AA, Judge DP, Padala M, Dal-Bianco JP, Aikawa E, *et al.* Mitral valve disease—morphology and mechanisms. *Nat Rev Cardiol* 2015;**12**:689–710.
47. Ellinor PT, Sasse-Klaassen S, Probst S, Gerull B, Shin JT, Toeppel A, *et al.* A novel locus for dilated cardiomyopathy, diffuse myocardial fibrosis, and sudden death on chromosome 10q25-26. *J Am Coll Cardiol* 2006;**48**:106–111.
48. Brauch KM, Karst ML, Herron KJ, de Andrade M, Pellikka PA, Rodeheffer RJ, *et al.* Mutations in ribonucleic acid binding protein gene cause familial dilated cardiomyopathy. *J Am Coll Cardiol* 2009;**54**:930–941.