

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

3

4 Sergiy Kyryachenko¹, PhD*, Adrien Georges¹, PhD*, Mengyao Yu¹, PhD*, Takiy Barrassou¹,
5 PhD, Patrick Bruneval¹, MD, PhD, Tony Rubio^{2,3}, MSc, Judith Gronwald^{2,3}, Hassina Baraki^{3,4},
6 MD, Ingo Kutschka^{3,4}, MD, Russell. A Norris^{5,6}, PhD, Niels Voigt^{2,3}, MD, Nabila Bouatia-Naji¹,
7 PhD

8

9 1. Université de Paris, PARCC, Inserm, Paris, France

10 2. Institute of Pharmacology and Toxicology, University Medical Center Göttingen, Germany

11 3. DZHK (German Center for Cardiovascular Research), Partner Site Göttingen, Germany

12 4. Department of Thoracic and Cardiovascular Surgery, University Medical Center, Göttingen,
13 Germany

14 5. Department of Regenerative Medicine and Cell Biology, Medical University of South
15 Carolina, Charleston, SC, USA;

16 6. Department of Medicine, Medical University of South Carolina, Charleston, SC, USA

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

18

19 **Short title:** Regulatory profiles of mitral valves

20

21 Corresponding author: Nabila Bouatia-Naji, email: nabila.bouatia-naji@inserm.fr, address: Paris

22 Cardiovascular Research Center Inserm U970, 56 rue Leblanc, F-75015 Paris, France

23

24 **Total word count:** 8,141

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

26

27 **ABSTRACT**

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

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

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

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

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

54

55 **Introduction**

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

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

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

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

95

96 **Materials and Methods**

97 **Sample acquisition and patient consent**

98 Mitral valve tissue was acquired at University Medical Center Göttingen, Georg-August
99 University of Göttingen, Germany with written and informed consent through Institutional
100 Review Board (IRB) protocol N° 4/11/18. Freshly obtained mitral valve tissue from 11 patients
101 who underwent mitral valve replacement surgery (Supplementary Table S1) was snap frozen in
102 liquid nitrogen and stored at -80°C.

103

104 **Cell culture**

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

111

112 **Isolation of nuclei from frozen mitral valves for ATAC-Seq**

113 Frozen mitral valve tissues were mechanically dissociated in liquid nitrogen using mortar and
114 pestle. Grinded tissue was transferred into 15 ml Falcon tubes containing 5 ml of cold
115 homogenization buffer (HB): 10 mM Tris-HCl pH 7.6, 10 mM NaCl, 3 mM MgCl₂, 0.1% NP40,
116 0.1% Tween-20, 0.01% Digitonin, 250mM Sucrose, and EDTA-free complete protease inhibitors
117 (Roche, Pleasanton, CA). The homogenate was filtered through 400 µm cell strainer

118 (pluriStrainer, Leipzig, Germany). Then, the isolation of nuclei was performed as described¹⁸
119 with modifications. Specifically, after Dounce homogenization, the tissue homogenate was
120 pelleted at 500g for 10 minutes at 4°C. Tissue homogenate (700-800 µl) was transferred to a 14
121 ml round bottom tube (Greiner Bio-One, Frickenhausen, Germany) and mixed with an equal
122 volume of 50% iodixanol in HB to obtain a final 25% iodixanol. 1200 µl of 30% iodixanol in HB
123 was layered underneath the 25% mixture, and then 1200 µl of 40% iodixanol in HB was layered
124 below the 30% iodixanol. Nuclei were enriched at 3,000g for 20 minutes at 4°C in a swinging-
125 bucket centrifuge. Nuclei that accumulated at the 30%/40% interface were collected into a 1.5 ml
126 Lo-Bind Eppendorf tube, stained with Trypan Blue, and manually counted under microscope.

127

128 **Library preparation for ATAC-Seq**

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

139

140 **Sequencing of ATAC-Seq libraries and peak calling**

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

156

157 **Analysis of chromatin accessibility profiles**

158 To perform sample correlation and principal component analysis, a common list of enriched
159 regions was generated using bedtools multiple intersect (Galaxy Version 2.29.0) in “cluster”
160 mode, and average read coverage on these regions was computed using deepTools
161 multiBamsummary (Galaxy Version 3.3.2.0.0). deepTools plotCoverage and plotPCA functions
162 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 ENCF052BXS, ENCF172RFM, ENCF208DZK, ENCF222SST, ENCF222WPT,
184 ENCF283CXG, ENCF332RJQ, ENCF385NNX, ENCF401EEA, ENCF491RWJ,

185 ENCF617BNF, ENCF670ZUM, ENCF686LUV, ENCF824SPW, ENCF946SRJ,
186 ENCF964UAX, and ENCF984MJS) and RegulomeDB score.²⁹

187

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

189 Using genotype data from our previous study¹², we performed an updated association analysis on
190 chromosome 2 and chromosome 17 using newly imputed SNPs obtained from the HRC³⁰ as
191 imputation reference panel through Michigan imputation server³¹. We used results from our
192 GWAS meta-analysis as described previously.¹² We used the GCTA-COJO function to perform
193 conditional analyses to identify secondary or causal signals on loci.³² The COJO analysis used
194 the MVP-France case control study as reference panel to calculate linkage disequilibrium (LD)
195 between SNPs. To avoid collinearity issues, SNPs in high LD ($r^2 > 0.9$) with the tested SNP were
196 excluded before analyses.

197

198 **Circular chromatin conformation capture followed by high-throughput sequencing (4C- 199 Seq)**

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

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

221

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

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

230 (Berthold Technologies, Bad Wildbad, Germany) was used to measure the luciferase activity
231 with the Dual-Luciferase Reporter Assay System (Promega, Madison, WI). All experiments were
232 performed according to the instructions recommended by the manufacturer. The luciferase
233 activity data were obtained from at least eight replicate wells, and the experiments were repeated
234 3 times. Two-tailed Student's t-test was used to compare if the difference was significant and the
235 significance threshold was set at $P < 0.05$.

236

237

238

239 **Results**

240 **Mapping the chromatin accessibility profiles of human mitral valve tissue**

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

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

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

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

273

274 **Chromatin accessibility profiles present regulatory features specific to mitral valve**

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

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

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

297 Mitral valve ATAC-Seq data provide potential regulatory elements that may participate to the
298 pathogenesis of MVP. To investigate further this idea, we tested the enrichment for associated
299 SNPs from existing MVP GWAS results among different ATAC-Seq datasets.¹² We used the
300 Genomic Regulatory Elements and GWAS Overlap algorithm (GREGOR)²⁸ that compares the
301 number of trait-associated SNPs overlapping predefined genomic regions (here, ATAC-Seq
302 peaks) with randomized SNP matched by regions (1Mb window) and frequency. Figure 4A
303 shows that MVP-associated SNPs ($p < 10^{-5}$) were significantly enriched ($p < 0.05$) in ATAC-Seq
304 peaks from all mitral valve samples whereas significant enrichment was found for only one of
305 the other samples (Heart LV-2). Maximal enrichment was seen for MV1830, for which ATAC-
306 Seq peaks overlapped over 4 times more MVP-associated SNPs than expected ($p=6 \times 10^{-6}$, Figure
307 4A). Comparatively, ATAC-Seq peaks from fibroblasts samples did not significantly overlap
308 with MVP-associated SNPs, despite the fact that most peaks from mitral valves were also
309 observed in fibroblast ATAC-Seq peaks.

310

311 **Identification of potential causal variants at MVP-associated loci**

312 Given the enrichment for regulatory elements near genes involved in myxomatous valve biology,
313 we sought to use the mitral valve ATAC-Seq datasets to narrow down a list of potential causal
314 SNPs at MVP-associated loci. We selected associated SNPs ($p < 0.001$) from the six loci
315 reaching genome-wide significance in our previous study¹² and filtered for colocalization with
316 mitral valve ATAC-Seq peaks (Table 1). Between three and 20 SNPs were identified at each
317 locus, representing a list of potential causal SNPs. We annotated these SNPs using RegulomeDB

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

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

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

341 18 of these SNPs also overlap H3K27ac marks in at least one tissue, most of them in all the
342 tested tissues. Among these SNPs, rs2641440 was ranked by RegulomeDB as the most likely
343 regulatory SNP (RegulomeDB score=1f), and it belongs to a strongly active region in all assayed
344 tissues (Table 1). We also selected rs9899330 for *in vitro* tests, given its RegulomeDB high score
345 2b (Table 1), and low linkage disequilibrium with rs2641440 ($r^2=0.21$, CEU 1000 genomes),
346 suggesting it may belong to a secondary association signal. We then tested the effect of
347 rs2641440 and rs9899330 genotypes on the regulatory activity of the surrounding region in HDF.
348 The vector with the alternative C allele for rs2641440 had significantly higher luciferase activity
349 compared with the construct containing the reference G allele ($p=0.0017$) (Supplementary Figure
350 S1A). Both constructs that contained alternative or reference allele of rs9899330 significantly
351 enhanced the activity of luciferase compared with the control construct, but had no allelic effect
352 on the regulatory activity of surrounding region (Supplementary Figure S1B). Interestingly,
353 conditional analyses both on rs2641440 and rs9899330 at *SMG6/SRR* locus (Supplementary
354 Table S4) caused a decrease of the association signal (Supplementary Figure S1C-F), supporting
355 the existence of two independent signals at this locus.

356

357 **Identification of potential target genes at MVP-associated loci**

358 To get further insight into the biology of MVP, we sought to determine potential target genes at
359 each of the six confirmed MVP-associated loci¹². First, we examined potential causal SNPs
360 (Table 1) for eQTL association with genes in any human tissue using the Genotype-Tissue
361 Expression (GTEx) portal (Table 2). Then, we performed circular chromatin conformation
362 capture (4C) followed by high-throughput sequencing (4C-Seq) in human dermal fibroblasts to
363 verify the potential target genes. We selected one to four viewpoints at each locus focusing on

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

386

387

388 **Discussion**

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

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

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

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

420 Our approach to fine-map GWAS loci using chromatin accessibility data derived from mitral
421 valve tissue, combined with the visualization of long-range genomic interaction using 4C-Seq
422 method, allowed us to identify potential causal variants and to narrow down the list of target
423 genes at risk loci, especially on chromosomes 2 and 17. At the *IGFBP5/TNSI* MVP-associated
424 locus we identified rs6723013 as a functional enhancer *in vitro* and demonstrated regulatory
425 allelic effect. Presence of the minor T allele (T=0.24, frequency in TOPMed panel) can modulate
426 the expression of potential target genes by altering transcription factor binding onto DNA.
427 Several TFs, such as *OVOL2* and *NFE2L1*, are predicted to bind more to the major G allele.
428 *OVOL2* suppresses epithelial-to-mesenchymal transition and thus maintains epithelial lineages,
429 and *NFE2L1* regulates expression of target genes in response to various stresses. Further studies
430 are needed to explore their involvement in gene regulation by this enhancer region. Potential
431 target genes at this locus were *IGFBP5*, *IGFBP2*, and *TNPI1* that are located upstream from the

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

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

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

484

485 **ACKNOWLEDGMENTS**

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

487

488 **FUNDING SOURCES**

489 This study was supported by a Ph.D. scholarship from the China Scholarship Council to MY, French
490 Agency of Research (ANR-16-CE17-0015-02) to NB-N. AG, SK and NB-N are supported by a European
491 Research Council grant (ERC-Stg-ROSALIND-716628). This work was supported by grants from the
492 Deutsche Forschungsgemeinschaft (DFG) to NV (VO 1568/3-1, IRTG1816, and SFB1002
493 project A13), from the Else-Kröner-Fresenius Foundation to NV (EKFS 2016_A20) and from
494 the German Center for Cardiovascular Research to NV (DZHK GOE MD3 and SE181). This
495 work was supported in part by grants from the National Institutes of Health (GM103444 to RAN;
496 R01HL131546, P20GM103444, and R01HL127692 to RAN; and American Heart Association
497 (19TPA34850095 to RAN, 17CSA33590067 to RAN).

498

499 **DISCLOSURES**

500 None

501

502 REFERENCES

- 503 1. Huk DJ, Austin BF, Horne TE, Hinton RB, Ray WC, Heistad DD, Lincoln J. Valve endothelial
504 cell-derived tgfbeta1 signaling promotes nuclear localization of sox9 in interstitial cells
505 associated with attenuated calcification. *Arteriosclerosis, thrombosis, and vascular biology*.
506 2016;36:328-338
- 507 2. Menon V, Lincoln J. The genetic regulation of aortic valve development and calcific disease.
508 *Frontiers in cardiovascular medicine*. 2018;5:162
- 509 3. Shapero K, Wylie-Sears J, Levine RA, Mayer JE, Jr., Bischoff J. Reciprocal interactions between
510 mitral valve endothelial and interstitial cells reduce endothelial-to-mesenchymal transition and
511 myofibroblastic activation. *Journal of molecular and cellular cardiology*. 2015;80:175-185
- 512 4. Hinton RB, Yutzey KE. Heart valve structure and function in development and disease. *Annual*
513 *review of physiology*. 2011;73:29-46
- 514 5. Rabkin-Aikawa E, Farber M, Aikawa M, Schoen FJ. Dynamic and reversible changes of
515 interstitial cell phenotype during remodeling of cardiac valves. *The Journal of heart valve*
516 *disease*. 2004;13:841-847
- 517 6. Levine RA, Hagege AA, Judge DP, et al. Mitral valve disease--morphology and mechanisms.
518 *Nature reviews. Cardiology*. 2015;12:689-710
- 519 7. Hjortnaes J, Camci-Unal G, Hutcheson JD, Jung SM, Schoen FJ, Kluin J, Aikawa E,
520 Khademhosseini A. Directing valvular interstitial cell myofibroblast-like differentiation in a
521 hybrid hydrogel platform. *Advanced healthcare materials*. 2015;4:121-130
- 522 8. Nishimura RA, McGoon MD. Perspectives on mitral-valve prolapse. *The New England journal of*
523 *medicine*. 1999;341:48-50
- 524 9. Avierinos JF, Gersh BJ, Melton LJ, 3rd, Bailey KR, Shub C, Nishimura RA, Tajik AJ, Enriquez-
525 Sarano M. Natural history of asymptomatic mitral valve prolapse in the community. *Circulation*.
526 2002;106:1355-1361
- 527 10. Spartalis M, Tzatzaki E, Spartalis E, Athanasiou A, Moris D, Damaskos C, Garmpis N, Voudris
528 V. Mitral valve prolapse: An underestimated cause of sudden cardiac death-a current review of
529 the literature. *Journal of thoracic disease*. 2017;9:5390-5398
- 530 11. Iung B, Vahanian A. Epidemiology of acquired valvular heart disease. *The Canadian journal of*
531 *cardiology*. 2014;30:962-970
- 532 12. Dina C, Bouatia-Naji N, Tucker N, et al. Genetic association analyses highlight biological
533 pathways underlying mitral valve prolapse. *Nature genetics*. 2015;47:1206-1211
- 534 13. Bryois J, Garrett ME, Song L, et al. Evaluation of chromatin accessibility in prefrontal cortex of
535 individuals with schizophrenia. *Nature communications*. 2018;9:3121
- 536 14. Kyono Y, Kitzman JO, Parker SCJ. Genomic annotation of disease-associated variants reveals
537 shared functional contexts. *Diabetologia*. 2019;62:735-743
- 538 15. Fagny M, Paulson JN, Kuijjer ML, Sonawane AR, Chen CY, Lopes-Ramos CM, Glass K,
539 Quackenbush J, Platig J. Exploring regulation in tissues with eqtl networks. *Proceedings of the*
540 *National Academy of Sciences of the United States of America*. 2017;114:E7841-E7850
- 541 16. Simonis M, Klous P, Splinter E, Moshkin Y, Willemsen R, de Wit E, van Steensel B, de Laat W.
542 Nuclear organization of active and inactive chromatin domains uncovered by chromosome
543 conformation capture-on-chip (4c). *Nature genetics*. 2006;38:1348-1354
- 544 17. Vermunt MW, Zhang D, Blobel GA. The interdependence of gene-regulatory elements and the 3d
545 genome. *The Journal of cell biology*. 2019;218:12-26
- 546 18. Corces MR, Trevino AE, Hamilton EG, et al. An improved atac-seq protocol reduces background
547 and enables interrogation of frozen tissues. *Nature methods*. 2017;14:959-962
- 548 19. Buenrostro JD, Wu B, Chang HY, Greenleaf WJ. Atac-seq: A method for assaying chromatin
549 accessibility genome-wide. *Current protocols in molecular biology*. 2015;109:21 29 21-21 29 29

- 550 20. Afgan E, Baker D, Batut B, et al. The galaxy platform for accessible, reproducible and
551 collaborative biomedical analyses: 2018 update. *Nucleic acids research*. 2018;46:W537-W544
- 552 21. Amemiya HM, Kundaje A, Boyle AP. The encode blacklist: Identification of problematic regions
553 of the genome. *Scientific reports*. 2019;9:9354
- 554 22. Yu G, Wang LG, He QY. Chipseeker: An r/bioconductor package for chip peak annotation,
555 comparison and visualization. *Bioinformatics*. 2015;31:2382-2383
- 556 23. Ross-Innes CS, Stark R, Teschendorff AE, et al. Differential oestrogen receptor binding is
557 associated with clinical outcome in breast cancer. *Nature*. 2012;481:389-393
- 558 24. Yu G, Wang LG, Han Y, He QY. Clusterprofiler: An r package for comparing biological themes
559 among gene clusters. *Omics : a journal of integrative biology*. 2012;16:284-287
- 560 25. Supek F, Bosnjak M, Skunca N, Smuc T. Revigo summarizes and visualizes long lists of gene
561 ontology terms. *PLoS one*. 2011;6:e21800
- 562 26. Machanick P, Bailey TL. Meme-chip: Motif analysis of large DNA datasets. *Bioinformatics*.
563 2011;27:1696-1697
- 564 27. Freese NH, Norris DC, Loraine AE. Integrated genome browser: Visual analytics platform for
565 genomics. *Bioinformatics*. 2016;32:2089-2095
- 566 28. Schmidt EM, Zhang J, Zhou W, Chen J, Mohlke KL, Chen YE, Willer CJ. Gregor: Evaluating
567 global enrichment of trait-associated variants in epigenomic features using a systematic, data-
568 driven approach. *Bioinformatics*. 2015;31:2601-2606
- 569 29. Boyle AP, Hong EL, Hariharan M, et al. Annotation of functional variation in personal genomes
570 using regulomedb. *Genome research*. 2012;22:1790-1797
- 571 30. Vergara C, Parker MM, Franco L, Cho MH, Valencia-Duarte AV, Beaty TH, Duggal P.
572 Genotype imputation performance of three reference panels using african ancestry individuals.
573 *Human genetics*. 2018;137:281-292
- 574 31. Das S, Forer L, Schönherr S, et al. Next-generation genotype imputation service and methods.
575 *Nature genetics*. 2016;48:1284-1287
- 576 32. Yang J, Ferreira T, Morris AP, et al. Conditional and joint multiple-snp analysis of gwas
577 summary statistics identifies additional variants influencing complex traits. *Nature genetics*.
578 2012;44:369-375, s361-363
- 579 33. Matelot M, Noordermeer D. Determination of high-resolution 3d chromatin organization using
580 circular chromosome conformation capture (4c-seq). *Methods in molecular biology*.
581 2016;1480:223-241
- 582 34. Thongjuea S, Stadhouders R, Grosveld FG, Soler E, Lenhard B. R3cseq: An r/bioconductor
583 package for the discovery of long-range genomic interactions from chromosome conformation
584 capture and next-generation sequencing data. *Nucleic acids research*. 2013;41:e132
- 585 35. Sheffield NC, Furey TS. Identifying and characterizing regulatory sequences in the human
586 genome with chromatin accessibility assays. *Genes*. 2012;3:651-670
- 587 36. Nemer M, Horb ME. The klf family of transcriptional regulators in cardiomyocyte proliferation
588 and differentiation. *Cell cycle*. 2007;6:117-121
- 589 37. Kelsey L, Flenniken AM, Qu D, et al. Enu-induced mutation in the DNA-binding domain of klf3
590 reveals important roles for klf3 in cardiovascular development and function in mice. *PLoS*
591 *genetics*. 2013;9:e1003612
- 592 38. Desjardins CA, Naya FJ. The function of the mef2 family of transcription factors in cardiac
593 development, cardiogenomics, and direct reprogramming. *Journal of cardiovascular development*
594 *and disease*. 2016;3
- 595 39. Vasan RS, Glazer NL, Felix JF, et al. Genetic variants associated with cardiac structure and
596 function: A meta-analysis and replication of genome-wide association data. *Jama*. 2009;302:168-
597 178
- 598 40. Schunkert H, König IR, Kathiresan S, et al. Large-scale association analysis identifies 13 new
599 susceptibility loci for coronary artery disease. *Nature genetics*. 2011;43:333-338

- 600 41. Wild PS, Felix JF, Schillert A, et al. Large-scale genome-wide analysis identifies genetic variants
601 associated with cardiac structure and function. *The Journal of clinical investigation*.
602 2017;127:1798-1812
- 603 42. Tyan L, Foell JD, Vincent KP, et al. Long qt syndrome caveolin-3 mutations differentially
604 modulate kv 4 and cav 1.2 channels to contribute to action potential prolongation. *The Journal of*
605 *physiology*. 2019;597:1531-1551
- 606 43. Wallace E, Howard L, Liu M, O'Brien T, Ward D, Shen S, Prendiville T. Long qt syndrome:
607 Genetics and future perspective. *Pediatric cardiology*. 2019;40:1419-1430
- 608 44. Frieden LA, Townsend TA, Vaught DB, Delaughter DM, Hwang Y, Barnett JV, Chen J.
609 Regulation of heart valve morphogenesis by eph receptor ligand, ephrin-a1. *Developmental*
610 *dynamics : an official publication of the American Association of Anatomists*. 2010;239:3226-
611 3234
- 612 45. Fang M, Alfieri CM, Hulin A, Conway SJ, Yutzey KE. Loss of beta-catenin promotes
613 chondrogenic differentiation of aortic valve interstitial cells. *Arteriosclerosis, thrombosis, and*
614 *vascular biology*. 2014;34:2601-2608
- 615 46. Wu B, Wang Y, Lui W, Langworthy M, Tompkins KL, Hatzopoulos AK, Baldwin HS, Zhou B.
616 Nfatc1 coordinates valve endocardial cell lineage development required for heart valve formation.
617 *Circulation research*. 2011;109:183-192
- 618 47. Zanotti S, Canalis E. Notch suppresses nuclear factor of activated t cells (nfat) transactivation and
619 nfatc1 expression in chondrocytes. *Endocrinology*. 2013;154:762-772
- 620 48. Gunawan F, Gentile A, Gauvrit S, Stainier DYR, Bensimon-Brito A. Nfatc1 promotes interstitial
621 cell formation during cardiac valve development in zebrafish. *Circulation research*.
622 2020;126:968-984
- 623 49. Johnson EN, Lee YM, Sander TL, Rabkin E, Schoen FJ, Kaushal S, Bischoff J. Nfatc1 mediates
624 vascular endothelial growth factor-induced proliferation of human pulmonary valve endothelial
625 cells. *The Journal of biological chemistry*. 2003;278:1686-1692
- 626 50. Greenhouse DG, Murphy A, Mignatti P, Zavadil J, Galloway AC, Balsam LB. Mitral valve
627 prolapse is associated with altered extracellular matrix gene expression patterns. *Gene*.
628 2016;586:56-61
- 629 51. Prall OW, Menon MK, Solloway MJ, et al. An nkx2-5/bmp2/smad1 negative feedback loop
630 controls heart progenitor specification and proliferation. *Cell*. 2007;128:947-959
- 631 52. Hulin A, Hortells L, Gomez-Stallons MV, et al. Maturation of heart valve cell populations during
632 postnatal remodeling. *Development*. 2019;146
- 633 53. Robson MI, Ringel AR, Mundlos S. Regulatory landscaping: How enhancer-promoter
634 communication is sculpted in 3d. *Molecular cell*. 2019;74:1110-1122
- 635 54. Schoenfelder S, Fraser P. Long-range enhancer-promoter contacts in gene expression control.
636 *Nature reviews. Genetics*. 2019;20:437-455
- 637 55. Fishilevich S, Nudel R, Rappaport N, et al. Genehancer: Genome-wide integration of enhancers
638 and target genes in genecards. *Database : the journal of biological databases and curation*.
639 2017;2017
- 640 56. Xia M, Liu Y, Figueroa DJ, et al. Characterization and localization of a human serine racemase.
641 *Brain research. Molecular brain research*. 2004;125:96-104
- 642 57. Soliman H, Paylor B, Scott RW, et al. Pathogenic potential of hic1-expressing cardiac stromal
643 progenitors. *Cell stem cell*. 2020;26:205-220 e208
- 644 58. Uhlen M, Fagerberg L, Hallstrom BM, et al. Proteomics. Tissue-based map of the human
645 proteome. *Science*. 2015;347:1260419
- 646 59. Valenta T, Lukas J, Doubravska L, Fafilek B, Korinek V. Hic1 attenuates wnt signaling by
647 recruitment of tcf-4 and beta-catenin to the nuclear bodies. *The EMBO journal*. 2006;25:2326-
648 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

656 **Table 1. Potential causal SNPs at the MVP-associated loci.** Overlaps with indicated
657 epigenomic marks are indicated by grey cells. P values were reported for the association with
658 MVP obtained in the discovery meta-analysis of GWAS¹² 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	<i>LMCD1</i>	Heart - Atrial Appendage	rs165177	C	8×10^{-7}	0.27	Y
				<i>LMCD1-AS1</i>	Pituitary	rs165172	C	4×10^{-6}	0.47	Y
Chr14	rs17767392	2.3×10^{-8}	7	<i>SIPAIL1-AS1</i>	Esophagus - Muscularis	rs17767410	C	1×10^{-12}	-0.43	Y
				<i>SIPAIL1</i>	Esophagus - Muscularis	rs17767410	C	6×10^{-12}	-0.27	Y
				<i>PCNX1</i>	Skin - Sun Exposed (Lower leg)	rs7155259	C	1×10^{-6}	0.17	N
				<i>RP6-91H8.5</i>	Brain - Cortex	rs7155259	C	1×10^{-5}	0.36	N
				<i>COX16</i>	Whole Blood	rs56201715	G	9×10^{-5}	0.1	N
Chr17	rs216205	1.5×10^{-8}	20	<i>SRR</i>	Muscle - Skeletal	rs3744270	G	1×10^{-106}	0.8	Y
				<i>TSR1</i>	Skin - Sun Exposed (Lower leg)	rs4477758	T	3×10^{-26}	-0.22	Y
				<i>SGSM2</i>	Skin - Sun Exposed (Lower leg)	rs4477758	T	6×10^{-26}	-0.35	Y
				<i>SGSM2-AS1</i>	Skin - Not Sun Exposed (Suprapubic)	rs408067	C	1×10^{-15}	0.25	Y
				<i>HNRNPA1P16</i>	Nerve - Tibial	rs3744270	G	8×10^{-13}	-0.34	N
				<i>HIC1</i>	Cells - Cultured fibroblasts	rs3744270	G	1×10^{-7}	-0.13	Y
				<i>DPH1</i>	Pancreas	rs9899330	A	8×10^{-7}	-0.27	Y
				<i>SMG6</i>	Testis	rs12450028	C	9×10^{-7}	-0.14	Y
				<i>RTN4RL1</i>	Heart - Left Ventricle	rs3815469	G	3×10^{-5}	-0.25	N
Chr21	rs62229266	1.2×10^{-8}	4	<i>LINC01436</i>	Artery - Tibial	rs2255858	T	1×10^{-78}	0.93	Y
				<i>RPL23AP3</i>	Cells - Cultured fibroblasts	rs2255858	T	2×10^{-35}	0.65	Y
				<i>SETD4</i>	Cells - Cultured fibroblasts	rs2255858	T	5×10^{-19}	0.26	Y
				<i>CBR1</i>	Nerve - Tibial	rs2255858	T	5×10^{-10}	0.19	N
				<i>AP000688.29</i>	Artery - Tibial	rs2255858	T	8×10^{-9}	0.29	N
Chr22	rs11705555	1.4×10^{-8}	7	<i>MNI</i>	Whole Blood	rs73164492	C	3×10^{-5}	-0.21	Y
				<i>TTC28</i>	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.

666 **FIGURE LEGENDS**

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

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

672

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

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

683

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

686 **A:** Volcano plot representing negative logarithm of enrichment false discovery rate (FDR) on the
687 Y axis and logarithm of enrichment in heart samples over mitral valve samples on the X axis.
688 Each dot represents an ATAC-Seq peak in the mitral valve or heart sample. Pink dots represent
689 significantly enriched regions in mitral valve or heart ATAC-Seq experiments ($FDR \leq 0.05$). **B-C:**
690 Bubble graphs represent GOBP terms enriched among genes at proximity (distance to
691 $TSS \leq 10\text{kb}$) of heart- (**B**) and mitral valve- (**C**) specific peaks. X and Y axes are in arbitrary
692 semantic coordinates.²⁵ Bubble size represents number of semantically similar terms aggregated
693 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

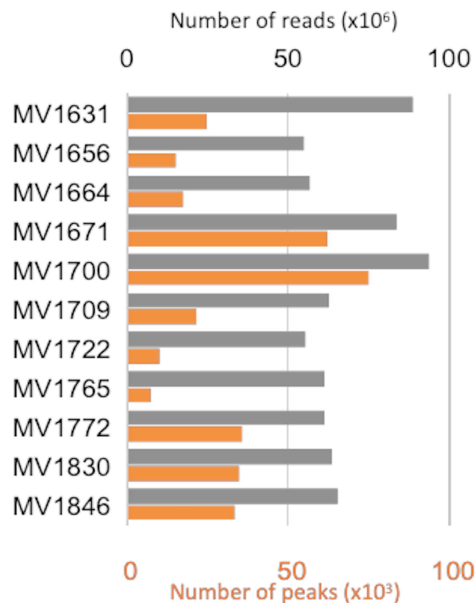
712 **Figure 5. 4C-Seq analysis of the *SMG6/SRR* MVP-associated locus.**

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.

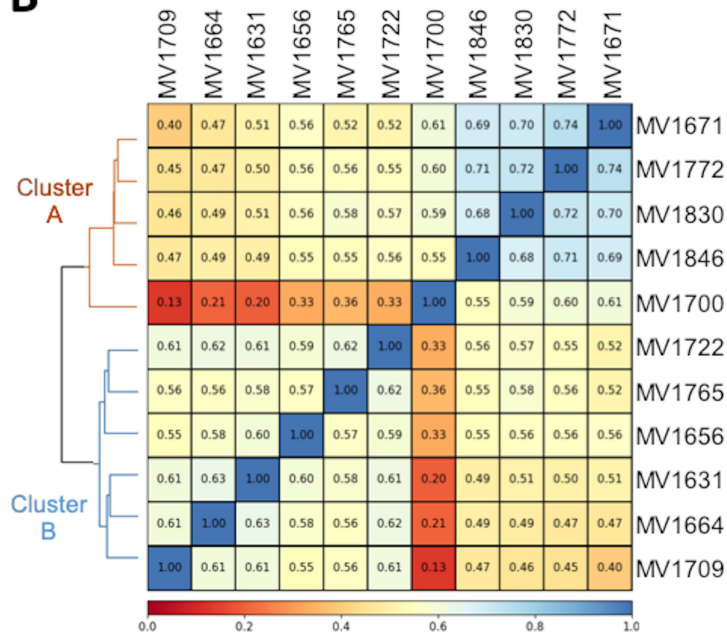
716

Fig 1

A



B



C

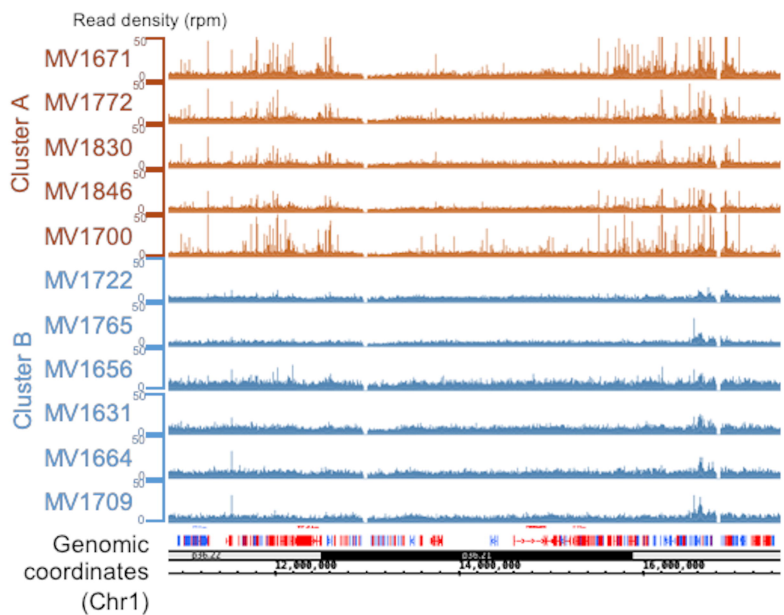
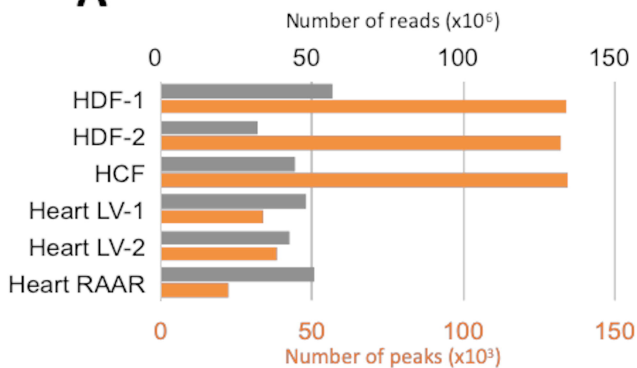
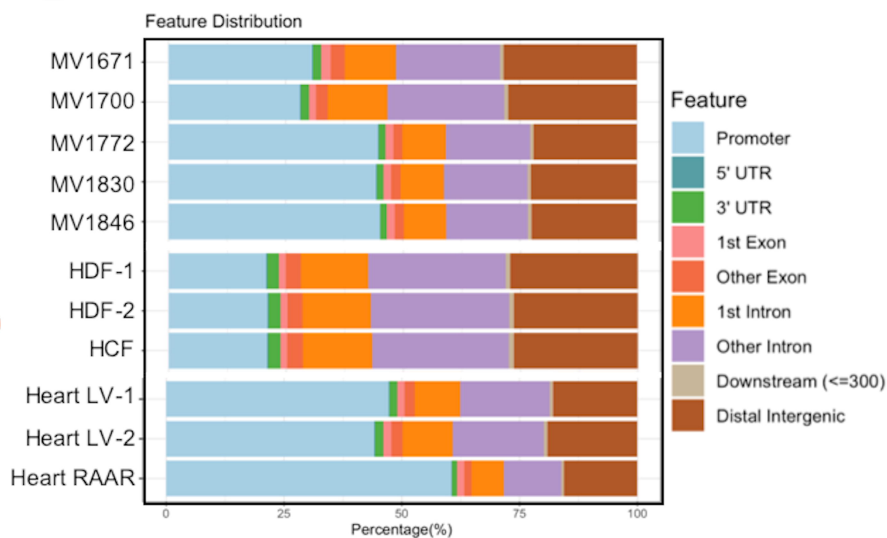


Fig 2

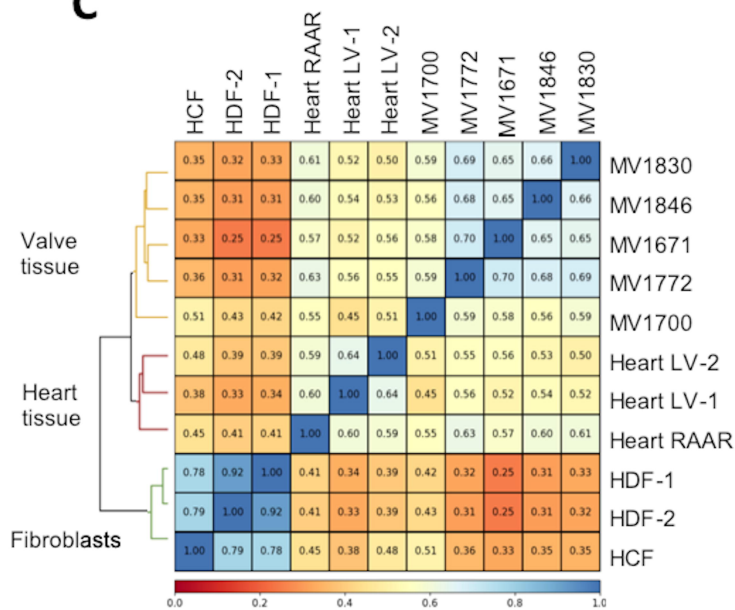
A



B



C



D

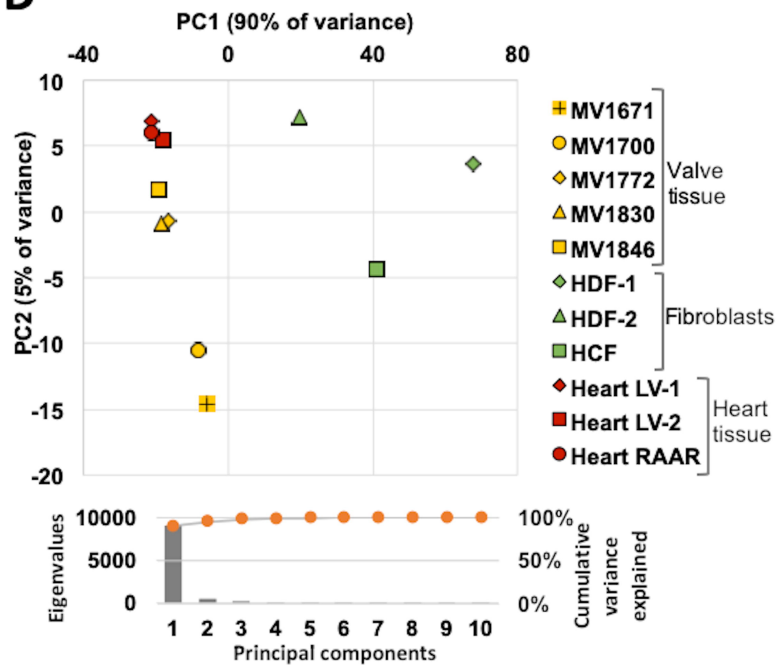
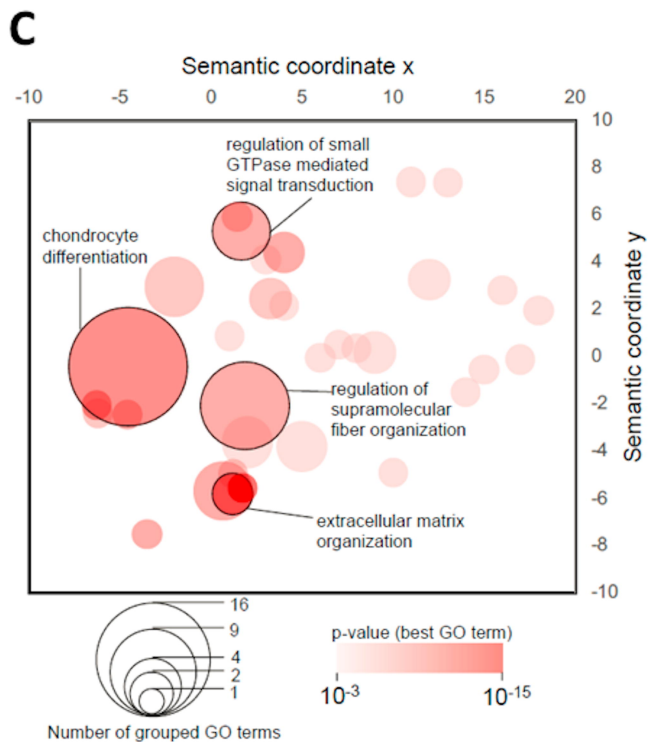
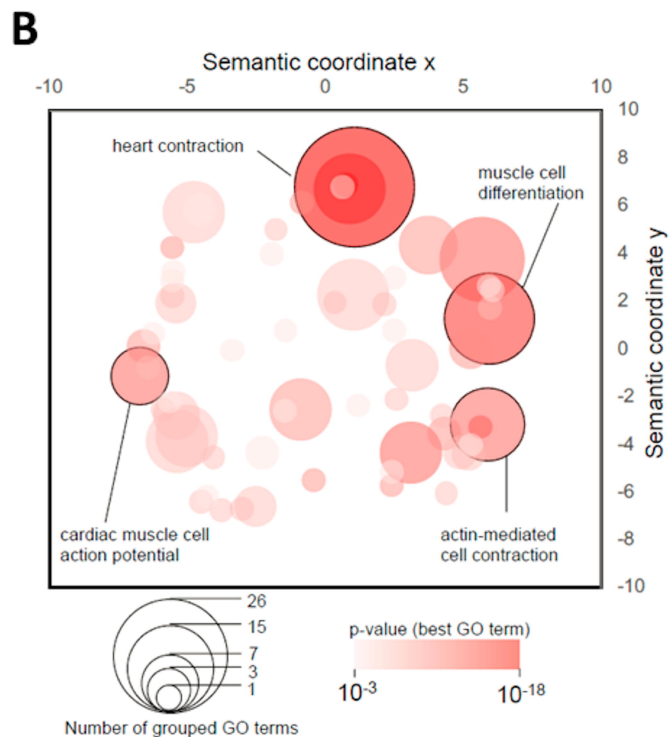
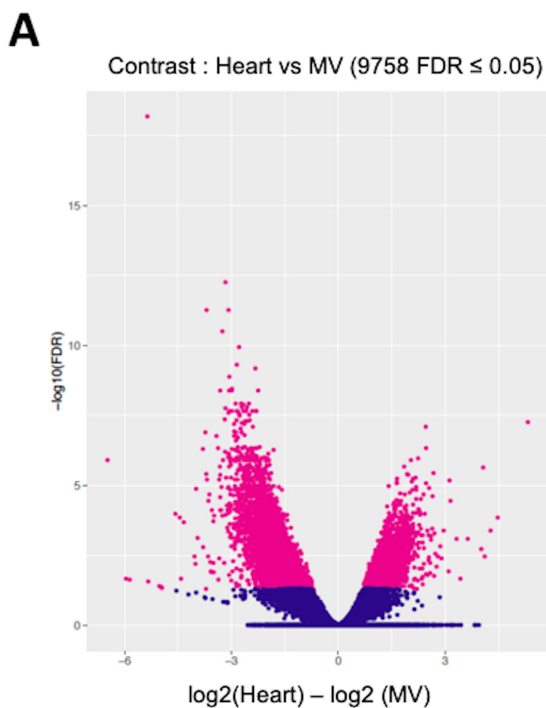


Fig 3



D

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

E

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

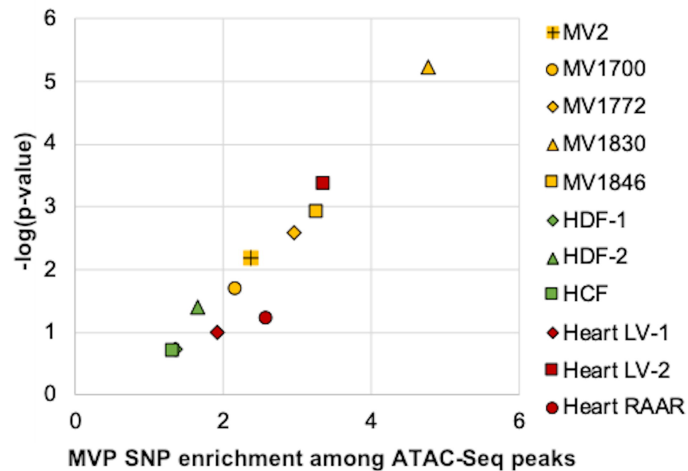
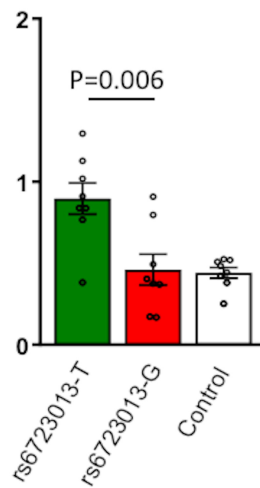
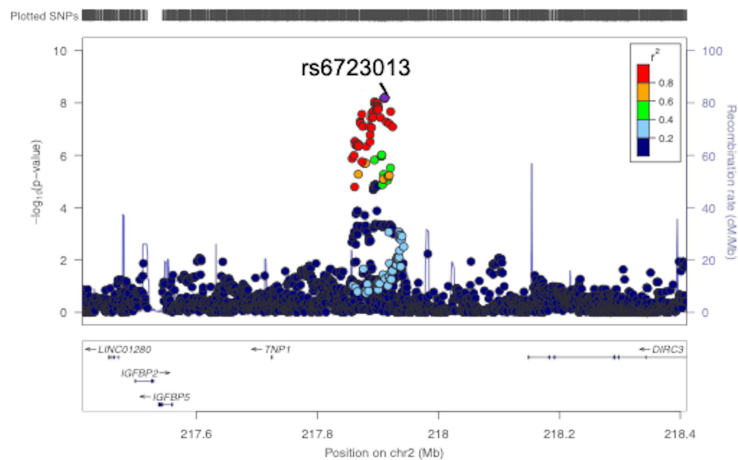
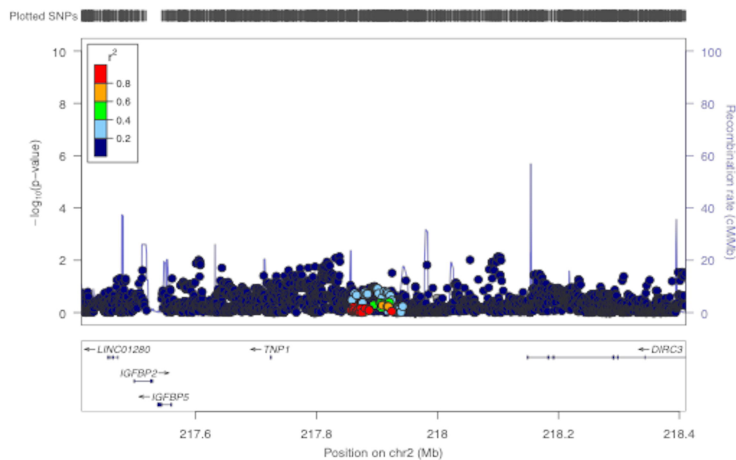
Fig 4**A****B****C****D**

Fig 5

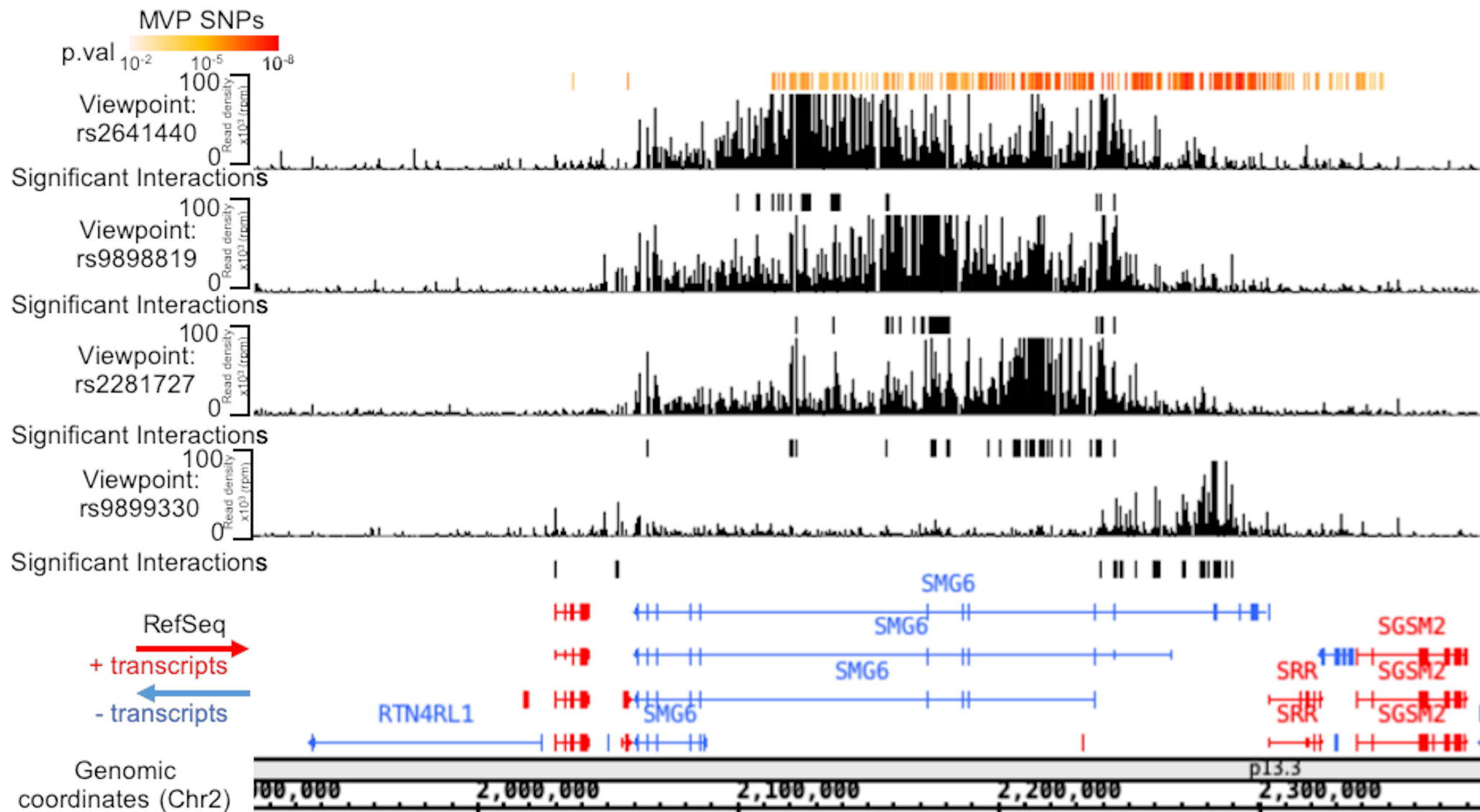


Fig S1

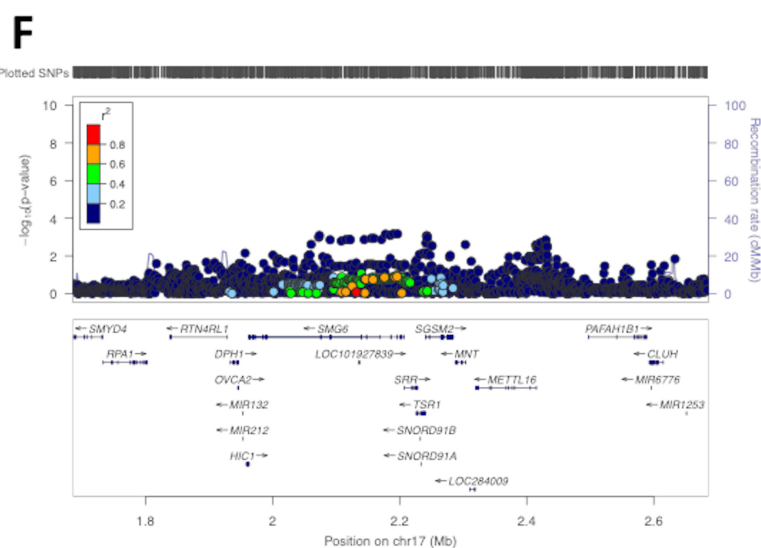
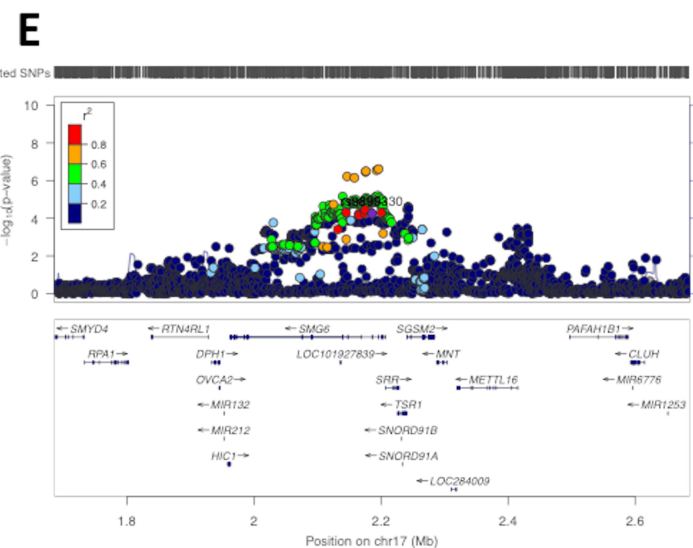
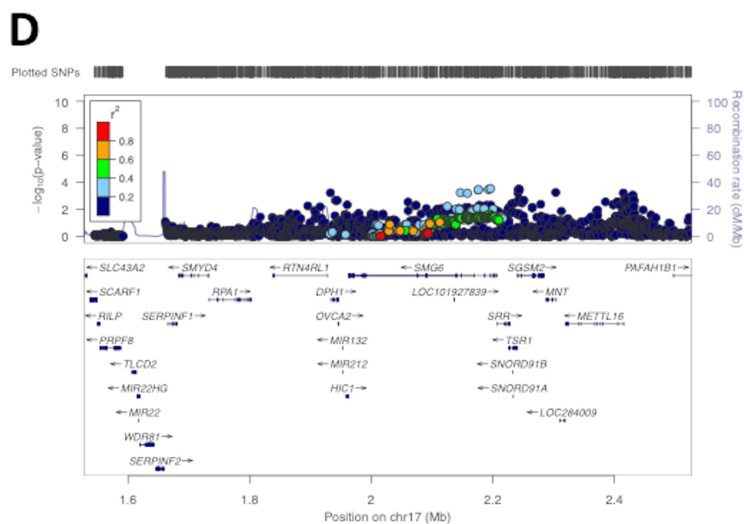
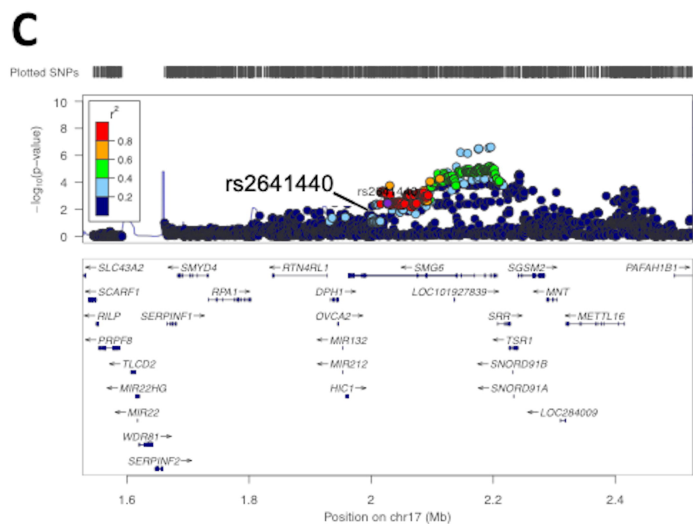
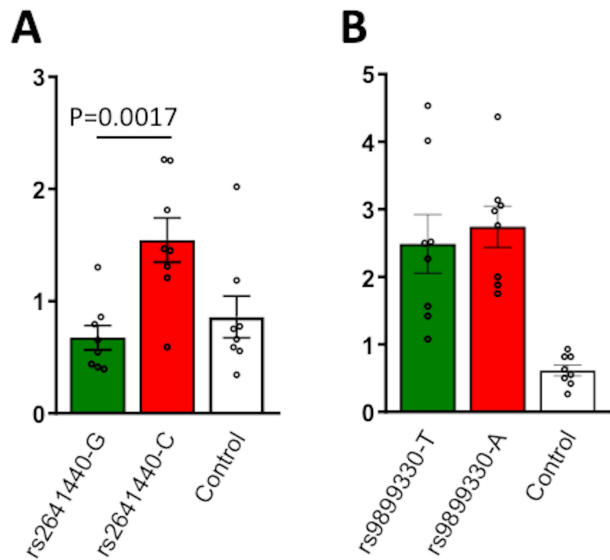
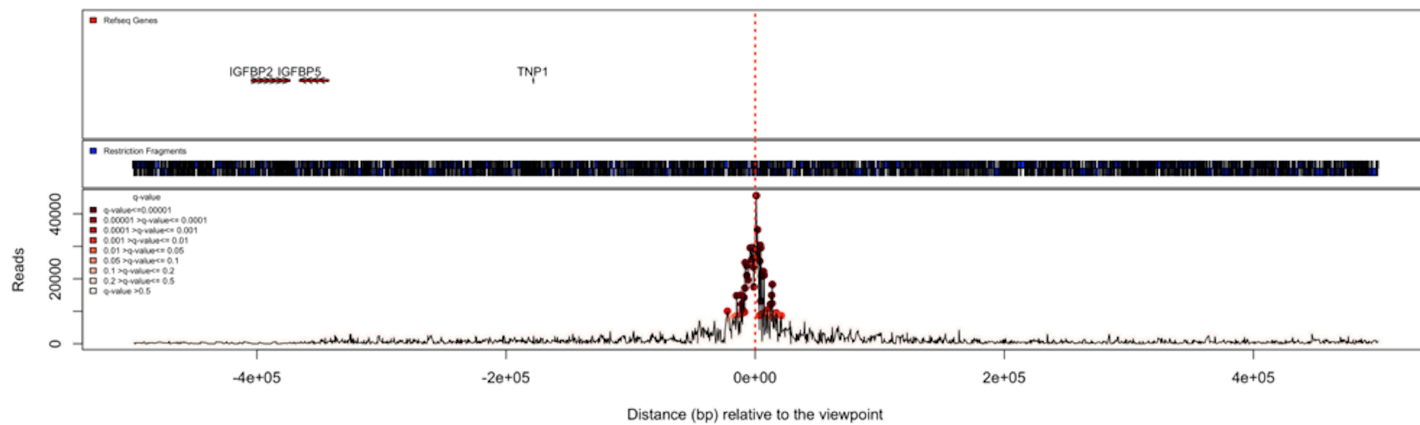
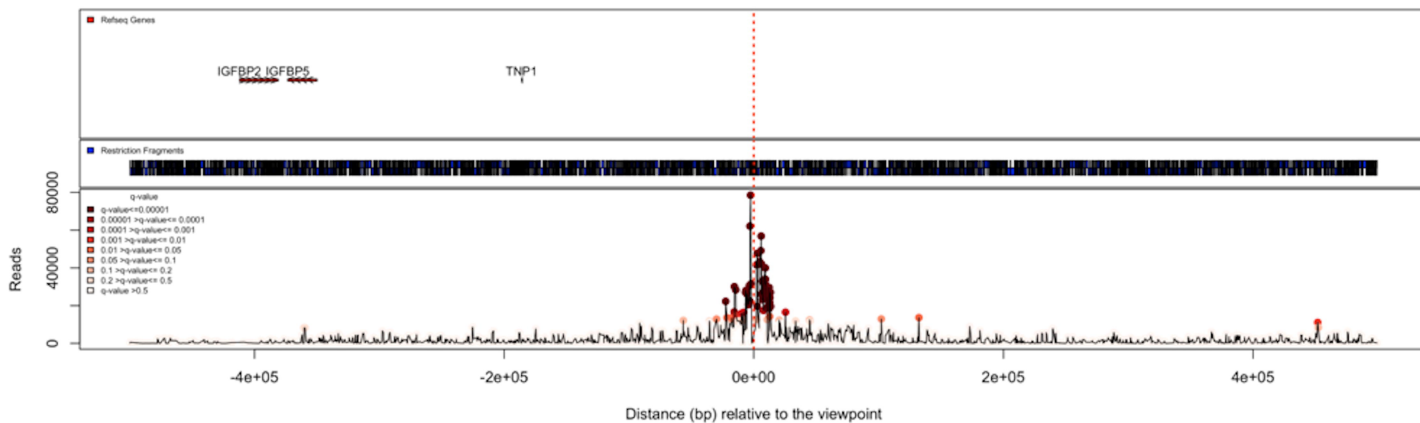


Fig S2

A



B



C

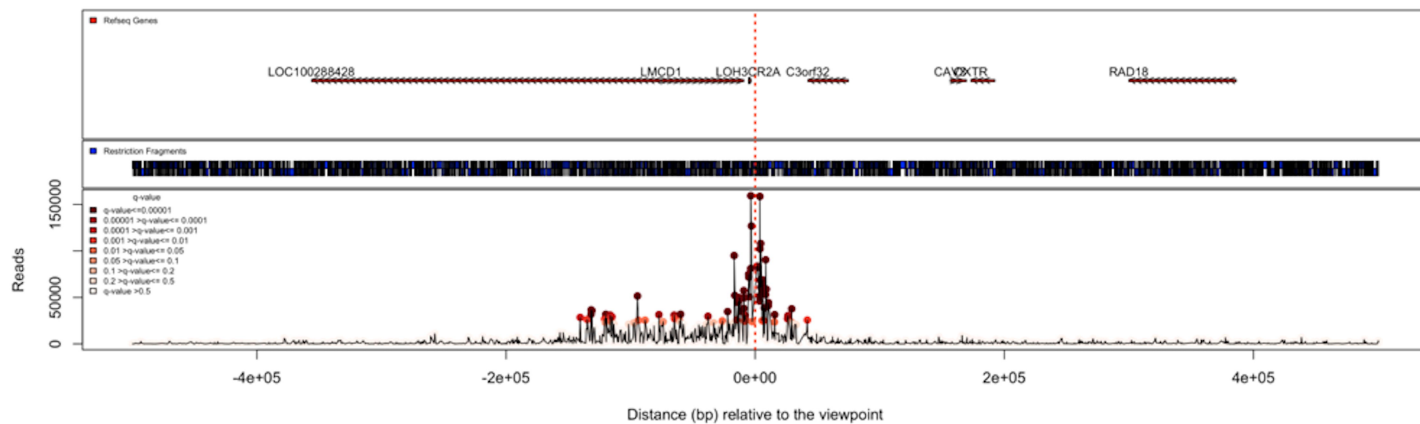
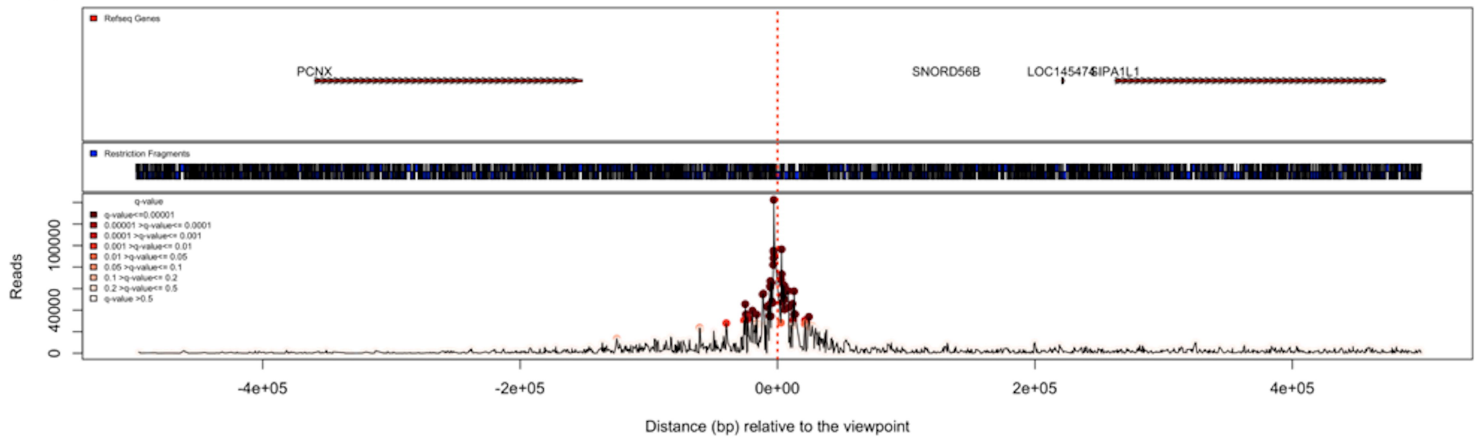
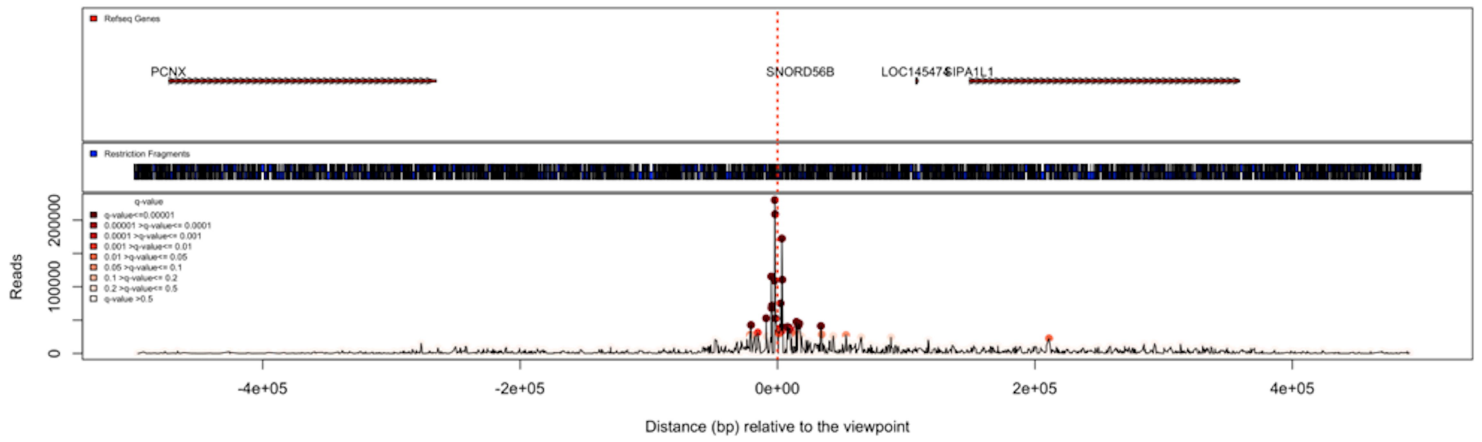


Fig S3

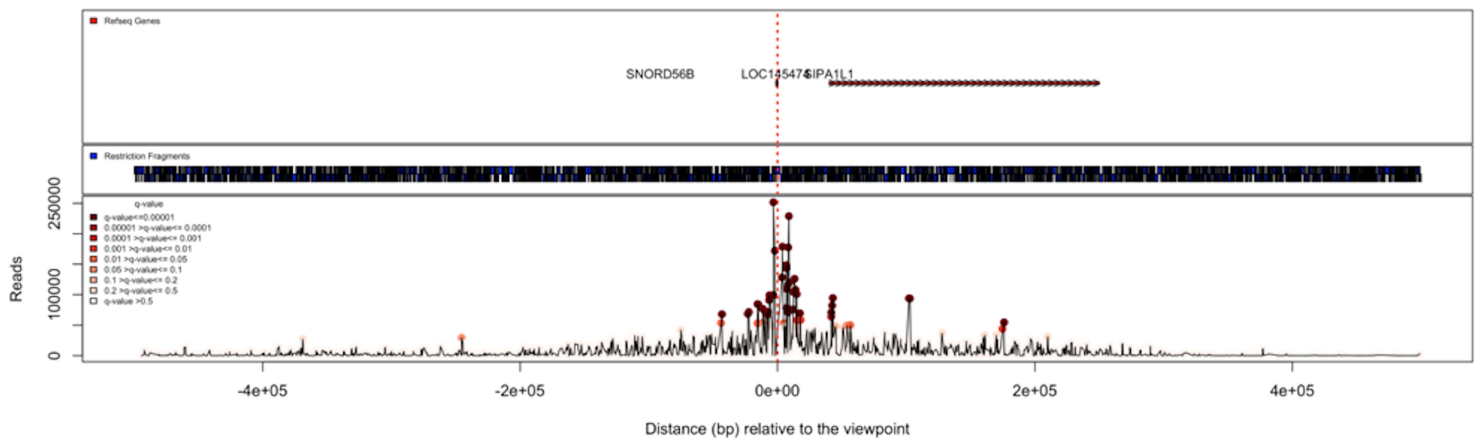
A



B



C



D

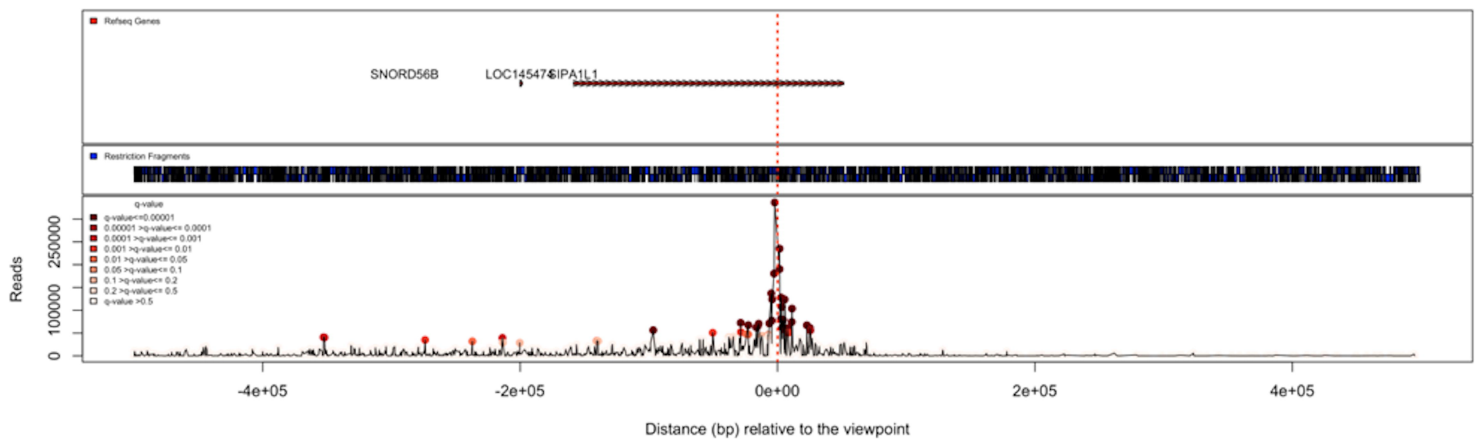
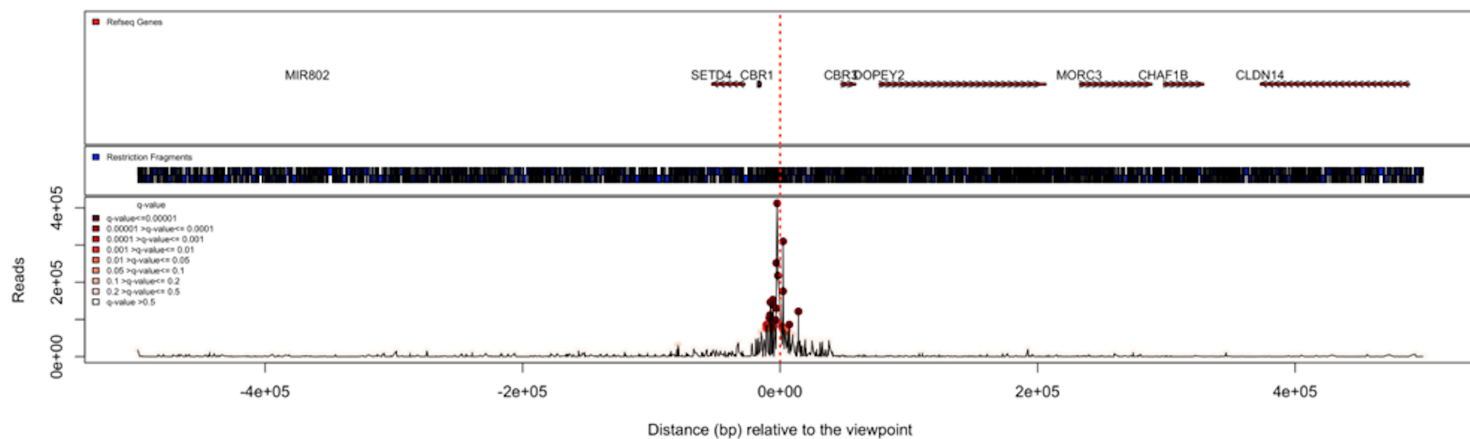
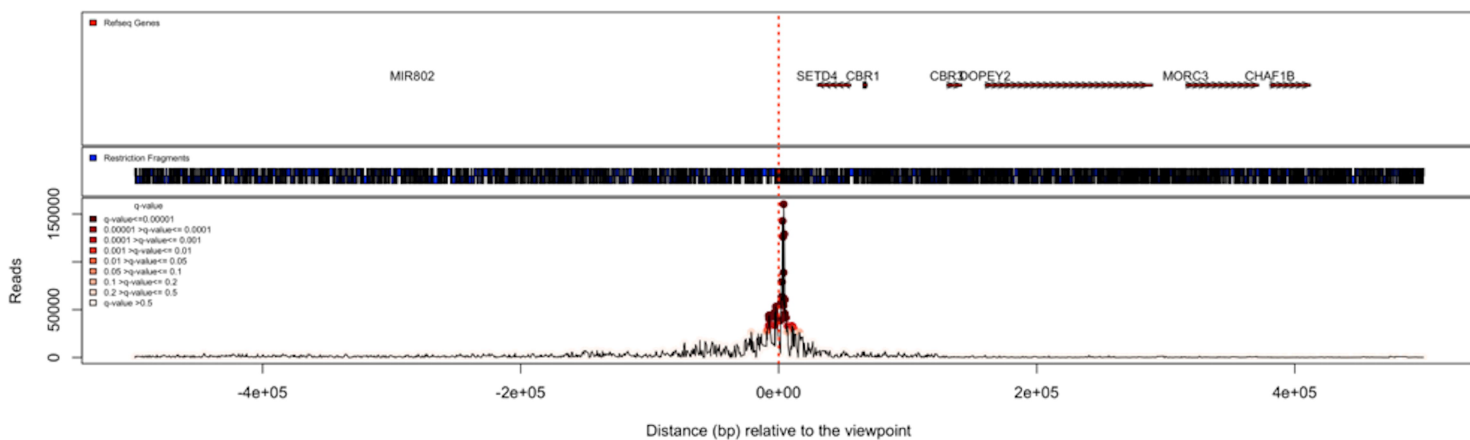


Fig S5

A



B



C

