

Genome-wide association meta-analysis supports genes involved in valve and cardiac development to associate with mitral valve prolapse

Mengyao Yu¹, PhD, Sergiy Kyryachenko¹, PhD, Stephanie Debette^{2,3}, MD, PhD, Philippe Amouyel,⁴ MD, PhD, Jean-Jacques Schott⁵, PhD Thierry Le Tourneau⁵, MD, PhD, Christian Dina,⁵ PhD, Russell A Norris^{6,7} PhD, Albert A. Hagège⁸, MD, PhD, Xavier Jeunemaitre¹, MD, PhD, Nabila Bouatia-Naji^{1*}, PhD

¹ PARCC, Inserm, Université de Paris, F-75015 Paris, France

² University of Bordeaux, Bordeaux Population Health Research Center, Inserm Center U1219, Bordeaux, France

³ Department of Neurology, Bordeaux University Hospital, Inserm U1219, Bordeaux, France

⁴ Univ. Lille, Inserm, CHU Lille, Institut Pasteur de Lille, U1167 - RID-AGE - Labex DISTALZ - Risk factors and molecular determinants of aging-related disease, F-59000 Lille, France

⁵ Université de Nantes, INSERM, CNRS, CHU Nantes, l'institut du thorax, F-44000 Nantes, France

⁶ Department of Regenerative Medicine and Cell Biology, Medical University of South Carolina, Charleston, SC, USA

⁷ Department of Medicine, Medical University of South Carolina, Charleston, SC, USA

⁸ Assistance Publique – Hôpitaux de Paris, Department of Cardiology, Hôpital Européen Georges Pompidou, 75015, Paris, France

Corresponding author

Nabila BOUATIA-NAJI, PhD

Paris Cardiovascular Research Center, INSERM UMR970, 56 Rue Leblanc, F-75015, Paris
FRANCE.

Tel. +33 1 53 98 79 95

Fax. +33 1 53 98 79 52

Email. nabila.bouatia-naji@inserm.fr

Keywords: heart valve disease, mitral valve prolapse, GWAS; Imputation; Computational
Biology.

Subject Terms: Valvular Heart Disease; Genetics; Association Studies; Meta-analysis;
Computational Biology.

Manuscript word count: 5875

Total number of figures and tables: 6 in the main text and 8 in the supplemental materials

TOC category - basic and population studies

TOC subcategory - Arteriosclerosis, Thrombosis, and Vascular Biology

Abstract

Objective. Mitral valve prolapse (MVP) is a common cardiac valve disease, which affects 1 in 40 in the general population. Previous GWAS have identified six risk loci for MVP. But these loci explained only partially the genetic risk for MVP. We aim to identify additional risk loci for MVP by adding a dataset from the UK Biobank.

Approaches and Results. We re-analyzed 1,007/1,469 cases and 479/862 controls from the MVP-France study and the MVP-Nantes study, respectively. We re-imputed genotypes using HRC and TOPMed, and found this latter to perform better in terms of accuracy in the lower ranges of minor allele frequency (MAF) below 0.1. We then incorporated 434 MVP cases and 4,527 controls from the UKBiobank and conducted a meta-analysis GWAS including ~2000 MVP cases and over 6,800 controls for ~8 million genotyped or imputed common SNPs (MAF>0.01). We replicated the association on chr2 and now provide a finer association map near *TNSI*. We identified three suggestive risk loci, all driven by common variants on Chr1 (*SYT2*), Chr8 (*MSRA*), and Chr19 (*FBXO46*). Gene-based association using MAGMA revealed 15 risk genes for MVP including *GLIS1*, *TGFB2*, *ID2*, *TBX5*, *MSRA*, and *DMPK*. Extensive functional annotation showed that genes associated with MVP are highly expressed in cardiovascular tissues, especially heart, and are involved in cardiac development and potentially aging.

Conclusions. We report an updated meta-analysis GWAS for MVP using dense imputation coverage and an improved case-control sample. We describe several loci and genes with MVP spanning biological mechanisms highly relevant to MVP, especially during valve and heart development.

Abbreviations

MVP:	Mitral Valve Prolapse
HRC:	Haplotype Reference Consortium
TOPMed:	Trans-Omics for Precision Medicine
SNP:	Single Nucleotide Polymorphism
GWAS:	Genome-Wide Association Study
eQTLs:	Expression Quantitative Trait Locus
MAF:	Minor allele frequency
RA:	Risk Allele
OR:	Odds Ratio
CI:	Confidence Interval

Introduction

Heart valves are key functional elements of the heart that display specific biological mechanisms in health and disease. During the heart morphogenesis, the mitral valve develops soon after cardiac looping. The complex shape of the mitral valve, with two leaves, allows a very precise balance of force to maintain unidirectional blood flow through the mitral orifice. The importance of valve development in the origin of mitral valve disease has been established through the discovery of the genetic causes of rare syndromes such as Marfan, Loeys-Dietz and Ehlers-Danlos and familial non-syndromic cases.^{1,2} The causal genes play key roles in extracellular matrix deposition and organization, being influenced by TGF beta and/or ciliogenic signaling nodes.

The adult valve can lose its flexibility under the action of permanent mechanical stress and a degenerative process gradually takes place leading to prolapse (MVP), and in many patients, the incapacity of the valve to close. The resulting mitral regurgitation requires surgery repair or replacement, as it greatly increases the risk of heart failure, arrhythmia and even sudden death.³ MVP is common, affecting approximately 1 in 40 individuals in the general population.³ As for many heart diseases, the establishment of MVP occurs as a result of mild dysfunctions of the many complex biological mechanisms required during development and/or valve function. We have shown through family studies the requirement of *DCHSI*, a member of the cadherin super family, for cell alignment during valve development.¹ More recently, we have shown that loss of primary cilia during development leads to progressive myxomatous degeneration of the mitral valve in mice and humans.² Using GWAS, we have identified predisposition loci,^{1,4} particularly those near *TNSI*, are involved in cell adhesion. This finding further supported the importance of cytoskeleton organization revealed by the study of the polyvalvulopathy syndrome caused by *FLNA* mutations.⁵ However, taken together, these loci explain only a small fraction of the genetic contribution to

MVP. For instance, the relevance of fine regulation of valve and cardiac development mechanisms is established to be at play for early onset syndromic and non-syndromic valve disease. However, this is unknown for late onset and aging-related valve disease. Identification of additional risk loci will likely provide a better and more complete understanding of the genetic and biological basis of MVP.

One of the limitations of the genetic study of MVP is the lack of large cohorts with genome-wide genotyping that would allow increasing the power to discover new predisposition genes of the current GWAS involving only 1,500 patients.¹ In addition, since our initial study, it has become possible to achieve much denser and more accurate GWAS through the high-density genetic imputation panels provided by HRC,⁶ and more recently TopMED⁷ consortia. These panels offer a theoretical increase in power, genomic coverage, and fine mapping, notably through the study of low-frequency variants.^{7, 8}

In this work, we first compare the imputation performance of the newly generated HRC and TOPMed panels in the context of our cohorts. We thus describe comparable results using these two panels, in favor of a better coverage for low-frequency variants for the TOPMed panel. Then, we performed a GWAS meta-analysis including a new case-control study defined using the UKbiobank resource.⁹ We replicate the *TNSI* locus, and described four new suggestive loci. The gene-based association analysis identifies several new loci that inform established and original mechanisms for the biology underlying the genetic risk for MVP.

Materials and Methods

Study populations

We studied 1,920 MVP cases and 6,858 controls in total from three case-control studies from France (MVP-France and MVP-Nantes) and the UK. MVP-France (MVP-F) cases were 1,007 patients recruited through a nation-wide clinical protocol as previously described¹ (mean age 63.2 years, 68% surgery diagnosis). Approvals were obtained from CPP Ile-de-France VI, (approval n° 60---08, June 25th, 2008), the “Commission Informatique et Libertés” (CNIL) (approval n° 908359, October 14th, 2008) and the French Ministry of Health (ID-RCB: 2008-A00568-47) and was registered on the ClinicalTrial.Gov website (protocol ID: 2008–01). Controls were 1,469 participants ascertained from the Three-City Study (3C), a French multisite population-based study investigating the association between dementia, cognitive impairment and vascular diseases¹⁰ (mean age 75 years; 79.04% and 10.13% individuals with hypertension and myocardial infarction, respectively). MVP-Nantes (MVP-N) study was described previously¹ and included 479 French MVP cases (mean age 63 years, 90% surgery diagnosis) and 862 controls (mean age 47 years). The MVP-Nantes study complies with the Declaration of Helsinki, the French guidelines for genetic research and was approved by local ethics committees. Written informed consent was obtained. For the purposes of this study, we ascertained an MVP case-control study from the UK Biobank⁹ (MVP-UKB), which included 434 cases and 4,527 controls. All participants from UK Biobank data were British, Irish and other white backgrounds.⁹ Inclusion criteria for the cases were the main or secondary diagnosis of MVP alone or combined with mitral regurgitation using health records (code = ICD10, I341, I341/I340) and genome-wide genotyping availability. The ethics research committee reference for UK Biobank is 11/NW/0382. We excluded patients with rare syndromes in who MVP usually occurs (e.g., Ehlers-Danlos syndrome and Marfan's syndrome),

and those with ischemic and rheumatic heart diseases. Controls were participants randomly selected among participants without heart disease reported, who were matched to MVP cases for age and sex in a ratio of 1:10.

Genotyping Data

Genome-wide genotypes were generated in MVP-F study using Illumina Human 660W-Quad and in MVP-N study using Affymetrix Axiom Genome Wide CEU-1. The MVP-UKB study genotypes were obtained using UK Biobank Axiom Array. We applied identical quality control (QC) procedures to the three case-control studies as recommended.¹¹ Briefly, individuals with genotyping missing rate $< 97\%$, an excess of ± 3 SD of the heterozygosity rate, and with non-European ancestry identified by plotting the first two genotype-driven principal components (PCs) were excluded from further analyses. SMARTPCA was used to identify ancestral outliers¹² and the background merging data came from HapMap3 (four ethnic populations) and 1000 genomes project (28 ethnic populations). The variants-level QC includes the removal of variants that met SNP missing rate $< 97\%$, MAF < 0.01 , and variants deviating from the Hardy-Weinberg disequilibrium (P-value $< 10^{-6}$).

Haplotype Phasing and Imputation

Imputation was conducted on the Michigan Imputation Server¹³ using minimac3/minimac4.¹⁴ Both HRC r1.1 2016⁶ and TOPMed Freeze5 on GRCh38⁷ were used as reference panels. Haplotypes phasing was performed using ShapeIT¹⁵ in the HRC panel imputation and Eagle for the TOPMed panel. We used bcftools (<https://samtools.github.io/bcftools/bcftools.html>) and vcftools¹⁶ to check for imputation quality. We excluded variants that did not meet the imputation quality defined as $Rsq < 0.3$ and $MAF < 0.01$.

Statistical analyses

We used SNPTEST¹⁷ to perform the association analyses (options -frequentist 1 and -method score), which applies logistic regression under the additive model to test the association with MVP status. The first five PCs were used as covariates during the association analyses. Hardy-Weinberg disequilibrium was checked again for variants in cases and controls and we excluded SNPs with P-value < 10⁻⁶. Meta-analysis was conducted using METAL,¹⁸ with the option that takes P-values, the direction of effect into account as well as the effective sample size, where Weight = 4/(1/N_{cases}+1/N_{controls}).

Genomic risk loci definition and functional annotation

We used FUMA v1.3.4b,¹⁹ a web-based platform providing diverse post-GWAS annotations to define the genomic risk loci and make the functional mapping of associated variants. As recommended, we defined independent significant SNPs by a GWAS P-value < 1×10⁻⁵ and moderate correlation ($r^2 < 0.6$, computed using 1000 genomes Phase 3 data) with any other genome-wide significant SNPs. SNPs in the meta-analysis that were in linkage disequilibrium (LD: $r^2 > 0.6$) with one of those independent significant SNPs were selected for annotation. In FUMA, the lead SNPs and those in LD with them were annotated using eQTLs from GTEX v8 (selected tissue types: artery aorta, artery coronary, heart atrial appendage, heart left ventricle),²⁰ combined annotation dependent depletion (CADD v1.4)²¹ and RegulomeDB (RDB v1.1).²² RNA *in situ* hybridization in mouse embryo for *Tgfb2*, *Id2*, *Tbx5*, *Smg6*, *Srr*, *Abcc3*, *Six5*, *Dmwd* and *Syt2* at E14.5 was obtained through Genepaint (<https://gp3.mpg.de>, accessed in June 2020).^{23, 24}

Regional association plots were created using LocusZoom.²⁵ Positions change between genome assembly GRCh38/hg38 and GRCh37/hg19 was completed using Liftover and Kaviar.²⁶

MAGMA gene-based association and gene-set analysis

The SNPs and P-values from the meta-analysis were used as the inputs for the gene-based association and gene-set analysis using MAGMA v1.07 in FUMA.^{19,27} MAGMA combines the P-values of SNPs that are within window sizes of 50kb upstream and 50kb downstream of the genes and takes LD between SNPs into account. We used 19,151 protein-coding genes from the NCBI 37.3 gene definitions, giving a Bonferroni-corrected P-value threshold of 2.611×10^{-6} .

We also used MAGMA to analyze 9,996 predefined gene sets (GO terms obtained from MsigDB v7.0).²⁸ The competitive analysis was used to test whether the combined effects of genes in a gene set are strongly associated with MVP than the combined effects of all other genes not in the gene set.

Results

Imputation accuracy between HRC and TopMED in MVP GWAS datasets

We first aimed to assess the usefulness of using larger imputation reference panels in the two French MVP case controls studies. MVP-France study (MVP-F) included 1,007 cases and 1,469 controls with 492,438 genotyped variants. MVP-Nantes study (MVP-N) included 479 cases and 862 controls with 370,697 genotyped variants. We generated 229,973,672 and 230,053,813 imputed variants in MVP-F and MVP-N, respectively, from the TOPMed reference panel, which represent 5.8-fold more variants compared to HRC (39,117,105 in MVP-F and MVP-N) (Figure 1a). However, most of these variants had an imputation score R_{sq} smaller than 0.3: 84% (MVP-F-TOPMed), 88% (MVP-N-TOPMed), 51% (MVP-F-HRC), and 60% (MVP-N-HRC) (Figure 1a). Compared to MVP-N, the relatively larger sample size of MVP-F study generates a slightly larger number of well-imputed variants (Figure 1a).

We then classified MAF into five regions ((0,0.01], (0.01,0.05], (0.05,0.1], (0.1,0.2] and (0.2,0.5]). Most of the variants with $R_{sq} < 0.3$ are rare variants which minor allele frequency (MAF) is smaller than 0.01, as illustrated for all SNPs generated on chromosome 22 (Supplementary figure I). We also found that TOPMed panel imputed variants present, on average, a better accuracy for low frequency ranges (0,0.01], (0.01,0.05] and (0.05,0.1], while relatively stable accuracy at MAF ranges (0.1,0.2] and (0.2,0.5] compared to HRC panel in both studies (Figure 1b). Taken together, we use TOPMed as the imputation panel to allow analyzing more well imputed variants, especially in the low frequency ($0.01 < \text{MAF} < 0.1$) category.

Genetic association analyses in French and UKBiobank case control studies

We meta-analyzed three GWAS involving a total of 1,920 MVP cases and 6,858 controls and ~8 million (8,021,974) genotyped or imputed common SNPs ($\text{MAF} > 0.01$). We confirmed a deviation from the expected levels of significance in this updated meta-analysis (Figure 2a).

The strongest association signal was observed on chromosome 2 at the *TNSI* locus, where we report 46 variants with a genome-wide significant P-value $< 5 \times 10^{-8}$ (Figure 2b). In addition to successfully replicating the previously reported top associated SNP rs12465515 (P-value = 2.8×10^{-10} ; OR = 1.30[1.19-1.42]), we found a new top associated variant on chromosome 2 (lead SNP: rs7595393; P-value = 1.71×10^{-10} ; OR = 1.31[1.20-1.42], Table 1, Supplementary figure II). Both SNPs are highly correlated ($r^2=95$, 1000G Phase 3: CEU). As expected, conditional analyses both on rs12465515 and rs7595393 at the *TNSI* locus resulted in disappearance of the association signal (Figure 3). Functional annotation of 103 SNPs in LD ($r^2 > 0.6$) with 4 independent significant SNPs (ISS) at this locus indicated the existence of 5 SNPs predicted to be deleterious alleles (CADD score > 12) and 7 SNPs likely to lie within regulatory elements according to the RDB score (RDB score $> 3a$) (Supplementary Table I).

We also report three suggestive and original association signals (Figure 2, Table 1). On chromosome 1, the lead SNP (rs199723025, effect allele frequent (EAF) = 0.05, P-value = 5.53×10^{-8} , OR = 1.69[1.40-2.04]) is an intergenic deletion variant of Synaptotagmin 2 gene (*SYT2*) (Table 1, Supplementary figure III). We found that the lead variant is a significant eQTL in atrial appendage and artery aorta (P-value: 2.4×10^{-5} and 1×10^{-8}) for the lysine demethylase 5B gene (*KDM5B*) involved in DNA stability and repair. The second locus on chromosome 8 (rs56028519, EAF = 0.74, P-value = 7.97×10^{-8} , OR = 1.29[1.17-1.41]) is an intronic variant of the methionine sulfoxide reductase A gene (*MSRA*) (Table 1, Supplementary figure IV). The lead SNP rs56028519 is an eQTL of a lncRNA gene (AF131215.2) in heart atrial appendage (P-value= 4.5×10^{-7}) and a lncRNA gene (RP11-981G7.6) in left ventricle (P-value= 1.3×10^{-7}). We found that several SNPs in LD ($r^2 > 0.6$) with the 3 ISS are located near the regulatory elements, especially rs11249991 and rs11781529 that showed high potential to be regulatory (RDB score $> 3a$). Three SNPs in total are likely to be deleterious, including rs11783281 (CADD score = 19.05, Supplementary Table I). The third suggestive association signal was located in a particularly gene-rich region on chromosome 19 (rs4802272, EAF = 0.55, P-value = 3.17×10^{-7} , OR = 1.24[1.14-1.34]) (Table 1, Supplementary figure V). Although the lead SNPs mapped to *FBXO46*, the association signal spans five genes in total, including, *SIX5*, *FOXA3*, *RSPH6A*, *DMPK* and *DMWD* (Supplementary figure 5). The lead SNP is an eQTLs of *DMPK* and *DMWD* in artery aorta (P-value: 2.1×10^{-9} , 3.4×10^{-11}). Two SNPs in high LD with the lead SNP showed high deleteriousness scores (rs62111759, CADD score = 12.93 and rs672348, CADD score = 12.43) (Supplementary Table I).

Genome-wide gene-based association and pathway analyses

We performed a genome-wide gene-based association analysis using MAGMA²⁷ to estimate the gene level association on the basis of all SNPs in a gene. This method is a complementary approach

to single SNP GWAS analyses and usually reveals genes with locally consistent associated SNPs that individually may do not reach genome-wide significance. In total, 15 genes reached genome-wide significant association (MAGMA P-value $< 2.611 \times 10^{-6}$) with MVP (Figure 4a, 4b). We highlight *GLIS1* (P= 2.50×10^{-6}) that we have previously identified to associate with MVP using single SNP and pathway analyses,⁴ *TGFB2* (P= 1.74×10^{-6}), a high-profile candidate gene for myxomatous valve disease, *MSRA* (P= 5.18×10^{-7}), one of the suggestive loci in the SNP GWAS, and *TBX5* (P = 1.10×10^{-6}), a key regulator of heart development (Table 2). The ten remaining genes mapped into two gene-rich loci on chromosome 17 and 19. On chromosome 17, five genes reached gene-wide significance including *SMG6* (P-value = 1.06×10^{-7}), *SRR* (P-value = 6.17×10^{-9}), in addition to *TSRI* (P-value = 1.04×10^{-8}), *SGSM2* (P-value = 3.98×10^{-8}) and *ABCC3* (P-value = 1.10×10^{-6}) (Table 2). On chromosome 19, we report *FBXO46* (P-value = 2.05×10^{-6}), *SIX5* (P-value = 9.25×10^{-9}), *DMPK* (P-value = 6.18×10^{-8}), *DMWD* (P-value = 9.86×10^{-8}) and *RSPH6A* (P-value = 6.42×10^{-7}) (Table 2). Tissue expression analyses using the GTEx data resource showed that four of genes (*DMPK*, *DMWD*, *TBX5* and *ID2*) are strongly expressed in cardiovascular tissues, especially in the heart atrial appendage and heart left ventricle (Figure 4c). We note that low frequency variant contributed little to the association of the genes above mentioned, where the associations were driven by common variants (Supplementary Figure VI). Section *in situ* hybridization validated developmental expression of many of these gene within the mitral valves (Supplementary Figure VII) further supporting their potential involvement in disease phenotype.

MAGMA pathway analysis using 9,996 gene sets (GO terms obtained from MsigDB) indicated that several significantly enriched gene sets for MVP associated genes are related to cardiac biology. We highlight cardiac ventricle formation (P-value = 3.09×10^{-6}), cardiac chamber formation (P-value = 8.94×10^{-6}), cardiac right ventricle morphogenesis (P-value = 1.11×10^{-4}),

cardioblast differentiation (P-value = 1.24×10^{-4}), apoptotic process involved in heart morphogenesis (P-value = 1.81×10^{-4}), atrial septum morphogenesis (P-value = 3.08×10^{-4}), cardiac atrium development (P-value = 3.12×10^{-4}) and a GO term related to ciliary rootlet (P-value = 4.05×10^{-4}) (Supplementary table II).

Discussion

Here we describe a high genetic coverage meta-analysis of GWAS based on TOPMed imputation involving ~8 million common variants in ~2000 MVP patients and ~6800 controls. In addition to replicate the association at the *TNSI* locus, we identified several associated variants and genes involving in established and original mechanisms for the biology underlying the genetic risk for MVP.

Our results using the TOPMed imputation panel provided higher resolution association map for the risk of MVP in a reasonably well powered dataset. We provide confirmatory results and fine mapping at the *TNSI* where we now report rs7595393 as the new lead associated SNP. The association signal on Chr2 is located in a gene-desert genomic region. We have previously provided solid biological evidence, which included a myxomatous valve phenotype in the heterozygote knockout mouse supporting the gene encoding Tensin 1, a focal adhesion protein, to be causal.¹ Functional annotations at this locus describe several SNPs belonging to the association block of rs7595393 as plausible causal variants. In the absence of eQTL, functional annotation for open chromatin and enhancer marks specifically in the mitral valve, we acknowledge that *in silico* annotation has limited ability to point at the causal variants and further experiments are needed to achieve this goal. One important addition in the current study is the improved coverage of low frequency variants ($0.01 < \text{MAF} < 0.10$). Our current data do not allow us to firmly conclude about the putative role of low frequency variants in MVP genetic risk, which needs to be explored further using larger datasets. However, the fact that the single-SNP and gene-based association signals we report were overwhelmingly driven by common variations does not support this hypothesis.

In addition to confirming the association on Chr2, we describe three unprecedented suggestive association signals in *SYT2* (Chr1), *MSRA* (Chr8) and *FBXO46* (Chr19) that will need to be replicated in larger studies. Two of these association signals (*MSRA* and *FBXO46* loci) were further supported by the gene-based association results. We also report gene association for *GLIS1* and several genes at the *SMG6/SRR* locus where we have previously described single SNP associations with MVP.^{1, 4}

Several newly associated genes with MVP are involved in established biological mechanisms and are highly relevant to mitral valve disease. This applied to the TGF-beta family member *TGFB2*, an essential growth factor for myocardial cells endothelial to mesenchymal transition and valve elongation during valve development.²⁹ It has been established that the myxomatous degeneration is the result of the response of endothelial cells to mechanical stress, aging, or excessive stimulation of TGF beta proteins.²⁹ Both in the Marfan mouse model³⁰ and human mitral valves,³¹ TGF beta induces pro-fibrotic function and promotes the proliferation and differentiation of endothelial cells into myofibroblasts.

Our gene-based association results strengthen the pivotal role that may play regulatory genes involved in cardiac development in susceptibility to MVP. This is supported by the association of inhibitor of DNA binding 2 gene (*ID2*) involved in cell differentiation and proliferation, which action is regulated by TGF-beta1 and BMP-7.³² Interestingly, *Id2* expression is lost in valve forming regions of Smad4-deficient endocardium in mice.³³ Moreover, *Id2* was shown to belong to the same molecular pathway that includes *Tbx5*, that coordinates ventricular conduction system lineage in mice.³⁴ *TBX5* association with genetic risk for MVP is intriguing. *TBX5* is a member of the T-box transcription factor and key regulator of cardiac development, specifically involved in cardiac conduction and maintenance of mature cardiomyocyte.³⁵ Mutations in *TBX5* cause the

Holt-Oram syndrome (OMIM #142900), a disorder characterized by abnormalities in upper limbs and congenital heart lesion. *TBX5* is also an established risk locus identified from GWAS for traits related to cardiac conduction, mainly atrial fibrillation,³⁶ QRS duration³⁷ and PR interval.³⁸ *TBX5* expression is not detected in human heart valves,³⁹ which may suggest the valve prolapse phenotype to result from defaults in the interplay between myocardium, conduction tissue and mitral valve apparatus during heart development or later as a consequence of valve aging.

Our genetic study also describes original mechanisms that may deserve future biological investigation. *MSRA* encodes methionine sulfoxide reductase A, a ubiquitously expressed and conserved enzyme, including in the heart, with the highest levels in the kidney and the nervous system. Several GWAS association signals near *MSRA* were reported for blood pressure,⁴⁰ neuroticism,⁴¹ and glomerular filtration.⁴² The biological implication of *MSRA* in the degenerative process of the valve is unclear, and could be through this enzyme protective role against oxidative stress during aging.⁴³

Our study presents several limitations. The inclusion of a new dataset from the UKBiobank and the generation of a dense association map did not compensate the limited power of our study and the original suggestive loci described will need to be replicated in future studies. The gene-based results are not able to detect association signals involving gene-desert genomic regions with long-range enhancers, as the one we observe on Chr2 upstream *TNSI*. Another limitation of this method is that gene-rich genomic regions provides redundant association signals involving the same sets of variants and do not allow to point specifically at a potential causal gene. Functional annotation is based on existing eQTLs and gene expression pattern in cardiovascular tissues, especially heart atrial appendage and left ventricle, where cell composition and gene expression may differ from gene expression in the mitral valve.

Conclusions

To summarize, we report an updated meta-analysis GWAS for MVP using dense imputation coverage and an increased case control sample. We describe several established and original associated loci and genes with MVP spanning biological mechanisms highly relevant to heart valve disease. Follow-up biological studies in cell and animal models are needed to better understand their direct effect on the valve degenerative process.

Acknowledgments

We acknowledge the contribution of the Leducq Foundation, Paris for supporting the genetic study in the French case control studies. We thank Carolina Roselli and Patrick Ellinor for providing access to TopMED panel data. This research has been conducted using the UK Biobank Resource under Application Number 32360.

Sources of funding

This study was supported by a Ph.D. scholarship from the China Scholarship Council to MY, and French Agency of Research (ANR-16-CE17-0015-02). AG, SK and NB-N are supported by a European Research Council grant (ERC-Stg-ROSALIND-716628). The recruitment of the MVP France cohort was supported by the French Society of Cardiology (SFC). The recruitment of the MVP Nantes cohort was supported by Fédération Française de Cardiologie, Fondation Coeur et Recherche, French Ministry of Health “PHRC-I 2012,” and INSERM Translational Research Grant. The genotyping of the controls from the Three-City Study (3C) was supported by the non-profit organization Fondation Alzheimer (Paris, France). This work was supported in part by grants

from the National Institutes of Health (GM103444 to RAN; R01HL131546, R01HL149696, P20GM103444, and R01HL127692 to RAN; and American Heart Association (19TPA34850095 to RAN, 17CSA33590067 to RAN).

Disclosures

None

REFERENCES

1. Dina C, Bouatia-Naji Ni, Tucker N, et. al. Genetic association analyses highlight biological pathways underlying mitral valve prolapse. *Nat Genet.* 2015;47:1206-1211
2. Toomer KA, Yu M, Fulmer D, et. al. Primary cilia defects causing mitral valve prolapse. *Sci Transl Med.* 2019;11
3. Freed LA, Levy D, Levine RA, Larson MG, Evans JC, Fuller DL, Lehman B, Benjamin EJ. Prevalence and clinical outcome of mitral-valve prolapse. *New England Journal of Medicine.* 1999;341:1-7
4. Yu M, Georges A, Tucker NR, et. al. Genome-wide association study-driven gene-set analyses, genetic, and functional follow-up suggest *glis1* as a susceptibility gene for mitral valve prolapse. *Circ Genom Precis Med.* 2019;12:e002497
5. Le Tourneau T, Le Scouarnec S, Cueff C, et. al. New insights into mitral valve dystrophy: A filamin-a genotype-phenotype and outcome study. *Eur Heart J.* 2018;39:1269-1277
6. McCarthy S, Das S, Kretzschmar W, et. al. A reference panel of 64,976 haplotypes for genotype imputation. *Nature genetics.* 2016;48:1279-1283
7. Taliun D, Harris DN, Kessler MD, Carlson J, Szpiech ZA, Torres R, Taliun SAG, Corvelo Ae, Gogarten SM, Kang HM, others. Sequencing of 53,831 diverse genomes from the nhlbi topmed program. *BioRxiv.* 2019:563866
8. Das S, Abecasis GR, Browning BL. Genotype imputation from large reference panels. *Annual review of genomics and human genetics.* 2018;19:73-96
9. Sudlow C, Gallacher J, Allen N, et. al. Uk biobank: An open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* 2015;12:e1001779
10. Group CS. Vascular factors and risk of dementia: Design of the three-city study and baseline characteristics of the study population. *Neuroepidemiology.* 2003;22:316-325
11. Anderson CA, Pettersson FH, Clarke GM, Cardon LR, Morris AP, Zondervan KT. Data quality control in genetic case-control association studies. *Nat Protoc.* 2010;5:1564-1573

12. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet.* 2006;38:904-909
13. Das S, Forer L, Schumacher S, Sebastian, Sidore C, Locke AE, Kwong A, Vrieze SI, Chew EY, Levy S, McGue M, others. Next-generation genotype imputation service and methods. *Nature Genetics.* 2016;48:1284-1287
14. Fuchsberger C, Abecasis GR, Hinds DA. Minimac2: Faster genotype imputation. *Bioinformatics (Oxford, England).* 2015;31:782-784
15. Delaneau O, Marchini J, Zagury J-F, et al. A linear complexity phasing method for thousands of genomes. *Nat Methods.* 2012;9:179-181
16. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R, Genomes Project Analysis G. The variant call format and vcfutils. *Bioinformatics.* 2011;27:2156-2158
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. Willer CJ, Li Y, Abecasis GR, et al. Metal: Fast and efficient meta-analysis of genomewide association scans. *Bioinformatics.* 2010;26:2190-2191
19. Watanabe K, Taskesen E, Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with fUMA. *Nature communications.* 2017;8:1826
20. Consortium GT, Laboratory DA, Coordinating Center -Analysis Working G, et al. Genetic effects on gene expression across human tissues. *Nature.* 2017;550:204-213
21. Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M. CADD: Predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res.* 2019;47:D886-D894
22. Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, Karczewski KJ, Park J, Hitz BC, Weng S, Cherry JM, Snyder M. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res.* 2012;22:1790-1797
23. Visel A, Thaller C, Eichele G. Genepaint.org: An atlas of gene expression patterns in the mouse embryo. *Nucleic Acids Res.* 2004;32:D552-556
24. Diez-Roux G, Banfi S, Sultan M, et al. A high-resolution anatomical atlas of the transcriptome in the mouse embryo. *PLoS Biol.* 2011;9:e1000582
25. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, Boehnke M, Abecasis G, et al., Willer CJ. LocusZoom: Regional visualization of genome-wide association scan results. *Bioinformatics.* 2010;26:2336-2337
26. Glusman G, Caballero J, Mauldin DE, Hood L, Roach JC. Kaviar: An accessible system for testing SNP novelty. *Bioinformatics.* 2011;27:3216-3217
27. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. Magma: Generalized gene-set analysis of GWAS data. *PLoS Comput Biol.* 2015;11:e1004219
28. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdottir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. *Bioinformatics.* 2011;27:1739-1740

29. Levine RA, Hagège AA, Judge DP, et. al. Mitral valve disease--morphology and mechanisms. *Nature reviews. Cardiology*. 2015;12:689-710
30. Ng CM, Cheng A, Myers LA, Martinez-Murillo F, Jie C, Bedja D, Gabrielson KL, Hausladen JM, Mecham RP, Judge DP, Dietz HC. Tgf-beta-dependent pathogenesis of mitral valve prolapse in a mouse model of marfan syndrome. *J Clin Invest*. 2004;114:1586-1592
31. Geirsson A, Singh M, Ali R, Abbas H, Li W, Sanchez JA, Hashim S, Tellides G. Modulation of transforming growth factor-beta signaling and extracellular matrix production in myxomatous mitral valves by angiotensin ii receptor blockers. *Circulation*. 2012;126:S189-197
32. Kowantetz M, Valcourt U, Bergstrom R, Heldin CH, Moustakas A. Id2 and id3 define the potency of cell proliferation and differentiation responses to transforming growth factor beta and bone morphogenetic protein. *Mol Cell Biol*. 2004;24:4241-4254
33. Moskowitz IP, Wang J, Peterson MA, Pu WT, Mackinnon AC, Oxburgh L, Chu GC, Sarkar M, Berul C, Smoot L, Robertson EJ, Schwartz R, Seidman JG, Seidman CE. Transcription factor genes *smad4* and *gata4* cooperatively regulate cardiac valve development. [corrected]. *Proc Natl Acad Sci U S A*. 2011;108:4006-4011
34. Moskowitz IP, Kim JB, Moore ML, Wolf CM, Peterson MA, Shendure J, Nobrega MA, Yokota Y, Berul C, Izumo S, Seidman JG, Seidman CE. A molecular pathway including *id2*, *tbx5*, and *nkx2-5* required for cardiac conduction system development. *Cell*. 2007;129:1365-1376
35. Steimle JD, Moskowitz IP. *Tbx5*: A key regulator of heart development. *Curr Top Dev Biol*. 2017;122:195-221
36. Nielsen JB, Thorolfsdottir RB, Fritsche LG, et. al. Biobank-driven genomic discovery yields new insight into atrial fibrillation biology. *Nat Genet*. 2018;50:1234-1239
37. Prins BP, Mead TJ, Brody JA, et. al. Exome-chip meta-analysis identifies novel loci associated with cardiac conduction, including *adamts6*. *Genome Biol*. 2018;19:87
38. Butler AM, Yin X, Evans DS, et. al. Novel loci associated with pr interval in a genome-wide association study of 10 african american cohorts. *Circ Cardiovasc Genet*. 2012;5:639-646
39. Hatcher CJ, Goldstein MM, Mah CS, Delia CS, Basson CT. Identification and localization of *tbx5* transcription factor during human cardiac morphogenesis. *Dev Dyn*. 2000;219:90-95
40. Giri A, Hellwege JN, Keaton JM, et. al. Trans-ethnic association study of blood pressure determinants in over 750,000 individuals. *Nat Genet*. 2019;51:51-62
41. Luciano M, Hagenaars SP, Davies G, et. al. Association analysis in over 329,000 individuals identifies 116 independent variants influencing neuroticism. *Nat Genet*. 2018;50:6-11
42. Wuttke M, Li Y, Li M, et. al. A catalog of genetic loci associated with kidney function from analyses of a million individuals. *Nat Genet*. 2019;51:957-972

43. Cabreiro F, Picot CR, Friguet B, Petropoulos I. Methionine sulfoxide reductases: Relevance to aging and protection against oxidative stress. *Ann N Y Acad Sci.* 2006;1067:37-44

Highlights

- Provide high coverage meta-analysis of GWAS based on TOPMed imputation involving ~8 million common variants in ~2000 MVP patients and ~6800 controls.
- Low frequency variant contributed little to the association of the genes mentioned, where the associations were driven by common variants
- Several association loci involve genes related to cardiac development and potentially aging.

Table 1. Associations of top SNPs with MVP obtained in the GWAS meta-analysis.

RSID	CHR	PosB38	RA*	Freq	MVP-Paris	MVP-Nantes	MVP-UKB	ALL
					P-value	P-value	P-value	P-value
					OR † [95%CI ‡]	OR [95%CI]	OR [95%CI]	OR [95%CI]
rs199723025	1	202712600	A	0.05	1.40×10^{-4}	0.31	1.92×10^{-5}	5.53×10^{-8}
					1.61 [1.25-2.08]	1.26[0.88-1.83]	1.72[1.34-2.22]	1.69[1.40-2.04]
rs7595393	2	217006531	G	0.38	1.40×10^{-6}	9.69×10^{-3}	7.40×10^{-4}	1.71×10^{-10}
					1.34 [1.19-1.52]	1.44 [1.22-1.70]	1.27 [1.10-1.48]	1.31 [1.20-1.42]
rs56028519	8	10341263	A	0.74	2.94×10^{-3}	1.21×10^{-2}	1.11×10^{-4}	7.97×10^{-8}
					1.19 [1.04-1.35]	1.31 [1.10-1.57]	1.34 [1.16-1.56]	1.29 [1.17-1.41]
rs4802272	19	45732692	A	0.55	2.10×10^{-3}	3.39×10^{-4}	1.97×10^{-2}	3.17×10^{-7}
					1.18 [1.06-1.33]	1.25 [1.07-1.47]	1.18 [1.03-1.36]	1.24 [1.14-1.34]

*RA: risk allele; †OR: odds ratio; CI: confidence interval.

Table 2. Genes significantly associated with MVP in gene-based association tests (GWGAS) using MAGMA.

CHR	START	STOP	NSNPS*	ZSTAT	P-value	SYMBOL
1	53921910	54249877	827	4.5652	2.50×10^{-6}	<i>GLIS1</i>
1	218469577	218667961	420	4.6398	1.74×10^{-6}	<i>TGFB2</i>
2	8768975	8874583	332	4.8734	5.49×10^{-7}	<i>ID2</i>
8	9861778	10336401	2023	4.8846	5.18×10^{-7}	<i>MSRA</i>
12	114741736	114896247	527	4.7335	1.10×10^{-6}	<i>TBX5</i>
17	1913133	2257065	895	5.1892	1.06×10^{-7}	<i>SMG6</i>
17	2156677	2278554	358	5.6949	6.17×10^{-9}	<i>SRR</i>
17	2175797	2290801	330	5.6058	1.04×10^{-8}	<i>TSR1</i>
17	2190792	2334352	391	5.368	3.98×10^{-8}	<i>SGSM2</i>
17	48662138	48819613	358	4.7339	1.10×10^{-6}	<i>ABCC3</i>
19	46163887	46284162	227	4.606	2.05×10^{-6}	<i>FBXO46</i>
19	46218043	46322484	182	5.6254	9.25×10^{-9}	<i>SIX5</i>
19	46222975	46335810	235	5.2882	6.18×10^{-8}	<i>DMPK</i>
19	46236205	46346060	265	5.2019	9.86×10^{-8}	<i>DMWD</i>
19	46248968	46368577	332	4.8423	6.42×10^{-7}	<i>RSPH6A</i>

Note: GWGAS tests were performed in MAGMA and represent the aggregate effect of all SNPs in a gene (defined by GRCh37 reference genome locations). The Bonferroni threshold of significance for 19,151 genes was 2.611×10^{-6} . The position in the table is B37. *NSNPS: the number of SNPs annotated to that gene that were found in the data and were not excluded based on internal SNP QC.

Figures

Figure 1. Comparison of the imputation quality of the results for MVP-F and MVP-N using HRC and TOPMed as reference panel, respectively. (a). The numbers of variants generated using different imputation accuracy thresholds; (b). The imputation quality at different MAF region using all the cleaned SNPs ($R_{sq} > 0.3$) in cohorts MVP-Paris and MVP-Nantes. R_{sq} : imputation accuracy; MVP-F: MVP-France cases-control study. MVP-N: MVP-Nantes case-control study.

Figure 2. SNP-based association results with MVP in a GWAS meta-analysis involving MVP-F, MVP-N and MVP-UKB case-control studies ($\lambda_{GC} = 1.08$). (a) The Q-Q plot represents the expected (x-axis) versus the observed (y-axis) P-values; (b) Manhattan plot summarizes the $-\log_{10}(P)$ of each SNP by chromosome obtained from the GWAS meta-analysis. The blue line indicates the threshold for suggestive association ($P\text{-value} < 1 \times 10^{-6}$) and the red line indicates the genome-wide significance threshold ($P < 5 \times 10^{-8}$). Both plots represent the association of ~ 8 million SNPs genotyped and imputed using TOPMed panel.

Figure 3. Conditional Analysis of the TNS1 locus. Locus Zoom plots of the TNS1 locus demonstrate conditional analysis of Dina reported lead SNP rs12465515, the top SNP rs7595393 using GCTA-COJO. Linkage disequilibrium of SNPs with the conditioned SNPs is based on data from MVP-F cohort and is shown by the color of the points. The best signals in each plot are marked as a purple rhombus. (a) Association of SNP with MVP; (b) MVP association conditioned

on Dina reported lead SNP rs12465515 (rs12465515 added as an additional covariate); (c) MVP association conditioned on the top SNP rs7595393

Figure 4. Genome-wide gene-based association analysis and the heatmap of expression of the highlighted genes. Genome-wide gene-based analysis of 19,151 genes that were tested for association with MVP using MAGMA. (a) Q-Q plot showing the expected (x-axis) versus the observed (y-axis) P-values. (b) Manhattan plot summarizing the $-\log_{10}(P)$ of each gene ordered by chromosome. The red line indicates the Bonferroni corrected threshold for genome-wide significance ($P = 0.05/19151 = 2.611 \times 10^{-5}$). The top 15 most significant genes are labeled (Clone-based (Vega) gene and RNA Genes are not labeled). (c) The heatmap of tissue expression of the 18 highlighted genes using the GTEx v8 54 tissue types data resource.

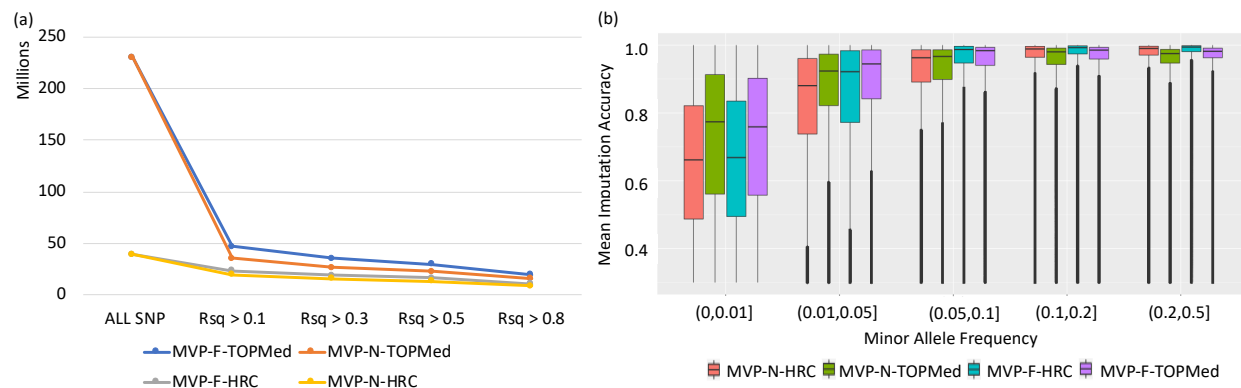


Figure 1. Comparison of the imputation quality of the results for MVP-F and MVP-N using HRC and TOPMed as reference panel, respectively. (a). The numbers of variants generated using different imputation accuracy thresholds; (b). The imputation quality at different MAF region using all the cleaned SNPs ($R_{sq} > 0.3$) in cohorts MVP-Paris and MVP-Nantes. R_{sq} : imputation accuracy; MVP-F: MVP-France cases-control study. MVP-N: MVP-Nantes case-control study.

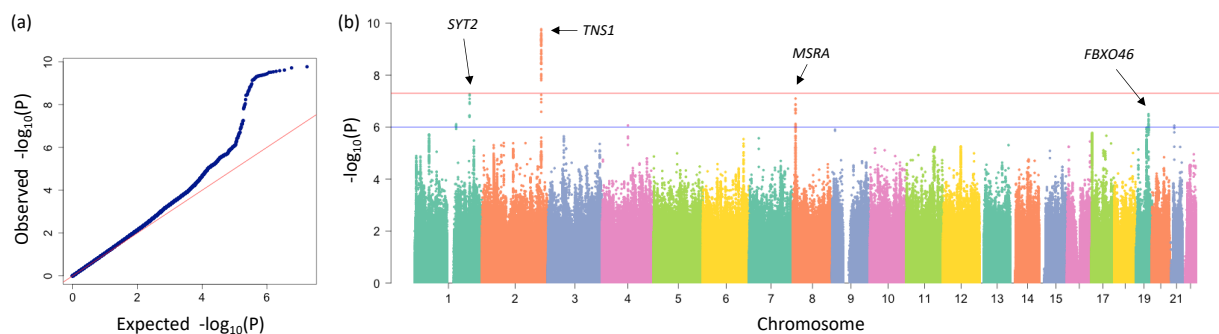


Figure 2. SNP-based association results with MVP in a GWAS meta-analysis involving MVP-F, MVP-N and MVP-UKB case-control studies ($\lambda_{GC} = 1.08$). (a) The Q-Q plot represents the expected (x-axis) versus the observed (y-axis) P-values; (b) Manhattan plot summarizes the $-\log_{10}(P)$ of each SNP by chromosome obtained from the GWAS meta-analysis. The blue line indicates the threshold for suggestive association ($P\text{-value} < 1 \times 10^{-6}$) and the red line indicates the genome-wide significance threshold ($P < 5 \times 10^{-8}$). Both plots represent the association of ~ 8 million SNPs genotyped and imputed using TOPMed panel.

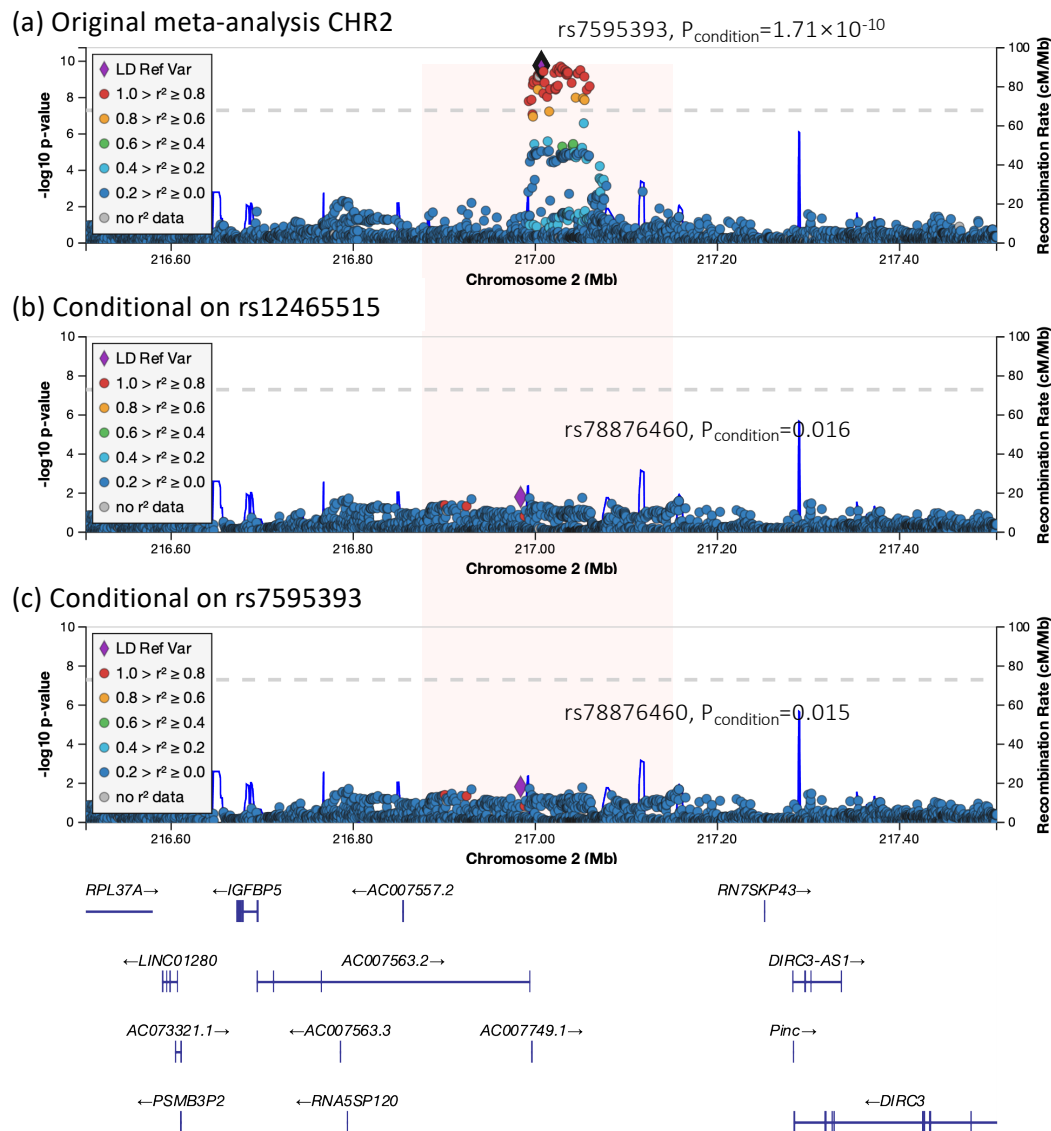


Figure 3. Conditional Analysis of the *TNSI* locus. Locus Zoom plots of the *TNSI* locus demonstrate conditional analysis of Dina reported lead SNP rs12465515, the top SNP rs7595393 using GCTA-COJO. Linkage disequilibrium of SNPs with the conditioned SNPs is based on data from MVP-F cohort and is shown by the color of the points. The best signals in each plot are marked as a purple rhombus. (a) Association of SNP with MVP; (b) MVP association conditioned on Dina reported lead SNP rs12465515 (rs12465515 added as an additional covariate); (c) MVP association conditioned on the top SNP rs7595393

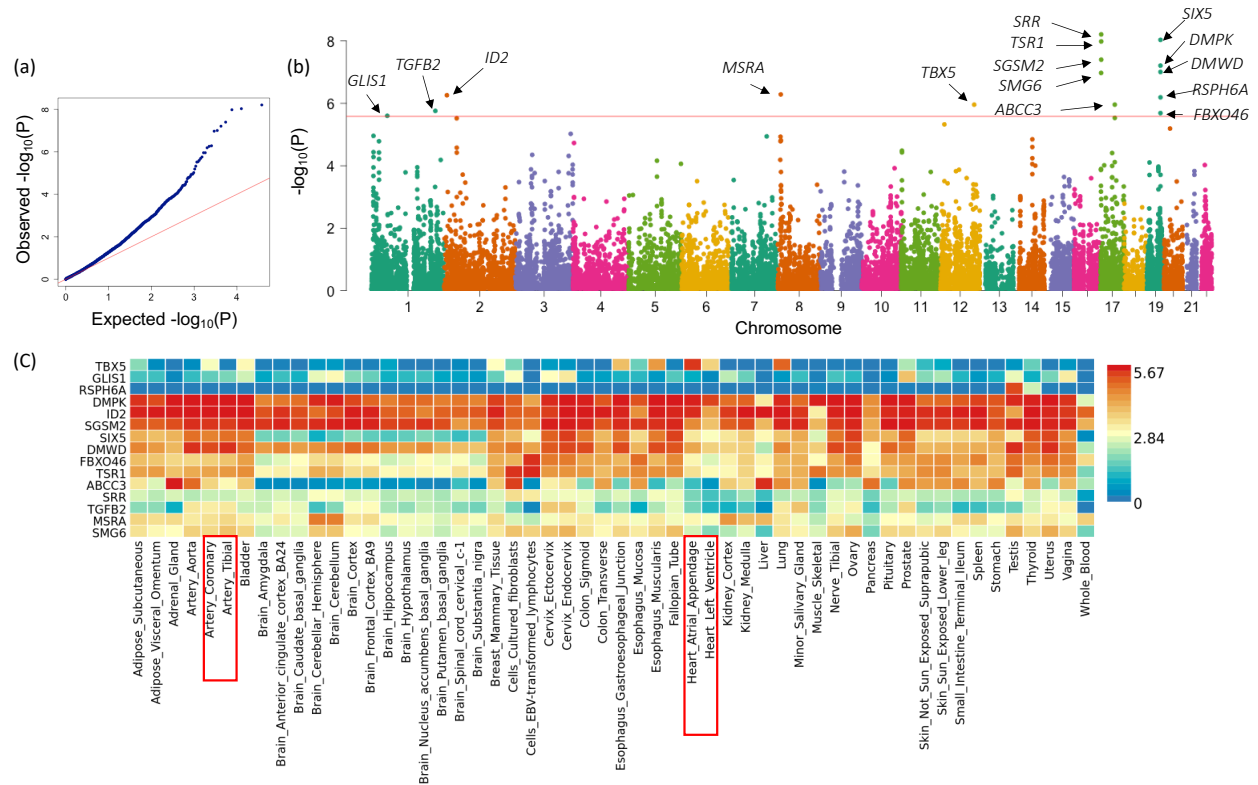


Figure 4. Genome-wide gene-based association analysis and the heatmap of expression of the highlighted genes. Genome-wide gene-based analysis of 19,151 genes that were tested for association with MVP using MAGMA. (a) Q-Q plot showing the expected (x-axis) versus the observed (y-axis) P-values. (b) Manhattan plot summarizing the $-\log_{10}(P)$ of each gene ordered by chromosome. The red line indicates the Bonferroni corrected threshold for genome-wide significance ($P = 0.05/19151 = 2.611 \times 10^{-5}$). The top 15 most significant genes are labeled (Clone-based (Vega) gene and RNA Genes are not labeled). (c) The heatmap of tissue expression of the 18 highlighted genes using the GTEx v8 54 tissue types data resource.