Characterization of genomic regulation profiles in human mitral valve whole tissue to annotate genetic risk loci for mitral valve prolapse

Sergiy Kyryachenko[1], PhD[*], Adrien Georges[1], PhD[*], Mengyao Yu[1], PhD[*], Takiy Barrandou[1], PhD, Patrick Bruneval[1], MD, PhD, Tony Rubio[2,3], MSc, Judith Gronwald[2,3], Hassina Baraki[3,4], MD, Ingo Kutschka[3,4], MD, Russell. A Norris[5,6], PhD, Niels Voigt[2,3], MD, Nabila Bouatia-Naji[1], PhD

1. Université de Paris, PARCC, Inserm, Paris, France
2. Institute of Pharmacology and Toxicology, University Medical Center Göttingen, Germany
3. DZHK (German Center for Cardiovascular Research), Partner Site Göttingen, Germany
4. Department of Thoracic and Cardiovascular Surgery, University Medical Center, Göttingen, Germany
5. Department of Regenerative Medicine and Cell Biology, Medical University of South Carolina, Charleston, SC, USA;
6. Department of Medicine, Medical University of South Carolina, Charleston, SC, USA

[*] Drs Kyryachenko, Georges, and Yu contributed equally to this work as joint first authors.

Short title: Regulatory profiles of mitral valves

Corresponding author: Nabila Bouatia-Naji, email: nabila.bouatia-naji@inserm.fr, address: Paris Cardiovascular Research Center Inserm U970, 56 rue Leblanc, F-75015 Paris, France

Total word count: 8,141

Subject terms: Valvular Heart Disease, Functional Genomics, Genetics, Etiology

ABSTRACT

Rationale: Mitral valve prolapse (MVP) is a common valve disease that leads to mitral insufficiency, heart failure and sudden death. The identification of risk loci provided insight into its genetic architecture, although the causal variants and target genes need to be fully characterized.

Objective: To establish the chromatin accessibility profiles and gene regulation specificities of human mitral valve and identify functional variants and target genes at MVP loci.

Methods and Results: We mapped the open chromatin accessible regions in nuclei from 11 human mitral valves by an assay for transposase-accessible chromatin with high-throughput sequencing (ATAC-Seq). Compared to the heart tissue and cardiac fibroblasts, we found that mitral valve-specific ATAC-Seq peaks were enriched near genes involved in extracellular matrix organization, chondrocyte differentiation, and connective tissue development. The most enriched motif in mitral valve-specific open chromatin peaks was for the nuclear factor of activated T cells (NFATC) family of transcription factors, involved in valve endocardial and interstitial cells formation. We also found that MVP-associated variants ($p < 10^{-5}$) observed in the current MVP GWAS were significantly enriched ($p<0.05$) in mitral valve open chromatin peaks. Integration of the ATAC-Seq data with GWAS loci, extensive functional annotation, and gene reporter assay revealed plausible causal variants at two risk loci: rs6723013 at the *IGFBP5/TNS1* locus and rs2641440 at the *SMG6/SRR* locus. Circular chromosome conformation capture followed by high-throughput sequencing provided evidence for several target genes, including *SRR*, *HIC1*, and *DPH1* at the *SMG6/SRR* locus and further supported *TNS1* as the most likely target gene on Chr2.

Conclusions: Here we describe unprecedented genome-wide open chromatin profiles from human mitral valves that indicates specific gene regulation profiles, compared to the heart. We also report *in vitro* functional evidence for potential causal variants and target genes at MVP risk loci involving established and new biological mechanisms relevant to mitral valve disease.

Key Words: Mitral valve prolapse, functional annotation at GWAS loci, ATAC-Seq, 4C-Seq.

Introduction

Mitral valve presents specific cellular and tissue organization, compared to the heart and the vessels. The mature mitral valve is mainly composed of valvular interstitial cells (VICs), and is covered by a layer of valvular endothelial cells (VECs). These cells communicate via paracrine signaling, and altered or damaged signaling within the VECs can lead to pathological changes in underlying VICs.[1,2,3] VICs are relatively quiescent non-contractile fibroblast-like cells that contribute to homeostasis of extracellular matrix (ECM).[4] During valve development or under mechanical stress, VICs change phenotype and become activated myofibroblast-like cells that produce matrix metalloproteinases and inflammatory cytokines, which increase matrix production and remodel the ECM.[5,6,7] Under excessive constant stress or TGF-β stimulation, this process leads to myxomatous degeneration of the valve causing mitral valve prolapse (MVP), a common valve disease. MVP affects nearly one in 40 adults and predisposes to mitral regurgitation, arrhythmias, and sudden cardiac death.[8,9,10] Mitral regurgitation is the second most frequent indication for valve surgery, where valve reparation and replacement represent an important public health cost.[11] The causes of MVP development and evolution are still poorly understood.

In a previous genome-wide association study (GWAS) for MVP, we identified six validated genetic risk loci, all located in intergenic or intronic noncoding regions.[12] Associated variants are potentially located near regulatory elements that may be involved in the adjustment of target genes expression, specific to mitral valve. In the context of GWAS association signals, typically including large number of highly correlated genetic variants, the identification of potential causal variants and target genes requires the annotation of loci with epigenomic data generated from disease-relevant tissues.[13,14] However, the most commonly used techniques, such as DNase-seq

or histone-ChIP, require relatively high amount of cells, which excludes many tissues, including heart valve, from such analyses. On the other hand, expression quantitative trait loci (eQTL) in disease-relevant tissues are widely used to narrow down causal variants and potential target genes at GWAS loci,[15] but such datasets are not available for the mitral valve.

Here we aimed to fill in the existing gap in functional annotation datasets for human mitral valve using the assay for transposase accessible chromatin using high-throughput sequencing (ATAC-Seq) approach, a genomic technique that provides a high-resolution map of open chromatin from samples with low cell content, as is the case of mitral valve tissue. Using this genomic annotation technique, we provide valuable information about active gene expression and main regulatory elements within human mitral valve cells. As a complementary approach, and in the absence of eQTL datasets for mitral valve, we performed circular chromatin conformation capture analyzed by high-throughput sequencing (4C-Seq) to provide evidence for physical proximity between MVP-associated variants and promoters as an indication of possible regulatory event between genetic variants and target genes.[16,17] Finally, we provide two successful examples of the application of the genomic annotation, which we have generated, combined with chromatin architecture and gene reporter enhancer assay *in vitro* to propose candidate genetic variants for causality and their potential target genes at MVP GWAS loci.

Materials and Methods

Sample acquisition and patient consent

Mitral valve tissue was acquired at University Medical Center Göttingen, Georg-August University of Göttingen, Germany with written and informed consent through Institutional Review Board (IRB) protocol Nº 4/11/18. Freshly obtained mitral valve tissue from 11 patients who underwent mitral valve replacement surgery (Supplementary Table S1) was snap frozen in liquid nitrogen and stored at -80°C.

Cell culture

Primary adult human dermal fibroblasts (HDF) (ATCC, Manassas, VA) and human cardiac fibroblasts (HCF) (Cell Applications, San Diego, CA) cells were purchased at passage 2 from LGC Standards, France and Tebu-bio, France respectively, and cultured in 5% CO2 in a 37°C incubator. HDF were maintained in fibroblast basal medium (ATCC, Manassas, VA) with fibroblast growth kit-low serum (ATCC, Manassas, VA), and HCF were maintained in HCF growth medium (Cell Applications, San Diego, CA) according to the manufactures' instructions.

Isolation of nuclei from frozen mitral valves for ATAC-Seq

Frozen mitral valve tissues were mechanically dissociated in liquid nitrogen using mortar and pestle. Grinded tissue was transferred into 15 ml Falcon tubes containing 5 ml of cold homogenization buffer (HB): 10 mM Tris-HCl pH 7.6, 10 mM NaCl, 3 mM MgCl2, 0.1% NP40, 0.1% Tween-20, 0.01% Digitonin, 250mM Sucrose, and EDTA-free complete protease inhibitors (Roche, Pleasanton, CA). The homogenate was filtered through 400 µm cell strainer

(pluriStrainer, Leipzig, Germany). Then, the isolation of nuclei was performed as described[18] with modifications. Specifically, after Dounce homogenization, the tissue homogenate was pelleted at 500g for 10 minutes at 4°C. Tissue homogenate (700-800 µl) was transferred to a 14 ml round bottom tube (Greiner Bio-One, Frickenhausen, Germany) and mixed with an equal volume of 50% iodixanol in HB to obtain a final 25% iodixanol. 1200 µl of 30% iodixanol in HB was layered underneath the 25% mixture, and then 1200 µl of 40% iodixanol in HB was layered below the 30% iodixanol. Nuclei were enriched at 3,000g for 20 minutes at 4°C in a swinging-bucket centrifuge. Nuclei that accumulated at the 30%/40% interface were collected into a 1.5 ml Lo-Bind Eppendorf tube, stained with Trypan Blue, and manually counted under microscope.

Library preparation for ATAC-Seq

HDF and HCF cells at passage 5 were trypsinized, counted using a Nucleocounter NC-100 (Chemometec, Allerod, Denmark), and 50,000 cells were collected for each experiment. Isolated from mitral valves nuclei were transferred into a 1.5 ml Lo-Bind Eppendorf tube containing 1 ml ATAC-resuspension buffer with 0.1% Tween-20 and pelleted at 500 g for 10 minutes at 4°C.[18] Depending on the number of nuclei (30,000-100,000) that were obtain, the transposition reaction was scaled up or down. The transposition reaction was performed as described in the Omni-ATAC protocol.[18] Then, the transposition reaction was purified with MinElute PCR Purification Kit (QIAGEN, Hilden, Germany). Library amplification was performed as described previously.[19] Amplified DNA was purified using Agencourt AMPure XP beads (Beckman Coulter, Brea, CA), according to manufacturer's instructions.

Sequencing of ATAC-Seq libraries and peak calling

ATAC-seq libraries were sequenced using 42 paired-end sequencing cycles on an Illumina NextSeq500 system at the high throughput sequencing core facility of Institute for Integrative Biology of the Cell (CNRS, France). Reads were demultiplexed using bcl2fastq2-2.18.12, and 31 to 95 million reads (fragments) were obtained per sample. Adapter sequences were trimmed using CutAdapt v1.15. For heart samples from ENCODE, raw reads from ATAC-Seq experiments on human heart left ventricle (ENCSR117PYB, ENCSR851EBF) and human right atrium auricular region (ENCSR062SVK) were downloaded from ENCODE webserver. Read length was adjusted to 42bp using Cutadapt. Further analyses were performed on the Galaxy webserver.[20] Reads were mapped on GRCh38 (hg38) genome using Bowtie2 v2.3.4.3 with default settings, except reads could be paired at up to 2kb distance. Aligned reads were filtered using BAM filter v0.5.9, keeping only mapped, properly paired reads, and removing secondary alignment and PCR duplicate reads as well as blacklisted regions.[21] ATAC peaks were called using MACS2 callpeak v2.1.1.20160309.6 with default settings. Binary read density files (bigwig) were created using bamCoverage v3.3.0.0.0, normalized on hg38 genome. BEDTool AnnotateBed v2.29.0 was used to compare peak files.

Analysis of chromatin accessibility profiles

To perform sample correlation and principal component analysis, a common list of enriched regions was generated using bedtools multiple intersect (Galaxy Version 2.29.0) in “cluster” mode, and average read coverage on these regions was computed using deepTools multiBamsummary (Galaxy Version 3.3.2.0.0). deepTools plotCoverage and plotPCA functions were used to calculate Pearson correlation between samples and Principal Component Analysis,

163 respectively. Global peak annotation was performed using ChIPSeeker v1.22.0.[22] We used
164 Diffbind (Galaxy Version 2.10.0) to detect differentially accessible regions between heart and
165 mitral valve samples.[23] clusterProfiler v3.14.0 was used to annotate genes at proximity of
166 ATAC-Seq peaks and identify enriched gene ontology terms.[24] Identified GOBP terms were
167 clustered using REVIGO webserver (http://revigo.irb.hr/), with "medium" settings.[25] We
168 analyzed ATAC-seq peaks for enrichment in putative TF binding motifs using MEME-ChIP tool
169 on MEME webserver (http://meme-suite.org/)[26] set with HOCOMOCO Human v11 motif
170 database. We used Integrated Genome Browser (IGB) to visualize read density profiles and peak
171 positions in the context of human genome.[27]

172

173 **Prioritization of functional SNPs using annotation tools**

174 Analysis of SNP enrichment among ATAC-Seq peaks was performed using GREGOR.[28] The
175 sentinel SNPs from loci associated with p-value < 10^{-5} were used as reference for MVP-
176 associated SNPs. We included in the analysis SNPs in high LD (r^2<0.7) in European samples
177 from 1000 Genomes. To use homogenous peak sets and mitigate the difference in signal-to-noise
178 ratio between samples, 500bp windows centered on peak summits were used for this analysis.
179 Next, to prioritize the most possible functional SNPs at each of the risk loci, we annotated each
180 SNP with overlapping mitral valve ATAC-Seq peaks (narrowpeak from MCAS2 output + 100bp
181 on each side), presence of overlapping H3K27ac histone mark from heart left ventricle, heart
182 right atrium, ascending aorta or primary fibroblasts (peaks from ENCODE datasets:
183 ENCFF052BXS, ENCFF172RFM, ENCFF208DZK, ENCFF222SST, ENCFF222WPT,
184 ENCFF283CXG, ENCFF332RJQ, ENCFF385NNX, ENCFF401EEA, ENCFF491RWJ,

ENCFF617BNF, ENCFF670ZUM, ENCFF686LUV, ENCFF824SPW, ENCFF946SRJ, ENCFF964UAX, and ENCFF984MJS) and RegulomeDB score.[29]

Conditional analyses using GWAS data at *IGFBP5/TNS1* and *SMG6/SRR* loci

Using genotype data from our previous study[12], we performed an updated association analysis on chromosome 2 and chromosome 17 using newly imputed SNPs obtained from the HRC[30] as imputation reference panel through Michigan imputation server[31]. We used results from our GWAS meta-analysis as described previously.[12] We used the GCTA-COJO function to perform conditional analyses to identify secondary or causal signals on loci.[32] The COJO analysis used the MVP-France case control study as reference panel to calculate linkage disequilibrium (LD) between SNPs. To avoid collinearity issues, SNPs in high LD (r^2>0.9) with the tested SNP were excluded before analyses.

Circular chromatin conformation capture followed by high-throughput sequencing (4C-Seq)

The 4C template was prepared as previously described.[33] In summary, 10 million of HDF were fixed in 2% formaldehyde, and then the cells were lysed to isolate nuclei. The chromatin was digested *in situ* with the four-base cutter MboI (New England Biolabs, Ipswich, MA) for 16 hours at 37°C followed by *in situ* ligation with T4 ligase (Thermo Fisher Scientific, Waltham, MA). Chromatin was decross-linked at 65°C overnight, followed by a second digestion with the four-base cutter NlaIII (New England Biolabs, Ipswich, MA). Digested fragments were diluted and self-ligated using T4 ligase to generate circular chromatin conformation capture (4C) library.

The efficiency of each digestion and ligation step was validated on agarose gels. Viewpoints were selected on the basis of the MVP susceptibility loci found in the GWAS.[12] To study the chromatin interactions of MVP-associated susceptibility loci with 4C, primers (Supplementary Table S2) were designed for each viewpoint as described previously.[33] 4C libraries were sequenced using single-end 75bp reads on an Illumina NextSeq500 system at the high throughput sequencing core facility of Institute for Integrative Biology of the Cell (CNRS, France). The raw sequencing reads were demultiplexed on the basis of viewpoint-specific primer sequences. Reads were trimmed of sequences preceding MboI (GATC) site on 5' end and following NlaIII site (CATG) on 3'end (if site was present in the read) using Cutadapt (Galaxy Version 1.16.6) on Galaxy webserver.[20] Trimmed reads were mapped on GRCH37 (hg19, for r3Cseq package) and GRCH38 (hg38, for visualization) genomes using Bowtie2 v2.3.4.3 with default settings. Unmapped reads and secondary alignments were filtered out using Filter bam v0.5.9. Kept reads were processed using the R-package r3Cseq to detect significant interactions.[34] All 4C-seq images were generated using default parameters of the pipeline.

Vector construction, cell transfection, and dual-luciferase reporter gene assay

Details about primers and vector construction are presented in the supplementary appendix and in Supplementary Table S3. The constructed vector that contained the test SNP (90 ng for 96-well plate) and internal control plasmid pRL-TK (Promega, Madison, WI) (10 ng for 96-well plate) were co-transfected into HDF. The cells were plated at 4000 cells/well into a 96-well white plate 24 hours before transfection. The cells were transfected with a TransfeX transfection reagent (ATCC, Manassas, VA) according to the manufacturer's instructions. Dual-luciferase reporter gene assays were performed after 48h post-transfection. Mithras LB 940 instrument

(Berthold Technologies, Bad Wildbad, Germany) was used to measure the luciferase activity with the Dual-Luciferase Reporter Assay System (Promega, Madison, WI). All experiments were performed according to the instructions recommended by the manufacturer. The luciferase activity data were obtained from at least eight replicate wells, and the experiments were repeated 3 times. Two-tailed Student's t-test was used to compare if the difference was significant and the significance threshold was set at $P < 0.05$.

Results

Mapping the chromatin accessibility profiles of human mitral valve tissue

We sought to get insight into the regulatory features of valve tissue by determining the chromatin accessibility profiles of surgically removed mitral valves using ATAC-Seq. We obtained between 55 and 95 million paired-end reads from the generated ATAC-Seq libraries of isolated nuclei from 11 valves. Peak calling ranged between 7,165 and 74,837 peaks, with five samples having at least 30,000 peaks (Figure 1A). We calculated Spearman correlation between samples to assess the variability in our experiment. We identified two main clusters: cluster A formed by the five samples where we obtained at least 30,000 peaks, and Cluster B formed by the six remaining samples, although these were relatively less correlated, indicating more inter-sample heterogeneity (Figure 1B). The visual examination of read density profiles showed that samples from Cluster A represented optimal quality ATAC-Seq data (Figure 1C). We therefore selected Cluster A samples (MV1671, MV1700, MV1772, MV1830, and MV1846) for further analyses.

Comparison of chromatin accessibility profiles between mitral valve, heart tissues, and fibroblast cells

To determine whether mitral valve ATAC-Seq allows the identification of valve specific regulatory elements, we compared our data to primary cultures of fibroblasts and heart tissue. We determined chromatin accessibility in human dermal fibroblasts (HDF, two samples) and human cardiac fibroblasts (HCF, one sample), and reanalyzed through our pipeline ATAC-Seq experiments on heart left ventricle (Heart LV, two samples) and heart right atrium auricular region (Heart RAAR, one sample) from ENCODE. We obtained more than 130,000 peaks in all three fibroblasts samples, and 20,000 to 40,000 peaks from heart tissue (Figure 2A). We then annotated the genomic localization of ATAC-Seq peaks obtained in all samples. A large proportion of peaks observed in mitral valves were located in promoter regions (25 to 50%, according to samples, Figure 2B). Overall, 75 to 80% of peaks were located in the promoter or transcribed regions of genes, corresponding to usual localization of regulatory elements in the human genome identified by ATAC-Seq.[35]

We calculated Spearman correlation between all samples and identified three main clusters corresponding to mitral valve, heart samples, and primary cultured fibroblasts (Figure 2C). Heart samples from ENCODE clustered closer to mitral valve than to primary fibroblasts. Similarly, principal component analyses showed that tissue samples clustered together and separately from fibroblasts along the first principal component, which explains about 90% of the observed variance, whereas heart and mitral valve samples were separated by second principal component, accounting for about 5% of the total variance (Figure 2D).

Chromatin accessibility profiles present regulatory features specific to mitral valve

To detect mitral valve-specific regulatory elements, we compared mitral valve samples to heart tissue samples. We identified 9758 peaks specifically enriched in the heart or mitral valve (FDR≤0.05, Figure 3A). To assess the global gene function controlled by these putative regulatory regions, we considered genes for which the transcription starting site (TSS) was located at 10kb or less distance from the enriched peaks (promoters or proximal regulatory elements), and we performed their functional annotation using Gene Ontology Biological Process (GOBP) terms. Genes at proximity of heart-specific ATAC-Seq peaks present strong enrichment for genes involved in heart contraction ($p=6\times10^{-19}$), muscle cell differentiation ($p=3\times10^{-16}$), and other pathways related to heart function (Figure 3B, Supplementary Table 1 and 2). Mitral valve-specific ATAC-Seq peaks were strongly enriched at proximity of genes involved in different biological pathways including ECM organization ($p=5\times10^{-15}$), chondrocyte differentiation ($p=6\times10^{-11}$), and terms associated to connective tissue development (Figure 3C, Supplementary Tables 3 and 4). We also note among top enriched terms those related to cell-cell adhesion and communication (e.g actin filament organization and the ephrin receptor signaling pathway) (Supplementary Tables 3 and 4).

Results of the enrichment analyses for transcription factor motifs also differ between heart and mitral valve. The most enriched motifs among heart-specific peaks were motifs for transcription factors involved in heart development and function[36,37,38], mainly the Kruppel-like factor (KLF) and myocyte enhancer factor-2 (MEF2) families (Figure 3D). In the mitral valve-specific peaks, the top enriched motif corresponds to transcription factors of the nuclear factor of activated T cells (NFATC) family (Figure 3E).

Mitral valve specific open chromatin peaks are enriched for MVP associated loci

Mitral valve ATAC-Seq data provide potential regulatory elements that may participate to the pathogenesis of MVP. To investigate further this idea, we tested the enrichment for associated SNPs from existing MVP GWAS results among different ATAC-Seq datasets.[12] We used the Genomic Regulatory Elements and GWAS Overlap algorithm (GREGOR)[28] that compares the number of trait-associated SNPs overlapping predefined genomic regions (here, ATAC-Seq peaks) with randomized SNP matched by regions (1Mb window) and frequency. Figure 4A shows that MVP-associated SNPs ($p < 10^{-5}$) were significantly enriched ($p < 0.05$) in ATAC-Seq peaks from all mitral valve samples whereas significant enrichment was found for only one of the other samples (Heart LV-2). Maximal enrichment was seen for MV1830, for which ATAC-Seq peaks overlapped over 4 times more MVP-associated SNPs than expected ($p=6\times10^{-6}$, Figure 4A). Comparatively, ATAC-Seq peaks from fibroblasts samples did not significantly overlap with MVP-associated SNPs, despite the fact that most peaks from mitral valves were also observed in fibroblast ATAC-Seq peaks.

Identification of potential causal variants at MVP-associated loci

Given the enrichment for regulatory elements near genes involved in myxomatous valve biology, we sought to use the mitral valve ATAC-Seq datasets to narrow down a list of potential causal SNPs at MVP-associated loci. We selected associated SNPs ($p < 0.001$) from the six loci reaching genome-wide significance in our previous study[12] and filtered for colocalization with mitral valve ATAC-Seq peaks (Table 1). Between three and 20 SNPs were identified at each locus, representing a list of potential causal SNPs. We annotated these SNPs using RegulomeDB

scores from 1 (most likely to affect TF binding and expression of a target gene) to 6 (least likely), and for the presence of H3K27ac enhancer/promoter marks in heart tissue (left ventricle or right atrium), ascending aorta, and dermal fibroblasts (Table 1).

To assess the validity of our approach, we sought to confirm the potential of selected SNPs as causal for MVP association. At the *IGFBP5/TNS1* locus on Chr2, where we had observed the strongest association signal in the discovery GWAS meta-analysis,[12] only three SNPs overlapped ATAC-Seq peaks, and among them rs6723013 had the most significant association. Rs6723013 directly overlapped an H3K27ac histone mark in the heart left ventricle tissue, indicating it may be a part of an active enhancer region (Table 1). To further verify the regulatory effect of this SNP, we cloned 696bp DNA surrounding rs6723013, containing either rs6723013 reference (T) or alternative allele (G), upstream of *firefly luciferase* gene in a reporter vector. We then transfected these constructs in HDF, in which the region appeared to be active, and we measured the expression of luciferase. Interestingly, we found that T allele of rs6723013 was associated with significantly higher expression levels of firefly luciferase (reporter gene) than the alternative G allele of SNP rs6723013 ($p=0.006$), whereas the construct containing G allele did not differ significantly from a control construct containing 751bp DNA from an inactive region (Figure 4B). These results show that rs6723013 genotype affects the regulatory properties of this enhancer and support it as a causal variant. In support of this, the conditional analysis on rs6723013, caused the association signal to totally disappear and indicated that this variant to totally drive the association at this locus (Supplementary Table S4, Figure 4C-D), (Figure 4D).

We applied the same strategy to the *SMG6/SRR* locus on Chr17, which is also known as a risk locus for other cardiovascular traits, such as aortic root size and coronary artery disease.[39,40,41] Colocalization analysis showed that 21 SNPs directly overlap mitral valve ATAC-Seq peaks and

18 of these SNPs also overlap H3K27ac marks in at least one tissue, most of them in all the tested tissues. Among these SNPs, rs2641440 was ranked by RegulomeDB as the most likely regulatory SNP (RegulomeDB score=1f), and it belongs to a strongly active region in all assayed tissues (Table 1). We also selected rs9899330 for *in vitro* tests, given its RegulomeDB high score 2b (Table 1), and low linkage disequilibrium with rs2641440 (r^2=0.21, CEU 1000 genomes), suggesting it may belong to a secondary association signal. We then tested the effect of rs2641440 and rs9899330 genotypes on the regulatory activity of the surrounding region in HDF. The vector with the alternative C allele for rs2641440 had significantly higher luciferase activity compared with the construct containing the reference G allele (p=0.0017) (Supplementary Figure S1A). Both constructs that contained alternative or reference allele of rs9899330 significantly enhanced the activity of luciferase compared with the control construct, but had no allelic effect on the regulatory activity of surrounding region (Supplementary Figure S1B). Interestingly, conditional analyses both on rs2641440 and rs9899330 at *SMG6*/SRR locus (Supplementary Table S4) caused a decrease of the association signal (Supplementary Figure S1C-F), supporting the existence of two independent signals at this locus.

Identification of potential target genes at MVP-associated loci

To get further insight into the biology of MVP, we sought to determine potential target genes at each of the six confirmed MVP-associated loci[12]. First, we examined potential causal SNPs (Table 1) for eQTL association with genes in any human tissue using the Genotype-Tissue Expression (GTEx) portal (Table 2). Then, we performed circular chromatin conformation capture (4C) followed by high-throughput sequencing (4C-Seq) in human dermal fibroblasts to verify the potential target genes. We selected one to four viewpoints at each locus focusing on

potentially functional SNPs according to the functional annotation, or at the closest genomic regions where we could design primers. Then, we used r3Cseq package to determine fragments that were significantly enriched in the 4C-Seq library (Supplementary Figure S2-5). Genes were considered as potential target genes if their promoter was present within 5kb of an enriched fragment or of a potentially functional SNP (Table 2). Neither eQTL nor 4C-seq analyses showed statistically significant association with target genes at the *IGFBP5/TNS1* MVP-associated locus (Supplementary Figure S2A, B, Table 2). At the Chr3 MVP-associated locus, we found interactions between the viewpoint and the promoter region of *LMCD1* gene (Supplementary Figure S2C) and no evidence for long range regulation for the nearby *CAV3*, which is highly expressed in the heart, and rare mutations may cause long QT syndrome.[42,43] Our analyses also pointed out potential target genes at the Chr14, Chr21, and Chr22 MVP-associated loci, and further studies should explore their role in the MVP disease (Table 2, Supplementary Figure S3 and S5). On Chr17, variants at the *SMG6/SRR* locus had eQTL association with many potential target genes, including *SMG6, SRR, TSR1, SGSM2, HIC1,* and *DPH1* (Table 2). 4C-seq analyses from rs9899330 viewpoint showed significant interactions with promoters of *HIC1* and *DPH1*, which are located upstream of *SMG6*. (Figure 5, Supplementary Figure S4D). Interestingly, we detected numerous significant interactions between the selected viewpoints and the open chromatin regions within *SMG6/SRR* locus (Figure 5, Supplementary Figure S4). In particular, all viewpoints interacted with the same region containing strong enhancer signals close to a *SMG6* alternative promoter. These results suggest that active regions at this locus strongly interact together, and support the presence of multiple causal SNPs participating to the associations targeting multiple genes at this locus.

Discussion

In this study, we used epigenomic approaches to unravel the regulatory profile of human mitral valve tissue and leveraged this data to investigate functional variants and target genes at MVP-associated loci. We showed that mitral valves genome-wide open chromatin profiles are distinct from those of the cardiac tissue and fibroblasts, and are enriched near genes that are highly relevant and specific to the development of the valve and to the myxomatous process. Combining *in vitro* enhancer activity and chromatin tri-dimensional architecture, we were able to identify functional variants at the *IGFBP5/TNS1* and *SMG6/SRR* loci and narrow down the list of target genes at several MVP confirmed loci.

Our data support that open chromatin mapping of diseased mitral valves provides a useful resource on cell differentiation and regulation of the myxomatous process. We found that ATAC-Seq peaks were highly enriched for genes relevant to valve biology and disease, such as ECM and connective tissue development and cytoskeleton organization, including actin filament and cell adhesion. Of particular interest are the open chromatin peaks near genes related to the ephrin receptor signaling pathway, an important family of transmembrane tyrosine kinase receptors that ensure short cell-cell communication. This pathway is involved in the regulation of epithelial-to-mesenchymal transformation during heart valve morphogenesis, and *Efna1* knockout mice exhibit the thickening of both aortic and mitral valves.[44] Another relevant process is chondrogenic differentiation, which is part of the valve myxomatous degeneration process of VICs.[45] We found the NFATC family of transcription factors to be the most enriched motif in mitral valve-specific open chromatin regions. NFATC1, in particular, is known to be a key factor

in heart valve and VICs development and its regulation may play a role in myxomatous degeneration.[46,47,48] NFATC1 is also expressed in adult human pulmonary VECs, and VEGF-induced NFATC1 activation can participate in the valve maintenance by increasing proliferation of VECs.[49] A study reported microarray gene expression to be increased for NFATC2 and NFATC4 in myxomatous mitral valves compared to normal valves.[50] Our results encourage future studies to understand the specific role of NFATCs in the establishment of MVP.

Our study shows that MVP-associated SNPs are significantly enriched among identified mitral valve open chromatin peaks, but not among regions accessible in dermal or cardiac fibroblasts. Despite the fibroblast-like phenotype of VICs and VECs, our results are in favor of open chromatin and active regulation in mitral valve tissue to significantly differ from the one observed in cultured fibroblasts derived from the skin or the heart.

Our approach to fine-map GWAS loci using chromatin accessibility data derived from mitral valve tissue, combined with the visualization of long-range genomic interaction using 4C-Seq method, allowed us to identify potential causal variants and to narrow down the list of target genes at risk loci, especially on chromosomes 2 and 17. At the *IGFBP5/TNS1* MVP-associated locus we identified rs6723013 as a functional enhancer *in vitro* and demonstrated regulatory allelic effect. Presence of the minor T allele (T=0.24, frequency in TOPMed panel) can modulate the expression of potential target genes by altering transcription factor binding onto DNA. Several TFs, such as OVOL2 and NFE2L1, are predicted to bind more to the major G allele. OVOL2 suppresses epithelial-to-mesenchymal transition and thus maintains epithelial lineages, and NFE2L1 regulates expression of target genes in response to various stresses. Further studies are needed to explore their involvement in gene regulation by this enhancer region. Potential target genes at this locus were *IGFBP5*, *IGFBP2*, and *TNP1* that are located upstream from the

signal, and *DIRC1* and *TNS1* located downstream from the signal. According to the chromatin organization observed from our 4C-Seq data, *IGFBP2*, *IGFBP5*, and *TNS1* were all part of the same TAD, which argues for high probability of contact between these genes promoters and the enhancer containing the functional variant. *IGFBP2* and *IGFBP5* are two members of the insulin growth factor binding protein family. Igfbp5 is expressed in the developing mouse heart,[51] and differentially expressed in VECs during postnatal ECM remodeling and leaflet morphogenesis in P7 and P30 murine valves.[52] However, we have previously shown the absence of atrioventricular regurgitation properties in zebrafish knockdown experiments of *igfbp2* and *igfbp5* fish orthologs.[12] In the same study, significant atrioventricular regurgitation was found in zebrafish knockdown, and *Tns1*$^{-/-}$ mice exhibited myxomatous phenotype of mitral valves.[12] In light of these elements, we conclude that *TNS1* is the most plausible causal gene at the *IGFBP5/TNS1* locus through a potential long-range gene regulation mechanism, similar to previously described genomic organizations.[53,54]

The *SMG6/SRR* locus is an intriguing and genetically complex locus reported to be associated with several cardiovascular traits, including coronary artery disease.[40,41] Our updated association analysis, using a more dense imputed set of variants, provides several hundreds of variants in high LD to significantly associate with MVP at this locus. Filtering for the overlap with ATAC-Seq peaks in mitral valves and enhancer marks, prior to reporter assay experiments, provided evidence for enhancer capacity *in vitro* for rs2641440, with the G allelic variant (G=0.64, TOPMed) at risk for MVP. Nonetheless, the conditional analysis supported the presence of several secondary signals at this locus, and we cannot exclude additional functional effects from the untested SNPs. The *SMG6/SRR* locus is located in a gene-rich region that contains overlapping promoters. Our 4C-Seq showed that all fours viewpoints that we tested as potential

455 enhancers were located within open chromatin in mitral valves, suggesting synergic regulation of
456 several nearby genes.[55] We have previously excluded *SMG6* and *SGSM2* candidacy for causality
457 at this locus as neither *smg6* nor *sgsm2* knockdown in zebrafish led to an abnormal cardiac
458 phenotype.[12] Our chromatin interaction results now highlight new potential target genes,
459 including HIC ZBTB transcriptional repressor 1 (*HIC1*), diphthamide biosynthesis 1 (*DPH1*),
460 and serine racemase (*SRR)* genes, all expressed in heart tissues.[56,57,58] HIC1 is a partner of TCF-
461 4, and this complex attenuates Wnt/β-catenin signaling, an important mechanism for heart valve
462 development and myxomatous degeneration.[59,45] On the other hand, patients with mutations in
463 *DPH1* present heart abnormalities, including ventricular septal defect and aortic stenosis.[60]
464 Finally, several associated variants are located at the promoter of *SRR*, which catalyzes the
465 synthesis of D-serine from L-serine, and among its related pathways is calcium signaling.[61]

466

467 Our work presents several limitations. First, it is the absence of non-pathogenic valves as
468 controls, which limited direct comparison of the regulatory profiles between healthy and
469 myxomatous valves. Second, open chromatin profiles obtained from ATAC-seq methods provide
470 indirect assessment of transcriptional activity. Direct assessment of transcription using RNA-Seq
471 would be useful to determine the specific regulatory features of mitral valves, although their poor
472 cellular content represents a technical challenge. Third, enhancer activity tested *in vitro* in
473 dermal fibroblasts may differ from *in vivo* native chromatin activity in VICs and VECs. Finally,
474 our results are conditioned by the limited resolution of the 4C-Seq experiments, which do not
475 allow refinement at gene-rich loci as the *SMG6/SGSM2*.

476

In conclusion, we provide for the first time mitral valve genome-wide chromatin accessibility profiles and demonstrate that this valuable data source to substantially contribute to deciphering MVP etiology. We show that MVP-associated variants are significantly and specifically enriched in open chromatin of myxomatous valves. Using *in vitro* functional validation, in combination with long range chromatin interaction experiments, we provide evidence for causal SNPs at two MVP-associated loci and several target genes of high relevance to valve development and degenerative process.

ACKNOWLEDGMENTS

The authors thank Ines Müller and Stefanie Kestel for excellent technical assistance.

FUNDING SOURCES

This study was supported by a Ph.D. scholarship from the China Scholarship Council to MY, French Agency of Research (ANR-16-CE17-0015-02) to NB-N. AG, SK and NB-N are supported by a European Research Council grant (ERC-Stg-ROSALIND-716628). This work was supported by grants from the Deutsche Forschungsgemeinschaft (DFG) to NV (VO 1568/3-1, IRTG1816, and SFB1002 project A13), from the Else-Kröner-Fresenius Foundation to NV (EKFS 2016_A20) and from the German Center for Cardiovascular Research to NV (DZHK GOE MD3 and SE181). This work was supported in part by grants from the National Institutes of Health (GM103444 to RAN; R01HL131546, P20GM103444, and R01HL127692 to RAN; and American Heart Association (19TPA34850095 to RAN, 17CSA33590067 to RAN).

DISCLOSURES

None

REFERENCES

1. Huk DJ, Austin BF, Horne TE, Hinton RB, Ray WC, Heistad DD, Lincoln J. Valve endothelial cell-derived tgfbeta1 signaling promotes nuclear localization of sox9 in interstitial cells associated with attenuated calcification. *Arteriosclerosis, thrombosis, and vascular biology*. 2016;36:328-338
2. Menon V, Lincoln J. The genetic regulation of aortic valve development and calcific disease. *Frontiers in cardiovascular medicine*. 2018;5:162
3. Shapero K, Wylie-Sears J, Levine RA, Mayer JE, Jr., Bischoff J. Reciprocal interactions between mitral valve endothelial and interstitial cells reduce endothelial-to-mesenchymal transition and myofibroblastic activation. *Journal of molecular and cellular cardiology*. 2015;80:175-185
4. Hinton RB, Yutzey KE. Heart valve structure and function in development and disease. *Annual review of physiology*. 2011;73:29-46
5. Rabkin-Aikawa E, Farber M, Aikawa M, Schoen FJ. Dynamic and reversible changes of interstitial cell phenotype during remodeling of cardiac valves. *The Journal of heart valve disease*. 2004;13:841-847
6. Levine RA, Hagege AA, Judge DP, et al. Mitral valve disease--morphology and mechanisms. *Nature reviews. Cardiology*. 2015;12:689-710
7. Hjortnaes J, Camci-Unal G, Hutcheson JD, Jung SM, Schoen FJ, Kluin J, Aikawa E, Khademhosseini A. Directing valvular interstitial cell myofibroblast-like differentiation in a hybrid hydrogel platform. *Advanced healthcare materials*. 2015;4:121-130
8. Nishimura RA, McGoon MD. Perspectives on mitral-valve prolapse. *The New England journal of medicine*. 1999;341:48-50
9. Avierinos JF, Gersh BJ, Melton LJ, 3rd, Bailey KR, Shub C, Nishimura RA, Tajik AJ, Enriquez-Sarano M. Natural history of asymptomatic mitral valve prolapse in the community. *Circulation*. 2002;106:1355-1361
10. Spartalis M, Tzatzaki E, Spartalis E, Athanasiou A, Moris D, Damaskos C, Garmpis N, Voudris V. Mitral valve prolapse: An underestimated cause of sudden cardiac death-a current review of the literature. *Journal of thoracic disease*. 2017;9:5390-5398
11. Iung B, Vahanian A. Epidemiology of acquired valvular heart disease. *The Canadian journal of cardiology*. 2014;30:962-970
12. Dina C, Bouatia-Naji N, Tucker N, et al. Genetic association analyses highlight biological pathways underlying mitral valve prolapse. *Nature genetics*. 2015;47:1206-1211
13. Bryois J, Garrett ME, Song L, et al. Evaluation of chromatin accessibility in prefrontal cortex of individuals with schizophrenia. *Nature communications*. 2018;9:3121
14. Kyono Y, Kitzman JO, Parker SCJ. Genomic annotation of disease-associated variants reveals shared functional contexts. *Diabetologia*. 2019;62:735-743
15. Fagny M, Paulson JN, Kuijjer ML, Sonawane AR, Chen CY, Lopes-Ramos CM, Glass K, Quackenbush J, Platig J. Exploring regulation in tissues with eqtl networks. *Proceedings of the National Academy of Sciences of the United States of America*. 2017;114:E7841-E7850
16. Simonis M, Klous P, Splinter E, Moshkin Y, Willemsen R, de Wit E, van Steensel B, de Laat W. Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4c). *Nature genetics*. 2006;38:1348-1354
17. Vermunt MW, Zhang D, Blobel GA. The interdependence of gene-regulatory elements and the 3d genome. *The Journal of cell biology*. 2019;218:12-26
18. Corces MR, Trevino AE, Hamilton EG, et al. An improved atac-seq protocol reduces background and enables interrogation of frozen tissues. *Nature methods*. 2017;14:959-962
19. Buenrostro JD, Wu B, Chang HY, Greenleaf WJ. Atac-seq: A method for assaying chromatin accessibility genome-wide. *Current protocols in molecular biology*. 2015;109:21 29 21-21 29 29

20. Afgan E, Baker D, Batut B, et al. The galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic acids research*. 2018;46:W537-W544
21. Amemiya HM, Kundaje A, Boyle AP. The encode blacklist: Identification of problematic regions of the genome. *Scientific reports*. 2019;9:9354
22. Yu G, Wang LG, He QY. Chipseeker: An r/bioconductor package for chip peak annotation, comparison and visualization. *Bioinformatics*. 2015;31:2382-2383
23. Ross-Innes CS, Stark R, Teschendorff AE, et al. Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature*. 2012;481:389-393
24. Yu G, Wang LG, Han Y, He QY. Clusterprofiler: An r package for comparing biological themes among gene clusters. *Omics : a journal of integrative biology*. 2012;16:284-287
25. Supek F, Bosnjak M, Skunca N, Smuc T. Revigo summarizes and visualizes long lists of gene ontology terms. *PloS one*. 2011;6:e21800
26. Machanick P, Bailey TL. Meme-chip: Motif analysis of large DNA datasets. *Bioinformatics*. 2011;27:1696-1697
27. Freese NH, Norris DC, Loraine AE. Integrated genome browser: Visual analytics platform for genomics. *Bioinformatics*. 2016;32:2089-2095
28. Schmidt EM, Zhang J, Zhou W, Chen J, Mohlke KL, Chen YE, Willer CJ. Gregor: Evaluating global enrichment of trait-associated variants in epigenomic features using a systematic, data-driven approach. *Bioinformatics*. 2015;31:2601-2606
29. Boyle AP, Hong EL, Hariharan M, et al. Annotation of functional variation in personal genomes using regulomedb. *Genome research*. 2012;22:1790-1797
30. Vergara C, Parker MM, Franco L, Cho MH, Valencia-Duarte AV, Beaty TH, Duggal P. Genotype imputation performance of three reference panels using african ancestry individuals. *Human genetics*. 2018;137:281-292
31. Das S, Forer L, Schönherr S, et al. Next-generation genotype imputation service and methods. *Nature genetics*. 2016;48:1284-1287
32. Yang J, Ferreira T, Morris AP, et al. Conditional and joint multiple-snp analysis of gwas summary statistics identifies additional variants influencing complex traits. *Nature genetics*. 2012;44:369-375, s361-363
33. Matelot M, Noordermeer D. Determination of high-resolution 3d chromatin organization using circular chromosome conformation capture (4c-seq). *Methods in molecular biology*. 2016;1480:223-241
34. Thongjuea S, Stadhouders R, Grosveld FG, Soler E, Lenhard B. R3cseq: An r/bioconductor package for the discovery of long-range genomic interactions from chromosome conformation capture and next-generation sequencing data. *Nucleic acids research*. 2013;41:e132
35. Sheffield NC, Furey TS. Identifying and characterizing regulatory sequences in the human genome with chromatin accessibility assays. *Genes*. 2012;3:651-670
36. Nemer M, Horb ME. The klf family of transcriptional regulators in cardiomyocyte proliferation and differentiation. *Cell cycle*. 2007;6:117-121
37. Kelsey L, Flenniken AM, Qu D, et al. Enu-induced mutation in the DNA-binding domain of klf3 reveals important roles for klf3 in cardiovascular development and function in mice. *PLoS genetics*. 2013;9:e1003612
38. Desjardins CA, Naya FJ. The function of the mef2 family of transcription factors in cardiac development, cardiogenomics, and direct reprogramming. *Journal of cardiovascular development and disease*. 2016;3
39. Vasan RS, Glazer NL, Felix JF, et al. Genetic variants associated with cardiac structure and function: A meta-analysis and replication of genome-wide association data. *Jama*. 2009;302:168-178
40. Schunkert H, Konig IR, Kathiresan S, et al. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nature genetics*. 2011;43:333-338

41. Wild PS, Felix JF, Schillert A, et al. Large-scale genome-wide analysis identifies genetic variants associated with cardiac structure and function. *The Journal of clinical investigation*. 2017;127:1798-1812
42. Tyan L, Foell JD, Vincent KP, et al. Long qt syndrome caveolin-3 mutations differentially modulate kv 4 and cav 1.2 channels to contribute to action potential prolongation. *The Journal of physiology*. 2019;597:1531-1551
43. Wallace E, Howard L, Liu M, O'Brien T, Ward D, Shen S, Prendiville T. Long qt syndrome: Genetics and future perspective. *Pediatric cardiology*. 2019;40:1419-1430
44. Frieden LA, Townsend TA, Vaught DB, Delaughter DM, Hwang Y, Barnett JV, Chen J. Regulation of heart valve morphogenesis by eph receptor ligand, ephrin-a1. *Developmental dynamics : an official publication of the American Association of Anatomists*. 2010;239:3226-3234
45. Fang M, Alfieri CM, Hulin A, Conway SJ, Yutzey KE. Loss of beta-catenin promotes chondrogenic differentiation of aortic valve interstitial cells. *Arteriosclerosis, thrombosis, and vascular biology*. 2014;34:2601-2608
46. Wu B, Wang Y, Lui W, Langworthy M, Tompkins KL, Hatzopoulos AK, Baldwin HS, Zhou B. Nfatc1 coordinates valve endocardial cell lineage development required for heart valve formation. *Circulation research*. 2011;109:183-192
47. Zanotti S, Canalis E. Notch suppresses nuclear factor of activated t cells (nfat) transactivation and nfatc1 expression in chondrocytes. *Endocrinology*. 2013;154:762-772
48. Gunawan F, Gentile A, Gauvrit S, Stainier DYR, Bensimon-Brito A. Nfatc1 promotes interstitial cell formation during cardiac valve development in zebrafish. *Circulation research*. 2020;126:968-984
49. Johnson EN, Lee YM, Sander TL, Rabkin E, Schoen FJ, Kaushal S, Bischoff J. Nfatc1 mediates vascular endothelial growth factor-induced proliferation of human pulmonary valve endothelial cells. *The Journal of biological chemistry*. 2003;278:1686-1692
50. Greenhouse DG, Murphy A, Mignatti P, Zavadil J, Galloway AC, Balsam LB. Mitral valve prolapse is associated with altered extracellular matrix gene expression patterns. *Gene*. 2016;586:56-61
51. Prall OW, Menon MK, Solloway MJ, et al. An nkx2-5/bmp2/smad1 negative feedback loop controls heart progenitor specification and proliferation. *Cell*. 2007;128:947-959
52. Hulin A, Hortells L, Gomez-Stallons MV, et al. Maturation of heart valve cell populations during postnatal remodeling. *Development*. 2019;146
53. Robson MI, Ringel AR, Mundlos S. Regulatory landscaping: How enhancer-promoter communication is sculpted in 3d. *Molecular cell*. 2019;74:1110-1122
54. Schoenfelder S, Fraser P. Long-range enhancer-promoter contacts in gene expression control. *Nature reviews. Genetics*. 2019;20:437-455
55. Fishilevich S, Nudel R, Rappaport N, et al. Genehancer: Genome-wide integration of enhancers and target genes in genecards. *Database : the journal of biological databases and curation*. 2017;2017
56. Xia M, Liu Y, Figueroa DJ, et al. Characterization and localization of a human serine racemase. *Brain research. Molecular brain research*. 2004;125:96-104
57. Soliman H, Paylor B, Scott RW, et al. Pathogenic potential of hic1-expressing cardiac stromal progenitors. *Cell stem cell*. 2020;26:205-220 e208
58. Uhlen M, Fagerberg L, Hallstrom BM, et al. Proteomics. Tissue-based map of the human proteome. *Science*. 2015;347:1260419
59. Valenta T, Lukas J, Doubravska L, Fafilek B, Korinek V. Hic1 attenuates wnt signaling by recruitment of tcf-4 and beta-catenin to the nuclear bodies. *The EMBO journal*. 2006;25:2326-2337

649 60. Urreizti R, Mayer K, Evrony GD, et al. Dph1 syndrome: Two novel variants and structural and
650 functional analyses of seven missense variants identified in syndromic patients. *European journal*
651 *of human genetics : EJHG*. 2020;28:64-75
652 61. McGee MA, Abdel-Rahman AA. N-methyl-d-aspartate receptor signaling and function in
653 cardiovascular tissues. *Journal of cardiovascular pharmacology*. 2016;68:97-105

654

655

Table 1. Potential causal SNPs at the MVP-associated loci. Overlaps with indicated epigenomic marks are indicated by grey cells. P values were reported for the association with MVP obtained in the discovery meta-analysis of GWAS[12] NA: not-available.

Chr	ID	P value	ATAC-seq					RegulomeDB	H3K27ac mark			
			MV1700	MV1772	MV1830	MV1846	MV1671		Heart LV	Heart RAAR	Asc. Aorta	Fibroblasts
chr2	rs17778091	$7.1x10^{-4}$	■	■		■	■	5	■			
	rs6723013	$1.8x10^{-8}$	■			■	■	3a				
	rs17778329	$5.7x10^{-4}$	■			■		2b		■	■	
chr3	rs165177	$5.2x10^{-5}$	■	■			■	5	■	■	■	
	rs164966	$1.7x10^{-5}$	■					4		■	■	
	rs165172	$1.6x10^{-5}$	■				■	4	■	■	■	
chr14	rs7155259	$3.2x10^{-4}$					■	2b				
	rs7156194	$3.1x10^{-4}$					■	1b				
	rs56201715	$7.2x10^{-4}$					■	5				
	rs61991204	$3.2x10^{-5}$	■				■	5				
	rs61991205	$3.4x10^{-5}$	■		■		■	5				
	rs17767410	$3.4x10^{-5}$	■		■		■	5				
	rs35609708	$3.8x10^{-5}$					■	2b				
chr17	rs2641440	$5.0x10^{-4}$	■		■		■	1f	■	■	■	■
	rs2760734	$1.5x10^{-4}$		■				4	■	■	■	■
	rs8068642	$1.4x10^{-4}$	■					3a	■	■	■	■
	rs8082485	$1.4x10^{-4}$	■					4	■		■	■
	rs898749	$1.3x10^{-4}$	■	■	■	■	■	3a	■	■	■	■
	rs898748	$1.3x10^{-4}$			■		■	3a	■	■	■	■
	rs7213715	$1.7x10^{-4}$	■					4	■	■	■	■
	rs9895551	$5.5x10^{-6}$					■	4	■	■	■	■
	chr17:2117652:D	$6.9x10^{-6}$					■	NA	■	■	■	■
	chr17:2117658:D	$1.6x10^{-6}$					■	NA	■	■	■	■
	rs11655813	$3.4x10^{-6}$	■		■		■	4	■	■	■	■
	rs216172	$3.3x10^{-6}$	■	■		■	■	4	■	■	■	■
	rs9899330	$3.2x10^{-6}$					■	2b				■
	rs408067	$9.2x10^{-5}$	■	■	■	■	■	4	■	■	■	
	rs3744270	$1.6x10^{-4}$			■		■	4	■	■	■	■
	rs12450028	$1.1x10^{-4}$			■		■	2a	■	■	■	■
	rs113423156	$7.6x10^{-4}$			■			3a				
	chr17:2220814:D	$1.6x10^{-4}$			■			NA				
	rs4477758	$3.4x10^{-4}$	■					6				
	rs3815469	$6.7x10^{-5}$		■	■		■	2b	■	■	■	
chr21	rs112479655	$5.8x10^{-4}$					■	NA				
	rs2255858	$3.9x10^{-5}$	■					4			■	
	rs75196102	$3.8x10^{-4}$		■	■	■	■	4	■	■	■	■
	rs7280136	$2.6x10^{-5}$	■	■	■	■	■	3a	■	■	■	■
chr22	rs1859136	$4.9x10^{-5}$					■	4				
	chr22:28206144:D	$4.5x10^{-8}$	■				■	NA	■	■	■	
	rs6005631	$3.2x10^{-6}$	■	■			■	4	■	■	■	
	rs73164492	$8.4x10^{-8}$	■	■	■	■	■	4	■	■	■	■
	rs11705555	$4.4x10^{-8}$	■	■	■	■	■	2b	■	■	■	
	rs13057169	$8.4x10^{-8}$	■	■	■	■	■	4	■	■	■	
	rs5997307	$3.5x10^{-6}$		■			■	4	■	■	■	

659 **Table 2. Potential target genes at MVP-associated loci.**

Chr	Index SNP	P value	Potential causal SNPs	eQTL genes	Tissue with highest eQTL association	Best eQTL SNP	Risk allele	eQTL p value	Effect size	Topological association
Chr2	rs12465515	3.1×10^{-11}	3							
Chr3	rs171408	1.3×10^{-11}	3	*LMCD1*	Heart - Atrial Appendage	rs165177	C	8×10^{-7}	0.27	Y
				LMCD1-AS1	Pituitary	rs165172	C	4×10^{-6}	0.47	Y
Chr14	rs17767392	2.3×10^{-8}	7	*SIPA1L1-AS1*	Esophagus - Muscularis	rs17767410	C	1×10^{-12}	-0.43	Y
				SIPA1L1	Esophagus - Muscularis	rs17767410	C	6×10^{-12}	-0.27	Y
				PCNX1	Skin - Sun Exposed (Lower leg)	rs7155259	C	1×10^{-6}	0.17	N
				RP6-91H8.5	Brain - Cortex	rs7155259	C	1×10^{-5}	0.36	N
				COX16	Whole Blood	rs56201715	G	9×10^{-5}	0.1	N
Chr17	rs216205	1.5×10^{-8}	20	*SRR*	Muscle - Skeletal	rs3744270	G	1×10^{-106}	0.8	Y
				TSR1	Skin - Sun Exposed (Lower leg)	rs4477758	T	3×10^{-26}	-0.22	Y
				SGSM2	Skin - Sun Exposed (Lower leg)	rs4477758	T	6×10^{-26}	-0.35	Y
				SGSM2-AS1	Skin - Not Sun Exposed (Suprapubic)	rs408067	C	1×10^{-15}	0.25	Y
				HNRNPA1P16	Nerve - Tibial	rs3744270	G	8×10^{-13}	-0.34	N
				HIC1	Cells - Cultured fibroblasts	rs3744270	G	1×10^{-7}	-0.13	Y
				DPH1	Pancreas	rs9899330	A	8×10^{-7}	-0.27	Y
				SMG6	Testis	rs12450028	C	9×10^{-7}	-0.14	Y
				RTN4RL1	Heart - Left Ventricle	rs3815469	G	3×10^{-5}	-0.25	N
Chr21	rs62229266	1.2×10^{-8}	4	*LINC01436*	Artery - Tibial	rs2255858	T	1×10^{-78}	0.93	Y
				RPL23AP3	Cells - Cultured fibroblasts	rs2255858	T	2×10^{-35}	0.65	Y
				SETD4	Cells - Cultured fibroblasts	rs2255858	T	5×10^{-19}	0.26	Y
				CBR1	Nerve - Tibial	rs2255858	T	5×10^{-10}	0.19	N
				AP000688.29	Artery - Tibial	rs2255858	T	8×10^{-9}	0.29	N
Chr22	rs11705555	1.4×10^{-8}	7	*MN1*	Whole Blood	rs73164492	C	3×10^{-5}	-0.21	Y
				TTC28	Muscle - Skeletal	rs11705555	A	6×10^{-5}	-0.15	N

660

661 Candidate SNPs for causality from Table 1 were used at each locus to interrogate potential eQTL association from GTEx v8 database. For each
662 eGene identified, the strongest association is reported (strongest tissue, SNP eQTL p-value, and effect size). Effect size sign was adjusted to
663 represent the effect of MVP risk allele. P-values are for MVP association reported from the discovery and replication results of each locus.[12] Last
664 column indicates whether the TSS of indicated eGenes is located either within 5kb of a potential causal SNP or within 5kb of a fragment that
665 interacts topologically with potential causal SNPs, from our 4C-Seq experiments.

FIGURE LEGENDS

Figure 1. Quality controls of mitral valve ATAC-Seq experiments.

A: Number of reads (grey) and number of peaks (orange) obtained for mitral valve ATAC-Seq libraries. **B**: Spearman correlation and hierarchical clustering of mitral valve ATAC-Seq datasets. **C**: Representative read density profile of mitral valve ATAC-Seq datasets on a gene rich portion of chr1 (11,000,000 to 17,500,000, hg38 coordinates).

Figure 2. Comparison of mitral valve, heart, and fibroblast cells ATAC-Seq datasets.

A: Number of reads (grey) and number of peaks (orange) obtained for HDF and HCF ATAC-Seq libraries as well as heart tissues: left ventricle (Heart LV) and right atrium auricular region (Heart RAAR) raw reads obtained from ENCODE database. **B**: Proportion of ATAC-Seq peaks located to different genomic regions. **C**: Spearman correlation and hierarchical clustering of ATAC-Seq datasets from 5 mitral valve samples, 3 fibroblasts samples, and 3 heart tissue samples from left ventricle (Heart LV) or right atrium auricular region (Heart RAAR). **D**: Principal Component Analysis of 5 mitral valve, 3 fibroblast, and 3 heart ATAC-Seq datasets. Upper panel shows the position of samples with respect to first 2 principal components. Lower panel indicates the eigenvalues and explained variance of the first 10 principal components.

Figure 3. Analysis of ATAC-Seq identifies mitral valve-specific regulatory elements and pathways.

A: Volcano plot representing negative logarithm of enrichment false discovery rate (FDR) on the Y axis and logarithm of enrichment in heart samples over mitral valve samples on the X axis. Each dot represents an ATAC-Seq peak in the mitral valve or heart sample. Pink dots represent significantly enriched regions in mitral valve or heart ATAC-Seq experiments (FDR≤0.05). **B-C**: Bubble graphs represent GOBP terms enriched among genes at proximity (distance to TSS≤10kb) of heart- (**B**) and mitral valve- (**C**) specific peaks. X and Y axes are in arbitrary semantic coordinates.[25] Bubble size represents number of semantically similar terms aggregated under the same index GO term. Bubble color represents the p-value of GO term enrichment for

694 the index GO term. **D**-**E**: Motif sequence and logo of most enriched motifs in heart- (**D**) and
695 mitral valve- (**E**) specific peaks. Motifs were detected *de novo* using DREME algorithm. Top 3
696 motifs are represented. E-val indicates the erased expected value (E-value) calculated by
697 DREME. "Transcription factors" indicate the top transcription factor or transcription factor
698 family detected by TOMTOM algorithm as possible factors binding the detected motif. "Central
699 enrichment" represents the enrichment of motifs with respect to ATAC-Seq peak summits.

700

701 **Figure 4. rs6723013 is a potential causal variant at the *IGFBP5/TNS1* MVP-associated**
702 **locus**.

703 **A**: Representation of MVP SNPs fold-enrichment (X-axis) and enrichment p-value (log scale, Y-
704 axis) among indicated ATAC-Seq samples. MVP SNPs overlap with ATAC-Seq peaks was
705 compared to 500 pools of randomized matched SNPs to calculate the indicated enrichments. **B**:
706 Luciferase reporter gene assay comparing the regulatory activity of constructs containing
707 rs6723013-region (T and G alleles) and a control region. The p-value of a student's t-test
708 comparing luciferase values of rs6723013 alleles is indicated. **C**-**D**: Conditional analysis at
709 *IGFBP5/TNS1* locus. LocusZoom plots represent MVP association before (**C**) and after (**D**)
710 conditioning on rs6723013.

711

712 **Figure 5. 4C-Seq analysis of the *SMG6/SRR* MVP-associated locus.**

713 Interacting fragment density profiles from SNP rs2641440, rs9898819, rs2281727, and
714 rs9899330 viewpoints using 4C-Seq. Fragments in significant interaction with the viewpoint are
715 indicated by black bars under the interaction plot. Genomic coordinates, GRCh38.

716

Fig 1

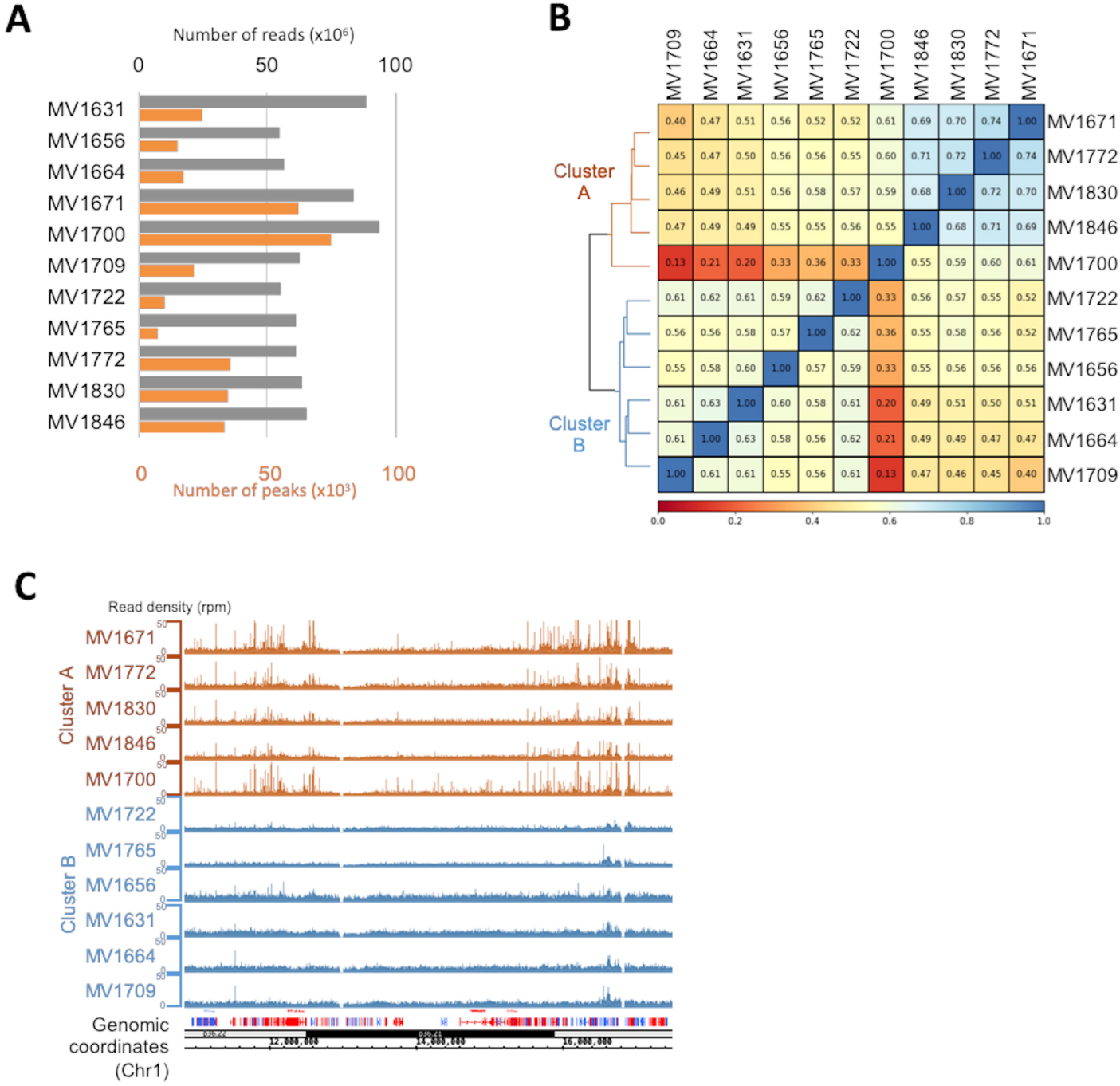

A
Number of reads (x10⁶)
0
50
100
MV1631
MV1656
MV1664
MV1671
MV1700
MV1709
MV1722
MV1765
MV1772
MV1830
MV1846
0
50
100
Number of peaks (x10³)
B
MV1709
MV1664
MV1631
MV1656
MV1765
MV1722
MV1700
MV1846
MV1830
MV1772
MV1671
Cluster A
Cluster B
0.0
0.2
0.4
0.6
0.8
1.0
C
Read density (rpm)
Cluster A
Cluster B
MV1671
MV1772
MV1830
MV1846
MV1700
MV1722
MV1765
MV1656
MV1631
MV1664
MV1709
Genomic coordinates (Chr1)
12,000,000
14,000,000
16,000,000

Fig 2

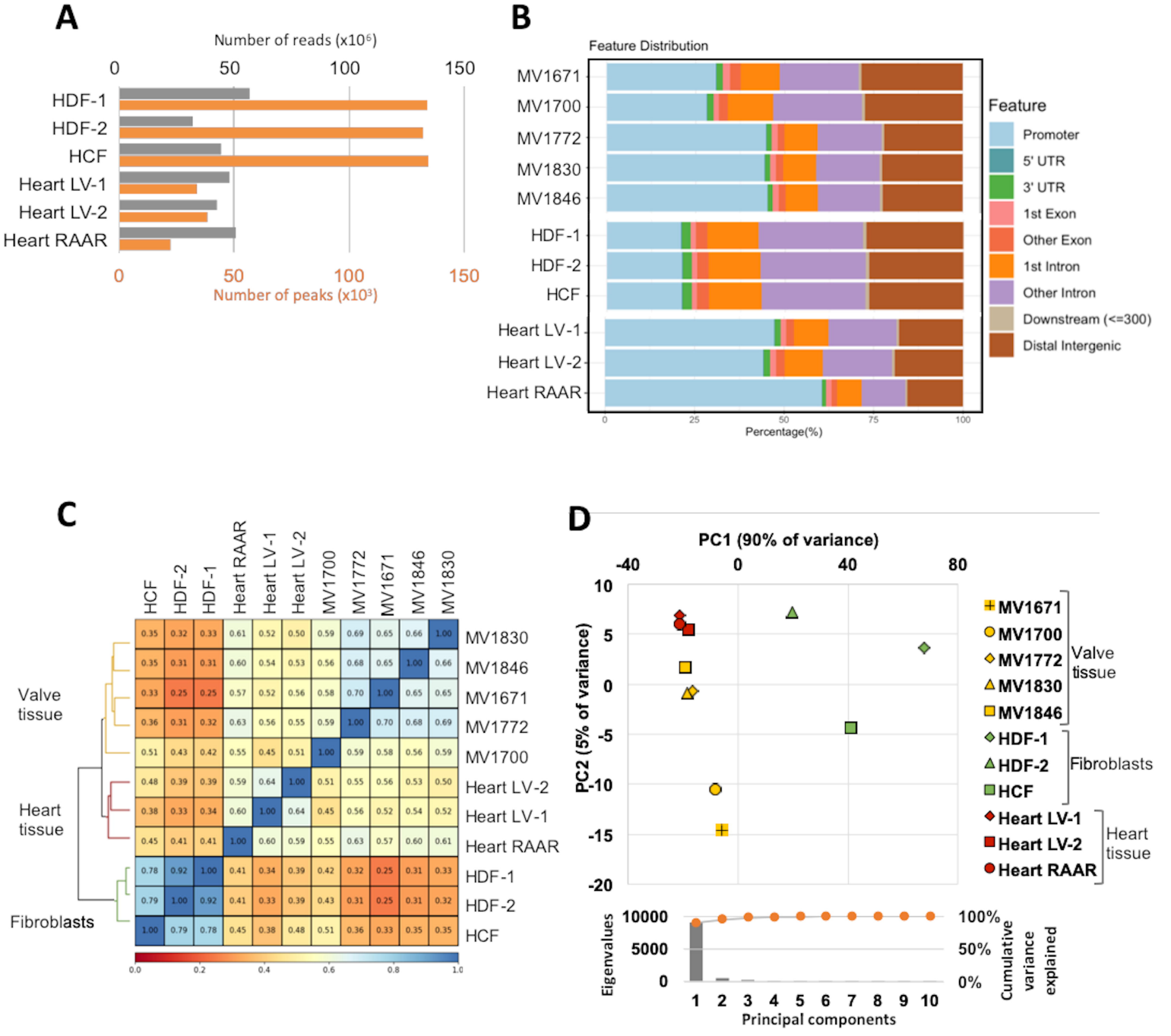
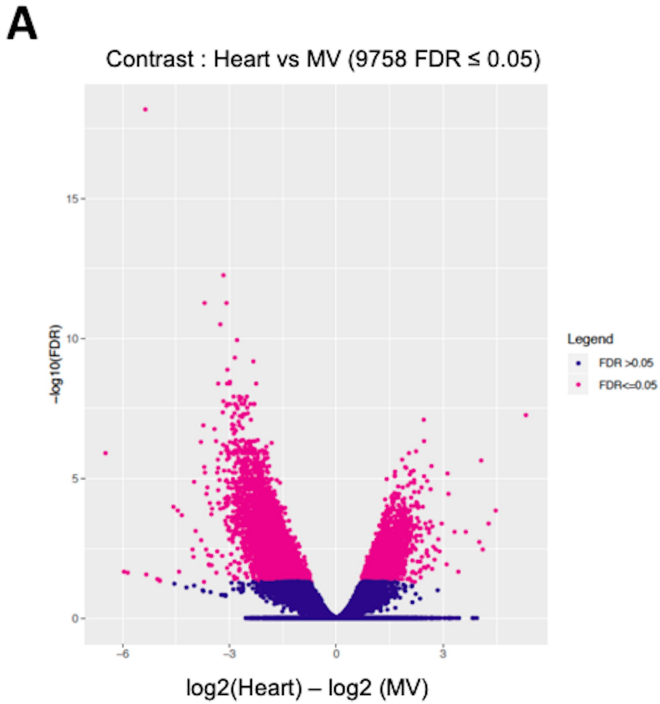
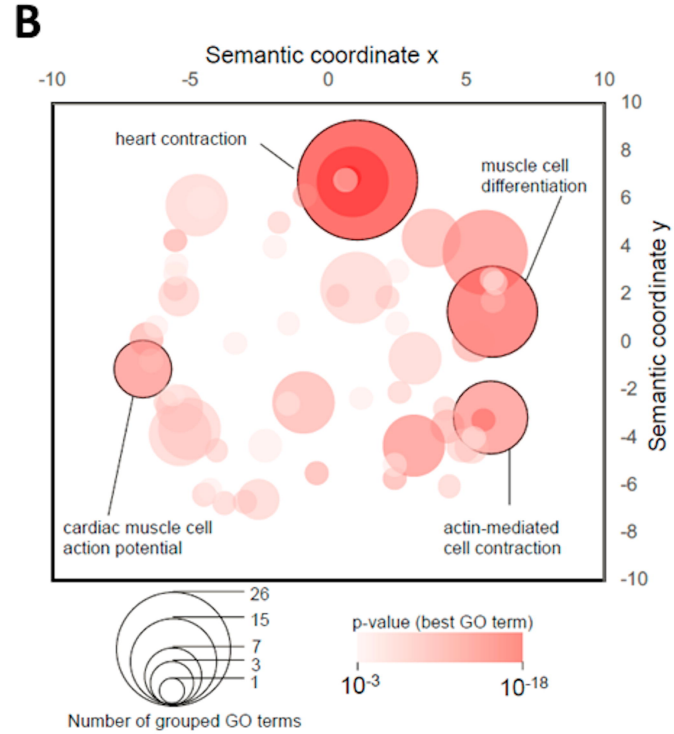
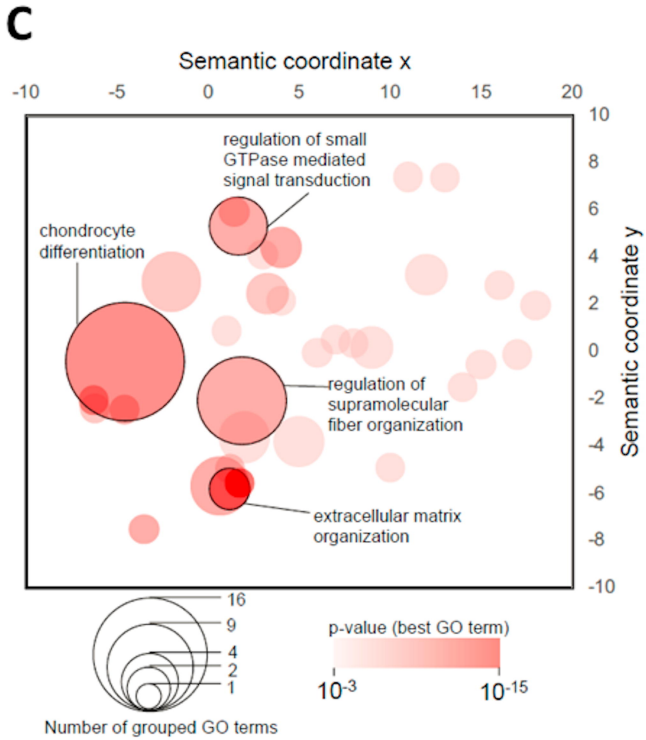

A
Number of reads (x10⁶)
0
50
100
150
HDF-1
HDF-2
HCF
Heart LV-1
Heart LV-2
Heart RAAR
0
50
100
150
Number of peaks (x10³)
B
Feature Distribution
MV1671
MV1700
MV1772
MV1830
MV1846
HDF-1
HDF-2
HCF
Heart LV-1
Heart LV-2
Heart RAAR
0
25
50
75
100
Percentage(%)
Feature
Promoter
5' UTR
3' UTR
1st Exon
Other Exon
1st Intron
Other Intron
Downstream (<=300)
Distal Intergenic
C
HCF
HDF-2
HDF-1
Heart RAAR
Heart LV-1
Heart LV-2
MV1700
MV1772
MV1671
MV1846
MV1830
Valve tissue
Heart tissue
Fibroblasts
0.0
0.2
0.4
0.6
0.8
1.0
D
PC1 (90% of variance)
-40
0
40
80
PC2 (5% of variance)
10
5
0
-5
-10
-15
-20
MV1671
MV1700
MV1772
MV1830
MV1846
Valve tissue
HDF-1
HDF-2
HCF
Fibroblasts
Heart LV-1
Heart LV-2
Heart RAAR
Heart tissue
Eigenvalues
10000
5000
0
100%
50%
0%
Cumulative variance explained
1
2
3
4
5
6
7
8
9
10
Principal components

Fig 3







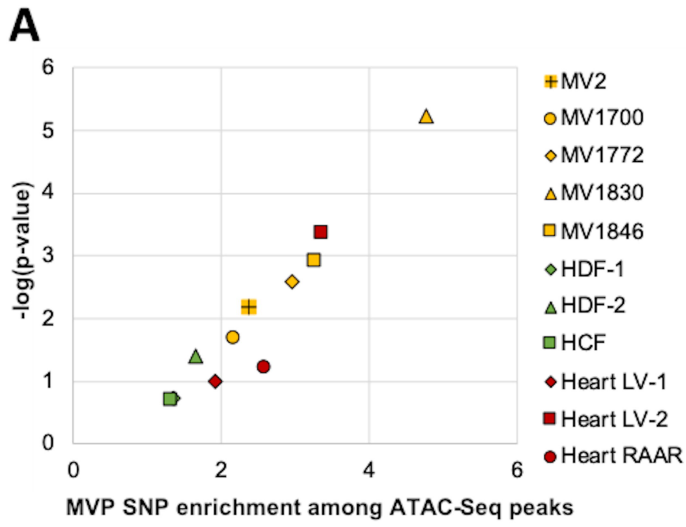
D

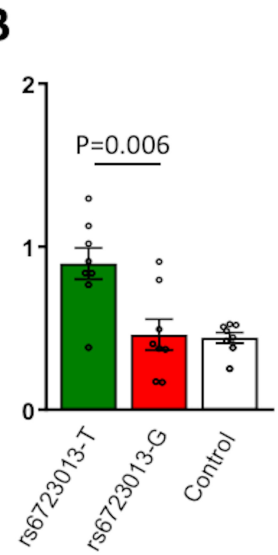
	Motif	Logo fw	Logo rv	E-val	Transcription factors	Central enrichment
1.	GGCGGRGM			1.7e-056	KLF	
2.	AWATAR			3.8e-032	MEF2	Not Centrally Enriched
3.	HGTCAY			2.2e-023	PAX2/ JUN/…	

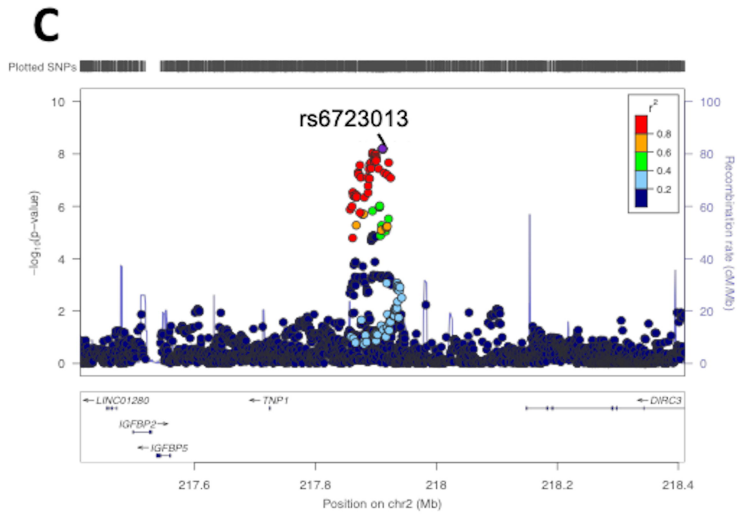
E

	Motif	Logo fw	Logo rv	E.val	Transcription factors	Central enrichment
1.	DGGAAW			4.5e-061	NFATC	Not Centrally Enriched
2.	DGCCAR			7.3e-052	NFIC/ NFIB/ …	
3.	AMAYA			7.7e-037	FUBP1	Not Centrally Enriched

Fig 4


A
-log(p-value)
MVP SNP enrichment among ATAC-Seq peaks
MV2
MV1700
MV1772
MV1830
MV1846
HDF-1
HDF-2
HCF
Heart LV-1
Heart LV-2
Heart RAAR


B
P=0.006
rs6723013-T
rs6723013-G
Control


C
Plotted SNPs
rs6723013
Recombination rate (cM/Mb)
LINC01280
TNP1
DIRC3
IGFBP2
IGFBP5
Position on chr2 (Mb)

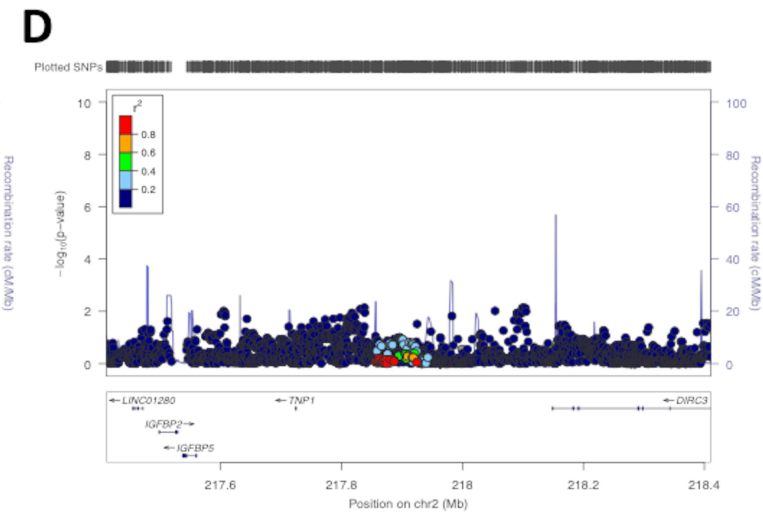

D
Plotted SNPs
Recombination rate (cM/Mb)
LINC01280
TNP1
DIRC3
IGFBP2
IGFBP5
Position on chr2 (Mb)

Fig 5

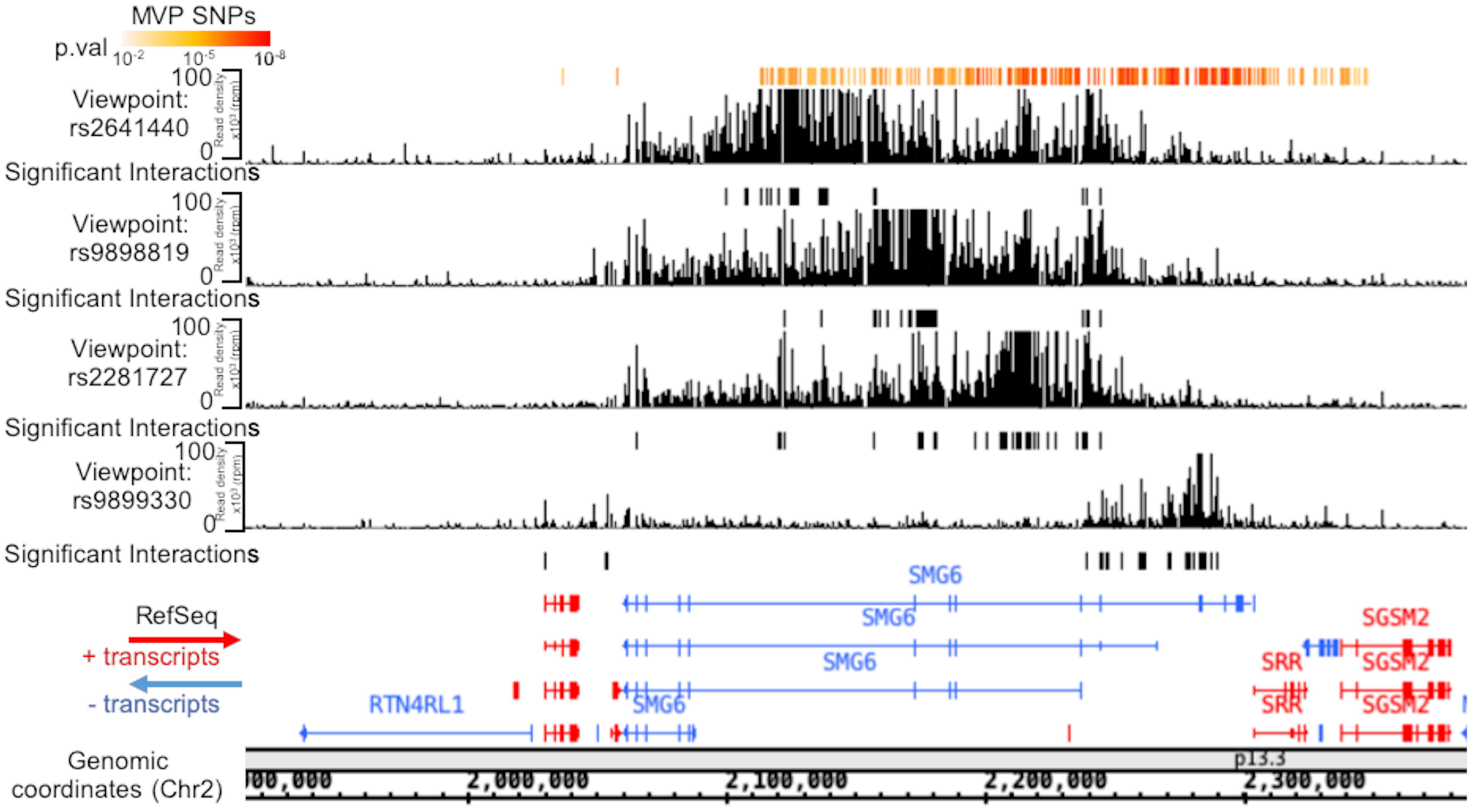
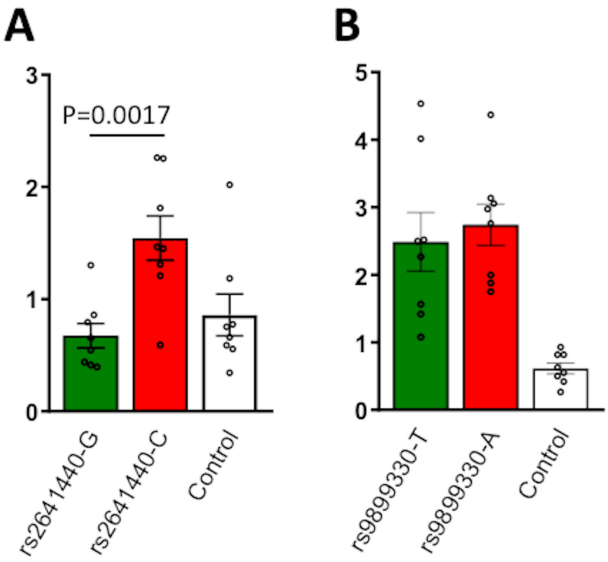
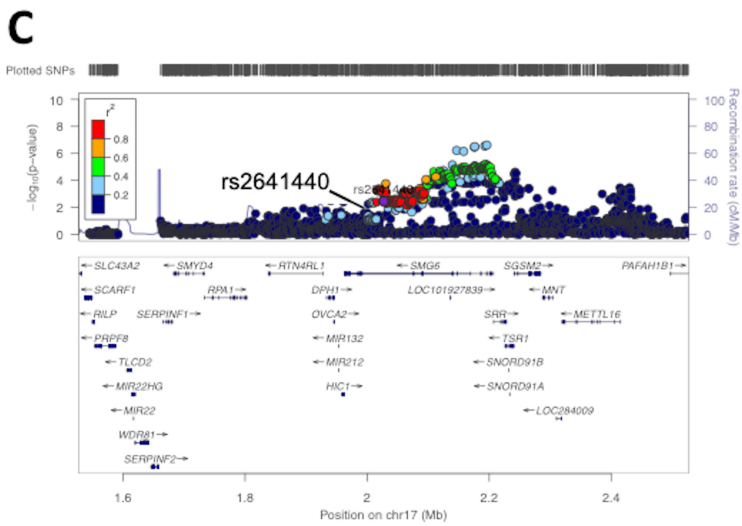
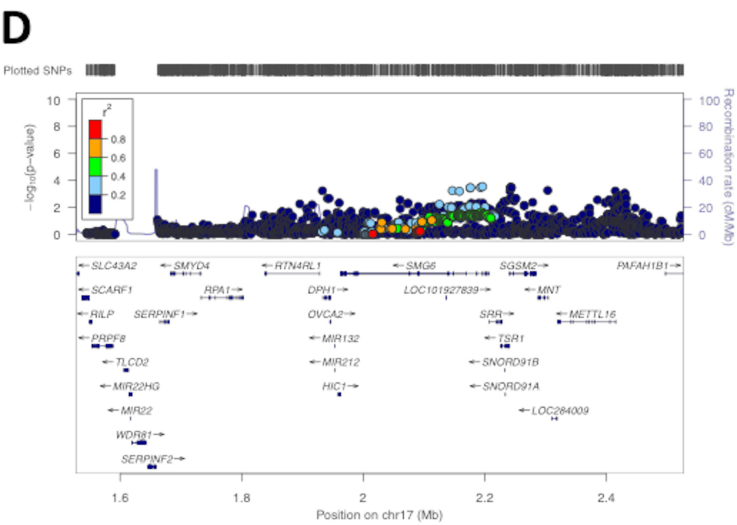
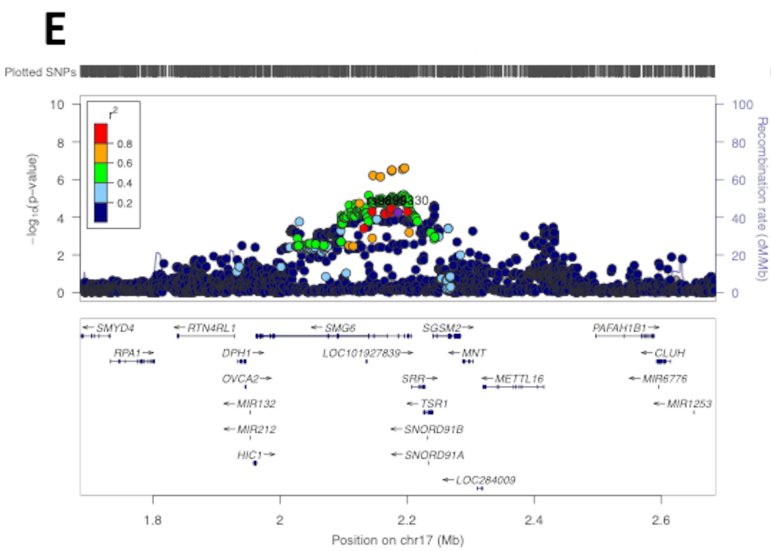

MVP SNPs
p.val
10-2
10-5
10-8
100
Viewpoint:
rs2641440
0
Read density
x10³ (rpm)
Significant Interactions
100
Viewpoint:
rs9898819
0
Read density
x10³ (rpm)
Significant Interactions
100
Viewpoint:
rs2281727
0
Read density
x10³ (rpm)
Significant Interactions
100
Viewpoint:
rs9899330
0
Read density
x10³ (rpm)
Significant Interactions
RefSeq
+ transcripts
- transcripts
SMG6
SMG6
SMG6
SMG6
RTN4RL1
SGSM2
SRR
SGSM2
SRR
SGSM2
p13.3
Genomic
coordinates (Chr2)
000,000
2,000,000
2,100,000
2,200,000
2,300,000

Fig S1


A
P=0.0017
rs2641440-G
rs2641440-C
Control
B
rs9899330-T
rs9899330-A
Control


C
Plotted SNPs
rs2641440
Position on chr17 (Mb)


D
Plotted SNPs
Position on chr17 (Mb)


E
Plotted SNPs
Position on chr17 (Mb)

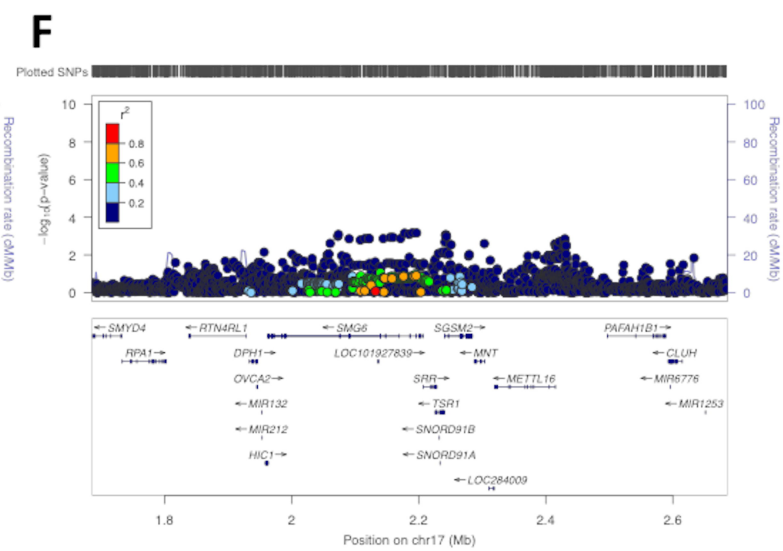

F
Plotted SNPs
Position on chr17 (Mb)

Fig S2

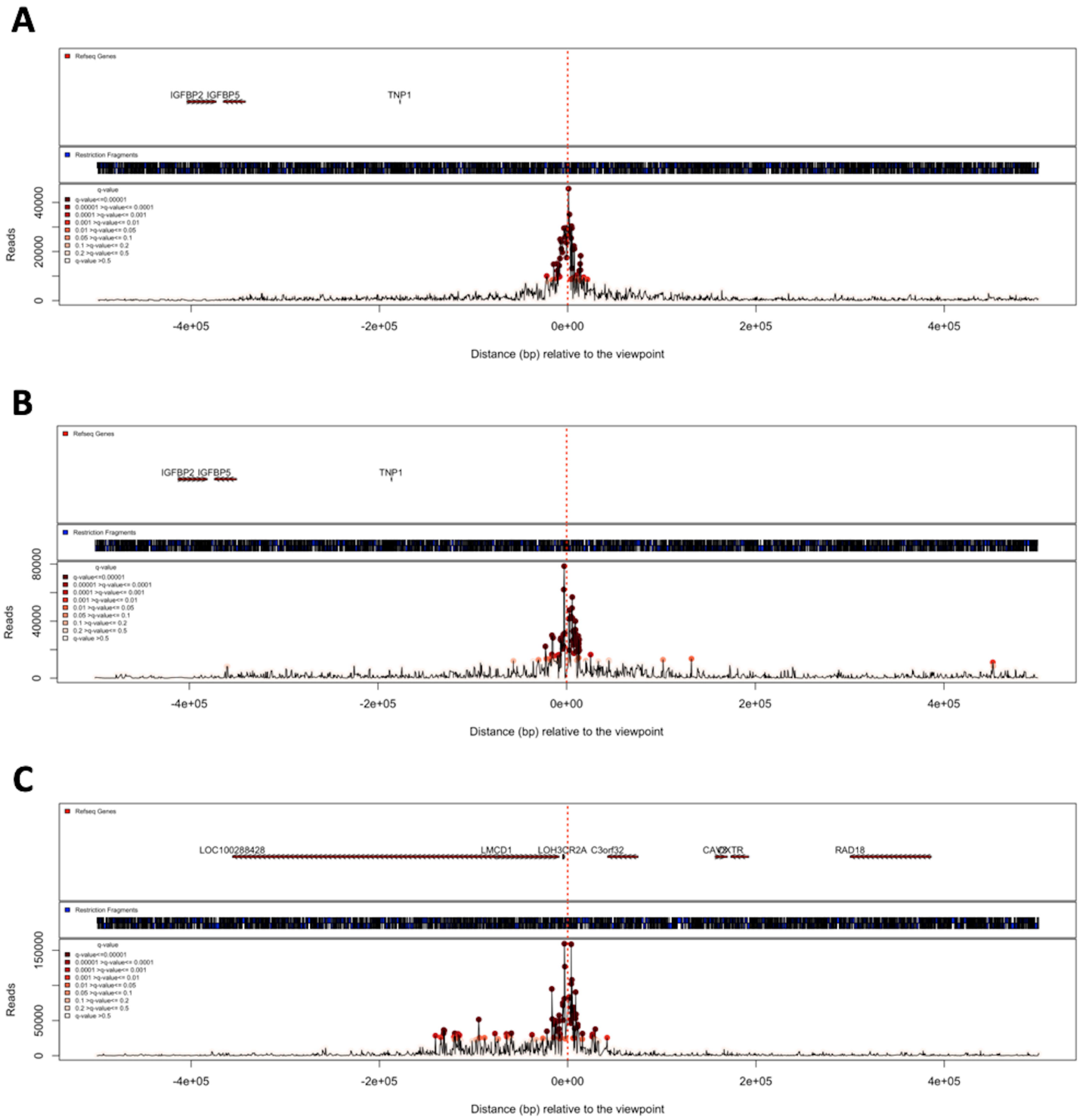
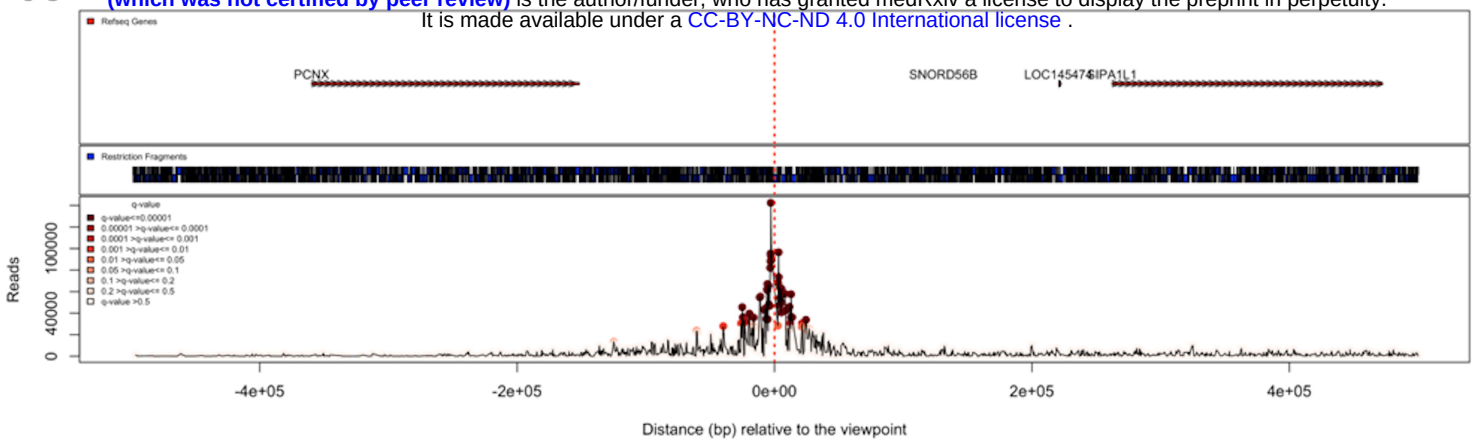
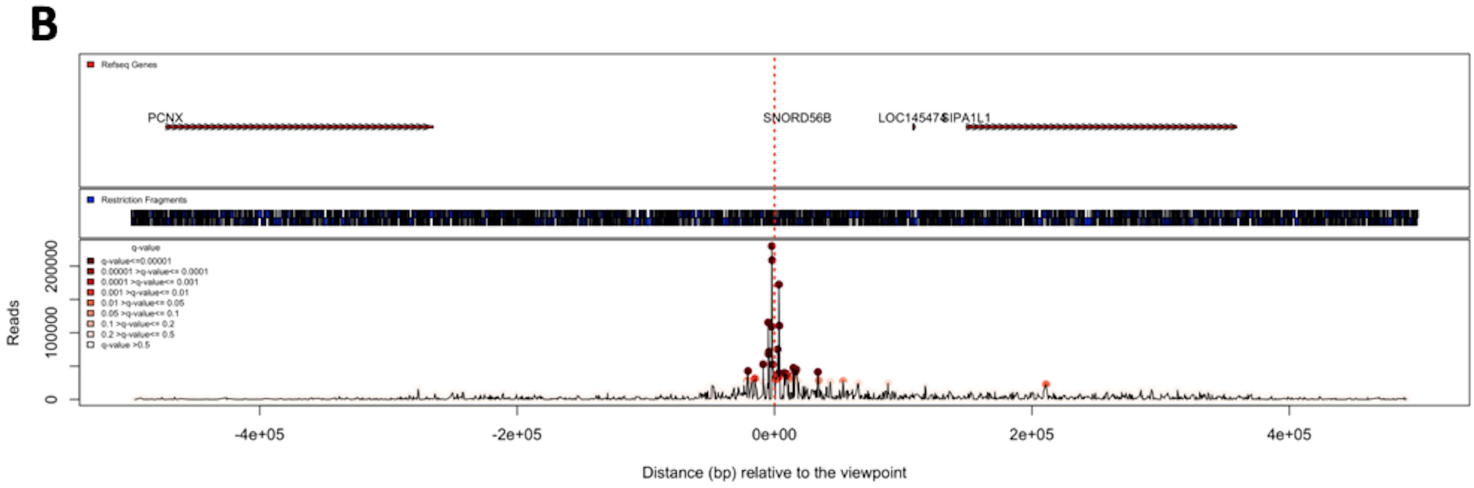
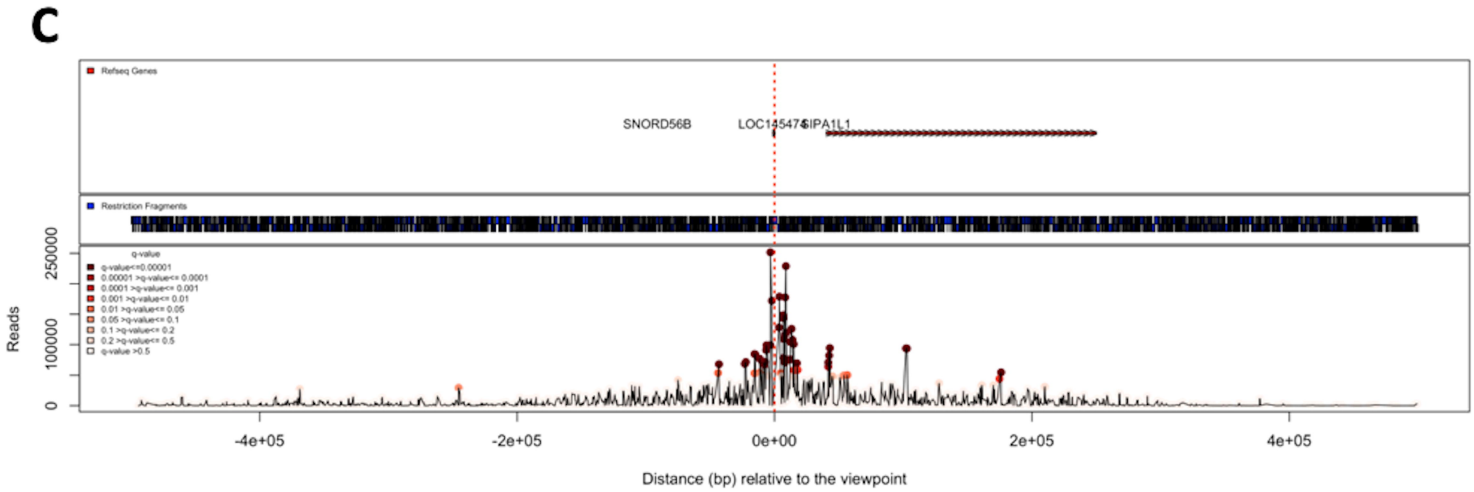

A
Refseq Genes
IGFBP2 IGFBP5
TNP1
Restriction Fragments
q-value
q-value<=0.00001
0.00001 >q-value<= 0.0001
0.0001 >q-value<= 0.001
0.001 >q-value<= 0.01
0.01 >q-value<= 0.05
0.05 >q-value<= 0.1
0.1 >q-value<= 0.2
0.2 >q-value<= 0.5
q-value >0.5
Reads
0
20000
40000
-4e+05
-2e+05
0e+00
2e+05
4e+05
Distance (bp) relative to the viewpoint
B
Refseq Genes
IGFBP2 IGFBP5
TNP1
Restriction Fragments
q-value
q-value<=0.00001
0.00001 >q-value<= 0.0001
0.0001 >q-value<= 0.001
0.001 >q-value<= 0.01
0.01 >q-value<= 0.05
0.05 >q-value<= 0.1
0.1 >q-value<= 0.2
0.2 >q-value<= 0.5
q-value >0.5
Reads
0
40000
80000
-4e+05
-2e+05
0e+00
2e+05
4e+05
Distance (bp) relative to the viewpoint
C
Refseq Genes
LOC100288428
LMCD1
LOH3CR2A
C3orf32
CAV2
XTR
RAD18
Restriction Fragments
q-value
q-value<=0.00001
0.00001 >q-value<= 0.0001
0.0001 >q-value<= 0.001
0.001 >q-value<= 0.01
0.01 >q-value<= 0.05
0.05 >q-value<= 0.1
0.1 >q-value<= 0.2
0.2 >q-value<= 0.5
q-value >0.5
Reads
0
50000
150000
-4e+05
-2e+05
0e+00
2e+05
4e+05
Distance (bp) relative to the viewpoint

Fig S3







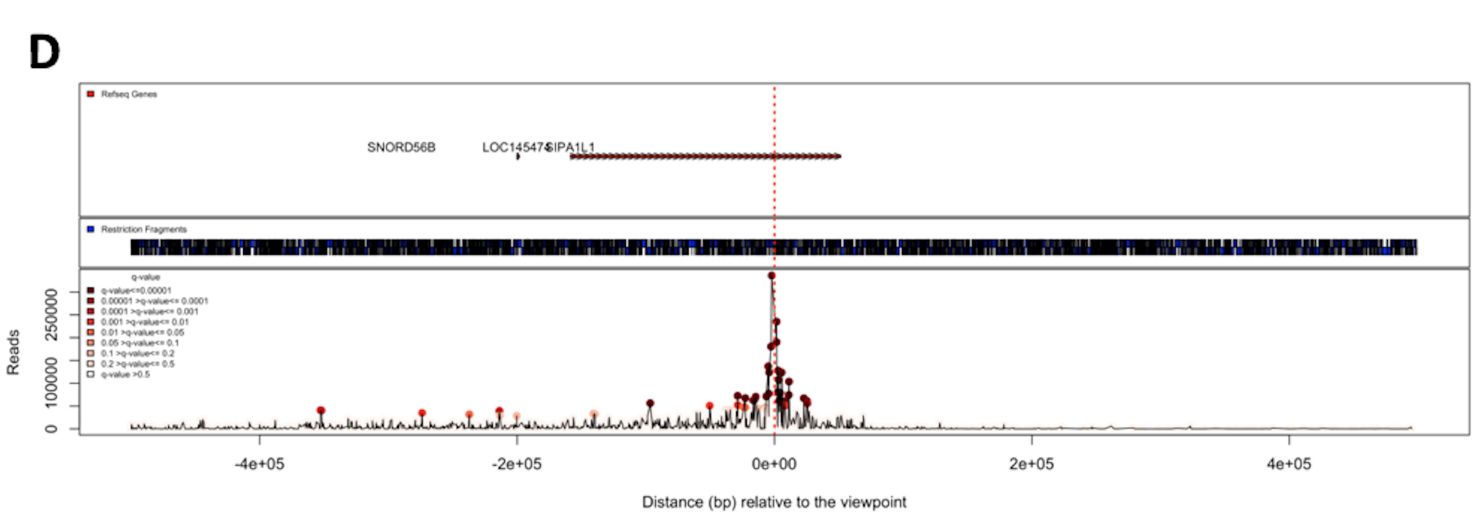
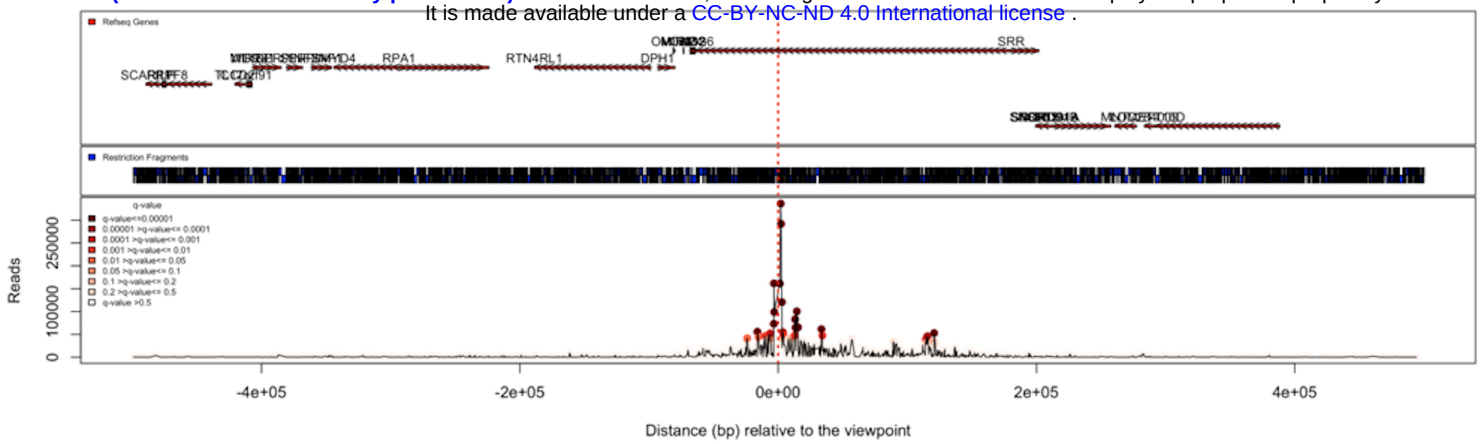
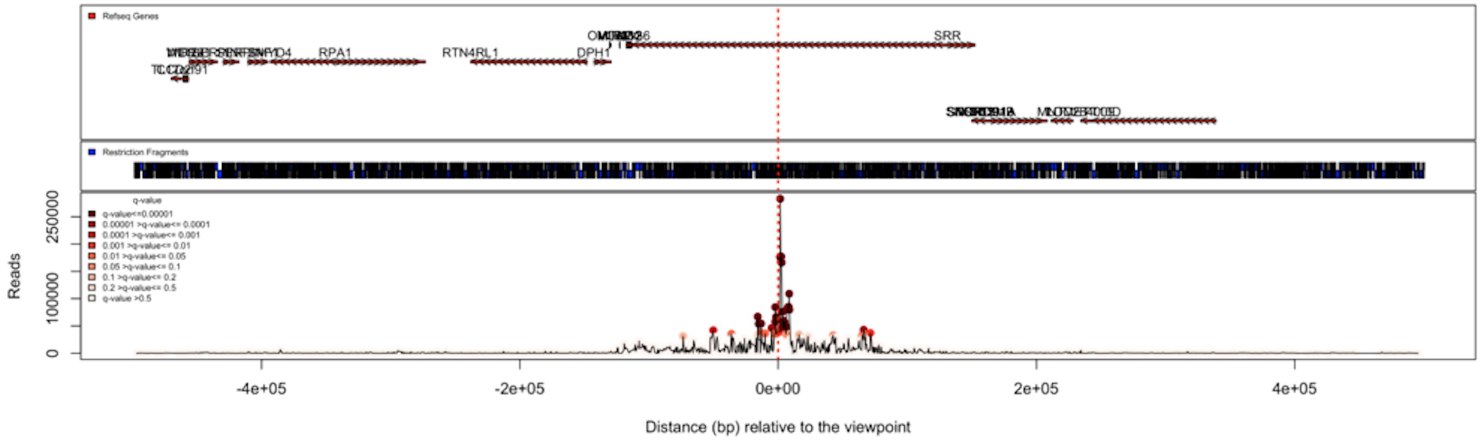
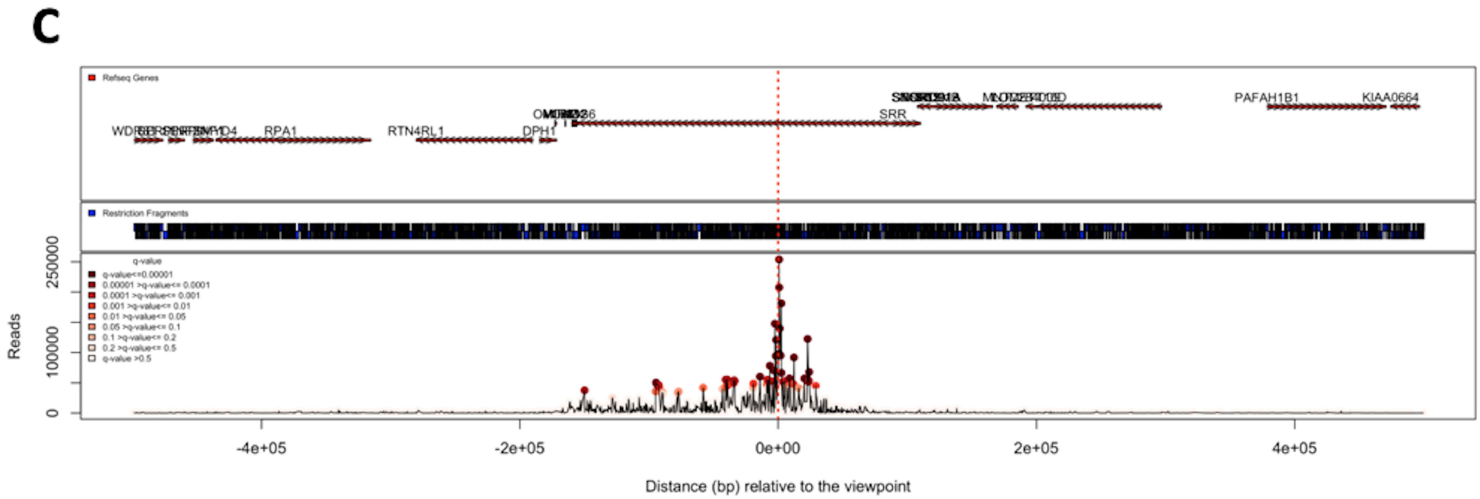


Fig S4







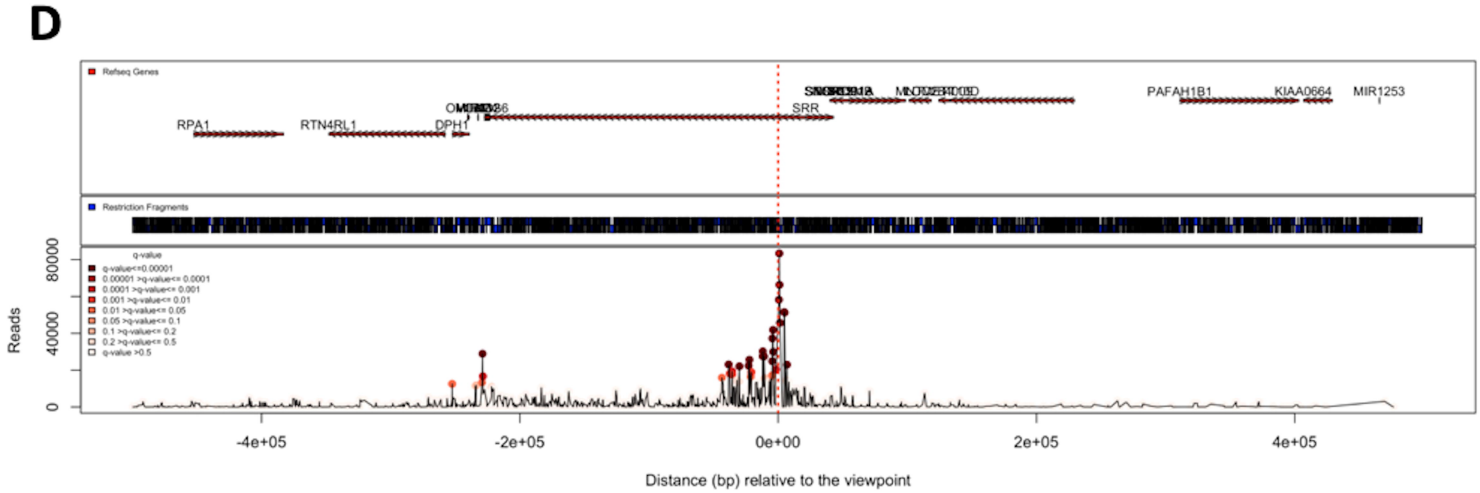
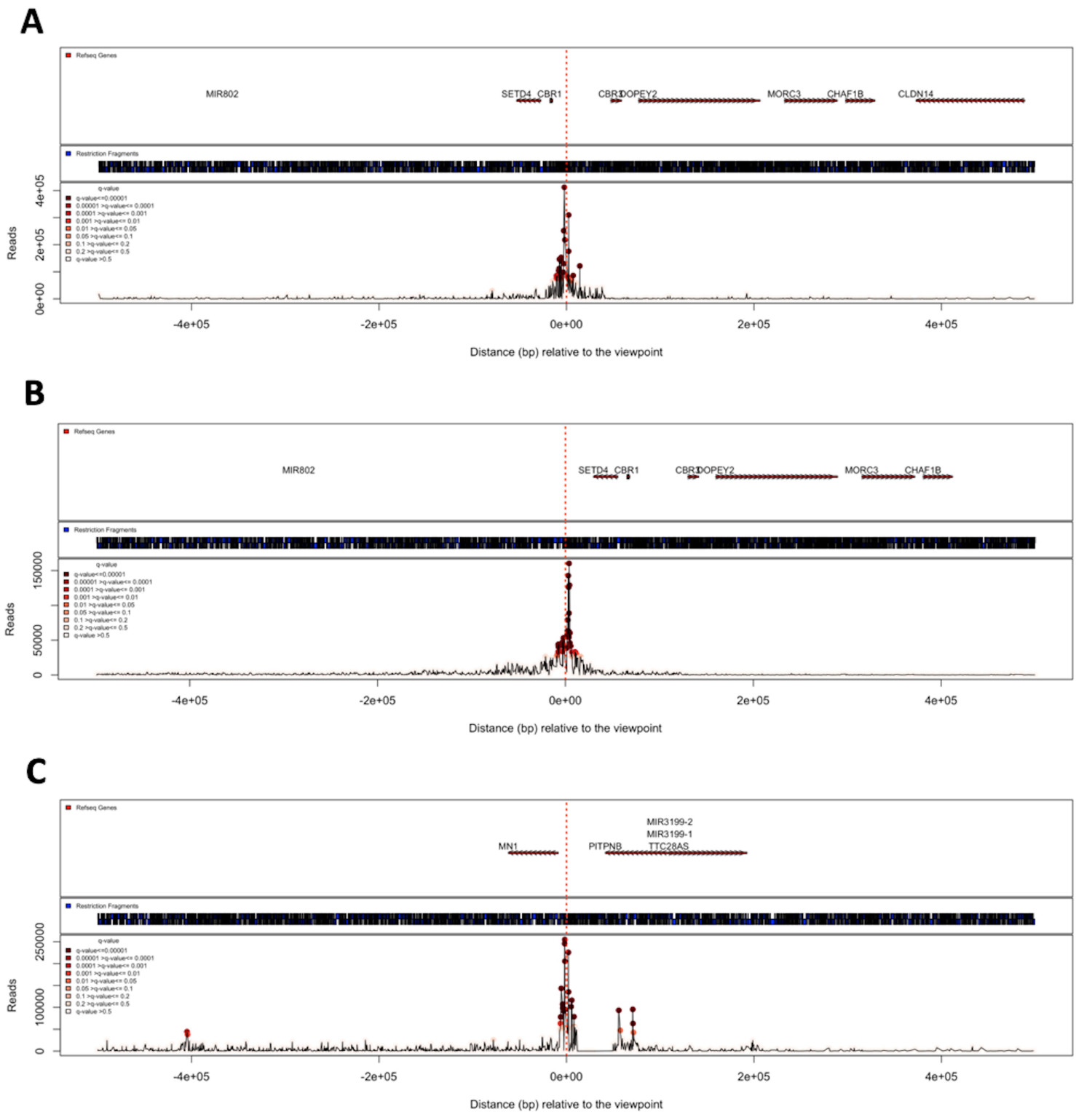


Fig S5


A
Refseq Genes
MIR802
SETD4
CBR1
CBR3
DOPEY2
MORC3
CHAF1B
CLDN14
Restriction Fragments
q-value
q-value<=0.00001
0.00001 >q-value<= 0.0001
0.0001 >q-value<= 0.001
0.001 >q-value<= 0.01
0.01 >q-value<= 0.05
0.05 >q-value<= 0.1
0.1 >q-value<= 0.2
0.2 >q-value<= 0.5
q-value >0.5
Reads
0e+00
2e+05
4e+05
-4e+05
-2e+05
0e+00
2e+05
4e+05
Distance (bp) relative to the viewpoint
B
Refseq Genes
MIR802
SETD4
CBR1
CBR3
DOPEY2
MORC3
CHAF1B
Restriction Fragments
q-value
q-value<=0.00001
0.00001 >q-value<= 0.0001
0.0001 >q-value<= 0.001
0.001 >q-value<= 0.01
0.01 >q-value<= 0.05
0.05 >q-value<= 0.1
0.1 >q-value<= 0.2
0.2 >q-value<= 0.5
q-value >0.5
Reads
0
50000
150000
-4e+05
-2e+05
0e+00
2e+05
4e+05
Distance (bp) relative to the viewpoint
C
Refseq Genes
MIR3199-2
MIR3199-1
MN1
PITPNB
TTC28AS
Restriction Fragments
q-value
q-value<=0.00001
0.00001 >q-value<= 0.0001
0.0001 >q-value<= 0.001
0.001 >q-value<= 0.01
0.01 >q-value<= 0.05
0.05 >q-value<= 0.1
0.1 >q-value<= 0.2
0.2 >q-value<= 0.5
q-value >0.5
Reads
0
100000
250000
-4e+05
-2e+05
0e+00
2e+05
4e+05
Distance (bp) relative to the viewpoint