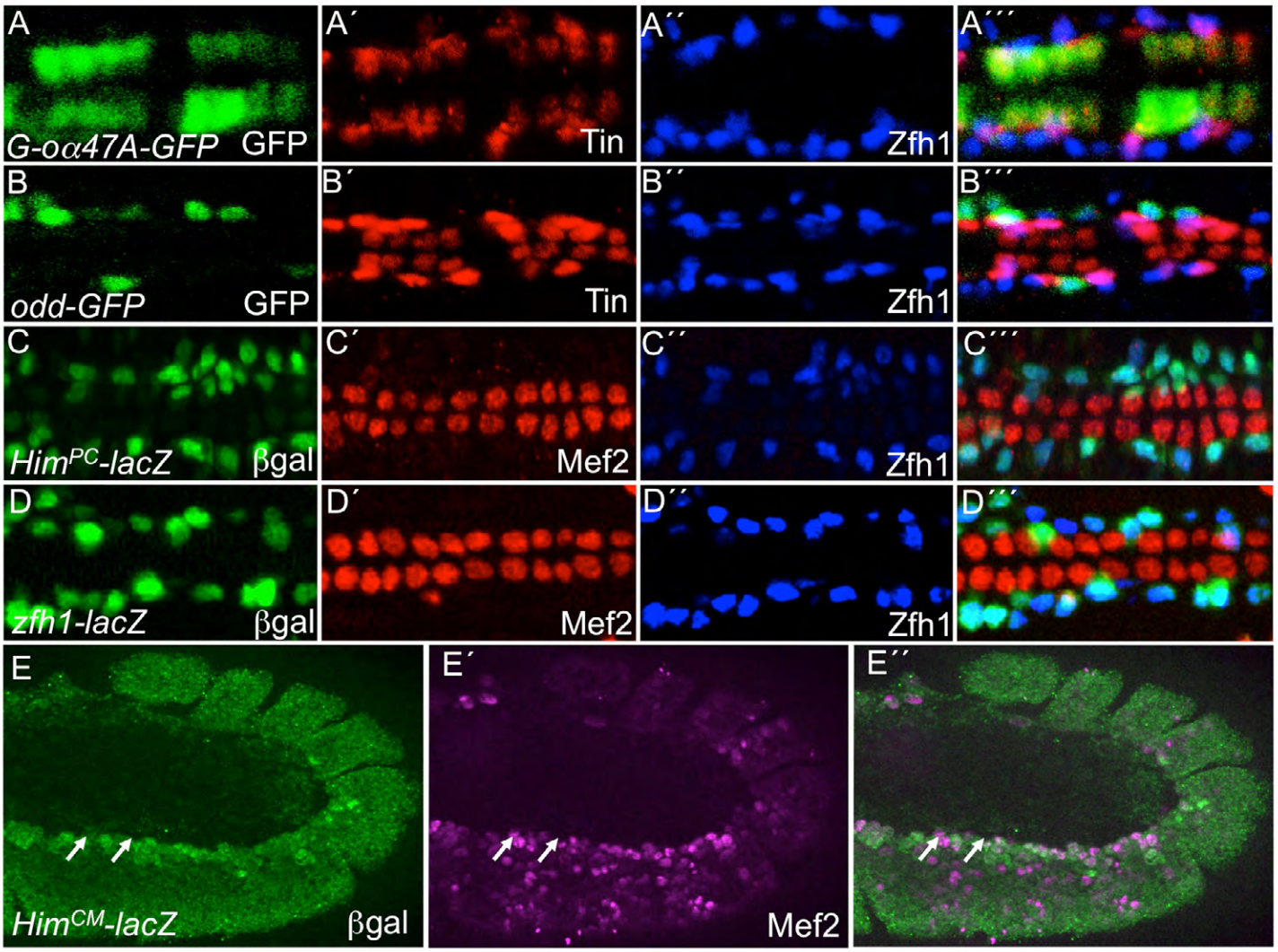
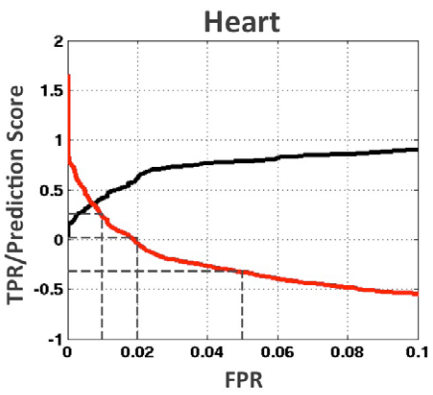
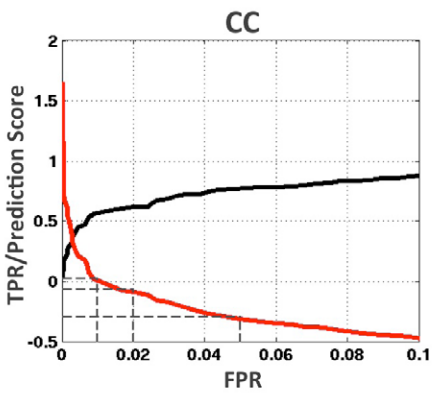


Supplementary Fig. S1. Building a training set of cardiac enhancers. (A-E) Empirical validation of candidate enhancers containing matches to Twi and Tin TFBS motifs and located in the flanking or intronic sequences associated with previously characterized heart genes (Warner et al., 2008). Enhancer coordinates are described in supplementary material Table S1. Antibody staining of stage 16 (A-D) and stage 11 (E) embryos containing *G-oα47A-GFP* (A), *odd-GFP* (B), *Him*^{PC}*-lacZ* (C), *zfh1-lacZ* (D) and *Him*^{CM}*-lacZ* (E) transgenes using antibodies against GFP (A,B), β-galactosidase (C-E), Tin (A-B), Mef2 (C-E) and Zfh1 (A-D). Panels A-E represent the activity of the relevant *lacZ* or *GFP* reporter, while Tin (Tin-expressing CCs and PCs), Zfh1 (all PCs) and Mef2 (all CCs) were used to stain and distinguish different cardiac cell types. *zfh1-lacZ* is active in all PCs whereas the enhancers for *G-oα47A-GFP* and *odd-GFP* are active in subsets of the PCs, with *G-oα47A-GFP* restricted to the 4 Tin-expressing PCs and *odd-GFP* is present in all of the Odd-expressing PCs. Separate enhancers control the activity of the Him gene, as *Him*^{PC}*-lacZ* is active in all PCs at stage 16 whereas *Him*^{CM}*-lacZ* is only active in cells of the cardiac mesoderm (arrows) at stage 11.

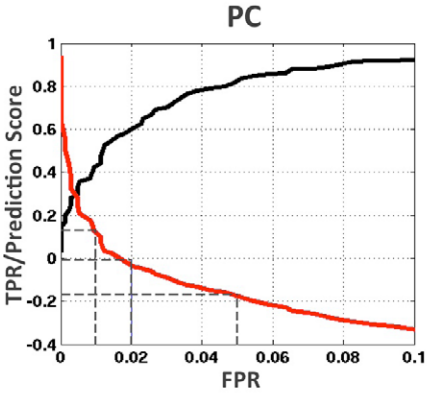
A) motif+ChIP



False Positive Rate	Score	Number of positive predictions
0.01	0.26	2682
0.02	0.02	5925
0.05	-0.25	12294

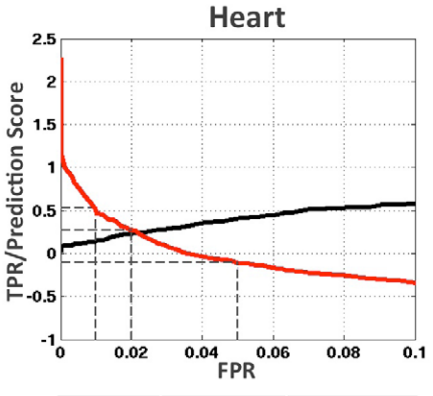


False Positive Rate	Score	Number of positive predictions
0.01	0.03	2962
0.02	-0.07	4303
0.05	-0.30	10804

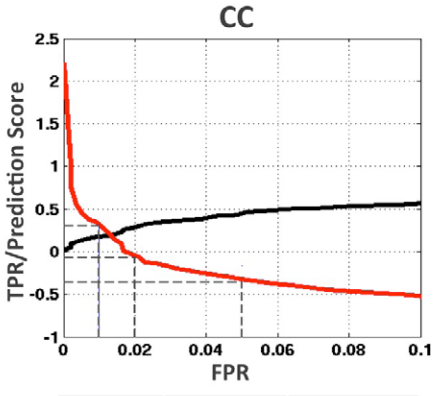


False Positive Rate	Score	Number of positive predictions
0.01	0.17	1907
0.02	-0.01	3330
0.05	-0.19	7124

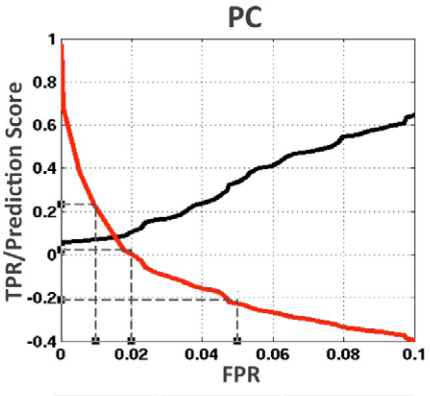
B) motif-alone



False Positive Rate	Score	Number of positive predictions
0.01	0.53	1787
0.02	0.27	4832
0.05	-0.1	9619

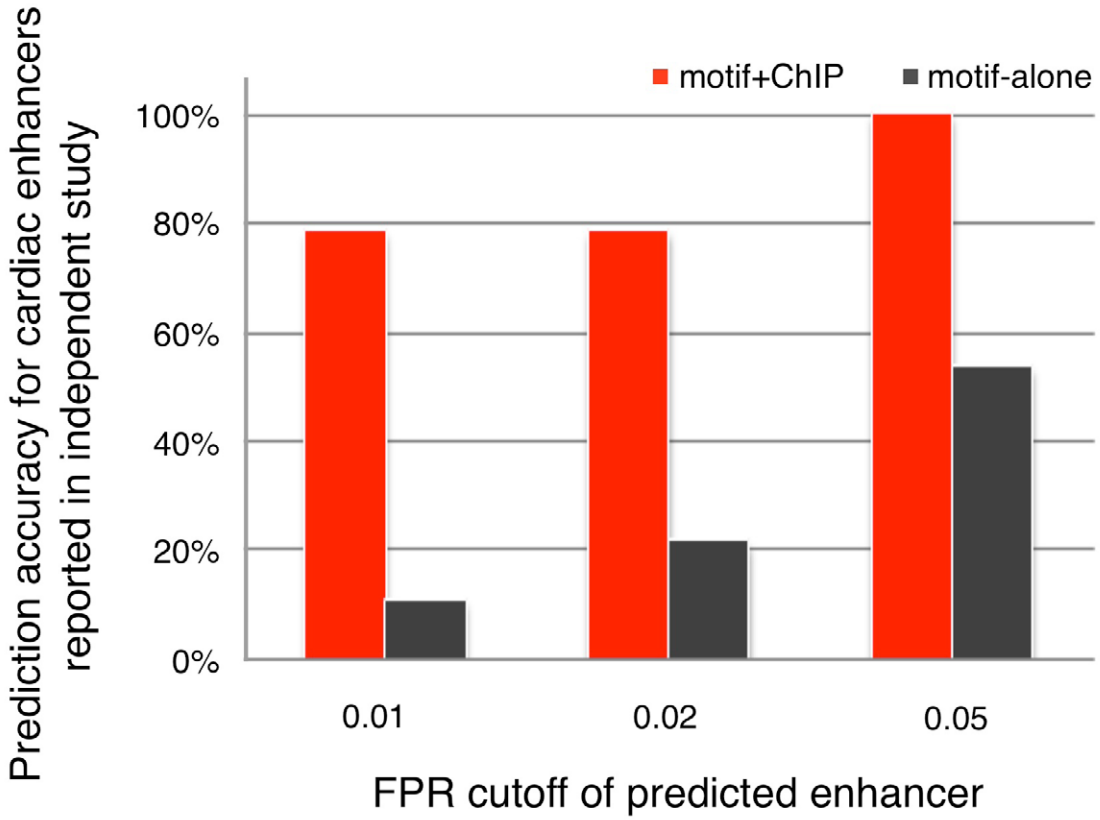


False Positive Rate	Score	Number of positive predictions
0.01	0.31	1285
0.02	-0.03	6339
0.05	-0.35	18457

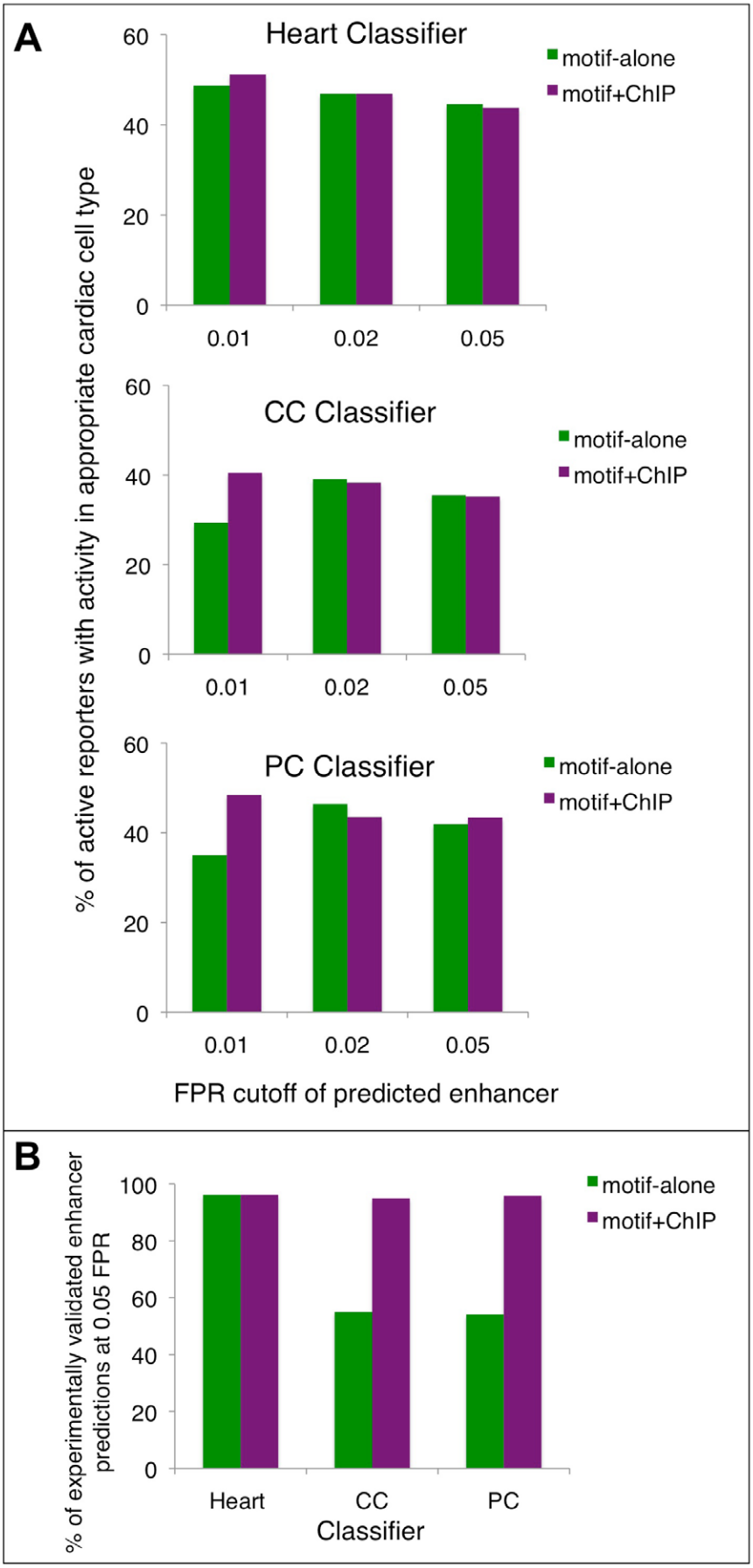


False Positive Rate	Score	Number of positive predictions
0.01	0.22	2097
0.02	0.01	5161
0.05	-0.20	12009

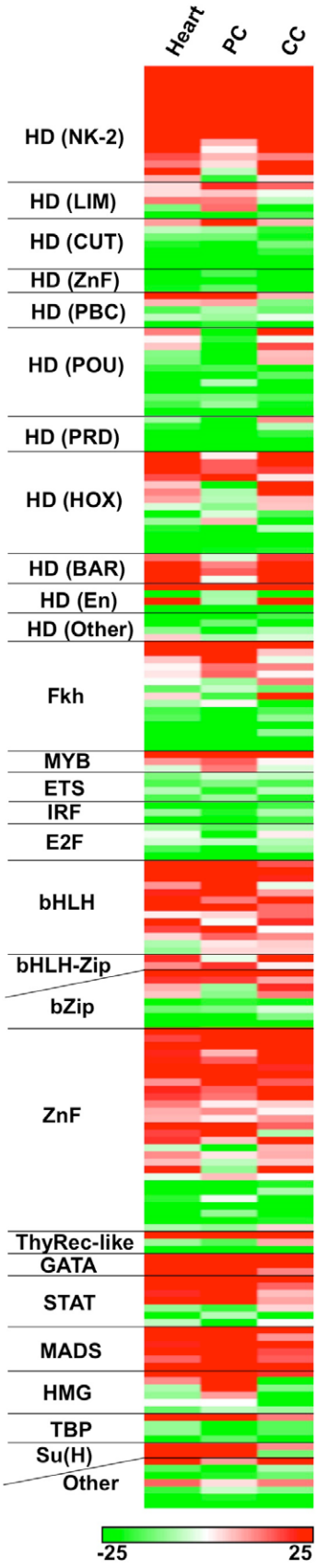
Supplementary Fig. S2. Prediction of cardiac enhancers at different false positive rates (FPRs). Prediction score cutoffs and true positive rates (TPRs) were plotted for different FPRs. In the tables, three typical FPRs (i.e. 0.01, 0.02 and 0.05) and the corresponding prediction results are shown.



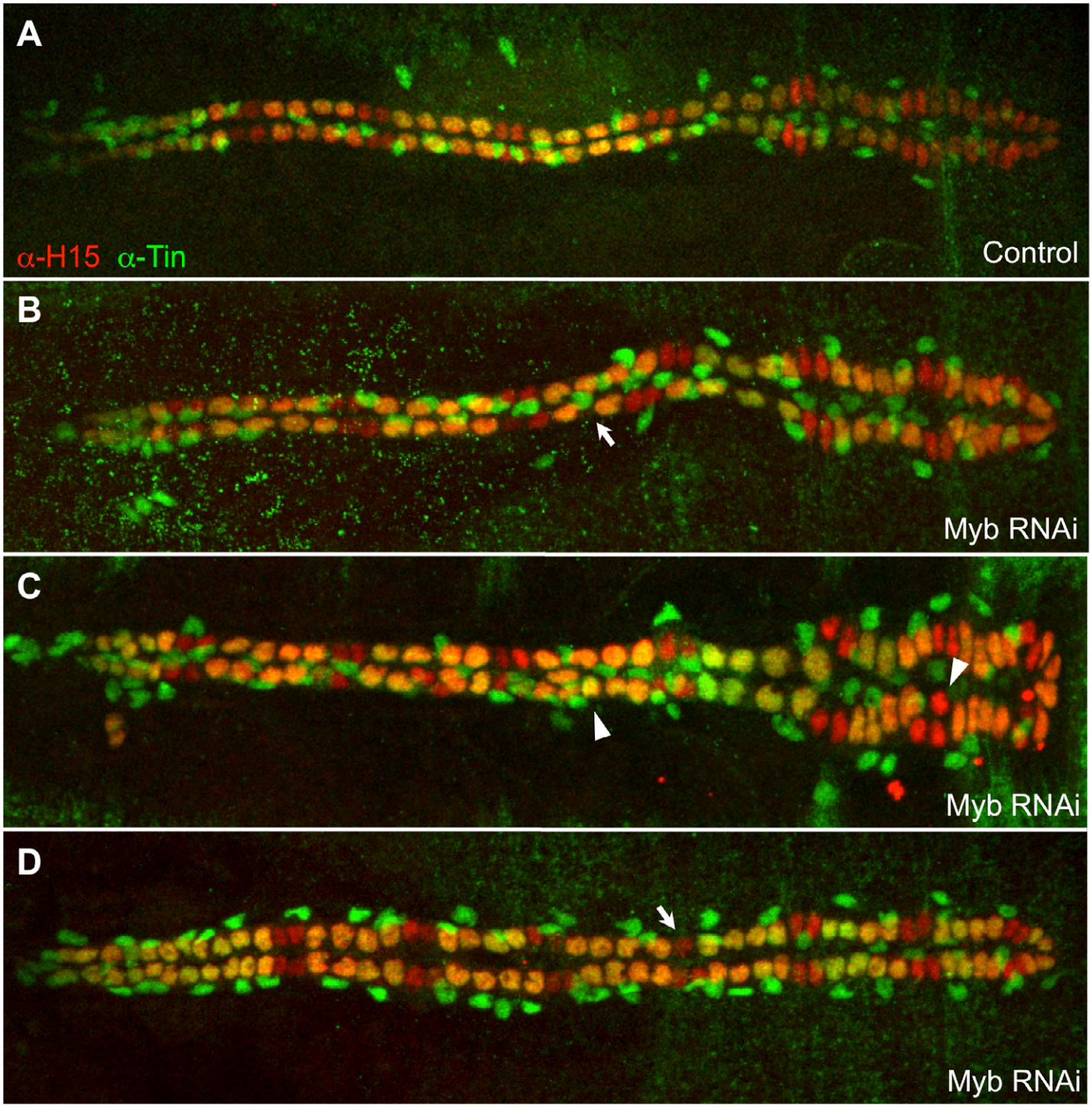
Supplementary Fig. S3. Prediction accuracy of heart classifiers for cardiac enhancers reported in an independent study. Histogram showing the fraction of *in vivo* verified cardiac enhancers from an independent study (Jin et al., 2013) that were accurately predicted by the motif+ChIP and motif-alone heart classifiers at different FPRs.



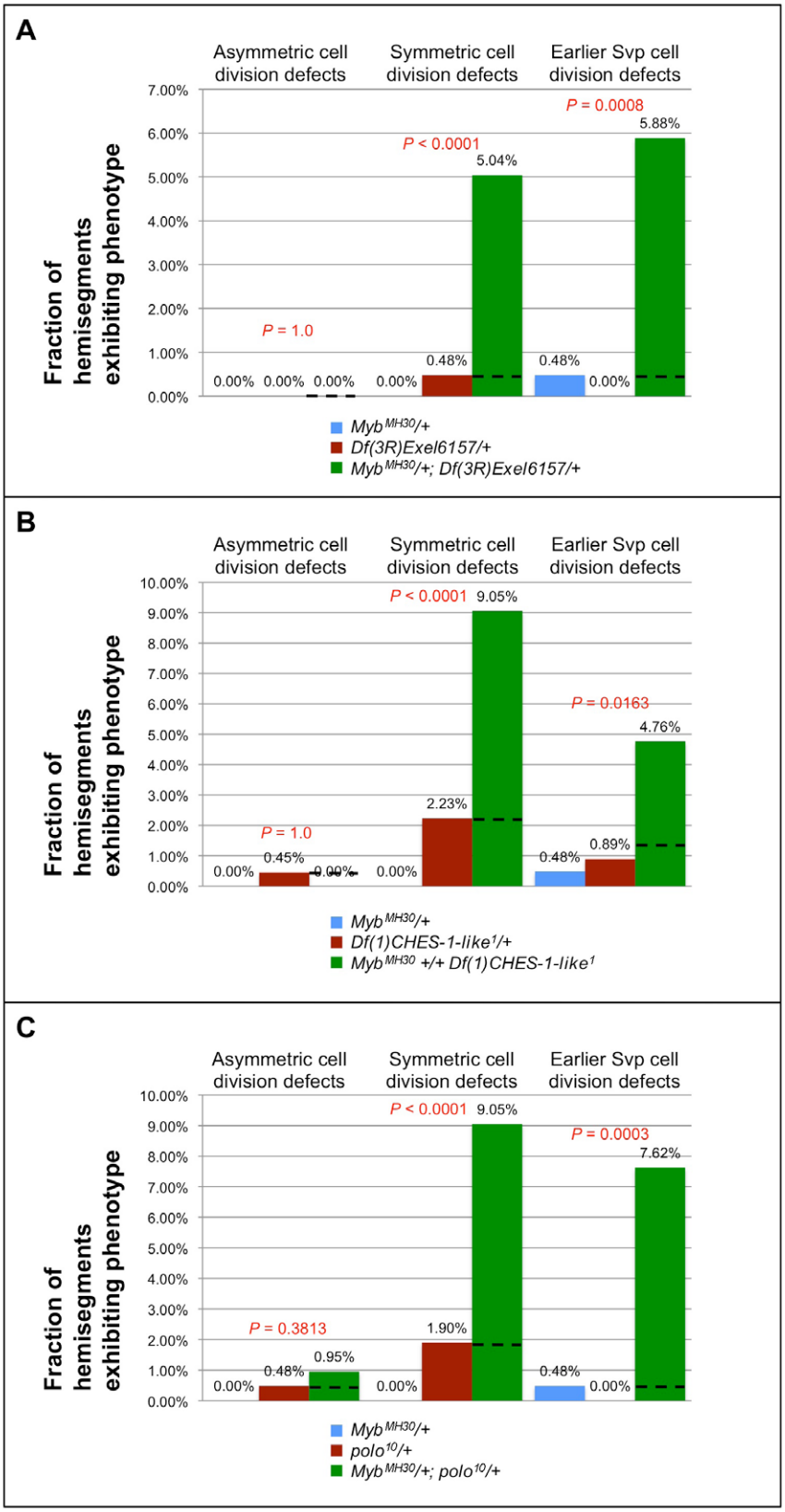
Supplementary Fig. S4. Empirical validation of classifier predictions. (A) Percentages of predicted enhancers with activity in the appropriate cardiac cell types for the motif-alone and motif+ChIP classifiers at different FPRs. (B) Percentages of successful enhancer predictions with activity in the appropriate cardiac cell types for the motif-alone and motif+ChIP classifiers at 0.05 FPR



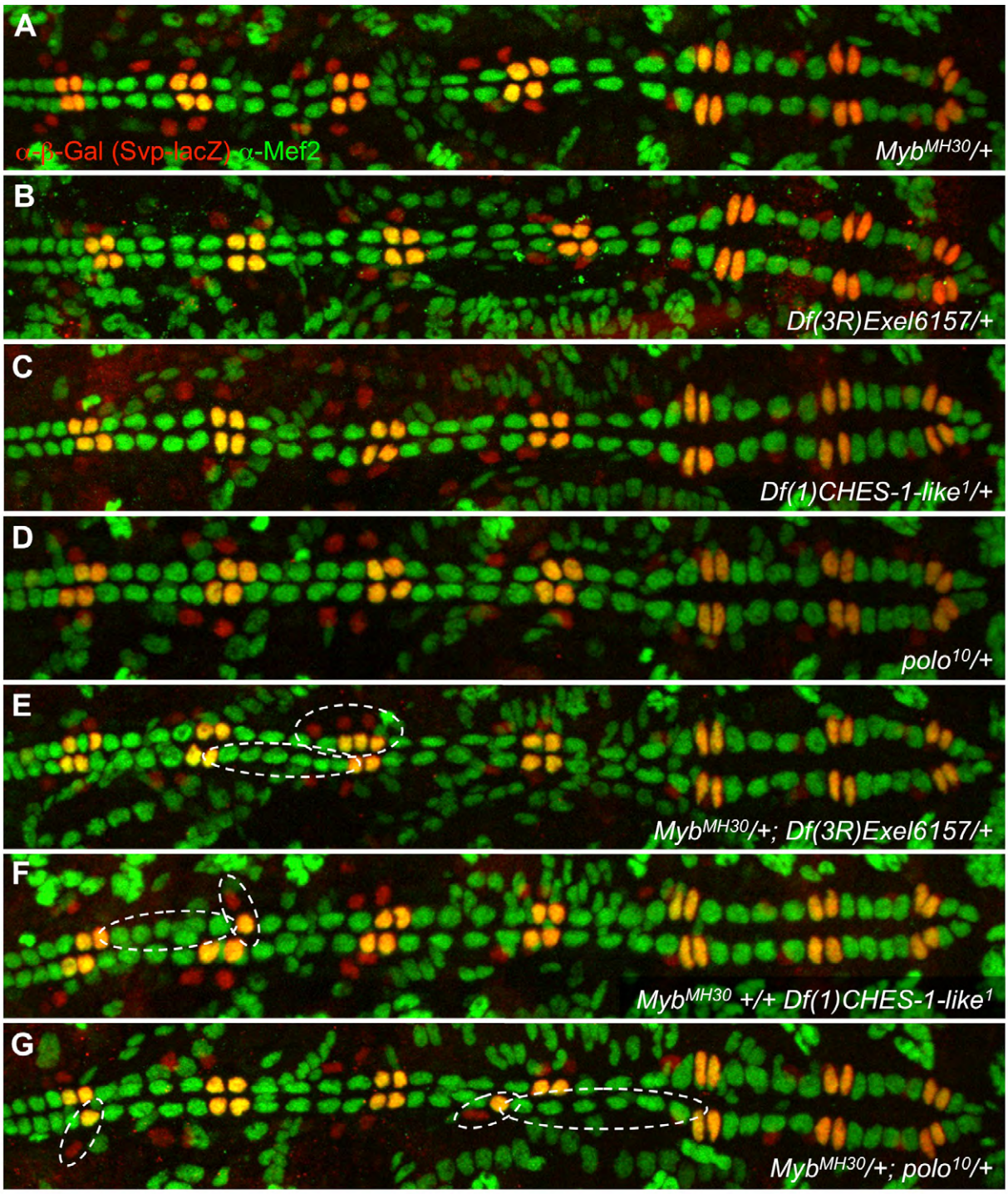
Supplementary Fig. S5. Machine learning reveals motif features associated with cardiac enhancer subtypes. A linear SVM was trained for each cardiac training set using only TF binding motifs (motif-alone classifier). Motifs are represented according to their linear SVM weights (a reflection of the discriminating power of these motifs). For each cardiac training set, the binding motifs were grouped according to the DNA-binding domain of the TF. See supplementary material Table S3 for motifs and motif weights. Similar results were seen for the motif+ChIP classifier (see Fig. 3).



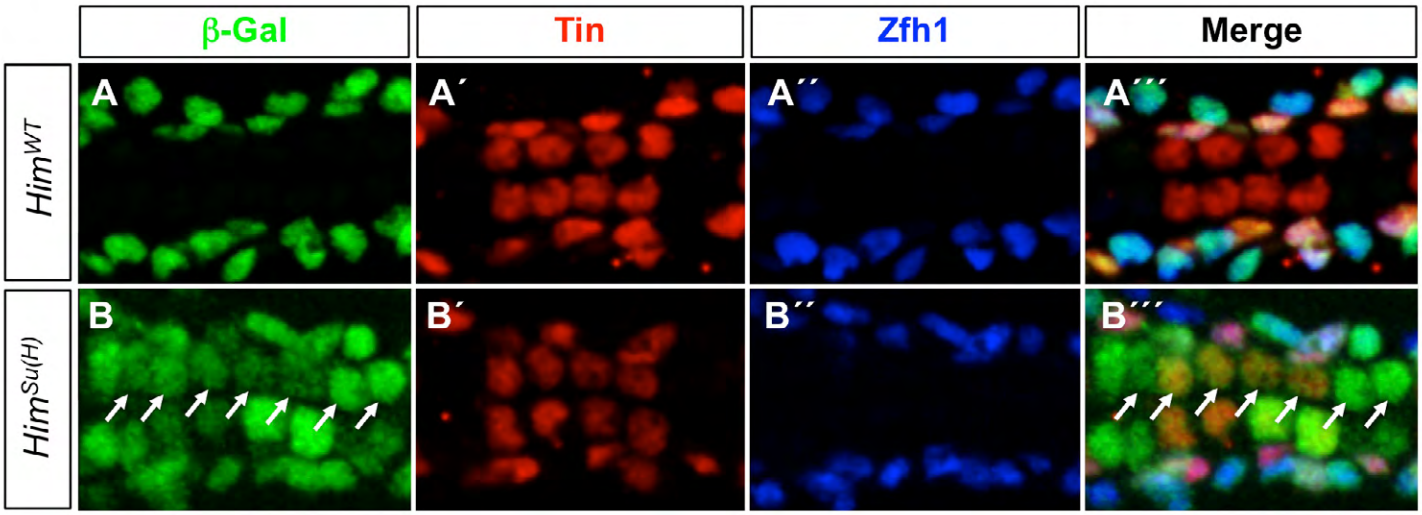
Supplementary Fig. S6. Cardiac phenotypes associated with knocking down Myb levels specifically in the developing heart using a targeted RNAi-based strategy. (A-D) Hearts from embryos stained with antibodies against Tin and with a CC-specific antibody against H15 such that Tin-CCs (yellow) can be distinguished from Svp-CCs (red). Anterior is to the left. (A) Heart from a control embryo. Note that every hemisegment except the posteriormost (A8) hemisegments includes four Tin-CCs and two Svp-CCs. (B-D) Hearts from embryos in which RNAi driven by both the cardiac-speccific drivers TinD-GAL4 and Hand-Gal4 is utilized to knock down Myb levels specifically in the developing heart. Localized reductions (arrows) in the number of both Tin-CCs (B) and Svp-CCs (D), as well as localized increases in both Tin-CCs and Svp-CCs (arrowheads in C), are detected.



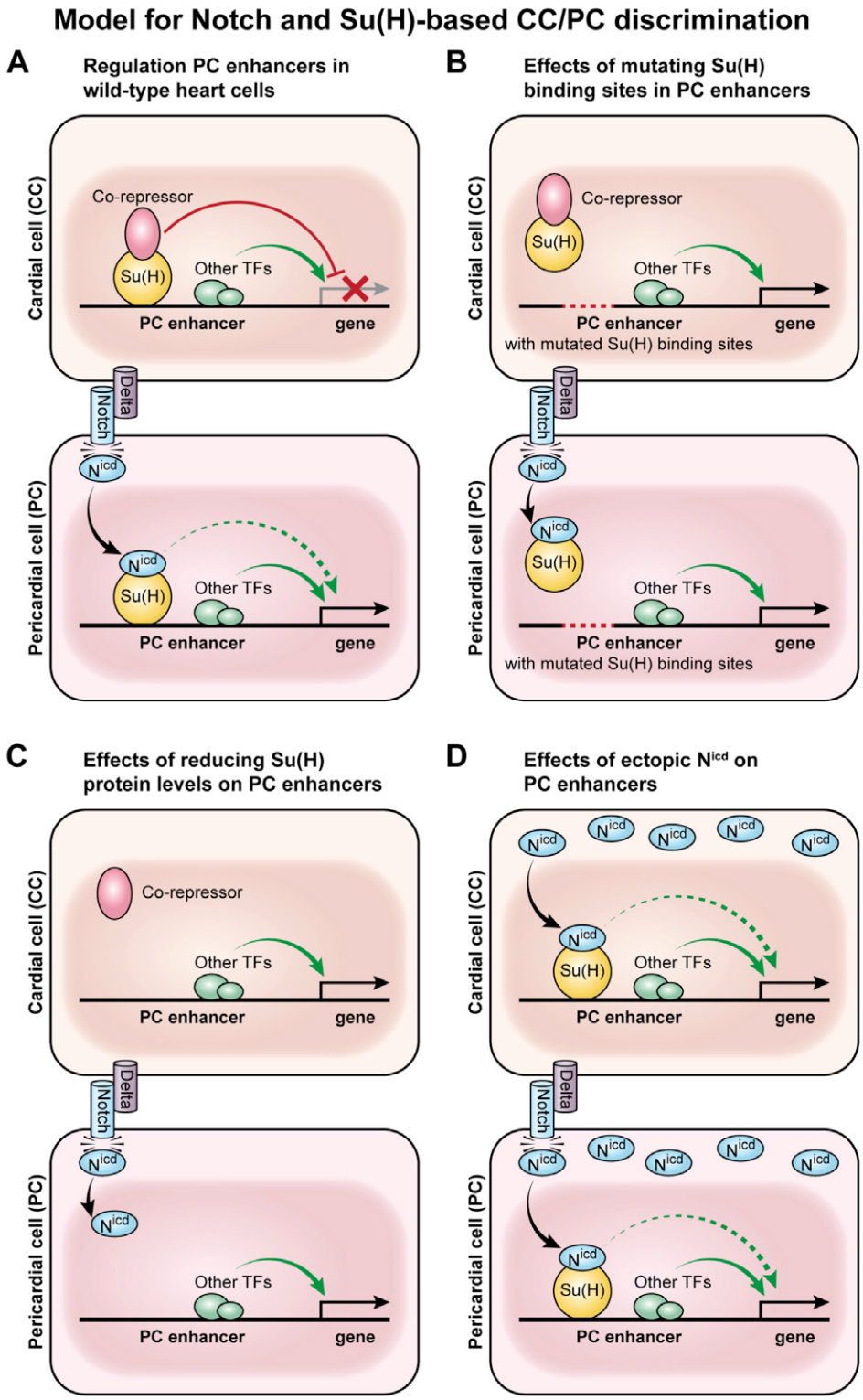
Supplementary Fig. S7. Synergistic interactions between the genes encoding Myb, Jumu, CHES-1-like and Polo proteins. (A) Fraction of hemisegments exhibiting asymmetric, symmetric, and earlier cell division defects affecting Svp progenitor numbers for single and double heterozygotes of a deficiency, *Df(3R)Exel6157*, which completely eliminates the *jumu* gene, and Myb^{MH30}, a null mutation in *Myb*. (B) Fraction of hemisegments exhibiting the three types of cardiac progenitor cell division defects for single and double heterozygotes of null mutations in *Myb* and *CHES-1-like*. (C) Fraction of hemisegments exhibiting the cardiac progenitor cell division defects for single and double heterozygotes of the *Myb* null mutation, and a strong hypomorphic mutation in *polo*. In each case, the black dashed line indicates the expected results in the double heterozygotes if the phenotypes were purely additive. See Supplementary Fig. S8 for representative images of these cardiac progenitor cell division defects.



Supplementary Fig. S8. Cardiac phenotypes associated with single and double heterozygotes for mutations in *Myb*, *jumu*, *CHES-1-like* and *polo*. (A-D) Representative hearts from embryos that are individually heterozygous for a null mutation in *Myb* (A), a deficiency that completely eliminates *jumu* (B), a null mutation in *CHES-1-like* (C), or a strong hypomorphic mutation in *polo* (D) are shown. Note that these single heterozygotes typically do not exhibit cardiac progenitor cell division defects. (E-G) In contrast, representative hearts from embryos doubly heterozygous for mutations in *Myb* and either *jumu* (E), *CHES-1-like* (F) or *polo* (G) exhibit significant cardiac phenotypes (dashed ovals) associated with defects in symmetric cell division and earlier cell divisions affecting the number of Svp progenitors. See Supplementary Fig. S7 for quantitation.



Supplementary Fig. S9. The Su(H) binding site is utilized to repress expression of the *Him* PC enhancer in CCs. The posterior-most four CCs are marked by Tin expression (red), and the PCs are marked by Zfh1 expression (blue). (A-A´´´) A β-galactosidase reporter (green) driven by the wild-type *Him* enhancer (Him^{WT}) is expressed only in the Zfh1-expressing PCs. (B-B´´´) When the Su(H) binding site is mutated in the *Him* enhancer ($Him^{Su(H)}$), the reporter is ectopically expressed in CCs (arrows). Its expression in all Zfh1-expressing PCs remains unaltered.



Supplementary Fig. S10. Schematic of the involvement of the Notch signaling pathway in the lineage decision between PCs and CCs for PC enhancers. Modes of regulation activating and repressing target genes are shown as green and red arrows respectively. (A) In cardial cells, the enhancers of PC genes such as *Him* are repressed by the Su(H)-co-repressor complex. The Delta ligand expressed by CCs activates Notch receptor in neighboring PCs, with the resulting cleaved N^{icd} fragment associating with Su(H) and displacing the co-repressor. The consequent elimination of repressor complex binding in PCs is sufficient to initiate transcription due to the presence of other local TF activators, which is enhanced further by the N^{icd}-Su(H) complex in PCs. (B) Mutating the Su(H) sites in PC enhancers prevents the Su(H)-co-repressor complex from binding to the enhancers in CCs. The resulting alleviation of repression is sufficient to ectopically transcribe the associated gene in CCs. (C) Similarly, minimizing the level of Su(H) protein by RNA interference reduces the formation and subsequent binding of the Su(H)-co-repressor complex to PC enhancers in CCs, leading to de-repression of the associated PC genes in CCs. (D) Ubiquitous Notch signaling drives expression of target genes for PC enhancers in all cells of the heart, both as a consequence of the alleviation of repression by the elimination of the Su(H)-co-repressor complex, and due to direct activation by the N^{icd}-Su(H) complex.

Supplementary Table S1. The genomic coordinates of heart, PC and CC enhancers, and orthologous sequences comprising the training sets utilized in this study, along with a list of genes with validated expression in the heart and its subsets (Ahmad et al., 2012). Enhancer orthologs were extracted from the 11 other sequenced *Drosophila* species by searching for regions with at least 50% but less than 80% sequence identity and similar length, GC-content and repeat density as their *D. melanogaster* counterparts (Busser et al., 2012a).

Download Table S1

Supplementary Table S2. Design of the motif-alone and motif+ChIP classifiers, the genomic coordinates and rank of all predicted enhancers from the classifiers, and descriptions of the activities of tested predicted enhancers. To build cell type-specific enhancer prediction models, we generated controls that were randomly sampled from *D. melanogaster* non-coding regions and had similar length, GC- and repeat-content to the training enhancers. Ten control sequences were retrieved for each training enhancer. Each sequence (either enhancer or control) was then scanned using MAST (Bailey and Gribskov, 1998) for 1019 TF binding motifs that were collected from TRANSFAC, JASPAR and uniProbe libraries (Wingender et al., 2001; Sandelin et al., 2004; Berger et al., 2006) and were present among our sequences. To this end, each DNA sequence was converted into a 1019-dimension vector in which a value indicates the counts of TF binding motif per 1000 bp along the considered sequence. We then used a linear support vector machine (SVM) (Cortes and Vapnik, 1995) provided in the libSVM library (Chang and Lin, 2011) to discriminate between enhancers and controls. We used a standard 10-fold cross-validation procedure to assess the discrimination capability of the constructed classifiers. In this procedure, the training set enhancers and corresponding controls were randomly partitioned into 10 disjointed and equal-sized subsets, with each subset being used in turn to test a cell type-specific classifier built with the remaining 9 subsets. In order to evaluate the classification performance reliably, we ran this 10-fold cross-validation procedure 20 times with independent partitioning of samples. See Fig. 2 for the results. In addition, to avoid information leak during cross-validation procedure, an enhancer and its orthologs were always put into the same sample set (either training or test sample set). With the trained classifiers (heart, CC, and PC), we scanned the genome of *D. melanogaster* (BDGP Release 5 assembly) to predict new enhancers. A 500 bps sliding window with a 250 bps incremental step was used for the genome scan. In total, we scored 376,586 sequences. The enhancer cutoff score was set according to the FPR established using a 10-fold cross-validation. Using the setting of FPR=0.01, we detected 2682 heart enhancers, 2962 CC enhancers and 1907 PC enhancers by using motif+ChIP classifiers (Supplementary Fig. S2). Prediction accuracy for each classifier was evaluated by examining the fraction of cardiac enhancers reported in an independent study (Jin et al., 2013) at different FPR cutoffs (Supplementary Fig. S3) and by utilizing genomic site-specific transgenic reporter assays to test 80 enhancer predictions with varying scores in the classifier rankings for the different enhancer models (Fig. 3, Supplementary Fig. S4).

Download Table S2

Supplementary Table S3. DNA sequence motifs and weighting factors identified by the motif-alone and motif+ChIP classifiers.

Download Table S3

Supplementary Table S4. Quantitative summary and statistical significance of the effects of mutating Myb binding sites in the *Ndg* enhancer, the effects of loss-of-function of Myb on the activity of the WT *Ndg* enhancer, and detailed quantitation of the cell division defects associated with the different genotypes examined in this study. Confidence intervals for the mean number of *Ndg* enhancer-expressing CCs per hemisegment were computed using bootstrap methods (Davison and Hinkley, 1997) (see also Fig. 5). Specifically, for a given genotype, the embryos were sampled with replacement to construct a sample of the original size and the mean number of *Ndg* enhancer-expressing CCs per hemisegment was calculated. This process was repeated 10,000 times and an empirical distribution of mean values was obtained. The 95% confidence interval is given by ($mean_{0.025}$, $mean_{0.975}$) where $mean_{0.025}$ designates the 250th smallest of the 10,000 empirical means and $mean_{0.975}$ designates the 9570th smallest of the means. Permutation testing (Good, 1994) was used to obtain p-values for testing for differences between *Ndg* enhancers (supplementary material Table S4A). Permutation testing (Good, 1994) was also used to obtain p-values for testing the hypothesis that the average number of defects per hemisegment is equivalent in two genotypes (supplementary material Table S4C, Row 1). A bootstrap approach (Davison and Hinkley, 1997) was used to obtain p-values for determining whether non-additive interactions exist among mutation types (supplementary material Table S4C, Rows 2-4). A bootstrap sample was drawn from the genotype with both mutations and the proportion of cell division errors for the genotype was calculated. This average was subtracted from the sum of averages obtained from bootstrap samples of each of the two genotypes with one mutation. From this subtraction a single estimate of the interaction was obtained. The procedure was repeated 10,000 times. P-values for the hypothesis of no interaction were obtained by examination of the proportion of 10,000 bootstrapped interactions above and below 0.

Download Table S4