

## Enhancer and promoter usage in the normal and failed human heart

Anthony M. Gacita BS<sup>1</sup>, Lisa Dellefave-Castillo MS CGC<sup>1</sup>, Patrick G. T. Page BS<sup>1</sup>, David Y. Barefield PhD<sup>1</sup>, J. Andrew Waserstrom PhD<sup>2</sup>, Megan J. Puckelwartz PhD<sup>1</sup>, Marcelo A. Nobrega MD PhD<sup>3</sup>, and Elizabeth M McNally MD PhD<sup>1</sup>

<sup>1</sup>Center for Genetic Medicine, Northwestern University Feinberg School of Medicine, Chicago IL

<sup>2</sup>Department of Medicine (Cardiology), Northwestern University Feinberg School of Medicine, Chicago IL.

<sup>3</sup>Department of Human Genetics, The University of Chicago, Chicago IL

**Keywords:** heart failure, cardiomyopathy, gene expression, promoter, enhancer, regulatory sequences

**Word count:** 5331 (excluding references and legends)

**Short title:** *Promoter and enhancer shifts in heart failure*

## ABSTRACT

The failed heart is characterized by re-expression of a fetal gene program, which contributes to adaptation and maladaptation in heart failure. To define genomewide enhancer and promoter use in heart failure, Cap Analysis of Gene Expression (CAGE-seq) was applied to healthy and failed human left ventricles to define short RNAs associated with both promoters and enhancers. Integration of CAGE-seq data with RNA sequencing identified a combined ~17,000 promoters and ~1,500 enhancers active in healthy and failed human left ventricles. Comparing promoter usage between healthy and failed hearts highlighted promoter shifts which altered amino-terminal protein sequences. Comparing enhancer usage between healthy and failed hearts revealed a majority of differentially utilized heart failure enhancers were intronic and primarily localized within the first intron, identifying this position as a common feature associated with tissue-specific gene expression changes in the heart. This dataset defines the dynamic genomic regulatory landscape underlying heart failure and serves as an important resource for understanding genetic contributions to cardiac dysfunction.

## INTRODUCTION

Heart failure is a leading cause of death worldwide.<sup>1, 2</sup> The failed heart is characterized by reduced function, impaired filling, and altered metabolism, all of which contribute to an inability to meet the body's demands for normal activity. Heart failure is associated with global changes in gene expression and splicing, some of which directly drive pathological and adaptive remodeling.<sup>3</sup> For example, the failed heart shifts its metabolism towards glycolysis, driven in part by gene expression changes.<sup>4, 5</sup> Within the sarcomere, isoform switches alter the composition of myosin heavy chain in the failed heart,<sup>6, 7</sup> and alternatively spliced isoforms *TNNT2* and *TTN*, encoding troponin T and titin, respectively, directly modify contractility and compliance,<sup>8-10</sup> and importantly many of these switches differ between the human and mouse heart. In addition, mutations within the coding region of many of these genes directly lead to cardiomyopathy and heart failure.<sup>11, 12</sup> Although specific genetic regulatory regions have been well characterized,<sup>13-15</sup> comparatively few genomewide analyses have been conducted. Genomewide epigenomic profiles have been used to infer regulatory regions in the developing mouse heart and embryonic stem cell derived-cardiomyocytes.<sup>16, 17</sup> Chromatin capture integrated with CTCF binding site maps was applied to mouse cardiomyocytes subjected to pressure overload.<sup>18</sup> However, much less is known about the promoter and enhancer shifts underlying human heart failure.

Gene expression is driven by transcription factor binding at promoters and enhancers, which interact in three-dimensional space to increase gene expression. Estimates of number of active heart enhancers vary from several thousand to tens of thousands depending on the approach used.<sup>19, 20</sup> One assessment of active cardiac enhancers defined p300/CBP binding sites from human fetal and one adult failed heart, requiring candidate enhancer regions position > 2.5kb from any annotated transcriptional start site.<sup>19</sup> This analysis suggested there were 5,000 enhancers active in fetal tissue and over 2,000 enhancers active in adult tissue, with approximately half of adult heart enhancers also active in fetal heart and underscoring the fetal

re-expression program.<sup>19</sup> A similar approach used normal human and mouse hearts integrating p300/CBP binding sites with H3K27ac marks and requiring candidate enhancers regions position > 1.5kb from any annotated transcriptional start site.<sup>20</sup> This integrated approach described more than 80,000 potential heart enhancers.<sup>20</sup> These studies provide a valuable dataset of enhancers in the normal developing and adult heart, but do not identify the genomic alterations seen in pathologic states, such as heart failure.

In addition to the use of alternative regulatory elements during heart failure, alternative promoter usage represents a distinct mechanism to regulate gene expression. Typically driven by the inclusion of alternative first exons, alternative promoters are estimated to affect 30-50% of human genes.<sup>21, 22</sup> Alternative promoters may affect the amino-terminal amino acids of proteins and/or the 5'UTR of transcripts, both of which can mediate functional consequences. Alternative promoters can also influence the effect of genetic variants on protein function and thus are vital for accurate variant effect predictions.<sup>23</sup> Despite the potential of broad proteome differences due to alternative promoter usage, a genomewide view of promoter and enhancer shifts in human heart failure is lacking.

Next generation sequencing technologies including Cap-Analysis of Gene Expression (CAGE) have made it possible to assay transcriptome usage by determining RNA transcriptional start sites at single base pair resolution.<sup>24</sup> Enhancer regions are transcribed into low-abundance enhancer RNAs (eRNAs) in a bidirectional pattern,<sup>25, 26</sup> contrasting with the unidirectional transcriptional expression seen near gene promoters, which produce high-abundance signals. Because of the precision with which these RNAs can be mapped, it is possible to accurately map enhancer and promoter signals at high resolution and without using blanket arbitrary filters that ultimately limit identification. For example, enhancers throughout the genome can be identified, as this analysis does not require removal of promoter-proximal intervals.

To define alternative promoter and enhancer use in heart failure, we generated CAGE sequence datasets from healthy and failed human left ventricles. CAGE sequencing information was integrated with RNA sequencing from these same samples, to improve sensitivity in detecting low-abundance eRNAs and rarely used gene promoters. We relied on a no-amplification non-tagging CAGE sequencing protocol, which allows for more robust and less biased detection of transcriptional start sites.<sup>24</sup> These data identified unique signatures of housekeeping and tissue specific promoters, as well as a pattern of enhancers within first introns that regulate tissue specific expression. In addition to identifying differential enhancer use in heart failure, we cataloged 129 genes with differential promoter usage in heart failure. These alternative promoters have the potential to encode proteins with unique amino-termini, highlighting potential protein composition shifts in the failed heart.

## RESULTS

**CAGE sequence clusters identify promoters and enhancers of the left ventricle.** CAGE sequencing identifies promoters and enhancer regions. Given the known gene expression differences that characterize failed hearts, we generated CAGE datasets from left ventricle (LV) taken from three healthy and four failed hearts. Healthy LV samples were those acquired but not used for transplant due to age or other incompatibility. Failed hearts were obtained at the time of transplant from patients with a range of primary mutations and ages (**Table 1**). Each library was sequenced to high depth and libraries demonstrated comparable alignment rates (**Supplemental Table 1**). To generate a comprehensive list of all potential promoters and enhancers, the initial analysis included the combination of healthy and failed hearts. CAGE sequence analysis identified 23,676 unidirectional sequence clusters, indicative of promoter regions, and 5,647 bidirectional sequence clusters, indicative of enhancer regions.

Unidirectional CAGE sequence clusters were annotated using Ensembl designations to map their position relative to annotated genes. Unidirectional sequence clusters that mapped

$\pm 100$ bp of transcriptional start sites constituted 70.1% of the clusters, consistent with their putative role as promoters (**Figure 1A**). An additional 8.1% of unidirectional clusters, mapped between 100 and 1000bp upstream of transcriptional start sites. The remaining 21.8% of sequence clusters mapped to 5'UTR, 3'UTR, exons, introns or intergenic regions.

We next analyzed clusters for their transcription factor binding motif composition. The 70.1% of clusters mapping within 100bp of transcriptional start sites were highly enriched for GFY-Staf, Sp1, and Elk/ETS binding motifs, which are known promoter binding transcription factors.<sup>27</sup> Clusters mapping into other gene regions showed minimal enrichment of these motifs (**Figure 1B**). To provide additional support for the promoter-enriched sequence clusters, ATAC sequencing and H3K4me3 ChIP-seq datasets were compared; **Supplemental Table 2** provides source information on comparison datasets. Clusters overlapping promoters overlapped considerably with ATAC-seq and H3K4me3 ChIP-seq signals, indicative of open chromatin and active promoter regions (**Figure 1C**). These clusters also showed high CTCF and Pol2A binding as well as a reduction of H3K4me1 histone modifications, consistent with their role as promoters and not enhancers (**Supplemental Figure 3**). The bimodal shape of histone methylation patterns is consistent with open chromatin signals being flanked by promoter histone marks. Taken together, these data identify unidirectional CAGE sequence clusters as bearing the genomic signatures of active promoters.

The bidirectional CAGE clusters were similarly annotated with Ensembl designations. Only 44.5% of bidirectional clusters mapped  $\pm 100$ bp within transcriptional start sites. In contrast to unidirectional clusters, 24.3% of clusters mapped to gene introns and 7.6% were intergenic (**Figure 1D**). When analyzed for known transcription factor motifs, intergenic and intronic clusters showed enrichment of GATA, GRE, and MEF2 motifs, essential transcription factors for cardiomyocyte specification and maintenance (**Figure 1E**).<sup>28, 29</sup> Intergenic bidirectional clusters showed enrichment of open chromatin signals (ATAC-seq), H3K4me1, and H3K27Ac histone

modifications in human left ventricles. Intronic clusters also showed a similar pattern, but with lower magnitude (**Figure 1F**). The intergenic and intronic bidirectional clusters showed enrichment of CTCF and Pol2A binding as well as reduced H3K4me3 modifications (**Supplemental Figure 3**). These patterns are highly consistent with the role of bidirectional CAGE sequence cluster regions as being enhancers, rather than promoters.

**CAGE sequence-defined promoters show shape divergence.** Mammalian promoters initiate transcription across broad or narrow genomic regions and correlate with distinct transcriptional regulatory mechanisms.<sup>30</sup> We evaluated cardiac promoters for these two major types of promoters by calculating the interquartile range (IQR) of promoter CAGE clusters by determining the base pair distance between 10% and 90% of a promoter's total signal. We observed the expected two distinct populations, defined as sharp (IQR < 10bp) and broad promoters (IQR  $\geq$  10bp) (**Figure 2A**). Subjecting these two types of promoters to gene ontology analysis, we observed that broad promoter genes were those associated many cellular functions, while genes regulated by sharp (narrow) promoters were significantly enriched for muscle and cardiac gene ontology terms (**Figure 2B**). Thus, tissue specific genes important for left ventricular specification and function were more likely to have sharp promoters.

Sharp and broad promoters also demonstrated differential enrichment of upstream sequence DNA-binding motifs. Sharp promoters had TATA motifs at positions 30-33 upstream of the predominant transcriptional start site, representing canonical TATA boxes. Broad promoters were devoid of TATA motifs, but did show enrichment of GC nucleotides, which likely represents CpG-islands.<sup>31</sup> Both classes of promoters showed a preference of transcription initiation at a G nucleotide and an upstream CC sequence at the -2 and -3 positions (**Figure 2C**). Sharp, tissue specific promoters were also more highly expressed compared to broad promoters, and this observation was driven by a smaller population of very highly expressed sharp promoters (**Figure 2D**). We compared promoter shape between healthy and failed hearts

and found a modest but significant genomewide increase in promoter IQR in failed hearts (**Figure 2E**).

**Intronic Enhancers map within the first Intron.** A large proportion of bidirectional CAGE clusters are located within gene introns. Intronic clusters shared transcription factors and epigenetic marks with intergenic CAGE clusters, suggestive of their roles as enhancers (**Figure 1E**). While a typical human gene contains on average 7.8 introns,<sup>32</sup> we observed that the majority (69%) of intronic enhancers in this dataset mapped to the first intron (**Figure 3A**). First intron enhancers generated more eRNA than enhancers in other introns, but were not wider and did not differ in their balance of bidirectionality (**Figure 3B**). Notably, all intronic enhancers mapped within genes enriched for cardiac and muscle gene ontology terms (**Figure 3C**), suggesting the importance of first intron enhancers for determining tissue specificity.

**Correlation of CAGE-sequencing and RNA-sequencing.** RNA sequencing was carried out on the same left ventricles and compared to CAGE sequencing. Since CAGE sequencing quantifies promoter expression, it also can measure overall gene expression, and this was consistent with the tight correlation between CAGE sequencing and RNA sequencing (**Figure 4A**). Additionally, we assessed correlations between pairs of samples. In general, healthy hearts correlated best with other healthy hearts, and failed hearts compared best with failed hearts. The RNA sequence and CAGE sequence expression estimates were most correlated for matched samples except for Failed Heart 4, which likely reflects the lower CAGE sequence read depth in this sample (**Figure 4B**). Gene expression differences were identified using both CAGE and RNA sequence datasets. RNA-seq identified more upregulated and downregulated genes, and approximately half of the genes identified by CAGE-seq were also identified by RNA-seq (**Figure 4C**). Gene ontology analyses on differentially expressed genes were similar in both sequence datasets. Genes associated with developmental pathways and extracellular

matrix organization were upregulated in heart failure while genes associated with catabolism were downregulated in heart failure (**Figure 4D**), consistent with prior reports.<sup>3, 4</sup>

**CAGE sequencing-defined enhancer regions validated by other enhancer datasets.** Of the ~1,800 enhancer regions identified by CAGE sequencing, data was available from 45 of these in the Vista Enhancer Browser, a database of enhancers tested in an *in vivo* reporter assay using transgenic mouse embryos.<sup>33</sup> Of the 45 present in Vista, 31 (70%) demonstrated reproducible activity in the developing mouse heart) (**Supplementary Figure 4A**). We also compared CAGE sequence-predicted enhancers to those predicted by H3K27Ac and p300 ChIP sequencing from developing and adult human and mouse tissues.<sup>20</sup> CAGE-sequence defined enhancers showed significantly higher overlap to H3K27Ac/p300 ChIP regions compared to length-matched scrambled control regions (**Supplementary Figure 4B**). An additional study reporting H3K27Ac ChIP-seq data from healthy and failed human hearts was similarly compared and showed significant overlap (**Supplementary Figure 4C**).<sup>34</sup> Finally, we compared CAGE sequence-defined enhancer predictions from the FANTOM consortium, which used CAGE sequencing across many tissues to define enhancers.<sup>35</sup> The CAGE sequencing in this current study showed significant overlap with FANTOM predictions with the highest percentage of overlap for ubiquitous enhancers but also overlap with left ventricle predicted enhancers (**Supplementary Figure 4D**). Many enhancers described here were not present in FANTOM predictions because of higher depth sequencing in our study (**Supplementary Table 1**). Taken together, the intersection of these orthogonal datasets with CAGE sequence data corroborate the notion that we have uncovered cardiac enhancers both in healthy hearts and in heart failure.

**Alternative promoter usage in heart failure.** In the LV, 3,032 (23%) expressed genes have evidence for more than one promoter (**Figure 5A**). We used CAGE sequencing data to

calculate the percentage of total transcripts coming from each promoter in multi-promoter genes. We compared the average percent-usage of each promoter in healthy and failed hearts and found 609 promoters in 325 genes with a shift  $\geq 10\%$  (**Figure 5B**). Of these, 149 promoters in 124 genes occurred after the exon containing the start codon, indicating the potential to alter the amino-terminal amino acid sequence of the resulting protein (**Figure 5C**). Of the 124 genes identified as having alternative promoters that occur after start codons in heart failure, many are associated with sarcomere regulation or muscle structure development, including *TNNT*, *MYOT*, and *SPEG*. We annotated a significant promoter switch in *PRKAG2*, a gene linked to hypertrophic cardiomyopathy and critical to heart metabolism.<sup>36-38</sup> Three major *PRKAG2* promoters were identified, encoding three different isoforms-  $\gamma 2a$ ,  $\gamma 2-3b$ , and  $\gamma 2b$  (**Figure 5D**). In healthy hearts, the relative expression of these three transcripts is 53%  $\gamma 2b$ , 28%  $\gamma 2-3b$ , and 17%  $\gamma 2a$ . In heart failure, these percentages significantly shift with 29%  $\gamma 2b$ , 59%  $\gamma 2-3b$ , and 10%  $\gamma 2a$  isoform (**Figure 5E**). Notably, the  $\gamma 2-3b$  isoform encodes a unique 32 amino acid sequence at the amino-terminus (**Figure 5F**).

**Enhancer usage shifts in heart failure.** The CAGE sequence analysis identified  $\sim 1800$  enhancer regions actively transcribed in human left ventricle (**Figure 1A**). Multidimensional scaling of normalized expression levels showed an overall similar profile of enhancer usage in healthy heart samples, but more disparate enhancer usage across the four failed hearts (**Figure 6A**). Comparing enhancer usage across healthy and failed hearts reveals 264 enhancers changing significantly in heart failure (raw p-value  $\leq 0.05$ ). To assess whether differential enhancer transcription was associated with differential transcription factor binding site profiles, we compared transcription factor motif instances across enhancers in healthy and failed hearts. We found that motifs for SMAD2, NFIX, NFAT, TCF7L2, ZNF740, and AR were enriched in enhancers that change in heart failure. SMAD2, NFIX, TCF7L2, and AR motifs were found

more in downregulated enhancers. While NFAT and ZNF740 motifs were found more in upregulated enhancers. RNA-sequencing demonstrated that *SMAD2* and *NFAT5* were significantly upregulated in heart failure (**Figure 6C**). **Figure 6D** illustrates alternative enhancer use within the first intron of *TRPM7*, which encodes the transient receptor potential cation channel subfamily M member, a gene implicated in ischemic cardiomyopathy and cardiac rhythm.<sup>39, 40</sup> This intronic enhancer showed significantly lower eRNA expression in heart failure (**Figure 6E**), concomitant with a significantly lower expression of *TRPM7* in failed hearts (**Figure 6F**).

## DISCUSSION

**Defining the promoterome of the human LV in health and disease.** Using CAGE-sequencing, we provide the most detailed genome-wide analysis of promoter usage in human left ventricle to date. The majority of CAGE sequence clusters overlapped the promoter regions of known annotated RNAs. Additionally, clusters mapping to promoter regions were enriched for promoter-bound transcription factors and epigenetic modifications compatible with their annotation as promoters. In total, we report ~17,000 high likelihood promoters active in the adult human heart. We observed two major promoter types, the sharp TATA-box-associated and the broad CpG island-associated.<sup>31</sup> Sharp promoters had single or a few transcriptional start sites and were linked to highly-expressed, tissue-specific genes like *MYH7*, *TTN*, and *MYL2*. Broad promoters had a wider distribution of transcriptional start sites and included both housekeeping and some tissue specific genes. We observed an increase in genome-wide promoter width in failed left ventricles, suggesting a loss of tight regulation of transcriptional start sites. This widening of promoters may reflect epigenetic modifications or transcriptional factor profile differences, both of which are known to occur in heart failure and hypertrophy.<sup>41, 42</sup>

**Alternative promoter usage in heart failure.** The CAGE sequence data indicate that ~20% of genes active in the human left ventricle have multiple active promoter regions, correlating well with previous estimates from different cell types.<sup>43, 44</sup> Promoter switches can generate unique 5'UTRs without changing the resulting protein, or result in novel amino-termini if the switch activates an alternative promoter downstream of the translation start site. Altering the 5'UTR may affect translational efficiency, imparting developmental and tissue specificity.<sup>45</sup> We demonstrated a significant shift in promoter usage in the *PRKAG2* gene. Mutations in *PRKAG2* have been linked to hypertrophic cardiomyopathy and arrhythmias.<sup>36</sup> *PRKAG2* is highly expressed in the heart, and it was reported that the  $\gamma$ 2b-3b isoform was most highly expressed in the heart.<sup>46</sup> In healthy left ventricles, we found that ~55% of transcripts originated from the  $\gamma$ 2b promoter and ~35% originate from the  $\gamma$ 2-3b promoter. In failed left ventricles, ~60% of *PRKAG2* transcripts represent the  $\gamma$ 2b-3b isoform, which has a unique 32 residue amino-terminal domain that may affect the cellular localization of *PRKAG2*. Data from a yeast two-hybrid screen indicated that the amino-terminus of the longer  $\gamma$ 2b, which shares some homology with the  $\gamma$ 2b-3b isoform, interacts with troponin I.<sup>47</sup> Upregulation of the  $\gamma$ 2b-3b isoform in failed hearts may also confer the cardiac-specific phenotype of *PRKAG2* mutations. This data indicates that alternative promoter usage may have broad implications on the cellular proteome.

**Enhancer enrichment in the first introns.** A significant portion of CAGE-defined enhancers were located within intronic regions of overlapping annotated transcripts, showing similar epigenetic modifications and transcription factor motifs to intergenic enhancer regions. The majority of intronic enhancers were located in the first intron. Genes with highly expressed first intron enhancers included *MYLK3*, *TPM1*, *SLC8A1*, and *FLNC*, which are all genes important for cardiac performance. Classic intronic enhancers have been described in mammalian systems.<sup>48, 49</sup> A study of copy number variants within intronic regions saw significant

correlations between intronic deletions and gene expression in lymphoblastoid cell lines.<sup>50</sup>

Additionally, a study using histone modifications to identify enhancer regions active in stem-cell derived cardiomyocytes identified a large proportion of enhancers within the first intron.<sup>16</sup>

**Differential enhancer usage in heart failure.** Heart failure is associated with both structural and transcriptional changes.<sup>3, 51</sup> We found that enhancer usage was more variable in failed ventricles, which may indicate genome-wide dysregulation of gene expression. SMAD2 binding motifs were enriched in differentially used heart failure enhancers, and this is highly consistent with the known upregulation of TGF- $\beta$  signaling in failing hearts.<sup>52, 53</sup> The enrichment of this motif in differential enhancers may be due to increased TGF- $\beta$ /SMAD activation in cardiomyocytes and/or a larger proportion of cardiac fibroblasts in the failed left ventricle tissues. SMAD motifs were found more often in downregulated enhancers, implying a repressive role for TGF- $\beta$ /SMAD signaling in heart failure. We highlighted a specific differential enhancer located within the first intron of the *TRPM7* gene. *TRPM7* encodes kinase domain-containing cation channel. Deletion of *Trpm7* in mice disrupts cardiac automaticity and causes cardiac hypertrophy and fibrosis.<sup>40, 54</sup> In ischemic cardiomyopathy, *TRPM7* was significantly down regulated in the left atria and ventricle.<sup>39</sup> However, in non-ischemic dilated cardiomyopathy, *TRPM7* was shown to be upregulated in patients with ventricular tachycardia.<sup>55</sup> We found a larger magnitude reduction in enhancer eRNA levels compared to reduction in mRNA levels, and this may indicate that additional *TRPM7* enhancer regions are active. Our findings support a reduction in *TRPM7* in the setting of end stage heart failure.

**Study Limitations and Conclusions.** This study used CAGE sequencing to define a broad spectrum of cardiac promoters and enhancers, with focus on their differential use in heart failure. We observed variability in differential promoter and enhancer usage in failed heart, as

the normal control hearts showed tighter correlations. This variability may reflect the end stage process of heart failure. While a larger dataset may be more revealing, the diversity of response in the failed hearts mirrors what has been observed when RNA sequencing was used to define transcripts produced for *TTN*, a large gene that has been examined in multiple failed hearts.<sup>56, 57</sup> The wide array of transcripts produced from even this single gene may underscore that a lack of uniform response itself could contribute to heart failure.

## METHODS

**Materials, Code and Data Availability.** All scripts used in this analysis are uploaded to the github page (in progress). Sequence data has been uploaded to the NCBI-GEO under accession number (in progress).

**RNA-Extraction, Library Preparation, and Sequencing.** Healthy and failed left ventricle samples were obtained from failed transplants or as discarded tissue, respectively. Living subjects provided consent. Healthy left ventricular samples were obtained from hearts provided by the Gift of Hope of Illinois and were found to be unsuitable for transplant due to age or prior cardiac surgeries. All patients were declared to be brain dead as the result of cerebral hemorrhage and familial consent was obtained for organ use in research. Tissues were snap-frozen in liquid nitrogen and stored at -80°C until use. Approximately 50mg of frozen tissue was ground into a fine powder using a mortar and pestle under liquid nitrogen. Ground powder was added to 1mL TRIzol (Invitrogen) containing 250ul of silica zirconium beads. Samples were placed in a bead homogenizer for 1 minute, allowed to cool on wet ice, and centrifuged at 12,000xg to remove any unhomogenized tissue pieces followed by chloroform extraction. Phases were separated by centrifugation and the upper aqueous layer was added to fresh 70% ethanol. The RNA-ethanol mix was used as in input to the Aurum Total RNA Mini Kit (BIORAD). RNA was isolated (including on-column DNase digestion) following manufacturer's instructions. Concentration was measured using a NanoDrop spectrophotometer and quality was assessed using an Aligent Bio analyzer. Only RNA extractions with RIN values  $\geq 7$  were used. If necessary, RNA extraction was repeated until  $\sim 10\mu\text{g}$  of RNA was obtained.

Custom nAnT-iCAGE-seq (no-Amplification-no-Tagging Illumina Cap Analysis of Gene Expression libraries) libraries were prepared by DNAFORM (Japan) following a previously-described protocol<sup>24</sup>. Briefly, 5 $\mu\text{g}$  of RNA was reverse transcribed using random primers. 5'-methyl-caps were biotinylated and enriched using streptavidin beads. cDNA was released and sequencing adapters were added using blunt-ended ligation. Following second strand synthesis, the libraries were quantified using qPCR.  $\sim 50\text{pM}$  of pooled libraries were loaded into an entire run on the NextSeq 500 (Illuminia) to yield  $\sim 400$  million total 75bp single end reads (**Supplemental Table 1**).

RNA-seq libraries were prepared using the TruSeq mRNA-seq library preparation kit (Illumina). Libraries were pooled in equimolar amounts and loaded on the HiSeq 4000 (Illumina) to generate  $\sim 40$  million 150bp paired-end reads/sample.

**CAGE-Seq Alignment and Clustering.** Raw CAGE-seq reads were checked with fastQC(v0.11.5) and aligned to the human genome (hg19) using STAR (v2.5.2) with default settings.<sup>58</sup> Uniquely aligning reads were inputted into CAGEr and converted into quantified CAGE transcriptional start site (CTSS) coordinates with removal of first G nucleotide mismatches.<sup>59</sup> CTSS coordinates and counts were outputted as bigwig files for input to CAGEfightR.<sup>60</sup> CTSSs from mitochondrial chromosomes and CTSSs only present in a single sample were removed. We clustered CTSSs from all samples into unidirectional and bidirectional clusters. CTSS's were required to have  $\geq 5$  pooled counts be included in a unidirectional cluster and all CTSSs within 20bp were merged into the same cluster. Unidirectional clusters also were required to have  $> 1$  TPM (tag per million) in at least 2 samples. Bidirectional clusters were required to have a balance score  $\geq 0.95$  and a 200bp window on either side of the midpoint was used to quantify each cluster, as described in.<sup>35</sup> Bidirectional clusters were also required to be bidirectional in at least one sample and have  $\geq 2$  counts in at least one sample. These clusters were annotated with Ensembl GTF file version 87 annotations

(downloaded May 2016), which includes known coding and noncoding RNA transcripts (**Supplemental Figure 2**). Unidirectional clusters overlapping known rRNA genes were also removed.

**CAGE Cluster Epigenetic and Transcription Factor Overlaps.** Epigenetic datasets of interest were downloaded from their respective locations (**Supplemental Table 2**). Bam files were converted into tag directories using HOMER.<sup>61</sup> HOMER annotatePeaks.pl was used to determine the read depth of each epigenetic dataset (normalized for cluster length and number) for each cluster annotation type ( $\pm 1000$ bp of the cluster midpoint). HOMER findMotifsGenome.pl was used to check for enrichment of known transcription factors for each annotation type. For unidirectional clusters, the cluster midpoint  $\pm 200$ bp was used as input. For bidirectional clusters, the cluster boundaries were used and no additional nucleotides were added. Homer generated background sequences generated from a masked genome were used.

**Promoter Width Analysis.** CAGEfightR was used to calculate the 0.1 to 0.9 inter-quantile range (IQR) for unidirectional clusters overlapping known promoter regions. A cutoff of 10bp was used to define a sharp and broad populations of promoters. The genomic coordinates of each promoter's predominant TSS with 100bp added upstream and 50bp added downstream were inputted into bedtools getfasta to obtain genomic sequences.<sup>62</sup> These sequences were inputted into the WebLogo tool to generate visual representations of nucleotide enrichment at each position relative to the predominant TSS.<sup>63</sup> The genes of each promoter type were also inputted into the PantherGO online tool to check for enrichment of gene ontology terms<sup>64</sup>. The R package ggplot2 was used to generate violin plots of sharp and broad promoter pooled expression levels and basepair width. To compare promoter IQR values across failed and healthy hearts, we first filtered out any promoters that were not present in all hearts. We used CAGEfightR with sample-specific scores to calculate the IQR of each promoter in each sample. We compared the average IQR across all promoters in the three healthy samples to the average IQR in the four failed samples using the nonparametric Wilcoxon rank sum test in R.

**Intronic Enhancer Analysis.** A custom script was written to generate an annotation file of first intron locations for all transcripts present in the Ensembl GTF file version 87 (Downloaded May 2016). We used custom Python script to evaluate if bidirectional intronic completely overlapped the first intron of any transcript. ggPlot2 was used to generate violin plots of enhancer eRNA expression levels, base pair width, and bidirectionality scores. Genes with first intron and other intron enhancers were inputted into the PantherGO online software tool to check for enrichment of gene ontology terms.<sup>64</sup>

**RNA-Seq Data Analysis and Comparison with CAGE-Seq Data.** Raw RNA-seq reads were trimmed with trimmomatic (v0.36) and aligned to the human genome (hg19) using STAR with default settings<sup>58</sup>. Uniquely aligned reads were assigned to genes using htseq-count using the Ensembl GTF file version 87 (Downloaded May 2016) as annotations.<sup>65</sup> Raw count matrices were inputted into EdgeR for normalization, dispersion estimation, and glm-model approach measures of differential expression between healthy and failed hearts.<sup>66</sup> Genes with <1 count per million were removed from the analysis. We defined differentially expressed genes as any gene with an FDR-corrected p-value of < 0.05. The read counts of CAGE-seq unidirectional clusters overlapping gene promoters were used to quantify overall gene expression. Expression values from multiple promoters of the same gene were collapsed into a gene-level value. These count values were inputted into EdgeR and analyzed similar to the RNA-seq data above. ggPlot2 was used to graph the log-normalized and depth-normalized gene expression values generated by CAGE-seq and RNA-seq. The R package corrplot was used on the normalized count matrix to generate a correlation matrix across all samples.

Significantly downregulated and upregulated genes were separated based on the sign of their log fold-change value. Ensembl gene IDs were inputted into the PantherGO online tool to check for enrichment of gene ontology terms.<sup>64</sup>

**Differential Enhancer Analysis.** Raw count values representing eRNA expression levels for bidirectional enhancers annotated as intragenic or intronic were exported as a counts table. This counts table was imported into EdgeR for differential expression analysis.<sup>66</sup> Counts were normalized to library size, dispersion estimated, and differential enhancer usage called using glm-models. EdgeR was also used to generate an MDS plot of normalized enhancer counts. Due to the low count numbers associated with eRNA expression, expression values are subject to high levels of variation. EdgeR, like other RNA-seq analysis tools, detects this increased variation and reports higher p-values for calling differential expression. After multiple testing correction, there are too few enhancers surviving for downstream analysis. Therefore, raw p-value cutoffs were used. Enhancers with raw p-values  $\leq 0.05$  were used as input to HOMER findmotifsGenomewide.pl to find *de novo* motif enrichments in differential enhancers (enhancers with raw p-values  $> 0.05$  were used as background sequences).<sup>61</sup>

**Alternative Promoter Usage Analysis.** Unidirectional CAGE clusters overlapping annotated promoters were used as our promoter set. We required that an individual promoter make up at least 1% of total gene counts in all samples to be included in our analysis. A python script was written to count the number of promoters per gene. A python script was also written to calculate the percent usage of each promoter. The percent usage was averaged for the 3 healthy hearts and 4 failed hearts and the difference was calculated. To assess the alternative promoters' effect on gene protein sequence, a custom annotation of all transcripts' start codons was generated using the Ensembl GTF file version 87 (Downloaded May 2016).

**Enhancer Validation with Other Methods.** Enhancer files from published sources were downloaded. To determine overlap with the Vista enhancer browser, enhancers with heart signal, enhancers with any positive signal, and enhancers with no signal were downloaded.<sup>33</sup> For Dickel et al. 2016, the “Putative human heart enhancers identified by integrative analysis” table was used as enhancer predictions.<sup>20</sup> For Spurrell et al. 2019, the “Predicted Enhancers in any 2 Samples” file was used as enhancer predictions. For FANTOM data, the “Ubiquitous enhancers organs” file for enhancer regions active in all organs tested was used. The FANTOM left ventricle and cerebellum enhancer sets were determined by requiring that the enhancer have non-zero expression in each tissue.<sup>35</sup> As a negative control, genomic coordinates of enhancer regions were scrambled 500 times, avoiding placement in repeats or gap sequences. Negative control regions and CAGE sequence defined enhancer regions were intersected with downloaded predictions using bedtools intersect requiring at least 1bp overlap<sup>62</sup>. Significance was determined using a fisher exact test.

**Acknowledgements:** We thank the patients for their participation. We also thank Dr. Xinkun Wang from the Northwestern University sequencing core and Yujiro Takegami from DNAFORM for their excellent technical support. AG conducted the analysis and drafted the manuscript. LDC secured patient consent and genotype information. PP assisted with genotyping. JAW provided access to control samples. DYB and MJP provided helpful advice and commentary and assisted with interpretation. MAN and EMM assisted with analysis, writing and editing the manuscript.

**Sources of funding:** National Institutes of Health NIH HL128075, NIH HL142187, NIH HL141698, American Heart Association AHA 18CDA34110460.

**Disclosures:** The authors have no conflicts of interest related to the content of this work.

## REFERENCES

1. Savarese G and Lund LH. Global Public Health Burden of Heart Failure. *Card Fail Rev.* 2017;3:7-11.
2. Jackson SL, Tong X, King RJ, Loustalot F, Hong Y and Ritchey MD. National Burden of Heart Failure Events in the United States, 2006 to 2014. *Circ Heart Fail.* 2018;11:e004873.
3. Heinig M, Adriaens ME, Schafer S, van Deutekom HWM, Lodder EM, Ware JS, Schneider V, Felkin LE, Creemers EE, Meder B, Katus HA, Ruhle F, Stoll M, Cambien F, Villard E, Charron P, Varro A, Bishopric NH, George AL, Jr., Dos Remedios C, Moreno-Moral A, Pesce F, Bauerfeind A, Ruschendorf F, Rintisch C, Petretto E, Barton PJ, Cook SA, Pinto YM, Bezzina CR and Hubner N. Natural genetic variation of the cardiac transcriptome in non-diseased donors and patients with dilated cardiomyopathy. *Genome Biol.* 2017;18:170.
4. Razeghi P, Young ME, Alcorn JL, Moravec CS, Frazier OH and Taegtmeyer H. Metabolic gene expression in fetal and failing human heart. *Circulation.* 2001;104:2923-31.
5. Taegtmeyer H. Genetics of energetics: transcriptional responses in cardiac metabolism. *Ann Biomed Eng.* 2000;28:871-6.
6. Miyata S, Minobe W, Bristow MR and Leinwand LA. Myosin heavy chain isoform expression in the failing and nonfailing human heart. *Circ Res.* 2000;86:386-90.
7. Yin Z, Ren J and Guo W. Sarcomeric protein isoform transitions in cardiac muscle: a journey to heart failure. *Biochimica et biophysica acta.* 2015;1852:47-52.
8. Anderson PA, Malouf NN, Oakeley AE, Pagani ED and Allen PD. Troponin T isoform expression in humans. A comparison among normal and failing adult heart, fetal heart, and adult and fetal skeletal muscle. *Circ Res.* 1991;69:1226-33.
9. Makarenko I, Opitz CA, Leake MC, Neagoe C, Kulke M, Gwathmey JK, del Monte F, Hajjar RJ and Linke WA. Passive stiffness changes caused by upregulation of compliant titin isoforms in human dilated cardiomyopathy hearts. *Circ Res.* 2004;95:708-16.
10. Beqqali A. Alternative splicing in cardiomyopathy. *Biophys Rev.* 2018;10:1061-1071.
11. Mestroni L, Brun F, Spezzacatene A, Sinagra G and Taylor MR. Genetic Causes of Dilated Cardiomyopathy. *Prog Pediatr Cardiol.* 2014;37:13-18.
12. Konno T, Chang S, Seidman JG and Seidman CE. Genetics of hypertrophic cardiomyopathy. *Curr Opin Cardiol.* 2010;25:205-9.
13. Angrisano T, Schiattarella GG, Keller S, Pironti G, Florio E, Magliulo F, Bottino R, Pero R, Lembo F, Avvedimento EV, Esposito G, Trimarco B, Chiariotti L and Perrino C. Epigenetic switch at atp2a2 and myh7 gene promoters in pressure overload-induced heart failure. *PLoS One.* 2014;9:e106024.
14. Akerberg BN, Gu F, VanDusen NJ, Zhang X, Dong R, Li K, Zhang B, Zhou B, Sethi I, Ma Q, Wasson L, Wen T, Liu J, Dong K, Conlon FL, Zhou J, Yuan GC, Zhou P and Pu WT. A reference map of murine cardiac transcription factor chromatin occupancy identifies dynamic and conserved enhancers. *Nat Commun.* 2019;10:4907.
15. He A, Gu F, Hu Y, Ma Q, Ye LY, Akiyama JA, Visel A, Pennacchio LA and Pu WT. Dynamic GATA4 enhancers shape the chromatin landscape central to heart development and disease. *Nat Commun.* 2014;5:4907.
16. Wamstad JA, Alexander JM, Truty RM, Shrikumar A, Li F, Eilertson KE, Ding H, Wylie JN, Pico AR, Capra JA, Erwin G, Kattman SJ, Keller GM, Srivastava D, Levine SS, Pollard KS, Holloway AK, Boyer LA and Bruneau BG. Dynamic and coordinated epigenetic regulation of developmental transitions in the cardiac lineage. *Cell.* 2012;151:206-20.
17. Paige SL, Thomas S, Stoick-Cooper CL, Wang H, Maves L, Sandstrom R, Pabon L, Reinecke H, Pratt G, Keller G, Moon RT, Stamatoyannopoulos J and Murry CE. A temporal chromatin signature in human embryonic stem cells identifies regulators of cardiac development. *Cell.* 2012;151:221-32.
18. Rosa-Garrido M, Chapski DJ, Schmitt AD, Kimball TH, Karbassi E, Monte E, Balderas E, Pellegrini M, Shih TT, Soehalim E, Liem D, Ping P, Galjart NJ, Ren S, Wang Y, Ren B and

Vondriska TM. High-Resolution Mapping of Chromatin Conformation in Cardiac Myocytes Reveals Structural Remodeling of the Epigenome in Heart Failure. *Circulation*. 2017;136:1613-1625.

19. May D, Blow MJ, Kaplan T, McCulley DJ, Jensen BC, Akiyama JA, Holt A, Plajzer-Frick I, Shoukry M, Wright C, Afzal V, Simpson PC, Rubin EM, Black BL, Bristow J, Pennacchio LA and Visel A. Large-scale discovery of enhancers from human heart tissue. *Nat Genet*. 2011;44:89-93.

20. Dickel DE, Barozzi I, Zhu Y, Fukuda-Yuzawa Y, Osterwalder M, Mannion BJ, May D, Spurrell CH, Plajzer-Frick I, Pickle CS, Lee E, Garvin TH, Kato M, Akiyama JA, Afzal V, Lee AY, Gorkin DU, Ren B, Rubin EM, Visel A and Pennacchio LA. Genome-wide compendium and functional assessment of in vivo heart enhancers. *Nat Commun*. 2016;7:12923.

21. Kimura K, Wakamatsu A, Suzuki Y, Ota T, Nishikawa T, Yamashita R, Yamamoto J, Sekine M, Tsuritani K, Wakaguri H, Ishii S, Sugiyama T, Saito K, Isono Y, Irie R, Kushida N, Yoneyama T, Otsuka R, Kanda K, Yokoi T, Kondo H, Wagatsuma M, Murakawa K, Ishida S, Ishibashi T, Takahashi-Fujii A, Tanase T, Nagai K, Kikuchi H, Nakai K, Isogai T and Sugano S. Diversification of transcriptional modulation: large-scale identification and characterization of putative alternative promoters of human genes. *Genome Res*. 2006;16:55-65.

22. Landry JR, Mager DL and Wilhelm BT. Complex controls: the role of alternative promoters in mammalian genomes. *Trends Genet*. 2003;19:640-8.

23. Zou J, Tran D, Baalbaki M, Tang LF, Poon A, Pelonero A, Titus EW, Yuan C, Shi C, Patchava S, Halper E, Garg J, Movsesyan I, Yin C, Wu R, Wilsbacher LD, Liu J, Hager RL, Coughlin SR, Jinek M, Pullinger CR, Kane JP, Hart DO, Kwok PY and Deo RC. An internal promoter underlies the difference in disease severity between N- and C-terminal truncation mutations of Titin in zebrafish. *Elife*. 2015;4:e09406.

24. Murata M, Nishiyori-Sueki H, Kojima-Ishiyama M, Carninci P, Hayashizaki Y and Itoh M. Detecting expressed genes using CAGE. *Methods Mol Biol*. 2014;1164:67-85.

25. De Santa F, Barozzi I, Mietton F, Ghisletti S, Polletti S, Tusi BK, Muller H, Ragoussis J, Wei CL and Natoli G. A large fraction of extragenic RNA pol II transcription sites overlap enhancers. *PLoS Biol*. 2010;8:e1000384.

26. Kim TK, Hemberg M, Gray JM, Costa AM, Bear DM, Wu J, Harmin DA, Laptewicz M, Barbara-Haley K, Kuersten S, Markenscoff-Papadimitriou E, Kuhl D, Bito H, Worley PF, Kreiman G and Greenberg ME. Widespread transcription at neuronal activity-regulated enhancers. *Nature*. 2010;465:182-7.

27. Benner C, Konovalov S, Mackintosh C, Hutt KR, Stunnenberg R and Garcia-Bassets I. Decoding a signature-based model of transcription cofactor recruitment dictated by cardinal cis-regulatory elements in proximal promoter regions. *PLoS Genet*. 2013;9:e1003906.

28. Schlesinger J, Schueler M, Grunert M, Fischer JJ, Zhang Q, Krueger T, Lange M, Tonjes M, Dunkel I and Sperling SR. The cardiac transcription network modulated by Gata4, Mef2a, Nkx2.5, Srf, histone modifications, and microRNAs. *PLoS Genet*. 2011;7:e1001313.

29. Cruz-Topete D, Oakley RH, Carroll NG, He B, Myers PH, Xu X, Watts MN, Trosclair K, Glasscock E, Dominic P and Cidlowski JA. Deletion of the Cardiomyocyte Glucocorticoid Receptor Leads to Sexually Dimorphic Changes in Cardiac Gene Expression and Progression to Heart Failure. *J Am Heart Assoc*. 2019;8:e011012.

30. Fantom, Riken, Forrest AR, Kawaji H, Rehli M, Baillie JK, de Hoon MJ, Haberle V, Lassmann T, Kulakovskiy IV, Lizio M, Itoh M, Andersson R, Mungall CJ, Meehan TF, Schmeier S, Bertin N, Jorgensen M, Dimont E, Arner E, Schmidl C, Schaefer U, Medvedeva YA, Plessy C, Vitezic M, Severin J, Semple C, Ishizu Y, Young RS, Francescato M, Alam I, Albanese D, Altschuler GM, Arakawa T, Archer JA, Arner P, Babina M, Rennie S, Balwierz PJ, Beckhouse AG, Pradhan-Bhatt S, Blake JA, Blumenthal A, Bodega B, Bonetti A, Briggs J, Brombacher F, Burroughs AM, Califano A, Cannistraci CV, Carbojo D, Chen Y, Chierici M, Ciani Y, Clevers HC, Dalla E, Davis CA, Detmar M, Diehl AD, Dohi T, Drablos F, Edge AS, Edinger M, Ekwall K,

Endoh M, Enomoto H, Fagiolini M, Fairbairn L, Fang H, Farach-Carson MC, Faulkner GJ, Favorov AV, Fisher ME, Frith MC, Fujita R, Fukuda S, Furlanello C, Furino M, Furusawa J, Geijtenbeek TB, Gibson AP, Gingeras T, Goldowitz D, Gough J, Guhl S, Guler R, Gustincich S, Ha TJ, Hamaguchi M, Hara M, Harbers M, Harshbarger J, Hasegawa A, Hasegawa Y, Hashimoto T, Herlyn M, Hitchens KJ, Ho Sui SJ, Hofmann OM, Hoof I, Hori F, Huminiecki L, Iida K, Ikawa T, Jankovic BR, Jia H, Joshi A, Jurman G, Kaczkowski B, Kai C, Kaida K, Kaiho A, Kajiyama K, Kanamori-Katayama M, Kasianov AS, Kasukawa T, Katayama S, Kato S, Kawaguchi S, Kawamoto H, Kawamura YI, Kawashima T, Kempfle JS, Kenna TJ, Kere J, Khachigian LM, Kitamura T, Klinken SP, Knox AJ, Kojima M, Kojima S, Kondo N, Koseki H, Koyasu S, Krampitz S, Kubosaki A, Kwon AT, Laros JF, Lee W, Lennartsson A, Li K, Lilje B, Lipovich L, Mackay-Sim A, Manabe R, Mar JC, Marchand B, Mathelier A, Mejhert N, Meynert A, Mizuno Y, de Lima Morais DA, Morikawa H, Morimoto M, Moro K, Motakis E, Motohashi H, Mummery CL, Murata M, Nagao-Sato S, Nakachi Y, Nakahara F, Nakamura T, Nakamura Y, Nakazato K, van Nimwegen E, Ninomiya N, Nishiyori H, Noma S, Noma S, Noazaki T, Ogishima S, Ohkura N, Ohimiya H, Ohno H, Ohshima M, Okada-Hatakeyama M, Okazaki Y, Orlando V, Ovchinnikov DA, Pain A, Passier R, Patrikakis M, Persson H, Piazza S, Prendergast JG, Rackham OJ, Ramiłowski JA, Rashid M, Ravasi T, Rizzu P, Roncador M, Roy S, Rye MB, Sajjyo E, Sajantila A, Saka A, Sakaguchi S, Sakai M, Sato H, Savvi S, Saxena A, Schneider C, Schultes EA, Schulze-Tanzil GG, Schwegmann A, Sengstag T, Sheng G, Shimoji H, Shimoni Y, Shin JW, Simon C, Sugiyama D, Sugiyama T, Suzuki M, Suzuki N, Swoboda RK, t Hoen PA, Tagami M, Takahashi N, Takai J, Tanaka H, Tatsukawa H, Tatum Z, Thompson M, Toyodo H, Toyoda T, Valen E, van de Wetering M, van den Berg LM, Verado R, Vijayan D, Vorontsov IE, Wasserman WW, Watanabe S, Wells CA, Winteringham LN, Wolvetang E, Wood EJ, Yamaguchi Y, Yamamoto M, Yoneda M, Yonekura Y, Yoshida S, Zabierowski SE, Zhang PG, Zhao X, Zucchelli S, Summers KM, Suzuki H, Daub CO, Kawai J, Heutink P, Hide W, Freeman TC, Lenhard B, Bajic VB, Taylor MS, Makeev VJ, Sandelin A, Hume DA, Carninci P and Hayashizaki Y. A promoter-level mammalian expression atlas. *Nature*. 2014;507:462-70.

31. Carninci P, Sandelin A, Lenhard B, Katayama S, Shimokawa K, Ponjavic J, Semple CA, Taylor MS, Engstrom PG, Frith MC, Forrest AR, Alkema WB, Tan SL, Plessy C, Kodzius R, Ravasi T, Kasukawa T, Fukuda S, Kanamori-Katayama M, Kitazume Y, Kawaji H, Kai C, Nakamura M, Konno H, Nakano K, Mottagui-Tabar S, Arner P, Chesi A, Gustincich S, Persichetti F, Suzuki H, Grimmond SM, Wells CA, Orlando V, Wahlestedt C, Liu ET, Harbers M, Kawai J, Bajic VB, Hume DA and Hayashizaki Y. Genome-wide analysis of mammalian promoter architecture and evolution. *Nat Genet*. 2006;38:626-35.

32. Sakharkar MK, Chow VT and Kanguane P. Distributions of exons and introns in the human genome. *In Silico Biol*. 2004;4:387-93.

33. Visel A, Minovitsky S, Dubchak I and Pennacchio LA. VISTA Enhancer Browser--a database of tissue-specific human enhancers. *Nucleic Acids Res*. 2007;35:D88-92.

34. Spurrell CH, Barozzi I, Mannion BJ, Blow MJ, Fukuda-Yuzawa Y, Afzal SY, Akiyama JA, Afzal V, Tran S, Plajzer-Frick I, Novak CS, Kato M, Lee E, Garvin TH, Pham QT, Harrington AN, Lisgo S, Bristow J, Cappola TP, Morley MP, Margulies KB, Pennacchio LA, Dickel DE and Visel A. Genome-Wide Fetalization of Enhancer Architecture in Heart Disease. *bioRxiv*. 2019:591362.

35. Andersson R, Gebhard C, Miguel-Escalada I, Hoof I, Bornholdt J, Boyd M, Chen Y, Zhao X, Schmidl C, Suzuki T, Ntini E, Arner E, Valen E, Li K, Schwarzfischer L, Glatz D, Raithel J, Lilje B, Rapin N, Bagger FO, Jorgensen M, Andersen PR, Bertin N, Rackham O, Burroughs AM, Baillie JK, Ishizu Y, Shimizu Y, Furuhata E, Maeda S, Negishi Y, Mungall CJ, Meehan TF, Lassmann T, Itoh M, Kawaji H, Kondo N, Kawai J, Lennartsson A, Daub CO, Heutink P, Hume DA, Jensen TH, Suzuki H, Hayashizaki Y, Muller F, Forrest ARR, Carninci P, Rehli M and Sandelin A. An atlas of active enhancers across human cell types and tissues. *Nature*. 2014;507:455-461.

36. Porto AG, Brun F, Severini GM, Losurdo P, Fabris E, Taylor MRG, Mestroni L and Sinagra G. Clinical Spectrum of PRKAG2 Syndrome. *Circ Arrhythm Electrophysiol*. 2016;9:e003121.
37. Banankhah P, Fishbein GA, Dota A and Ardehali R. Cardiac manifestations of PRKAG2 mutation. *BMC Med Genet*. 2018;19:1.
38. Ben Jehuda R, Eisen B, Shemer Y, Mekies LN, Szantai A, Reiter I, Cui H, Guan K, Haron-Khun S, Freimark D, Sperling SR, Gherghiceanu M, Arad M and Binah O. CRISPR correction of the PRKAG2 gene mutation in the patient's induced pluripotent stem cell-derived cardiomyocytes eliminates electrophysiological and structural abnormalities. *Heart Rhythm*. 2018;15:267-276.
39. Ortega A, Rosello-Lleti E, Tarazon E, Gil-Cayuela C, Lago F, Gonzalez-Juanatey JR, Martinez-Dolz L, Portoles M and Rivera M. TRPM7 is down-regulated in both left atria and left ventricle of ischaemic cardiomyopathy patients and highly related to changes in ventricular function. *ESC Heart Fail*. 2016;3:220-224.
40. Sah R, Mesirca P, Van den Boogert M, Rosen J, Mably J, Mangoni ME and Clapham DE. Ion channel-kinase TRPM7 is required for maintaining cardiac automaticity. *Proc Natl Acad Sci U S A*. 2013;110:E3037-46.
41. Prasher D, Greenway SC and Singh RB. The Impact of Epigenetics on Cardiovascular Disease. *Biochem Cell Biol*. 2019.
42. Kohli S, Ahuja S and Rani V. Transcription factors in heart: promising therapeutic targets in cardiac hypertrophy. *Curr Cardiol Rev*. 2011;7:262-71.
43. Kim TH, Barrera LO, Zheng M, Qu C, Singer MA, Richmond TA, Wu Y, Green RD and Ren B. A high-resolution map of active promoters in the human genome. *Nature*. 2005;436:876-80.
44. Cooper SJ, Trinklein ND, Anton ED, Nguyen L and Myers RM. Comprehensive analysis of transcriptional promoter structure and function in 1% of the human genome. *Genome Res*. 2006;16:1-10.
45. Davuluri RV, Suzuki Y, Sugano S, Plass C and Huang TH. The functional consequences of alternative promoter use in mammalian genomes. *Trends Genet*. 2008;24:167-77.
46. Pinter K, Grignani RT, Czibik G, Farza H, Watkins H and Redwood C. Embryonic expression of AMPK gamma subunits and the identification of a novel gamma2 transcript variant in adult heart. *J Mol Cell Cardiol*. 2012;53:342-9.
47. Oliveira SM, Zhang YH, Solis RS, Isackson H, Bellahcene M, Yavari A, Pinter K, Davies JK, Ge Y, Ashrafi H, Walker JW, Carling D, Watkins H, Casadei B and Redwood C. AMP-activated protein kinase phosphorylates cardiac troponin I and alters contractility of murine ventricular myocytes. *Circ Res*. 2012;110:1192-201.
48. Khandekar M, Brandt W, Zhou Y, Dagenais S, Glover TW, Suzuki N, Shimizu R, Yamamoto M, Lim KC and Engel JD. A Gata2 intronic enhancer confers its pan-endothelia-specific regulation. *Development*. 2007;134:1703-12.
49. Yoshida T and Delafontaine P. An Intronic Enhancer Element Regulates Angiotensin II Type 2 Receptor Expression during Satellite Cell Differentiation, and Its Activity Is Suppressed in Congestive Heart Failure. *J Biol Chem*. 2016;291:25578-25590.
50. Rigau M, Juan D, Valencia A and Rico D. Intronic CNVs and gene expression variation in human populations. *PLoS Genet*. 2019;15:e1007902.
51. Schaper J, Kostin S, Hein S, Elsasser A, Arnon E and Zimmermann R. Structural remodelling in heart failure. *Exp Clin Cardiol*. 2002;7:64-8.
52. Khalil H, Kanisicak O, Prasad V, Correll RN, Fu X, Schips T, Vagnozzi RJ, Liu R, Huynh T, Lee SJ, Karch J and Molkentin JD. Fibroblast-specific TGF-beta-Smad2/3 signaling underlies cardiac fibrosis. *J Clin Invest*. 2017;127:3770-3783.
53. Chen H, Moreno-Moral A, Pesce F, Devapragash N, Mancini M, Heng EL, Rotival M, Srivastava PK, Harmston N, Shkura K, Rackham OJL, Yu WP, Sun XM, Tee NGZ, Tan ELS,

Barton PJR, Felkin LE, Lara-Pezzi E, Angelini G, Beltrami C, Pravenec M, Schafer S, Bottolo L, Hubner N, Emanueli C, Cook SA and Petretto E. WWP2 regulates pathological cardiac fibrosis by modulating SMAD2 signaling. *Nat Commun.* 2019;10:3616.

54. Rios FJ, Zou ZG, Harvey AP, Harvey KY, Nosalski R, Anyfanti P, Camargo LL, Lacchini S, Ryazanov AG, Ryazanova L, McGrath S, Guzik TJ, Goodyear CS, Montezano AC and Touyz RM. Chanzyme TRPM7 protects against cardiovascular inflammation and fibrosis. *Cardiovasc Res.* 2019.

55. Parajuli N, Valtuille L, Basu R, Famulski KS, Halloran PF, Sergi C and Oudit GY. Determinants of ventricular arrhythmias in human explanted hearts with dilated cardiomyopathy. *Eur J Clin Invest.* 2015;45:1286-96.

56. Haggerty CM, Damrauer SM, Levin MG, Birtwell D, Carey DJ, Golden AM, Hartzel DN, Hu Y, Judy R, Kelly MA, Kember RL, Lester Kirchner H, Leader JB, Liang L, McDermott-Roe C, Babu A, Morley M, Nealy Z, Person TN, Pulenthiran A, Small A, Smelser DT, Stahl RC, Sturm AC, Williams H, Baras A, Margulies KB, Cappola TP, Dewey FE, Verma A, Zhang X, Correa A, Hall ME, Wilson JG, Ritchie MD, Rader DJ, Murray MF, Fornwalt BK and Arany Z. Genomics-First Evaluation of Heart Disease Associated With Titin-Truncating Variants. *Circulation.* 2019;140:42-54.

57. Roberts AM, Ware JS, Herman DS, Schafer S, Baksi J, Bick AG, Buchan RJ, Walsh R, John S, Wilkinson S, Mazzarotto F, Felkin LE, Gong S, MacArthur JA, Cunningham F, Flannick J, Gabriel SB, Altshuler DM, Macdonald PS, Heinig M, Keogh AM, Hayward CS, Banner NR, Pennell DJ, O'Regan DP, San TR, de Marvao A, Dawes TJ, Gulati A, Birks EJ, Yacoub MH, Radke M, Gotthardt M, Wilson JG, O'Donnell CJ, Prasad SK, Barton PJ, Fatkin D, Hubner N, Seidman JG, Seidman CE and Cook SA. Integrated allelic, transcriptional, and phenomic dissection of the cardiac effects of titin truncations in health and disease. *Sci Transl Med.* 2015;7:270ra6.

58. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M and Gingeras TR. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics.* 2013;29:15-21.

59. Haberle V, Forrest AR, Hayashizaki Y, Carninci P and Lenhard B. CAGEr: precise TSS data retrieval and high-resolution promoterome mining for integrative analyses. *Nucleic Acids Res.* 2015;43:e51.

60. Thodberg MT, A..Vitting-Seerup, K.. Andersson, R.. Sandelin, A. CAGEfightR: Cap Analysis of Gene Expression (CAGE) in R/Bioconductor. *bioRxiv.* 2018.

61. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H and Glass CK. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell.* 2010;38:576-89.

62. Quinlan AR and Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics.* 2010;26:841-2.

63. Crooks GE, Hon G, Chandonia JM and Brenner SE. WebLogo: a sequence logo generator. *Genome Res.* 2004;14:1188-90.

64. Thomas PD, Campbell MJ, Kejariwal A, Mi H, Karlak B, Daverman R, Diemer K, Muruganujan A and Narechania A. PANTHER: a library of protein families and subfamilies indexed by function. *Genome Res.* 2003;13:2129-41.

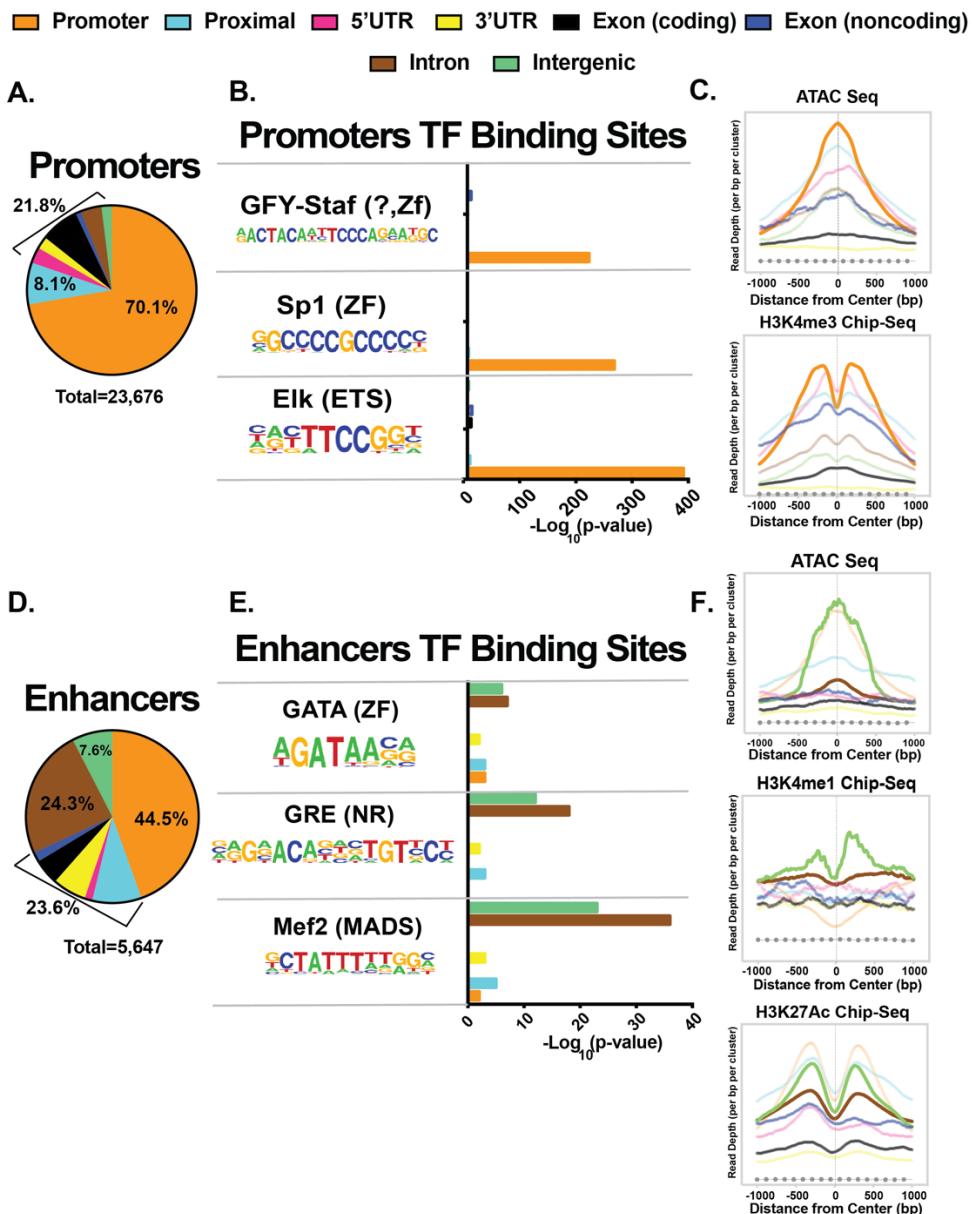
65. Anders S, Pyl PT and Huber W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics.* 2015;31:166-9.

66. Robinson MD, McCarthy DJ and Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics.* 2010;26:139-40.

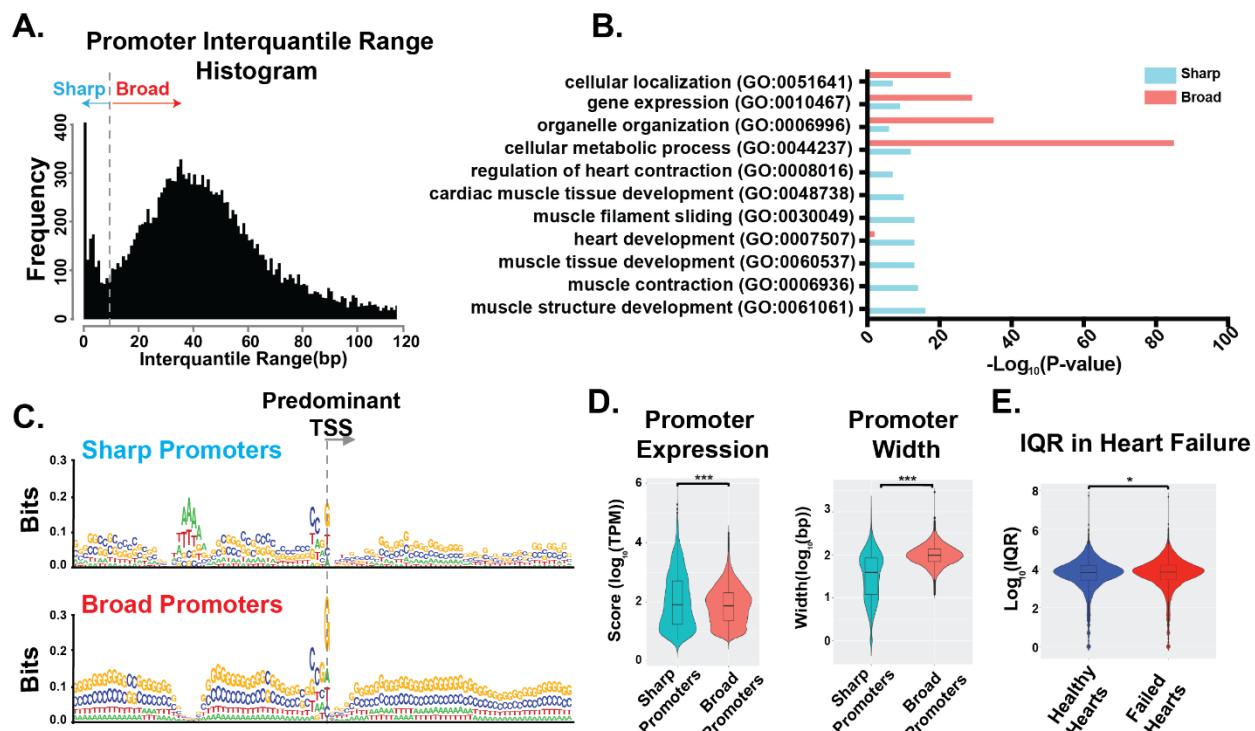
**TABLE 1. Left Ventricle Tissue Source Demographics, Phenotypes, and Mutations.**

Sample	Primary phenotype	Additional phenotypes	Sex	Race	Age	Primary gene mutation(s)
Healthy 1	Healthy	-	M	Caucasian	62	N/A
Healthy 2	Healthy	-	M	Caucasian	47	N/A
Healthy 3	Healthy	-	F	Caucasian	76	N/A
Heart Failure 1	Cardiomyopathy	Ventricular Tachycardia	M	Caucasian	20	<i>TPM1</i> D230N
Heart Failure 2	Cardiomyopathy	-	M	Hispanic	16	<i>TTN</i> c.42521-5 C>G, <i>TNNT2</i> K210del
Heart Failure 3	Cardiomyopathy	Becker Muscular Dystrophy	M	Caucasian	54	<i>DMD</i> IVS +1 G>T
Heart Failure 4	Cardiomyopathy	Limb Girdle Muscular Dystrophy	F	Caucasian	26	<i>LMNA</i> c.1142-1157+1del17

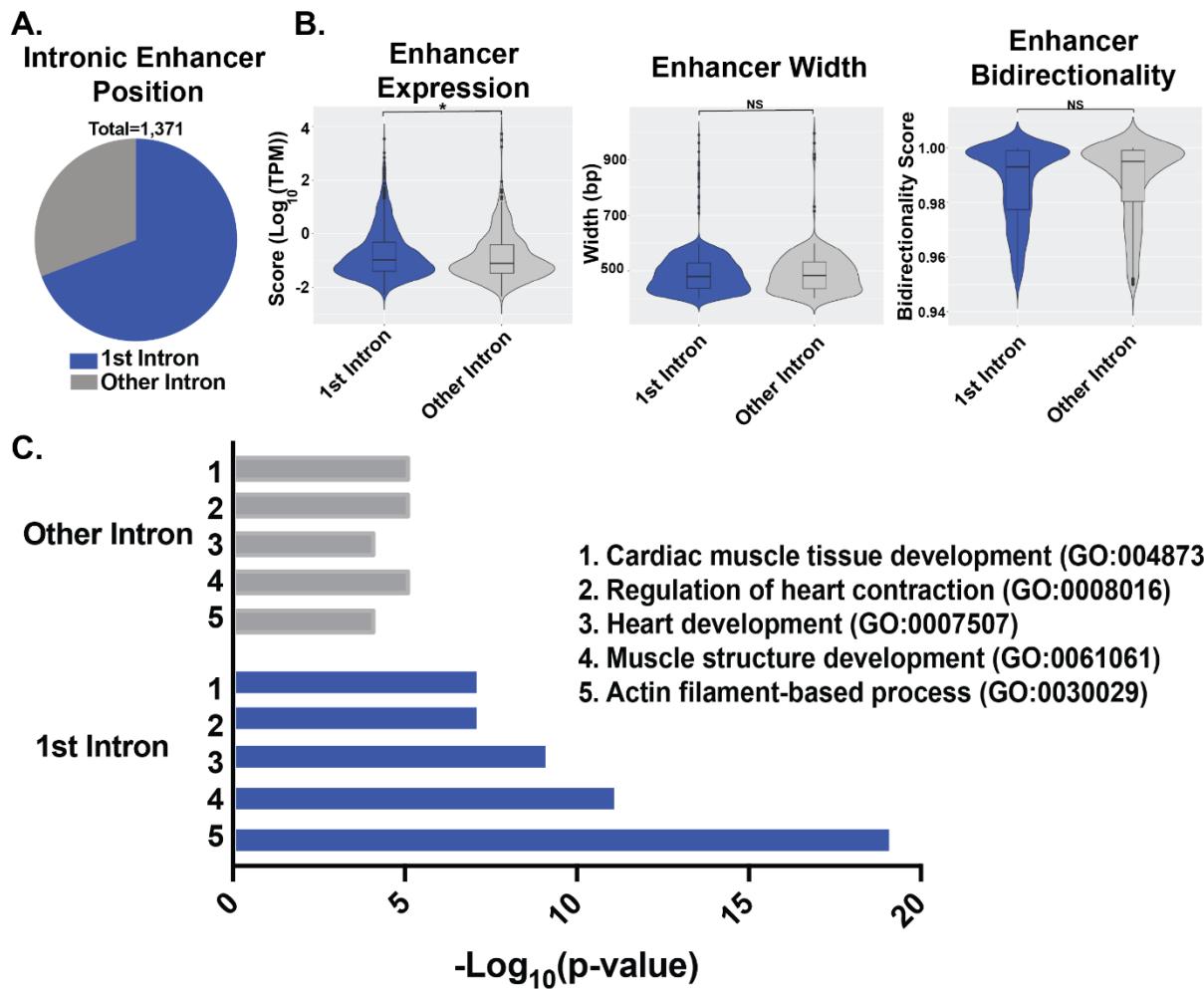
## FIGURES AND LEGENDS



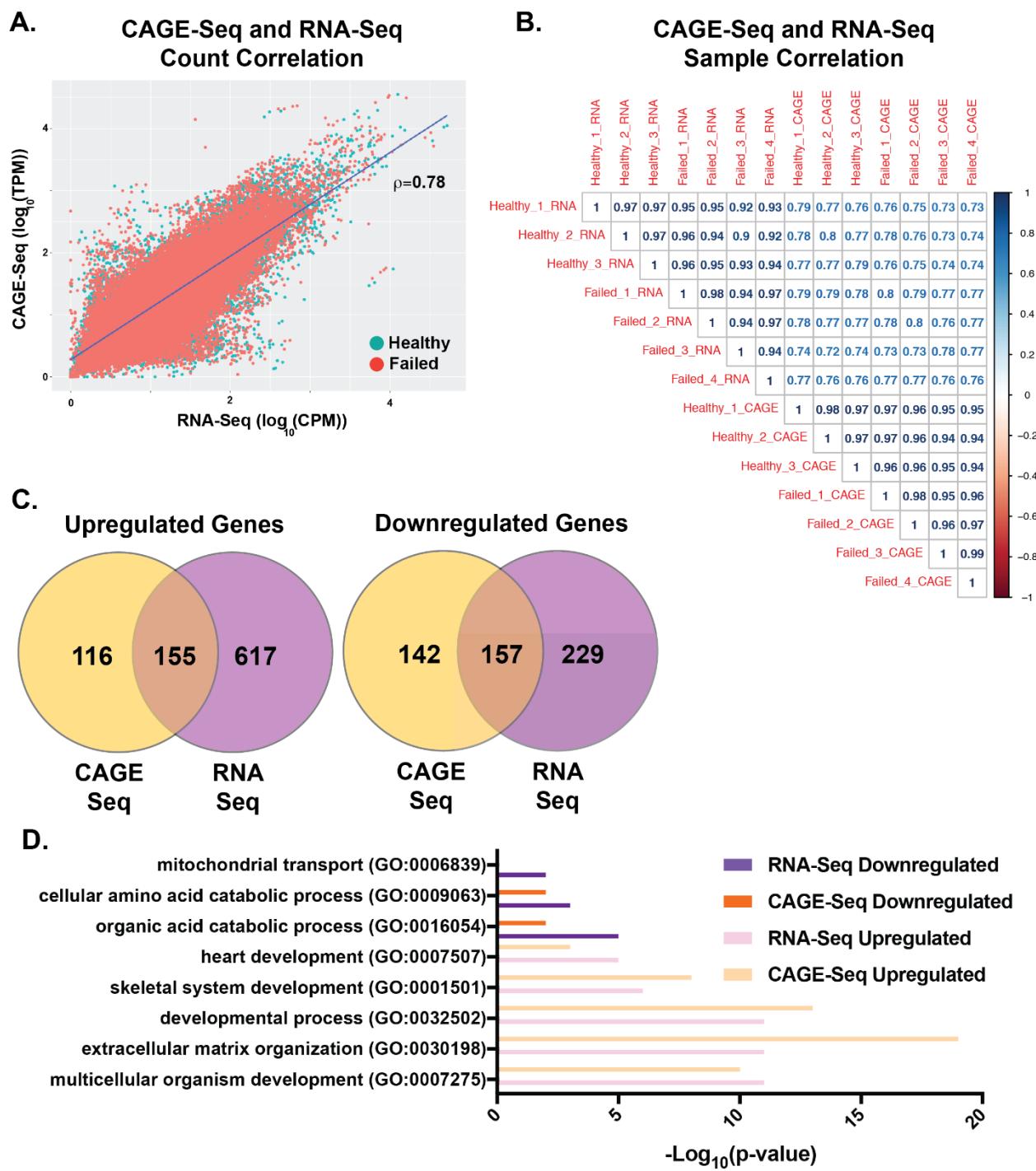
**Figure 1. CAGE sequencing identified promoters and enhancers active in human left ventricular tissue.** **A.** Shown is the distribution annotation classes of unidirectional CAGE clusters. **B.** Enrichment values for three promoter-bound transcription factors in unidirectional clusters from all annotation classes. **C.** Left ventricle signals of open chromatin (ATAC-seq) and a promoter-associated histone mark (H3K4me3) across unidirectional clusters from all annotation classes. **D.** Venn diagram indicating the distribution annotation classes of bidirectional CAGE clusters. **E.** Enrichment values for three cardiac transcription factors in bidirectional clusters from all annotation classes. **F.** Left ventricle signals of open chromatin (ATAC-seq) and two enhancer-associated histone marks (H3K4me1 and H3K27Ac) across bidirectional clusters from all annotation classes. Dashed lines in **C** and **F** represent signals from genomic regions created by scrambling the location of unidirectional and bidirectional clusters, respectively. *TF*, transcription factor.



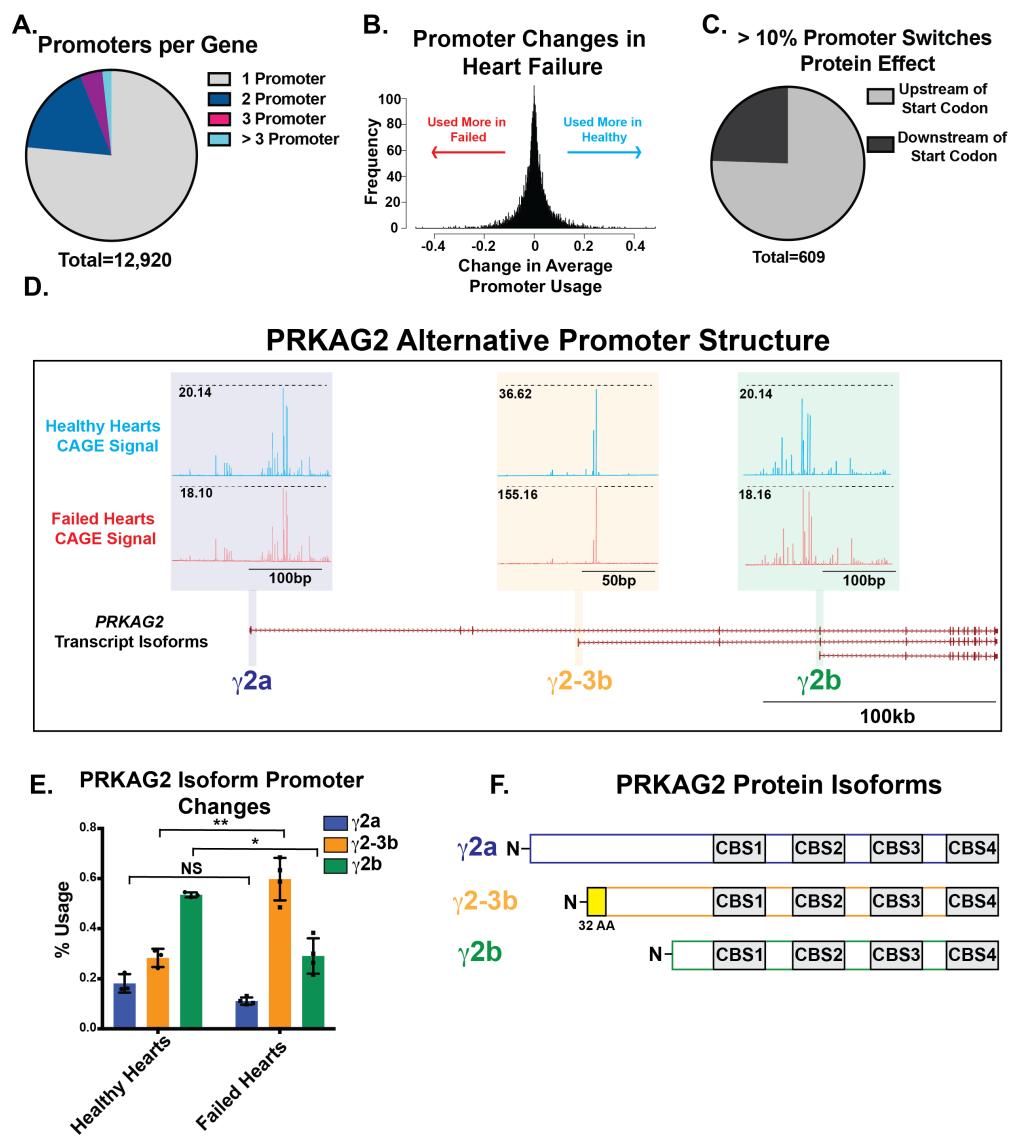
**Figure 2. Sharp and broad transcriptional start sites in cardiac promoters.** **A.** Histogram of interquartile range (number of basepairs between 10% and 90% of a total signal from a given promoter) of all promoters revealed sharp and broad promoter classes. **B.** Gene ontology analysis of genes driven by sharp or broad promoters indicates broad promoters have housekeeping functions while sharp promoters are found across all gene ontology categories including tissue specific genes. **C.** Relative nucleotide compositions of the upstream and downstream sequences from the transcriptional start site for sharp and broad promoters. **D.** Violin plots comparing the expression level and width of sharp and broad promoters. **E.** Violin plot comparing the interquartile range of promoters in healthy and failed hearts. Significance determined by two-tailed nonparametric Mann Whitney Test ( $p \leq 0.05$ (\*),  $\leq 0.005$ (\*\*),  $\leq 0.0005$ (\*\*\*)). TSS, transcriptional start site. IQR, interquartile range. bp, basepair.



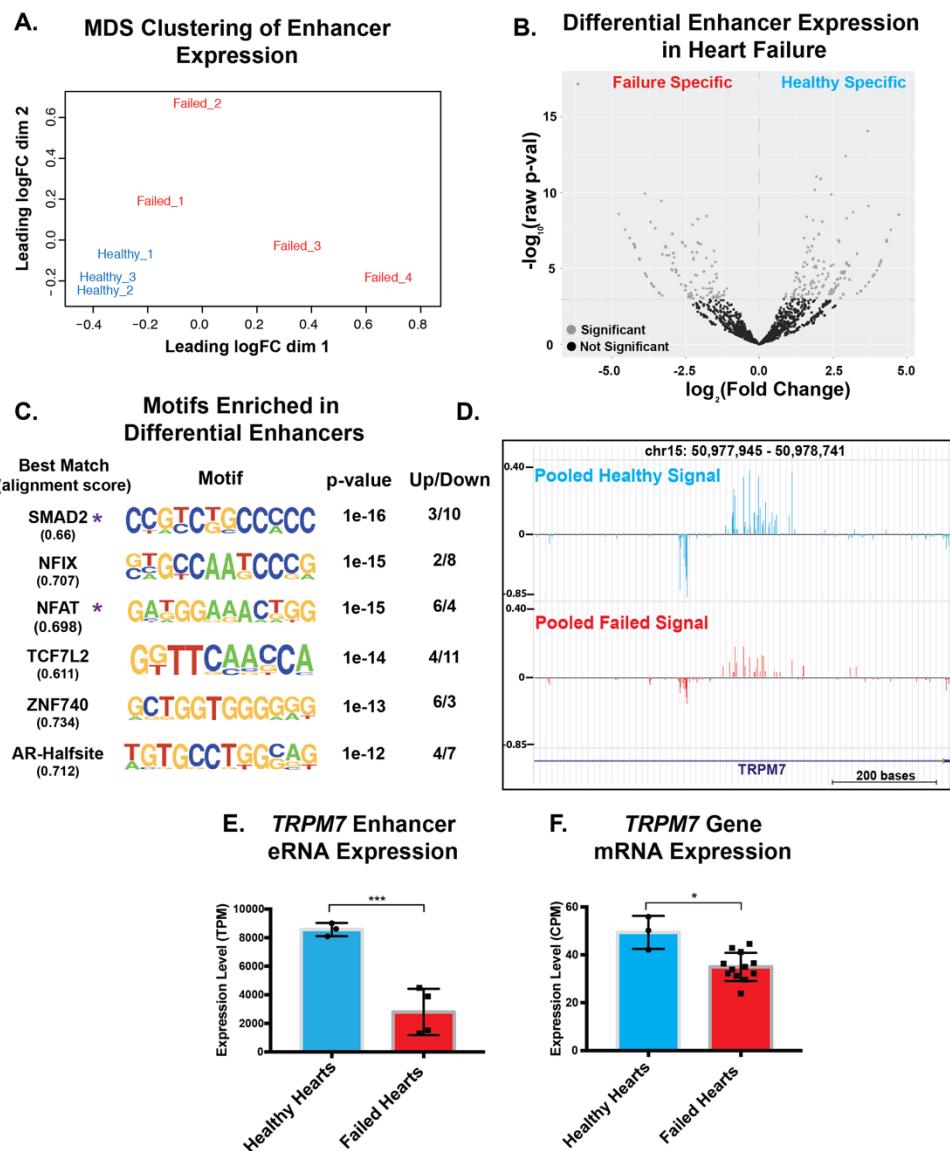
**Figure 3. The majority of intronic cardiac enhancers localize to the first intron.** **A.** Venn diagram demonstrating that ~70% of intronic enhancers were within the first intron of an overlapping transcript. **B.** Violin plots comparing enhancer expression levels, enhancer width, and enhancer bidirectionality score between enhancers in the first intron and enhancers in other introns. **C.** Gene ontology analysis of genes with first intron enhancers and genes with other intron enhancers indicates tissue specific genes are more likely to have first intron enhancers. Significance determined by two-tailed nonparametric Mann Whitney Test ( $p \leq 0.05(*)$ ,  $\leq 0.005(**)$ ,  $\leq 0.0005 (***)$ ).



**Figure 4. Comparison of CAGE-seq and RNA-seq gene expression levels.** **A.** Scatter plot of CAGE-seq gene expression values versus RNA-seq expression values for healthy (teal) and failed (red) samples demonstrating tight correlation. **B.** Sample level correlation matrix of Spearman's correlation coefficient of genome-wide gene expression levels. **C.** Venn diagrams displaying the number of differentially upregulated and downregulated genes determined by CAGE-seq and RNA-seq. **D.** Gene ontology analysis of genes identified as differentially upregulated or downregulated by CAGE-seq and RNA seq.



**Figure 5. Heart failure was associated with significant changes in left ventricle promoter usage.** **A.** Venn diagram of the number of left ventricle promoters per gene determined by CAGE-seq. **B.** Histogram displaying the distribution of average promoter percent usage changes in heart failure. The x-axis represents the difference between a promoter's average percent usage in three healthy left ventricles and four failed left ventricles. Left-shifted promoters make up a greater percentage of gene expression in failed ventricles and right-shifted promoters are a greater percentage in healthy ventricles. **C.** Venn diagram of the relationship to an overlapping transcript's start codon for promoters that undergo a  $\geq 10\%$  shift in heart failure. **D.** Genome browser representation of the alternative promoter structure of the PRKAG2 gene. Three known isoforms of PRKAG2 are represented at the bottom. Above the promoter of each transcript, the CAGE-seq signal for healthy (blue) and failed (red) hearts is shown and the scales of each representation are indicated in black. **E.** Quantification of the CAGE-seq signals shown in D indicating the promoter percent usage of each isoform in healthy and failed hearts. Significance determined by a two-tailed Student's t-test. **F.** Schematic of the predicted amino acid sequences translated from each PRKAG2 isoform. ( $p \leq 0.05$ (\*),  $\leq 0.005$ (\*\*),  $\leq 0.0005$  (\*\*\*)). PRKAG2, Protein Kinase AMP-Activated Non-Catalytic Subunit Gamma.



**Figure 6. CAGE-Seq detected differential left ventricle enhancer usage in heart failure. A.** Multidimensional scaling plot of enhancer expression levels in healthy and failed left ventricle samples. **B.** Volcano plot indicating the results of differential enhancer usage analysis. Differential enhancers are in light gray. Left shifted enhancers are expressed higher in failed hearts and right shifted enhancers are expressed higher in healthy hearts. **C.** *De novo* transcription factor motif enrichment analysis comparing differentially changed enhancers to unchanged enhancers. The best match of enriched motifs is listed to the left. A purple asterisk indicates that the matching transcription factor was differentially expressed by RNA-seq. Up/down indicates the instances of the identified motif in upregulated and downregulated enhancers, respectively. **D.** Genome browser representation of a differentially expressed enhancer within the first intron of the *TRPM7* gene. The gene annotation is at the bottom and the healthy and failed CAGE-seq signals are graphed above on the same scale. **E.** Quantification of the healthy and failed CAGE-seq signals for the intronic *TRPM7* enhancer in D. **F.** Quantification of *TRPM7* overall gene expression by RNA-seq. Significance determined by EdgeR using a generalized linear model approach. ( $p \leq 0.05$ (\*),  $\leq 0.005$ (\*\*),  $\leq 0.0005$  (\*\*\*)). FC, fold change. *TRPM7*, transient receptor potential cation channel subfamily M member 7.

## SUPPLEMENTAL MATERIAL for Gacita et al.

**Supplementary Table 1.** Sequencing Read Yields and Mapping Rates for CAGE-seq and RNA-seq Libraries.

Sample	pM	CAGE-Seq		RNA-Seq	
		Total reads (75bp SE)	Uniquely Aligned Reads (%)	Total Reads (150bp PE)	Uniquely Aligned Reads (%)
Healthy 1	8.10	59,865,536	45,191,501 (75.49)	53,274,910	41,113,982 (77.17)
Healthy 2	7.34	56,226,409	44,062,633 (78.37)	58,609,578	44,742,094 (76.34)
Healthy 3	5.07	32,763,023	25,314,481 (77.27)	49,620,253	38,276,453 (77.14)
Heart Failure 1	5.30	39,295,214	30,906,318 (78.65)	49,587,645	38,077,307 (76.79)
Heart Failure 2	8.25	45,584,689	36,344,157 (79.09)	43,147,599	32,813,912 (76.05)
Heart Failure 3	8.03	46,398,662	35,734,514 (77.02)	47,529,437	35,808,836 (75.34)
Heart Failure 4	4.53	22,668,805	18,101,857 (79.85)	40,737,687	31,579,044 (77.52)

*pM*, picomolar. *SE*, single end. *PE*, paired end

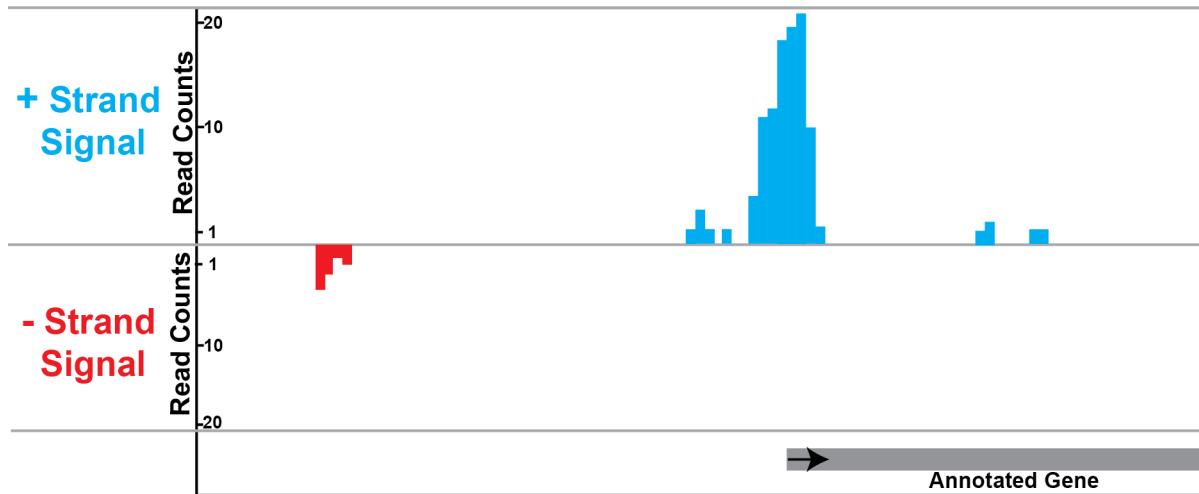
**Supplementary Table 2.** Epigenetic Datasets used for CAGE-Cluster Functional Annotation

Target	Assay Type	Tissue Source	Accession #
Open Chromatin	ATAC-Seq	Female adult (51 years) LV Tissue	ENCSR117PYB
Open Chromatin	DNase-Seq	Female adult (53 years) LV Tissue	ENCFF702IJE
H3K4me1	ChIP-Seq	Male child (3 years) LV Tissue	ENCFF901JPP
H3K4me3	ChIP-Seq	Male Adult (34 years) LV Tissue	ENCFF527LGE
H3K27Ac	ChIP-Seq	Female Adult (51 years) LV Tissue	ENCFF625XET
CTCF	ChIP-Seq	Female Adult (53 years) LV tissue	ENCFF738KRH
POL2A	ChIP-Seq	Female Adult (53 years) LV tissue	ENCFF318MWF

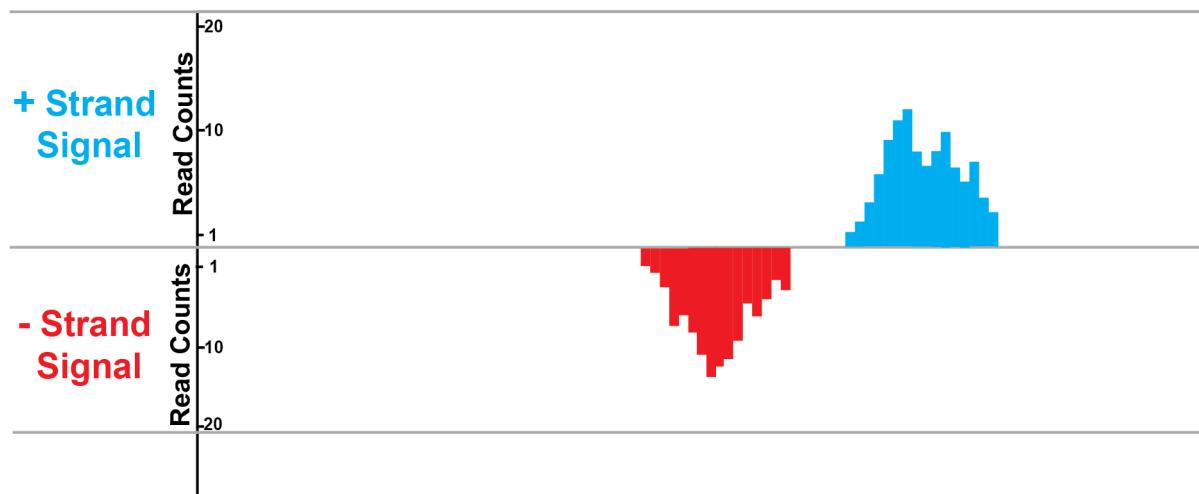
ATAC, assay for transposase-accessible chromatin. ChIP, chromatin immunoprecipitation.

## Supplementary Figures

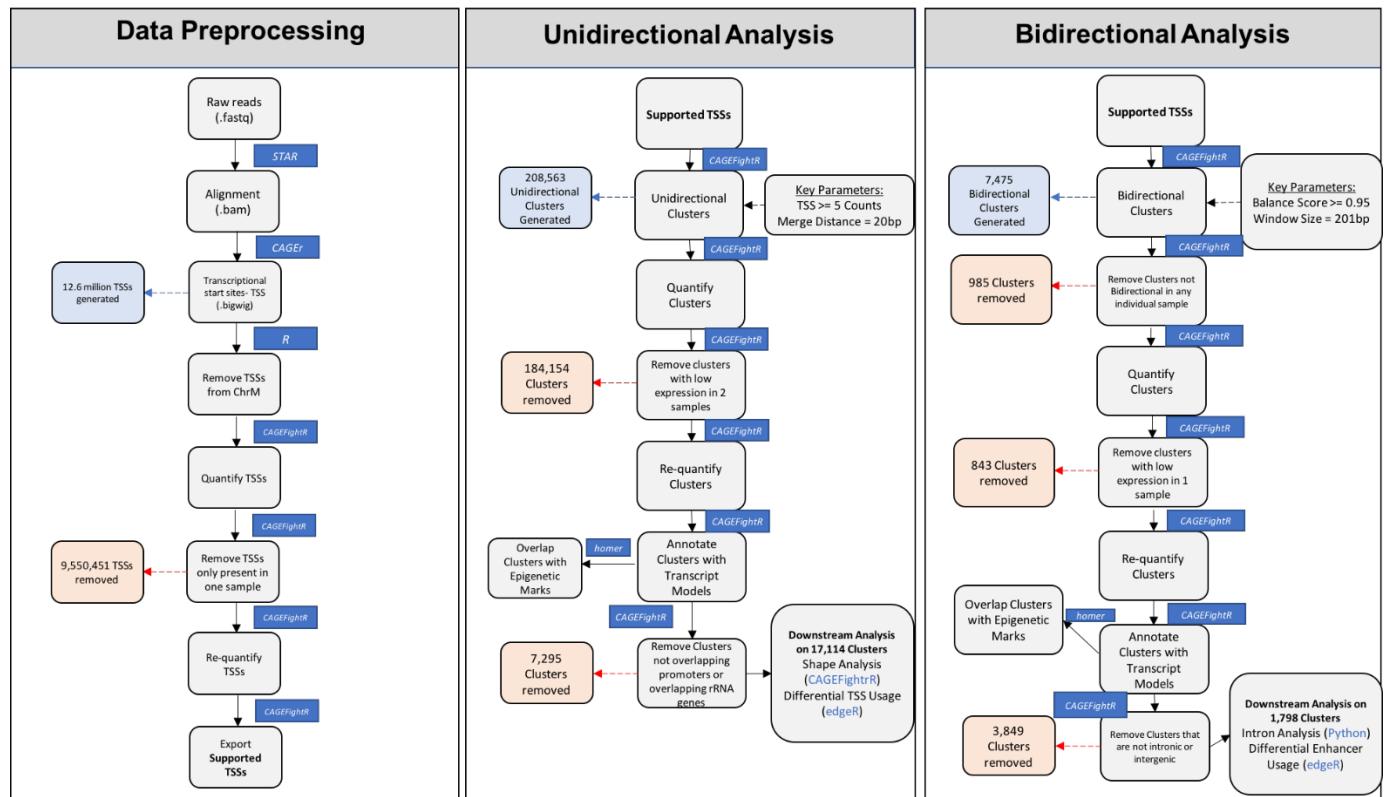
### Unidirectional Cluster (Promoter) Signal



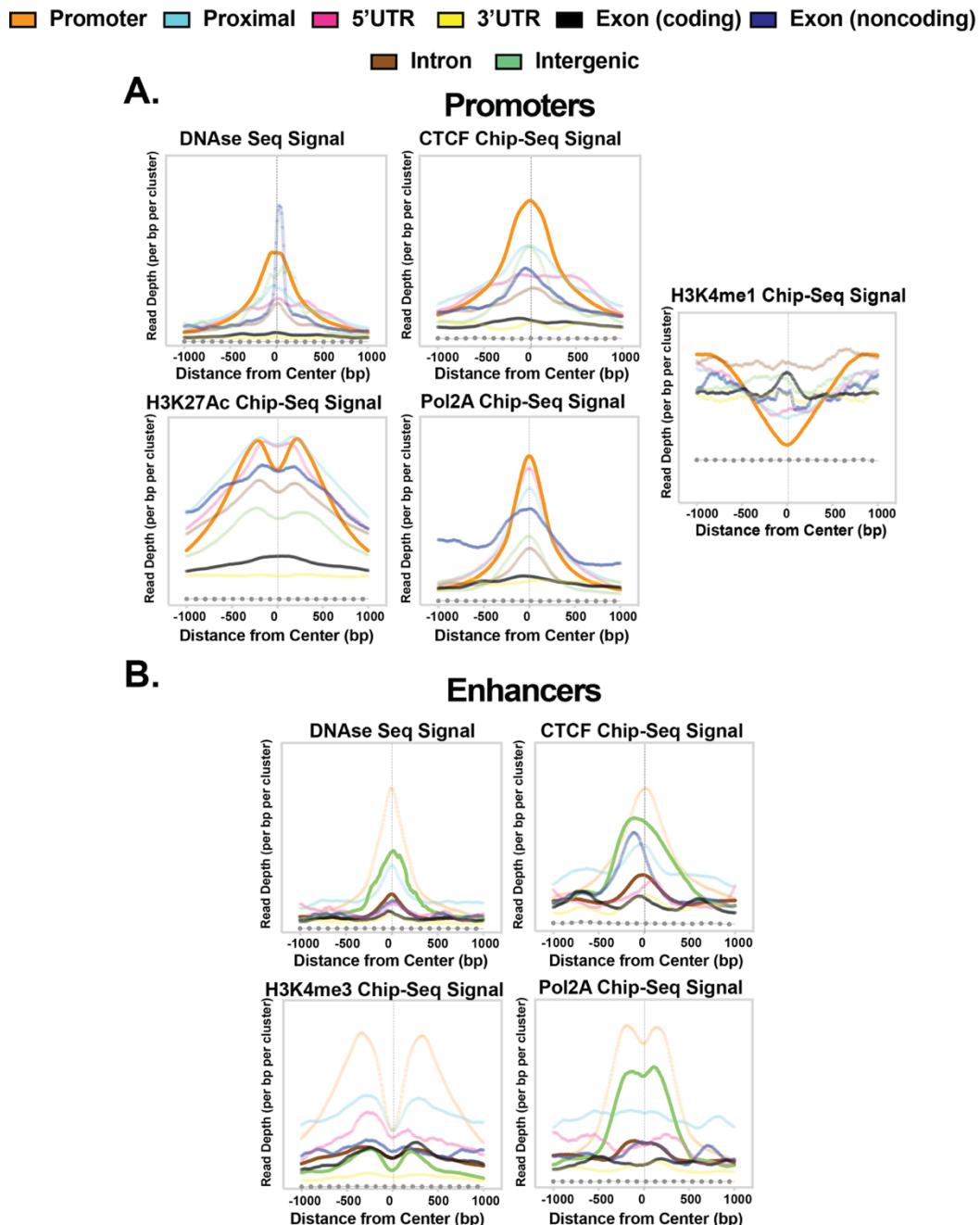
### Bidirectional Cluster (Enhancer) Signal



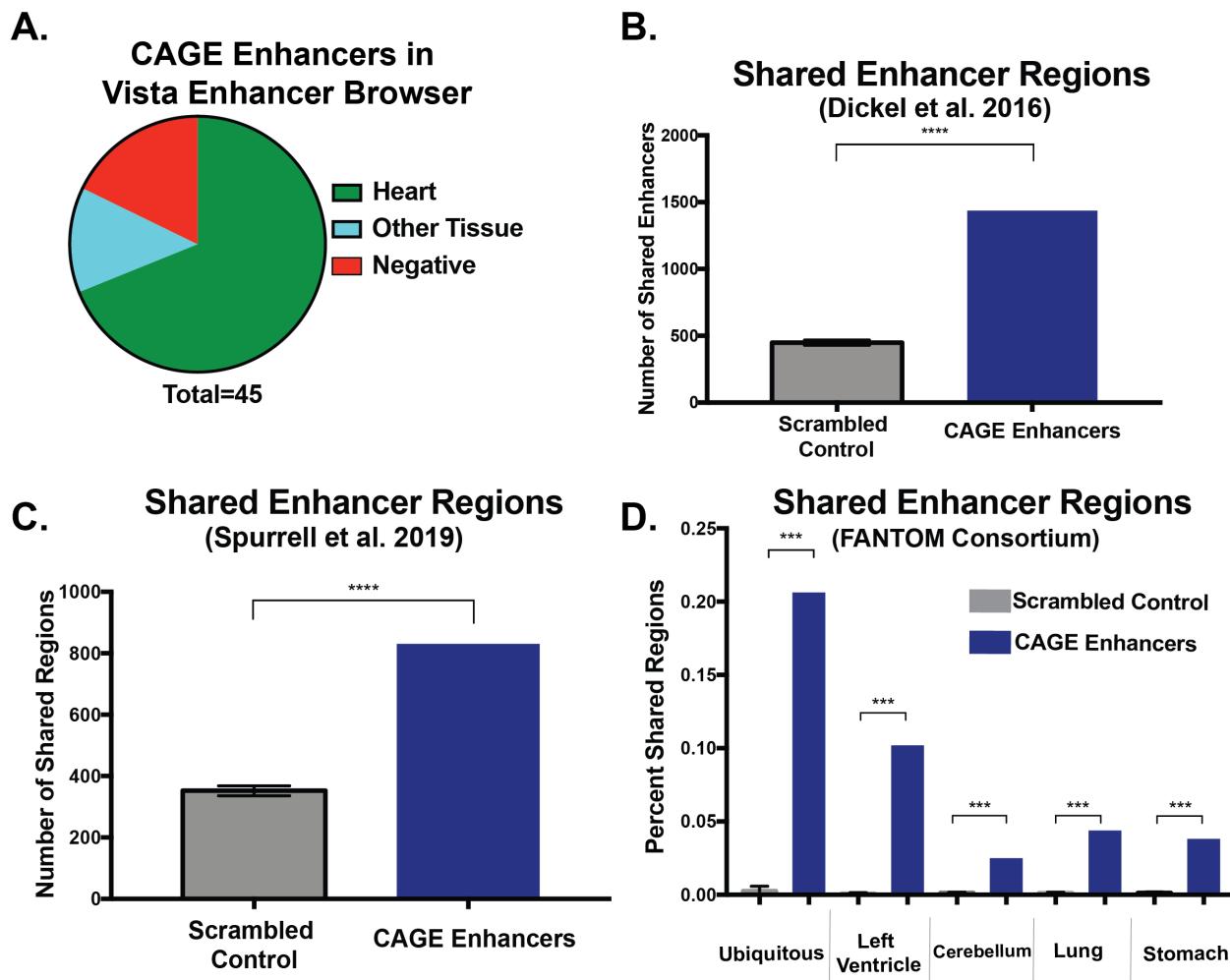
**Supplementary Figure 1.** Example schematic of unidirectional (top) and bidirectional (bottom) CAGE clusters representing promoters and enhancer regions, respectively. Positive (sense) strand signals are shown in blue and minus (antisense) signals are shown in red.



**Supplementary Figure 2.** Data analysis pipeline for identifying and analyzing promoters and enhancers from CAGE-seq data. TSS, transcriptional start sites.



**Supplementary Figure 3.** Additional left ventricle epigenetic signals of unidirectional and bidirectional CAGE clusters. **A.** Left ventricle open chromatin (DNAse-seq), protein binding (CTCF, Pol2A), and enhancer histone marks (H3K27Ac and H3K4me1) signals for unidirectional CAGE clusters of all annotation classes. **B.** Left ventricle open chromatin (DNAse-seq), protein binding (CTCF, Pol2A), and promoter histone marks (H3K4me3) signals for bidirectional CAGE clusters. Dashed lines in A and B represent signals from genomic regions created by scrambling the location of unidirectional and bidirectional clusters, respectively.



**Supplementary Figure 4.** CAGE-defined enhancers overlapped enhancers determined by independent methods. **A.** Venn diagram of Vista Enhancer Browser data displaying the results of functional testing of 45 CAGE enhancers. **B & C.** Bar charts showing the number of overlapping enhancers when comparing CAGE and two independent methods. **D.** Bar charts indicating the percentage of FANTOM enhancers that overlap CAGE enhancers for five different groups of FANTOM enhancers. Scrambled controls represent the number of overlaps obtained when randomly shuffling the genomic location of CAGE enhancers. Significance determined by fisher's exact test. ( $p \leq 0.05(*)$ ,  $\leq 0.005(**)$ ,  $\leq 0.0005 (***)$ ).