A promoter interaction map for cardiovascular disease genetics

Lindsey E. Montefiori[1], Débora R. Sobreira[1], Noboru J. Sakabe[1], Ivy Aneas[1], Amelia C. Joslin[1], Grace T. Hansen[1], Grazyna Bozek[1], Ivan P. Moskowitz[1,2], Elizabeth M. McNally[3], Marcelo A. Nóbrega[1]

[1]Department of Human Genetics, The University of Chicago, Chicago IL; [2]Department of Pediatrics and Pathology, The University of Chicago, Chicago IL; [3]Center for Genetic Medicine, Northwestern University Feinberg School of Medicine, Chicago IL

Address for Correspondence:

Marcelo A. Nóbrega	Lindsey Montefiori
Department of Human Genetics	Department of Human Genetics
920 E. 58th St.	920 E 58th St.
CLSC Room 515	CLSC Room 515
773-702-1808	773-702-1808
Email: nobrega@uchicago.edu	Email: lem@uchicago.edu

24   **Abstract**

25   Over 500 genetic loci have been associated with risk of cardiovascular diseases (CVDs),

26   however most loci are located in gene-distal non-coding regions and their target genes are not known.

27   Here, we generated high-resolution promoter capture Hi-C (PCHi-C) maps in human induced

28   pluripotent stem cells (iPSCs) and iPSC-derived cardiomyocytes (CMs) to provide a resource for

29   identifying and prioritizing the functional targets of CVD associations. We validate these maps by

30   demonstrating that promoters preferentially contact distal sequences enriched for tissue-specific

31   transcription factor motifs and are enriched for chromatin marks that correlate with dynamic changes in

32   gene expression. Using the CM PCHi-C map, we linked 1,999 CVD-associated SNPs to 347 target

33   genes. Remarkably, more than 90% of SNP-target gene interactions did not involve the nearest gene,

34   while 40% of SNPs interacted with at least two genes, demonstrating the importance of considering

35   long-range chromatin interactions when interpreting functional targets of disease loci.

36

37

38

39

40

41

42

43

44

45

46   **Introduction**

47   A major goal in human genetics research is to understand genetic contributions to complex

48   diseases, specifically the molecular mechanisms by which common DNA variants impact disease

49   etiology. Most genome-wide association studies (GWAS) implicate non-coding variants that are far from

50   genes, complicating interpretation of their mode of action and correct identification of the target gene

51   (Maurano et al. 2012). Mounting evidence suggests that disease variants disrupt the function of *cis*-

52   acting regulatory elements, such as enhancers, which in turn affects expression of the specific gene or

53   genes that are functional targets of these elements (Wright et al. 2010; Musunuru et al. 2010; Cowper-

54   Sal·lari et al. 2012; Smemo et al. 2014; Claussnitzer et al. 2015). However, because *cis*-acting

55   regulatory elements can be located kilobases (kb) away from their target gene(s), identifying the true

56   functional targets of regulatory elements remains challenging (Smemo et al. 2014).

57   Chromosome conformation capture techniques such as Hi-C (Lieberman-Aiden et al. 2009)

58   enable the genome-wide mapping of long-range chromatin contacts and therefore represent a

59   promising strategy to identify distal gene targets of disease-associated genetic variants. Recently, Hi-C

60   maps have been generated in numerous human cell types including embryonic stem cells and early

61   embryonic lineages (Dixon et al. 2012, 2015), immune cells (Rao et al. 2014), fibroblasts (Jin et al.

62   2013) and other primary tissue types (Schmitt et al. 2016). However, despite the increasing abundance

63   of Hi-C maps, most datasets are of limited resolution (>40 kb) and do not precisely identify the genomic

64   regions in contact with gene promoters.

65   More recently, promoter capture Hi-C (PCHi-C) was developed which greatly increases the

66   power to detect interactions involving promoter sequences (Schoenfelder et al. 2015; Mifsud et al.

67   2015). PCHi-C in different cell types identified thousands of enhancer-promoter contacts and revealed

68   extensive differences in promoter architecture between cell types and throughout differentiation

69   (Schoenfelder et al. 2015; Mifsud et al. 2015; Javierre et al. 2016; Freire-Pritchett et al. 2017; Rubin et

70   al. 2017; Siersbæk et al. 2017). These studies collectively demonstrated that genome architecture

3

71   reflects cell identity, suggesting that disease-relevant cell types are critical for successful interrogation

72   of the gene regulatory mechanisms of disease loci.

73   In support of this notion, several recent studies utilized high-resolution promoter interaction

74   maps to identify tissue-specific target genes of GWAS associations. Javierre *et al*. generated promoter

75   capture Hi-C data in 17 primary human blood cell types and identified 2,604 potentially causal genes for

76   immune- and blood-related disorders, including many genes with unannotated roles in those diseases

77   (Javierre et al. 2016). Similarly, Mumbach *et al*. interrogated GWAS SNPs associated with autoimmune

78   diseases using HiChIP where they identified ~10,000 promoter-enhancer interactions that linked

79   several hundred SNPs to target genes, most of which were not the nearest gene (Mumbach et al.

80   2017). Importantly, both studies reported cell-type specificity of SNP-target gene interactions.

81   Cardiovascular diseases, including cardiac arrhythmia, heart failure, and myocardial infarction,

82   continue to be the leading cause of death world-wide. Over 50 GWAS have been conducted for these

83   specific cardiovascular phenotypes alone, with more than 500 loci implicated in cardiovascular disease

84   risk (NHGRI GWAS catalog, https://www.ebi.ac.uk/gwas/), most of which map to non-coding genomic

85   regions. To begin to dissect the molecular mechanisms by which genetic variants contribute to CVD

86   risk, a comprehensive gene regulatory map of human cardiac cells is required. Here, we present high

87   resolution promoter interaction maps of human iPSCs and iPSC-derived cardiomyocytes (CMs). Using

88   PCHi-C, we identified hundreds of thousands of promoter interactions in each cell type. We

89   demonstrate the physiological relevance of these datasets by functionally interrogating the relationship

90   between gene expression and long-range promoter interactions, and demonstrate the utility of long-

91   range chromatin interaction data to resolve the functional targets of disease-associated loci.

92   **Results**

93   **iPSC-derived cardiomyocytes provide an effective model to study the architecture of CVD**

94   **genetics**

4

95   We used iPSC-derived CMs (Burridge et al. 2014) as a model to study cardiovascular gene

96   regulation and disease genetics. The CMs generated in this study were 86-94% pure based on cardiac

97   Troponin T protein expression and exhibited spontaneous, uniform beating (Figure 1—figure

98   supplement 1A, Supplementary Mov. 1). To demonstrate that iPSCs and CMs recapitulate

99   transcriptional and epigenetic profiles of matched primary cells, we conducted RNA-seq and ChIP-seq

100   for the active enhancer mark H3K27ac in both cell types and compared these data with similar cell

101   types from the Epigenome Roadmap Project (Roadmap Epigenomics Consortium et al. 2015). RNA-

102   seq profiles of iPSCs clustered tightly with H1 embryonic stem cells, whereas CMs clustered with both

103   left ventricle (LV) and fetal heart (FH) profiles (Figure 1—figure supplement 1B). Furthermore, we

104   observed that matched cell types exhibited three-fold greater overlap in the number of promoter-distal

105   H3K27ac ChIP-seq peaks than non-matched cell types (Figure 1—figure supplement 1C,D), indicating

106   that both iPSCs and CMs recapitulate tissue-specific epigenetic states of human stem cells and primary

107   cardiomyocytes, respectively.

108   To further validate our system, we analyzed differentially expressed genes between iPSCs and

109   CMs. Among the top 10% of over-expressed genes in CMs were genes directly related to cardiac

110   function including essential cardiac transcription factors (*GATA4*, *MEIS1*, *TBX5*, and *TBX20*) and

111   differentiation products (*TNNT2*, *MYH7B*, *MYL7*, *ACTN2*, *NPPA*, *HCN4*, and *RYR2*) (fold-change > 1.5,

112   P_{adj} < 0.05, Figure 1—figure supplement 2A-C). Gene Ontology (GO) enrichment analysis for genes

113   over-expressed in CMs relative to iPSCs further confirmed the cardiac-specific phenotypes of these

114   cells with top terms relating to the development of the cardiac conduction system and cardiac muscle

115   cell contraction (Figure 1—figure supplement 2D).

116   **Promoter-capture Hi-C identifies distal regulatory elements in iPSCs and CMs**

117   To comprehensively map long-range regulatory elements in iPSCs and CMs, we performed *in-*

118   *situ* Hi-C (Rao et al. 2014) in triplicate iPSC-CM differentiations; importantly, we used the 4-cutter

119   restriction enzyme MboI which generates ligation fragments with an average size of 422 bp, enabling

120   enhancer-level resolution of promoter contacts. We enriched iPSC and CM *in situ* Hi-C libraries for

121   promoter interactions through hybridization with a set of 77,476 biotinylated RNA probes ("baits")

122   targeting 22,600 human RefSeq protein-coding promoters (see Methods) and sequenced each library

123   to an average depth of ~413 million (M) paired-end reads. After removing duplicates and read-pairs that

124   did not map to a bait, we obtained an average of 31M and 41M read-pairs per replicate for iPSC and

125   CM, respectively. We used CHiCAGO (Cairns et al. 2016), a computational pipeline which accounts for

126   bias from the sequence capture, to identify significant interactions and further filtered for those

127   significant in at least two out of three replicates (see Methods). Finally, we exclusively focused on

128   interactions that were separated by a distance of at least 10 kb. This criterion addresses the high

129   frequency of close-proximity ligation evets in Hi-C data, which are difficult to distinguish as random

130   Brownian contacts or functional chromatin interactions (Cairns et al. 2016). In total, we identified

131   350,062 promoter interactions in iPSCs and 401,098 in CMs. A large proportion (~55%) of interactions

132   were shared between the two cell types, indicating that even at high resolution many long-range

133   interactions are stable across cell types (Figure 1A). Approximately 20% of all interactions were

134   between two promoters, demonstrating the high connectivity between genes and supporting the

135   recently suggested role of promoters acting as regulatory inputs for distal genes (Dao et al. 2017; Diao

136   et al. 2017) (Figure 1B). Most interactions were promoter-distal, with a median of ~170 kb between the

137   promoter and the distal-interacting region (Figure 1C).
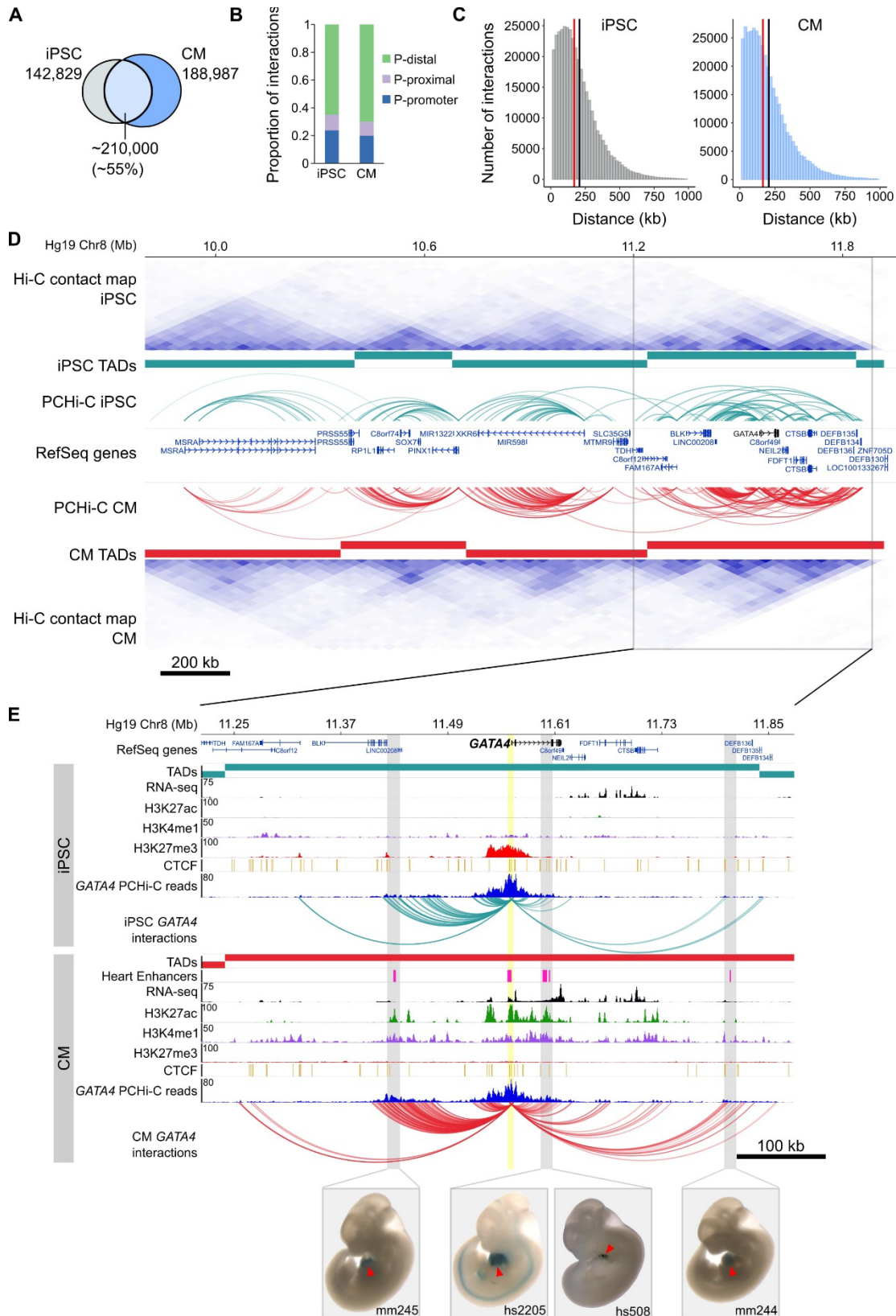
138   To compare the PCHi-C maps with known features of genome organization, we sequenced our

139   pre-capture Hi-C libraries to an average depth of 665M reads per cell type and identified topologically

140   associating domains (TADs) with TopDom (see Methods). TADs are organizational units of

141   chromosomes defined by <1 megabase (Mb) genomic blocks that exhibit high self-interacting

142   frequencies with a very low interaction frequency across TAD boundaries (Dixon et al. 2012; Nora et al.

143   2012). Notably, this organization is thought to constrain the activity of *cis*-regulatory elements to target

144   genes with the same TAD, as disruption of TAD boundaries has been shown to lead to aberrant

6

145 activation of genes in neighboring TADs (Nora et al. 2012; Lupiáñez et al. 2015; Franke et al. 2016;

146 Symmons et al. 2016; Tsujimura et al. 2015). We found that the majority of PCHi-C interactions

147 occurred within TADs (73% and 77% in iPSCs and CMs, respectively; Figure 1D and Figure 1—figure

148 supplement 3A). TAD-crossing interactions ("inter-TAD") contained proportionally more promoter-

149 promoter interactions than intra-TAD interactions, and were more likely to overlap promoter-distal

150 CTCF sites; however, they were similarly enriched for looping to distal H3K27ac sites, a mark of active

151 chromatin (Figure 1—figure supplement 3B-D). Inter-TAD interactions had slightly lower CHiCAGO

152 scores, reflecting a lower number of reads supporting these interactions, and spanned greater genomic

153 distances than intra-TAD interactions (Figure 1—figure supplement 3E,F). Additionally, promoters with

154 inter-TAD interactions were preferentially located close to TAD boundaries (Figure 1—figure

155 supplement 3G) and had higher expression levels compared to promoters with intra-TAD interactions,

156 particularly in CMs (Figure 1—figure supplement 3H). These observations are consistent with previous

157 studies which demonstrated that highly expressed genes, specifically housekeeping genes, are

158 enriched at TAD boundaries (Dixon et al. 2012).

159   To illustrate the utility of high-resolution PCHi-C interaction maps, we highlight the *GATA4* locus

160 in Figures 1D and 1E. *GATA4* is a master regulator of heart development (Watt et al. 2004; Pikkarainen

161 et al. 2004) and the *GATA4* gene is located in a TAD structure that is relatively stable between iPSCs

162 and CMs (Figure 1D). However, PCHi-C identified increased interaction frequencies between the

163 *GATA4* promoter and several H3K27ac-marked regions, including four *in vivo* validated heart

164 enhancers from the Vista enhancer browser (Visel et al. 2007), specifically in CMs and coincident with

165 strong up-regulation of *GATA4* (Figure 1 – figure supplement 2C). Although TAD-based analyses help

166 define a gene's *cis*-regulatory landscape, high resolution promoter interaction data provides the

167 resolution necessary to precisely map enhancer-promoter interactions in the context of cellular

168 differentiation.

169

A iPSC 142,829, CM 188,987, ~210,000 (~55%)

B Proportion of interactions — P-distal, P-proximal, P-promoter (iPSC, CM)

C Number of interactions vs Distance (kb), iPSC and CM

D Hg19 Chr8 (Mb): Hi-C contact map iPSC, iPSC TADs, PCHi-C iPSC, RefSeq genes, PCHi-C CM, CM TADs, Hi-C contact map CM. 200 kb scale bar.

E Hg19 Chr8 (Mb): RefSeq genes, GATA4. iPSC tracks: TADs, RNA-seq, H3K27ac, H3K4me1, H3K27me3, CTCF, GATA4 PCHi-C reads, iPSC GATA4 interactions. CM tracks: TADs, Heart Enhancers, RNA-seq, H3K27ac, H3K4me1, H3K27me3, CTCF, GATA4 PCHi-C reads, CM GATA4 interactions. 100 kb scale bar. mm245, hs2205, hs508, mm244.

Figure 1. General features of promoter interactions. (A) Venn diagram displaying the number of cell type-specific and shared promoter interactions in each cell type. **(B)** Proportion of interactions in each distance category: promoter (P)-promoter (both interacting ends overlap a transcription start site (TSS)); P-proximal (non-promoter end overlaps captured region but not the TSS); P-distal (non-promoter end is outside of captured region). Note that all promoter interactions are separated by at least 10 kb. **(C)** Distribution of the distances spanning each interaction in iPSCs and CMs. The red line depicts the median (170 kb in iPSCs, 164 kb in CMs); the black line depicts the mean (208 kb in iPSCs, 206 kb in CMs). **(D)** A ~2 Mb region of chromosome 8 encompassing the *GATA4* gene is shown along with pre-capture (whole genome) Hi-C interaction maps at 40 kb resolution for iPSCs (top) and CMs (bottom). TADs called with TopDom are shown as colored bars (median TAD size = 640 kb in both cell types, mean TAD size = 742 kb in iPSCs and 743 kb in CMs) and significant PCHi-C interactions as colored arcs. **(E)** Zoomed-in view of the *GATA4* locus (promoter highlighted in yellow) in iPSCs (top) and CMs (bottom) along with corresponding RNA-seq data generated as part of this study, and ChIP-seq data for H3K27ac, H3K4me1, H3K27me3 and CTCF from the Epigenome Roadmap Project/ENCODE (H1 and left ventricle for iPSC and CM, respectively). Filtered *GATA4* read counts used by CHiCAGO are displayed in blue with the corresponding significant interactions shown as arcs. For clarity, only *GATA4* interactions are shown. Gray highlighted regions show interactions overlapping *in vivo* validated heart enhancers (pink boxes), with representative E11.5 embryos for each enhancer element (Visel et al. 2007). Red arrowhead points to the heart.

To validate the CM interaction map as a resource for cardiovascular disease genetics we next extensively characterized several important aspects of genetic architecture in CMs. We compared CMs with iPSCs in each analysis as a measure of cell-type specificity. These analyses serve as benchmarks that build on established features of genome organization and aid interpretations of the roles that long range interactions play in gene regulation.

Promoter interactions are enriched for tissue-specific transcription factor motifs

Distal enhancers activate target genes through DNA looping, a mechanism that enables distally bound transcription factors to contact the transcription machinery of target promoters (Pennacchio et al. 2013; Miele and Dekker 2008; Deng et al. 2012). To assess whether this feature of gene regulation was reflected in the iPSC and CM interactions, we conducted motif analysis using HOMER (Heinz et al. 2010) on the set of promoter-distal interacting sequences in each cell type. We initially focused on interactions for genes differentially expressed between iPSCs and CMs (fold-change > 1.5, P_{adj} < 0.05).

9

204 We identified CTCF as the most enriched motif in each case (Figures 2A,B), consistent with the known

205 role of this factor in mediating long-range genomic interactions (Phillips and Corces 2009; Phillips-

206 Cremins et al. 2013; Nora et al. 2017). Among the other top motifs, we identified the pluripotency factor

207 motifs OCT4-SOX2-TCF-NANOG (OSN) and SOX2 as preferentially enriched in distal sequences

208 looping to genes over-expressed in iPSCs (Figures 2A,C), whereas top motifs in distal sequences

209 looping to genes over-expressed in CMs included TBX20, ESRRB and MEIS1 (Figures 2B,C). TBX20

210 and MEIS1 transcription factors are important regulators of heart development and function (Cai et al.

211 2005; Sakabe et al. 2012; Mahmoud et al. 2013) and ESRRB was previously identified as a potential

212 binding partner of TBX20 in adult mouse cardiomyocytes (Shen et al. 2011). We also observed that

213 distal interactions unique to either iPSCs or CMs were similarly enriched for tissue-specific transcription

214 factor motifs (Figure 2D). In line with a recent report that AP-1 contributes to dynamic loop formation

215 during macrophage development (Phanstiel et al. 2017), both iPSC- and CM-specific interactions were

216 enriched for AP-1 motifs (Figure 2D), suggesting that AP-1 transcription factors may represent a

217 previously unrecognized genome organizing complex.
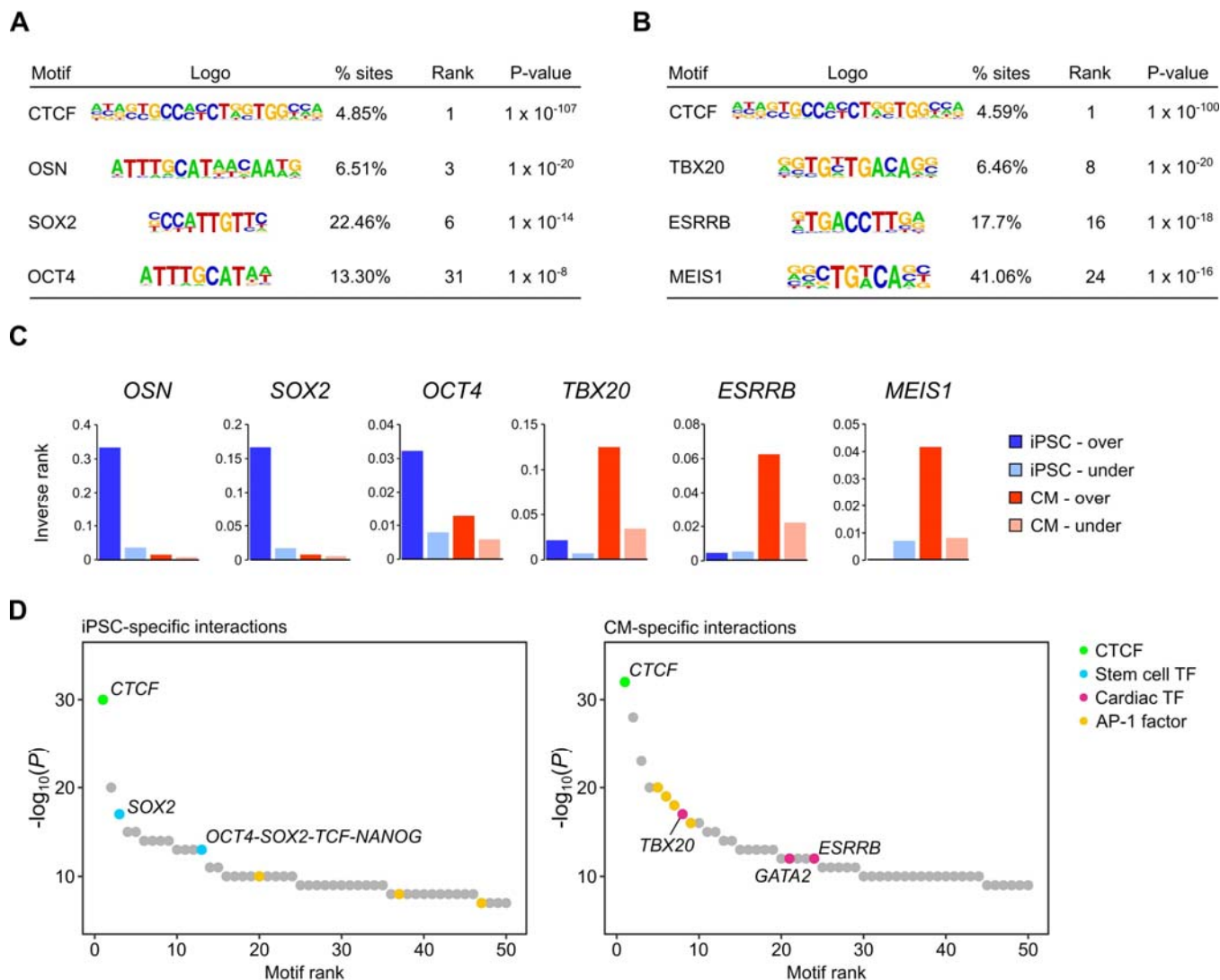
218

219

220

221

222

223

224

225

226

10

Figure 2. Transcription factor motif enrichment in distal interacting regions. (A,B) Selected transcription factor (TF) motifs identified using HOMER in the promoter-distal interacting sequences for all over-expressed genes in **(A)** iPSCs and **(B)** CMs (fold change > 1.5, P_{adj} < 0.05). "% sites" refers to the percent of distal interactions overlapping the motif; rank is based on P-value significance. **(C)** To compare motif ranks across gene sets, the inverse of the rank is plotted for selected motifs identified in distal interactions from over- or under-expressed genes in both iPSCs and CMs. **(D)** The top 50 motifs identified in cell type-specific interactions. *OSN*, OCT4-SOX2-TCF-NANOG motif.

Long-range promoter interactions are enriched for active *cis*-regulatory elements and correspond to gene expression dynamics

11

238   Functionally active *cis*-regulatory elements are characterized by the presence of specific histone

239   modifications; active enhancers are generally associated with H3K4me1 and H3K27ac (Creyghton et

240   al. 2010; Heintzman et al. 2009), whereas inactive (e.g. poised or silenced) elements are often

241   associated with H3K27me3 (Rada-Iglesias et al. 2011; Erceg et al. 2017). In support of the gene-

242   regulatory function of long-range interactions, we found that the promoter-distal MboI fragments

243   involved in significant promoter interactions were enriched for these three histone modifications in both

244   iPSCs and CMs (Figures 3A-C). When promoters were grouped by expression level, we observed that

245   this enrichment increased with increasing expression for H3K27ac and H3K4me1, and decreased with

246   increasing expression for H3K27me3, consistent with an additive nature of enhancer-promoter

247   interactions (Schoenfelder et al. 2015; Javierre et al. 2016), and validating that PCHi-C enriches for

248   likely functional long-range chromatin contacts.

249   A strong correlation (Pearson correlation coefficient $r > 0.7$) between the degree of histone

250   modifications and gene expression was first reported nearly ten years ago (Karlić et al. 2010), however

251   that analysis only considered histone modifications within 2 kb of promoters. To understand whether

252   this relationship extends beyond promoter-proximal regions, we correlated the number of histone ChIP-

253   seq peaks within 300 kb of promoters with the promoter's expression level (Figure 3—figure

254   supplement 1A,B). H3K27ac and H3K4me1 both positively correlated with expression level

255   (Spearman's $\rho = 0.22$ and 0.16, respectively in iPSC and $\rho = 0.23$ and 0.24, respectively in CMs, $P <$

256   2.2^{-16}); in contrast, H3K27me3 negatively correlated with expression level in CMs (Spearman's $\rho = -$

257   0.20, $P < 2.2^{-16}$), however this relationship was not present in iPSCs (Spearman's $\rho = 0.02$, $P = 0.06$).

258   Although moderate, these correlations could partially explain why higher expressed genes show

259   stronger enrichment for promoter interactions overlapping histone peaks when using a genome-wide

260   background model (see Methods), and lends support to the notion that active genes are located in

261   generally active genomic environments (Stevens et al. 2017; Gilbert et al. 2004).

12

262   We next investigated the relationship between cell type-specific interactions and enrichment for

263   tissue-specific CTCF, H3K27ac, and H3K27me3 marks, hypothesizing that interactions unique to

264   iPSCs or CMs would be most enriched for tissue-specific chromatin features. Indeed, we observed that

265   cell type-specific interactions preferentially involved H3K27ac peaks from the matched cell type, and

266   were either not enriched (iPSC) or depleted (CM) for H3K27ac marks that were specific to the non-

267   matched cell type (Figure 3E, middle panel). However, the strongest enrichment was for cell type-

268   specific interactions to overlap chromatin features that were present in both cell types (Figure 3E).

269   Additionally, interactions that were shared between iPSCs and CMs were most enriched for shared

270   chromatin features. These results suggest that all interactions, whether shared or unique to one cell

271   type, preferentially contact regulatory regions that are active in both cell types, whereas cell type-

272   specific interactions are not likely to occur in regions specifically marked in the non-matched cell type.

273   An example of a gene that encompasses these observations is the atrial natriuretic peptide

274   gene *NPPA* (Figure 3F) which is specifically expressed in cells of the heart atrium and is upregulated in

275   CMs (Figure 1—figure supplement 2C). *NPPA* makes numerous cell type-specific interactions to a

276   distal region that is only marked with active chromatin (H3K27ac and H3K4me1) in CMs; furthermore,

277   functional characterization showed that this region corresponds to an *in vivo* enhancer recapitulating

278   *NPPA*'s endogenous expression in the developing heart (Visel et al. 2007). Taken together, these

279   results illuminate the complex relationship between long-range promoter interactions and gene

280   regulation and provide evidence that promoter architecture reflects cell type-specific gene expression.

281

282

283

13

284
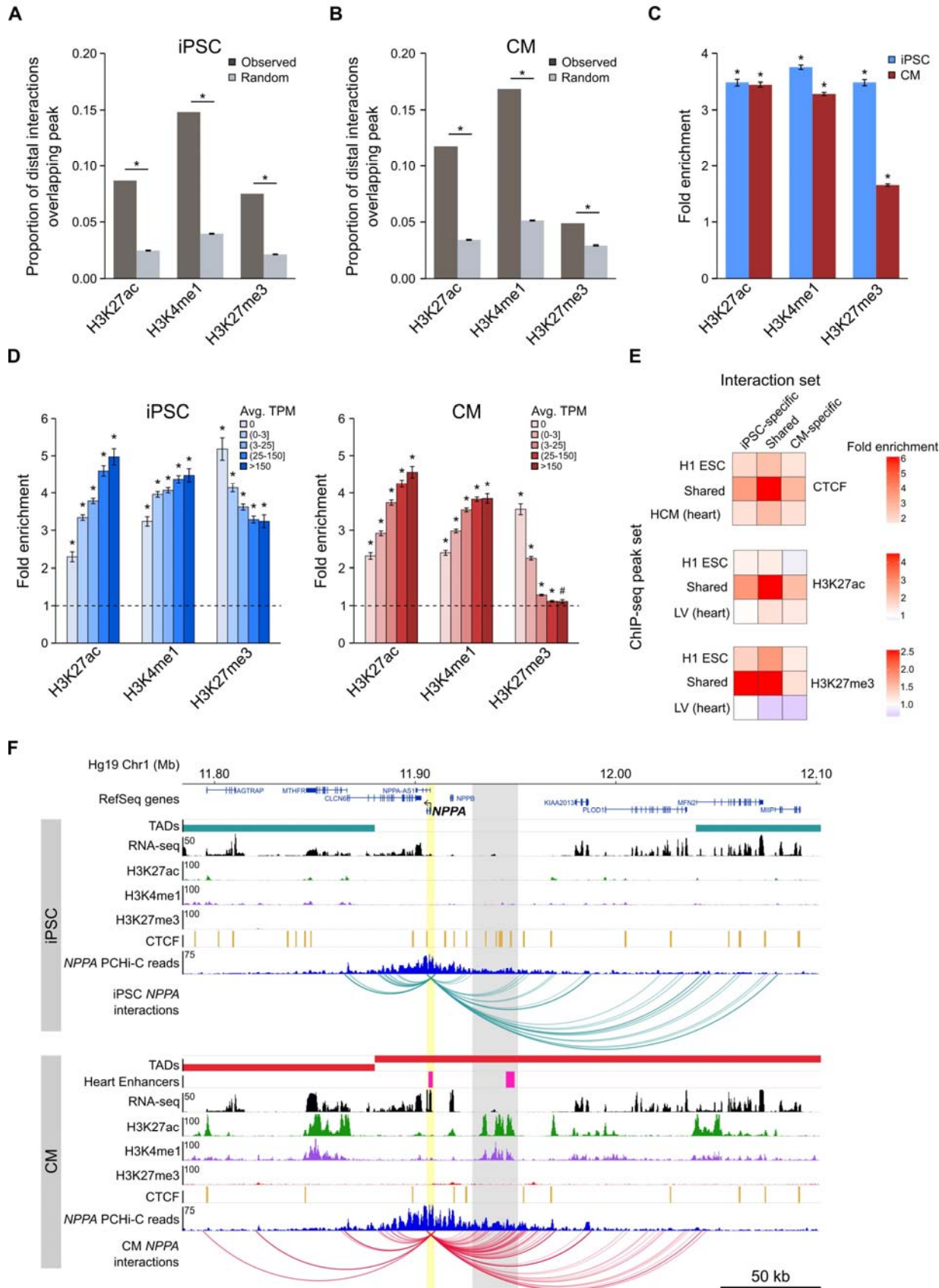
285 **Figure 3. Enrichment of promoter interactions to distal regulatory features. (A,B)** Proportion of
286 promoter-distal interactions overlapping a histone ChIP-seq peak compared to random control MboI
287 fragments (see Methods). iPSC interactions were overlapped with H1 ESC ChIP-seq data; CM
288 interactions were overlapped with left ventricle ChIP-seq data from the Epigenome Roadmap Project
289 (Supplementary File 10). **(C)** Fold enrichment of the data presented in **(A)** and **(B)**. **(D)** Fold enrichment
290 of promoter-distal interactions based on the expression level of the promoter. Promoters were grouped
291 into 5 bins according to their average TPM values. Dashed line indicates no enrichment. **(E)** Fold
292 enrichment of cell type-specific and shared interactions (columns) to tissue-specific and shared
293 chromatin features (rows). **(F)** Example of the *NPPA* gene in iPSCs (top) and CMs (bottom). Gray box
294 highlights CM-specific interactions to CM-specific chromatin marks and an *in vivo* heart enhancer (Visel
295 et al. 2007). For clarity, only interactions for *NPPA* are shown. *$P < 0.00001$, [#]$P = 0.0017$, Z-test.

296

297 **Dynamic changes in genomic compartmentalization involve a subset of cardiac-specific genes**

298   As a final benchmark of our datasets, we analyzed large-scale differences in genome

299 organization between iPSCs and CMs. The first Hi-C studies revealed that the genome is organized in

300 two major compartments, A and B, that correspond to open and closed regions of chromosomes,

301 respectively (Lieberman-Aiden et al. 2009; Rao et al. 2014). Although most compartments are stable

302 across different cell types, some compartments switch states in a cell type-specific manner which may

303 reflect important gene regulatory changes (Dixon et al. 2015). To assess whether capture Hi-C data,

304 which is more cost-effective for capturing promoter-centered interactions, is able to identify A/B

305 compartments, we compared our capture Hi-C data with pre-capture, genome-wide Hi-C libraries. A/B

306 compartments identified using HOMER (Heinz et al. 2010) were remarkably similar in the whole-

307 genome and PCHi-C datasets (97% correspondence, Figure 4A, top panel, and Figure 4—figure

308 supplements 1 and 2), demonstrating that PCHi-C data contains sufficient information to identify

309 broadly active and inactive regions of the genome. As an example, we highlight a 10 Mb region on

310 chromosome 4 containing the *CAMK2D* gene locus (Figure 4A). Compartments were relatively stable

311 across this region in iPSCs and CMs, however the *CAMK2D* gene itself was located in a dynamic

312 compartment that switched from inactive in iPSCs to active in CMs. Correspondingly, this gene was

313 highly upregulated during differentiation to CMs (Figure 4A, inset).

15

314   We observed this effect on a global level, as genes located in A compartments were expressed

315   at significantly higher levels than genes located in the B compartments in both iPSCs and CMs (Figure

316   4B). Additionally, genes that switched A/B compartments between cell types were correspondingly up-

317   or down-regulated (Figure 4C). GO analysis of the 1,008 genes that switched from B to A

318   compartments during iPSC-CM differentiation revealed enrichment for terms such as "cardiovascular

319   system development" and "heart contraction" (Figure 4D, Supplementary File 5). Importantly, these

320   genes were identified based solely on their location in a dynamic genomic compartment and not from

321   gene expression data. GO analysis for genes that switched from A to B compartments during iPSC-CM

322   differentiation related to non-cardiac processes, such as skin development, epithelial cell differentiation

323   and sex determination (Figure 4—figure supplement 3, Supplementary Files 5 and 6). These data show

324   that PCHi-C accurately captured tissue-specific interactions and indicate that compartmentalization of

325   genes in spatially regulated regions of the nucleus may be one mechanism to ensure tissue-specific

326   gene expression (Dixon et al. 2015). In summary, our analyses demonstrated that CM promoter

327   interactions recapitulate key features of cardiac gene regulation and function, validating the CM map as

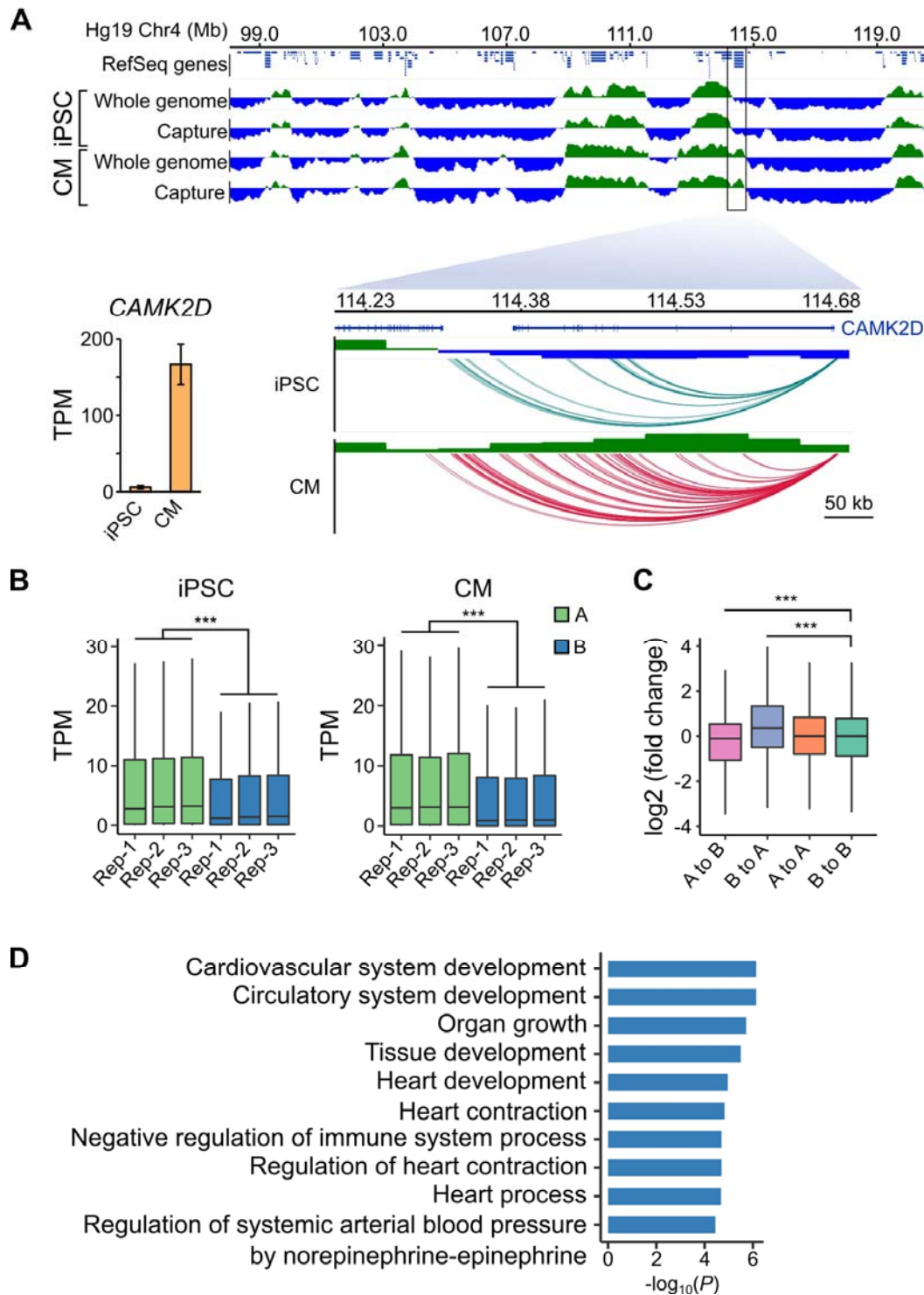328   an important tool to investigate CVD genetics.

329

330

331

332

333

334

335

336

337

Figure 4. **A/B compartment switching corresponds to activation of tissue-specific genes. (A)** Top panel: 10 Mb region on chromosome 4 showing A (green) and B (blue) compartments based on the first principle component analysis calculated by HOMER (Heinz et al. 2010) of the whole-genome Hi-C and capture Hi-C interaction data. Bottom panel: zoomed in on the *CAMK2D* locus; only capture Hi-C A/B

342 compartments shown. Inset: expression level of *CAMK2D* in iPSCs and CMs across the three
343 replicates. **(B)** Expression level (TPM) of genes located in the A (green) or B (blue) compartment in
344 each replicate of iPSC (left) or CM (right). **(C)** Difference in expression level (log2 fold change relative
345 to iPSCs) of genes switching compartments from iPSC to CM or remaining in stable compartments. **(D)**
346 Gene Ontology analysis of biological processes associated with genes switching from B to A
347 compartments during iPSC-CM differentiation. *** $P < 2.2 \times 10^{-16}$, Wilcoxon rank-sum test.

348

CM promoter interactions link GWAS SNPs to target genes

350 A particularly relevant application of high resolution promoter interaction maps is to guide post-

351 GWAS studies by identifying the target genes of disease-associated variants. We employed this

352 approach to link GWAS SNPs for several major cardiovascular diseases to their target gene(s) using

353 the CM interaction map. We compiled 524 lead SNPs from the NHGRI database

354 (https://www.ebi.ac.uk/gwas/) for three important classes of CVDs: cardiac arrhythmias, heart failure,

355 and myocardial infarction (Table 1, Supplementary Files 7 and 8). Because of linkage disequilibrium

356 (LD) patterns, the true causal SNP could be any SNP in high LD with the lead variant. Therefore, we

357 expanded this set of SNPs to include all variants in high LD ($r^2 > 0.9$, within 50 kb of lead SNP),

358 increasing the number of putatively causal variants to 10,475 (hereafter called LD SNPs). We found

359 that 1,999 (19%) of the LD SNPs were located in promoter-distal MboI fragments that interacted with

360 the promoters of 347 genes in CMs (Supplementary File 8), hereafter referred to as target genes. The

361 majority (89%) of LD SNP-target gene pairs were located within the same TAD, with a median distance

362 of 185kb between each SNP-target gene pair (Figure 5A). Importantly, 90.4% of SNP-target gene

363 interactions skipped at least one gene promoter and 42% of SNPs interacted with at least two different

364 promoters (Figure 5B).

365

366

367

368

369 **Table1. Summary of the SNPs and target genes characterized in each disease class**

	Arrhythmia	Myocardial Infarction	Heart Failure	Combined
Number of studies	30	11	11	50
Tag SNPs	358	86	80	524
SNPs in LD	6,555	1,822	2,098	10,475
SNPs looping to genes	1,152	357	490	1,999
Target genes	237	72	53	347

370 Summary values for each disease group are depicted along with the total number of GWAS, SNPs, and
371 target genes ("Combined" column). Tag SNPs were identified from the published GWAS in the NHGRI-
372 EBI database; SNPs in LD are the total number of non-promoter SNPs (including tag SNPs) in LD (r^2
373 >0.9) with the tag SNPs in each disease group; SNPs looping to genes are the SNPs in LD that are in a
374 distal promoter interaction; Target genes are all genes with an interaction to a promoter-distal SNP.
375 See Supplementary File 8 for a complete list of all GWAS, coordinates of each SNP and its assigned
376 target gene, expression level in iPSC and CM, and mouse knock-out phenotype where available.

377

378 　　　　To confirm that the CM PCHi-C interactions linked SNPs to CVD-relevant target genes, we

379 performed GO analysis and found that target genes were highly and specifically enriched for biological

380 processes related to cardiac function, such as membrane repolarization and cardiac conduction (Figure

381 5C, left panel and Supplementary Files 5 and 6). As a control, we used iPSC interactions to link the

382 same SNPs to target genes and observed a completely different set of unrelated biological processes

383 for these genes (Figure 5C, right panel). To further characterize the biological relevance of target

384 genes, we mined mouse knock-out data from the Mouse Genome Informatics (MGI) database (Blake et

385 al. 2017), which revealed that a statistically significant number of target genes resulted in a

386 cardiovascular phenotype when knocked-out in the mouse (78 genes (22.4%), $P = 1\times10^{-5}$, Figure 5D).

387 Finally, we examined expression quantitative trait loci (eQTL) data from human left ventricle (LV) tissue

388 and found that of the 1,999 LD SNPs in interactions, 410 (20.5%) corresponded to LV eQTLs; in

389 comparison, only 12.2% of the full set of LD SNPs corresponded to LV eQTLs ($P < 0.00001$, Figure

390 5E). We next assessed whether eQTLs loop to their associated gene. For this analysis, we considered

391 the full set of LV eQTLs, as the 410 LD SNP eQTLs represent too small of a proportion of the full set

392 (<0.1% of all LV eQTLs) to fully ascertain significance. On a genome-wide level, LV eQTLs in promoter-

393 distal interactions were significantly more likely to loop to their associated gene than expected by

394 chance ($P < 0.00001$, Figure 5F, left panel). Importantly, this significance decreased when LV eQTLs

395 were analyzed with iPSC promoter interactions ($P = 0.035$, Figure 5F, right panel). Taken together,

396 these results indicate that CM promoter interactions identify a subset of disease-relevant SNPs most

397 likely to be functional and support the use of the CM map to assign distal CVD-associated SNPs to

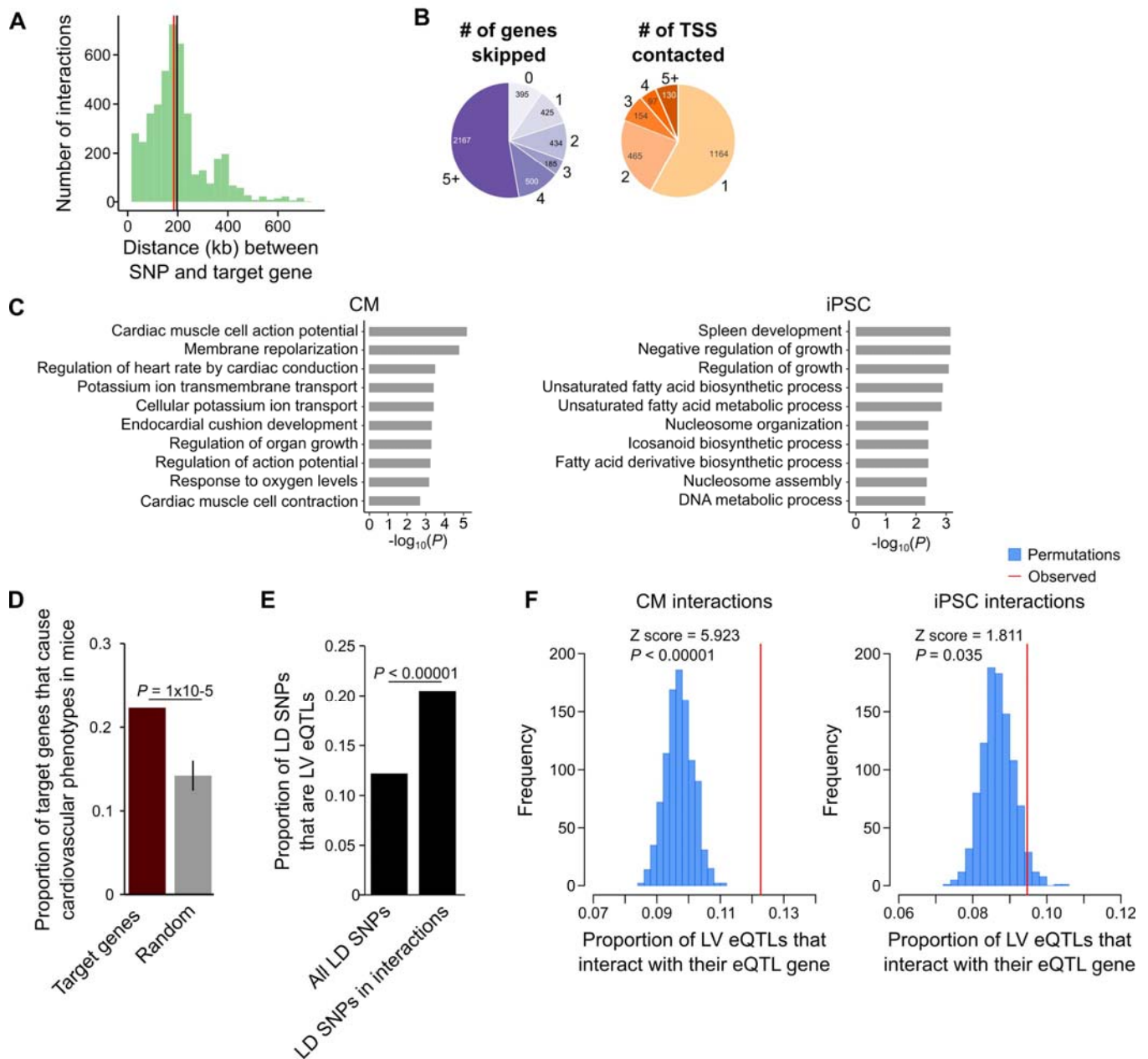398 putative target genes.

399

400

401

402

403

404

405

406

407

408

409

410

411

412

413

Figure 5. CM promoter interactions link CVD GWAS SNPs to target genes. (A) Distribution of genomic distances separating SNP-target gene interactions (red line, median = 185 kb; black line, mean = 197 kb). **(B)** Pie chart showing the number of TSS's skipped for each SNP-target gene interaction (left) and the number of genes contacted by each SNP (right). **(C)** GO enrichment analysis for genes looping to LD SNPs using the CM promoter interaction data (left panel) or the iPSC promoter interaction data (right panel). **(D)** Proportion of target genes that result in a cardiovascular phenotype when knocked-out in the mouse (MGI database (Blake et al. 2017)), compared to a random control set.

421   P-value calculated with a Z-test. **(E)** Proportion of GWAS LD SNPs that are eQTLs in left ventricle (LV)

422   when considering either the full set of LD SNPs, or the subset that overlap CM promoter interactions. P-

423   value calculated with Fisher's exact test. **(F)** Proportion of LV eQTLs (genome-wide) that map within a

424   promoter interaction for the eQTL-associated gene (indicated by the red line). Random permutations

425   were obtained by re-assigning each promoter's set of interactions to a new promoter and calculating

426   the proportion of eQTLs in random interactions that interact with their eQTL-associated gene.

427   Proportions only consider eQTLs that overlap a promoter-distal interaction. P-values calculated with a

428   Z-test.

429

430   **Using gene expression as a metric for interpreting disease-relevance of newly identified target**

431   **genes**

432   Based on an enrichment of target genes with known cardiac function, we next assessed

433   whether expression level is an informative metric to further prioritize functional follow-up studies. We

434   examined the expression level of the 347 target genes and found that they were moderately over-

435   expressed in CMs compared to iPSCs (median log2 fold change = 1.08, mean log2 fold change = 1.44,

436   mean TPM values were 40.6 in iPSCs and 60.1 in CMs, P = 0.12, Figures 6A and 6B). Although not

437   significant, this result reflects the enrichment of known cardiac-related genes that interact with CVD

438   loci. However, because a subset of target genes was over-expressed in iPSCs relative to CMs (Figure

439   6C), we predicted that gene expression level alone may be an insufficient metric to gauge the

440   relevance of target genes to CVD biology. Indeed, we found that 21 of the 78 target genes (27%) that

441   cause cardiovascular phenotypes when knocked-out in mice were overexpressed in iPSCs compared

442   to CMs (Supplementary File 8). This result indicates that putatively causal genes may not appear as

443   obvious candidates based solely on gene expression data.

444   To illustrate this point, we highlight two genes: *TBX5*, a gene directly linked to cardiac

445   arrhythmia (Figure 6D) (Smemo et al. 2012; Arnolds et al. 2012), and *LITAF*, a gene that, until recently,

446   had no obvious role in cardiac biology (Moshal et al. 2017) (Figure 6E). Both genes formed long-range

447   interactions to LD SNPs identified in arrhythmia GWAS, making both genes candidate functional targets

448   of the GWAS associations. *TBX5*, which is over-expressed in CMs (Figure 6C), is the most likely target

449 gene of the LD SNPs nearby based on the interaction data but also because of its known role in

450 directing proper development of the cardiac conduction system. *LITAF*, on the other hand, was over-

451 expressed in iPSCs compared to CMs (Figure 6C), and was not known to contribute to cardiac function

452 until a recent study identified this gene as a regulator of cardiac excitation in zebrafish hearts (Moshal

453 et al. 2017).
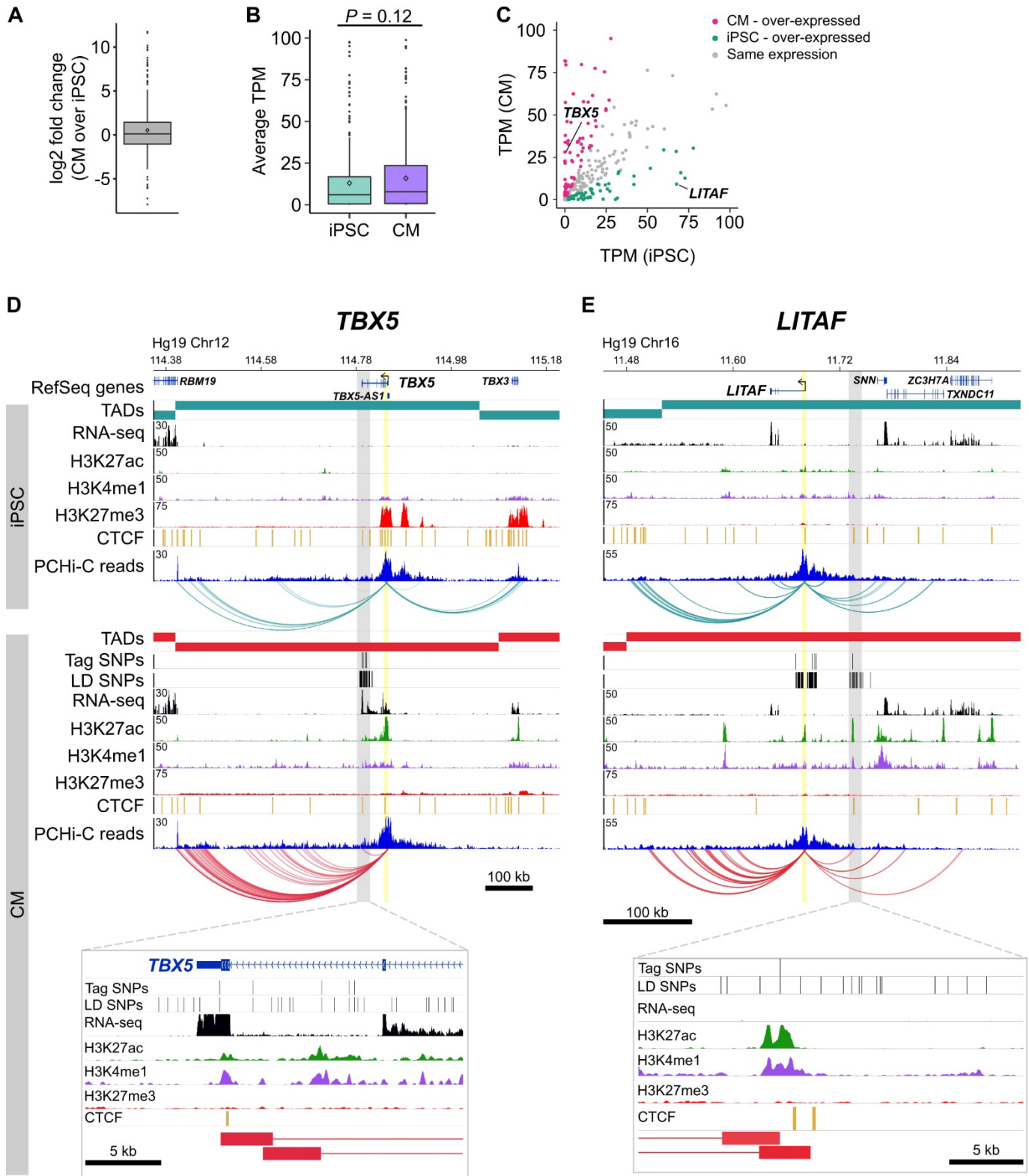
454

455

456

457

458

459

460

461

462

463

464

465

466

467

468

469

470

471

472   **Figure 6. Characterizing target genes based on expression level. (A)** Log2 fold change of the
473   expression level of target genes in CMs compared to iPSCs (horizontal bar indicates median, 1.08;
474   diamond indicates mean, 1.44). **(B)** Average TPM values of target genes in iPSCs and CMs ($P = 0.12$,
475   Wilcoxon rank-sum test). Diamonds indicate the mean value (40.6 for iPSC, 60.1 for CM). **(C)**
476   Comparison of average TPM values for target genes in CMs and iPSCs. See Supplementary File 8 for
477   full list of genes and TPM values. **(D,E)** Examples of genes looping to cardiac arrhythmia GWAS SNPs
478   in CMs. **(D)** The *TBX5* gene interacts with a functionally validated arrhythmia locus (Smemo et al.
479   2012). **(E)** The *LITAF* gene interacts with a locus identified in (Arking et al. 2014). Yellow highlighted
480   region indicates the promoter; gray box and zoom panel show the promoter-interacting regions
481   overlapping arrhythmia SNPs. For clarity, only interactions for the indicated promoter are shown.

482

483   **CM promoter interactions are informative to cardiovascular associations that do not directly**

484   **involve cardiomyocytes**

485   Because the three disease classes that we analyzed represent diverse pathologies, we

486   predicted that the target genes identified for each class individually may relate to different biological

487   processes. Specifically, we considered that cardiac arrhythmias – which directly result from defects in

488   cardiomyocytes specialized for electrical conduction – may uncover the most cardiac-relevant target

489   genes compared to heart failure and myocardial infarction, two CVDs that also involve non-cardiac

490   systems. When broken down into the respective disease classes, we confirmed that the majority of the

491   GO enrichment for cardiac terms was driven by the cardiac arrhythmia SNPs (Figure 7A), with terms

492   directly related to the cardiac conduction system. Myocardial infarction (Figure 7B) and heart failure

493   (Figure 7C) analyses uncovered a set of genes that were slightly enriched for regulation of growth and

494   morphogenesis, respectively.

495   Despite these seemingly non-specific processes, each set of target genes contained important

496   disease-relevant candidates. For example, one of the strongest associations for myocardial infarction

497   lies in-between the *CELSR2* and *PSRC1* genes on chromosome 1p13, but a careful screen of genes

498   whose expression was affected by the risk allele implicated the more distal *SORT1* gene (Musunuru et

499   al. 2010). *SORT1* encodes a sorting receptor that is expressed in many tissues and has been shown to

500   act in the liver to regulate cholesterol levels (Petersen et al. 1997; Musunuru et al. 2010). Despite

25

501 functioning in the liver, we identified multiple promoter interactions between *SORT1* and the myocardial

502 infarction GWAS locus in CMs (Figure 7D), directly implicating *SORT1* as the target gene and lending

503 further support to experimental validation of this locus as a *SORT1* enhancer (Musunuru et al. 2010).

504 Additionally, the *ACTA2* gene is located 220 kb away from the heart failure GWAS locus proximal to the

505 *CH25H* and *LIPA* genes on chromosome 10q21 (Smith et al. 2010) (Figure 7E). *ACTA2* encodes the

506 smooth muscle cell-specific actin protein and mutations in this gene have been shown to cause

507 coronary artery disease, among other vascular diseases (Guo et al. 2009). Despite its location at a

508 considerable distance from the GWAS association, chromatin interactions provide an important level of

509 evidence that *ACTA2* is a putative causal gene in the development of heart failure. Therefore, the CM

510 interaction map is not only useful to interrogate diseases directly related to cardiomyocytes, as in the

511 case of cardiac arrhythmias, but also aids interpretation of target genes that may act in non-cardiac

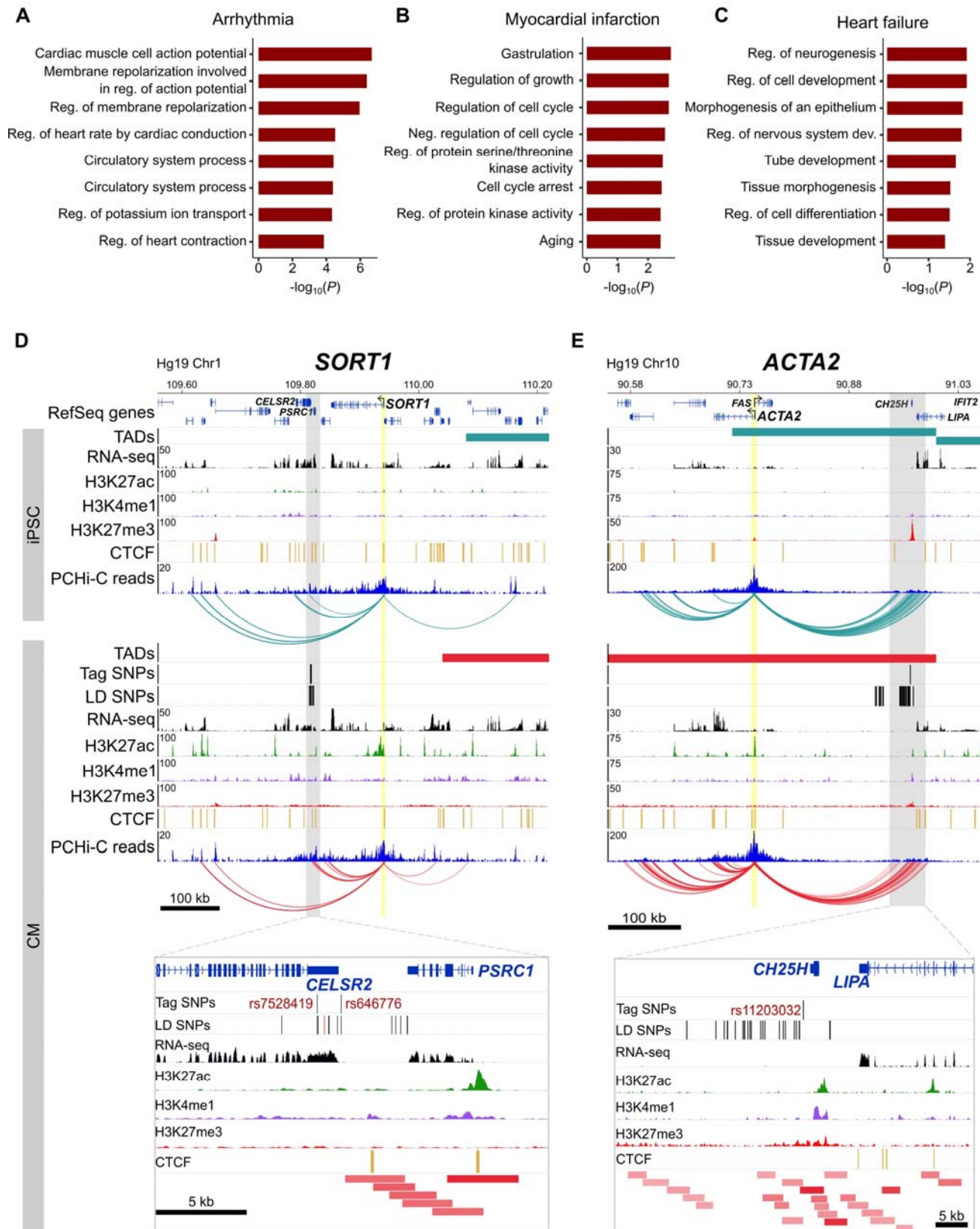512 tissues.

513

514

515

516

517

518

519

520

521

522

26

523

524

525 **Figure 7. Relevance of CM promoter interactions for cardiac arrhythmia, myocardial infarction**
526 **and heart failure. (A-C)** Gene Ontology analysis for target genes looping to **(A)** cardiac arrhythmia
527 SNPs, **(B)** myocardial infarction SNPs, and **(C)** heart failure SNPs. **(D)** The *SORT1* promoter loops to a
528 distal myocardial infarction locus (Musunuru et al. 2010). The rs12740374 SNP shown to disrupt a
529 C/EBP binding site in (Musunuru et al. 2010) is colored red. **(E)** The *ACTA2* promoter loops to the
530 10q21 heart failure locus (Smith et al. 2010). Zoom plots depict the full interacting region overlapping
531 GWAS LD SNPs. For clarity, only interactions for the indicated gene are shown.

532

533 **Discussion**

534   Incomplete understanding of long-range gene regulation is a major roadblock in the translation

535 of GWAS-associated loci to disease biology. Major challenges in this process include identifying

536 putatively causal variants mapping within regulatory elements and functionally connecting these

537 regulatory elements to their target genes. To delineate gene-regulatory interactions between CVD-

538 associated SNPs and putative causal genes, we generated high-resolution maps of promoter

539 interactions in human iPSCs and iPSC-derived CMs. We demonstrated that promoters interact with a

540 diverse set of distal DNA elements in both cell types, including known enhancer sequences, which

541 reflect cell identity and correspond to tissue-specific gene expression. To demonstrate the utility of the

542 CM map, we linked 1,999 CVD-associated SNPs to putative causal target genes which identified both

543 validated and potentially novel genes important for cardiovascular disease biology. To validate the

544 biological relevance of our maps, we addressed several important features of long-range chromatin

545 interactions in comparative analyses.

546 *Promoters contact distal regions enriched for tissue-specific transcription factor motifs*

547   Gene regulation by distant regulatory elements involves the bridging of linearly separated DNA

548 sequences, for example between a promoter and its distal enhancers, through chromatin looping

549 mechanisms (Spitz and Furlong 2012). In support of this model, we report an enrichment of tissue-

550 defining transcription factor motifs in the distally interacting sequences of differentially expressed

551 promoters both for CMs and iPSCs, providing an important level of evidence to validate the functional

552   relevance of iPSC and CM interactions. One explanation for this enrichment is that our interaction maps

553   are high resolution. We generated Hi-C libraries with the 4-bp cutter MboI, which generates fragments

554   with an average size of 422 bp; this increased specificity of the captured region likely leads to better

555   resolution of the underlying enhancer sequence and, consequently, increased power to detect short

556   transcription factor binding motifs.

557   *Influence of active and repressive promoter interactions on gene expression level*

558   The majority of capture Hi-C studies to date have reported that gene expression level correlates

559   with enrichment for various histone marks. We observed the same trend in our data, with highly

560   expressed genes exhibiting strong enrichment for looping to distal H3K4me1 and H3K27ac-marked

561   regions, and lowly expressed genes exhibiting strong enrichment for looping to H3K27me3-marked

562   regions. These data are consistent with a model in which the number of long-range interactions to

563   enhancers or repressors additively contributes to gene expression level (Schoenfelder et al. 2015;

564   Javierre et al. 2016). The forces that drive increased association between promoters and distal *cis*-

565   regulatory elements are not completely understood and have been topics of investigation in the

566   genome organization and chromatin biology fields for several years (Dekker and Mirny 2016; Calo and

567   Wysocka 2013). One possibility is that this increasing enrichment is driven by genomic

568   compartmentalization of active and inactive chromatin. We showed that a gene's expression level

569   correlates with the number of histone ChIP-seq peaks within a large window (300 kb) surrounding each

570   promoter. Thus, highly expressed genes are more likely to contact active chromatin regions compared

571   to lowly expressed genes, corresponding to the observed increasing enrichment we and others have

572   reported. This local increase in active or repressive chromatin may be one driving force underlying the

573   expression level-dependent increase in association between promoters and *cis*-regulatory elements,

574   akin to a phase separation-mediated model of enhancer-promoter interactions (Hnisz et al. 2017).

575   *A promoter interaction map for cardiovascular disease genetics*

576   We demonstrated several ways in which promoter interaction data can be used to better

577   understand disease genetics, specifically addressing the major requirement for a high-resolution map of

578   the gene-regulatory network in human cardiomyocytes. Although iPSC-derived CMs are known to be

579   relatively immature and do not fully reflect the diverse structural and functional aspects of adult cardiac

580   cells (Gherghiceanu et al. 2011; Karakikes et al. 2015), the difficulty in obtaining pure sub-populations

581   of primary cardiomyocytes with high integrity necessitates the use of an *in vitro* system. We showed

582   that the CMs used in this study were highly pure and recapitulate known gene regulatory properties of

583   primary cardiomyocytes. Because of this purity, we were able to integrate CVD-associated SNPs with

584   CM promoter interactions with high confidence, assigning nearly 20% of the variants in high LD with

585   these associations to 347 target genes.

586   Supporting the physiological relevance of CMs to the cardiac conduction system, we found that

587   target genes were most relevant for GWAS loci associated with cardiac arrhythmias, in line with

588   previous findings in immune cells that many target gene interactions were unique to relevant immune

589   cell subtypes (Javierre et al. 2016; Mumbach et al. 2017). Our data also revealed that even for

590   diseases whose etiology involves cell types other than cardiomyocytes, such as myocardial infarction

591   and heart failure, we identified interactions involving loci associated with these diseases that

592   recapitulate the enhancer-promoter interactions in non-cardiac cell types. As an example, we showed

593   that a validated myocardial infarction locus interacts with the distal *SORT1* promoter in CMs even

594   though this locus has been extensively characterized in the context of cholesterol metabolism in

595   hepatocytes. Therefore, the promoter interactions we observe linking the disease locus to *SORT1* may

596   represent tissue-invariant genome architecture, likely reflecting that genome organization in general is

597   relatively stable (Dixon et al. 2015; Jin et al. 2013; Ghavi-Helm et al. 2014). While we advocate the use

598   of the CM map for investigating gene regulatory mechanisms of diseases related to cardiomyocyte

599   biology, we also emphasize that, where identified, any interaction between a promoter and a putative

30

600 disease-associated genomic region serves as an important level of evidence to prioritize that gene for

601 future follow-up studies.

602 *Limitations of the PCHi-C maps*

603 The PCHi-C technique holds great promise to identify with high resolution and throughput all

604 regulatory elements for all genes in any tissue or developmental stage of interest. However, due to

605 technical and biological limitations, there are important caveats to PCHi-C that should be considered

606 when interpreting the iPSC or CM interaction data. The most important caveat is that there are likely to

607 be many false negatives, or "missing" interactions. Although the capture step greatly enriches for

608 promoter-containing ligation fragments in a Hi-C library, the total landscape of promoter contacts in a

609 population of cells is still under-sampled, even with a sequencing depth of ~400M reads per replicate

610 conducted for this study. This is due to several factors, including the hybridization efficiency of each

611 bait, ability to design sufficient baits per promoter, and the transient nature of many regulatory

612 interactions. This latter issue is confounded by the distance-dependent effect on ligation frequency: as

613 the distance between two fragments increases, the read-depth required to robustly identify that

614 interaction also increases. Because many regulatory elements are located far from their target gene(s),

615 very deep sequencing is required to identify them with high confidence. The feasibility of deeper

616 sequencing and modifications to computational pipelines will continue to improve the coverage and

617 resolution of Hi-C data. Nevertheless, the data sets we provide here represent a highly enriched set of

618 ~350,000 and ~400,000 promoter interactions in iPSC and CMs, respectively; although there are likely

619 missing interactions, the interactions that we did identify should be considered as very high confidence,

620 as they were independently identified in at least two biological replicates and show strong signal of

621 enrichment for known features of genome architecture and gene regulation.

622 In conclusion, the promoter interaction maps we generated in this study represent important

623 resources for any investigation into the gene regulatory mechanisms underlying cardiovascular disease

624 traits. The list of candidate regulatory variants and their target genes may serve as an entry point for

31

625   several hypotheses related to CVD GWAS, and can be readily tested in experimental settings. To

626   provide both the iPSC and CM maps as an accessible resource, we have hosted the full set of data

627   presented in this study at the WashU EpiGenome Browser. Additionally, we provide the significant

628   PCHi-C interaction files used in all analyses in the Supplementary Material (Supplementary Files 1 and

629   2); these can be applied to future multi-omics analyses of gene regulation and disease genetics.

630

631   **Methods**

632   ***Tissue culture of iPSCs***

633   We used the Yoruban iPSC line 19101, kindly provided by the laboratory of Yoav Gilad. This

634   iPSC line was reprogrammed from lymphoblastoid cells as part of a previous study, where it was shown

635   to differentiate into all three germ layers, displayed a normal karyotype, and expressed markers

636   characteristic of pluripotency (Banovich et al. 2018). iPSCs were grown in Essential 8 (E8) Medium

637   (Thermo Fisher #A1517001) supplemented with 1X Penicillin-Streptomycin (Pen/Strep, Gibco) on

638   Matrigel-coated tissue culture dishes (Corning #354277). Cells were passaged when they were ~80%

639   confluent using enzyme-free dissociation solution (30mM NaCl, 0.5mM EDTA, 1X PBS minus

640   Magnesium and Calcium) and maintained in E8 Medium with 10µM Y-27632 dihydrochloride (Abcam

641   #ab120129) for 24 hours. Medium was replaced daily. iPSC cultures routinely tested negative for

642   mycoplasma contamination using the Universal Mycoplasma Detection Kit (ATCC #30-1012K).

643   ***Cardiomyocyte differentiation***

644   Cardiomyocyte differentiations were based on the protocol of Burridge *et al.* (Burridge et al.

645   2014) with modifications described in Banovich *et al.* (Banovich et al. 2018). iPSCs were expanded in

646   60 mm dishes in E8 media until they reached 60-70% confluency at which time the differentiation was

647   started (day 0). On day 0, E8 media was replaced with 10mL of basic heart media/12µM GSK-3

648   inhibitor CHIR-99021 trihydrochloride (Tocris #4953)/Matrigel overlay [basic heart media: RPMI 1640

649 minus L-glutamine (HyClone #SH30096.01) with 1X GlutaMax (Life Technologies #11879020)

650 supplemented with 1X B27 minus insulin (Thermo Fisher #A1895601) and 1X Pen/Strep; Matrigel

651 overlay was accomplished by dissolving Matrigel in 50mL basic heart media at a concentration of 0.5X

652 according to the lot-specific dilution factor]. After 24 hours (day 1), the GSK-3 inhibitor was removed by

653 replacing media with 10 mL basic heart media. On day 3, media was replaced with 10 mL basic heart

654 media supplemented with 2 µM Wnt-C59 (Tocris #5148). On day 5 (48 hours later), media was

655 replaced with 10 mL basic heart media. On day 7, cells were washed once with 1X PBS and then 15

656 mL basic heart media was added. Media was replaced every other day in this way until day 15 at which

657 time cardiomyocytes were selected for by replacing basic heart media with 10mL lactate media (RPMI

658 1640 minus D-glucose, plus L-glutamine (Life Technologies #11879020), supplemented with 0.5 mg/mL

659 recombinant human albumin (Sigma 70024-90-7), 5 mM sodium DL-lactate (Sigma 72-17-3), 213

660 µg/mL L-ascorbic acid 2-phosphate (Sigma 70024-90-7) and 1X Pen/Strep). Lactate media was

661 replaced every other day until day 20 at which point cardiomyocytes were harvested. Cells from

662 successful differentiations exhibited spontaneous beating around days 7-10.

663   Cardiomyocytes were harvested by washing once with 1X PBS followed by incubation in 4 mL

664 TrypLE (Life Technologies 12604-021) at 37°C for 5 minutes. After incubation, 4 mL lactate media was

665 added to the TrypLE and a 1 mL pipet was used to dislodge cells. Cells were strained once with a

666 100µM strainer and then once with a 40µM strainer. Cells were pelleted at 500xg and then

667 resuspended in PBS and counted. For each batch of differentiation, 5 million cells were taken for

668 promoter-capture Hi-C and 1 million cells were taken for RNA-seq. To assess purity, 2 million cells

669 were taken for flow cytometry analysis using an antibody for cardiac Troponin T (BD Biosciences

670 564767). All cells used in downstream experiments were at least 86% Troponin T positive (Figure 1—

671 figure supplement 1A). We carried out three independent differentiations of the same iPSC line and

672 generated promoter-capture Hi-C and RNA-seq libraries in iPSCs and CMs from each triplicate.

673 **_Promoter capture Hi-C_**

33

674 ***-Crosslinking cells***

675   iPSCs or cardiomyocytes were harvested from tissue culture dishes and counted. Cells were

676 resuspended in 1X PBS at a concentration of 1 million cells/mL and 37% formaldehyde was added to a

677 final concentration of 1%. Crosslinking was carried out for 10 minutes at room temperature on a rocking

678 platform. Glycine was added to a final concentration of 0.2 M to quench the reaction. The cells were

679 pelleted, snap frozen in liquid nitrogen and stored at -80°C until ready for Hi-C processing.

680 ***-in situ Hi-C***

681   We prepared all promoter capture Hi-C libraries in one batch using three crosslinked pellets of 5

682 million cells for both iPSCs and iPSC-derived cardiomyocytes, representing three independent

683 cardiomyocyte differentiations. The *in situ* Hi-C step was performed as in Rao *et al.* (Rao et al. 2014)

684 with a single modification in which NEBNext reagents from the NEBNext Multiplex Oligos for Illumina kit

685 were used (NEB #E7335S) instead of Illumina adapters, following the manufacturer's instructions. Hi-C

686 libraries were amplified directly off of T1 beads (Life Technologies #65602) using NEBNext primers and

687 6 cycles of PCR.

688 ***-Promoter capture – probe design and generation***

689   Hi-C capture probes were designed to target four MboI restriction fragment ends (120 bp) near

690 the TSS of protein coding RefSeq genes (O'Leary et al. 2016) mapped to hg19 in the UCSC Genome

691 Browser (Speir et al. 2016). To select restriction fragments, we only kept MboI restriction fragments

692 longer than 200 bp and overlapping 10 kb around a RefSeq TSS. TSSs closer than 1 kb from each

693 other were excluded, as their interactions were likely to be captured by the other RefSeq TSS. The four

694 MboI restriction fragments ends closest to each RefSeq TSS were selected as putative probes. The

695 120 bp sequences were submitted to Agilent's SureDesign proprietary software for probe selection,

696 which can slightly shift the location and remove probes. In total, we ordered a library of 77,476 single-

697 stranded DNA oligos from CustomArray, Inc. (www.customarrayinc.com). Each oligo consisted of the

34

698   sequence 5'-ATCGCACCAGCGTGTN$_{120}$CACTGCGGCTCCTCA-3' (Gnirke et al. 2009) where N$_{120}$

699   represents the 120 nucleotides adjacent to the MboI cut site. The complete list of oligo probes and

700   their corresponding gene name is provided in Supplementary File 9.1.

701   The oligos arrived as a pool containing 1000 ng of material. We used 16 ng of the oligo pool in a

702   PCR reaction to make them double stranded using primers 5'-CTGGGAATCGCACCAGCGTGT-3'

703   (Primer A), and 5'-CGTGGATGAGGAGCCGCAGTG-3' (Primer B) as in (Gnirke et al. 2009). The PCR

704   reaction was cleaned using AMPure XP beads (Agencourt #A6388) and eluted with 20µl of water. To

705   add the full T7 promoter to the 5' end of the oligos, a second PCR reaction was carried out using 10ng

706   of the cleaned-up first-round PCR product with the forward primer 5'-

707   GGATTCTAATACGACTCACTATAGGGATCGCACCAGCGTGT-3' (Primer A T7). We purified the PCR

708   product corresponding to 176 bp using a Qiagen gel extraction kit (#28704). To generate biotinylated

709   RNA baits, we performed *in vitro* transcription on the double-stranded library using the

710   MEGAshortscript T7 Transcription Kit (Thermo Fisher #AM135) with Biotin-16-dUTP (Sigma

711   #11388908910). After DNase treatment the transcription reaction was cleaned using the MEGAclear kit

712   (Thermo Fisher #AM1908) and eluted with 50µl elution buffer. We confirmed the correct bait size on a

713   denaturing gel.

714   ***-Promoter capture – hybridization with Hi-C library***

715   To isolate promoter-containing fragments from the whole-genome *in situ* Hi-C library, we

716   hybridized the biotinylated RNA bait pool with the Hi-C library as follows. A mix containing 500ng of the

717   Hi-C library, 2.5µg of human Cot-1 DNA (Invitrogen #15279-011), 2.5µg of salmon sperm DNA

718   (Invitrogen #15632-011), 0.5µl blocking primer P5 (IDT #1016184), and 0.5µl blocking primer P7 (IDT

719   #1016186) was heated for 5 min. at 95°, held at 65° and mixed with 13µl pre-warmed hybridization

720   buffer (10X SSPE, 10X Denhardt's, 10 mM EDTA and 0.2% SDS) and a 6 µl pre-warmed mix of 500ng

721   of the biotinylated RNA bait and 20U SUPERase-In (Thermo Fisher #AM2694). The hybridization mix

35

722 was incubated for 24h at 65°C. To isolate captured fragments, we prepared 500ng of streptavidin-

723 coated magnetic beads (Dynabeads MyOne Streptavidin T1, Thermo Fisher #65601) in 200µl of

724 Binding buffer (1M NaCl, 10mM Tris-HCl pH 7.5, 1mM EDTA). The hybridization mix was added to the

725 Streptavidin beads and rotated for 30 minutes at room temperature. The beads containing the captured

726 Hi-C fragments were washed with 1X SSC, 0.1% SDS for 15 minutes at room temperature, followed by

727 three washes (10 min each) at 65°C with 0.1X SSC/0.1% SDS. After the final wash, the beads were

728 resuspended in 22µl of water and proceeded to post-capture PCR. The PCR reaction was performed

729 with 11µl of the "capture Hi-C beads" and 8 cycles of amplification. An AMPure XP bead purification

730 was used to clean the PCR reaction and DNA was quantified using the QuantiFluor dsDNA System

731 (Promega #E2670) and a High Sensitivity Bioanalyzer. Final capture Hi-C libraries were subjected to

732 100bp paired-end sequencing on an Illumina HiSeq 4000 machine. Read count summaries are

733 provided in Supplementary File 9.2.

734 ***Interaction calling***

735 We used HiCUP v0.5.9 (Wingett et al. 2015) to align and filter Hi-C reads (total and filtered read

736 counts are presented in Supplementary File 9.2). Unique reads were given to CHiCAGO version 1.2.0

737 (Cairns et al. 2016) and significant interactions were called with default parameters. In this study, we

738 focused exclusively on *cis*-interactions as the evidence that *trans*-chromosomal interactions contribute

739 to gene expression regulation is limited. CHiCAGO reports interactions for each captured restriction

740 fragment; to summarize interactions by gene, we considered the interval spanning all captured MboI

741 fragments (i.e. the set of probes spanning each TSS) as the promoter region ("merged TSS"). This

742 means the promoter regions created have variable lengths. In cases where multiple genes were

743 annotated to the same promoter region, we report the interaction for each gene individually. This

744 annotation allowed us to perform gene-level analyses, for example based on expression level. We

745 removed this redundancy as necessary, for example in motif enrichment analyses of the promoter-

746 interacting fragments. Using the "merged TSS" interaction files, we filtered interactions to retain those

36

747   that mapped within 1kb of each other in at least two replicates. Specifically, we extended each

748   promoter-interacting fragment by 1kb on each end and then used BEDTools (Quinlan and Hall 2010)

749   pairToPair functionality to identify interactions where both ends matched across replicates. To identify

750   cell type-specific interactions, we required that the interaction (with the 1 kb extension) was not present

751   in any of the three replicates of the other cell type. The number of read-pairs per promoter and the

752   corresponding number of significant interactions identified is presented in Supplementary File 9.3. The

753   TAD analyses, motif enrichment, ChIP-seq peak enrichment, and eQTL analyses (related to Figures 1,

754   2, 3 and 5) were conducted with fragment-level interactions (no 1kb extension). The GWAS SNP

755   analyses were conducted with 1kb-extended interactions, as we aimed to be as inclusive as possible

756   when linking CVD SNPs to target genes.

757   *4C-style plots*

758   To generate the by-gene read counts displayed in the genome-browser figures, all read-pairs

759   mapping to captured MboI fragments for a given promoter were summed across replicates. Specifically,

760   we summed reads for each MboI fragment where the read was part of a paired-read that mapped to a

761   bait for the given gene. The arcs that are displayed underneath the 4C-style plot represent significant

762   interactions that were identified in at least two replicates as detailed above in "Interaction calling".

763   *TAD analysis*

764   To identify TADs, we pooled reads across replicates for each cell type using the pre-capture Hi-

765   C data (600M reads for iPSC and 733M reads for CM) and used HiCUP v0.5.9 (Wingett et al. 2015) to

766   align and filter Hi-C reads. HOMER v4.8.3 (Heinz et al. 2010) was used to generate normalized

767   interaction matrices at a resolution of 40 kb and then TopDom v0.0.2 (Shin et al. 2016) was used with a

768   window size w=10 to identify topological domains, boundaries and gaps. We only considered domains

769   for the analyses in this paper. We considered a promoter capture Hi-C interaction to be "intra-TAD" if

37

770   the entire span of the interaction was fully contained in a single domain. "Inter-TAD" interactions are

771   defined as interactions where each end maps to a different domain.

A/B compartments

773   The program runHiCpca.pl from the HOMER (Heinz et al. 2010) v4.8.3 package was used to

774   call A/B compartments with -res 50000 for both whole-genome and capture Hi-C data.

RNA-seq

776   Total RNA was extracted from flash-frozen pellets of 1 million cells using TRI Reagent (Sigma

777   #T9424) and a homogenizer followed by RNA isolation and clean-up using the Direct-zol RNA Kit

778   (Zymo Research #11-331). RNA-seq libraries were generated with the Illumina TruSeq V2 kit (Illumina,

779   RS-122-2001) and 1µg of RNA, following manufacturer's instructions. Libraries were made from RNA

780   isolated from three independent iPSC-CM differentiations (triplicates of iPSC and of cardiomyocytes).

781   Libraries were sequenced on an Illumina HiSeq 4000.

782   Gene counts were quantified with Salmon 0.7.2 (Patro et al. 2017) and imported with tximport

783   1.2.0 (Soneson et al. 2015) into DESeq2 1.12.4 (Love et al. 2014) to call differentially expressed genes.

784   A minimum 1.5-fold-difference between CMs and iPSC triplicates and a minimum adjusted P-value of

785   0.05 were required to select differentially expressed genes for downstream analyses. TPMs (transcripts

786   per million) were also estimated by Salmon. Because the samples clearly clustered according to their

787   known tissues of origin (Figure 1 – figure supplement 2A), no correction for batch effects was

788   performed.

H3K27ac ChIP-seq for comparison with Epigenome Roadmap samples

790   We performed ChIP-seq on 2.5 million cells each for iPSCs and CMs using H3K27ac antibodies

791   (Wako #306-34849). Briefly, cells were crosslinked with 1% formaldehyde for 10 minutes at room

792   temperature, quenched with 0.2M glycine for 5 minutes, pelleted and snap-frozen in liquid nitrogen.

793   Cells were lysed in Lysis Buffer 1 (50 mM HEPES-KOH, pH 7.5, 140 mM NaCl, 1 mM EDTA, 10%

38

794 glycerol, 0.5% NP-40, 0.25% Triton X-100). Crosslinked chromatin was sheared to an average size of

795 300 bp using a Bioruptor with 30" on/30" off at high setting and then incubated overnight at 4ºC with 1

796 µg antibody. Dynabeads M-280 Sheep Anti-Mouse IgG (ThermoFisher #11201D) were used to pull

797 down chromatin and ChIP DNA was eluted and prepared for sequencing using the NEBNext Ultra II

798 DNA Library prep kit (NEB #E7645S). ChIP-seq reads were aligned with Bowtie 2-2.2.3 (Langmead

799 and Salzberg 2012) and peaks were called with HOMER (Heinz et al. 2010) v4.8.3 on unique reads

800 with mapping quality > 10 using the –region and –style histone parameters. Significant peaks were

801 overlapped with H3K27ac peaks from Epigenome Roadmap samples which demonstrated high

802 concordance between matched tissue types (Figure 1—figure supplement 2C,D). Because we

803 performed a low level of sequencing, we did not identify as many peaks as the Roadmap samples.

804 Therefore we used Roadmap ChIP-seq data in all of our analyses.

805 ***Gene Ontology analysis***

806 The human Gene Ontology (GO) associations of GO terms (Ashburner et al. 2000) to genes

807 and the GO database were downloaded on January 22, 2016 from http://geneontology.org/gene-

808 associations. GO terms were associated with RefSeq genes via gene symbols. Using the GO

809 annotation graph, all parent terms were assigned to the terms annotated to a gene. A hypergeometric

810 test was used to calculate the statistical significance of the difference of the number of genes

811 associated with a given GO term in a particular gene set and the universe of all RefSeq genes ($P <$

812 0.05). P-values were corrected with the R package p.adjust function using the "fdr" method.

813 For two of the GWAS disease groups (heart failure and myocardial infarction), the list of genes

814 looping to LD SNPs included many histone genes. This is because there is a tag SNP located in the

815 middle of a histone gene cluster (containing > 30 histone genes located close together) in each case.

816 After expanding the tag SNP to all SNPs in LD, many of the histone genes in that cluster looped to the

817 LD SNPs, resulting in a high representation of these genes in the final gene list. The resulting Gene

818 Ontology enrichment analysis gave terms relating to nucleosome and chromatin organization because

819   of this over-representation. We therefore chose to remove these genes from the final gene lists of heart

820   failure and myocardial infarction target genes.

821   *Motif analysis*

822   The program findMotifsGenome.pl from the HOMER (Heinz et al. 2010) v4.8.3 package was

823   used with –size given parameter to identify overrepresented motifs in the distal (non-promoter)

824   interacting sequences of promoter interactions. As stated above, this analysis was performed on

825   fragment-level promoter-interacting sequences.

826   *Histone ChIP-seq enrichment analysis*

827   We obtained publicly available ChIP-seq data in the form of processed peak calls for H3K27ac,

828   H3K4me1 and H3K27me3 from the Roadmap Epigenomics Project (Roadmap Epigenomics

829   Consortium et al. 2015), and for CTCF from ENCODE (ENCODE Project Consortium 2012)

830   (Supplementary File 10). We only considered peaks that mapped outside of the captured region of

831   promoters to ensure our results were not driven by the strong peak signal over most promoters. As a

832   proxy for iPSCs, we used data from the H1 embryonic stem cell line and for CMs we used data from

833   Left Ventricle tissue. We grouped genes into five expression categories based on the average TPM

834   values: group 1 (0 TPM), group 2 (TPM 0-3), group 3 (TPM 3-25), group 4 (TPM 25-150) and group 5

835   (TPM >150) and for each group of genes, we calculated the enrichment for promoter interactions to

836   overlap a given feature. To calculate enrichment of interactions overlapping an epigenetic feature, we

837   compared the observed proportion of MboI fragments in significant interactions overlapping a feature to

838   the proportion of random MboI fragments overlapping the feature. Specifically, we randomly selected

839   MboI fragments from a set that excluded fragments mapping within captured regions (promoters) or

840   within unmappable genomic regions (gaps). The number of randomly selected fragments matched the

841   number of interacting fragments considered for the analysis. We performed 100 iterations of

842   overlapping random fragments with a feature and report the average fold-enrichment. We refer to this

40

843   method of enrichment as a "genome-wide" background model because for each gene expression

844   group, the observed proportion of fragments containing a peak is compared to randomly selected

845   fragments from the whole genome.

846   To calculate the correlation between expression and histone ChIP-seq peak density, we

847   calculated the Spearman's rank correlation between the expression value for each gene (the average

848   TPM value) and the number of peaks mapping within 300 kb of each gene TSS. We only considered

849   genes with at least one significant interaction in the respective cell type to allow for generalizations to

850   the enrichment analysis presented in Figure 3.

851   **GWAS analysis**

852   We compiled genome-wide significant SNPs associated with GWAS for cardiac arrhythmia, heart

853   failure, and myocardial infarction from the NHGRI-EBI database (http://www.ebi.ac.uk/gwas/); see

854   Supplementary File 7 for list of terms used to identify specific GWAS. We expanded each set of SNPs

855   to all SNPs in high LD ($r^2 > 0.9$) using phase 3 data of the 1,000 genomes project (Nikpay et al. 2015)

856   (Supplementary File 3). For each lead SNP from the GWAS we analyzed, we selected a 100 kb interval

857   centered on the SNP (SNP +/- 50kb). For each 100 kb interval, Tabix (Li 2011) was used to retrieve

858   genotypes. We then used PLINK (Purcell et al. 2007) v1.90p on phase 3 data from the 1,000 genomes

859   project (Nikpay et al. 2015) (ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502, v5a) to select SNPs

860   in LD ($r^2 > 0.9$) with the tag SNP and a minimum allele frequency of 0.01. We only included the

861   populations primarily studied in the GWASs: CEU (central European), ASW (African American) and

862   JPT (Japanese). We assigned all SNPs in promoter-distal interactions to their interacting gene(s)

863   ("target genes") using cardiomyocyte promoter capture Hi-C data. We did not require the SNP to map

864   to regions associated with open chromatin or enhancer marks as these types of data are highly cell-

865   type specific and we did not wish to exclude SNPs in regions that may be active in non-assayed cell

866   types.

867   We note that one major GWAS for dilated cardiomyopathy was not included in the NHGRI-EBI

868   database (Meder et al. 2014), likely because there is an error obtaining the online methods of the

869   paper. After careful inspection of the study, we concluded that the GWAS met the NHGRI-EBI criteria

870   and included the associations from that study in our analysis. A complete list of all studies used in this

871   analysis can be found in Supplementary File 8.

872   ***MGI analysis***

873   To calculate enrichment of target genes to cause cardiovascular phenotypes when deleted in

874   mice (Mouse Genome Informatics database), we randomly selected 347 genes from the list of starting

875   genes (i.e. genes with at least one promoter-distal interaction in CMs, meaning it could be a target

876   gene), and calculated the proportion that caused a cardiovascular phenotype in mice. We performed

877   this randomized selection for 100 iterations to generate the randomized (expected) values. Random

878   genes were not required to be expressed, as the set of target genes contains genes that are not

879   expressed. P-value was calculated with a Z test.

880   ***eQTL analysis***

881   For eQTLs used in comparisons with GWAS variants and Hi-C interactions, we used the set of

882   GTEx v7 eQTLs identified as significant in the left ventricle of the heart (Carithers et al. 2015). eQTLs

883   were called significant if $q<0.05$ after false discovery rate correction (Storey and Tibshirani 2003). We

884   only considered promoter-distal eQTLs that were at least 10 kb from their associated gene to allow for

885   that eQTL to map to an interaction with it's associated gene.

886   To calculate enrichment for eQTLs to loop to their associated gene, we used a background

887   model whereby each promoter's set of interactions were re-mapped to a different promoter, keeping the

888   distance and strand orientation consistent. We performed this re-mapping of all promoter interactions

889   1000 times and calculated the proportion of all eQTLs that mapped to interactions for their eQTL-

890   associated gene in each permutation. We either used the CM interactions or the iPSC interactions with

891   the same set of left ventricle eQTLs to compare cell-type specificity of the promoter interaction data.

892   **Data Availability:**

893   Raw and processed sequencing data are provided at ArrayExpress through accession numbers E-

894   MTAB-6014 (Hi-C) and E-MTAB-6013 (RNA-seq).

895   **Acknowledgements**

898   **Competing Interests**

899   None

900   **Sources of Funding**

904

905   **References**

906   Arking DE, Pulit SL, Crotti L, van der Harst P, Munroe PB, Koopmann TT, Sotoodehnia N, Rossin EJ,

907   Morley M, Wang X, Johnson AD, Lundby A, Gudbjartsson DF, Noseworthy PA, Eijgelsheim M,

908   Bradford Y, Tarasov K V, Dörr M, Müller-Nurasyid M, et al. 2014. Genetic association study of QT

909   interval highlights role for calcium signaling pathways in myocardial repolarization. *Nat Genet* **46**:

910   826–836.

911   Arnolds DE, Liu F, Fahrenbach JP, Kim GH, Schillinger KJ, Smemo S, McNally EM, Nobrega MA, Patel

912   V V., Moskowitz IP. 2012. TBX5 drives Scn5a expression to regulate cardiac conduction system

913   function. *J Clin Invest* **122**: 2509–2518.

914 Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS,

915   Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE,

916   Ringwald M, Rubin GM, et al. 2000. Gene ontology: tool for the unification of biology. The Gene

917   Ontology Consortium. *Nat Genet* **25**: 25–9.

918 Banovich NE, Li YI, Raj A, Ward MC, Greenside P, Calderon D, Tung PY, Burnett JE, Myrthil M,

919   Thomas SM, Burrows CK, Romero IG, Pavlovic BJ, Kundaje A, Pritchard JK, Gilad Y. 2018.

920   Impact of regulatory variation across human iPSCs and differentiated cells. *Genome Res* **28**: 122–

921   131.

922 Blake JA, Eppig JT, Kadin JA, Richardson JE, Smith CL, Bult CJ, the Mouse Genome Database Group.

923   2017. Mouse Genome Database (MGD)-2017: community knowledge resource for the laboratory

924   mouse. *Nucleic Acids Res* **45**: D723–D729.

925 Burridge PW, Matsa E, Shukla P, Lin ZC, Churko JM, Ebert AD, Lan F, Diecke S, Huber B, Mordwinkin

926   NM, Plews JR, Abilez OJ, Cui B, Gold JD, Wu JC. 2014. Chemically defined generation of human

927   cardiomyocytes. *Nat Methods* **11**: 855–60.

928 Cai C-L, Zhou W, Yang L, Bu L, Qyang Y, Zhang X, Li X, Rosenfeld MG, Chen J, Evans S. 2005. T-box

929   genes coordinate regional rates of proliferation and regional specification during cardiogenesis.

930   *Development* **132**: 2475–2487.

931 Cairns J, Freire-Pritchett P, Wingett SW, Várnai C, Dimond A, Plagnol V, Zerbino D, Schoenfelder S,

932   Javierre B-M, Osborne C, Fraser P, Spivakov M. 2016. CHiCAGO: robust detection of DNA

933   looping interactions in Capture Hi-C data. *Genome Biol* **17**: 127.

934 Calo E, Wysocka J. 2013. Modification of enhancer chromatin: what, how, and why? *Mol Cell* **49**: 825–

935   37.

936   Carithers LJ, Ardlie K, Barcus M, Branton PA, Britton A, Buia SA, Compton CC, DeLuca DS, Peter-

937   Demchok J, Gelfand ET, Guan P, Korzeniewski GE, Lockhart NC, Rabiner CA, Rao AK, Robinson

938   KL, Roche N V., Sawyer SJ, Segrè A V., et al. 2015. A Novel Approach to High-Quality

939   Postmortem Tissue Procurement: The GTEx Project. *Biopreserv Biobank* **13**: 311–319.

940   Claussnitzer M, Dankel SN, Kim K-H, Quon G, Meuleman W, Haugen C, Glunk V, Sousa IS, Beaudry

941   JL, Puviindran V, Abdennur NA, Liu J, Svensson P-A, Hsu Y-H, Drucker DJ, Mellgren G, Hui C-C,

942   Hauner H, Kellis M. 2015. *FTO* Obesity Variant Circuitry and Adipocyte Browning in Humans. *N

943   Engl J Med* **373**: 895–907.

944   Cowper-Sal·lari R, Zhang X, Wright JB, Bailey SD, Cole MD, Eeckhoute J, Moore JH, Lupien M. 2012.

945   Breast cancer risk–associated SNPs modulate the affinity of chromatin for FOXA1 and alter gene

946   expression. *Nat Genet* **44**: 1191–1198.

947   Creyghton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, Hanna J, Lodato MA,

948   Frampton GM, Sharp PA, Boyer LA, Young RA, Jaenisch R. 2010. Histone H3K27ac separates

949   active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci* **107**: 21931–

950   21936.

951   Dao LTM, Galindo-Albarr?n AO, Castro-Mondragon JA, Andrieu-Soler C, Medina-Rivera A, Souaid C,

952   Charbonnier G, Griffon A, Vanhille L, Stephen T, Alomairi J, Martin D, Torres M, Fernandez N,

953   Soler E, van Helden J, Puthier D, Spicuglia S. 2017. Genome-wide characterization of mammalian

954   promoters with distal enhancer functions. *Nat Genet* **49**: 1073–1081.

955   Dekker J, Mirny L. 2016. The 3D Genome as Moderator of Chromosomal Communication. *Cell* **164**:

956   1110–1121.

957   Deng W, Lee J, Wang H, Miller J, Reik A, Gregory PD, Dean A, Blobel G a. 2012. Controlling long-

45

958   range genomic interactions at a native locus by targeted tethering of a looping factor. *Cell* **149**:

959   1233–1244.

960   Diao Y, Fang R, Li B, Meng Z, Yu J, Qiu Y, Lin KC, Huang H, Liu T, Marina RJ, Jung I, Shen Y, Guan

961   K-L, Ren B. 2017. A tiling-deletion-based genetic screen for cis-regulatory element identification in

962   mammalian cells. *Nat Methods* **14**: 629–635.

963   Dixon JR, Jung I, Selvaraj S, Shen Y, Antosiewicz-Bourget JE, Lee AY, Ye Z, Kim A, Rajagopal N, Xie

964   W, Diao Y, Liang J, Zhao H, Lobanenkov V V, Ecker JR, Thomson JA, Ren B. 2015. Chromatin

965   architecture reorganization during stem cell differentiation. *Nature* **518**.

966   Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B. 2012. Topological domains in

967   mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**: 376–380.

968   ENCODE Project Consortium TEP. 2012. An integrated encyclopedia of DNA elements in the human

969   genome. *Nature* **489**: 57–74.

970   Erceg J, Pakozdi T, Marco-Ferreres R, Ghavi-Helm Y, Girardot C, Bracken AP, Furlong EEM. 2017.

971   Dual functionality of cis-regulatory elements as developmental enhancers and Polycomb response

972   elements. *Genes Dev* **31**: 590–602.

973   Franke M, Ibrahim DM, Andrey G, Schwarzer W, Heinrich V, Schöpflin R, Kraft K, Kempfer R, Jerković

974   I, Chan W-L, Spielmann M, Timmermann B, Wittler L, Kurth I, Cambiaso P, Zuffardi O, Houge G,

975   Lambie L, Brancati F, et al. 2016. Formation of new chromatin domains determines pathogenicity

976   of genomic duplications. *Nature*.

977   Freire-Pritchett P, Schoenfelder S, Várnai C, Wingett SW, Cairns J, Collier AJ, García-Vílchez R,

978   Furlan-Magaril M, Osborne CS, Fraser PJ, Rugg-Gunn PJ, Spivakov M. 2017. Global

979   reorganisation of *cis* -regulatory units upon lineage commitment of human embryonic stem cells.

980   *Elife* **6**: e21926.

981 Ghavi-Helm Y, Klein FA, Pakozdi T, Ciglar L, Noordermeer D, Huber W, Furlong EEM. 2014. Enhancer

982   loops appear stable during development and are associated with paused polymerase. *Nature* **512**:

983   96.

984 Gherghiceanu M, Barad L, Novak A, Reiter I, Itskovitz-Eldor J, Binah O, Popescu LM. 2011.

985   Cardiomyocytes derived from human embryonic and induced pluripotent stem cells: comparative

986   ultrastructure. *J Cell Mol Med* **15**: 2539–2551.

987 Gilbert N, Boyle S, Fiegler H, Woodfine K, Carter NP, Bickmore WA. 2004. Chromatin Architecture of

988   the Human Genome: Gene-Rich Domains Are Enriched in Open Chromatin Fibers. *Cell* **118**: 555–

989   566.

990 Gnirke A, Melnikov A, Maguire J, Rogov P, LeProust EM, Brockman W, Fennell T, Giannoukos G,

991   Fisher S, Russ C, Gabriel S, Jaffe DB, Lander ES, Nusbaum C. 2009. Solution hybrid selection

992   with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nat Biotechnol* **27**:

993   182–9.

994 Guo D-C, Papke CL, Tran-Fadulu V, Regalado ES, Avidan N, Johnson RJ, Kim DH, Pannu H, Willing

995   MC, Sparks E, Pyeritz RE, Singh MN, Dalman RL, Grotta JC, Marian AJ, Boerwinkle EA, Frazier

996   LQ, LeMaire SA, Coselli JS, et al. 2009. Mutations in smooth muscle alpha-actin (ACTA2) cause

997   coronary artery disease, stroke, and Moyamoya disease, along with thoracic aortic disease. *Am J*

998   *Hum Genet* **84**: 617–27.

999 Heintzman ND, Hon GC, Hawkins RD, Kheradpour P, Stark A, Harp LF, Ye Z, Lee LK, Stuart RK,

1000   Ching CW, Ching KA, Antosiewicz-Bourget JE, Liu H, Zhang X, Green RD, Lobanenkov V V,

1001   Stewart R, Thomson JA, Crawford GE, et al. 2009. Histone modifications at human enhancers

1002   reflect global cell-type-specific gene expression. *Nature* **459**: 108–12.

1003 Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK.

1004   2010. Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory

1005   Elements Required for Macrophage and B Cell Identities. *Mol Cell* **38**: 576–589.

1006   Hnisz D, Shrinivas K, Young RA, Chakraborty AK, Sharp PA. 2017. Perspective A Phase Separation

1007   Model for Transcriptional Control. *Cell* **169**: 13–23.

1008   Javierre BM, Sewitz S, Cairns J, Wingett SW, Várnai C, Thiecke MJ, Freire-Pritchett P, Spivakov M,

1009   Fraser P, Burren OS, Cutler AJ, Todd JA, Wallace C, Wilder SP, Kreuzhuber R, Kostadima M,

1010   Zerbino DR, Stegle O, Kreuzhuber R, et al. 2016. Lineage-Specific Genome Architecture Links

1011   Enhancers and Non-coding Disease Variants to Target Gene Promoters. *Cell* **167**: 1369–

1012   1384.e19.

1013   Jin F, Li Y, Dixon JR, Selvaraj S, Ye Z, Lee AY, Yen C-A, Schmitt AD, Espinoza C a, Ren B. 2013. A

1014   high-resolution map of the three-dimensional chromatin interactome in human cells. *Nature* **503**:

1015   290–4.

1016   Karakikes I, Ameen M, Termglinchan V, Wu JC. 2015. Human Induced Pluripotent Stem Cell–Derived

1017   Cardiomyocytes. *Circ Res* **117**.

1018   Karlić R, Chung H-R, Lasserre J, Vlahovicek K, Vingron M. 2010. Histone modification levels are

1019   predictive for gene expression. *Proc Natl Acad Sci U S A* **107**: 2926–31.

1020   Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**: 357–359.

1021   Li H. 2011. Tabix: fast retrieval of sequence features from generic TAB-delimited files. *Bioinformatics*

1022   **27**: 718–9.

1023   Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR,

1024   Sabo PJ, Dorschner MO, Sandstrom R, Bernstein B, Bender MA, Groudine M, Gnirke A,

1025   Stamatoyannopoulos J, Mirny LA, Lander ES, Dekker J. 2009. Comprehensive mapping of long-

1026   range interactions reveals folding principles of the human genome. *Science* **326**: 289–93.

1027   Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq

1028   data with DESeq2. *Genome Biol* **15**: 550.

1029   Lupiáñez DG, Kraft K, Heinrich V, Krawitz P, Brancati F, Klopocki E, Horn D, Kayserili H, Opitz JM,

1030   Laxova R, Santos-Simarro F, Gilbert-Dussardier B, Wittler L, Borschiwer M, Haas SA, Osterwalder

1031   M, Franke M, Timmermann B, Hecht J, et al. 2015. Disruptions of Topological Chromatin Domains

1032   Cause Pathogenic Rewiring of Gene-Enhancer Interactions. *Cell* **161**: 1012–1025.

1033   Mahmoud AI, Kocabas F, Muralidhar SA, Kimura W, Koura AS, Thet S, Porrello ER, Sadek HA. 2013.

1034   Meis1 regulates postnatal cardiomyocyte cell cycle arrest. *Nature* **497**: 249–253.

1035   Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, Reynolds AP, Sandstrom R, Qu

1036   H, Brody J, Shafer A, Neri F, Lee K, Kutyavin T, Stehling-Sun S, Johnson AK, Canfield TK, Giste

1037   E, Diegel M, et al. 2012. Systematic Localization of Common Disease-Associate Variation in

1038   Regulatorty DNA. *Science (80-)* **337**: 1190–1195.

1039   Meder B, Rühle F, Weis T, Homuth G, Keller A, Franke J, Peil B, Lorenzo Bermejo J, Frese K, Huge A,

1040   Witten A, Vogel B, Haas J, Völker U, Ernst F, Teumer A, Ehlermann P, Zugck C, Friedrichs F, et

1041   al. 2014. A genome-wide association study identifies 6p21 as novel risk locus for dilated

1042   cardiomyopathy. *Eur Heart J* **35**: 1069–1077.

1043   Miele A, Dekker J. 2008. Long-range chromosomal interactions and gene regulation. *Mol Biosyst* **4**:

1044   1046.

1045   Mifsud B, Tavares-Cadete F, Young AN, Sugar R, Schoenfelder S, Ferreira L, Wingett SW, Andrews S,

1046   Grey W, Ewels P a, Herman B, Happe S, Higgs A, LeProust E, Follows G a, Fraser P, Luscombe

1047   NM, Osborne CS. 2015. Mapping long-range promoter contacts in human cells with high-

1048   resolution capture Hi-C. *Nat Genet* **47**: 598–606.

1049   Moshal K, Roder K, Werdich AA, Dural NT, Kim TY, Cooper LL, Lu YC, Choi B-R, Terentyev D,

1050   MacRae C, Koren G. 2017. LITAF, A Novel Regulator of Cardiac Excitation. *FASEB J* **31**: 686.3-

1051   686.3.

1052   Mumbach MR, Satpathy AT, Boyle EA, Dai C, Gowen BG, Cho SW, Nguyen ML, Rubin AJ, Granja JM,

1053   Kazane KR, Wei Y, Nguyen T, Greenside PG, Corces MR, Tycko J, Simeonov DR, Suliman N, Li

1054   R, Xu J, et al. 2017. Enhancer connectome in primary human cells identifies target genes of

1055   disease-associated DNA elements. *Nat Genet* **49**: 1602–1612.

1056   Musunuru K, Strong A, Frank-Kamenetsky M, Lee NE, Ahfeldt T, Sachs K V, Li X, Li H, Kuperwasser

1057   N, Ruda VM, Pirruccello JP, Muchmore B, Prokunina-Olsson L, Hall JL, Schadt EE, Morales CR,

1058   Lund-Katz S, Phillips MC, Wong J, et al. 2010. From noncoding variant to phenotype via SORT1 at

1059   the 1p13 cholesterol locus. *Nature* **466**: 714–9.

1060   Nikpay M, Goel A, Won H-H, Hall LM, Willenborg C, Kanoni S, Saleheen D, Kyriakou T, Nelson CP,

1061   Hopewell JC, Webb TR, Zeng L, Dehghan A, Alver M, Armasu SM, Auro K, Bjonnes A, Chasman

1062   DI, Chen S, et al. 2015. A comprehensive 1000 Genomes–based genome-wide association meta-

1063   analysis of coronary artery disease. *Nat Genet* **47**: 1121–1130.

1064   Nora EP, Goloborodko A, Valton AL, Gibcus JH, Uebersohn A, Abdennur N, Dekker J, Mirny LA,

1065   Bruneau BG. 2017. Targeted Degradation of CTCF Decouples Local Insulation of Chromosome

1066   Domains from Genomic Compartmentalization. *Cell* **169**: 930–944.e22.

1067   Nora EP, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, Servant N, Piolot T, van Berkum NL, Meisig J,

1068   Sedat J, Gribnau J, Barillot E, Blüthgen N, Dekker J, Heard E. 2012. Spatial partitioning of the

1069   regulatory landscape of the X-inactivation centre. *Nature* **485**: 381–385.

1070   O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, Rajput B, Robbertse B, Smith-

1071   White B, Ako-Adjei D, Astashyn A, Badretdin A, Bao Y, Blinkova O, Brover V, Chetvernin V, Choi

1072   J, Cox E, Ermolaeva O, et al. 2016. Reference sequence (RefSeq) database at NCBI: current

1073   status, taxonomic expansion, and functional annotation. *Nucleic Acids Res* **44**: D733-45.

1074   Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. 2017. Salmon provides fast and bias-aware

1075   quantification of transcript expression. *Nat Methods* **14**: 417–419.

1076   Pennacchio LA, Bickmore W, Dean A, Nobrega MA, Bejerano G. 2013. Enhancers: five essential

1077   questions. *Nat Rev Genet* **14**: 288–295.

1078   Petersen CM, Nielsen MS, Nykjaer A, Jacobsen L, Tommerup N, Rasmussen HH, Roigaard H,

1079   Gliemann J, Madsen P, Moestrup SK. 1997. Molecular identification of a novel candidate sorting

1080   receptor purified from human brain by receptor-associated protein affinity chromatography. *J Biol*

1081   *Chem* **272**: 3599–605.

1082   Phanstiel DH, Van Bortle K, Spacek D, Hess GT, Shamim MS, Machol I, Love MI, Aiden EL, Bassik

1083   MC, Snyder MP. 2017. Static and Dynamic DNA Loops form AP-1-Bound Activation Hubs during

1084   Macrophage Development. *Mol Cell* **67**: 1037–1048.e6.

1085   Phillips-Cremins JE, Sauria MEG, Sanyal A, Gerasimova TI, Lajoie BR, Bell JSK, Ong C-T, Hookway

1086   TA, Guo C, Sun Y, Bland MJ, Wagstaff W, Dalton S, McDevitt TC, Sen R, Dekker J, Taylor J,

1087   Corces VG. 2013. Architectural Protein Subclasses Shape 3D Organization of Genomes during

1088   Lineage Commitment. *Cell* **153**: 1281–1295.

1089   Phillips JE, Corces VG. 2009. CTCF: master weaver of the genome. *Cell* **137**: 1194–211.

1090   Pikkarainen S, Tokola H, Kerkelä R, Ruskoaho H. 2004. GATA transcription factors in the developing

1091   and adult heart. *Cardiovasc Res* **63**: 196–207.

1092   Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker

1093   PIW, Daly MJ, Sham PC. 2007. PLINK: a tool set for whole-genome association and population-

1094   based linkage analyses. *Am J Hum Genet* **81**: 559–75.

1095   Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features.

1096   *Bioinformatics* **26**: 841–2.

51

1097 Rada-Iglesias A, Bajpai R, Swigut T, Brugmann SA, Flynn RA, Wysocka J. 2011. A unique chromatin

1098   signature uncovers early developmental enhancers in humans. *Nature* **470**: 279–283.

1099 Rao SSP, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I,

1100   Omer AD, Lander ES, Aiden EL. 2014. A 3D Map of the Human Genome at Kilobase Resolution

1101   Reveals Principles of Chromatin Looping. *Cell* **159**: 1665–1680.

1102 Roadmap Epigenomics Consortium A, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-

1103   Moussavi A, Kheradpour P, Zhang Z, Wang J, Ziller MJ, Amin V, Whitaker JW, Schultz MD, Ward

1104   LD, Sarkar A, Quon G, Sandstrom RS, Eaton ML, et al. 2015. Integrative analysis of 111 reference

1105   human epigenomes. *Nature* **518**: 317–30.

1106 Rubin AJ, Barajas BC, Furlan-Magaril M, Lopez-Pajares V, Mumbach MR, Howard I, Kim DS, Boxer

1107   LD, Cairns J, Spivakov M, Wingett SW, Shi M, Zhao Z, Greenleaf WJ, Kundaje A, Snyder M,

1108   Chang HY, Fraser P, Khavari PA. 2017. Lineage-specific dynamic and pre-established enhancer–

1109   promoter contacts cooperate in terminal differentiation. *Nat Genet.*

1110 Sakabe NJ, Aneas I, Shen T, Shokri L, Park S-Y, Bulyk ML, Evans SM, Nobrega MA. 2012. Dual

1111   transcriptional activator and repressor roles of TBX20 regulate adult cardiac structure and function.

1112   *Hum Mol Genet* **21**: 2194–2204.

1113 Schmitt AD, Hu M, Jung I, Xu Z, Qiu Y, Tan CL, Li Y, Lin S, Lin Y, Barr CL, Ren B. 2016. A

1114   Compendium of Chromatin Contact Maps Reveals Spatially Active Regions in the Human

1115   Genome. *Cell Rep* **17**: 2042–2059.

1116 Schoenfelder S, Furlan-Magaril M, Mifsud B, Tavares-Cadete F, Sugar R, Javierre B-M, Nagano T,

1117   Katsman Y, Sakthidevi M, Wingett SW, Dimitrova E, Dimond A, Edelman LB, Elderkin S, Tabbada

1118   K, Darbo E, Andrews S, Herman B, Higgs A, et al. 2015. The pluripotent regulatory circuitry

1119   connecting promoters to their long-range interacting elements. *Genome Res* **25**: 582–597.

1120 Shen T, Aneas I, Sakabe N, Dirschinger RJ, Wang G, Smemo S, Westlund JM, Cheng H, Dalton N, Gu

1121 Y, Boogerd CJ, Cai C, Peterson K, Chen J, Nobrega MA, Evans SM. 2011. Tbx20 regulates a

1122 genetic program essential to adult mouse cardiomyocyte function. *J Clin Invest* **121**: 4640–54.

1123 Shin H, Shi Y, Dai C, Tjong H, Gong K, Alber F, Zhou XJ. 2016. TopDom: an efficient and deterministic

1124 method for identifying topological domains in genomes. *Nucleic Acids Res* **44**: e70.

1125 Siersbæk R, Madsen JGS, Javierre BM, Nielsen R, Bagge EK, Cairns J, Wingett SW, Traynor S,

1126 Spivakov M, Fraser P, Mandrup S. 2017. Dynamic Rewiring of Promoter-Anchored Chromatin

1127 Loops during Adipocyte Differentiation. *Mol Cell* **66**: 420–435.e5.

1128 Smemo S, Campos LC, Moskowitz IP, Krieger JE, Pereira AC, Nobrega MA. 2012. Regulatory variation

1129 in a TBX5 enhancer leads to isolated congenital heart disease. *Hum Mol Genet* **21**: 3255–63.

1130 Smemo S, Tena JJ, Kim K-H, Gamazon ER, Sakabe NJ, Gómez-Marín C, Aneas I, Credidio FL,

1131 Sobreira DR, Wasserman NF, Lee JH, Puviindran V, Tam D, Shen M, Son JE, Vakili NA, Sung H-

1132 K, Naranjo S, Acemel RD, et al. 2014. Obesity-associated variants within FTO form long-range

1133 functional connections with IRX3. *Nature* **507**: 371–5.

1134 Smith NL, Felix JF, Morrison AC, Demissie S, Glazer NL, Loehr LR, Cupples LA, Dehghan A, Lumley

1135 T, Rosamond WD, Lieb W, Rivadeneira F, Bis JC, Folsom AR, Benjamin E, Aulchenko YS,

1136 Haritunians T, Couper D, Murabito J, et al. 2010. Association of Genome-Wide Variation With the

1137 Risk of Incident Heart Failure in Adults of European and African Ancestry: A Prospective Meta-

1138 Analysis From the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE)

1139 Consortium. *Circ Cardiovasc Genet* **3**: 256–266.

1140 Soneson C, Love MI, Robinson MD. 2015. Differential analyses for RNA-seq: transcript-level estimates

1141 improve gene-level inferences. *F1000Research* **4**: 1521.

1142 Speir ML, Zweig AS, Rosenbloom KR, Raney BJ, Paten B, Nejad P, Lee BT, Learned K, Karolchik D,

1143   Hinrichs AS, Heitner S, Harte RA, Haeussler M, Guruvadoo L, Fujita PA, Eisenhart C, Diekhans M,

1144   Clawson H, Casper J, et al. 2016. The UCSC Genome Browser database: 2016 update. *Nucleic*

1145   *Acids Res* **44**: D717–D725.

1146   Spitz F, Furlong EEM. 2012. Transcription factors: from enhancer binding to developmental control. *Nat*

1147   *Rev Genet* **13**: 613–626.

1148   Stevens TJ, Lando D, Basu S, Atkinson LP, Cao Y, Lee SF, Leeb M, Wohlfahrt KJ, Boucher W,

1149   O'Shaughnessy-Kirwan A, Cramard J, Faure AJ, Ralser M, Blanco E, Morey L, Sansó M, Palayret

1150   MGS, Lehner B, Di Croce L, et al. 2017. 3D structures of individual mammalian genomes studied

1151   by single-cell Hi-C. *Nature* **544**: 59–64.

1152   Storey JD, Tibshirani R. 2003. Statistical significance for genomewide studies. *Proc Natl Acad Sci U S*

1153   *A* **100**: 9440–5.

1154   Symmons O, Pan L, Remeseiro S, Aktas T, Klein F, Huber W, Spitz F. 2016. The Shh Topological

1155   Domain Facilitates the Action of Remote Enhancers by Reducing the Effects of Genomic

1156   Distances. *Dev Cell* **39**: 529–543.

1157   Tsujimura T, Klein FA, Langenfeld K, Glaser J, Huber W, Spitz F. 2015. A Discrete Transition Zone

1158   Organizes the Topological and Regulatory Autonomy of the Adjacent Tfap2c and Bmp7 Genes ed.

1159   W.A. Bickmore. *PLoS Genet* **11**: e1004897.

1160   Visel A, Minovitsky S, Dubchak I, Pennacchio LA. 2007. VISTA Enhancer Browser--a database of

1161   tissue-specific human enhancers. *Nucleic Acids Res* **35**: D88-92.

1162   Watt AJ, Battle MA, Li J, Duncan SA. 2004. GATA4 is essential for formation of the proepicardium and

1163   regulates cardiogenesis. *Proc Natl Acad Sci* **101**: 12573–12578.

1164   Wingett S, Ewels P, Furlan-Magaril M, Nagano T, Schoenfelder S, Fraser P, Andrews S. 2015. HiCUP:

1165   pipeline for mapping and processing Hi-C data. *F1000Research* **4**: 1310.

1166   Wright JB, Brown SJ, Cole MD. 2010. Upregulation of c-MYC in cis through a Large Chromatin Loop

1167   Linked to a Cancer Risk-Associated Single-Nucleotide Polymorphism in Colorectal Cancer Cells.

1168   *Mol Cell Biol* **30**: 1411–1420.
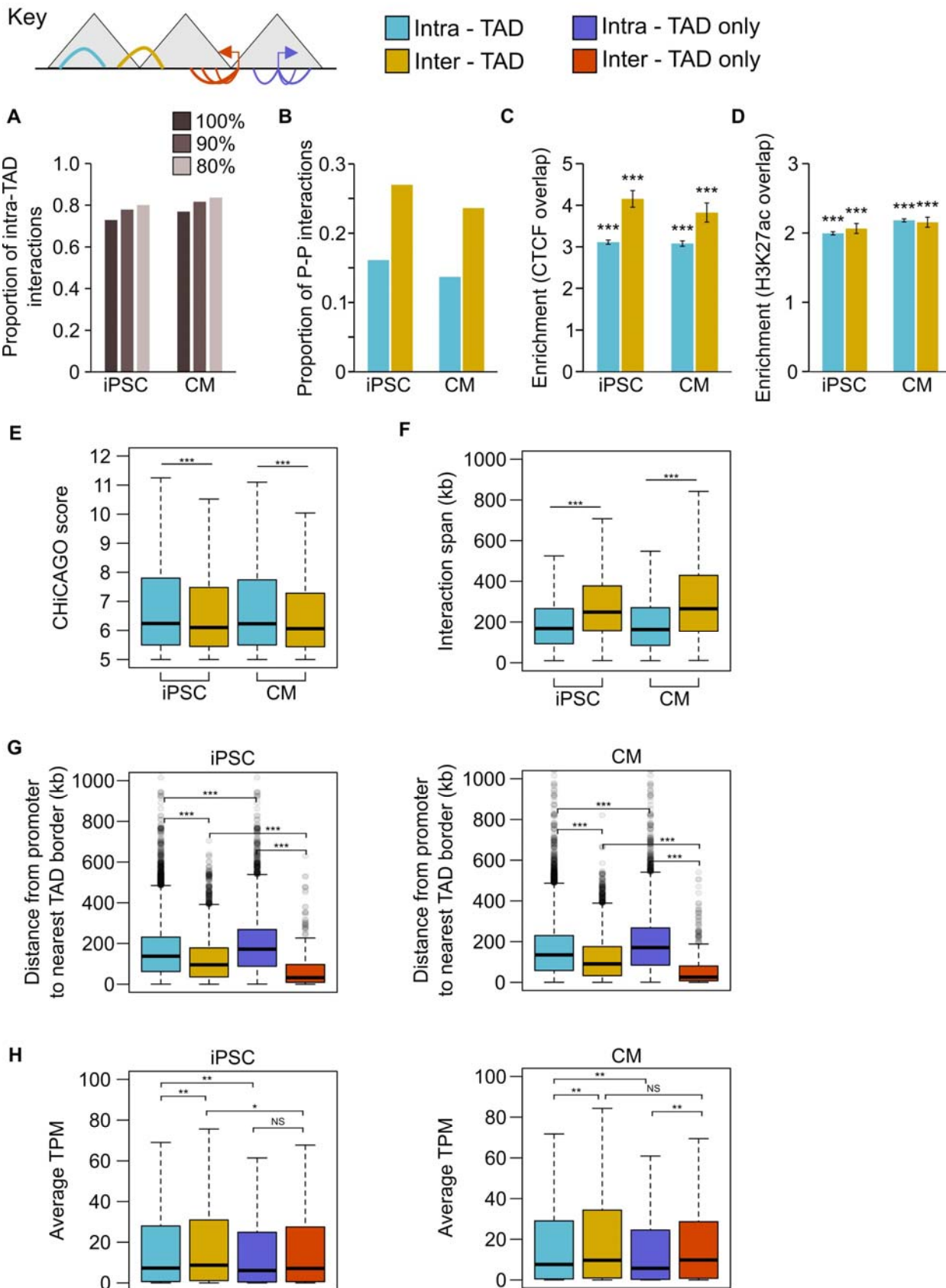
Figure 1 - Figure Supplement 1
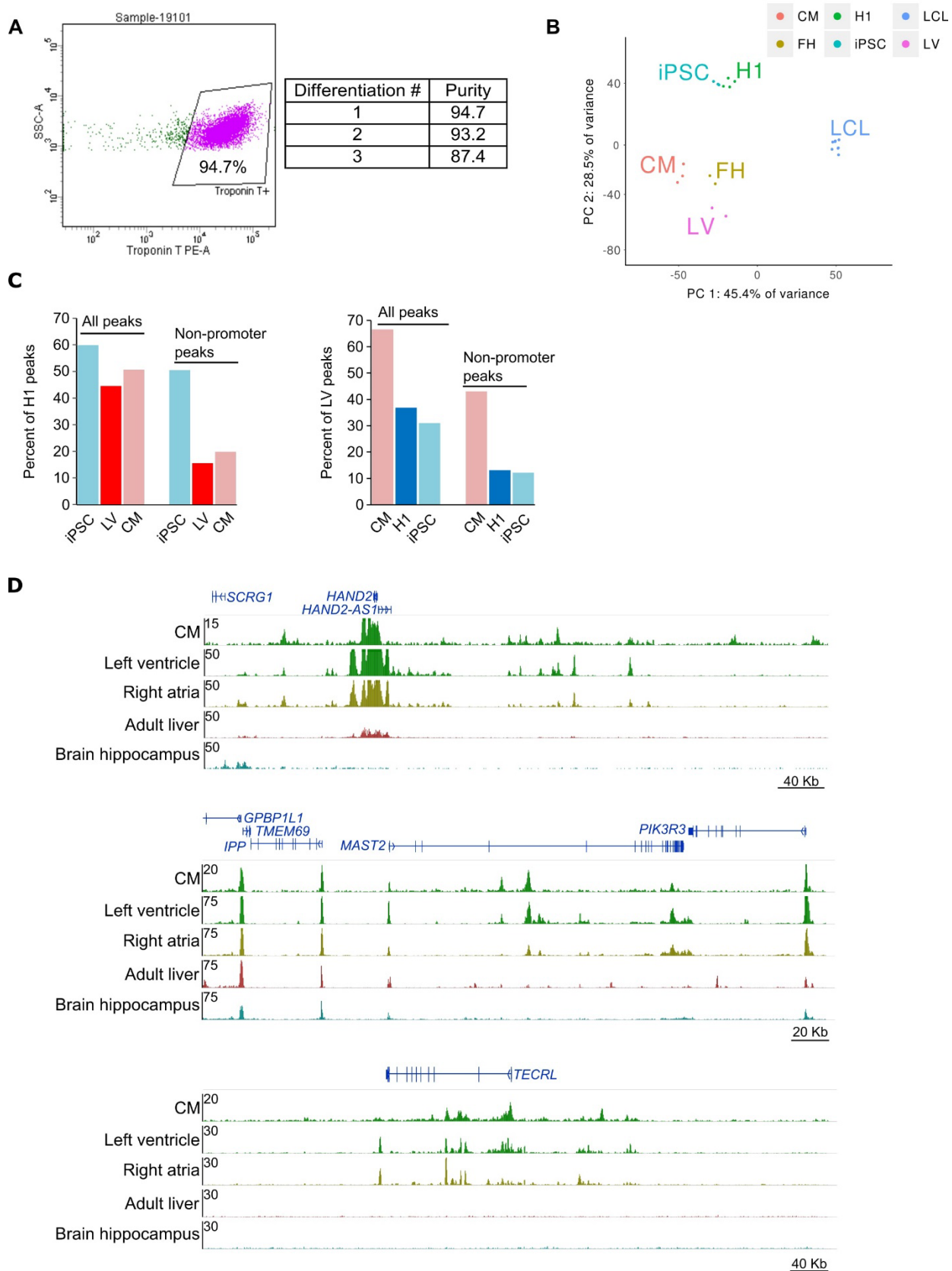
1170 **Figure 1—figure supplement 1**. **(A)** Flow cytometry of iPSC-derived cardiomyocytes. Representative

1171 image of flow data for cardiomyocytes (left) and percent cardiac troponin T (cTnT) positive for each

1172 differentiation (right). Cells were first gated on live/dead and then on cTnT staining. **(B)** Principle

1173 component analysis of RNA-seq data in iPSCs and CMs along with H1 embryonic stem cells, left

1174 ventricular cells (LV), fetal heart cells (FH), and lymphoblastoid cell line cells (LCL). LCLs cluster

1175 independently from iPSC and CM, indicating that iPSCs were faithfully reprogrammed. **(C)** Percentage

1176 of Epigenome Roadmap H3K27ac ChIP-seq peaks overlapping iPSC and CM H3K27ac peaks.

1177 Overlaps for all peaks and only non-promoter peaks are shown. LV, left ventricle; H1, H1 embryonic

1178 stem cell line. **(D)** Three genome browser snap-shots displaying the epigenetic landscape in CMs

1179 compared to left ventricle, right atria, adult liver and brain hippocampus from the Epigenome Roadmap.

Figure 1 - Figure Supplement 2

1180
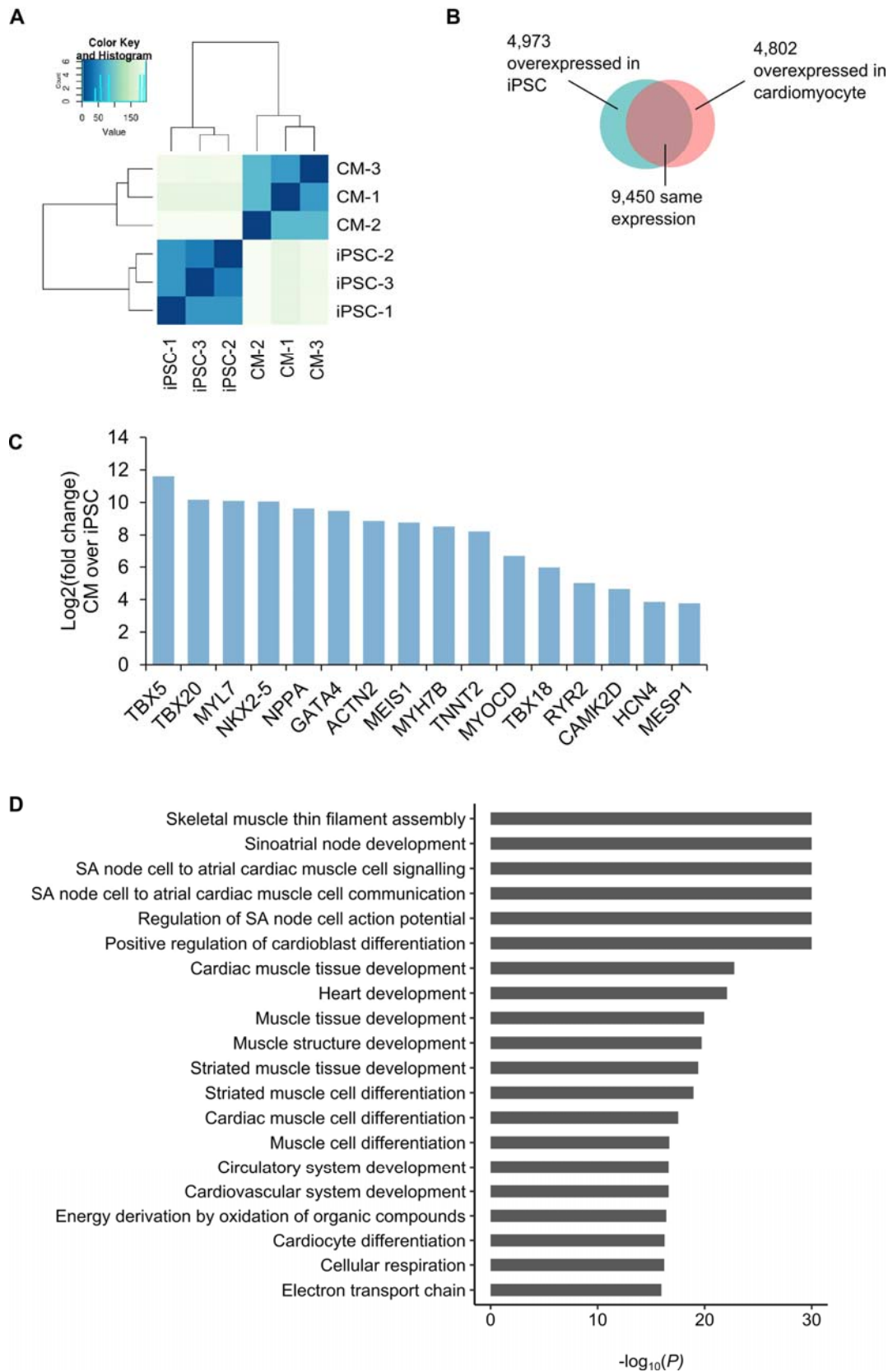
1181 **Figure 1—figure supplement 2. (A)** Cluster analysis of RNA-seq data from each triplicate of iPSC and

1182 CM. **(B)** Number of genes differentially expressed in each cell type. **(C)** Selected genes overexpressed

1183 in CMs relative to iPSCs. **(D)** Gene Ontology enrichment analysis of the biological processes

1184 associated with the 4,802 genes overexpressed in cardiomyocytes.
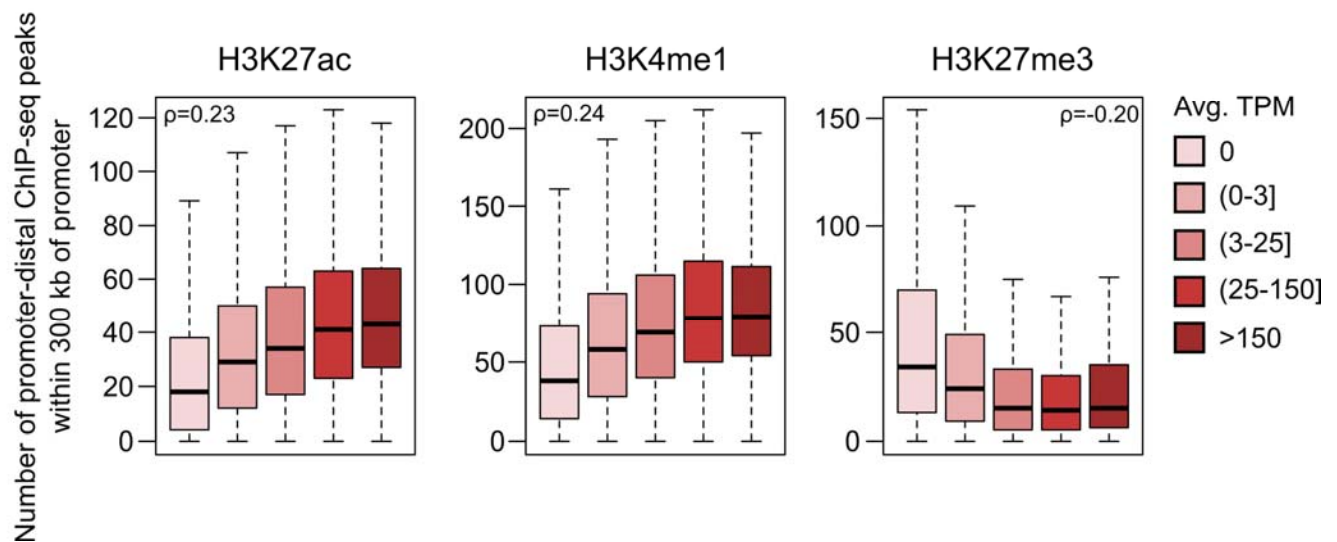
Figure 1 - Figure Supplement 3

1186   **Figure 1—figure supplement 3**. In this analysis, interactions were classified as intra-TAD (both ends

1187   of the interaction fully within a single TAD) or inter-TAD (each end of the interaction is in a different

1188   TAD). Interactions falling partially or wholly within TAD "boundaries" or "gaps" as defined by TopDom

1189   were omitted (see Methods). **(A)** Proportion of interactions that are intra-TAD at different cut-offs. All

1190   analyses used interactions that were 100% within a TAD. **(B)** Proportion of promoter-promoter

1191   interactions in the set of intra-TAD and inter-TAD interactions. **(C,D)** Fold enrichment for intra-TAD and

1192   inter-TAD interactions to overlap CTCF **(C)** or H3K27ac peaks **(D)**. Only promoter-distal ChIP-seq

1193   peaks were analyzed. ***$P < 2.2 \times 10^{-16}$, Z-test. **(E)** CHiCAGO score and **(F)** interaction span of intra- vs.

1194   inter-TAD interactions. ***$P < 2.2 \times 10^{-16}$, Wilcoxon rank-sum test. **(G,H)** Considering promoters with an

1195   intra-TAD interaction, an inter-TAD interaction, or exclusively intra-TAD or inter-TAD interactions: **(G)**

1196   distance from the promoter TSS to the nearest TAD boundary and **(H)** average TPM value of the

1197   promoter. ***$P < 2.2 \times 10^{-16}$, **$P < 0.01$, *$P < 0.05$, NS = not significant, Wilcoxon rank-sum test.

61
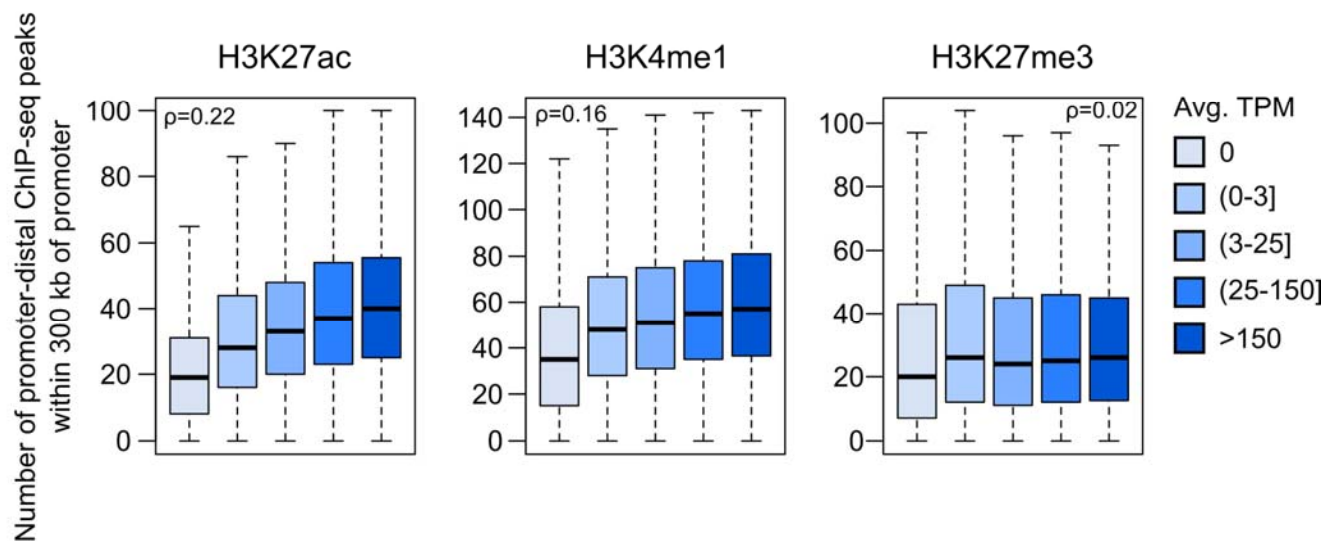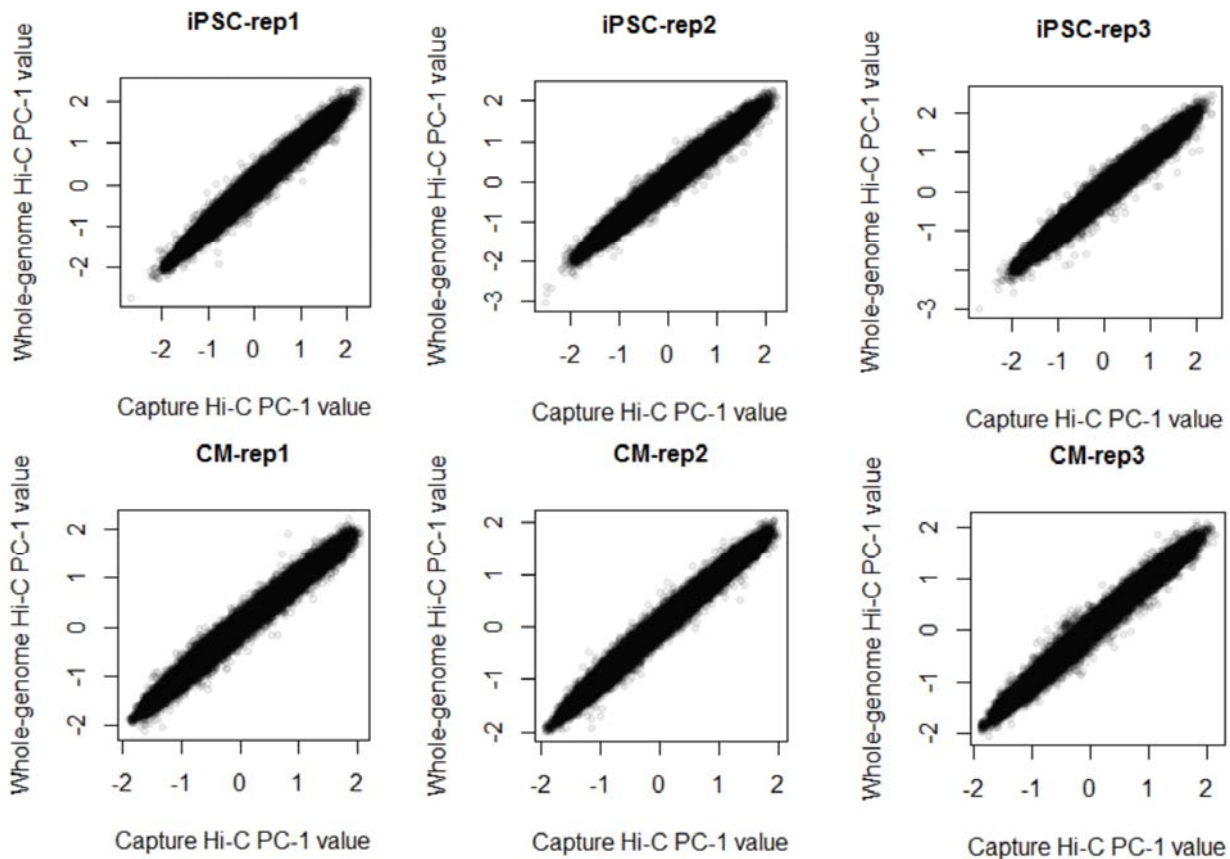
Figure 3 - Figure Supplement 1

Figure 3—figure supplement 1. Number of promoter-distal histone ChIP-seq peaks within 300 kb of promoters in iPSC **(A)** and CM **(B)**. Spearman's rho (ρ) was calculated on the full set of promoter expression values/peak counts for all promoters with at least one significant interaction in the respective cell type (12,926 genes for iPSC and 13,555 genes for CM; see Methods). Data are grouped by expression category to emphasize the trend. Horizontal bars indicate the median for each

1204 expression category. All correlation estimates are significant at $P < 2.2 \times 10^{-16}$ except for H3K27me3 in

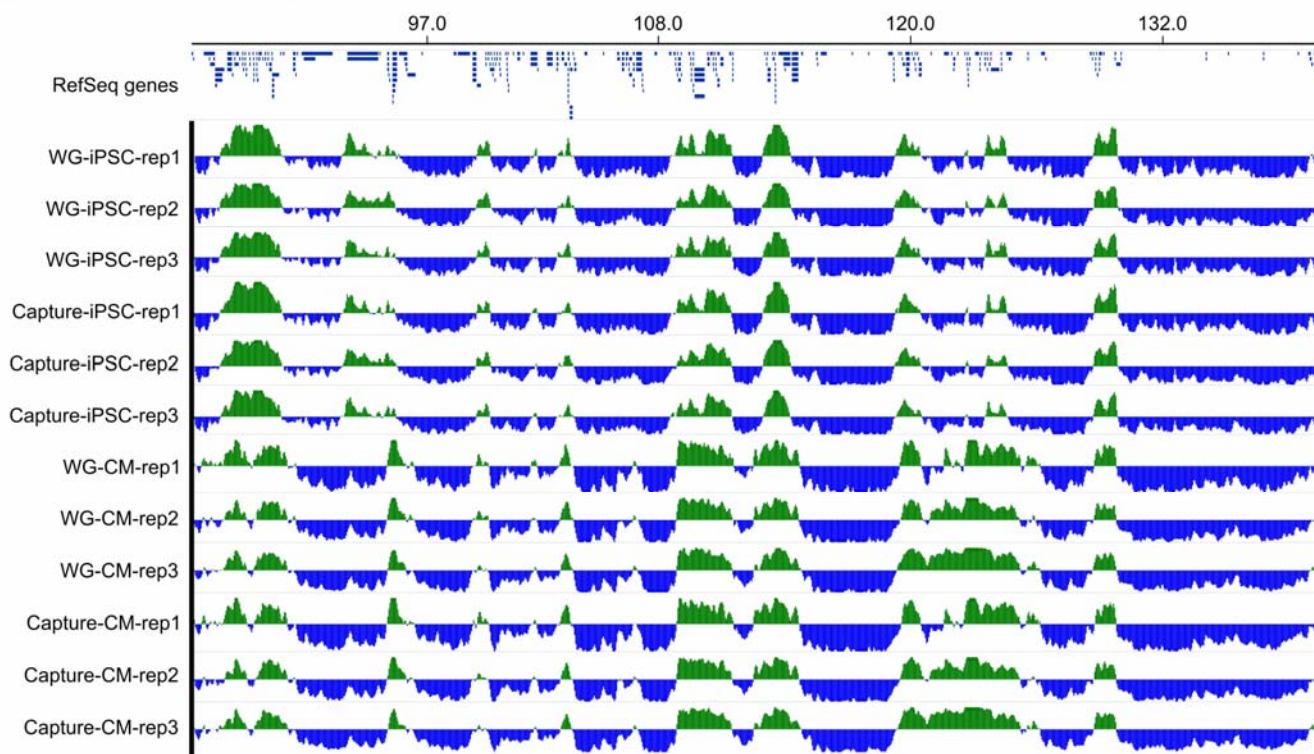1205 iPSCs ($P = 0.06$).

Figure 4 - Figure Supplement 1

1206

1207 **Figure 4—figure supplement 1.** Correlation between the A/B compartment score (principle

1208 component analysis of interaction data, PC-1) in whole-genome Hi-C (y-axis) and promoter capture Hi-

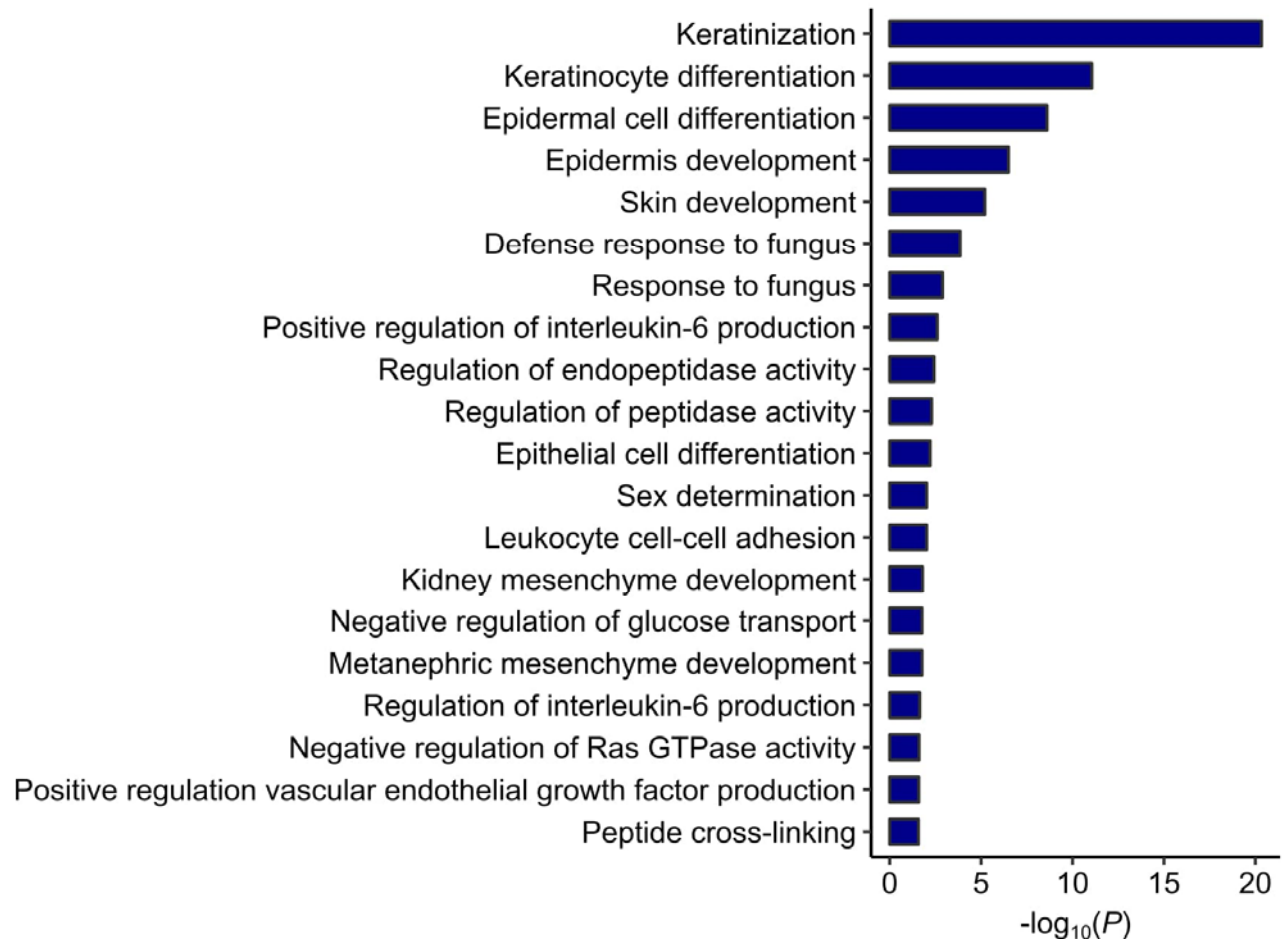1209 C (x-axis) in iPSCs (top) and CMs (bottom). Spearman's ρ > 0.98, $P < 2.2 \times 10^{-16}$ in all cases.

Figure 4 - Figure Supplement 2

Figure 4—figure supplement 2. Genome browser snapshot of a ~53 Mb region on chromosome 4 showing A/B compartments in all three replicates of iPSCs and CMs using both whole-genome (WG) and promoter capture Hi-C data.

Figure 4 - Figure Supplement 3

1214

Figure 4—figure supplement 3. GO analysis on the genes switching from active A compartments in iPSCs to inactive B compartments in CMs.

Supplementary Movie 1 – Video of iPSC-derived cardiomyocytes exhibiting spontaneous beating at day 20 of the differentiation (day of cell harvesting).

Supplementary File Legends

65

1222 **Supplementary File 1 – PCHi-C interactions for iPSC.** Significant interactions in iPSC (identified in at

1223 least two out of three replicates) are presented in paired bed format. Column 7 is the CHiCAGO score;

1224 Column 8 contains the gene information (gene name, identifier, strand, TSS position). If the interaction

1225 involves another promoter, the second gene name information is provided after the "|" symbol.

1226 **Supplementary File 2 – PCHi-C interactions for CM.** Significant interactions in CM (identified in at

1227 least two out of three replicates) are presented in paired bed format. Column 7 is the CHiCAGO score;

1228 Column 8 contains the gene information (gene name, identifier, strand, TSS position). If the interaction

1229 involves another promoter, the second gene name information is provided after the "|" symbol.

1230 **Supplementary File 3 – CVD SNPs.** All SNPs in high LD ($r^2 > 0.9$) with CVD tag SNPs are provided.

1231 The first 4 columns indicate the tag SNP position; columns 5-8 indicate the SNPs in LD with the tag

1232 SNP; column 9 is the degree of LD (r^2 value).

1233 **Supplementary File 4 – HOMER motif analysis for the distal interacting regions of promoter**

1234 **interactions**. For each analysis presented in Figure 2, the full output of motifs identified in interacting

1235 fragments are listed.

1236 **Supplementary File 5 – Gene Ontology enrichment output**. The full output of the GO enrichment

1237 analysis is provided.

1238 **Supplementary File 6 – Gene Ontology input gene lists.** The list of genes used as input for GO

1239 analysis is provided.

1240 **Supplementary File 7 – GWAS terms used to compile studies.** The list of trait terms used to filter

1241 GWAS studies is provided.

1242 **Supplementary File 8 – GWAS summary table.** Table 8.1 contains information related to the CVD

1243 GWAS used in this paper, including PubMed ID, first author, date of publication, journal, study title, tag

1244 SNP chromosome position, rsID, trait. Table 8.2 contains information on each LD SNP-target gene

1245 interaction (tag SNP and corresponding LD SNP, target gene, interaction coordinates, target gene

1246 expression in iPSC and CM, MGI cardiovascular phenotype information.

1247   **Supplementary File 9 – Hi-C read information.** Table 9.1 contains the probe sequences used for

1248   promoter capture, along with the corresponding gene name. Table 9.2 contains the total number of

1249   sequenced and processed/filtered reads for each Hi-C experiment. Table 9.3 contains the number of

1250   reads mapping to each promoter and the corresponding number of significant (present in at least two

1251   replicates) interactions called.

1252   **Supplementary File 10 – Public datasets used.** ChIP-seq and RNA-seq datasets used in our

1253   analyses are listed.

1254