Supplemental Material for

Human cardiac cis-regulatory elements, their cognate transcription factors and regulatory DNA sequence variants

Dongwon Lee*, Ashish Kapoor, Alexias Safi, Lingyun Song, Marc K. Halushka, Gregory E. Crawford, Aravinda Chakravarti*

*Corresponding authors: aravinda.chakravarti@nyumc.org, dongwon.lee@nyumc.org

This document includes:

Supplemental Methods

Supplemental Figures S1 to S11

Supplemental Tables S1 to S5

Legends for Supplemental Data S1 to S6

Supplemental Methods

DNase-seq on adult human heart samples

Two adult female human heart left ventricle (LV) samples, one from an autopsy accessed 5 hours postmortem and one from an explant accessed 12 hours post-surgery, were collected for DNase-seq experiments, under a Johns Hopkins University School of Medicine IRB-approved protocol. To evaluate the effect of postmortem or post-surgery time interval on DNase I hypersensitivity, two pieces were dissected from each sample, both snap frozen either immediately after dissection or after being left at room temperature for 19 hours and 12 hours, respectively. DNase-seq libraries were constructed as previously described (Song and Crawford 2010), with each library sequenced using a 50bp single-end protocol on an Illumina HiSeq 2000 instrument to an average depth of 25 million aligned reads per sample. We trimmed the DNase-seq reads to 20 bp to remove adapter sequences and aligned them to the hg19 human genome using Bowtie (Langmead et al. 2009) with parameters -n 2 –m 4. We identified minimal effects of postmortem or post-surgical time interval on DNase-seq signals, and, therefore, pooled them as technical replicates.

gkm-SVM training

To train gkm-SVM, we followed a previously described method with minor modifications (Lee et al. 2015; Lee 2016). Briefly, we defined a positive set by finding 600bp regions centered at each of the DHS summits, and a negative set as an equal number of random genomic regions matching GC-content, repeat fraction, and the length distribution of the positive set. Both positive and negative regions from Chromosome 9 were set aside for evaluation of classification performance. We trained gkm-SVM using the new software LS-GKM with default parameters: weighted gkm-kernel, L = 11, k = 7, and d = 3. We built final models by averaging gkm-SVMs trained on ten independently generated negative sets. For comparison of classification performance across models, we also tried 300bp regions and the original gkm-kernel (Ghandi et al. 2014) keeping other parameters fixed. Next, we tested its ability to predict new DHSs by focusing on a SVM model trained on only the Roadmap data. We

excluded DHSs overlapping the Roadmap DHSs from each of the remaining four data sets, to ensure that DHSs for validation were never seen in training. Then, we scored the remaining DHSs as well as random genomic regions and compared their distributions. To investigate the relationship between the number of samples used for model building and a model's accuracy, we predicted the top 10,000 regions using each of the models from all possible combinations of samples, comparing them to the reserved chromosome 9 DHSs (n = 7,070). The recall rate ("the number of recovered true DHSs by the prediction" / "the number of true DHSs") was then estimated. As expected, the recall rate increased as more samples were used for prediction (Supplemental Fig. S4B).

GRCh38 vs. hg19

To test the effect of the reference genome on gkm-SVM, we retrained the model using the Roadmap data and compared their SVM weights. Specifically, we downloaded the GRCh38 mapped reads from the ENCODE portal website (file name: ENCFF837QAV.bam) and recalled the peaks as described in Methods. Expectedly, most hg19 DHSs (~94%; 103,778/110,636) overlap GRCh38 DHSs lifted over to hg19. A new model trained on GRCh38 DHSs achieved similar AUC to the original hg19 model (AUC=0.913). Moreover, we achieved a very high correlation (C = 0.85) between the GRCh38 (single negative set) model vs. the hg19 (an average of 10 negative sets). Since the randomly generated negative sets also contribute to the variation, the imperfect correlation was expected. We therefore concluded that the reference genome would not significantly alter our gkm-SVM models and results.

Data sets for genomic annotations

To identify H3K27ac marked regions in heart tissues, we obtained the processed H3K27ac ChIP-seq peaks directly from the ENCODE portal (www.encodeproject.org). We downloaded 8 H3K27ac peak files (file accessions: ENCFF101VMB, ENCFF159NYF, ENCFF995GJW, ENCFF124AQU, ENCFF175SDM, ENCFF134ZIJ, ENCFF277TVL, ENCFF546TJN) from heart tissues (left/right ventricle, right atrium auricle, and right cardiac atrium), extended them 100bp at both ends, and

combined them, resulting in 118,597 distinct cardiac H3K27ac peaks. For the mappability analysis, we obtained the pre-calculated mappability (24bp and 36bp) (Derrien et al. 2012) from the ENCODE UCSC genome browser and calculated the proportion of ambiguous bases for each category of the predicted DHSs. As a marker for heterochromatin DNA, we used ENCODE H3K9me3 ChIP-seq data sets from multiple heart tissues (left/right ventricles and right cardiac atrium). We downloaded six H3K9me3 peak files from the ENCODE portal website (File accession: ENCFF964NVL, ENCFF951QMV, ENCFF495QNH, ENCFF253OXB, ENCFF524ELE, ENCFF137LHG) and merged them to define heart H3K9me3 regions. Then, we calculated proportion of predicted DHSs overlapping these regions (>1bp overlap). To calculate SNP frequencies, we divided all SNPs from the 1000 Genome project into common and rare variants using 1% MAF in European ancestry subjects as the threshold. We then calculated the SNP frequencies as "the number of (common or rare) SNPs in predicted DHSs" divided by "the total base coverage of the DHSs".

Transcription factor binding motifs

The full list of position weight matrices (PWMs) of human transcription factors (TF) were obtained from the Catalog of Inferred Sequence Binding Preferences (CIS-BP) (last updated: Apr 5th, 2015, build 1.02) (Weirauch et al. 2014). We excluded low confidence PWMs likely derived from false oligonucleotides (stem loop forms of single-stranded DNA) in HT-SELEX experiments (Jolma et al. 2013; Orenstein and Shamir 2014), and then selected one motif per TF to reduce redundancy based on their source types (MSource_Type) with the following order of preference: (1) JASPAR, (2) ChIP-seq, (3) HT-SELEX, (4) PBM, (5) HOCOMOCO, (6) B1H, and (7) TRANSFAC. This resulted in 775 distinct motifs associated with 811 TFs. We augmented this list with a newly identified set of 93 C2H2 Zinc Finger TF motifs added CIS-BP (Schmitges et al. 2016). The final set contained 868 distinct motifs associated with 904 TFs (Supplemental Data 4 and 5).

Gene expression analysis of transcription factors

We obtained median expression values (FPKM) of all genes for 53 tissues (V6P) from the GTEx project (The GTEx Consortium 2015). These values were normalized by the TMM (trimmed mean of M-values) method (Robinson and Oshlack 2010) implemented in edgeR (Robinson et al. 2010). We declared genes with normalized expression value > 1 as 'expressed.' Starting from the 904 TFs with known PWMs we identified, we constructed 2x2 contingency tables for each tissue using two binary variables: whether the corresponding motifs were enriched or not in heart CREs and whether they are expressed or not in a given tissue, and used one-sided Fisher's exact test to determine statistical significance. We also identified 350 commonly expressed TFs, defined as expressed in 90% or greater of GTEx tissues, and repeated the association tests after removing these TFs.

ENCODE ChIP-seq data analysis

To validate the predicted cardiac TFs, we analyzed ENCODE ChIP-seq data sets. We first downloaded all TF ChIP-seq data sets (n=2,287) from the ENCODE portal website with the following criteria: hg19 (genome build), bed narrowPeak (file type), optimal (output type). We then identified a subset of 1,367 data sets covering 302 distinct TFs whose binding specificities were available in our motif database. Among them, 176 TFs from 1,054 data sets matched the 334 potential cardiac TFs, while the remaining 126 TFs from 313 data sets did not. For each of these data sets, we calculated the proportion of TF ChIP-seq regions overlapping observed and predicted DHSs, and compared the distributions between these two classes (cardiac vs non-cardiac TFs).

Properties of common DHSs

DHSs from all ENCODE and Roadmap projects were aggregated using BEDTools multiinter command (Quinlan and Hall 2010) to find common open chromatin regions. We identified these as regions that are accessible in at least 30% of the samples tested; common heart DHSs were those present in heart tissues only. For the CTCF-bound regions, we obtained 71 data sets of the processed CTCF ChIP-seq peaks from a previous study (Ghandi et al. 2014) and merged them to identify all potential CTCF

binding sites (n = 118,579) in the human genome. For transcription start sites, we used GENCODE v19 transcripts (hg19) (Harrow et al. 2012). To identify sequence features differentially enriched in the common heart DHSs, we compared SVM weights between the generic and specific cardiac models. The 11-mer SVM weights were normalized to Z-scores (zero means, unit variance) for comparisons between models. We identified 11-mers with >6 Z-score differences as common DHS enriched 11-mers.

Preprocessing of DNase-seq data sets for allele-biased DHSs

We used twelve fetal heart DNase-seq data sets from the Roadmap project (GSM530661, GSM665809, GSM665811, GSM665814, GSM665817, GSM665824, GSM665830, GSM665831, GSM723023, GSM774203, GSM817220, and GSM878630) as well as the five adult heart DNase-seq data sets from this study. To reduce reference sequence mapping bias, we realigned DNase-seq reads using GSNAP with the "SNP-tolerant alignment" option (Wu and Watanabe 2005; Wu and Nacu 2010) using all common SNPs in dbSNP138 as a reference. After read mapping, we retained the properly mapped reads with at least 30 MAPQ scores for downstream analysis. We were only able to remap the DNase-seq reads generated by the UW protocol (12 fetal samples + 1 adult sample) because of their longer read lengths.

Heritability analysis of QT-interval and other phenotypes

We adapted the LD score regression method (Bulik-Sullivan et al. 2015; Finucane et al. 2015) to calculate partitioned heritability of the QTi, systolic blood pressure, diastolic blood pressure, pulse rate and BMI, based on our heart CRE annotations. We obtained the QTi summary statistics from the senior author (Arking et al. 2014), and the summary statistics of other phenotypes from the UK-biobank project (sites.google.com/broadinstitute.org/ukbbgwasresults). For LD score regression analysis, we used European ancestry 1KGP individuals (Phase 3) as a reference panel. Partitioned heritability was estimated for each annotation along with 53 baseline annotations, using HapMap3 common (MAF > 5%) SNPs as regression SNPs and the 1KGP common SNPs (MAF > 5%) as reference SNPs, as

recommended (Finucane et al. 2015). All the data for these LD score regression analyses were obtained from: data.broadinstitute.org/alkesgroup/LDSCORE.

References

Arking DE, Pulit SL, Crotti L, van der Harst P, Munroe PB, Koopmann TT, Sotoodehnia N, Rossin EJ, Morley M, Wang X, et al. 2014. Genetic association study of QT interval highlights role for calcium signaling pathways in myocardial repolarization. *Nat Genet* **46**: 826–836.

Bulik-Sullivan BK, Loh P-R, Finucane HK, Ripke S, Yang J, Schizophrenia Working Group of the Psychiatric Genomics Consortium, Patterson N, Daly MJ, Price AL, Neale BM. 2015. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* **47**: 291–295.

Derrien T, Estellé J, Marco Sola S, Knowles DG, Raineri E, Guigó R, Ribeca P. 2012. Fast Computation and Applications of Genome Mappability. *PLoS ONE* **7**: e30377.

Finucane HK, Bulik-Sullivan B, Gusev A, Trynka G, Reshef Y, Loh P-R, Anttila V, Xu H, Zang C, Farh K, et al. 2015. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet* **47**: 1228–1235.

Ghandi M, Lee D, Mohammad-Noori M, Beer MA. 2014. Enhanced Regulatory Sequence Prediction Using Gapped k-mer Features. *PLoS Comput Biol* **10**: e1003711.

Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, Aken BL, Barrell D, Zadissa A, Searle S, et al. 2012. GENCODE: The reference human genome annotation for The ENCODE Project. *Genome Res* **22**: 1760–1774.

Jolma A, Yan J, Whitington T, Toivonen J, Nitta KR, Rastas P, Morgunova E, Enge M, Taipale M, Wei G, et al. 2013. DNA-Binding Specificities of Human Transcription Factors. *Cell* **152**: 327–339.

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**: R25.

Lee D. 2016. LS-GKM: a new gkm-SVM for large-scale datasets. *Bioinformatics* **32**: 2196–2198.

Lee D, Gorkin DU, Baker M, Strober BJ, Asoni AL, McCallion AS, Beer MA. 2015. A method to predict the impact of regulatory variants from DNA sequence. *Nat Genet* **47**: 955–961.

Orenstein Y, Shamir R. 2014. A comparative analysis of transcription factor binding models learned from PBM, HT-SELEX and ChIP data. *Nucleic Acids Res* **42**: e63–e63.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841–842.

Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**: 139–140.

Robinson MD, Oshlack A. 2010. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol* **11**: R25.

Schmitges FW, Radovani E, Najafabadi HS, Barazandeh M, Campitelli LF, Yin Y, Jolma A, Zhong G, Guo H, Kanagalingam T, et al. 2016. Multiparameter functional diversity of human C2H2 zinc finger proteins. *Genome Res* **26**: 1742–1752.

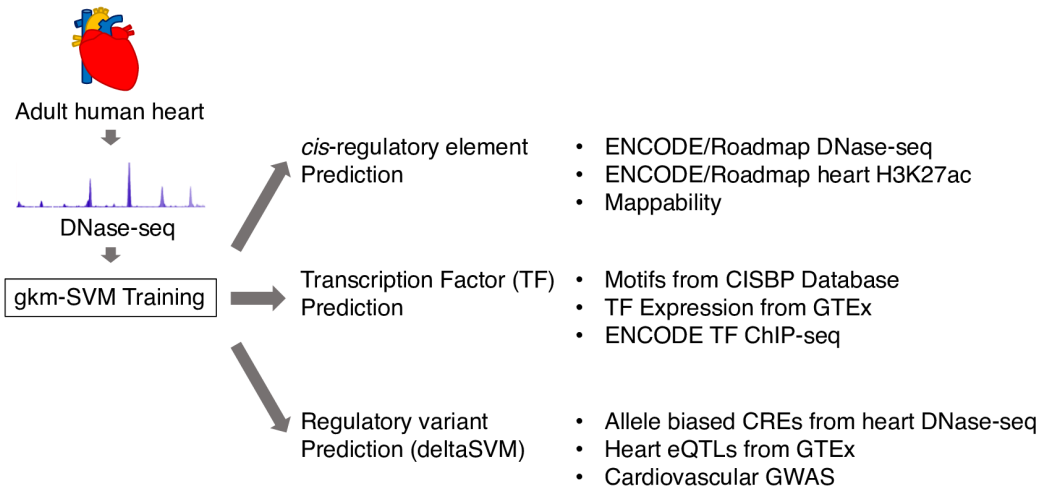
Song L, Crawford GE. 2010. DNase-seq: A High-Resolution Technique for Mapping Active Gene Regulatory Elements across the Genome from Mammalian Cells. *Cold Spring Harb Protoc* **2010**. http://cshprotocols.cshlp.org/content/2010/2/pdb.prot5384 (Accessed January 14, 2015).

The GTEx Consortium. 2015. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science* **348**: 648–660.

Weirauch MT, Yang A, Albu M, Cote AG, Montenegro-Montero A, Drewe P, Najafabadi HS, Lambert SA, Mann I, Cook K, et al. 2014. Determination and Inference of Eukaryotic Transcription Factor Sequence Specificity. *Cell* **158**: 1431–1443.

Wu TD, Nacu S. 2010. Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics* **26**: 873–881.

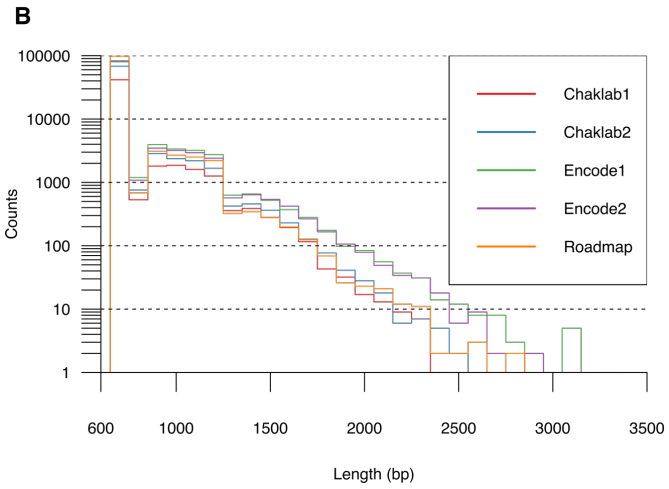
Wu TD, Watanabe CK. 2005. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* **21**: 1859–1875.



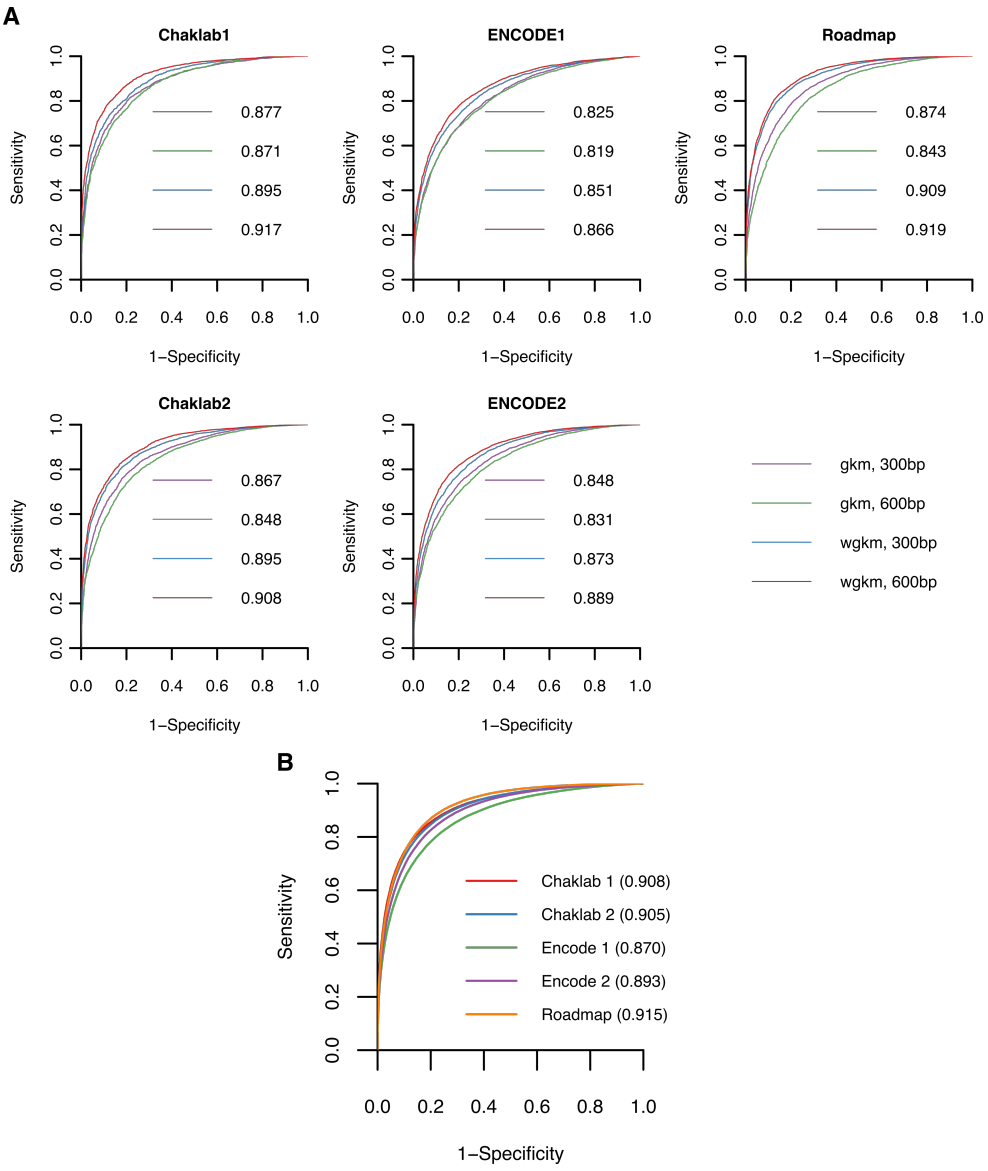
Supplemental Figure S1. A schematic overview of our study. Major data sets used for analyses are listed.

A

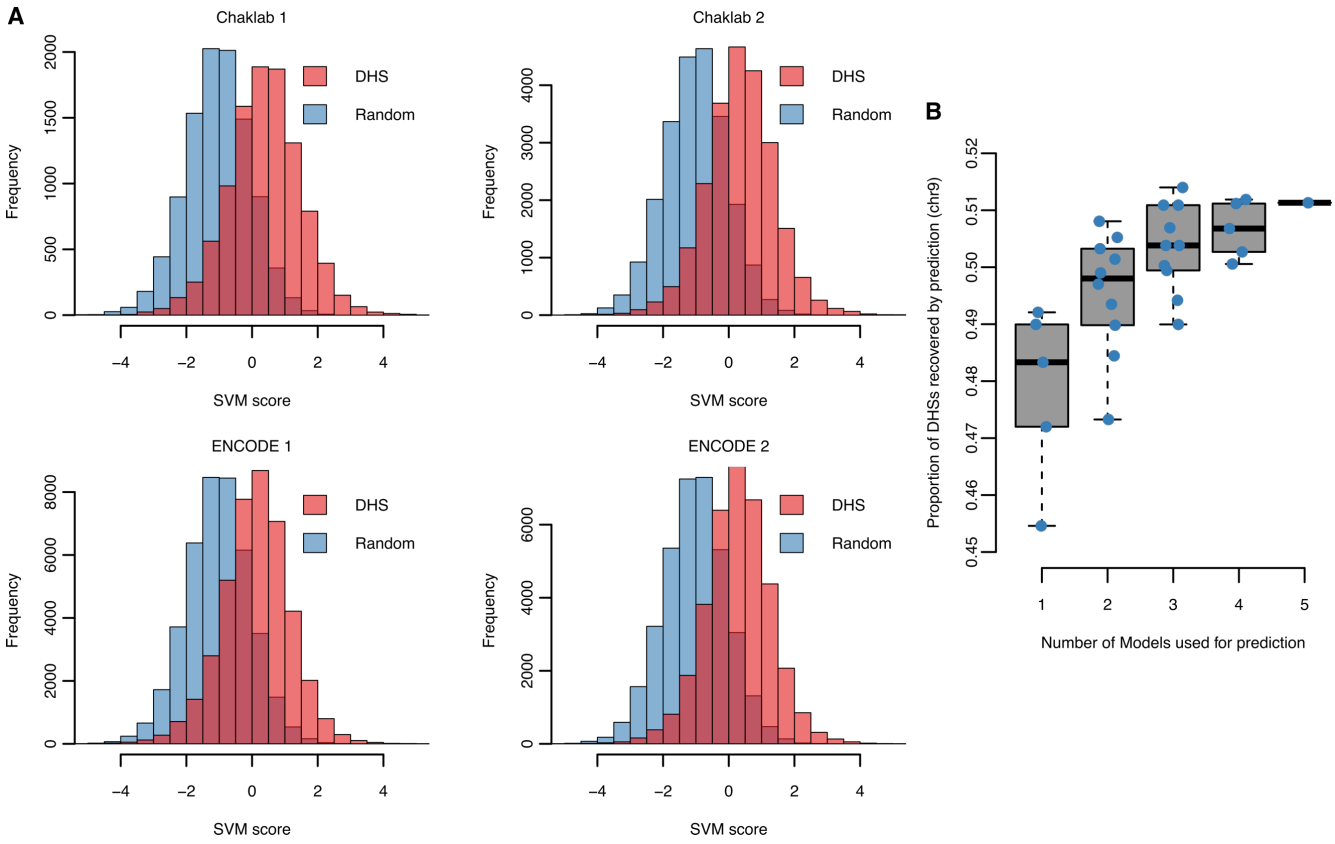
Project	DNase-seq Protocol	Rep	Read Size	# Total Reads	# Unique Reads	# of Summits	# of DHSs
Chaklab	Duke	1	20bp	45.0M	35.9M	61,191	50,190
Chaklab	Duke	2	20bp	51.6M	43.4M	94,460	79,802
ENCODE	Duke	1	20bp	154.3M	77.1M	120,386	97,621
ENCODE	Duke	2	20bp	167.1M	62.4M	120,161	99,022
Roadmap	UW	1	35bp	42.5M	38.2M	125,845	110,636



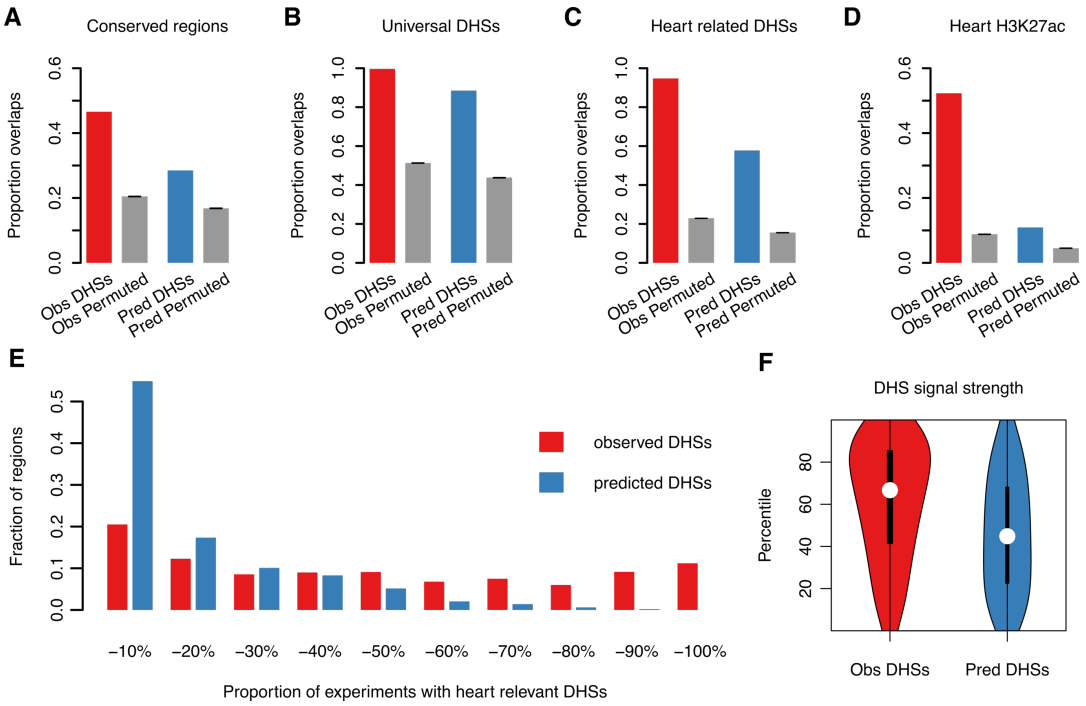
Supplemental Figure S2. DHSs in individual human tissues show the greatest similarity to their biological replicates than to other tissues. (A) Genomic statistics of five biological heart DNase-seq replicate data. **(B)** Distribution of DHS element lengths of five replicates. **(C)** Pairwise DHS overlap between different tissues: the top 50,000 regions from each sample were selected and overlap estimated by the Jaccard index (number of bases in their intersection over their union); C1: Chaklab 1, C2: Chaklab 2, E1: ENCODE 1, E2: ENCODE 2, R: Roadmap; the prefix 'f' stands for fetal tissues; other tissue samples data were obtained from the Roadmap Epigenomics project.



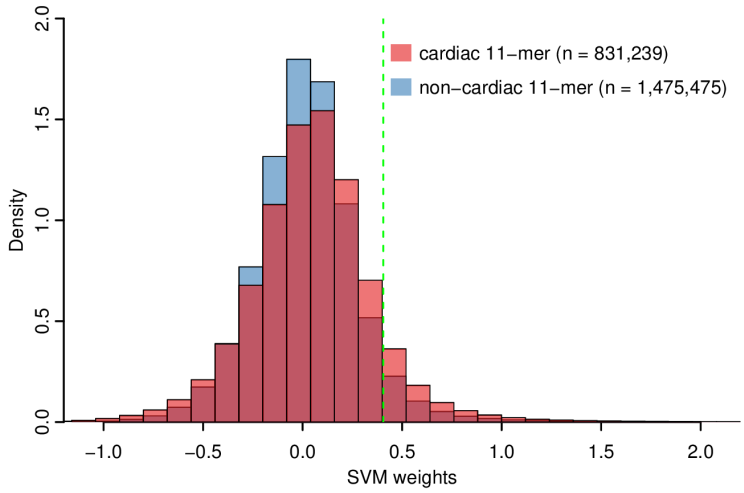
Supplemental Figure S3. New LS-GKM software generates more accurate models. (A) Receiver operating characteristic (ROC) curves of gkm-SVM trained on each of five heart DNase-seq replicate data sets with different parameters using the LS-GKM software. gkm: original gkm-kernel; wgkm: weighted gkm-kernel; DHS lengths of 300bp and 600bp. Reserved chromosome 9 test sets were used for evaluation. wgkm with 600bp achieved the best classification performance in all cases. **(B)** ROC curves with 5-fold cross validation.



Supplemental Figure S4. Sequence-based prediction identifies true DHSs missed by experiments. (A) Comparisons of SVM score distributions between DHSs and random genomic regions, with SVM trained on the Roadmap data set. **(B)** Box plots of the number of sequence model predicted DHSs as a function of the number of samples used for training and prediction.

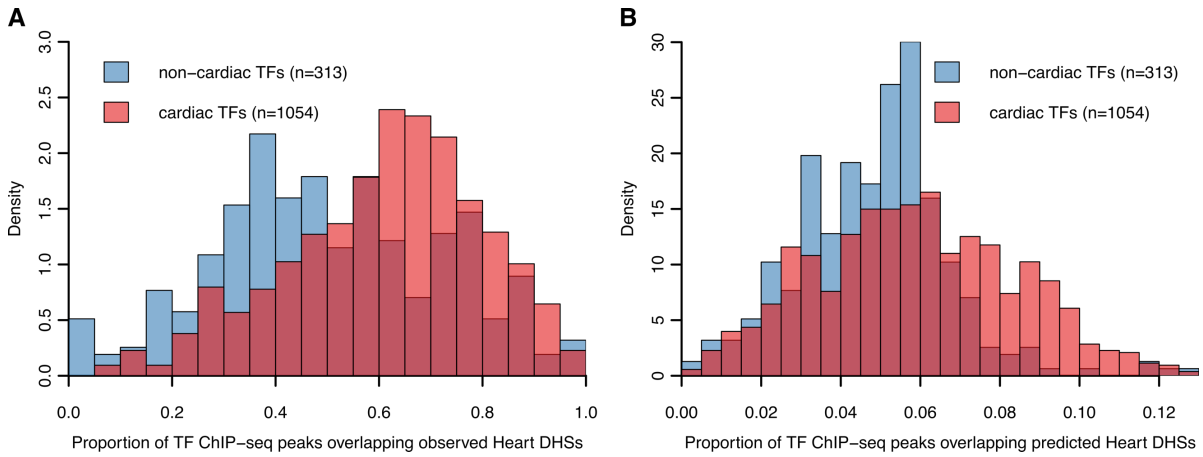


Supplemental Figure S5. Predicted DHSs are enriched for functional genomic annotations but with lower detection rates. The overlap of observed DHSs and predicted DHSs with **(A)** genome conserved elements (GERP++ elements were used for identifying conserved regions; average overlap of 100 randomly permuted regions for each set were used as negative controls); **(B)** universal DHSs; **(C)** heart related DHSs; **(D)** H3K27ac histone modification marks in heart tissues; **(E)** proportion of heart related samples with DHSs among all observed and predicted DHSs; **(F)** violin plots of DHS signal strength in heart-related tissues and cell-types for 5,000 randomly selected observed and predicted DHSs.

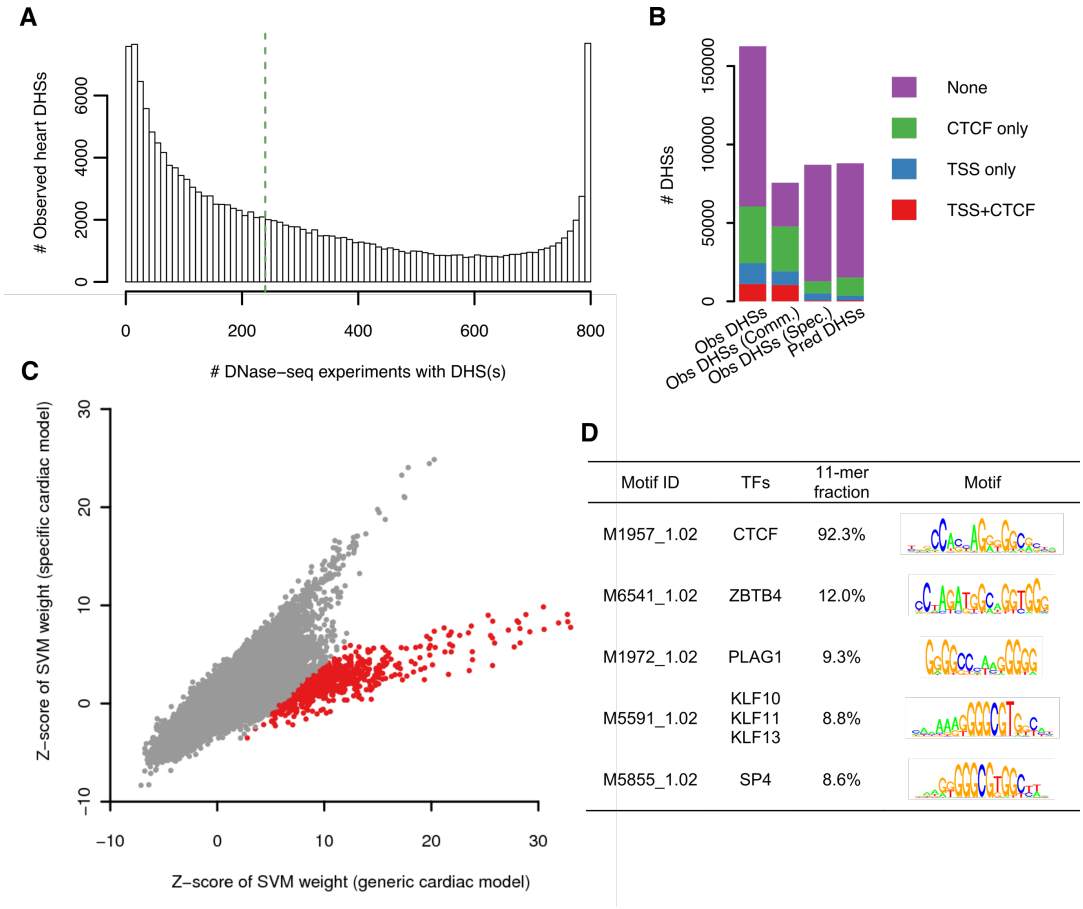


Supplemental Figure S6. 11-mers matching cardiac TF PWMs have larger SVM weights.

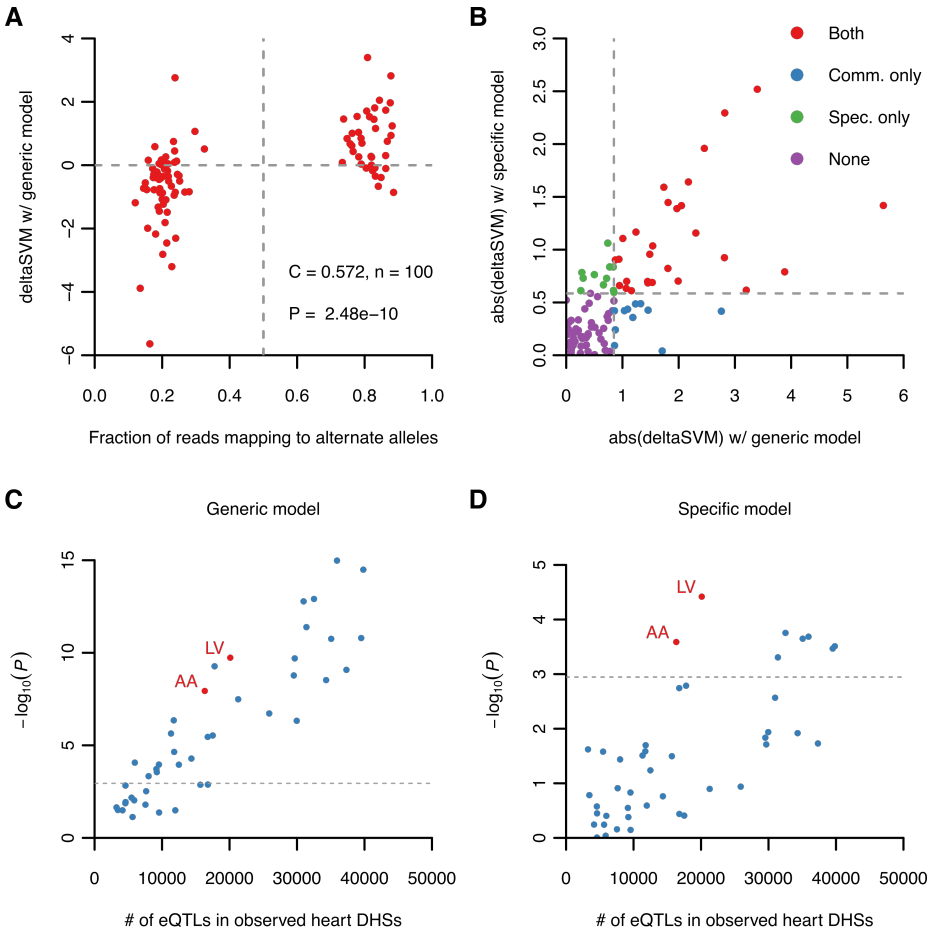
Two classes of 11-mers were defined based on their matches to predicted cardiac TF PWMs (cardiac 11-mer) and non-cardiac TF PWMs (non-cardiac 11-mer). Distributions of SVM weights for these two sets of 11-mers are shown.



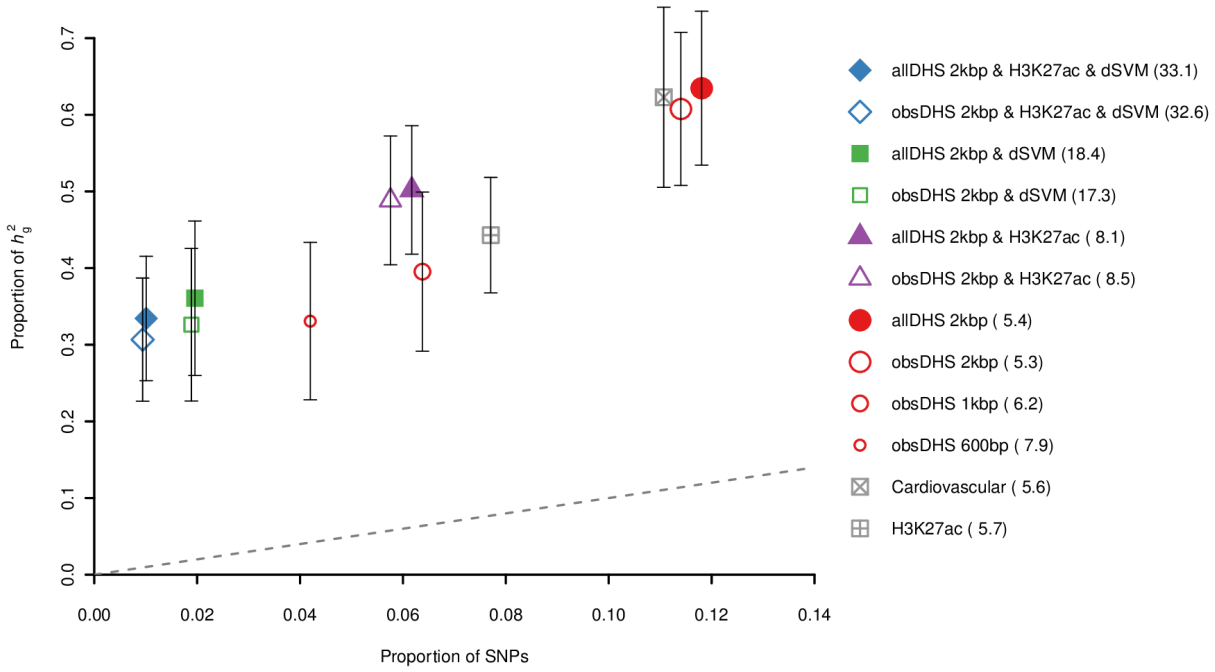
Supplemental Figure S7. Cardiac TF bound genomic regions are more enriched in heart DHSs. Proportions of TF ChIP-seq peaks that overlap **(A)** observed and **(B)** predicted heart DHSs are compared between the potential cardiac TFs and non-cardiac TFs. Regions bound by the predicted cardiac TFs exceed the overlap with heart DHSs compared to non-cardiac TFs ($P < 2.2 \times 10^{-16}$ for observed heart DHSs and $P < 2.84 \times 10^{-16}$ for predicted heart DHSs). P-values were calculated using one-tailed two sample Kolmogorov-Smirnov tests.



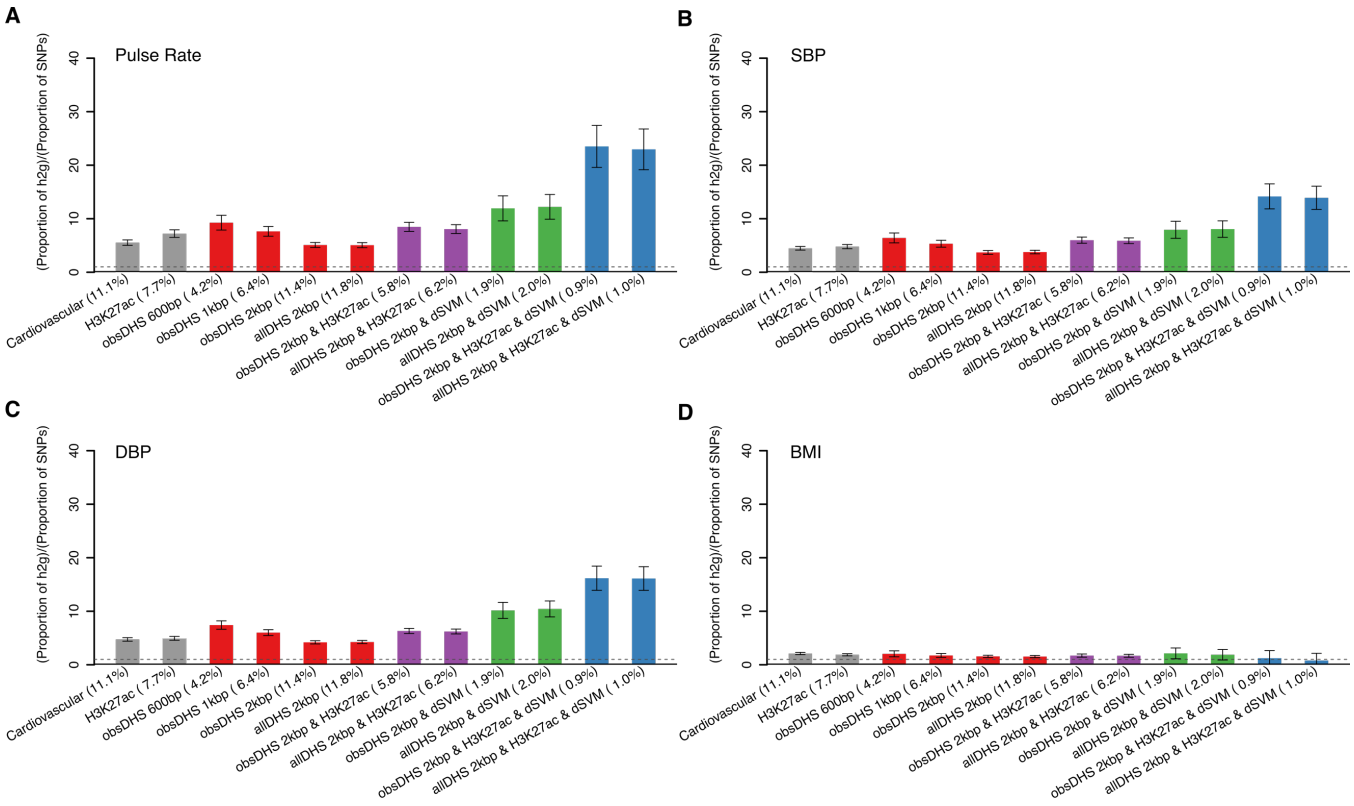
Supplemental Figure S8. Common DHSs are mostly CTCF bound regions and gene promoters. (A) Distribution of numbers of tissue samples in which cardiac DHSs are in open chromatin; the green dashed line is at 30% of the samples. **(B)** Classification of cardiac DHSs by overlapping CTCF-bound regions and/or transcription start sites (TSS). **(C)** Comparison of SVM weights between the generic and the specific cardiac model. The SVM weights were normalized to Z-scores with 11-mers with significantly reduced weights in the specific models (Z-score difference > 6) highlighted as red. **(D)** Table showing the top 5 TF motifs that are expressed in heart tissues as well as enriched in the highlighted *k*-mers in (C).



Supplemental Figure S9. Generic cardiac model predicts distinct regulatory variants, while deltaSVM variants restricted to observed DHSs only miss causal eQTLs. (A) deltaSVM scores calculated using the generic cardiac model are highly correlated with allele-biased chromatin accessibility; **(B)** 30% of deltaSVM variants predicted by the generic cardiac model are not predicted by the specific cardiac model. The dashed lines indicate the deltaSVM score cut-off for classifying candidate regulatory variants. **(C)** deltaSVM variants from the generic and **(D)** specific models are compared to GTEx eQTLs restricted to observed DHSs only. The two heart tissues (AA: Atrial Appendages, LV: Left Ventricles) are highlighted in red, and the dashed lines indicate a Bonferroni corrected $P < 0.05$.



Supplemental Figure S10. Putative cardiac regulatory variants impact QT trait variation significantly. The proportion of QT interval heritability explained by SNPs (h_G^2) within various DHS-based cardiac CRE annotations, with fold enrichment indicated in parenthesis; error bars denote standard errors estimated by a block Jackknife method; the dashed line indicates no enrichment.



Supplemental Figure S11. Putative cardiac regulatory variants significantly impact multiple cardiac phenotypes. Enrichment of SNP heritability was estimated for various cardiac functional annotations for three different cardiac phenotypes and one non-cardiac phenotype from the UK Biobank data set; **(A)** pulse rate; **(B)** systolic blood pressure; **(C)** diastolic blood pressure; **(D)** BMI.

Supplemental Table S1. Comparisons of predicted DHSs and potential factors affecting CRE detection.

DHS	Number of regions	Prop. of ambiguous bases (24bp)	Prop. of ambiguous bases (36bp)	Prop. of overlap with H3K9me3	Rare SNP Frequency (MAF < 1%)	Common SNP Frequency (MAF > 1%)	Prop. of FANTOM5 Enhancers
Observed	162,435	0.312	0.100	0.026	0.00702	0.00332	0.112
Tier 1	7,463	0.287	0.059	0.025	0.00683	0.00312	0.123
Tier 2	43,329	0.360	0.112	0.069	0.00703	0.00342	0.059
Tier 3	27,081	0.409	0.160	0.104	0.00734	0.00361	0.027
Tier 4	10,153	0.602	0.366	0.154	0.00640	0.00307	0.005

Supplemental Table S2. Comparisons of predicted DHSs and DHSs observed at lower stringency.

DHS	Number of regions	Prop. overlap with DHSs of FDR<0.1	Prop. overlap with DHSs of FDR<0.15	Prop. overlap with DHSs of FDR<0.2	Prop. overlap with DHSs of FDR<0.25
Tier 1	7,463	0.139	0.191	0.275	0.345
Tier 2	43,329	0.067	0.087	0.145	0.200
Tier 3	27,081	0.014	0.021	0.042	0.062
Tier 4	10,153	0.003	0.004	0.010	0.017

Supplemental Table S3. Comparisons of QT-interval heritability enrichment using various classes of DHS predictions.

DHS	Prop. SNP	Prop. h^2	Prop. h^2 SE	Enrichment	Enrichment SE	P-value
Observed	0.114	0.608	0.100	5.329	0.876	4.749×10^{-06}
Observed + Tier 1	0.118	0.635	0.100	5.376	0.850	1.774×10^{-06}
Observed + Tier 1/2	0.148	0.564	0.112	3.807	0.756	2.602×10^{-04}
Observed + Tier 1/2/3/4	0.173	0.558	0.117	3.232	0.678	9.528×10^{-04}

Supplemental Table S4. 149 predicted QT-interval regulatory variants.

regulatory SNP								Index SNP				
rsID	Chr	Position	Freq.	GWAS P-value	Obs DHS	R^2	D'	rsID	Position	Freq.	GWAS P-value	Locus
rs11583631	1	6147298	0.142	NA	1	0.976	0.992	rs2273042	6149122	0.141	1.48×10^{-08}	*RNF207*
rs11584419	1	6147341	0.143	NA	1	0.984	1.000	rs2273042	6149122	0.141	1.48×10^{-08}	*RNF207*
rs12069441	1	6264775	0.272	NA	1	0.148	0.397	rs846111	6279370	0.260	2.30×10^{-37}	*RNF207*
rs4522020	1	6319604	0.701	NA	1	0.484	-0.766	rs846111	6279370	0.260	2.30×10^{-37}	*RNF207*
rs7511979	1	23747425	0.861	NA	1	0.070	0.677	rs2298632	23710475	0.487	6.76×10^{-15}	*TCEA3*
rs10918444	1	161979533	0.176	NA	1	0.474	-0.937	rs6669543	161981025	0.717	1.57×10^{-38}	*NOS1AP*
rs7532680	1	161988206	0.213	5.19×10^{-35}	1	0.400	-0.765	rs6669543	161981025	0.717	1.57×10^{-38}	*NOS1AP*
rs6427645	1	161989418	0.906	NA	1	0.264	1.000	rs6669543	161981025	0.717	1.57×10^{-38}	*NOS1AP*
rs60340472	1	162006751	0.042	NA	1	0.142	0.699	rs16857031	162112910	0.130	2.41×10^{-60}	*NOS1AP*
rs12143842	1	162033890	0.258	8.97×10^{-210}	1	1.000	1.000	rs12143842	162033890	0.258	8.97×10^{-210}	*NOS1AP*
rs12084981	1	162034982	0.298	NA	1	0.482	0.769	rs12143842	162033890	0.258	8.97×10^{-210}	*NOS1AP*
rs12093845	1	162101829	0.083	NA	1	0.552	0.958	rs16857031	162112910	0.130	2.41×10^{-60}	*NOS1AP*
rs4657150	1	162102064	0.626	NA	1	0.416	-0.846	rs12143842	162033890	0.258	8.97×10^{-210}	*NOS1AP*
rs6704253	1	162102117	0.086	NA	1	0.528	0.920	rs16857031	162112910	0.130	2.41×10^{-60}	*NOS1AP*
rs66839222	1	162103614	0.083	NA	1	0.552	0.958	rs16857031	162112910	0.130	2.41×10^{-60}	*NOS1AP*
rs4427385	1	162103916	0.083	NA	1	0.552	0.958	rs16857031	162112910	0.130	2.41×10^{-60}	*NOS1AP*
rs4480334	1	162104081	0.083	NA	1	0.552	0.958	rs16857031	162112910	0.130	2.41×10^{-60}	*NOS1AP*
rs10918758	1	162116987	0.628	NA	1	0.339	-0.760	rs12143842	162033890	0.258	8.97×10^{-210}	*NOS1AP*
rs4445433	1	162148796	0.369	NA	1	0.259	-1.000	rs4657172	162179632	0.869	9.21×10^{-09}	*NOS1AP*
rs10918846	1	162162523	0.388	1.31×10^{-53}	1	0.239	-1.000	rs4657172	162179632	0.869	9.21×10^{-09}	*NOS1AP*
rs6692381	1	162167884	0.257	NA	1	0.985	1.000	rs3934467	162182677	0.259	4.76×10^{-129}	*NOS1AP*
rs17457880	1	162168154	0.026	NA	1	1.000	1.000	rs17457880	162168154	0.026	2.33×10^{-10}	*NOS1AP*
rs12123267	1	162199351	0.259	NA	1	0.990	0.995	rs3934467	162182677	0.259	4.76×10^{-129}	*NOS1AP*
rs4424487	1	162199362	0.583	NA	1	0.482	-0.993	rs3934467	162182677	0.259	4.76×10^{-129}	*NOS1AP*
rs12125138	1	162200616	0.055	NA	1	0.165	1.000	rs3934467	162182677	0.259	4.76×10^{-129}	*NOS1AP*
rs10918930	1	162200743	0.109	NA	1	0.667	-0.906	rs4657172	162179632	0.869	9.21×10^{-09}	*NOS1AP*
rs4657178	1	162210610	0.288	NA	0	0.757	0.935	rs3934467	162182677	0.259	4.76×10^{-129}	*NOS1AP*
rs10753784	1	162248920	0.454	2.18×10^{-31}	1	0.262	0.789	rs3934467	162182677	0.259	4.76×10^{-129}	*NOS1AP*
rs12030361	1	162253826	0.261	NA	1	0.421	0.652	rs3934467	162182677	0.259	4.76×10^{-129}	*NOS1AP*
rs12028141	1	162253908	0.261	NA	1	0.421	0.652	rs3934467	162182677	0.259	4.76×10^{-129}	*NOS1AP*
rs10919067	1	162254743	0.261	NA	1	0.421	0.652	rs3934467	162182677	0.259	4.76×10^{-129}	*NOS1AP*
rs347313	1	162304276	0.556	NA	1	0.199	1.000	rs347272	162318498	0.863	1.77×10^{-37}	*NOS1AP*
rs5778266	1	162304825	0.555	NA	1	0.198	1.000	rs347272	162318498	0.863	1.77×10^{-37}	*NOS1AP*
rs11577628	1	162319524	0.094	2.96×10^{-34}	1	0.640	-0.988	rs347272	162318498	0.863	1.77×10^{-37}	*NOS1AP*
rs79011457	1	169072992	0.130	NA	1	0.922	0.973	rs12061601	169070450	0.127	3.66×10^{-20}	*ATP1B1*
rs12079856	1	169073002	0.132	NA	1	0.957	1.000	rs12061601	169070450	0.127	3.66×10^{-20}	*ATP1B1*
rs7527703	1	169338046	0.368	NA	1	0.308	-0.942	rs1983546	169446183	0.374	6.15×10^{-16}	*ATP1B1*
rs67014132	2	40740353	0.041	NA	1	1.000	1.000	rs12997023	40752982	0.041	5.35×10^{-14}	*SLC8A1*
rs1473805	3	38577679	0.263	NA	1	0.588	0.953	rs6793245	38599037	0.356	2.42×10^{-26}	*SCN5A*
rs4073797	3	38590850	0.361	NA	1	0.776	0.891	rs6793245	38599037	0.356	2.42×10^{-26}	*SCN5A*
rs6793245	3	38599037	0.356	NA	1	1.000	1.000	rs6793245	38599037	0.356	2.42×10^{-26}	*SCN5A*
rs62245110	3	38630311	0.180	NA	1	0.263	0.561	rs11710077	38657899	0.208	9.79×10^{-13}	*SCN5A*
rs11710077	3	38657899	0.208	9.79×10^{-13}	1	1.000	1.000	rs11710077	38657899	0.208	9.79×10^{-13}	*SCN5A*
rs6782237	3	38696553	0.666	NA	1	0.908	0.977	rs6599234	38715300	0.677	2.50×10^{-10}	*SCN5A*
rs6801957	3	38767315	0.585	3.11×10^{-10}	1	1.000	1.000	rs6801957	38767315	0.585	3.11×10^{-10}	*SCN5A*
rs7636423	3	47309134	0.600	NA	1	0.876	-0.945	rs17784882	47544003	0.395	4.66×10^{-08}	*ELP6*
rs1076394	3	47322781	0.399	NA	1	0.943	0.979	rs17784882	47544003	0.395	4.66×10^{-08}	*ELP6*
rs8180040	3	47388947	0.399	1.80×10^{-07}	1	0.951	0.983	rs17784882	47544003	0.395	4.66×10^{-08}	*ELP6*
rs7613282	3	47389409	0.399	NA	1	0.951	0.983	rs17784882	47544003	0.395	4.66×10^{-08}	*ELP6*
rs807810	3	47403892	0.400	NA	1	0.947	0.983	rs17784882	47544003	0.395	4.66×10^{-08}	*ELP6*
rs4858890	3	47405456	0.415	NA	1	0.889	0.983	rs17784882	47544003	0.395	4.66×10^{-08}	*ELP6*
rs73081205	3	47517670	0.396	NA	1	0.996	1.000	rs17784882	47544003	0.395	4.66×10^{-08}	*ELP6*
rs11708770	3	47517774	0.396	NA	1	0.996	1.000	rs17784882	47544003	0.395	4.66×10^{-08}	*ELP6*
rs11716763	3	47518266	0.397	NA	1	0.992	1.000	rs17784882	47544003	0.395	4.66×10^{-08}	*ELP6*
rs62260710	3	47564462	0.395	NA	1	1.000	1.000	rs17784882	47544003	0.395	4.66×10^{-08}	*ELP6*
rs10516950	4	95254730	0.471	NA	1	0.394	-0.744	rs3857067	95026434	0.444	1.05×10^{-07}	*SMARCAD1*
rs2905583	5	137003769	0.164	NA	1	0.272	0.624	rs10040989	137573725	0.120	2.70×10^{-09}	*GFRA3*
rs10515496	5	137044526	0.163	NA	1	0.301	0.654	rs10040989	137573725	0.120	2.70×10^{-09}	*GFRA3*
rs77915370	5	137370990	0.154	NA	1	0.375	0.707	rs10040989	137573725	0.120	2.70×10^{-09}	*GFRA3*
rs283080	6	118606000	0.556	NA	1	0.365	0.696	rs11153730	118667522	0.485	5.17×10^{-67}	*PLN*
rs445099	6	118618252	0.536	NA	1	0.559	0.828	rs11153730	118667522	0.485	5.17×10^{-67}	*PLN*
rs141430267	6	118639045	0.148	NA	1	0.159	-0.986	rs11153730	118667522	0.485	5.17×10^{-67}	*PLN*
rs7764093	6	118692304	0.656	8.21×10^{-24}	1	0.442	0.946	rs11153730	118667522	0.485	5.17×10^{-67}	*PLN*
rs7453914	6	118692981	0.563	NA	1	0.672	0.958	rs11153730	118667522	0.485	5.17×10^{-67}	*PLN*
rs4551203	6	118693502	0.093	1.41×10^{-11}	1	0.172	0.876	rs3902035	119000232	0.315	4.17×10^{-16}	*PLN*
rs72967581	6	118699329	0.093	NA	1	0.172	0.876	rs3902035	119000232	0.315	4.17×10^{-16}	*PLN*
rs7758680	6	118719413	0.141	3.18×10^{-13}	1	0.170	0.986	rs11153730	118667522	0.485	5.17×10^{-67}	*PLN*
rs3951041	6	118722296	0.671	NA	1	0.406	0.938	rs11153730	118667522	0.485	5.17×10^{-67}	*PLN*
rs4593396	6	118722514	0.135	NA	1	0.166	0.699	rs3902035	119000232	0.315	4.17×10^{-16}	*PLN*
rs76372817	6	118785780	0.055	NA	1	0.638	0.849	rs12210733	118653075	0.049	2.61×10^{-22}	*PLN*
rs3798420	6	118786486	0.534	NA	1	0.738	0.947	rs11153730	118667522	0.485	5.17×10^{-67}	*PLN*
rs11967375	6	118823975	0.457	NA	1	0.835	0.966	rs11153730	118667522	0.485	5.17×10^{-67}	*PLN*
rs34479834	6	118826244	0.536	NA	1	0.732	0.947	rs11153730	118667522	0.485	5.17×10^{-67}	*PLN*
rs7761717	6	118848565	0.735	NA	1	0.300	0.938	rs11153730	118667522	0.485	5.17×10^{-67}	*PLN*
rs11758454	6	118848799	0.455	NA	1	0.821	0.962	rs11153730	118667522	0.485	5.17×10^{-67}	*PLN*
rs13192336	6	118875162	0.536	NA	1	0.725	0.943	rs11153730	118667522	0.485	5.17×10^{-67}	*PLN*
rs763254	6	118886615	0.458	3.70×10^{-61}	1	0.831	0.962	rs11153730	118667522	0.485	5.17×10^{-67}	*PLN*
rs59671368	6	118911430	0.537	NA	1	0.722	0.942	rs11153730	118667522	0.485	5.17×10^{-67}	*PLN*
rs9489455	6	118919126	0.717	NA	1	0.331	0.942	rs11153730	118667522	0.485	5.17×10^{-67}	*PLN*

rs9489457	6	118920285	0.178	NA	1	0.426	0.951	rs3902035	119000232	0.315	4.17×10^{-16}	*PLN*
rs3862828	6	118940770	0.717	NA	1	0.331	0.942	rs11153730	118667522	0.485	5.17×10^{-67}	*PLN*
rs80153509	6	118967463	0.178	NA	1	0.426	0.951	rs3902035	119000232	0.315	4.17×10^{-16}	*PLN*
rs62422236	6	118972633	0.173	NA	1	0.417	0.958	rs3902035	119000232	0.315	4.17×10^{-16}	*PLN*
rs7746210	6	119027325	0.365	6.61×10^{-42}	1	0.482	0.889	rs11153730	118667522	0.485	5.17×10^{-67}	*PLN*
rs10499088	6	119027955	0.260	NA	1	0.748	0.989	rs3902035	119000232	0.315	4.17×10^{-16}	*PLN*
rs2888675	7	150542081	0.521	NA	1	0.110	0.477	rs3807375	150667210	0.345	1.61×10^{-32}	*KCNH2*
rs877120	7	150543134	0.150	NA	1	0.207	0.700	rs2072413	150647969	0.294	7.53×10^{-49}	*KCNH2*
rs113986050	7	150557132	0.179	NA	1	0.357	0.827	rs2072413	150647969	0.294	7.53×10^{-49}	*KCNH2*
rs2071515	7	150557471	0.269	NA	1	0.123	0.420	rs3807375	150667210	0.345	1.61×10^{-32}	*KCNH2*
rs12179	7	150557622	0.269	NA	1	0.123	0.420	rs3807375	150667210	0.345	1.61×10^{-32}	*KCNH2*
rs1805120	7	150649531	0.213	NA	1	0.470	0.957	rs3807375	150667210	0.345	1.61×10^{-32}	*KCNH2*
rs2269000	7	150650792	0.213	NA	1	0.477	0.964	rs3807375	150667210	0.345	1.61×10^{-32}	*KCNH2*
rs2269001	7	150651816	0.210	NA	1	0.468	0.964	rs3807375	150667210	0.345	1.61×10^{-32}	*KCNH2*
rs4725984	7	150668514	0.599	2.94×10^{-23}	1	0.369	-0.769	rs2072413	150647969	0.294	7.53×10^{-49}	*KCNH2*
rs3778873	7	150669867	0.153	3.56×10^{-33}	1	0.336	0.990	rs3807375	150667210	0.345	1.61×10^{-32}	*KCNH2*
rs11771808	7	150675868	0.577	NA	1	0.293	-0.719	rs2072413	150647969	0.294	7.53×10^{-49}	*KCNH2*
rs6951150	7	150681914	0.577	NA	1	0.293	-0.719	rs2072413	150647969	0.294	7.53×10^{-49}	*KCNH2*
rs12703107	7	150683629	0.216	4.72×10^{-20}	1	0.425	0.902	rs3807375	150667210	0.345	1.61×10^{-32}	*KCNH2*
rs5888424	7	150685798	0.566	NA	1	0.295	-0.737	rs2072413	150647969	0.294	7.53×10^{-49}	*KCNH2*
rs7006399	8	71029421	0.119	2.53×10^{-07}	1	0.963	1.000	rs16936870	71189342	0.123	2.52×10^{-09}	*NCOA2*
rs4738079	8	71129687	0.877	1.34×10^{-08}	0	1.000	-1.000	rs16936870	71189342	0.123	2.52×10^{-09}	*NCOA2*
rs71517442	8	71315917	0.123	NA	1	1.000	1.000	rs16936870	71189342	0.123	2.52×10^{-09}	*NCOA2*
rs7820975	8	98810194	0.469	NA	1	0.455	0.715	rs11779860	98850330	0.440	1.15×10^{-09}	*LAPTM4B*
rs4735495	8	98817263	0.455	1.62×10^{-08}	1	0.895	0.975	rs11779860	98850330	0.440	1.15×10^{-09}	*LAPTM4B*
rs13261020	8	98824239	0.461	NA	1	0.873	0.975	rs11779860	98850330	0.440	1.15×10^{-09}	*LAPTM4B*
rs34054579	8	98824633	0.461	NA	1	0.873	0.975	rs11779860	98850330	0.440	1.15×10^{-09}	*LAPTM4B*
rs142159812	8	98834639	0.461	NA	1	0.873	0.975	rs11779860	98850330	0.440	1.15×10^{-09}	*LAPTM4B*
rs2513954	8	98835358	0.496	NA	1	0.699	-0.950	rs11779860	98850330	0.440	1.15×10^{-09}	*LAPTM4B*
rs2485378	10	104059139	0.405	NA	1	0.996	1.000	rs2485376	104050006	0.404	2.67×10^{-08}	*GBF1*
rs796427	10	104066117	0.404	8.61×10^{-08}	1	0.992	0.996	rs2485376	104050006	0.404	2.67×10^{-08}	*GBF1*
rs708155	11	2396995	0.826	7.30×10^{-08}	1	0.113	0.666	rs2301696	2426984	0.548	4.21×10^{-10}	*KCNQ1*
rs800349	11	2414702	0.813	NA	0	0.121	0.660	rs2301696	2426984	0.548	4.21×10^{-10}	*KCNQ1*
rs76503999	11	2486033	0.238	NA	1	0.973	0.989	rs7122937	2486550	0.237	5.26×10^{-54}	*KCNQ1*
rs5028646	11	2496057	0.776	1.38×10^{-08}	1	0.076	0.925	rs7122937	2486550	0.237	5.26×10^{-54}	*KCNQ1*
rs4246215	11	61564299	0.370	NA	1	0.825	0.922	rs174583	61609750	0.363	1.30×10^{-10}	*FADS2*
rs174561	11	61582708	0.303	NA	1	0.756	0.995	rs174583	61609750	0.363	1.30×10^{-10}	*FADS2*
rs174562	11	61585144	0.345	NA	1	0.916	0.995	rs174583	61609750	0.363	1.30×10^{-10}	*FADS2*
rs6606689	12	110975675	0.796	7.38×10^{-08}	1	0.424	-0.955	rs3026445	110723203	0.355	9.76×10^{-09}	*ATP2A2*
rs1886512	13	74520186	0.376	4.26×10^{-08}	1	0.975	0.996	rs728926	74513122	0.372	2.04×10^{-08}	*KLF12*
rs762105	14	102830175	0.334	NA	1	0.764	-0.943	rs2273905	102974999	0.631	4.74×10^{-11}	*ANKRD9*
rs12590993	14	102844064	0.336	NA	1	0.771	-0.944	rs2273905	102974999	0.631	4.74×10^{-11}	*ANKRD9*
rs12448294	16	11677434	0.147	NA	1	0.976	0.992	rs12930096	11670758	0.148	2.88×10^{-12}	*LITAF*
rs8048558	16	11678332	0.147	NA	1	0.976	0.992	rs12930096	11670758	0.148	2.88×10^{-12}	*LITAF*
rs12921437	16	11680148	0.147	NA	1	0.930	0.968	rs12930096	11670758	0.148	2.88×10^{-12}	*LITAF*
rs7198919	16	11688891	0.319	1.14×10^{-20}	0	0.539	0.955	rs735951	11693536	0.442	3.74×10^{-28}	*LITAF*
rs8063597	16	11690807	0.318	NA	1	0.543	0.961	rs735951	11693536	0.442	3.74×10^{-28}	*LITAF*
rs735951	16	11693536	0.442	NA	1	1.000	1.000	rs735951	11693536	0.442	3.74×10^{-28}	*LITAF*
rs12927069	16	11697954	0.451	NA	1	0.138	-0.693	rs12444261	11734642	0.258	1.82×10^{-12}	*LITAF*
rs12933316	16	11699720	0.447	NA	1	0.134	-0.690	rs12444261	11734642	0.258	1.82×10^{-12}	*LITAF*
rs4781128	16	11707571	0.446	1.71×10^{-12}	1	0.137	-0.698	rs12444261	11734642	0.258	1.82×10^{-12}	*LITAF*
rs9646268	16	11727552	0.271	NA	1	0.567	0.778	rs12444261	11734642	0.258	1.82×10^{-12}	*LITAF*
rs369202186	16	11735712	0.261	NA	1	0.974	0.995	rs12444261	11734642	0.258	1.82×10^{-12}	*LITAF*
rs368548487	16	11735713	0.261	NA	1	0.974	0.995	rs12444261	11734642	0.258	1.82×10^{-12}	*LITAF*
rs1704528	16	14388750	0.347	NA	1	0.869	0.972	rs246185	14395432	0.328	8.27×10^{-11}	*MRTFB*
rs181766	16	14394878	0.329	NA	1	0.987	0.995	rs246185	14395432	0.328	8.27×10^{-11}	*MRTFB*
rs4784935	16	58460342	0.747	NA	1	0.995	1.000	rs4784934	58459926	0.748	5.11×10^{-09}	*CNOT1*
rs3809596	16	58549014	0.320	8.10×10^{-10}	1	0.150	-1.000	rs246196	58574253	0.242	2.54×10^{-57}	*CNOT1*
rs12596546	16	58582949	0.758	NA	1	1.000	-1.000	rs246196	58574253	0.242	2.54×10^{-57}	*CNOT1*
rs7201946	16	58593522	0.437	NA	1	0.248	-1.000	rs246196	58574253	0.242	2.54×10^{-57}	*CNOT1*
rs17821573	16	58611885	0.438	NA	1	0.249	-1.000	rs246196	58574253	0.242	2.54×10^{-57}	*CNOT1*
rs9921370	16	58611935	0.758	NA	1	0.995	-1.000	rs246196	58574253	0.242	2.54×10^{-57}	*CNOT1*
rs9926577	16	58613431	0.758	NA	1	1.000	-1.000	rs246196	58574253	0.242	2.54×10^{-57}	*CNOT1*
rs27097	16	58664065	0.758	1.20×10^{-53}	1	0.968	-0.984	rs246196	58574253	0.242	2.54×10^{-57}	*CNOT1*
rs9635769	17	33288363	0.570	3.55×10^{-09}	1	0.279	0.640	rs1052536	33331575	0.474	2.92×10^{-24}	*LIG3*
rs11650512	17	33356907	0.479	NA	1	0.949	0.984	rs1052536	33331575	0.474	2.92×10^{-24}	*LIG3*
rs797989	17	33414758	0.478	NA	1	0.953	0.984	rs1052536	33331575	0.474	2.92×10^{-24}	*LIG3*
rs9910355	17	64315205	0.530	1.13×10^{-13}	1	0.972	0.992	rs9892651	64303793	0.533	5.86×10^{-14}	*PRKCA*
rs11653587	17	68316678	0.206	NA	1	0.630	1.000	rs10775360	68325868	0.291	6.32×10^{-13}	*KCNJ2*
rs1605749	17	68475145	0.532	2.65×10^{-22}	1	0.968	0.992	rs1396515	68430993	0.536	7.40×10^{-25}	*KCNJ2*

Supplemental Table S5. deltaSVM prediction of 14 variants tested by luciferase assays in iPSC-derived cardiomyocytes.

rsID	chr	start	end	dSVM specific	dSVM generic	luciferase p-value
rs10030238	chr4	141808804	141808805	-0.60	-0.68	9.72E-16
rs10089107	chr8	141698450	141698451	-0.45	0.10	2.90E-01
rs1044503	chr14	102965995	102965996	-0.11	1.30	6.07E-10
rs11119843	chr1	212248322	212248323	-0.52	-0.12	2.68E-03
rs11263841	chr1	35309308	35309309	0.24	-0.44	2.20E-01
rs1275988	chr2	26914363	26914364	0.10	0.06	1.50E-02
rs1451509	chr17	57443132	57443133	0.59	0.88	2.00E-03
rs17779853	chr17	30065712	30065713	-0.65	-0.14	4.33E-03
rs196067	chr22	38853676	38853677	0.45	1.13	1.60E-02
rs2305054	chr2	220506521	220506522	-0.44	-0.04	6.40E-02
rs3734637	chr6	126081318	126081319	0.24	0.34	1.96E-04
rs4904569	chr14	89875278	89875279	-0.57	-0.84	9.04E-08
rs6565060	chr16	82750050	82750051	-1.19	-1.48	5.00E-03
rs7783216	chr7	103255194	103255195	0.10	-0.37	3.50E-01

* Statistically significant values are highlighted in red.

Supplemental Data S1: Enrichment tests of PWM matched 11-mers in the top 5th percentile of the SVM weight distribution.

Supplemental Data S2: Significant deltaSVM SNPs overlapping heart CREs for both generic and specific models

Supplemental Data S3: Regional association plots for each of the 34 genome-wide significant loci for the QT interval GWAS meta-analysis, except for the KCNE1 locus that contains a rare missense mutation. For each locus, only the HapMap3 imputed SNPs used for GWAS meta-analysis (top) and those that also overlap heart DHSs (bottom) are shown. Purple circles denote the predicted cardiac regulatory variants that overlap DHSs as well as have significant deltaSVM scores from the specific model. All independent index SNPs are labeled and marked as diamonds, and neighboring SNPs are color-coded with different shapes based on their highest LD with these index SNPs. For LD, we calculated both r^2 (top left) and D' (top right) for comparison. Dashed lines indicate genome-wide significant P-values.

Supplemental Data S4: Position weight matrices from the CISBP database in the MEME format.

Supplemental Data S5: Position weight matrices for C2H2 Zinc-finger transcription factors in the MEME format.

Supplemental Data S6: Custom scripts for allele-biased DHS analysis, TF gene expression analysis, and DHS heatmap analysis.