Figures and figure supplements

A promoter interaction map for cardiovascular disease genetics

Lindsey E Montefiori et al
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)
Figure 1 continued

(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.

DOI: https://doi.org/10.7554/eLife.35788.003
Figure 1—figure supplement 1. Quality control of iPSC-CMs. (A) Flow cytometry of iPSC-derived cardiomyocytes. Representative image of flow data for cardiomyocytes (left) and percent cardiac troponin T (cTnT) positive for each differentiation (right). Cells were first gated on live/dead and then on
Figure 1—figure supplement 1 continued

cTnT staining. (B) Principle component analysis of RNA-seq data in iPSCs and CMs along with H1 embryonic stem cells, left ventricular cells (LV), fetal heart cells (FH), and lymphoblastoid cell line cells (LCL). LCLs cluster independently from iPSC and CM, indicating that iPSCs were faithfully reprogrammed. (C) Percentage of Epigenome Roadmap H3K27ac ChIP-seq peaks overlapping iPSC and CM H3K27ac peaks. Overlaps for all peaks and only non-promoter peaks are shown. LV, left ventricle; H1, H1 embryonic stem cell line. (D) Three genome browser snap-shots displaying the epigenetic landscape in CMs compared to left ventricle, right atria, adult liver and brain hippocampus from the Epigenome Roadmap.

DOI: https://doi.org/10.7554/eLife.35788.004
Figure 1—figure supplement 2. Analysis of RNA-seq in iPSCs and iPSC-CMs. (A) Cluster analysis of RNA-seq data from each triplicate of iPSC and CM. (B) Number of genes differentially expressed in each cell type. (C) Selected genes overexpressed in CMs relative to iPSCs. (D) Gene Ontology enrichment analysis of the biological processes associated with the 4802 genes overexpressed in cardiomyocytes.

DOI: https://doi.org/10.7554/eLife.35788.005
Analysis of PCHi-C interactions in the context of TADs. In this analysis, interactions were classified as intra-TAD (both ends of the interaction fully within a single TAD) or inter-TAD (each end of the interaction is in a different TAD). Interactions falling partially or wholly

Figure 1—figure supplement 3 continued on next page
within TAD ‘boundaries’ or ‘gaps’ as defined by TopDom were omitted (see Materials and methods). (A) Proportion of interactions that are intra-TAD at different cut-offs. All analyses used interactions that were 100% within a TAD. (B) Proportion of promoter-promoter interactions in the set of intra-TAD and inter-TAD interactions. (C,D) Fold enrichment for intra-TAD and inter-TAD interactions to overlap CTCF (C) or H3K27ac peaks (D). Only promoter-distal ChIP-seq peaks were analyzed. ***p<2.2 \times 10^{-16}, Z-test. (E) CHiCAGO score and (F) interaction span of intra-vs. inter-TAD interactions. ***p<2.2 \times 10^{-16}, Wilcoxon rank-sum test. (G,H) Considering promoters with an intra-TAD interaction, an inter-TAD interaction, or exclusively intra-TAD or inter-TAD interactions: (G) distance from the promoter TSS to the nearest TAD boundary and (H) average TPM value of the promoter. ***p<2.2 \times 10^{-16}, **p<0.01, *p<0.05, NS = not significant, Wilcoxon rank-sum test.

DOI: https://doi.org/10.7554/eLife.35788.006
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, \( P_{\text{adj}} < 0.05 \)). ‘% 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.

DOI: https://doi.org/10.7554/eLife.35788.008
Figure 3. Enrichment of promoter interactions to distal regulatory features. (A,B) Proportion of promoter-distal interactions overlapping a histone ChIP-seq peak compared to random control MboI fragments (see Materials and methods). iPSC interactions were overlapped with H1 ESC ChIP-seq data; Figure 3 continued on next page.
CM interactions were overlapped with left ventricle ChIP-seq data from the Epigenome Roadmap Project (Supplementary file 10). (C) Fold enrichment of the data presented in (A) and (B). (D) Fold enrichment of promoter-distal interactions based on the expression level of the promoter. Promoters were grouped into five bins according to their average TPM values. Dashed line indicates no enrichment. (E) Fold enrichment of cell-type-specific and shared interactions (columns) to tissue-specific and shared chromatin features (rows). (F) Example of the NPPA gene in iPSCs (top) and CMs (bottom). Gray box highlights CM-specific interactions to CM-specific chromatin marks and an in vivo heart enhancer (Visel et al., 2007). For clarity, only interactions for NPPA are shown. *p<0.00001, *p=0.0017, Z-test.

DOI: https://doi.org/10.7554/eLife.35788.009
Figure 3—figure supplement 1. Correlation between the number of histone ChIP-seq peaks within 300 kb of promoters and gene expression level. Number of promoter-distal histone ChIP-seq peaks within 300 kb of promoters in iPSC (A) and CM (B). Spearman’s rho (\( \rho \)) 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 Materials and methods). Data are grouped by expression category to emphasize the trend. Horizontal bars indicate the median for each expression category. All correlation estimates are significant at \( p<2.2 \times 10^{-16} \) except for H3K27me3 in iPSCs (\( p=0.06 \)).

DOI: https://doi.org/10.7554/eLife.35788.010
**Figure 4.** A/B compartment switching corresponds to activation of tissue-specific genes. (A) Top panel: 10 Mb region on chromosome four 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-MAP dataset. CAMK2D expression is shown in TPM across CM and iPSC conditions. (B) Statistical analysis of gene expression in iPSC and CM conditions showing significant differences in expression levels for genes A and B across multiple replicates. (C) Box plots showing log2 fold change for each gene expression condition. (D) Gene ontology enrichment analysis for A/B compartment switching and tissue-specific gene expression.
Figure 4 continued

C and capture Hi-C interaction data. Bottom panel: zoomed in on the CAMK2D locus; only capture Hi-C A/B compartments shown. Inset: expression level of CAMK2D in iPSCs and CMs across the three replicates. (B) Expression level (TPM) of genes located in the A (green) or B (blue) compartment in each replicate of iPSC (left) or CM (right). (C) Difference in expression level (log2 fold change relative to iPSCs) of genes switching compartments from iPSC to CM or remaining in stable compartments. (D) Gene Ontology analysis of biological processes associated with genes switching from B to A compartments during iPSC-CM differentiation. ***p<2.2 x 10^{-16}, Wilcoxon rank-sum test.

DOI: https://doi.org/10.7554/eLife.35788.011
Figure 4—figure supplement 1. Comparison of A/B compartments in Hi-C and PCHi-C. Correlation between the A/B compartment score (principal component analysis of interaction data, PC-1) in whole-genome Hi-C (y-axis) and promoter capture Hi-C (x-axis) in iPSCs (top) and CMs (bottom). Spearman's $\rho > 0.98$, $p < 2.2 \times 10^{-16}$ in all cases.

DOI: https://doi.org/10.7554/eLife.35788.012
Figure 4—figure supplement 2. Example of A/B compartments. Genome browser snapshot of a ~53 Mb region on chromosome four showing A/B compartments in all three replicates of iPSCs and CMs using both whole-genome (WG) and promoter capture Hi-C data.

DOI: https://doi.org/10.7554/eLife.35788.013
Figure 4—figure supplement 3. GO analysis on the genes switching from active A compartments in iPSCs to inactive B compartments in CMs.

DOI: https://doi.org/10.7554/eLife.35788.014
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. p-Value calculated with a Z-test. (E) Proportion of GWAS LD SNPs that are eQTLs in left ventricle (LV) when considering either the full set of LD SNPs, or the subset that overlap CM promoter interactions. p-Value calculated with Fisher’s exact test. (F) Proportion of LV eQTLs (genome-wide) that map within a promoter interaction for the eQTL-associated gene (indicated by the red line). Random permutations were obtained by re-assigning each promoter’s set of interactions to a new promoter and calculating the proportion of eQTLs in random interactions that interact with their eQTL-associated gene. Proportions only consider eQTLs that overlap a promoter-distal interaction. P-values calculated with a Z-test.

DOI: https://doi.org/10.7554/eLife.35788.015
Figure 6. Characterizing target genes based on expression level. (A) Log2 fold change of the expression level of target genes in CMs compared to iPSCs (horizontal bar indicates median, 1.08; diamond indicates mean, 1.44). (B) Average TPM values of target genes in iPSCs and CMs (p=0.12, Wilcoxon rank-sum test). Diamonds indicate the mean value (40.6 for iPSC, 60.1 for CM). (C) Comparison of average TPM values for target genes in CMs and iPSCs. See Supplementary file 8 for full list of genes and TPM values. (D,E) Examples of genes looping to cardiac arrhythmia GWAS SNPs in Figure 6 continued on next page.
CMs. (D) The TBX5 gene interacts with a functionally validated arrhythmia locus (Smemo et al., 2012). (E) The LITAF gene interacts with a locus identified in (Arking et al., 2014). Yellow highlighted region indicates the promoter; gray box and zoom panel show the promoter-interacting regions (pink boxes) overlapping arrhythmia SNPs. For clarity, only interactions for the indicated promoter are shown.

DOI: https://doi.org/10.7554/eLife.35788.017
Figure 7. Relevance of CM promoter interactions for cardiac arrhythmia, myocardial infarction and heart failure. (A–C) Gene Ontology analysis for target genes looping to (A) cardiac arrhythmia SNPs, (B) myocardial infarction SNPs, and (C) heart failure SNPs. (D) The SORT1 promoter loops to a
distal myocardial infarction locus (Musunuru et al., 2010). The rs12740374 SNP shown to disrupt a C/EBP binding site in (Musunuru et al., 2010) is colored red. (E) The ACTA2 promoter loops to the 10q21 heart failure locus (Smith et al., 2010). Zoom plots depict the full interacting region overlapping GWAS LD SNPs. For clarity, only interactions for the indicated gene are shown.

DOI: https://doi.org/10.7554/eLife.35788.018