SUPPLEMENTAL MATERIAL

Expanded Materials & Methods

Preparation of Human Tissue for ChIP-Seq and RNA-Seq

Use of human embryonic tissue was reviewed and approved by the Human Subjects Protection Program at UConn Health (UCHC 710-2-13-14-03. Human embryonic craniofacial tissue was collected, staged and provided by the Joint MRC/Wellcome Trust Human Developmental Biology Resource (www.hdbr.org). Tissues were flash frozen upon collection and stored at -80°C. For both ChIP-seq and RNA-seq frozen human embryonic heart tissues frozen were removed from tubes to a petri dish with cold PBS using 1ml of cold PBS and a cut pipette tip. Hearts were photographed from at least two aspects and the tube with embryo ID photographed for records. Hearts were homogenized by mechanical disruption and divided between samples for RNA-seq and ChIP-seq as necessary. For ChIP-seq, tissues were fixed by incubation in 1% formaldehyde for 15 minutes at room temperature with agitation before being quenched by addition of 2.5M glycine to a concentration of 150mM with rotation/agitation for 10 minutes. Tissues were spun down, washed once with PBS, and placed in a dry ice- ethanol slurry to flash freeze. For RNA-Seq homogenized tissue was added to Qiazol (Qiagen) in a non-stick 1.5ml eppendorf tube, inverted to mix and placed in a dry ice- ethanol slurry to flash freeze.

ChIP-Seq

Fixed cells pellets were processed for ChIP as previously described.[106] Briefly, samples were thawed in 1 mL of 1x Cell Lysis buffer and incubated on ice for 20 minutes. Cells were lysed with dounce homogenization and nuclei were collected by centrifugation (5 min, 2500g, 4°C). Nuclei were resuspended in 300 µL of 1x Nuclear Lysis buffer + 0.3% SDS + 2 mM sodium butyrate and incubated on ice for 20 minutes. Chromatin was sheared with a Qsonica Q800R1 sonicator system operating at amplitude 23 and 2°C for 30 minutes (10 seconds duty, 10 seconds rest). Samples were cleared by centrifugation (5 min, 20,000g, 4°C) and soluble chromatin was transferred equally into seven separate tubes with 10% reserved as an input control. SDS concentration was reduced to 0.18% with ChIP-seq Dilution buffer. Protein G Dynabeads (ThermoFisher) separately preloaded with 2.5-5 µg of antibodies were added to each chromatin aliquot. Antibodies used in this study were as follows: anti-H3K27ac (C15410196, Diagenode), anti-H3K4me1 (C15410194, Diagenode), anti-H3K4me2 (ab7766, Abcam), anti-H3K4me3 (C15410003, Diagenode), anti-H3K27me3 (C16410195, Diagenode), anti-H3K9me3 (C15410193, Diagenode) and anti-H3K36me3 (C15410192, Diagenode). All Diagenode antibodies came pre-validated for ChIP, and the antibody from abcam (H3K4me2) was validated using Absurance H3 Histone Peptide Array (16-667, Millipore). ChIP samples were incubated overnight at 4°C on a rotisserie. The chromatin was then immunoprecipitated on a magnet and the supernatant was discarded. Beads were washed 8 times with 1 mL of 500 mM LiCl ChIP-Seq Wash Buffer and once with 1 mL of TE. Chromatin was eluted from the beads twice with ChIP Elution buffer at 65°C for 10 minutes with constant agitation. Combined eluates for each ChIP were subjected to crosslink reversal overnight at 65°C. Samples were then sequentially treated with RNAse A and proteinase K, purified with a PCR Purification Kit (Qiagen), and eluted in 40 uL of EB. ChIP samples were then quantified with picoGreen (ThermoFisher) and ChIP-seq libraries were prepared (SMARTer® ThruPLEX® DNA-seq 48S Kit, R400427, Takara Bio USA), then quantified by qPCR (NEBNext Library Quant Kit for Illumina E7630L), multiplexed, and sequenced for 75 cycles across multiple flow cells on an Illumina NextSeq 500 instrument using a NextSeq 500/550 High Output v2 kit (75 cycles, Cat No. FC-404-2005).

ChIP-Seq Data Analysis

Quality control was performed on ChIP-seq reads using FastQC (version [v.] 0.11.5) and MultiQC (v.1.1). Trimming for adapters, quality and length was performed using Trimmomatic (v.0.36) for single end data. ChIP-seq reads were aligned to the human genome (hg19) using Bowtie2 (v. 2.2.5).[107] Fragment sizes of each library were estimated using PhantomPeakQualTools (v.1.14).[108] We then generated p value-based signal tracks relative to appropriate input controls based on estimated library fragment size using MACS2 (2.1.1.20160309).[109] Bedgraph files for all p-value signals from primary ChIP-Seq data were converted to 25 bp resolution and processed for model training and generation of imputed signals for all samples using ChromImpute (v1.0.3) as previously described.[37] Resulting imputed signal tracks were converted to bigWig format for display in UCSC genome browser and converted for use with ChromHMM (v1.12)[110], using ChromImpute's ExportToChromHMM. Signal files for individual chromosomes for each epigenome were binarized and segmentation was performed using the previously published 25-state chromatin models using ChromHMM as previously described.[19] Following segmentation, annotation of states and generation of genome browser files was performed based on annotations provided by Roadmap Epigenome. The accession number for the ChIP-seq signals, imputed signal files and chromatin state segmentations reported in this paper is GEO: GSE137731.

Global Multi-tissue Comparisons of ChIP-Seq Signals

To qualitatively assess the similarity per mark of human embryonic heart ChIP-signals generated and imputed in this paper across all time points, we extracted primary and imputed signals from all samples for all marks in 10kb bins across the genome using the multiBigwigSummary command from DeepTools (v2.0)[111] excluding all regions blacklisted by ENCODE, and outputting raw counts. The PCA plot was made using the prcomp() function on the transformed matrix of these counts using the built-in R stats (v3.4.1) package. To qualitatively assess similarity of human embryonic heart samples generated here to other tissues throughout the human body, we assembled a list of all enhancer states (states 13 through 18) from 127 tissues profiled by Roadmap Epigenome, craniofacial samples previously profiled from our laboratory and those identified from human embryonic heart profiled here.[19,112] We next extracted imputed H3K27ac signals from all samples at all enhancer regions using the multiBigwigSummary command from DeepTools (v2.0) excluding all regions blacklisted by ENCODE. The resulting signal matrix were filtered to remove regions were signal was low (>10) across all samples (n = 163) and log10 transformed. This transformed matrix was used to calculate the Euclidean distance between each sample. The resulting distance matrix was then processed for t-Distributed Stochastic Neighbor Embedding using the Rtsne package (v0.15, https://github.com/jkrijthe/Rtsne) using options "is_distance = true, perplexity = 10, theta = 0.5, dims = 2, max_iter = 1000". The x and y dimensions were combined with sample and group labels for plotting with ggplot2 in R. We also identified super-enhancer regions using H3K27ac ChIP-Seq reads at all heart enhancer segments with default parameters in ROSE.[50]

Differential Regulatory Site Activation and Motif Enrichment

To identify putative regulatory elements that are differentially utilized between Early (C13), Mid (CS16 and CS17) and Late (CS23) human embryonic heart samples, we compared H3K27ac and H3K4me2 signals at enhancer chromatin state segmentations independently using DiffBind (V2.6.6) in R (V3.4.1).[113] For a specific chromatin signal, uniquely aligned reads from two to four replicates of each time period were quantified and normalized for input signal at enhancer segments (states 13 through 18 from 25 state model) using fragment sizes determined by phantompeakqualtools (v.1.14)[108], and the DBA_SCORE_TMM_MINUS_FULL_CPM function of DiffBind. Differential signals were determined by DiffBind using DESeq2 and filtered for a false discovery rate less than 0.1. Known motif enrichment in differentially activated regions for each histone modification were determined using HOMER with the options "-size given -len 8,10,12,14 -mask -gc" (v4.9).[114] Resulting HOMER output

files were loaded into R using homerkit (https://github.com/slowkow/homerkit) and -log10 transformed p-values for each motif were compared between regions more active in each time point vs each other time point, comparisons were as follows: early up versus mid (EVM), early up versus late (EVL), mid up versus early (MVE), mid up versus late (MVL), late up versus mid (LVM), late up versus early (LVE). Z-score were calculated for each motif across the comparisons and plotted as heatmaps. Differentially enriched regions were assigned to the single nearest gene up to 1 Mb away and resulting gene lists were assessed for gene ontology enrichments using GREAT.[115] All results from GREAT were retrieved programmatically using rGREAT (V1.17.1 https://doi.org/doi:10.18129/B9.bioc.rGREAT). Z-score were calculated for each -log10 transformed p-value for each gene ontology enrichment across a single comparison and plotted as a heatmap.

Enrichment of Chromatin States at Cardiomyocyte Chromatin Loops

The non-promoter anchor points from published high-resolution promoter capture Hi-C (PCHi-C) data from iPSC-derived cardiomyocytes (CMs) were overlapped with the chromatin state segmentations from our human embryonic heart samples and chromatin states from Roadmap brain samples (E053, E054, E067, E068, E069, E070, E071, E072, E073, E074, E081, E082, E125). The fold enrichment of overlap was calculated by taking the fraction of overlap of the non-promoter anchors with the desired tissue type state annotations over the fraction of overlap of randomly selected segments to match the segments chromatin state annotations with the same non-promoter anchors. Significance of difference between human embryonic heart and adult brain fold enrichments were calculated using the Mann-Whitney test (p-value ≤ 0.05 = *, ≤ 0.01= **, ≤ 0.001= *** , ≤ 0.0001 = ****).

Enrichment of GWAS signals in enhancer chromatin state segmentations

Two linkage-disequilibrium (LD) aware approaches were used to determine enrichment of cardiovascular and cardiac development related GWAS signals in enhancer chromatin state signals. In the first approach only variants with genome-wide p-values < $5x10^{-8}$ were selected from the GWAS catalog or published literature related to atrial fibrillation, resting heart rate, QRS interval, and P-wave duration, aortic root size, congenital heart defects, and coronary artery disease (See Key Resources Table for Term Accessions). Enhancer states (ChromHMM states 13 through 18) identified in embryonic heart samples, embryonic craniofacial samples, and all tissues profiled by Roadmap Epigenome were combined into a single annotation per embryo or tissue. Variant enrichment was then determined using GREGOR for each embryo or tissue dataset. Testing was done by using $r^2 \geq 0.8$, maximum LD window of 1 Mb, and 500 minimum neighboring SNPs for each variant based on variants found in the 1000 Genome Project samples of European ancestry. In the second approach we first collected full GWAS summary statistics from cardiac trait related studies of atrial fibrillation, resting heart rate, QRS interval, and P-wave duration along with putative negative controls from immune disease related studies of systemic lupus erythematous and Crohn's disease (See STAR Methods for individual study accessions). GWAS summary statistics were prepared for processing by GARFIELD using garfield-create-input-gwas.sh script for each study. Individual chromatin state annotations for the same embryos and tissues described above were prepared for processing by GARFIELD using garfield_annotate_uk10k.sh. Finally GARFIELD was used to determine enrichment of SNPs at multiple GWAS significance thresholds (p<10-5 to p<10-8) in each chromatin state across all embryo and tissue samples using settings as previously described.[69]

ROC Curves

The ROC curves for each of the three types of data (ChromHMM, EMERGE, Dickel) used enhancers

verified by the Vista Enhancer browser. We used the full set of elements that had enhancer activity, n=281 for true heart positives and n=846 for true heart negative. For ChromHMM prediction values for enhancers in heart versus non heart, the posterior probabilities were used. Each chromatin state has a posterior probability output file from ChromHMM segmented into 200bp bins. The mean of the sum of the posterior probabilities for states 13,14,15 and 18 were calculated for the ROC curves. For the EMERGE data, we were provided a bedgraph prediction track from the authors. Peaks were called on this bedgraph using MACS. The command used was mac2 bdgpeakcall with -c = 0.05. The resulting peaks were then merged with a gap distance of 75bp. We subtracted out overlapping the blacklist from ENCODE and all peaks that intersected with a known TSS. To get the final prediction score for each EMERGE peak, the prediction scores from the bedgraph track that spanned the peaks were summed. The sum of the prediction scores from all peaks that intersected the true positive and true negative elements were used for the ROC curves. From the Dickel et al. resource, we used the *score_Prenatal* column of the provided putative enhancers with overlapping TSS rows removed. The sum of the prenatal score across the true positive and true negative elements were used for the ROC curves. The R package plotROC (v2.2.1) was used to create the ROC curves and calculate the AUC values.

RNA-Seq

RNA was extracted using miRNeasy RNA extraction kit with on-column DNAse treatment according to the manufacturer's protocol (Qiagen). RNA integrity was checked using Agilent Tapestation 2200. RNA-seq libraries were prepared from 100-200ng total RNA using the TruSeq stranded mRNA kit (Illumina). Libraries were quantified using NEBNext Library Quant Kit for Illumina and library quality checked using Agilent Tapestation 220. Libraries were pooled and diluted to 1.8pm and sequenced on the NextSeq500 Illumina platform using 75bp paired end sequencing according to manufacturer's recommendations. The accession number for the RNA-seq bigWigs, and counts matrices reported in this paper is GEO:138799.

RNA-Seq Data Processing

Quality control was performed on RNA-seq reads using FastQC (version [v.] 0.11.5) and MultiQC (v.1.1). Trimming for adapters, quality and length was performed using Trimmomatic (v.0.36). Trimmed fastqs were aligned with Rail-RNA[76] using human assembly GRCh38/hg38. The coverage bigWig files output by Rail-RNA were used as input for the generation of counts tables by following the instructions and pipeline from recount (https://github.com/leekgroup/recount-contributions), where the comprehensive gencode v.25 annotation was used.[75] The "level 3" genes defined by gencode were excluded. The recount rse_gene objects for each sample were combined into one rse_gene object and transformed with scale_counts() from recount (v1.8.2).

RNA-Seq Differential Expression

The scaled rse_gene recount object was made into a DESeq2 (v1.22.2) object. Low gene counts were filtered by removing all genes whose sum of counts across all samples was less than 100. This left a total of 26,122 genes for downstream analysis. Batch effects were mitigated by using the sva package in R. (v3.30.1) Pairwise, differential analysis of each of the eight carnegie stages was performed with DESeq2 including the surrogate variables identified from sva. For the heatmaps and PCA plots the counts table was transformed using the surrogate variables, see provided code (https://github.com/cotneylab) for these calculations. The PCA plots in the supplement were made using the prcomp() function from the built-in R stats (v3.6.3) package. The genes shown in the heatmaps and used for the PCA plot generation had a Benjamini Hochberg adjusted p-value less than 0.05 and a fold change greater than the absolute value of 1. heatmap.2() from the R library package

gplots(v3.0.1.1) was used to generate the heatmaps of the normalized SV corrected counts by setting the scale option to “row”. The distance matrix of one minus the transpose of the kendall correlation of the SV corrected counts was clustered using hclust() with method = “complete”. The resulting hierarchical clustering was used in the heatmap.2() option Rowv to organize the gene rows. clusterProfiler (v3.12.0) was used to obtain the gene ontology enrichments from the dendrogram groups using the function enrichGO() with standard options. The enrichment map plots were made using enrichplot (v1.6.0) function emaaplot(). Standard options were used with manual code changes to the color and p-value settings

GTEx tSNE Analysis

The rse_gene R data object, or counts table of all GTEx tissues was retrieved from the Recount2 database (https://jhubiostatistics.shinyapps.io/recount/). The GTEx counts table was generated using the same Rail-RNA & Recount pipeline we used to generate the counts tables for our embryonic heart data which is described above. The GTEx data contained 9,662 samples which we combined with the 24 embryonic heart samples to make one matrix containing 9,686 total samples by 58,037 genes. The meta-data for GTEx is provided in a link under the *phenotype* column from the Recount2 database, and the tissue assignments located under the column named *smts* was used, resulting in a total of 31 unique tissues. The counts were transformed using the *scale_counts()* function from the R library recount (v1.8.2) as is recommended from the workflow in the Recount2 F1000 paper.[116] Genes whose mean across all GTEx and embryonic heart tissues was lower than or equal to 1 were removed, resulting in 36,990 genes. The filtered counts matrix was transformed by *log10()* with a pseudo count of 1 added to all values. The transpose of the log10 transformed matrix was then converted to a distance matrix using the *dist()* function in R. This distance matrix was used as input for the tsne model generated by using the *Rtsne()* function from the R package Rtsne (v0.15). The parameters used in Rtsne() were the following: perplexity=10, max_iter=1000, theta=0.5, dims=2.

Tissue Specificity of Gene Expression (GINI)

The combined, filtered GTEx and human embryonic heart counts matrix used for the tsne analysis was also used to calculate the tissue specificity for each gene. The Gini Index for each gene was calculated using the Gini() function from R library Ineq (v0.2-13) on the average counts per tissue. A gene was given a tissue assignment based on the tissue with the maximum count for that gene. To create the heatmap of the various tissues, the distance matrix of one minus the transpose of the pearson correlation of the average expression per tissue for all genes with a Gini score of 0.5 and above was clustered using hclust() with method = “complete”. The z-scores of this average expression per tissue matrix was plotted using heatmap.2 with the gene rows organized by the dendrogram calculated from hclust(). The genes for the embryonic heart, spleen and adult heart were determined by cutree() such that it would result in 25 groups to correspond to the 25 tissues. RDAVIDWebService (v1.22.0) was used to obtain the gene ontology enrichments using the original 36,990 genes as the background.

Novel Heart Enhancer Effects on Gene Expression

Assignment of enhancers to genes was made using GREAT v.4.0.4 using human embryonic heart specific enhancers on hg19 with the whole genome as background and using the single nearest gene association rule setting. The line graph of the enhancers versus expression used the average expression per tissue type from GTEx and our heart enhancer data using geom_smooth() from ggplot2 with the method set to “loess”. The corresponding violin plot included only average gene expression from genes with more than 25 enhancers targeting. Stat_compare_means() was used to compare all tissues against the embryonic heart, using the Mann-Whitney test (p-value ≤ 0.05 = *, ≤ 0.01= **, ≤ 0.001= *** , $\leq$ 0.0001 = ****).

Weighted Gene Co-Expression Network Analysis

We generated co-expression networks using the WGCNA Rpackage based on recommendations put forth by the Horvath group.

Network Construction

For the WGCNA a soft-thresholding power of 8 was chosen assuming an unsigned network and based on recommendations put forth by the Horvath group for samples between 20 and 30. The modules were detected from the network from the cutreeDynamic() function from the WGCNA package with the following parameters, minClusterSize=100, deepSplit=2. Detected modules were merged based on their eigengene correlation. To do this a dendrogram of the module eigengenes was generated and a threshold value of 0.18 was chosen as input for the function mergeCloseModules(). The intra-modular connectivity of each gene was calculated using the intramodularConnectivity() function from the WGCNA library which determined the hub and non-hub designation. The resulting network contained 29 modules.

Plotting of Modules

A multidimensional scaling of the module eigenvectors output from WGCNA was generated to plot the modules in 2-D space using the function cmdscale() from the stats v3.6.1 package. A pearson correlation of the module eigenvectors was calculated for the edges. Positive correlations of 0.5 and greater were included. Modules were plotted that fulfilled the criteria of having significant adjusted p-values (< 0.05) from the GO analysis, significant permutation p-values (< 0.05) of embryonic heart specific enhancers, and/or of embryonic heart specific Gini genes, this resulted in exclusion of 9 modules from the plot. The module eigenvectors of each of the 20 modules were plotted using ggplot2, geom_smooth() function with the loess smoothing method. The confidence intervals were removed for ease of visualization.

Gene Ontology and Functional Enrichments

RDAVIDWebService (v1.22.0) was used to obtain gene ontology enrichment of the genes within each of the 29 WGCNA modules. The gene background list used was all the genes input into the WGCNA. The module enrichments of embryonic heart specific Gini genes, embryonic heart specific enhancers, various disease gene lists, and the NKX2-5 bound gene lists were determined by a permutation test with 1000 iterations. The single-cell RNA-seq differential expression for genes with an adjusted-p value <= 0.05 were used. For the LOEUF enrichment analysis among the hub vs non-hub genes, the non-hub genes were randomly sampled using the R function sample(). The number of non-hub genes sampled were the same number of total hub genes within the network- 10% of 26,122 genes or 2,612 from the 23,510 non-hub gene list. This process was iterated 1000 times to get a mean (gray bars) and standard deviation (shown as error bars) for each LOEUF decile. The LOEUF score and decile designation for each gene is freely available through gnomAD v2.1.1

Protein-protein interaction analysis

To generate the ppi histogram, 100 randomized versions of the WGCNA network were made. This was done by randomly assigning the 26,122 genes to 29 modules of equal gene sizes to the original network using the R function sample(). The ppi enrichment of up to 500 randomly chosen genes for each module of each of the 100 randomized versions was then determined using the STRINGdb (v1.24.0) package. Up to 500 genes were used due to constraints from the STRINGdb package. The ppi database was loaded by using STRINGdb$new with version =”10” and score_threshold=0.4. For each iteration, the output p-value of the STRINGdb call get_ppi_enrichment() was adjusted using the Bonferroni method. The number of modules that met the adjusted p-value cut off of 0.05 was counted for each iteration to produce frequency values.

Data display

Throughout this work we make heavy use of the ggplot2 package (v3.2.0) in R(v3.6.1) package.

Online Tables

Online Table I

Sample ID	Stage	Date of processing	Karyotype	Collection Site	Collection Date
14131	CS12	1/11/2019	rsa(13,15,16,18,21,22,X)x2	Newcastle	4/27/2018
12383	CS13	7/26/2018	46,XY	London	1/13/2015
12690	CS13	7/26/2018	46,XY	London	6/23/2015
14401	CS13	1/11/2019	rsa(13,15,16,18,21,22)x2,(X,Y)x1	London	10/16/2018
14479	CS13	1/11/2019	rsa(13,15,16,18,21,22)x2,(X,Y)x1	Newcastle	11/22/2018
12408	CS14	1/11/2019	46, XY	London	1/20/2015
14135	CS14	1/11/2019	rsa(13,15,16,18,21,22,X)x2	Newcastle	5/3/2018
12997	CS16	1/11/2019	46,XY	London	1/14/2016
14213	CS16	1/11/2019	rsa(13,15,16,18,21,22,X)x2	Newcastle	6/21/2018
14315	CS16	1/11/2019	rsa(13,15,16,18,21,22,X)x2	Newcastle	8/16/2018
14209	CS16	1/11/2019	rsa(13,15,16,18,21,22)x2,(X,Y)x1	Newcastle	6/19/2018
12291	CS17	6/21/2018	46,XY	Newcastle	1/1/2017
12331	CS17	6/21/2018	46,XX	Newcastle	1/1/2017
12752	CS17	6/21/2018	46,XX	Newcastle	1/1/2017
12059	CS18	7/23/2018	46,XY	London	5/21/2014
12456	CS18	7/23/2018	46,XY	London	12/6/2016
13474	CS18	7/23/2018	46,XX	London	2/11/2015
11914	CS19	6/18/2018	46,XX	London	1/8/2014
12135	CS19	6/18/2018	46,XY	London	7/23/2014
12248	CS19	6/20/2018	46,XY	London	10/8/2014
12448	CS20	7/24/2018	46,XY	London	2/11/2015
12451	CS20	7/24/2018	46,XX	London	2/11/2015
13068	CS20	7/24/2018	46,XX	London	2/23/2016
11849	CS21	7/25/2018	46,XY	London	10/2/2013
12093	CS21	7/25/2018	46,XX	London	6/18/2014
12210	CS21	7/25/2018	46,XX	London	9/10/2014
12058	CS23	7/27/2018	46,XY	London	5/21/2014
12151	CS23	7/27/2018	46,XX	London	8/6/2014
12193	CS23	7/27/2018	46,XY	London	9/3/2014

Online Table I. Embryo ID and Karyotype. The embryo ID and karyotype of every embryo used in the experiments in this paper as reported from HDBR. Related to Figure 1.

Online Table II. Bed files of enhancer coordinates. Tab 1 - Read me document. Tab 2- Reproducible Enhancers in human embryonic heart (HEH). All DNA segments called as any enhancer state in two or more samples. Tab 3- Strong HEH Enhancers not in Roadmap. All enhancer segments that do not overlap with any enhancer segment called in any tissue in Roadmap Epigenome known as EHEs. Tab 4- All super enhancers called by the ROSE algorithm from the human embryonic heart samples. Tab 5- Super enhancers that are called only in human embryonic heart samples and not in dbSuper database, annotated by GREAT version 4.0.4 on hg19 using association rule: Single nearest gene: 1000000 bp max extension, curated regulatory domains included. Tab 6 - Functional enrichments of gene residing in super enhancers as determined by DAVID. Related to Figure 2.

Online Table III. GREAT Results. Tab 1 - GREAT version 4.0.4 using single nearest gene association rule output for all enhancers not found in roadmap. Tab 2 - Assignment of human heart enhancers to potential regulatory gene targets using the same rule as in Tab 1. Related to Figure 2.

Online Table IV. Full Homer Results for TF Motifs In Human Embryonic Heart Enhancers. Related to Figure 2.

Online Table V. List of Gini Score and Tissue Specificity per Gene. Related to Figure 6.

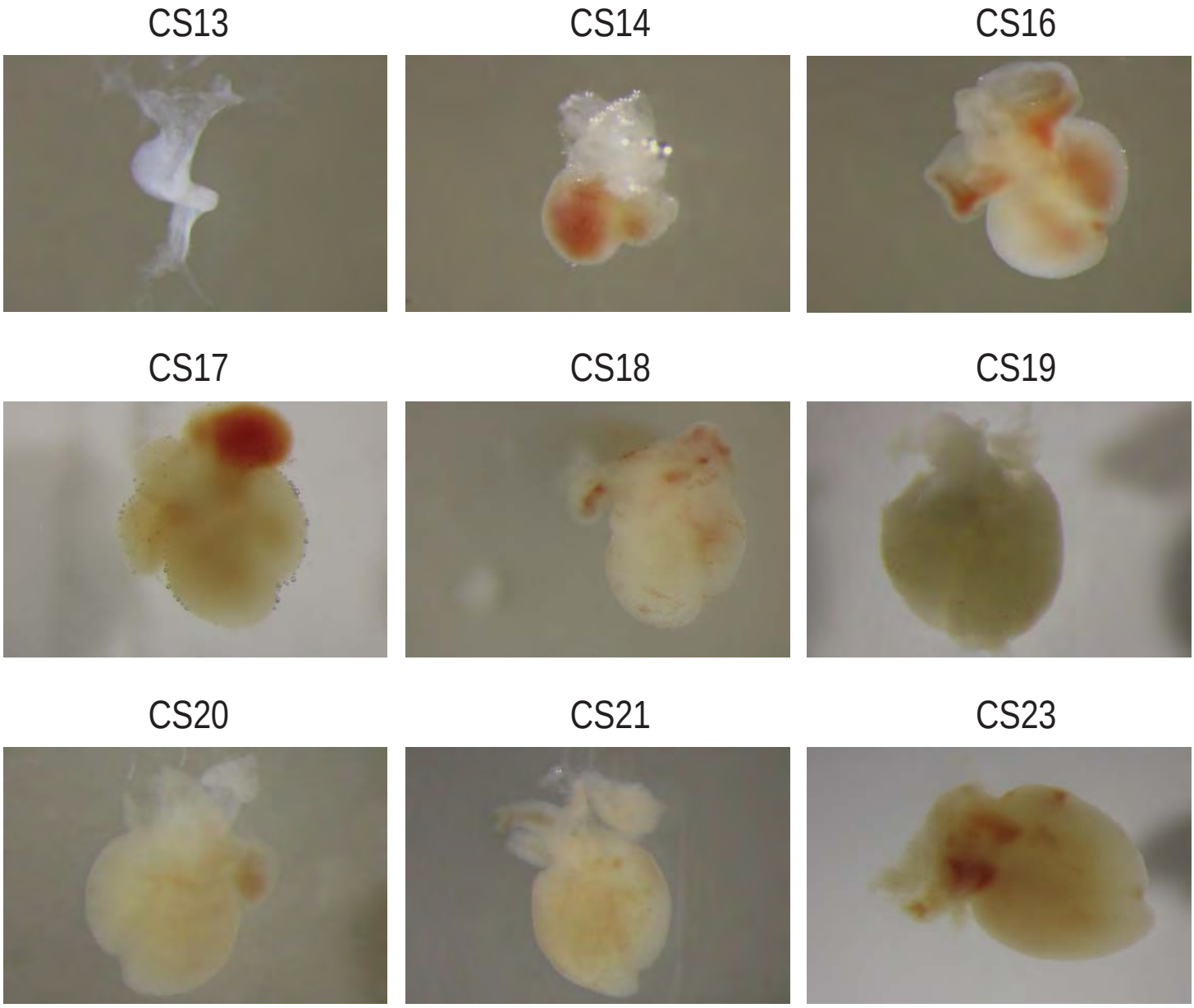
Online Table VI. Gini Heart oPOSSUM and IPA Results. Related to Figure 6.

Online Table VII. DAVID GO Differential Expression. Related to Figure 6.

Online Table VIII. GO of WGCNA Modules. Related to Figure 7.

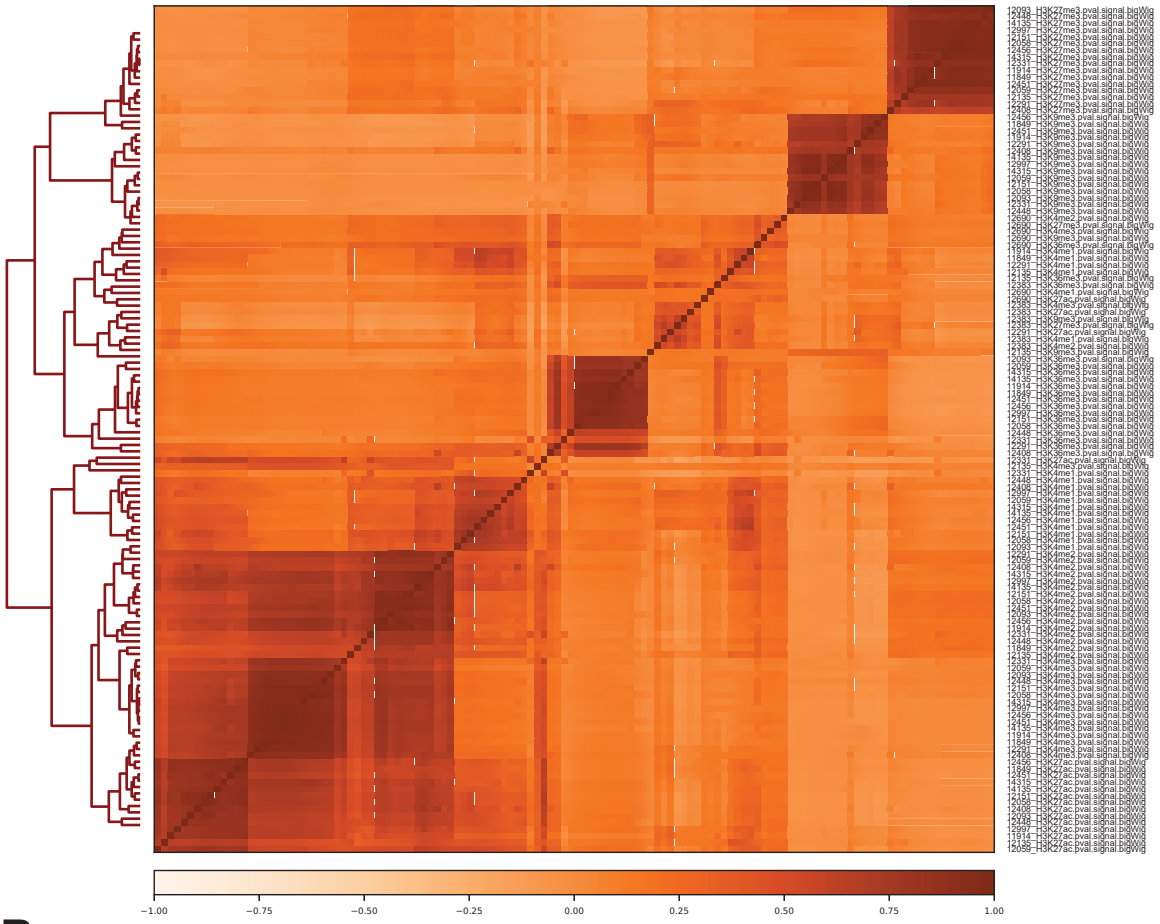
Online Table IX. Novel CHD genes. Related to Figure 6, 7 and 8.

Online Table X. WGCNA Results. Related to Figures 7 and 8.

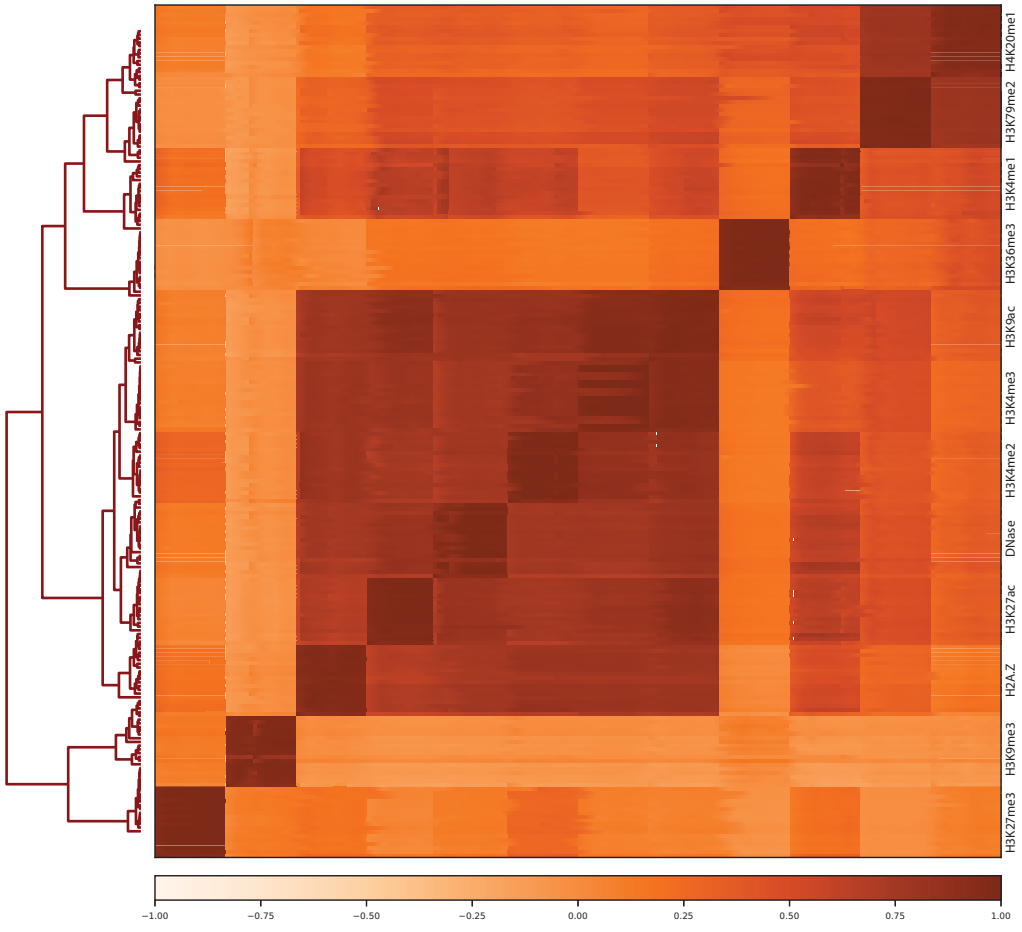


Online Figure I. Example Images of Embryos at Each Carnegie Stage. Images of one embryo for Carnegie Stages 13 through 23 that were used for experiments in this paper. Related to Figure 1.

A. Primary Signals

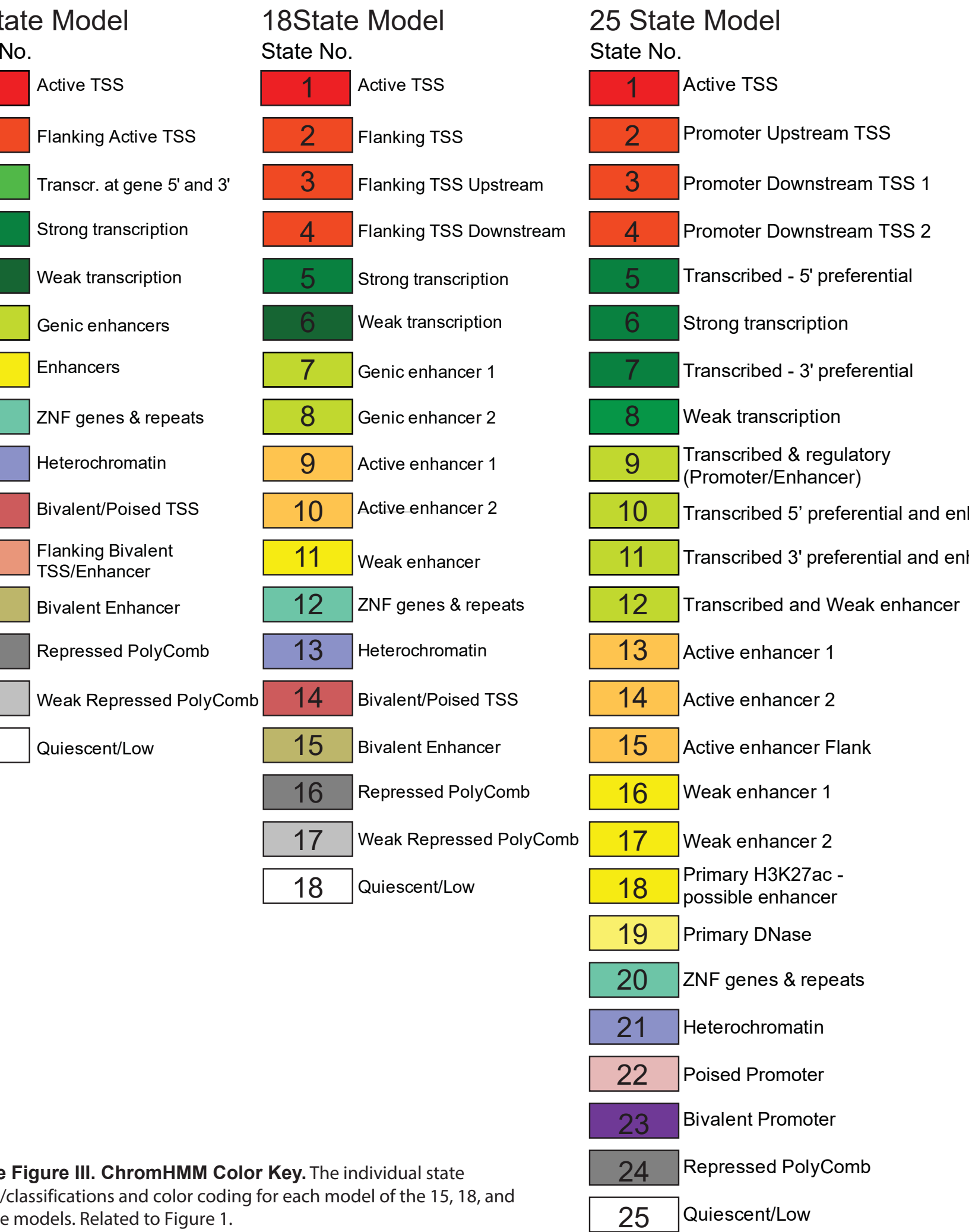


B. Imputed Signals

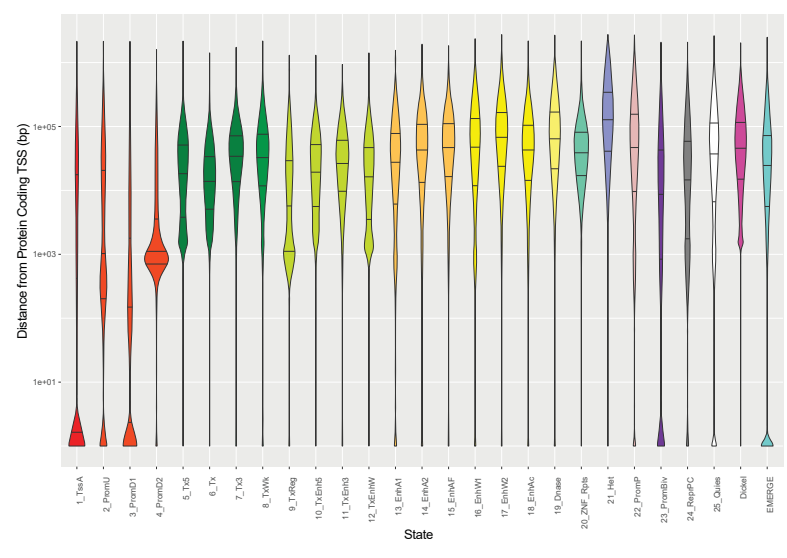


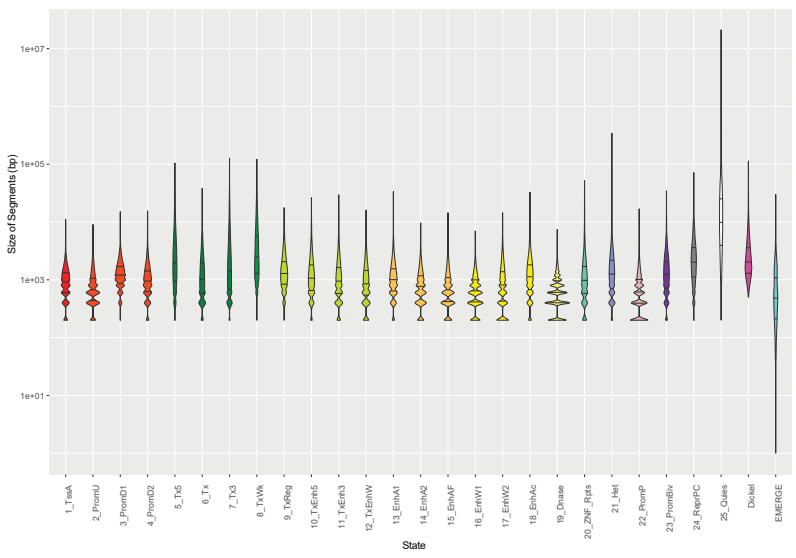
Online Figure II. Pearson Correlation Plots of Primary and Imputed ChIP-seq Signals. A. Pearson correlation of primary ChIP-Seq signals by 10kb bins across the genome showing general correlation by mark. B. Pear-

Figure III

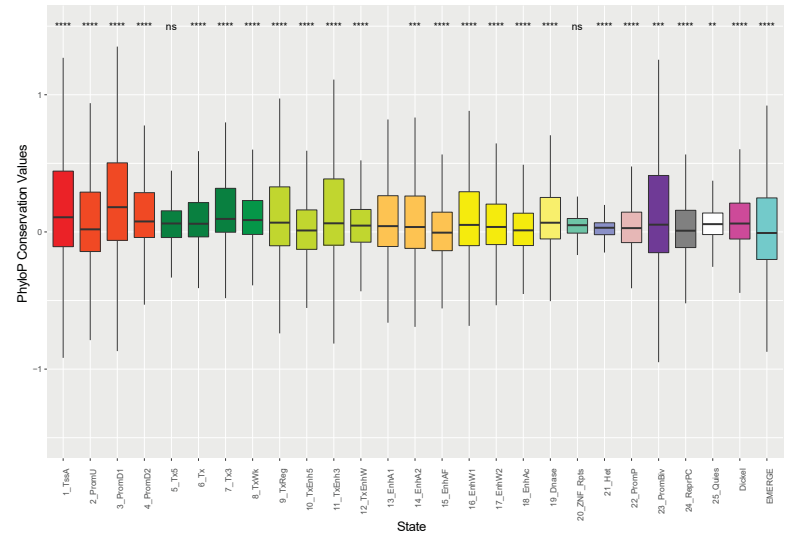


e Figure III. ChromHMM Color Key. The individual state /classifications and color coding for each model of the 15, 18, and e models. Related to Figure 1.

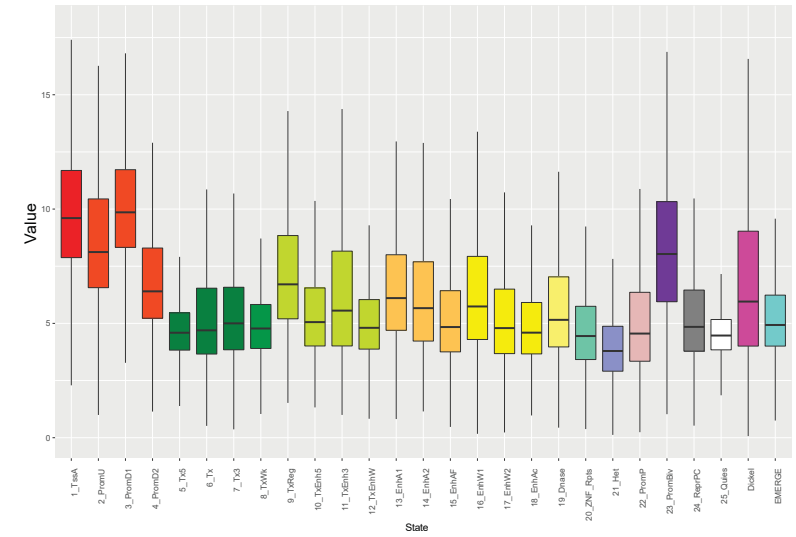




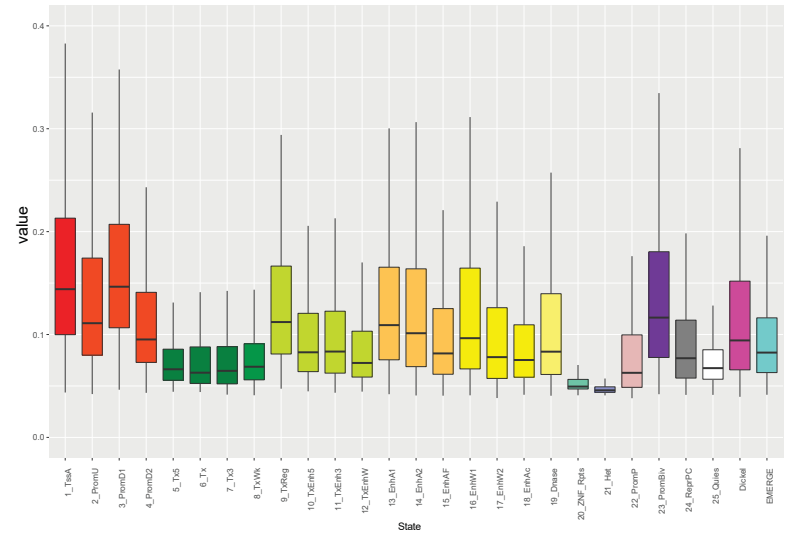
C. Conservation per state



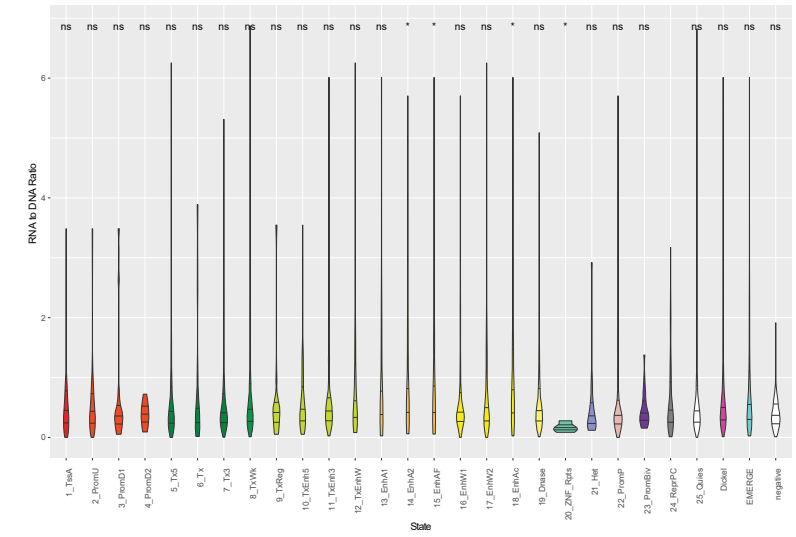
D. CADD



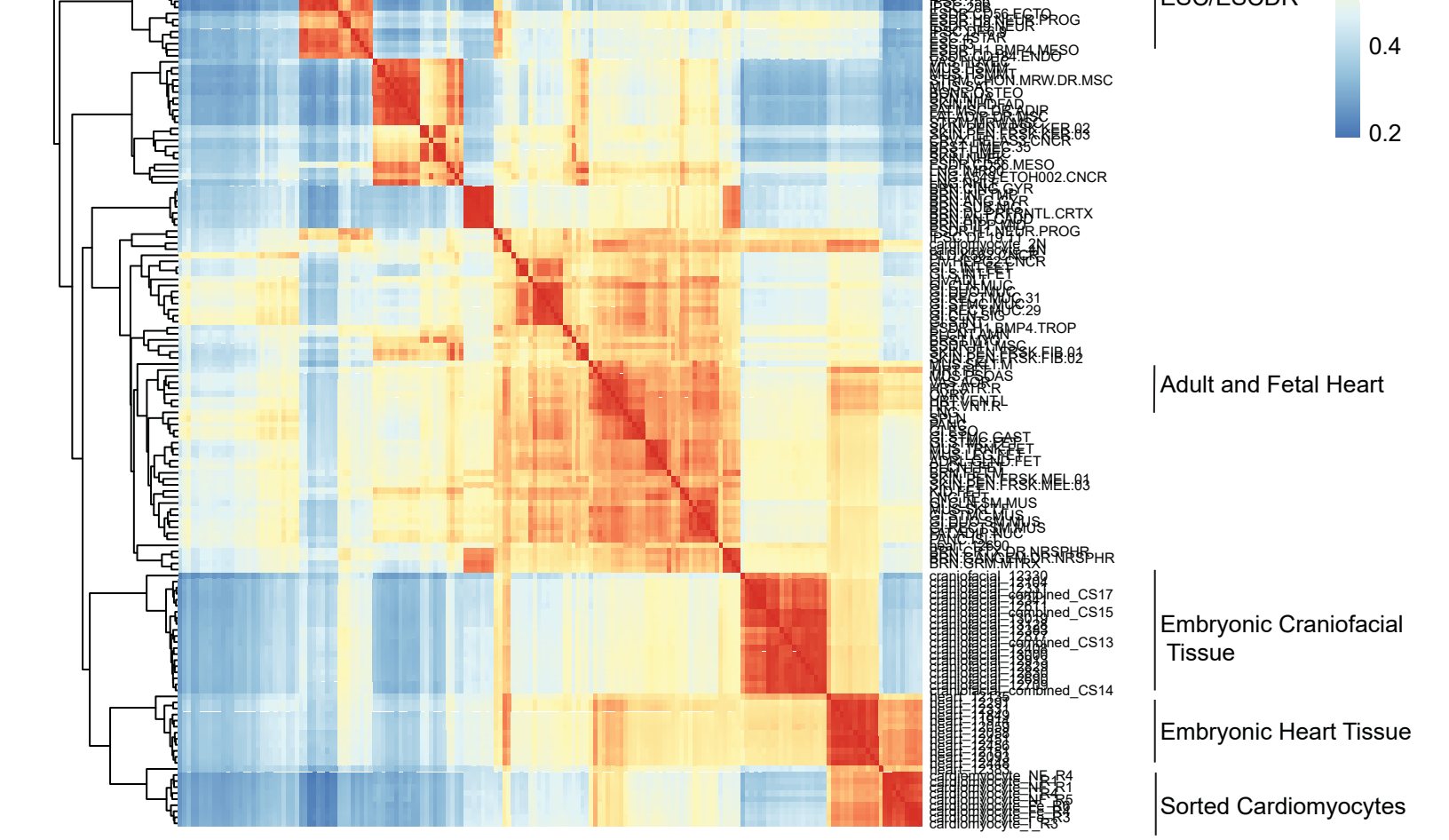
E. LINSIGHT



F. Activity of overlap with MPRA for *in vivo* mouse cardiomyocytes

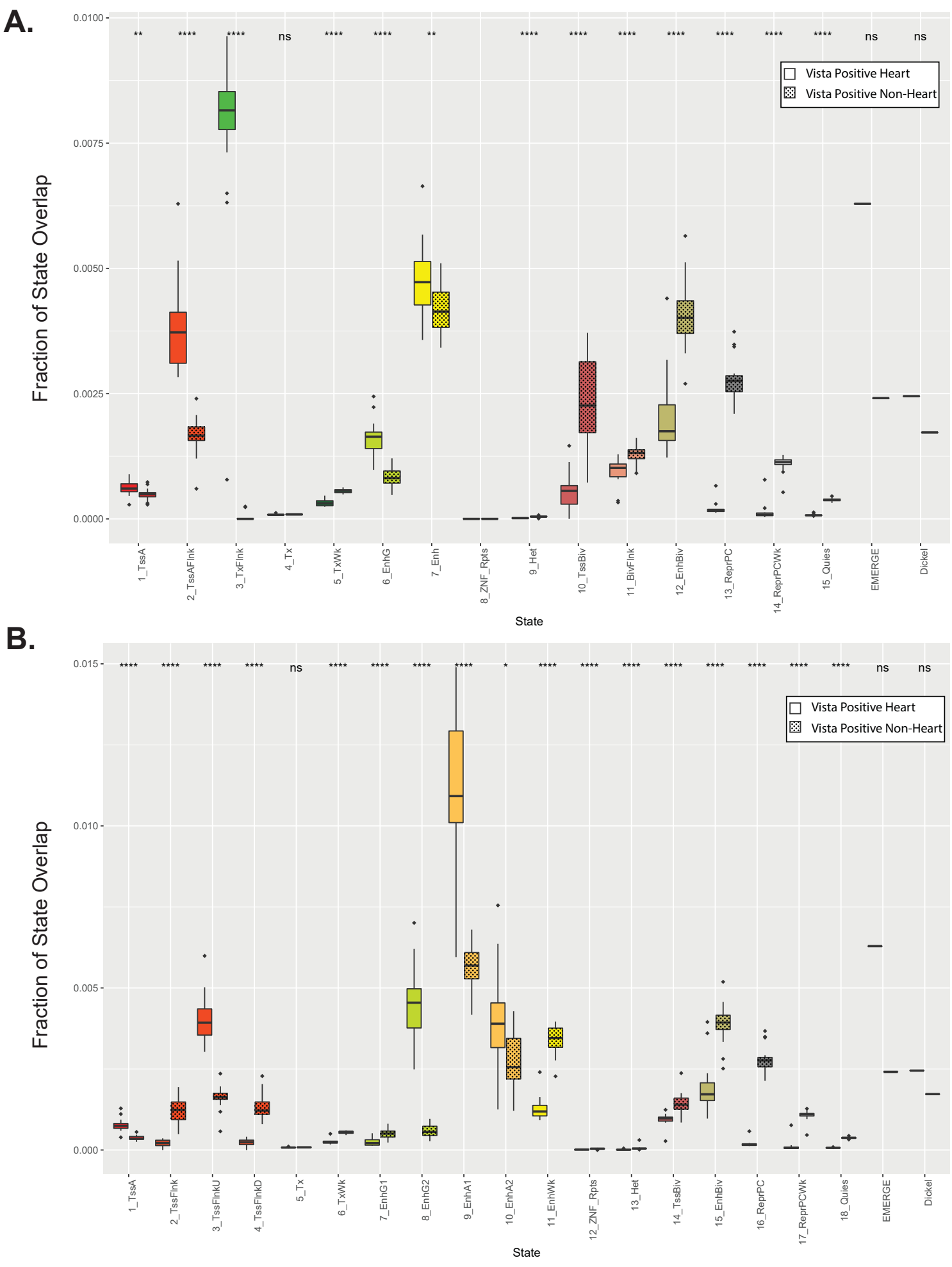


Online Figure IV. Exploration of the ChromHMM 25 state model. A..Distance of segments from protein coding TSS in base pairs per state with Dickel and EMERGE. B. Length of segments per state with Dickel and EMERGE in base pairs. C. Conservation based on phyloP (phylogenetic p-values) scores per state with Dickel and EMERGE from the PHAST package for multiple alignments of 99 vertebrate genomes to the human genome (100 way). Significance of difference of scores between strong enhancer state 13 and other states along with Dickel and EMERGE were calculated using the Mann-Whitney test and is shown at top (p-value ≤ 0.05 = *, ≤ 0.01= **, ≤ 0.001= *** , ≤ 0.0001 = ****). D CADD scores per state with Dickel and EMERGE. E. LINSIGHT scores per state with Dickel and EMERGE F. Activity of overlap with MPRA for in vivo mouse cardiomyocytes. Significance of difference of activity score between repressed state 24

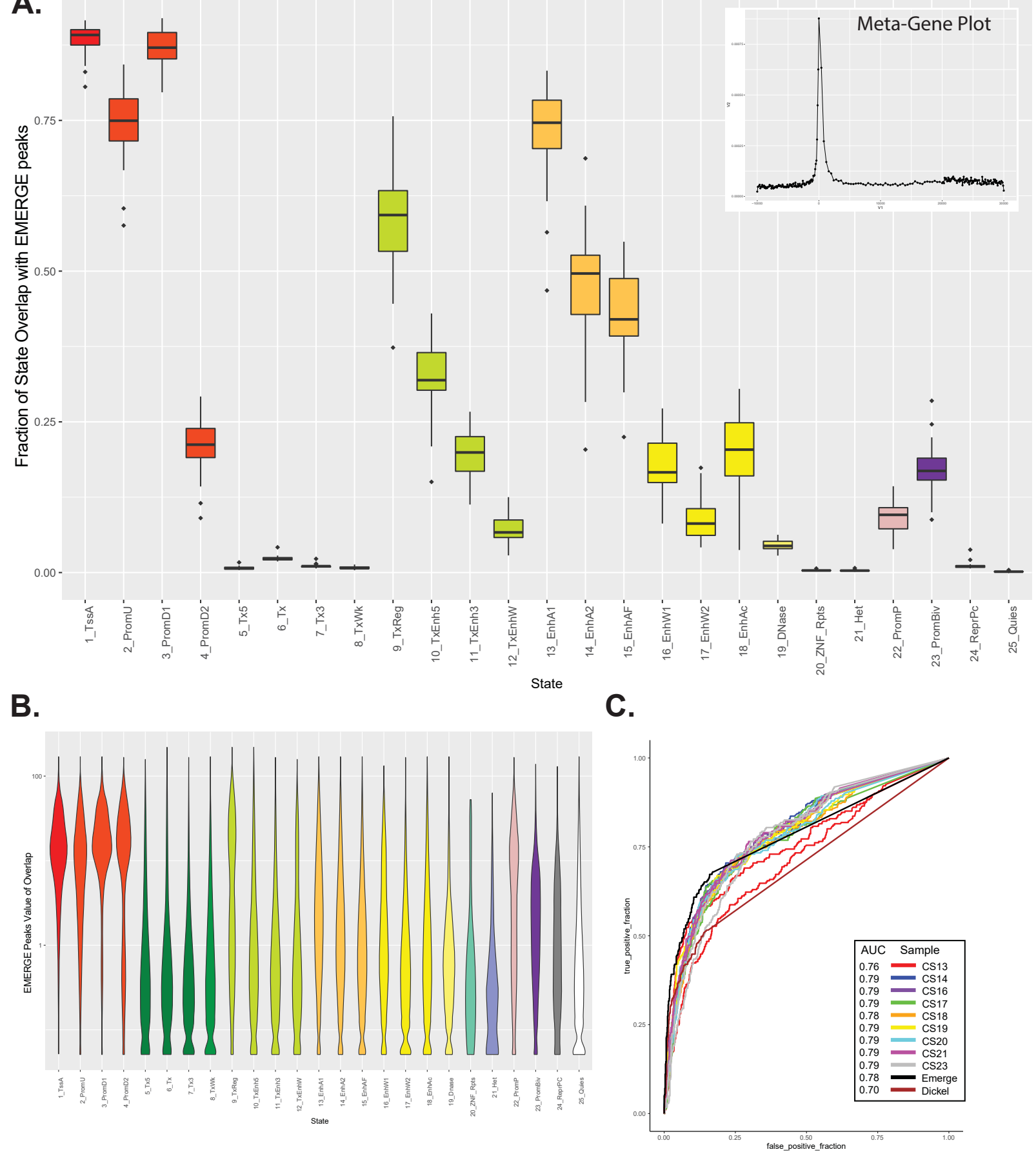

ESC/ESDR
0.4
0.2
Adult and Fetal Heart
Embryonic Craniofacial Tissue
Embryonic Heart Tissue
Sorted Cardiomyocytes

B.


group
Adipose
Blood_T_cell
Brain
Cardiomyocyte
Craniofacial
Digestive
ENCODE2012
Epithelial
ES_deriv
ESC
Heart
HSC_B_cell
IMR90
iPSC
Mesench
Muscle
Myosat
Neurosph
Other
Smooth_Muscle
Thymus
10
0
−10

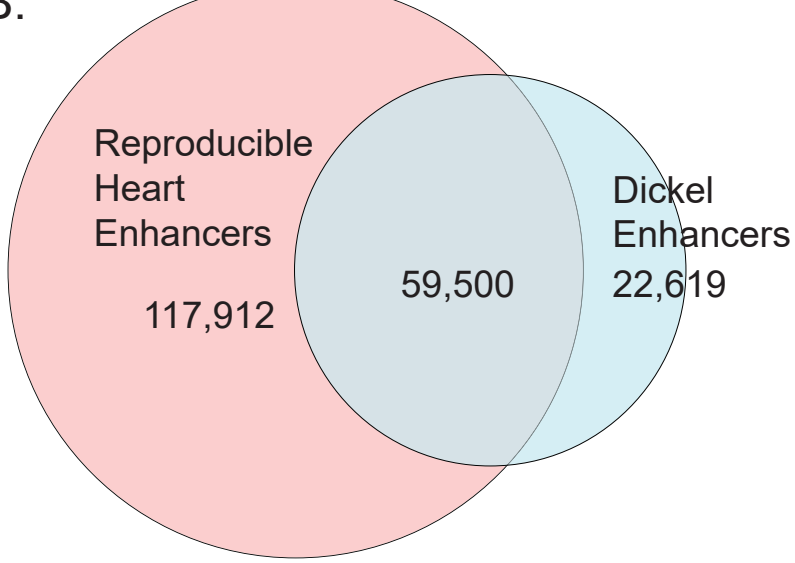


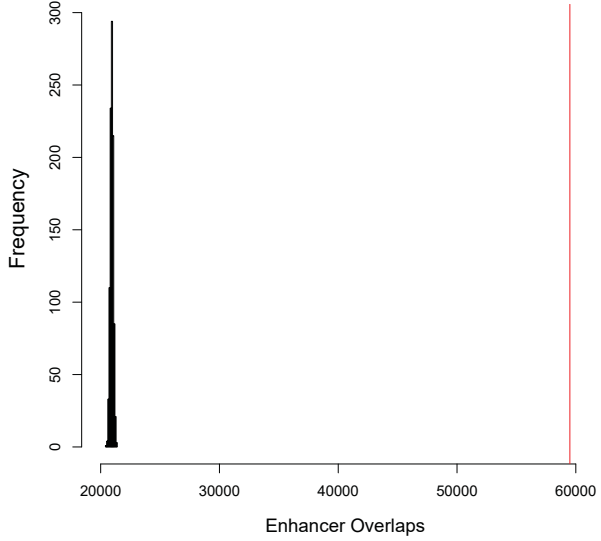
Online Figure VI. Overlap of Human Embryonic Heart States with VISTA Positive Heart and Non-Heart Enhancers for ChromHMM 15 and 18 State Models. A. 15 state ChromHMM model analysis of the embryonic heart data. Boxplot where solid bar indicates fraction of overlap of VISTA heart positive enhancers with each



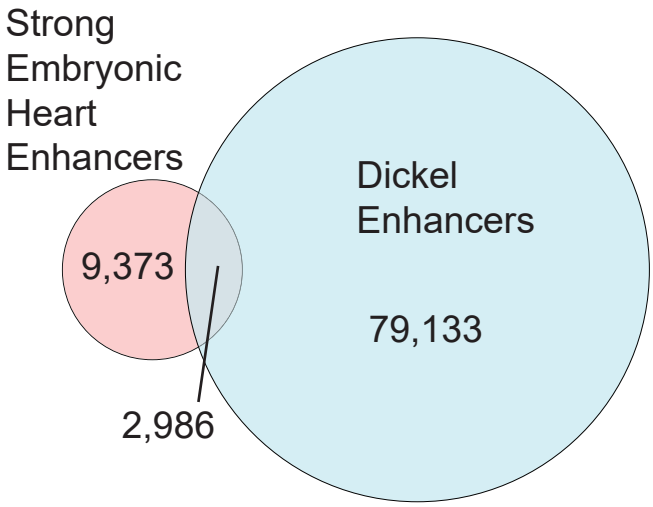
Online Figure VII. Examination of the EMERGE dataset and ROC Curves. A. Boxplot representing the fraction of each state in our embryonic heart samples that overlaps with peaks called from EMERGE bedgraph showing the greatest overlap between EMERGE and our TSS and Promoter States (specifically States 1-3), followed by strong enhancer states (States 13-15), and transcribed and regulatory states (States 9-10). In the upper right corner is an inset of a meta-gene plot of the distribution of EMERGE peaks, which also shows that a majority of peaks can be found very close to the TSS. B. Violin plot of the distribution of the scores calculated over EMERGE peaks from bedgraph signal, with again the highest concentration of higher scores being seen in the TSS and promoter states (States 1-4). C. ROC curves calculated for each of our embryonic heart samples, as well as the Dickel and EMERGE datasets, showing similar

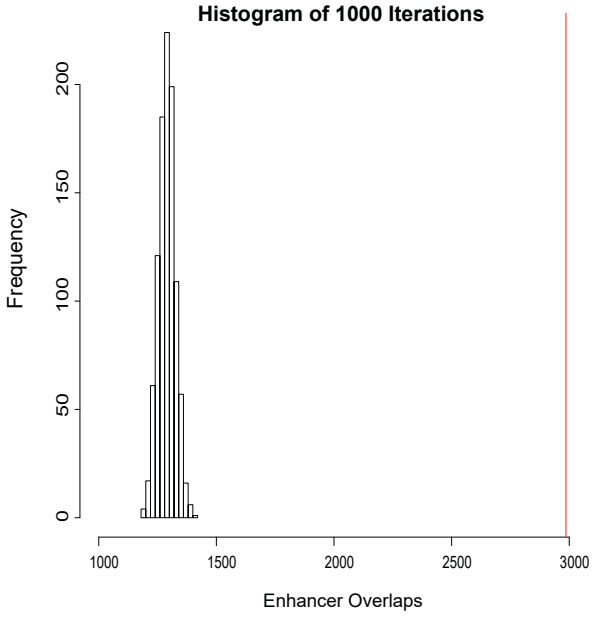
B.


Reproducible
Heart
Enhancers
117,912
59,500
Dickel
Enhancers
22,619

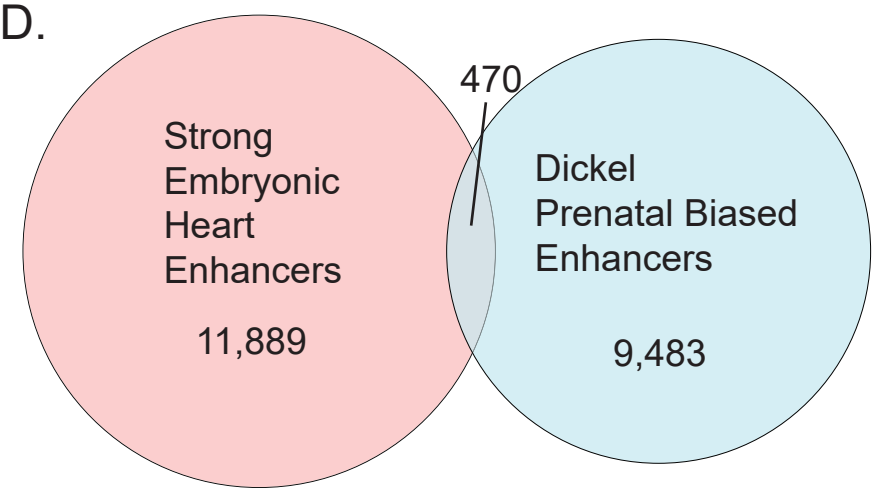

Frequency
0
50
100
150
200
250
300
20000
30000
40000
50000
60000
Enhancer Overlaps

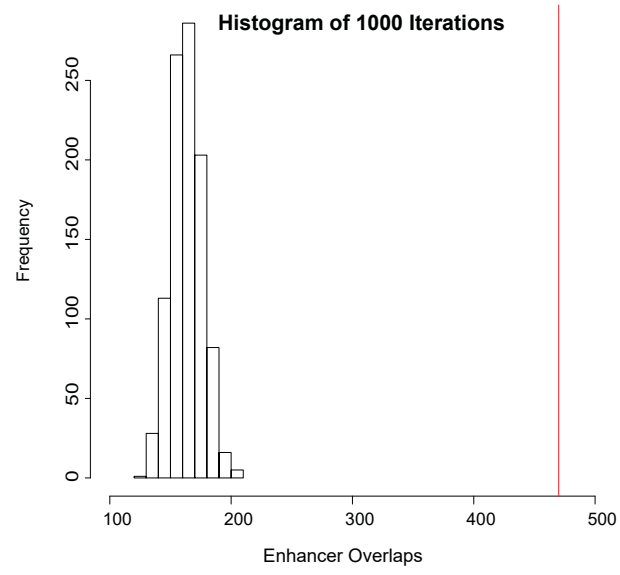
C.


Strong
Embryonic
Heart
Enhancers
9,373
2,986
Dickel
Enhancers
79,133

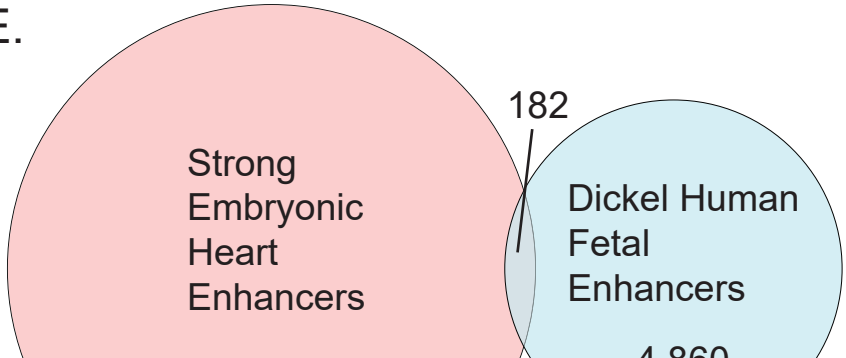

Histogram of 1000 Iterations
Frequency
0
50
100
150
200
1000
1500
2000
2500
3000
Enhancer Overlaps

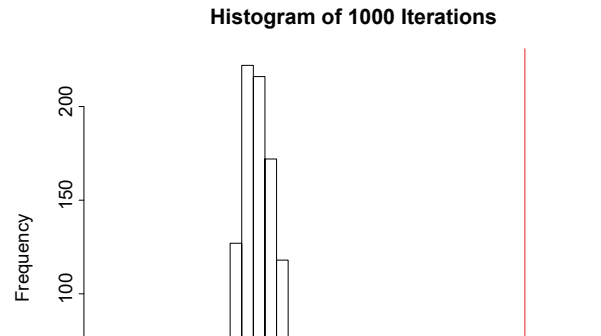
D.

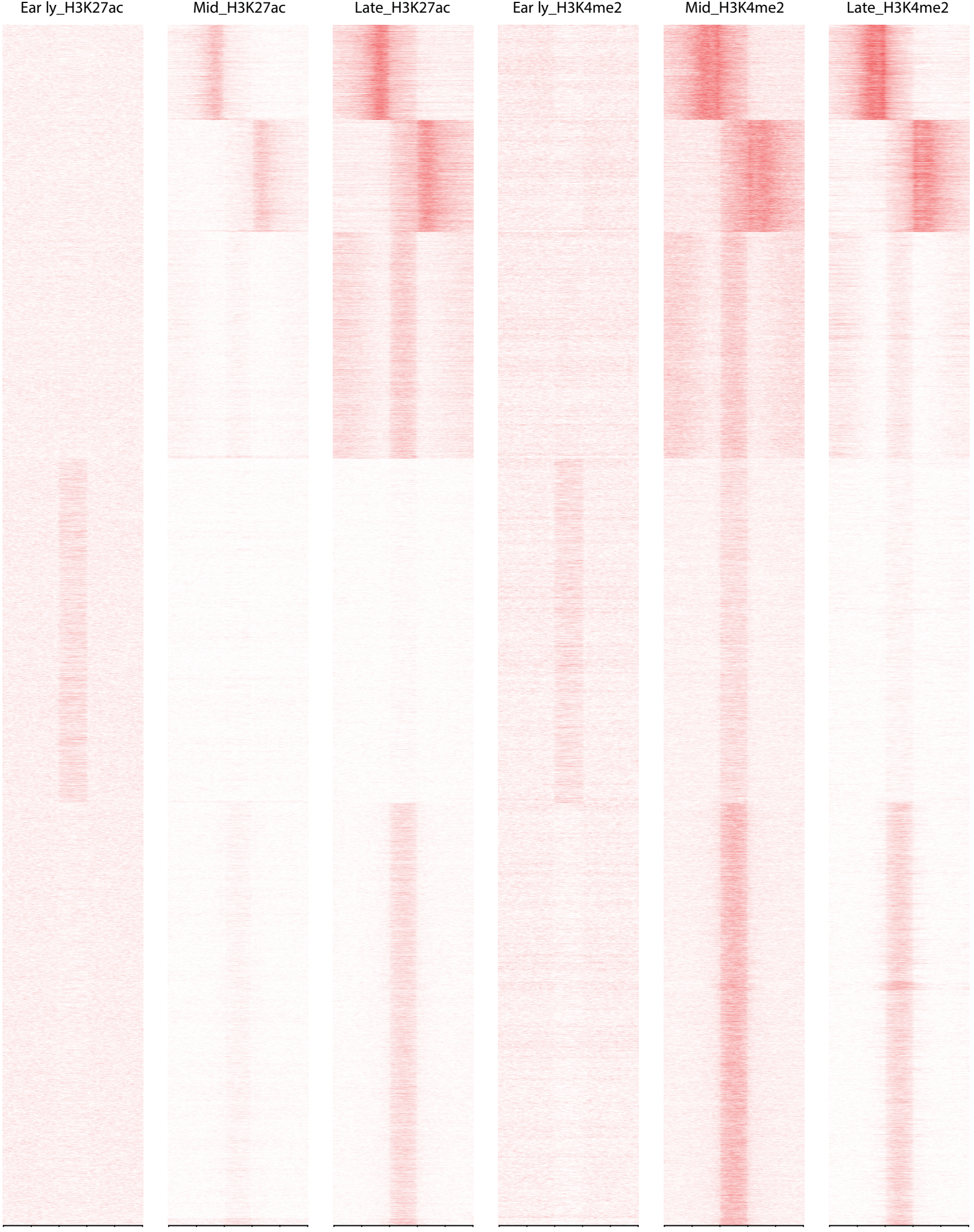

470
Strong
Embryonic
Heart
Enhancers
11,889
Dickel
Prenatal Biased
Enhancers
9,483


Histogram of 1000 Iterations
Frequency
0
50
100
150
200
250
100
200
300
400
500
Enhancer Overlaps

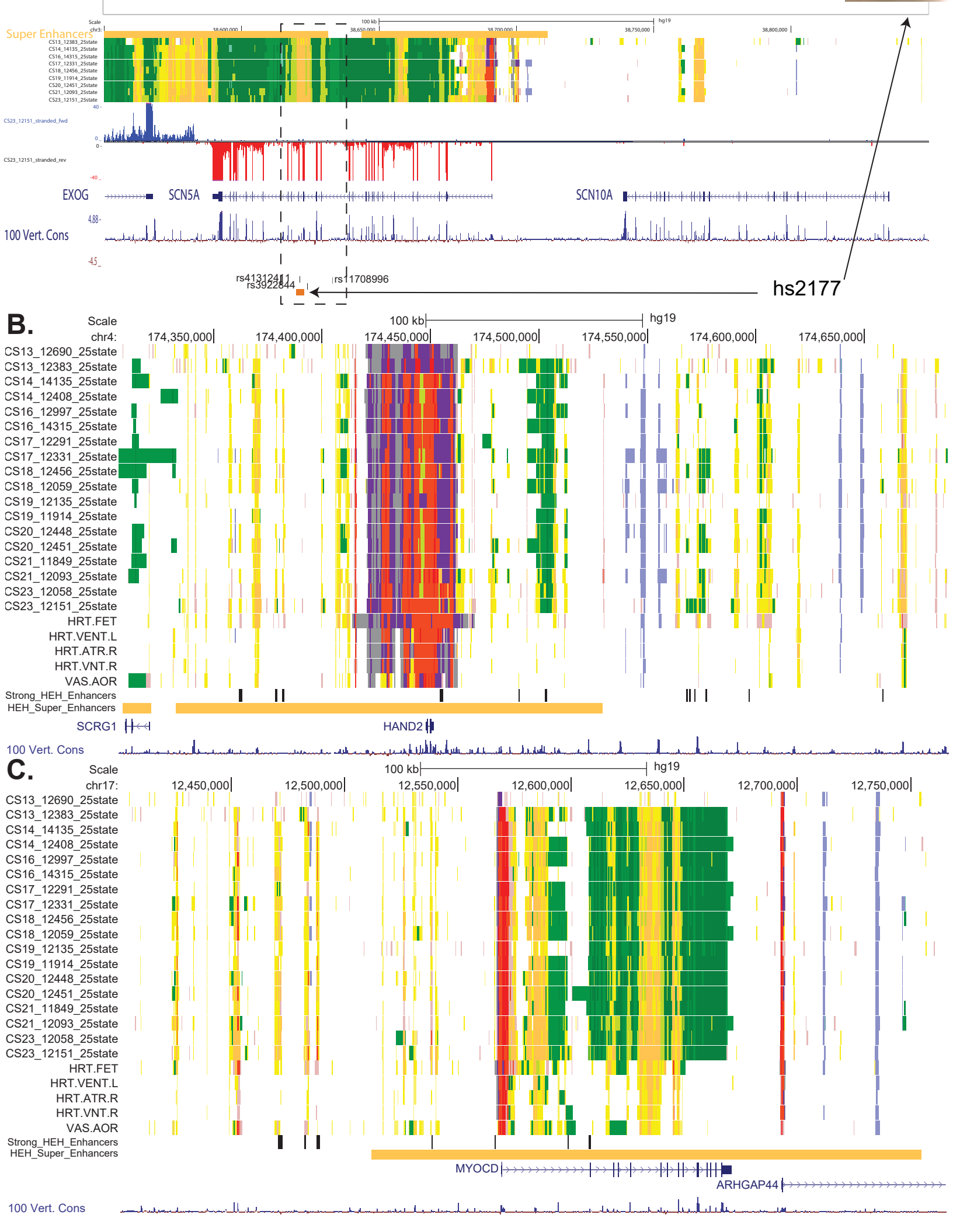
E.

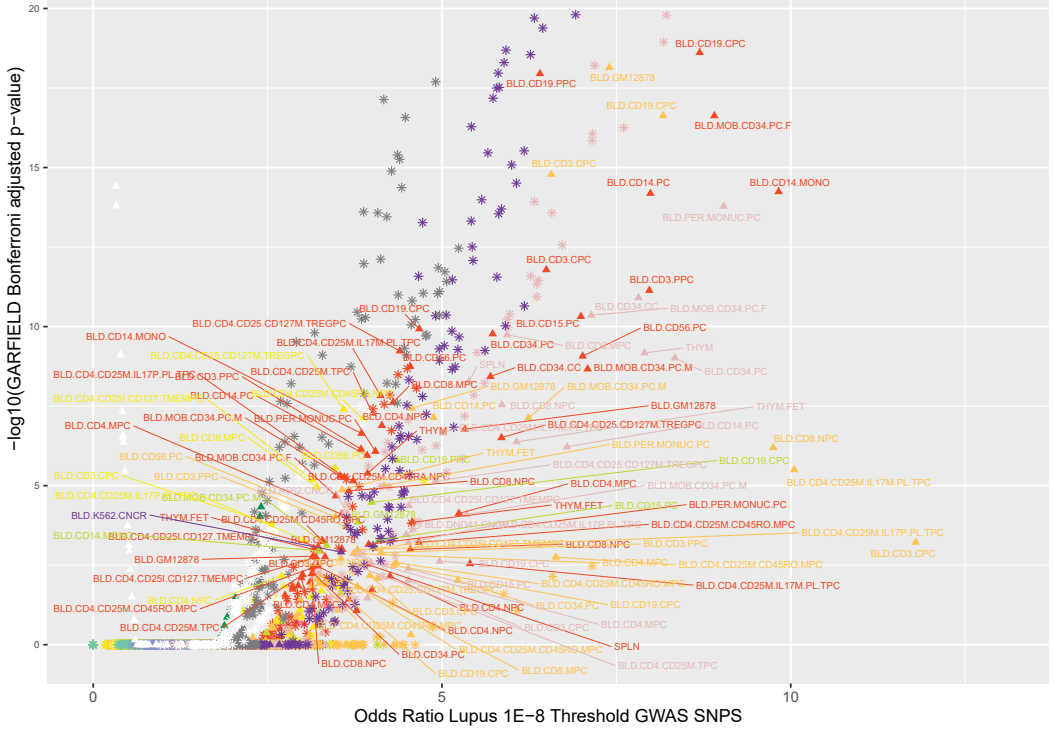

182
Strong
Embryonic
Heart
Enhancers
Dickel Human
Fetal
Enhancers


Histogram of 1000 Iterations
Frequency
100
150
200

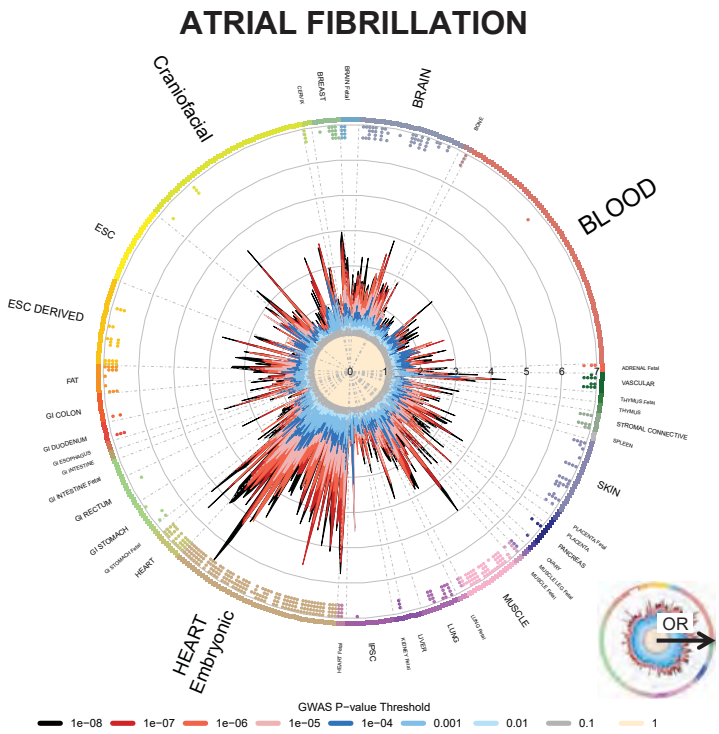


Online Figure IX. Heatmap of the Signal of Putative Enhancers Differentially Marked By H3K27ac and

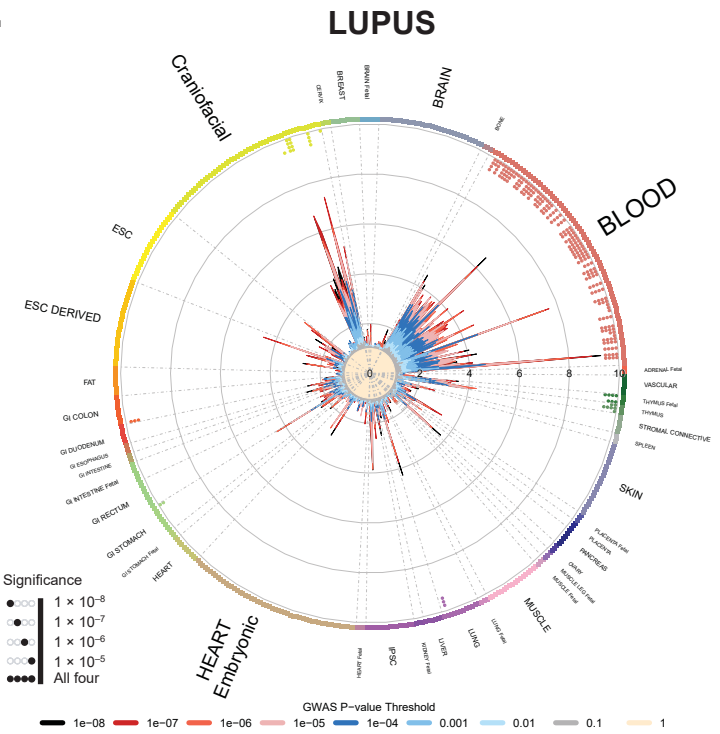

Super Enhancers
EXOG
SCN5A
SCN10A
100 Vert. Cons
rs41312411
rs3922844
rs11708996
hs2177
B.
chr4:
174,350,000
174,400,000
174,450,000
174,500,000
174,550,000
174,600,000
174,650,000
100 kb
hg19
CS13_12690_25state
CS13_12383_25state
CS14_14135_25state
CS14_12408_25state
CS16_12997_25state
CS16_14315_25state
CS17_12291_25state
CS17_12331_25state
CS18_12456_25state
CS18_12059_25state
CS19_12135_25state
CS19_11914_25state
CS20_12448_25state
CS20_12451_25state
CS21_11849_25state
CS21_12093_25state
CS23_12058_25state
CS23_12151_25state
HRT.FET
HRT.VENT.L
HRT.ATR.R
HRT.VNT.R
VAS.AOR
Strong_HEH_Enhancers
HEH_Super_Enhancers
SCRG1
HAND2
100 Vert. Cons
C.
chr17:
12,450,000
12,500,000
12,550,000
12,600,000
12,650,000
12,700,000
12,750,000
MYOCD
ARHGAP44



F.

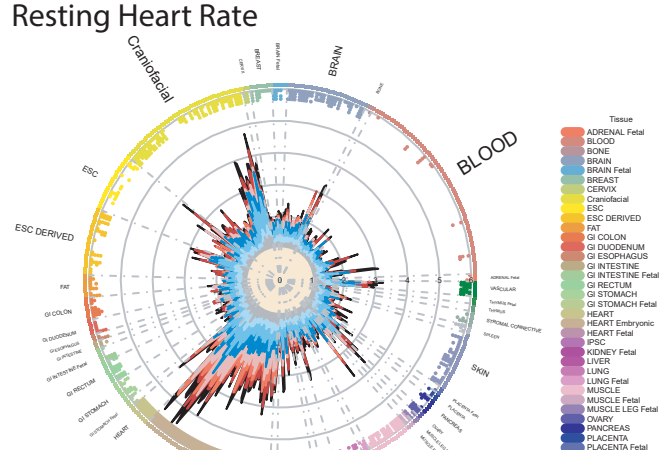


G.

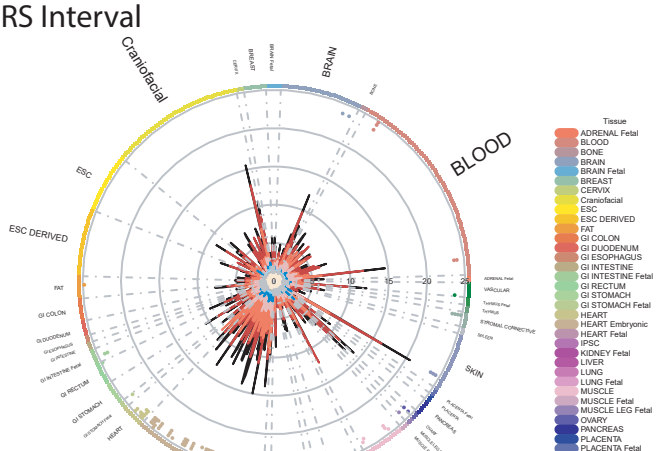


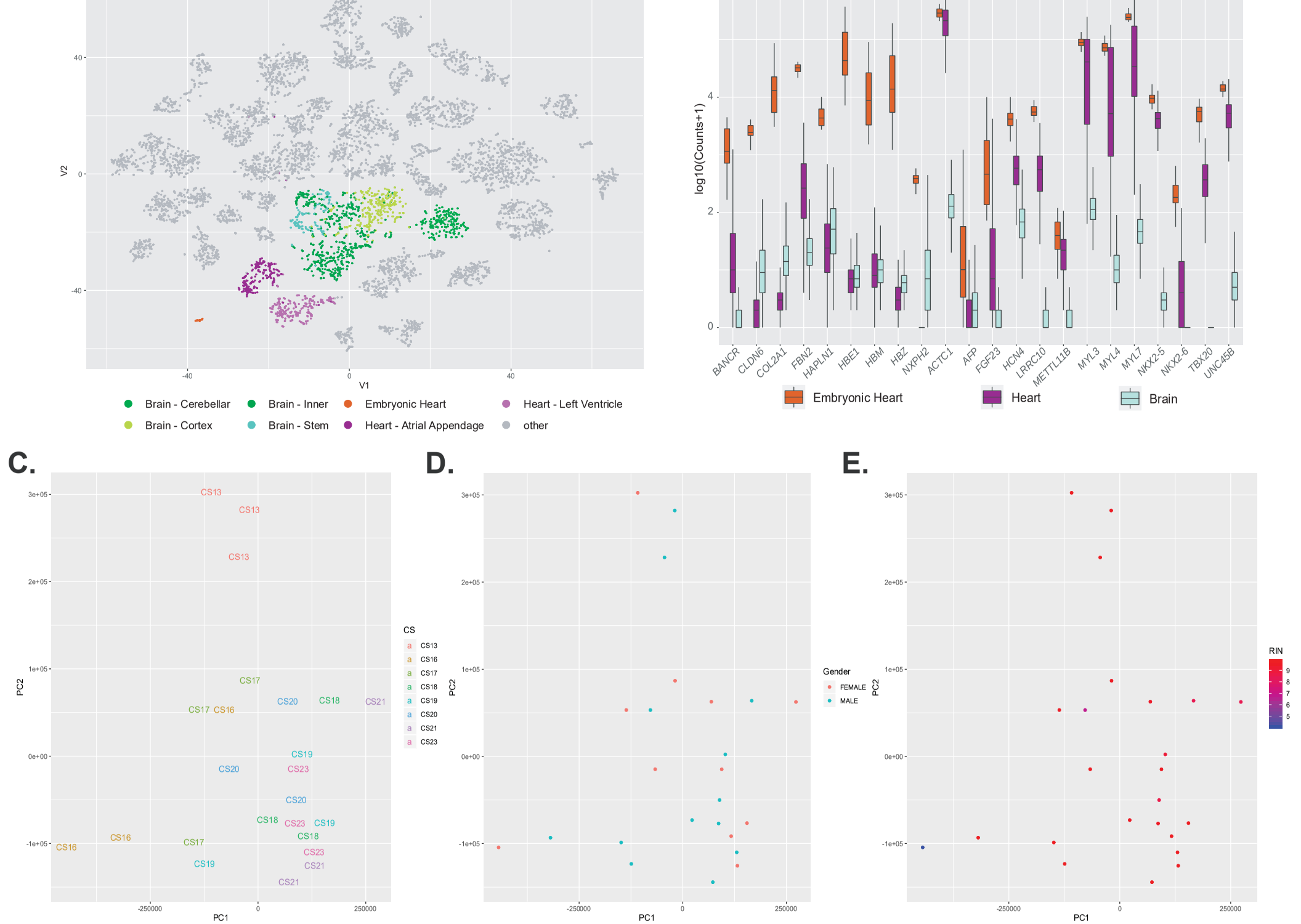
ADRENAL Fetal
BLOOD
BONE
BRAIN
BRAIN Fetal
BREAST
CERVIX
Craniofacial
ESC
ESC DERIVED
FAT
GI COLON
GI DUODENUM
GI ESOPHAGUS
GI INTESTINE
GI INTESTINE Fetal
GI RECTUM
GI STOMACH
GI STOMACH Fetal
HEART
HEART Embryonic
HEART Fetal
IPSC
KIDNEY Fetal
LIVER
LUNG
LUNG Fetal
MUSCLE
MUSCLE Fetal
MUSCLE LEG Fetal
OVARY
PANCREAS
PLACENTA
PLACENTA Fetal
SKIN
SPLEEN
STROMAL CONNECTIVE
THYMUS
THYMUS Fetal
VASCULAR

H. Resting Heart Rate

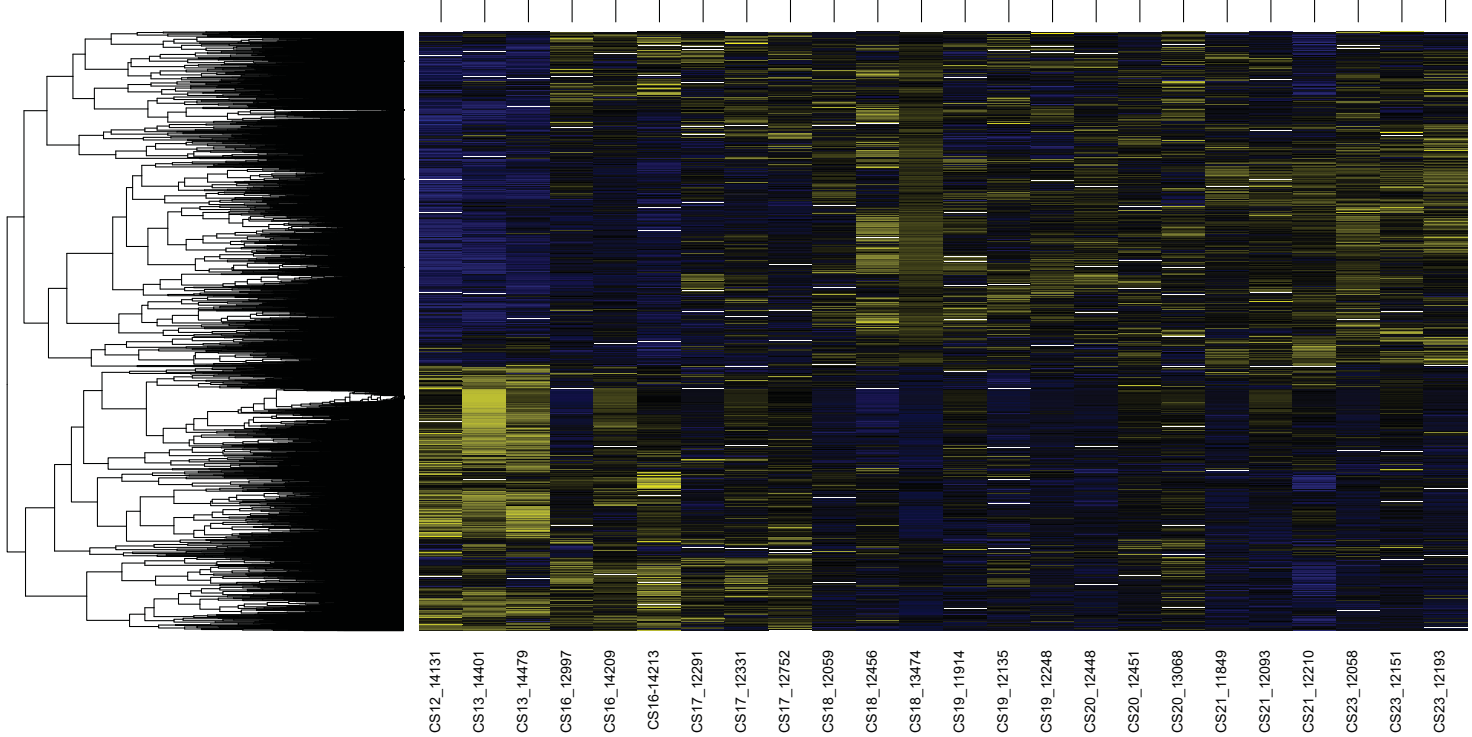


I. QRS Interval

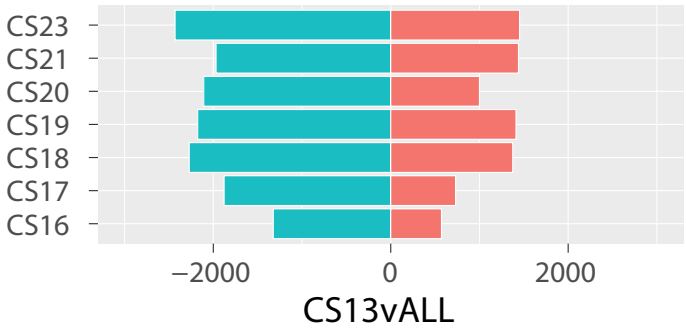


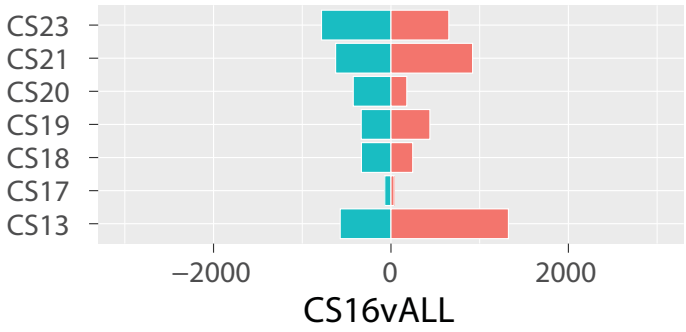


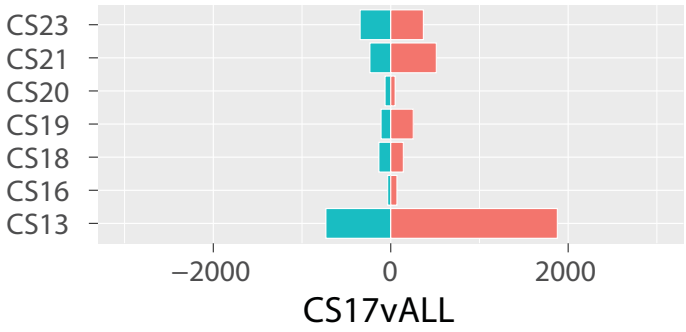
Online Figure XII. tSNE, Highest GINI box plot, and PCA for RNA-Seq Data. A. tSNE plot using all genes annotated by Gencode (v25) quantified by Rail-RNA form this study or GTEx and retrieved from recount2. All samples profiled in this study (red) cluster well with one another relative to other human tissues including adult heart (purples) and brain (greens). B. Box plots of expression values based on log10 scaled counts for samples

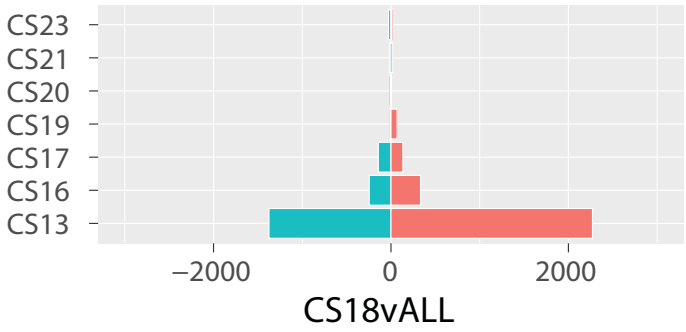

CS12_14131
CS13_14401
CS13_14479
CS16_12997
CS16_14209
CS16-14213
CS17_12291
CS17_12331
CS17_12752
CS18_12059
CS18_12456
CS18_13474
CS19_11914
CS19_12135
CS19_12248
CS20_12448
CS20_12451
CS20_13068
CS21_11849
CS21_12093
CS21_12210
CS23_12058
CS23_12151
CS23_12193

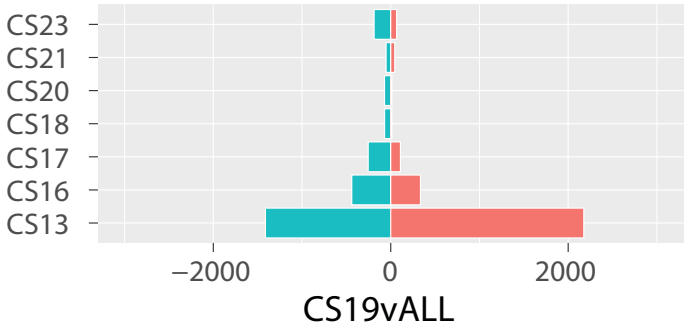
B.

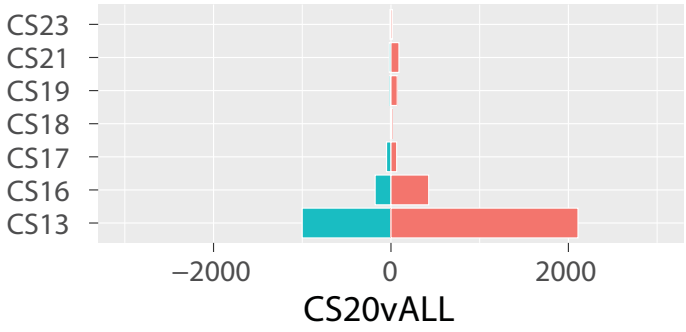

CS23
CS21
CS20
CS19
CS18
CS17
CS16
−2000
0
2000
CS13vALL

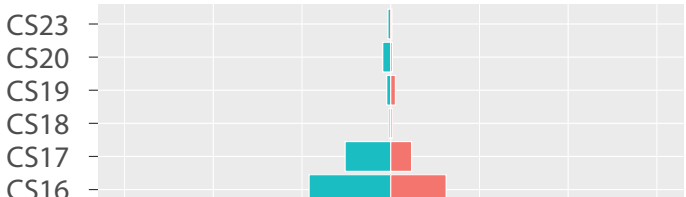

CS23
CS21
CS20
CS19
CS18
CS17
CS13
−2000
0
2000
CS16vALL

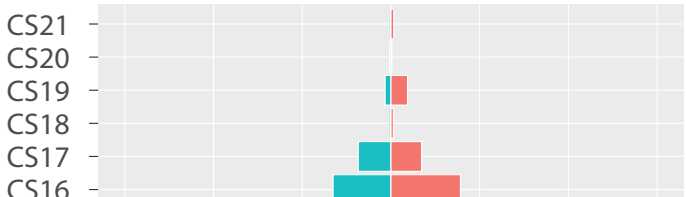

CS23
CS21
CS20
CS19
CS18
CS16
CS13
−2000
0
2000
CS17vALL

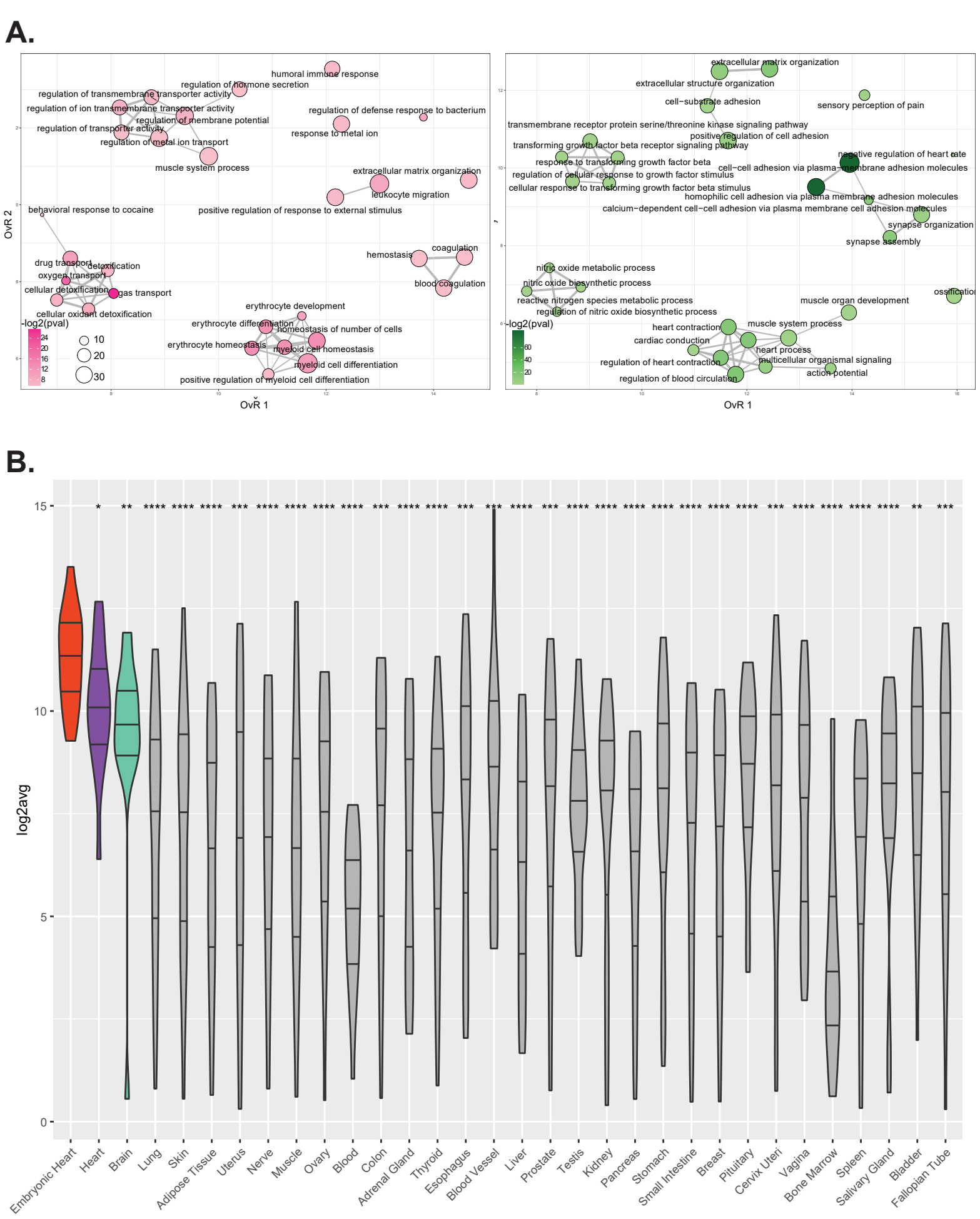

CS23
CS21
CS20
CS19
CS18
CS17
CS16
CS13
−2000
0
2000
CS18vALL


CS23
CS21
CS20
CS18
CS17
CS16
CS13
−2000
0
2000
CS19vALL

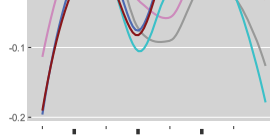

CS23
CS21
CS19
CS18
CS17
CS16
CS13
−2000
0
2000
CS20vALL


CS23
CS20
CS19
CS18
CS17
CS16


CS21
CS20
CS19
CS18
CS17
CS16



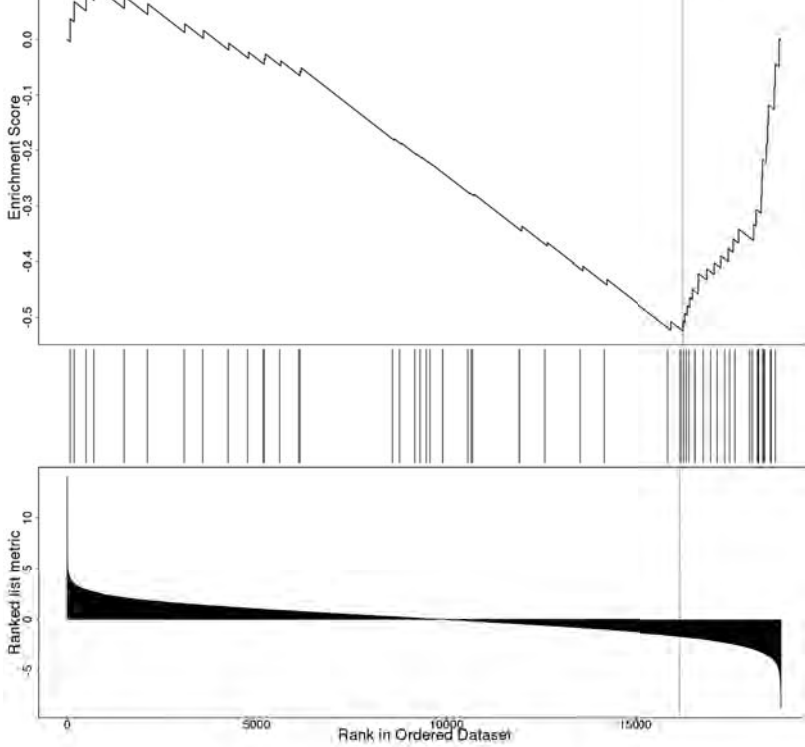
Online Figure XIV. Gene Ontology Dot Plots and Full Differential Expression Heatmap. A The gene ontologies for the CS16 (pink cluster, left panel) time point. The gene ontologies for the CS17 (green cluster, right panel) time periods. B. Violin plot of the expression of Genes with greater than 15 embryonic heart-specific enhancers in embryonic heart samples, as well as all other GTEx tissues. Significance of differences were calculated using the



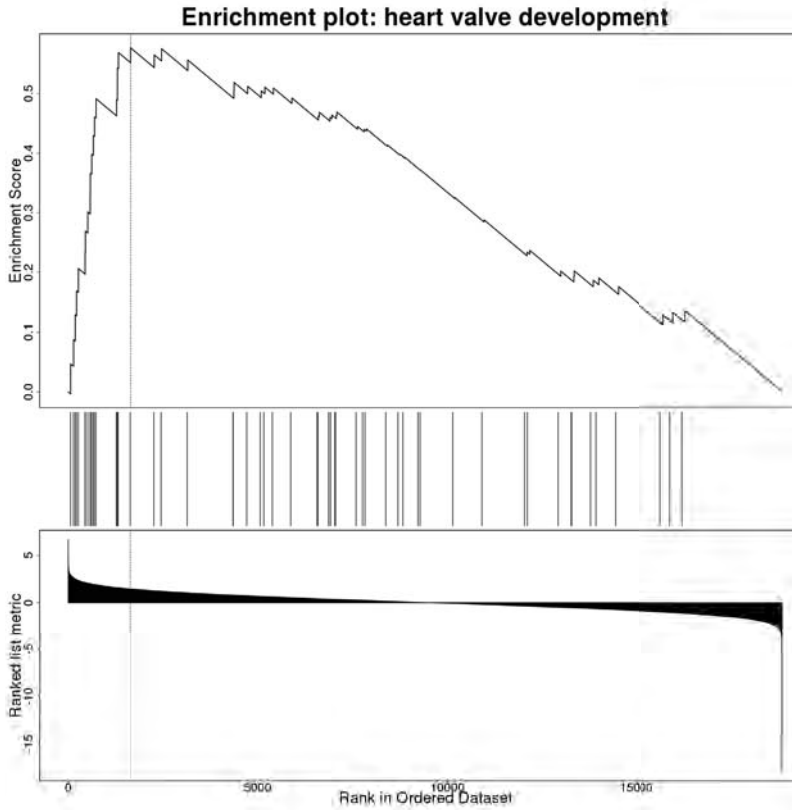
heart valve development
collagen metabolic process
response to interferon−gamma
angiogenesis
tissue migration
regulation of vasculature development
regulation of small GTPase mediated signal transduction
pri−miRNA transcription by RNA polymerase II
ameboidal−type cell migration
Ras protein signal transduction
response to interleukin−1
tissue remodeling
vascular endothelial growth factor receptor signaling pathway
coagulation
extracellular structure organization
endothelium development
bone development
inositol lipid−mediated signaling
interleukin−6 production
vasculogenesis
positive regulation of cell motility
cell−substrate adhesion
platelet morphogenesis
negative regulation of locomotion
aminoglycan metabolic process
pyridine−containing compound metabolic process
drug catabolic process
ribonucleoprotein complex subunit organization
protein localization to mitochondrion
coenzyme metabolic process
small molecule catabolic process
mitochondrial membrane organization
cytoplasmic translation
nucleoside phosphate biosynthetic process
purine−containing compound biosynthetic process
nucleoside diphosphate metabolic process
cellular modified amino acid metabolic process
ribonucleotide metabolic process
tricarboxylic acid metabolic process
ribose phosphate biosynthetic process
cellular amino acid metabolic process
antibiotic metabolic process
generation of precursor metabolites and energy
nucleotide phosphorylation
cofactor catabolic process
nucleoside triphosphate metabolic process
translational initiation
pyruvate metabolic process
nucleoside monophosphate metabolic process
mitochondrial respiratory chain complex assembly
mitochondrial gene expression
NADH dehydrogenase complex assembly
translational elongation

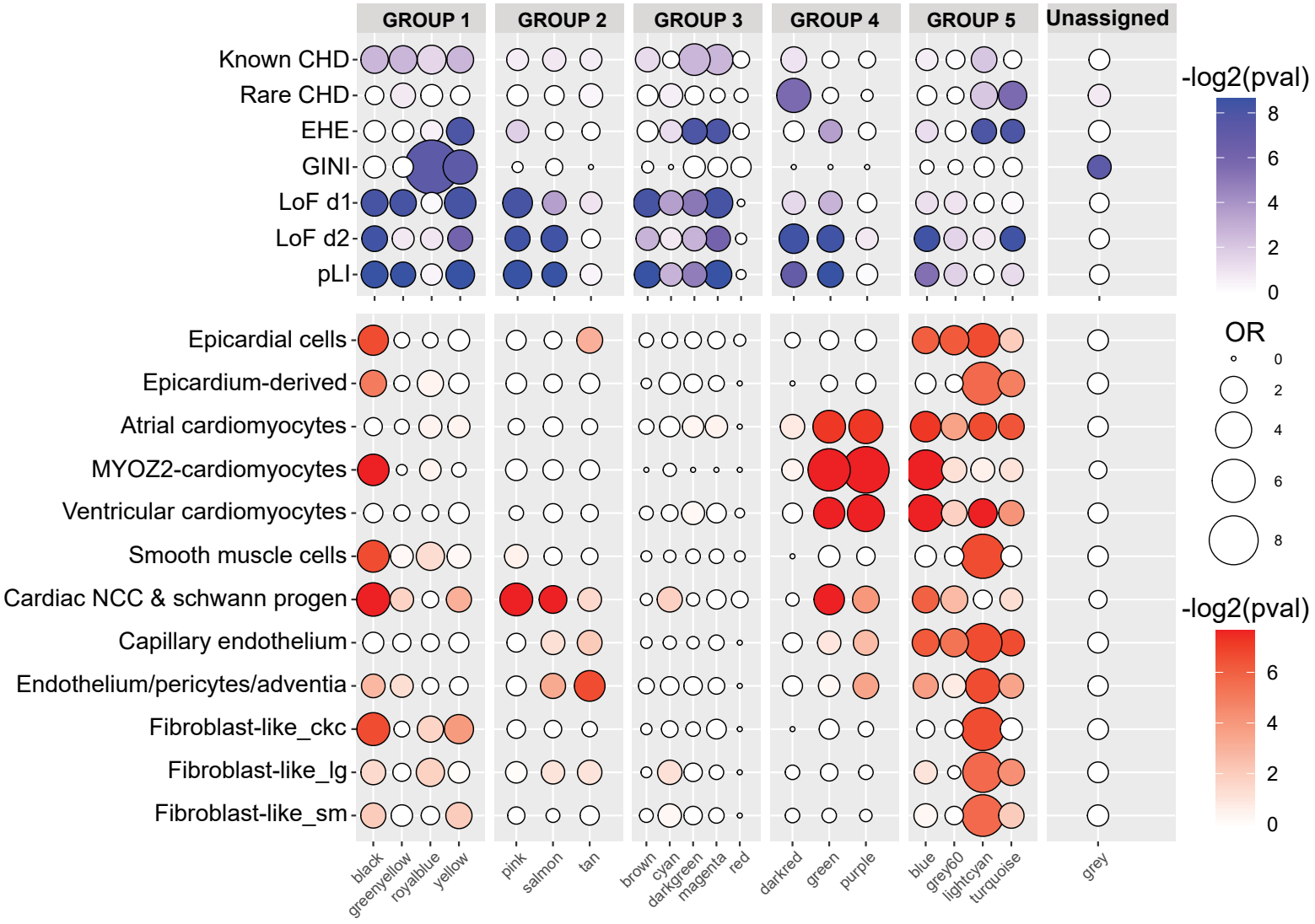
−2 −1 0 1 2 −2 −1 0 1 2

Normalized Enrichment Score



C CS18 → CS20





Online Figure XVI. Enrichment of embryonic heart specific gene lists in developing Brain WGCNA. Enrichment of curated gene lists in WGCNA of human time-series brain RNA-seq. The enrichment of heart specific lists (EHE, GINI) do not line up with cardiomyocytes like was observed for the embryonic heart network. Also there is significant enrichment of embryonic heart specific genes in their null module (unassigned, grey). As expect-ed, a subset of modules get enrichment for constrained genes (LoF d1/d2, pLI). Related to Figure 8.